POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Transition to an Intelligent Data Platform: Optimizing Energy Forecasting from Excel to Snowflake

Tesi di Laurea Magistrale in
Management Engineering - Ingegneria Gestionale

Author: **Nicolas Ojeda Franco**

Student ID: 10970476
Advisor: Prof. Davide Chiaroni
Co-advisors: Dr. Luca Pedicone, Dr Eugenio Mazza
Academic Year: 2025-2026

Abstract

The global energy transition is profoundly reshaping the energy sector, introducing an unprecedented level of complexity. The rapid growth of non-programmable renewable energy sources, the electrification of items such as electric stoves, heat pumps, e-mobility and the adoption of advanced energy-efficiency solutions are making the balance between supply and demand increasingly dynamic and difficult to predict.

In this evolving context, traditional forecasting methods based on linear models and aggregated historical data often prove insufficient for capturing the non-linear patterns and volatility that now characterize the energy market, especially over short- and medium-term horizons. For A2A, operating in Italy's energy ecosystem, the ability to forecast energy demand with high precision is a strategic capability that leads to competitiveness, cost efficiency, and grid stability.

This master's thesis, developed in collaboration between the Politecnico di Milano and the Data Management Area of A2A, aims to implement the transition from traditional Excel-based workflows to a modern, scalable data infrastructure (Snowflake) capable of supporting advanced analytics and continuous model improvement. The project also provides a literature review how advanced Machine Learning (ML) and Deep Learning (DL) methodologies can be applied to improve forecasting accuracy.

Keywords: Energy Data Modernization; Cloud-Native Data Platforms; Forecasting Digitalization; Snowflake; Smart Infrastructure; Data Governance; Advanced Analytics; A2A; Operational Scalability; Forecasting Reliability.

Abstract in lingua italiana

La transizione energetica globale sta profondamente trasformando il settore energetico, introducendo un livello di complessità senza precedenti. La rapida crescita delle fonti di energia rinnovabile non programmabili, l'elettrificazione di apparecchiature quali fornelli elettrici, pompe di calore e mobilità elettrica, nonché l'adozione di soluzioni avanzate di efficienza energetica stanno rendendo l'equilibrio tra domanda e offerta sempre più dinamico e difficile da prevedere.

In questo contesto in evoluzione, i metodi di previsione tradizionali basati su modelli lineari e dati storici aggregati spesso si rivelano insufficienti per cogliere la volatilità che ora caratterizzano il mercato energetico, soprattutto su orizzonti di breve e medio termine. Per A2A, che opera nell'ecosistema energetico italiano, la capacità di prevedere con elevata precisione la domanda di energia è una competenza strategica a competitività dei costi e della rete.

Questa tesi di laurea magistrale, sviluppata in collaborazione tra il Politecnico di Milano e l'Area Gestione Dati di A2A, mira a implementare la transizione dai tradizionali flussi di lavoro basati su Excel a un'infrastruttura dati moderna e scalabile (Snowflake) in grado di supportare analisi avanzate e il miglioramento continuo dei modelli. Il progetto fornisce anche una revisione della letteratura di di come le metodologie avanzate di Machine Learning (ML) e Deep Learning (DL) possano essere applicate per migliorare l'accuratezza delle previsioni.

Parole chiave: Modernizzazione dei dati energetici; Piattaforme dati cloud-native; Digitalizzazione dei processi previsionali; Snowflake; Smart Infrastructure; Governance dei dati; Analisi avanzate; A2A; Scalabilità operativa; Affidabilità previsionale.

Contents

Introduction

Background and Motivation

The Northern Italian energy sector faces rapid transformation, driven by widespread electrification and decarbonization mandates. Emerging load patterns introduced by heat pumps, residential electrification (e.g., electric stoves), e-mobility infrastructure, and mandated energy-efficiency solutions create a highly dynamic, non-linear, and complex dataset for energy consumption [22] . Forecasting accurate electric load is crucial for ensuring power system security and reliability. Failure to achieve accurate predictions results in increased operating costs; conversely, improved forecasting accuracy translates directly into significant savings and effective decision-making [23].

However, A2A's current forecasting processes rely mainly on Excel based systems that present several limitations. These include fragmented data sources, manual data handling, and limited scalability, which make it difficult to manage large and rapidly growing datasets from smart meters, IoT devices, and external sources. Moreover, Excel lacks real-time processing capabilities and robust version control, hindering data consistency, traceability, and collaboration between business units.

This thesis examines the transition from traditional Excel-based workflows to a scalable, cloud-native data platform (Snowflake), establishing the foundation for data-driven forecasting and analytics. Beyond the transition, this work also investigates how the adoption of advanced Machine Learning (ML) and Deep Learning (DL) methodologies, specifically tailored to handle the unique volatility and complexity of these new loads, can significantly enhance A2A's operational efficiency, strategic infrastructure planning, and compliance with the objectives of developing smart cities [18].

1 | Literature Review

1.1. Literature Retrieval and Selection Methodology

To ensure methodological rigor and transparency in the literature review process, this study adopted a systematic search strategy based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework. The objective was to identify, screen, and select relevant academic contributions aligned with the thesis focus on data modernization in the energy sector and the transition from legacy Excel-based systems to cloud-native platforms.

1.1.1. Database Selection

The primary database used for literature retrieval was **Scopus**, selected due to its broad coverage of peer-reviewed journals, conference proceedings, and technical publications in the fields of energy systems, digital transformation, data management, and cloud computing. Scopus ensures high-quality indexing and allows structured Boolean search queries across titles, abstracts, and keywords.

1.1.2. Search Strategy

The literature search was divided into two main thematic clusters:

Forecasting and Data Modernization in the Energy Sector

The first search cluster focused on the intersection between energy forecasting, data management modernization, and digital transformation within utilities. The following query was applied:

```

ALL (
    ( "data platform" OR "cloud platform" OR "cloud-based" OR "cloud-native"
      OR "data warehouse" OR "data lake" OR "data modernization"
```

```
      OR "data management" OR "data governance" OR "Snowflake"
      OR "data migration" OR "ETL" OR "ELT" OR "digital transformation"
      OR "data integration" OR "data architecture" )
    AND
    ( "energy sector" OR "energy industry" OR "utility"
      OR "power system" OR "distribution network"
      OR "energy forecasting" OR "load forecasting"
      OR "energy data" OR "grid data" )
    AND
    ( "legacy systems" OR "Excel" OR "spreadsheet-based"
      OR "manual workflow" OR "data silos"
      OR "unstructured data" OR "operational inefficiency" )
    AND
    ( "scalability" OR "automation" OR "data quality"
      OR "data reliability" OR "real-time data"
      OR "interoperability" OR "centralized data"
      OR "collaboration" OR "standardization" )
    AND
    ( "advanced analytics" OR "analytics enablement"
      OR "machine learning" OR "data-driven"
      OR "predictive analysis" OR "AI readiness" )
    AND
    ( "energy transition" OR "smart city"
      OR "sustainable cities" OR "urban energy systems"
      OR "digital energy" OR "smart infrastructure"
      OR "energy planning" OR "climate neutrality" )
    AND
    ( "Europe" OR "European Union" OR "Italy"
      OR "Italian" OR "Milan" OR "urban"
      OR "metropolitan" OR "city-scale" )
)
```

Listing 1.1: Search Query 1: Forecasting and Data Modernization in the Energy Sector (Scopus)

This query was designed to capture studies addressing the modernization of data systems in energy utilities, with particular attention to European and urban contexts. Resulting of 165 papers found in the database.

Transition to Cloud Platforms in the Energy Industry

The second search cluster focused specifically on cloud migration strategies and the replacement of legacy systems in the energy domain. The following query was used:

```
TITLE-ABS-KEY (
    ( "cloud data platform" OR "cloud computing"
      OR "data warehouse" OR "data lake"
      OR "cloud-based system" OR "data modernization"
      OR "digital transformation" OR "data governance"
      OR "data integration" OR "ETL" OR "ELT"
      OR "data pipeline" )
    AND
    ( "energy sector" OR "energy industry"
      OR "utility" OR "power system"
      OR "electricity" OR "grid data"
      OR "smart grid" OR "energy management" )
    AND
    ( "legacy system" OR "spreadsheet"
      OR "Excel" OR "manual process"
      OR "data silo" )
    AND
    ( "scalability" OR "automation"
      OR "data quality" OR "data reliability"
      OR "centralized data"
      OR "operational efficiency" )
)
```

Listing 1.2: Search Query 2: Transition to Cloud Data Platforms in the Energy Industry (Scopus)

This search aimed to identify studies discussing structural digital transformation processes, data governance models, and scalable cloud architectures in the utility sector. Resulting of 25 papers found in the database.

Screening and Selection Process

The retrieved records were exported from Scopus and imported into Rayyan, a web-based systematic review tool that facilitates blinded screening and collaborative selection. Titles and abstracts were independently screened according to predefined inclusion and exclusion criteria:

- Inclusion: Peer-reviewed articles addressing data platform modernization, cloud adoption, or forecasting systems within the energy or utility sector.
- Exclusion: Studies unrelated to energy systems, purely technical IT infrastructure papers without energy application, and non-peer-reviewed documents.

Full-text screening was subsequently conducted to ensure alignment with the thesis objectives. The PRISMA methodology was followed to document the number of identified, screened, excluded, and included studies, ensuring reproducibility and transparency of the review process.

This structured approach guarantees that the literature review is not only comprehensive but also methodologically sound, supporting the thesis objective of analyzing the transition from legacy spreadsheet-based forecasting systems to a scalable, cloud-native data platform within A2A. After the filtering by abstract, 26 papers were academically sound with the current research and therefore included in the current discussion.

1.2. Energy data management in smart utilities

Smart utilities operate energy infrastructures that are increasingly digital, interconnected and data-intensive. Across generation, transmission, distribution and end-use, utilities now collect vast amounts of data from SCADA systems, smart meters, IoT sensors, asset-management platforms, GIS, market systems and customer channels. Traditional, system-specific databases were not designed for this level of volume, variety and velocity, which leads to data silos, duplicated information and inconsistent semantics that limit the value that can be extracted from data for planning, operations and forecasting [10][9].

1.2.1. From legacy silos to integrated data architectures

The literature on power-system digitalization consistently highlights data silos and information asymmetry as core barriers to intelligent grid operation. In modern power systems, data spans real-time operational measurements, equipment status, historical logs and forecast series across all stages of the value chain. Conventional relational databases and application-centric data stores struggle with such heterogeneity and complex relationships, which “severely restrict the efficiency and decision-making capabilities of power systems” [10]

On the power-system side, digital transformation is confronted with massive, diverse and heterogeneous data generated by numerous business systems across generation, transmission, distribution and retail. Traditional database-centric approaches struggle with this growth, leading to data silos, information asymmetry and limited decision-making capability.

Several strands of recent work focus specifically on modernizing data management for power systems:

- **Knowledge-graph–based data asset management.** Knowledge graphs are proposed as a way to integrate structured, semi-structured and unstructured power-system data into a unified semantic model. By recognising entities (assets, measurements, locations) and relationships, performing knowledge fusion across heterogeneous systems, and enabling reasoning on top of this graph, utilities can move from fragmented databases to a global, queryable view of their data assets. This supports standardized data management, improves data quality, and provides intelligent decision support—for example for asset management, planning or fault analysis [10].

- **Middleware for heterogeneous data integration.** In parallel, middleware frameworks are being developed to integrate relational and NoSQL databases, file systems and APIs into unified analysis platforms. A recent Golang-based middleware for power-grid data, built around a rule-driven parent-table structure and ORM tooling, can ingest and transform tens of thousands of rows from external systems into a consistent target schema with zero conversion errors and acceptable latency. Such middleware directly targets the grid's digital-transformation bottleneck: connecting legacy systems and new applications without bespoke point-to-point integrations [10]. [14].

These approaches illustrate that "data modernization" in the energy sector is not only about moving data to the cloud, but about building semantic, scalable and interoperable data layers that can support near-real-time analytics, cross-functional workflows and new digital services.

1.3. Cloud data platforms for energy systems: architecture, integration, and governance

Cloud-native data platforms have become a central pillar of digital transformation in power and utility companies. They provide a scalable environment to integrate operational technology (OT) and information technology (IT) data, expose it via standardized interfaces, and enforce governance rules so that analytics and advanced applications can be developed on top in a sustainable way. Recent work in industrial data platforms, DataOps, middleware and knowledge-graph–based data management offers a set of architectural patterns that are directly relevant for smart utilities.

1.3.1. Platform architecture and “foundational digital twins”

Large industrial players have begun to replace fragmented, asset-centric databases with cloud-hosted industrial data platforms that collect time-series, events and master data from hundreds of thousands of sensors and control systems into a single, horizontally scalable store. Aker BP implemented an “industrial data platform” that ingests more than 800,000 measurements per second from over 200,000 data sources into a cloud environment, making this data accessible through standardized APIs to applications, analytics and digital-twin solutions. The platform constructs a “foundational digital twin” that organizes raw signals around equipment, tags and topology, decoupling applications from underlying source systems and enabling re-use of curated data across multiple use cases [14].

Mukherjee and Das with the concept of “Industrial Data Operations” (DataOps) as an architectural and organizational framework for managing data across the value chain. They propose a three-layered architecture comprising a Foundation layer (for ingestion, storage and semantic modeling), an Enablement layer (for data quality services, lineage, catalogs and security), and an Insights layer (for self-service analytics, AI and applications). Their case study on a global energy company shows how a cloud-hosted platform built on these principles can serve as a shared backbone for multiple business units while remaining compliant with local regulations and performance constraints. [17].

These architectures are aligned with the principles of cloud data platforms such as Snowflake or lakehouse solutions in the utility domain: separation of storage and compute, elastic scaling for spiky analytical workloads, and a strong semantic layer (via digital twins or data models) that abstracts away the complexity of underlying SCADA, historian, market and asset-management systems. For electricity utilities, adopting similar patterns allows load-forecasting, asset-health analytics and customer-centric applications to operate on a single, trusted representation of network and demand data instead of bespoke extracts.

1.3.2. Semantic layers and knowledge-graph–based data asset management

Beyond basic schema modeling, recent work shows that utilities can use knowledge graphs as a semantic backbone for their data platforms. An optimization method for power-system data asset management based on knowledge graphs that integrate structured, semi-structured and unstructured data (equipment status, operational data, sensor readings, documentation) into a unified graph representation [10][9]. This framework combines entity recognition, knowledge fusion and entity alignment to reconcile heterogeneous sources into a cohesive graph, which is then stored and queried using big-data technologies such as HDFS, NoSQL and parallel processing engines to support full lifecycle data management.

A key insight from this work is that the knowledge graph does not merely serve as a passive catalog, but actively supports intelligent search, data audit, and lifecycle tracking (access, flow, conversion and disposal of data resources). This enables utilities to implement granular governance policies, who can access which measurement in which context, while at the same time exposing a flexible semantic layer for advanced analytics and decision support. To support the theory, a graph model for the "Open Power System Data" platform, illustrate how technical, financial, geographic and operational data can be linked through a unified Common Information Model (CIM)-style schema to support complex queries on generation capacity, plants, time series and licensing constraints.

For a cloud data platform such as Snowflake, this suggests a reference design in which the physical storage and compute layer is complemented by an explicit semantic layer (implemented via views, data products or an external knowledge graph) that encodes relationships between meters, customers, grid assets, time-series channels and market entities.

1.3.3. Governance, quality and DataOps

A recurring theme across some literature is that architecture and integration alone are insufficient, governance and operational practices are equally important. In the Industrial Data Operations framework, governance is implemented through a data catalog, quality services, lineage tracking and standardized APIs that make data discoverable and trustworthy for analytics teams [17]. The authors emphasize that DataOps is not merely a tooling choice but an organizational capability: cross-functional teams own data products, monitor SLAs and continuously improve pipelines in response to business needs.

Knowledge-graph–based approaches embed governance directly in the data model: they support data-audit capabilities (real-time comparison and consistency checks) and lifecycle tracking across generation, storage, use and disposal, which is essential for complying with regulatory requirements and managing the risks associated with incorrect grid or customer data [10]. Middleware-based integration frameworks further highlight the need for security and privacy controls during data migration and sharing, especially when integrating external sources into a unified platform.

1.4. Forecasting Horizons and Strategic Relevance

Forecasting methodologies must be rigorously matched to the required prediction horizon, distinguishing between real-time operational needs and long-term strategic investments.

Table 1.1: Classification of forecasting horizons and their strategic role for A2A.

Horizon	Typical Duration	Strategic Focus	Relevance for A2A
Very Short-Term Load Forecasting (VSTLF)	Seconds to a few hours	Real-time operational planning and grid management	Supports control flow and cost optimization [16].
Short-term (STLF)	Hours $\rightarrow$ 1 week	Operational dispatching, microgrid balancing, market bidding	Represents the most important forecasting horizon in many studies. Essential for daily scheduling of production and effective demand response programs for flexible loads like EV charging. [13].
Medium-term (MTLF)	1 week $\rightarrow$ 1 year	Asset maintenance, tariff setting, storage scheduling	Required for planning renewable energy sources and maintaining the power grid infrastructure, often needing an hourly frequency [26]
Long-term (LTLF)	$>$ 1 year	Capacity expansion, electrification scenario planning, regulatory strategy	Aligns with decarbonization, electrification, and CAPEX decisions [26]

For A2A's operations, highly accurate STLF is necessary to manage the minute-to-minute volatility introduced by electric loads like heat pumps and e-mobility. Short-term ML/DL forecasting can reduce imbalance penalties by 10–15 % [11][7]. LTLF requires integrating macroeconomic factors (like GDP and population growth, relevant to the Italian context to guide strategic investment and capacity planning under the Fit-for-55 framework. [4].

- Operational: Accurate hourly load forecasts support real-time balancing, EV-charging management, and district heating optimization.

- Strategic: Long-term predictive analytics underpin heat-pump adoption scenarios, distributed PV penetration, and network digitalization planning.

1.5. Forecasting Methods and Classification

The complexity of energy consumption in modern Northern Italian cities necessitates a pivot from traditional statistical tools toward hybrid and deep learning models to capture highly **non-linear and complex patterns** [26].

Conventional/Statistical Models: These models, such as Time Series (e.g., ARIMA) and Regression (LR/MLR), are generally valued for their simplicity, interpretability, and ease of use [23][15] [1]. They are typically preferred for yearly consumption forecasting at the national level. LR models, using GDP and population, have historically been developed for **Italian electricity consumption forecasting** up to 2030 [4].

Machine Learning (ML/AI) Models: These methods, including Artificial Neural Networks (ANN), Support Vector Regression (SVR), and Random Forest (RF), are increasingly used (54% usage in one review) due to their enhanced ability to handle complex, **non-linear factors** like weather and indoor conditions better than conventional approaches [23].

• **Deep Learning (DL) Models:** DL, including Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, utilizes neural network technology to process high-dimensional data and capture complex non-linear relationships, improving stability and accuracy [6] [18]. DL is specifically applied for **forecasting residential load** and power consumption in buildings [2].

Some studies show that multivariate CNNs outperform LSTMs in monthly peak-load prediction, cutting RMSE by up to 12 % [11]. Also some analyses show that CNNs achieve comparable accuracy to ANN and superior to SVM at building level [3].

Table 1.2: Overview of statistical, ML and DL forecasting approaches.

Category	Typical Methods	Key Characteristics	Executive Takeaway
Statistical / Econometric	Linear Regression, ARIMA, SARIMA	Transparent, interpretable, less robust to non-linearities	Baseline models; suitable for regulatory reporting
Machine Learning (ML)	Random Forest, Gradient Boosting (LightGBM), Support Vector Machines	Capture non-linear dependencies; scalable; moderate interpretability	Best suited for medium-term demand and price forecasting
Deep Learning (DL)	CNN, LSTM, CNN–LSTM hybrids, Temporal Convolutional Networks (TCN)	Learn spatio-temporal patterns automatically; handle IoT and smart-meter data	High accuracy in volatile contexts such as EV and heat-pump loads

The current research direction favors **hybrid models**. Hybrid approaches, which combine two or more methodologies (e.g., CNN-LSTM), are the preferred technique due to their sustained accuracy, enhanced flexibility, and precision [23] [13].

- **DL Hybrid Superiority:** Hybrid models that combine CNN (for feature extraction) and LSTM (for modeling time information and trends) are studied as effective methods for overcoming the complexity of electricity demand forecasting. Studies show hybrid DL models generally yield better performance results than individual models [23].
- **E-mobility and Smart Cities:** DL is revolutionizing energy optimization in smart cities. AI/ML algorithms are employed to manage energy systems, including the optimization of resource allocation for **renewable energy**. Forecasting methods (like DL neural networks) are proposed to incorporate supply, demand, weather, and date information [5][18].

1.6. Key Input Variables and Data Integration

Across reviewed literature, the following drivers consistently influence forecasting performance:

Table 1.3: Key Input Variables and Their Application for A2A

Category	Key Variables and Relevance	Context and Emerging Trends
Historic Electrification & Residential Loads	Residential consumption, individual load profiles, historical load data with behavioral analysis [15]	Forecasting individual household consumption is an extremely challenging task due to high volatility. Load forecasting must account for factors like space heating, space cooling, water heating, and cooking, all directly affected by heat pumps and electric stoves [19].
Weather	Temperature, humidity, solar radiation, wind speed [2]	Some studies show that temperature alone explains **30–50 %** of city-scale electricity variability, with each +1°C leading to **5 % higher demand** [25]. Weather data must be incorporated into ML/DL models alongside historical demand.
Socio-economic	GDP, population, CO_2 emissions, and share of renewable energy (SRE) [22]	For long-term forecasting in the EU context, SRE and CO_2 emissions are identified as dominant predictors of fossil fuel reliance [22]. LTLF models for Italy have traditionally used GDP, GDP per capita, and population [4].
Technology: E-mobility and Grid Data	Number of EV chargers, PV installations, building retrofits, previous peak demand, load curves [8, 11]	The increase of EVs as consumer loads needs accurate forecasts for electric power transmission networks [11]. Smart-meter and sensor data can feed real-time ML pipelines, while cloud-based data lakes allow decentralized learning for local networks in Lombardy and Brescia [5, 8].

1.7. Performance Evaluation and Benchmarking

Some insights of traditional models to assess their accuracy needs to be rethink when including new characteristics of ML or Dl to the model. However, standardst ragrding some limits are widely spread and represent a starting point to analyse possible models:

- Statistical models (ARIMA) → RMSE 10 %
- Gradient Boosting / Random Forest → RMSE 6–8 %
- CNN-LSTM or TCN models → RMSE 3–5 %, particularly under high volatility and large datasets.

Table 1.4: Common metrics used to evaluate forecasting model accuracy and their executive-level interpretation.

Metric	Description	Executive Interpretation
MAPE (Mean Absolute Percentage Error)	The most popular metric (51.72% usage), provides a scale-independent measure of accuracy [1, 13].	5% = excellent forecast for operational use
RMSE (Root Mean Square Error) & MAE (Mean Absolute Error)	RMSE (18.97% usage) and MAE (17.24% usage) are commonly used to measure prediction error magnitude [13].	Used for model comparison and stability
R^2	Measures how well a model fits the observed data, indicating the proportion of variance explained by the model.	Values $>$ 0.95 indicate reliable model fit

1.8. Contexts and Challenges

A2A's mandate to support intelligent energy solutions in Italian smart cities is highly relevant. However, this advanced forecasting strategy is subject to specific technical and institutional hurdles.

1. **Non-Linearity and Complexity:** Accurate forecasting is inherently challenging due to the dynamic, non-linear, and complex nature of energy consumption data. Peak demand, critical for grid stability, is often very nonlinear [11].
2. **Data Scarcity and Quality:** Developing robust DL models is limited by the lack of sufficient high-quality, labeled data. The challenge of working with big data—specifically obtaining and organizing hourly meteorological and electricity demand data for a city—is significant [18].

3. **The Black Box Problem:** EML/AI and hybrid models are often considered "black boxes," making it challenging for regulators and operators to trust the model's outputs or understand the underlying decision-making processes [24].

4. **Regulatory Compliance:** Compliance with EU AI Act and transparency in automated forecasting pipelines [18].

5. **Integration Difficulties:**Integrating accurate demand forecasts with grid management systems requires technical challenges in communication protocols, data exchange, and system coordination [23].

2 | Context and Motivation

2.1. Context: digital transformation and data modernization in the energy sector

The energy sector is undergoing a profound shift characterized by the exponential growth of data across the value chain—from generation and transmission to distribution and consumption. For multi-utility companies like A2A, the complexity of modern power systems requires moving beyond traditional data management to ensure operational efficiency and strategic agility [10].

2.1.1. The Legacy Bottleneck: Data Silos and Manual Workflows

Despite the abundance of data generated by smart grids and IoT devices, the industry struggles with "data usefulness" due to fragmented infrastructure. The reliance on legacy systems, specifically spreadsheet-based workflows (e.g., Excel) and disparate operational databases, creates significant barriers to innovation [14]:

- **Data Silos and Information Asymmetry:** Traditional methods struggle to handle complex data relationships, resulting in silos where information remains locked within specific departments. This fragmentation restricts the decision-making capabilities of power systems and prevents a holistic view of operations [10][9].
- **Operational Inefficiency:** The reliance on manual data handling is costly. Industry surveys indicate that engineers and data scientists spend over 50% of their time identifying and formatting data rather than performing value-added analysis [17].
- **Lack of Standardization:** Legacy applications often utilize non-standardized, subjective approaches to data capture, leading to erroneous inputs and compromising the reliability of downstream analytics [12].

2.1.2. Implications for utilities and this thesis

For utilities navigating the transition to net-zero, digital transformation and data modernization have become strategic capabilities. European decarbonisation scenarios emphasise deep electrification, efficiency and renewables, all of which increase the need for granular forecasting. To respond, leading energy companies are:

- Consolidating operational and enterprise data into cloud-based industrial data platforms.
- Using DataOps and middleware to break down data silos and expose governed, high-quality data products.[17][14].
- Introducing semantic layers like knowledge graphs to standardise definitions across business domains [10].
- Building AI and advanced-analytics use cases such as load forecasting requiere worfflow digitalization [12].

Within this broader landscape, the transition to Snowflake can be seen as a key enabler for advanced electricity-demand forecasting explored in the rest of the document. It provides the scalable, interoperable and governed data foundation required to move from isolated pilots to production grade machine-learning workflows that support both operational excellence and the energy-transition objectives of the utility.

2.2. The strategic role of cloud-native platforms (Snowflake)

Traditional data-warehouse architectures used in many utilities suffer from well-known limitations: coupled storage and compute, rigid capacity planning and difficulty supporting heterogeneous workloads (BI, data science, ad-hoc analysis) without performance conflicts. Snowflake's architecture addresses these points through a three layer design:

- a **storage layer** that holds compressed, columnar data in cloud object storage;
- one or more **compute clusters ("virtual warehouses")** that can be started, resized or stopped independently;
- and a **cloud services layer** that manages metadata, security, transaction management and query optimization. [20]

Because compute is decoupled from storage and from other compute clusters, multiple teams (e.g. market analytics, network planning, customer analytics) can query the same data simultaneously using isolated warehouses, avoiding resource contention and "noisy neighbor" problems. Multi-cluster warehouses and automatic scaling further help to absorb peak workloads in large-scale scenario runs, without over-provisioning permanent capacity. [21]

For an energy utility, this means that operational reporting, regulatory dashboards, forecasting pipelines and experimental data-science workloads can all run on a single data platform, while still being isolated in terms of performance and cost accounting.

2.3. Research Objectives and Questions

This thesis is positioned at the intersection of data modernization and advanced analytics in the energy sector. Its primary focus is the transition from legacy, siloed data architectures to a cloud-native platform (Snowflake) in a multi-utility context, using ML and AI based electricity demand forecasting as a reference use case.

Specific objectives and research questions

Objective 1 – Diagnose the current state and requirements

To analyse the current data and analytics landscape of the case-study utility, with a focus on electricity-demand forecasting needs and constraints.

- **RQ1.1** What are the main technical, organizational and data-quality limitations of the existing data architecture with respect to electricity-demand forecasting and related analytics?
- **RQ1.2** Which functional, data, governance and performance requirements must a modern data platform satisfy to support forecasting, scenario analysis and regulatory reporting in this context?

Objective 2 – Design a Snowflake based architecture and operating model

To design a Snowflake-centric data architecture and operating model that addresses the identified requirements.

- **RQ2.1** How should data from heterogeneous sources (metering, network operations, customer systems, external data such as weather and markets) be modelled, ingested and governed within Snowflake to serve forecasting and planning use cases?
- **RQ2.2** Which DataOps and MLOps patterns (e.g. pipelines, environments, roles, governance mechanisms) are most suitable for operating forecasting workflows on

top of Snowflake in a sustainable way?

Objective 3 – Evaluate Snowflake's suitability for energy-demand analytics

To evaluate Snowflake's ability to support the utility's analytical needs, using representative non-ML workloads (e.g. descriptive analytics, aggregations, historical analyses) as proxies for more advanced future use cases.

- **RQ3.1** To what extent does a Snowflake-based architecture improve data accessibility, reuse, scalability, maintainability and reproducibility compared to the legacy setup, when applied to representative electricity-demand analytics?
- **RQ3.2** What are the main benefits and limitations observed in terms of performance, cost management, governance and user experience when operating these workloads on Snowflake?

3 | Methodology and Framework Design

This chapter explains how the research is structured to answer the questions defined in the introduction. The work combines an explorative **single-case study** of a multi-utility company A2A to propose and assess a Snowflake-based data platform for energy-demand analytics.

The methodology has four main components:

1. Research approach and case-study design
2. Data collection and analysis
3. Design framework for the target architecture
4. Evaluation framework and criteria

3.1. Research approach and case-study design

The thesis adopts a **qualitative, exploratory case-study** design. The case is A2A, a Italian multi-utility company that is in the process of modernizing its data landscape and has selected Snowflake as a strategic cloud platform. The case-study approach is appropriate because the phenomenon of data modernization for energy-demand analytics is complex, context-dependent, and tightly coupled to organizational processes; and rich empirical insights can be obtained from existing systems, internal documentation, and expert knowledge within the company.

Within this case-study setting, the work follows a design logic: starting from an observed problem (limitations of the current data architecture), it requirements, proposes a Snowflake-based solution, implements a limited proof of concept, and evaluates the design against explicit criteria.

3.2. Data collection and sources

Data for the study is drawn from multiple sources to ensure triangulation:

- **Technical documentation**: existing architecture diagrams, interface specifications, database schemas, ETL descriptions and internal reports on data flows related to electricity-demand analytics.
- **Operational artefacts**: sample datasets from metering, network, customer and external sources (e.g. weather), existing SQL scripts and dashboards that are currently used for reporting and analysis.
- **Informal interviews and expert input**: discussions with data engineers and business stakeholders involved in the current systems and in the Snowflake adoption. These inputs are used primarily to clarify requirements, constraints and implicit assumptions.
- **Tool and platform documentation**: vendor and internal documentation for Snowflake and related tools, used to understand available features and constraints relevant to the proposed architecture.

The analysis proceeds iteratively: initial readings of documentation and artefacts are used to map the **"as-is"** landscape; this mapping is validated and refined through expert input; and the refined understanding then feeds into the requirement definition and design of the **"to-be"** architecture.

3.3. Design framework for the target architecture

The design work is structured in three phases, each aligned with one or more research questions.

3.3.1. Phase 1 – Current-state assessment and requirements

- **System mapping**: identification of source systems (metering data, network operations, customer information, external datasets), their data models, frequencies and existing integration mechanisms.
- **Pain-point analysis**: documentation of technical and organisational limitations that affect electricity-demand analytics (e.g. data silos, duplicated pipelines, latency, lack of governance).

- **Obtaining requirements**: derivation of functional and non-functional requirements for a modern data platform, grouped into:
  - data and integration requirements;
  - governance and security requirements;
  - performance and scalability requirements;
  - analytical and usability requirements (what the platform must enable for analysts and data scientists).

Construction of the requirements catalogue Phase 1 concludes with the construction of a requirements catalogue, which serves as the design brief for the Snowflake-based architecture. The aim of the catalogue is to translate the current-state analysis (systems, data flows, pain points) into a structured set of requirements that can guide design decisions in a traceable way. Requirements are derived from technical documentation, operational artefacts and expert input, and are consolidated into a single artefact instead of being scattered across notes or diagrams.

For each requirement, the catalogue records: (i) a unique identifier, (ii) a short description, (iii) the primary source (document, interview, observation), (iv) a qualitative priority level, and (v) the related research question(s). An illustrative template is shown in the table 3.1; the fully populated catalogue for the case study is presented in full detail in the Appendix A.

The four requirements define the evaluation framework used throughout the remainder of this thesis. Together, they translate the observed limitations of the current data landscape into a coherent and traceable set of design criteria that guide the proposed Snowflake-based architecture. The catalogue captures what the data platform must enable in order to support reliable electricity-demand analytics and future advanced use cases. In subsequent chapters, these requirements are used both as a design reference for the target architecture and as qualitative evaluation dimensions to assess the extent to which the proposed solution addresses the needs of a smart-utility environment.

Table 3.1: Data and integration requirements catalogue

ID	Category	Description	Source	Priority	Related RQ
DI01	Data & integration	Energy consumption data (electricity, gas, distric heating) must be available in a single, harmonized schema across all points of delivery, independent of the originating source system.	System documentation; Source database SIDES	High	RQ1.1, RQ1.2
GS01	Governance & security	Access to energy consumption data must be controlled through role-based access mechanisms.	Internal access practices	High	RQ1.2
PS01	Performance & scalability	The platform must support interactive analytical queries over large volumes of historical energy-demand data without requiring manual aggregation.	Performance issues in current reports	High	RQ1.1, RQ3.1
AU01	Analytics & usability	Analysts must be able to access curated energy-demand datasets using SQL or Python queries languages without requiring changes to ingestion or transformation pipelines.	Analyst feedback; current workflow analysis	High	RQ3.1

3.3.2. Phase 2 – Snowflake-based architecture and operating model

The objective of Phase 2 is to formalize a Snowflake-based analytical architecture and operating model that is fully aligned with the existing enterprise data platform and architectural guidelines already adopted by A2A. This phase focuses on structuring, documenting, and rationalizing how the current Snowflake environment supports energy demand analysis, KPI computation, and scenario modelling, ensuring consistency with corporate standards, data governance practices, and business-intelligence workflows.

This phase therefore answers how an existing cloud-native data platform can be leveraged for energy-demand analytics.

Architectural principles and constraints The architecture described in this phase adheres to a set of principles derived from the current operational setup. These include a layered data model (L0–L1–L2) separating raw ingestion, historised integration, and business-oriented analytical entities; a clear separation between transactional systems and analytical workloads; read-only governance for official analytical layers; direct integration with enterprise BI tools (primarily Qlik Sense); and explicit support for manual business inputs as first-class, historised data sources.

Data platform architecture overview The Snowflake environment operates as a **data lakehouse**, consolidating heterogeneous energy-related data sources into a unified analytical platform. Source systems include operational and enterprise applications such as SIDES (energy consumption), CURIT (heating systems), CENED (energy certification), SAP, GIS, and domain-specific manual input interfaces.

Data flows are structured across three logical layers:

- **L0 – Landing layer**, which captures raw data ingested from source systems or APIs with minimal transformation.
- **L1 – Integration layer**, where data is historized, validated, and aligned to source-system semantics, enabling time-consistent analysis.
- **L2 – Business layer**, where domain-specific entities and KPIs are published as analytical facts and dimensions, explicitly designed to support reporting and modelling use cases.

This layered structure enables scalability, traceability, and reuse across analytical processes. The L2 layer constitutes the core analytical interface for energy-demand analysis. It exposes business-oriented fact tables that integrate multiple upstream sources and encapsulate domain logic defined through functional analyses. These entities support consolidated energy-consumption KPIs, installed-capacity and power indicators, energy-certification metrics, and aggregated consumption marts optimised for reporting and exploratory analysis.

L2 entities are designed to expose last year data, while retaining historical traceability when required.

Operating and access model Analytical outputs are produced through scheduled and manual execution of domain-specific procedures that implement deterministic business logic for KPI computation and scenario aggregation. The separation between intermediate calculations and final outputs supports reproducibility and controlled updates.

Access to the platform follows a role-based model aligned with analytical responsibilities. Viewer users consume certified datasets exclusively through BI tools; Analyst users combine BI self-service with read-only access to L2 data; and Pro users access raw and intermediate layers within controlled sandbox schemas for advanced analysis. This model balances governance requirements with analytical autonomy.

Positioning of advanced analytics Although the architecture is technically capable of supporting advanced analytics and machine-learning workloads, such capabilities are explicitly positioned as future extensions. Within the scope of this thesis, Snowflake serves as a stable foundation for data integration, KPI computation, and scenario analysis, upon which more advanced modelling approaches may be introduced without fundamental architectural changes

3.3.3. Phase 3 – Proof of concept and operationalization

In Phase 3, a limited proof of concept (PoC) is used to operationalize the framework defined in Phase 2 and to verify its applicability within the existing Snowflake environment. The PoC aim to to demonstrate that the proposed architectural and operational principles can be instantiated using real data and workflows.

Specifically, the PoC focuses on:

- the selection of a representative subset of data sources (e.g. energy consumption in Milano and technology-related inputs) relevant for energy-demand analytics
- the application of ingestion, transformation, and layering mechanisms (L0–L1–L2) to these datasets
- the execution of a small number of analytical queries and simple demand-oriented indicators to assess usability, consistency, and performance under realistic conditions

3.4. Evaluation framework and criteria

To assess the proposed Snowflake-based design and the proof of concept, the thesis adopts a structured evaluation framework organised into four dimensions. These dimensions directly correspond to the requirement categories defined in Phase 1 and ensure coherence between design, implementation, and validation.

1. **Data & integration**
    - The extent to which relevant data sources (e.g. consumption, technology mix,

certification data, manual inputs) are available in Snowflake in a consistent, harmonised, and well-documented form.

- The degree of reduction in data silos and manual data preparation compared to the legacy situation.
- The presence of historisation mechanisms, incremental ingestion processes, and traceable transformations.

2. **Governance & security**
   - The effectiveness of role-based access control mechanisms (Viewer, Analyst, Pro) and the enforcement of read-only governance on official analytical layers.
   - The presence of data classification, masking, and auditability mechanisms.
   - The clarity of data ownership, stewardship responsibilities, and lifecycle management processes.
3. **Performance & scalability**
   - The ability of the platform to handle expected data volumes and concurrent analytical workloads, assessed conceptually through Snowflake capabilities and empirically through PoC observations.
   - The flexibility to allocate and isolate compute resources across different workloads.
   - The stability of performance as data volumes grow over time.
4. **Analytics & usability**
   - The accessibility of curated datasets for SQL-based and BI-driven analysis without requiring changes to ingestion or transformation pipelines.
   - The reproducibility of analytical results through historised tables, deterministic procedures, and parameterised workflows.
   - The availability of clean, well-structured time-series and contextual data suitable for future forecasting and advanced analytics use cases.

Each dimension is assessed qualitatively by comparing legacy and Snowflake-based conditions, drawing on evidence from the proof of concept, platform documentation, and stakeholder feedback. The evaluation does not aim to produce quantitative performance benchmarks, but to provide a structured and transparent assessment of architectural alignment, operational robustness, and future-readiness.

3.5. Limitations

The chosen methodology has several limitations that are explicitly acknowledged:

- Single-case focus: results are grounded in A2A, one company context and may not generalise directly to all energy companies. The aim is analytical, not statistical, generalisation.
- Scope of implementation: the PoC covers only a subset of data sources and analytical use cases, and does not implement full machine-learning workflows. Conclusions about ML readiness are therefore based on architectural analysis and partial empirical evidence.
- Time and access constraints: the depth of interviews and access to production systems may be limited, which can leave some aspects of the existing landscape or future roadmap under-specified.

4 | Implementation and Proof of Concept

Building on the methodological framework defined in the previous chapter, the following chapter presents the concrete implementation of the proposed approach within the A2A data platform. The implementation instantiates the architectural principles, operating model, and evaluation criteria described in Phases 2 and 3, using a limited proof of concept based on real data sources and workflows.

The implementation focuses on demonstrating how the Snowflake-based framework can be applied in practice, and on providing empirical evidence to support the qualitative evaluation defined in the methodology.

4.1. Instantiation of the L0–L1–L2 framework

This section describes how the L0–L1–L2 framework defined in the Methodology is physically instantiated within the existing Snowflake production environment at A2A. The instantiation relies entirely on pre-existing databases, schemas, naming conventions, and orchestration mechanisms, and **therefore represents a faithful execution of the framework** rather than a redesign.

The framework is implemented through the explicit separation of responsibilities at schema level, with each layer supporting a distinct role in the data lifecycle: ingestion, integration, and analytical consumption.

4.1.1. Physical hierarchy and naming conventions

In Snowflake, all data objects are organised according to the following hierarchy:

```
DATABASE SCHEMA TABLE / VIEW / TASK / PROCEDURE
```

Listing 4.1: Snowflake physical object hierarchy

Within the PoC, the production Data Platform database is:

Database: `PRD_DP`

The methodological layers are instantiated as schemas within this database:

- |**L0**| – landing and raw ingestion
- |**L1**| – historized and integrated data
- |**L2**| – business-oriented analytical entities

As a result, an object such as:

```
PRD_DP.L1.L1_DCU_CATASTO_UNICO_REG_IMPIANTI_TERMICI_CRI_TD
```

Listing 4.2: Example of L1 object reference

must be interpreted as:

- `PRD_DP`: production Data Platform database.
- `L1`: integration and historization schema.
- `L1_DCU_CATASTO_UNICO_REG_IMPIANTI_TERMICI_CRI_TD`: physical table.

This explicit separation ensures that the logical layering defined in the methodology is enforced at physical level, supporting traceability and governance (RQ1).

4.1.2. L0 – Landing and raw ingestion layer

The L0 layer represents the entry point for data into the platform. It is used exclusively for raw ingestion and preserves the structure and semantics of source systems.

In the PoC, L0 is primarily used for external API ingestion (e.g. CENED), and manual business uploads via controlled interfaces.

```
PRD_DP LO LO_CEN_CERTIFICAZIONE_ENERGETICA_EDIFICI_TD
PRD_DP LO LO_INTFC_USR_INPUT_MODELLO_TE_TD
PRD_DP LO LO_INTFC_USR_CONTROL_PANEL_MODELLO_TE_TD
PRD_DP LO LO_INTFC_USR_PROFILI_CARICO_MODELLO_TE_TD
```

Listing 4.3: L0 landing examples in the PoC

Key characteristics of L0 instantiation include:

- no business logic is applied.
- append-only loading.
- explicit load timestamps.

- source-level traceability.

This layer ensures that the original data can always be recovered and reprocessed, directly addressing the data traceability requirement of RQ1.

4.1.3. L1 – Historised integration layer

The L1 layer constitutes the core integration layer of the framework. All datasets entering the PoC are transformed into historized, standardized structures at this level.

Integrated system datasets such as CURIT, is represented in L1 through historized tables:

```
PRD_DP L1 L1_DCU_CATASTO_UNICO_REG_IMPIANTI_TERMICI_CRI_TD
```

Listing 4.4: L1 system-integrated datasets

This table is updated on a monthly cadence and preserves historical snapshots of installed heating systems, technology and fuel classifications, linkage keys for downstream analytics.

Manual inputs are also fully historised at L1 level:

```
PRD_DP L1 L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD
PRD_DP L1 L1_INTFC_USR_CONTROL_PANEL_MODELLO_TE_CPM_TD
PRD_DP L1 L1_INTFC_USR_PROFILI_CARICO_MODELLO_TE_CMT_TD
```

Listing 4.5: L1 historised manual input datasets

Historization at L1 ensures that past scenarios remain reproducible, business assumptions are auditable, analytical outputs can be traced to specific parameter versions.

This design choice is fundamental to RQ1 (integration and reproducibility) and RQ2 (governed analytics).

4.1.4. L2 – Business-oriented analytical layer

The L2 layer exposes certified analytical entities derived from L1 data and aligned with business definitions. It represents the stable consumption interface for reporting and decision support.

Final analytical fact tables The PoC produces two primary L2 fact tables:

```
PRD_DP L2 L2_F_MODELLO_TE_TD_TEM
PRD_DP L2 L2_F_POTENZE_MODELLO_TE_TD_PTE
```

Listing 4.6: Primary L2 analytical entities

These tables consolidate multiple upstream domains, expose certified KPIs for energy consumption and power, and are consumed by Qlik Sense and strategic analyses.

Intermediate L2 entities To support modularity and transparency, several intermediate L2 tables are also maintained:

```
PRD_DP L2 L2_F_CONSUMI_ELE_MODELLO_TE_TD_MTE
PRD_DP L2 L2_F_MODELLO_TE_CONSUMI_GAS_TD_SAG
PRD_DP L2 L2_F_MODELLO_TE_CONSUMI_TRL_TD_TLR
PRD_DP L2 L2_F_POTENZA_IMPIANTI_CURIT_TD_PZA
PRD_DP L2 L2_F_CERTIFICAZIONE_ENERGETICA_EDIFICI_TD_CEN
```

Listing 4.7: Intermediate L2 KPI tables

This layered analytical design supports explainability of KPIs, isolates domain-specific logic, and reduces coupling between transformations.

It confirms the framework's suitability for scalable and maintainable analytics (RQ2).

4.1.5. Layer interaction and traceability

The interaction between L0, L1, and L2 is strictly unidirectional L0 $\rightarrow$ L1 $\rightarrow$ L2.

No reverse dependencies are allowed. This guarantees deterministic transformations, reproducibility of analytical outputs and clear ownership of data at each stage. From a methodological perspective, this instantiation demonstrates that the conceptual framework maps directly to physical Snowflake artefacts, thereby validating RQ1.

4.2. Scope and datasets

The proof of concept (PoC) is scoped as a limited but representative instantiation of the energy-demand analytics domain within the existing Snowflake-based enterprise data platform at A2A. The objective is to validate that the methodological framework defined in Chapter 3 can be concretely instantiated, governed, and executed using real production datasets.

The scope focuses on data domains that are:

- central to energy-demand modelling,
- heterogeneous in origin and structure,
- governed by different ownership and update mechanisms,

- jointly required to compute certified analytical KPIs.

The selected datasets are organised into four main data domains, described below.

4.2.1. Heating systems and installed power domain (CURIT)

The CURIT domain (Catasto Unico Regionale Impianti Termici) provides detailed information on heating systems installed in Lombardy, including technology, fuel type, and associated buildings. Within the Implementation, CURIT data is used to validate cross-domain integration between energy consumption and installed power.

The CURIT dataset is ingested and historised at L1 level, reflecting its role as a structured but externally governed source.

```
DATABASE SCHEMA TABLE

PRD_DP L1 L1_DCU_CATASTO_UNICO_REG_IMPIANTI_TERMICI_CRI_TD
```

Listing 4.8: CURIT – Heating systems and installed power domain

This dataset is updated on a monthly basis and supports historisation of installed technologies, linkage between consumption and infrastructure capacity and computation of power-related KPIs at L2 level.

4.2.2. Building energy certification domain (CENED)

The CENED domain provides energy certification data for buildings in Lombardy, retrieved through APIs exposed by Regione Lombardia. Within the Implementation, CENED data is used as **contextual analytical information** to enrich energy-demand analyses with building-level energy performance indicators.

The domain is instantiated across all three layers, from landing to certified analytical output.

```
DATABASE SCHEMA TABLE / PROCEDURE

PRD_DP L0 L0_CEN_CERTIFICAZIONE_ENERGETICA_EDIFICI_TD
PRD_DP L1 L1_CEN_CERTIFICAZIONE_ENERGETICA_EDIFICI_CEE_TD
PRD_DP L2 MODELLO_TE_L2_F_CERTIFICAZIONE_ENERGETICA_EDIFICI_TD_CEN
```

Listing 4.9: CENED – Building energy certification domain

This domain validates the framework's capability to integrate external API-based sources, enforce historisation and traceability, and expose certified KPIs for analytical consumption (RQ2).

4.2.3. Energy consumption domain (SIDES)

The SIDES domain represents the primary source of energy consumption data for the A2A Group, covering electricity, natural gas, and district heating. It constitutes the **core demand signal** of the PoC and is essential to validate the framework's ability to integrate and harmonise multi-supply consumption data.

The PoC relies on monthly, POD-level consumption data exposed through certified L2 analytical structures.

```
DATABASE SCHEMA TABLE / VIEW

PRD_DP L2 L2_F_CONSUMI_MENSILI_TD_SUM
PRD_DP L2 L2_F_CONSUMI_MENSILI_VQ_SUM
```

Listing 4.10: SIDES – Energy consumption domain (monthly consumption)

These datasets provide harmonized monthly consumption across three supplies, a stable analytical structure (L2) for downstream KPI computation, and certified access for analytical tools and business users (RQ2).

4.2.4. Manual business input domain

In addition to system-generated data, the PoC explicitly includes manual business inputs, which are critical for the execution of the Modello TE. These inputs are provided through controlled user interfaces and historised to ensure governance and reproducibility.

Three complementary manual input domains are included:

```
DATABASE SCHEMA TABLE

PRD_DP L1 L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD
PRD_DP L1 L1_INTFC_USR_CONTROL_PANEL_MODELLO_TE_CPM_TD
PRD_DP L1 L1_INTFC_USR_PROFILI_CARICO_MODELLO_TE_CMT_TD
```

Listing 4.11: Manual business input domains

These datasets contain scenario-specific assumptions and change factors, global control parameters (e.g. conversion factors, constants), hourly load profiles used for peak power calculations (RQ2).

4.3. Implementation of analytical workflows

This section describes the concrete implementation of the end-to-end analytical workflows instantiated within the proof of concept. The objective is to demonstrate how the datasets structured across the L0–L1–L2 layers are operationalized to produce certified analytical outputs in a deterministic, reproducible, and governed manner.

The workflows presented below correspond to the Modello TE domain and are executed using the existing procedural architecture already in production within the A2A Snowflake environment. Figure B.1 illustrates the end-to-end instantiation of the MODELLO_TE analytical workflow, including data ingestion from SIDES, CURIT and CENED, intermediate transformations, and the final KPI computation procedures. Following there is a detailled explanation of each of its items.

4.3.1. Energy consumption analytical workflow

This subsection details the end-to-end analytical workflow used to compute annual energy consumption KPIs for the Modello di Transizione Energetica (Modello TE). The workflow operationalizes the L0–L1–L2 framework by transforming monthly SIDES consumption records into certified annual KPIs for three supplies (electricity, natural gas, district heating) and by integrating historized manual business inputs when required.

Workflow flow and operational trigger

The workflow starts from SIDES consumption data and uses a certified monthly MART structure in L2 as the base analytical input. Annual KPI computation is triggered via a scheduled task, which executes a stored procedure that produces three L2 KPI tables (one per supply).

```

SIDES (external source)
  |
  v
PRD_DP L2 L2_F_CONSUMI_MENSILI_TD_SUM   (MART, daily update; monthly POD
    -level records)
  |
  v
PRD_DP L2 L2_F_CONSUMI_ANNUALI_MODELLO_TE_TK
CRON: 0 11 1 1 * UTC  (1 January, 11:00)
  |
  v
```

```
PRD_DP L2 L2_F_CONSUMI_ANNUALI_MODELLO_TE_SP_EGT()
 |
 +--> PRD_DP L2 L2_F_CONSUMI_ELE_MODELLO_TE_TD_MTE   (electricity KPIs)
 +--> PRD_DP L2 L2_F_MODELLO_TE_CONSUMI_GAS_TD_SAG   (natural gas KPIs)
 +--> PRD_DP L2 L2_F_MODELLO_TE_CONSUMI_TRL_TD_TLR   (district heating
   KPIs)
```

Listing 4.12: Energy consumption workflow (SIDES → annual KPIs)

Two design choices are central for reproducibility:

- **Temporal determinism**: the procedure computes KPIs for the *consuntivated year*, defined as `YEAR(CURRENT_DATE) - 1`, ensuring that the annual run always targets a closed historical period.
- **Certified input**: the procedure reads from `L2_F_CONSUMI_MENSILI_TD_SUM`, which acts as the stable input interface for SIDES-derived consumption records.

Input dataset: SIDES monthly consumption MART

The table `PRD_DP.L2.L2_F_CONSUMI_MENSILI_TD_SUM` stores monthly consumption at POD level for the three supplies managed by the A2A Group (electricity, natural gas, district heating). It is refreshed with a daily cadence while keeping monthly granularity, enabling timely availability of the latest consolidated monthly values.

Within the procedure, the table is used through:

- **Year filtering** (`ID_SUM_ANNO_CONSUMO = YEAR(CURRENT_DATE)-1`),
- **Geographic filtering** (e.g., `DS_SUM_COMUNE ILIKE 'MILANO'`),
- **Usage classification** (residential vs non-residential),
- **Technical segmentation** (e.g., voltage levels for electricity: low/medium/high).

This standardization of filters and classifications ensures that KPI definitions are consistent across executions and consumers.

Orchestration: annual task execution

The annual aggregation is operationalized via the scheduled task:

```
PRD_DP L2 L2_F_CONSUMI_ANNUALI_MODELLO_TE_TK
0 11 1 1 * UTC
```

Listing 4.13: Scheduled task for annual consumption KPI computation

This setup ensures that the KPI computation is executed in a controlled, repeatable manner, aligned with governance and data availability assumptions for consuntivated annual values.

Core transformation logic

The procedure `PRD_DP.L2.L2_F_CONSUMI_ANNUALI_MODELLO_TE_SP_EGT()` implements the deterministic transformation from monthly consumption records to annual KPI outputs. Conceptually, it performs the following steps:

1. selects the reference year (`YEAR(CURRENT_DATE)-1`);
2. applies domain classification rules (e.g., residential vs non-residential, voltage tiers);
3. aggregates monthly consumption values to annual totals;
4. normalizes units (e.g., dividing by 1000 to align with KPI unit conventions);
5. writes results into the tables shown in the listing 4.12 by type of energy in the L2 KPI tables, tagging the data source (`DS_FONTE = 'SIDES'`).

A representative excerpt (electricity domain) illustrates how KPI rows are generated through controlled `INSERT` statements and explicit classification filters:

```
CREATE OR REPLACE PROCEDURE PRD_DP.L2.L2_F_CONSUMI_ANNUALI_MODELLO_TE_SP_EGT()
RETURNS VARCHAR
LANGUAGE SQL
EXECUTE AS OWNER
AS '
BEGIN

  -- Electricity: domestic use (example)
  INSERT INTO L2.L2_F_CONSUMI_ELE_MODELLO_TE_TD_MTE (
      CD_ANNO_RIFERIMENTO,
      COMUNE,
      DS_CATEGORIA_USO,
      KPI_NAME,
      NM_UTENTI,
      DS_FONTE
  )
  SELECT
      ID_SUM_ANNO_CONSUMO,
      ''MI'',
      ''DOMESTICO'',
      ''Usi domestici - energia distribuita'',
```

```
      SUM(NM_SUM_CONSUMO)/1000,
      ''SIDES''
  FROM L2.L2_F_CONSUMI_MENSILI_TD_SUM
  WHERE ID_SUM_ANNO_CONSUMO = YEAR(CURRENT_DATE) - 1
    AND DS_SUM_COMUNE ILIKE ''MILANO''
    AND DS_SUM_TIPOLOGIAUTENZA_NEW = ''RESIDENZIALE''
    AND UPPER(DS_SUM_TENSIONE) = ''BASSA''
  GROUP BY ID_SUM_ANNO_CONSUMO

END;'
```

Listing 4.14: Excerpt of `L2_F_CONSUMI_ANNUALI_MODELLO_TE_SP_EGT()` (electricity KPI examples)

Two implementation aspects are particularly relevant for governance and reproducibility:

- **Execution context**: the procedure runs `EXECUTE AS OWNER`, ensuring that execution privileges are controlled and consistent with the production governance model.
- **Explicit KPI semantics**: each output row includes `KPI_NAME`, `DS_CATEGORIA_USO`, and `DS_FONTE`, which stabilizes business interpretation and supports downstream traceability.

Analytical outputs: certified KPI tables by supply These outputs represent the standardized analytical interface for annual consumption within the Modello TE domain. They are designed to be further consolidated into the final model consumption entity and to be directly consumable by downstream analytics. The procedure writes certified results into three L2 KPI tables:

```
PRD_DP L2 L2_F_CONSUMI_ELE_MODELLO_TE_TD_MTE
PRD_DP L2 L2_F_MODELLO_TE_CONSUMI_GAS_TD_SAG
PRD_DP L2 L2_F_MODELLO_TE_CONSUMI_TRL_TD_TLR
```

Listing 4.15: Annual consumption KPI outputs by supply

RQ linkage and PoC relevance

The energy consumption analytical workflow serves as a concrete validation artifact for the research questions defined in Chapter 1.

- **RQ1 (data accessibility and integration)**: Addressed through the structured ingestion of heterogeneous external data (SIDES) into certified L2 KPI tables. The transformation from raw external feeds to standardized MART entities demonstrates the platform's ability to integrate, harmonize, and expose consumption data in a consistent analytical format.

- **RQ2 (governance and maintainability)**: Validated through production-grade operational patterns, including scheduled execution (CRON-based orchestration), controlled privilege management via `EXECUTE AS OWNER`, and deterministic yearly parametrization using `YEAR(CURRENT_DATE)-1`. These mechanisms ensure repeatability, security, and reduced operational fragility compared to legacy or manually managed processes.
- **RQ3 (analytical scalability and platform suitability)**: Demonstrated by executing representative non-ML analytical workloads—annual aggregations, KPI computations, and historical consumption modeling—directly within Snowflake. The workflow illustrates improvements in reproducibility, data reuse, workload isolation, and operational scalability compared to the legacy architecture.

4.3.2. Installed power analytical workflow (CURIT)

This subsection details the analytical workflow used to compute power-related KPIs for heating systems based on CURIT (Catasto Unico Regionale Impianti Termici) data. The workflow operationalizes the L0–L1–L2 framework by transforming monthly CURIT snapshots historized in L1 into certified L2 KPIs that describe the composition of installed power by service (space heating vs domestic hot water), building typology, and Modello TE technology classes.

Workflow flow and operational trigger

The workflow starts from CURIT data, which is integrated in the Data Platform and historized at L1 level with a monthly refresh cadence (first day of each month). Annual KPI computation is triggered by a scheduled task, which executes a stored procedure that reads the relevant historized CURIT snapshot and reconciles technology attributes through static mapping tables.

```
CURIT (external source, integrated in DP)
  |
  v
PRD_DP L1 L1_DCU_CATASTO_UNICO_REG_IMPIANTI_TERMICI_CRI_TD
(monthly update; first day of the month)
  |
  v
PRD_DP L2 L2_POTENZA_IMPIANTI_CURIT_TK
```

```
CRON: 0 11 1 2 * Europe/Rome  (1 February, 11:00)
  |
  v
PRD_DP L2 L2_F_POTENZA_IMPIANTI_CURIT_SP_PZA()
  |
  +--> reads reconciliation tables (static) on schema
   DATALAB_A2A_DGE_BSI_BDT
  |
  v
PRD_DP L2 L2_F_POTENZA_IMPIANTI_CURIT_TD_PZA
```

Listing 4.16: Installed power workflow (CURIT → L2 power KPIs)

Two implementation choices are central for reproducibility and governance:

- **Reference year definition**: the procedure computes KPIs for the *consuntivated year*, defined as `YEAR(CURRENT_DATE)-1`. This aligns the KPI computation with closed-year reporting.
- **Upstream dependency check**: before computing KPIs, the procedure verifies that the L1 CURIT integration has been executed for the referring month by reading the monitoring event log.

Input dataset (L1 CURIT snapshot)

The workflow reads from the historized CURIT table:

```
PRD_DP L1 L1_DCU_CATASTO_UNICO_REG_IMPIANTI_TERMICI_CRI_TD
```

Listing 4.17: L1 historized CURIT dataset used for KPI computation

This dataset contains information on heating systems (e.g., installed power, technology, services, fuel) and is filtered within the procedure through explicit validity rules (e.g., `FL_VALIDITY = true`) and perimeter constraints (e.g., removal of outliers such as extreme power values).

Static reconciliation tables (technology mapping) To standardize CURIT categorical attributes into Modello TE technology classes, the procedure joins a set of static reconciliation tables available in the DATALAB schema (already existing within the enterprise environment). These mapping tables enable standardization of heterogeneous CURIT values (technology labels, fuels, generator typologies, building typologies, services) into a controlled taxonomy.

```
DATALAB_A2A_DGE_BSI_BDT CFG_MTE_RICONCILIAZIONE_TECNOLOGIE_TD
```

```
DATALAB_A2A_DGE_BSI_BDT CFG_MTE_RICONCILIAZIONE_SERVIZI_TD
DATALAB_A2A_DGE_BSI_BDT CFG_MTE_RICONCILIAZIONE_TIPOLOGIA_EDIFICIO_TD
DATALAB_A2A_DGE_BSI_BDT CFG_MTE_RICONCILIAZIONE_TIPOLOGIA_GENERATORE_TD
DATALAB_A2A_DGE_BSI_BDT CGF_MTE_RICONCILIAZIONE_COMBUSTIBILE_TD
DATALAB_A2A_DGE_BSI_BDT CFG_MTE_TECNOLOGIA_TEMP_TD
```

Listing 4.18: Static reconciliation tables used for CURIT standardization

The reconciliation step is essential to ensure that the resulting KPIs are consistent with Modello TE analytical conventions, and that downstream consumers do not depend on raw CURIT categorical variability.

Orchestration: annual task execution

The workflow is operationalized through a scheduled task:

```
PRD_DP L2 L2_POTENZA_IMPIANTI_CURIT_TK
0 11 1 2 * Europe/Rome
```

Listing 4.19: Scheduled task for annual CURIT power KPI computation

This scheduling choice aligns with the operational assumption that CURIT is refreshed monthly and that annual KPIs are produced after the year-end period is fully closed.

Core transformation logic: procedure

The procedure `PRD_DP.L2.L2_F_POTENZA_IMPIANTI_CURIT_SP_PZA()` is implemented in JavaScript and executed under `EXECUTE AS CALLER`. It performs three main functions: (i) validates prerequisites through monitoring metadata, (ii) standardizes and aggregates CURIT power values into Modello TE technology classes, and (iii) persists certified annual KPIs into the L2 target table under transactional control.

Prerequisite validation through monitoring events. Before computing KPIs, the procedure checks whether the L1 CURIT integration has been executed for the referring month. This check queries the event log in the monitoring database and compares the month of the latest recorded load event with the current month (RQ2).

Perimeter extraction and standardization (temporary base table). The procedure then creates a temporary base structure (e.g., `L2.DATA_BASE`) by selecting CURIT records for the consuntivated year and joining the static reconciliation tables. In this step, key filters are applied, including:

- `FL_VALIDITY = true`,

- exclusion of non-relevant technologies (e.g., "Teleriscaldamento", "Altro"),
- exclusion of outliers (e.g., `GENERATORE_POTENZA < 10000`),
- non-null constraints on reconciled services and Modello TE technology classes.

This step represents the core standardization logic that translates CURIT categorical attributes into Modello TE-compatible categories.

Service-specific KPI computation (Riscaldamento vs ACS). From the standardized base table, the procedure computes KPIs for:

- **District heating** ("Riscaldamento"): excluding services such as domestic hot water and summer cooling.
- **Domestic hot water** ("ACS"): selecting records whose reconciled services contain water-related attributes (e.g., `RICONCILIAZIONE_SERVIZI ILIKE '%acqua%'`).

For each service, the procedure computes the power composition by Modello TE technology class as a *relative share* (percentage). This is implemented by:

- aggregating power by technology and perimeter,
- computing a service-level total (excluding specific excluded technology classes such as "Caldaia ad altri combustibili"),
- calculating the ratio `SUM(POTENZA)/SUM(TOT_POTENZA)` and expressing it as a percentage.

```sql
-- Total power for the perimeter (excluding "Caldaia ad altri combustibili")
CREATE OR REPLACE TEMPORARY TABLE L2.RISCALDAMENTO_D_ND_TOT AS (
  SELECT
      ANNO_CONSUNTIVATO_RIF,
      UBICAZIONE_COMUNE,
      RICONCILIAZIONE_TIPOLOGIA_EDIFICIO,
      SUM(POTENZA) AS POTENZA
  FROM L2.RISCALDAMENTO_D_ND
  WHERE TECNOLOGIA_MODELLO_TE != 'Caldaia ad altri combustibili'
  GROUP BY
      ANNO_CONSUNTIVATO_RIF,
      UBICAZIONE_COMUNE,
      RICONCILIAZIONE_TIPOLOGIA_EDIFICIO
);

-- KPI = technology power share (%) within the perimeter
```

```
CREATE OR REPLACE TEMPORARY TABLE L2.RISCALDAMENTO_KPI AS (
  SELECT
      N.ANNO_CONSUNTIVATO_RIF,
      N.UBICAZIONE_COMUNE,
      N.RICONCILIAZIONE_TIPOLOGIA_EDIFICIO,
      N.TECNOLOGIA_MODELLO_TE,
      ROUND(SUM(N.POTENZA) / SUM(D.POTENZA) * 100, 3) AS POTENZA
  FROM L2.RISCALDAMENTO_D_ND N
  LEFT JOIN L2.RISCALDAMENTO_D_ND_TOT D
    ON  N.ANNO_CONSUNTIVATO_RIF = D.ANNO_CONSUNTIVATO_RIF
    AND N.UBICAZIONE_COMUNE = D.UBICAZIONE_COMUNE
    AND N.RICONCILIAZIONE_TIPOLOGIA_EDIFICIO = D.
   RICONCILIAZIONE_TIPOLOGIA_EDIFICIO
  WHERE N.TECNOLOGIA_MODELLO_TE != 'Caldaia ad altri combustibili'
  GROUP BY
      N.ANNO_CONSUNTIVATO_RIF,
      N.UBICAZIONE_COMUNE,
      N.RICONCILIAZIONE_TIPOLOGIA_EDIFICIO,
      N.TECNOLOGIA_MODELLO_TE
);
```

Listing 4.20: Percentage share computation (heating power KPIs by Modello TE technology)

In analytical terms, for a given year, municipality, and building typology, the resulting KPI rows can be interpreted as a share (%) of the percentage of installed power mapped to the given technology (eg. traditional boiler) with a consistent denominator (the service-level total after exclusions).

Analytical output: certified L2 power KPI table The procedure writes certified annual KPIs into the following L2 entity:

```
PRD_DP L2 L2_F_POTENZA_IMPIANTI_CURIT_TD_PZA
```

Listing 4.21: L2 output – CURIT installed power KPIs

RQ linkage and PoC relevance

The CURIT workflow extends the validation of the proposed architecture by introducing historized regulatory data and domain-specific perimeter logic.

- **RQ1 (data accessibility and integration)**: Validated by integrating the external CURIT registry into a historized L1 table, preserving temporal consistency and traceability. Certified KPIs are then exposed through an L2 analytical entity, demonstrating the platform's ability to standardize heterogeneous regulatory data into reusable analytical structures.

- **RQ2 (governance, maintainability, and reproducibility)**: Demonstrated through prerequisite validation checks, controlled execution via `EXECUTE AS CALLER`, explicit perimeter enforcement rules, and idempotent transactional persistence. These mechanisms ensure traceable execution, controlled data exposure, and reproducible KPI generation across temporal updates.
- **RQ3 (analytical scalability and platform suitability)**: Validated through the production of standardized, time-consistent power-related KPIs that can be reused across descriptive analytics and regulatory reporting workloads. The workflow illustrates the platform's capacity to manage historized datasets while maintaining performance, consistency, and semantic clarity.

4.3.3. Building energy certification workflow (CENED)

This subsection details the analytical workflow used to compute building energy certification KPIs based on CENED data provided through Regione Lombardia APIs. The workflow operationalizes the L0–L1–L2 framework by ingesting certification records into landing (L0), historizing and validating them in L1, and producing certified L2 KPI entities designed for the Modello TE consumption logic.

Workflow flow and operational trigger

The workflow is orchestrated via a scheduled task, which triggers a final procedure. The final procedure coordinates three layer-specific procedures (L0, L1, L2), ensuring that ingestion, historization, and KPI computation remain traceable and reproducible.

```
CENED API (Regione Lombardia)
  |
  v
PRD_DP L2 L2_CERTIFICAZIONE_ENERGETICA_EDIFICI_TK
CRON: 0 8 1 2 * Europe/Rome  (1 February, 08:00)
  |
  v
PRD_DP L2 L2_CERTIFICAZIONE_ENERGETICA_EDIFICI_SP_FINAL()
  |
  +--> calls: PRD_DP L0 L0_CEN_CERTIFICAZIONE_ENERGETICA_EDIFICI_SP()
  |        output: PRD_DP L0 L0_CEN_CERTIFICAZIONE_ENERGETICA_EDIFICI_TD
  |
```

```
+--> calls: PRD_DP L1 L1_CEN_CERTIFICAZIONE_ENERGETICA_EDIFICI_CEE_SP
 ()
|         output: PRD_DP L1
 L1_CEN_CERTIFICAZIONE_ENERGETICA_EDIFICI_CEE_TD
|
+--> calls: PRD_DP L2 L2_F_CERTIFICAZIONE_ENERGETICA_EDIFICI_SP_CEN()
          output: PRD_DP L2
 L2_F_CERTIFICAZIONE_ENERGETICA_EDIFICI_TD_CEN
```

Listing 4.22: CENED workflow (API → L0 → L1 → L2 KPIs)

The key design choice for reproducibility is that the workflow computes KPIs for the *consuntivated year* defined as `YEAR(CURRENT_DATE) - 1`, ensuring that annual outputs always refer to a closed historical period.

Input dataset (L2 CENED snapshot)

The workflow instantiates the methodological layers as follows:

```
PRD_DP L0 L0_CEN_CERTIFICAZIONE_ENERGETICA_EDIFICI_TD
PRD_DP L1 L1_CEN_CERTIFICAZIONE_ENERGETICA_EDIFICI_CEE_TD
PRD_DP L2 L2_F_CERTIFICAZIONE_ENERGETICA_EDIFICI_TD_CEN
```

Listing 4.23: CENED layer artifacts used in the PoC

- **L0 (landing)** stores the raw payload obtained from the external API, enabling reprocessing and auditability.
- **L1 (historized)** stores validated and historized certification records, including data quality flags (e.g., `FL_VALIDITY`).
- **L2 (analytical)** exposes standardized KPI outputs aligned with Modello TE naming conventions and consumption typologies (domestic vs non-domestic).

Core transformation logic: procedure

The procedure `PRD_DP.L2.L2_F_CERTIFICAZIONE_ENERGETICA_EDIFICI_SP_CEN()` is implemented in JavaScript and executed under `EXECUTE AS CALLER`. It performs:

Step 1 – Base table construction (feature engineering on L1). The procedure creates a temporary base table from L1 historized records and derives additional measures used for KPI computation.

```
CREATE OR REPLACE TEMPORARY TABLE L2.CENED_BASE_TABLE AS
SELECT
```

```
  CEE.*,
  CEE.EP_GL_NREN * CEE.SUPERFICIE_DISPERDENTE AS NM_KWH_ASSORBITI,
  CASE
    WHEN CLASSE_ENERGETICA IN ('G','F') THEN 'Bassa'
    WHEN CLASSE_ENERGETICA IN ('E','D','C') THEN 'Media'
    ELSE 'Alta'
  END AS CLUSTERIZZAZIONE
FROM L1.L1_CEN_CERTIFICAZIONE_ENERGETICA_EDIFICI_CEE_TD CEE
WHERE CEE.FL_VALIDITY = TRUE
  AND (RESIDENZIALE = TRUE OR NON_RESIDENZIALE = TRUE);
```

Listing 4.24: Base table construction and efficiency clustering (excerpt)

The explicit condition on residential / non-residential flags is included to handle cases where both flags are `FALSE`, ensuring perimeter consistency in the KPI computation.

Step 2 – Pivot aggregation by system type, energy vector, cluster, and use. The procedure aggregates the base table into a pivoted structure by system type (e.g., `CI_TIPO_IMPIANTO_1`), energy vector, efficiency cluster, municipality, and domestic vs non-domestic use.

```
CREATE OR REPLACE TEMPORARY TABLE L2.CENED_PIVOT AS
SELECT
  CI_TIPO_IMPIANTO_1,
  CI_VETTORE_ENERGETICO_1,
  CLUSTERIZZAZIONE,
  COMUNE,
  CASE
    WHEN RESIDENZIALE = TRUE THEN 'DOMESTICO'
    WHEN NON_RESIDENZIALE = TRUE THEN 'NON DOMESTICO'
  END AS DS_D_ND,
  SUM(NM_KWH_ASSORBITI) AS TOT_KWH_ASSORBITI
FROM L2.CENED_BASE_TABLE
GROUP BY
  CI_TIPO_IMPIANTO_1,
  CI_VETTORE_ENERGETICO_1,
  CLUSTERIZZAZIONE,
  COMUNE,
  DS_D_ND;
```

Listing 4.25: Step 2 excerpt: pivoted aggregation and DS_D_ND derivation

Step 3 – Totalization for technology families. A second aggregation computes totals per municipality and use for three main technology families relevant to Modello

TE: Boilers (Caldaia), Heat pumps (PDC) and District heating (Teleriscaldamento).

Each family is computed separately by efficiency cluster (Alta/Media/Bassa), enabling subsequent ratio computation.

```
CREATE OR REPLACE TEMPORARY TABLE L2.CENED_TOT AS
SELECT
  COMUNE,
  DS_D_ND,
  SUM(CASE
        WHEN CI_TIPO_IMPIANTO_1 = 'Generatore a combustione'
         AND CI_VETTORE_ENERGETICO_1 = 'Gas naturale'
         AND CLUSTERIZZAZIONE = 'Alta'
        THEN TOT_KWH_ASSORBITI ELSE 0
      END) AS TOTAL_CALDAIA_ALTA,
  SUM(CASE
        WHEN CI_TIPO_IMPIANTO_1 = 'Generatore a combustione'
         AND CI_VETTORE_ENERGETICO_1 = 'Gas naturale'
         AND CLUSTERIZZAZIONE = 'Media'
        THEN TOT_KWH_ASSORBITI ELSE 0
      END) AS TOTAL_CALDAIA_MEDIA,
  SUM(CASE
        WHEN CI_TIPO_IMPIANTO_1 = 'Generatore a combustione'
         AND CI_VETTORE_ENERGETICO_1 = 'Gas naturale'
         AND CLUSTERIZZAZIONE = 'Bassa'
        THEN TOT_KWH_ASSORBITI ELSE 0
      END) AS TOTAL_CALDAIA_BASSA
FROM L2.CENED_PIVOT
GROUP BY COMUNE, DS_D_ND;
```

Listing 4.26: Step 3 excerpt: totals for boilers (Caldaia) by cluster

Step 4 – Mapping to legacy KPI naming conventions. The procedure creates and populates a mapping table that aligns the computed ratios (e.g., `PER_CALDAIA_ALTA`) with the KPI identifiers and descriptions expected by downstream Modello TE logic and reporting (e.g., `MIX_CLASSE_ENERGIA_CALDAIA_ALTA_EFF_D`). This step ensures backward-compatible semantics without changing the governance model.

```
CREATE OR REPLACE TEMPORARY TABLE L2.CENED_MAPPING (
  CD_CEN_INPUT_NAME VARCHAR(),
  DS_D_ND VARCHAR(),
  CD_INPUT_NAME VARCHAR(),
  DS_INPUT_NAME VARCHAR()
);

```

```
INSERT INTO L2.CENED_MAPPING VALUES
('PER_CALDAIA_ALTA','DOMESTICO','MIX_CLASSE_ENERGIA_CALDAIA_ALTA_EFF_D',
 'Mix classe energetica - Caldaia - Alta efficienza - Domestico');
```

Listing 4.27: Step 4 excerpt: mapping table and representative values

Step 5 – Final ratio computation and unpivoting. The final KPI values are computed as shares within each technology family, using `NULLIF` to prevent division-by-zero. For example, the boiler efficiency mix is computed as:

```
CREATE OR REPLACE TEMPORARY TABLE L2.CENED_FINAL AS
SELECT
  UPV.COMUNE,
  UPV.CD_CEN_INPUT_NAME,
  UPV.DS_D_ND,
  UPV.NM_VALORE,
  CASE
    WHEN UPV.DS_D_ND = 'DOMESTICO' THEN 'C_RISC-D'
    WHEN UPV.DS_D_ND = 'NON DOMESTICO' THEN 'C_RISC-ND'
  END AS DS_TIPOLOGIA_CONSUMO,
  MAP.CD_INPUT_NAME,
  MAP.DS_INPUT_NAME
FROM (
  SELECT
    COMUNE,
    DS_D_ND,
    TOTAL_CALDAIA_ALTA / NULLIF(TOTAL_CALDAIA_ALTA+TOTAL_CALDAIA_BASSA+
   TOTAL_CALDAIA_MEDIA,0) AS PER_CALDAIA_ALTA
  FROM L2.CENED_TOT
) UNPIVOT (NM_VALORE FOR CD_CEN_INPUT_NAME IN (PER_CALDAIA_ALTA)) UPV
INNER JOIN L2.CENED_MAPPING MAP
  ON MAP.CD_CEN_INPUT_NAME = UPV.CD_CEN_INPUT_NAME
 AND MAP.DS_D_ND = UPV.DS_D_ND;
```

Listing 4.28: Step 5 excerpt: ratio computation and UNPIVOT into normalized KPI rows

The result is then `UNPIVOT`ed into a normalized KPI table and enriched with mapped KPI names and descriptions.

Analytical output: certified L2 KPI table The output stores, for each municipality and for domestic/non-domestic consumption typology, the standardized KPI ratios representing the efficiency mix for boilers, heat pumps, and district heating. The workflow produces the following certified analytical entity:

```
PRD_DP L2 L2_F_CERTIFICAZIONE_ENERGETICA_EDIFICI_TD_CEN
```

Listing 4.29: L2 output – CENED certification KPIs for Modello TE

RQ linkage and PoC relevance

The CENED workflow represents the most structurally advanced validation artifact within the PoC, integrating API-based data ingestion, layered orchestration, and modeling-oriented KPI construction.

- **RQ1 (data accessibility and integration)**: Validated through the ingestion of an external API-based source into a historized L1 structure, followed by the production of certified L2 KPI entities. This demonstrates the platform's ability to operationalize semi-structured external data streams and convert them into standardized analytical assets usable across energy-demand use cases.
- **RQ2 (governance and reproducibility)**: Demonstrated through caller-based execution control, explicit separation of responsibilities via layered procedures (L0–L1–L2), and idempotent annual persistence under transactional guarantees. This layered orchestration ensures modularity, auditability, and long-term maintainability of the data pipeline.
- **RQ3 (analytical scalability and platform suitability)**: Validated by producing standardized municipality-level KPI ratios aligned with energy-demand analytical requirements. The workflow illustrates the platform's ability to support feature-oriented transformations and consistent territorial aggregations while preserving reproducibility and performance.

Overall, the CENED workflow demonstrates that the Snowflake-based architecture can support API-driven ingestion, layered procedural orchestration, and modeling-oriented analytical preparation in a governed and scalable environment.

4.3.4. Manual business inputs and parameterization workflows

This subsection describes how manual business inputs are integrated into the Snowflake-based L0–L1–L2 framework through three governed interfaces. These interfaces enable controlled human intervention in the energy-demand modeling process while preserving traceability, historization, and reproducibility.

INTFC_USR_INPUT_MODELLO_TE – Manual model inputs

This interface captures expert-defined inputs required by the energy transition model, including change factors, renovation rates, and other scenario- and city-specific parameters that cannot be derived from operational systems. These inputs represent deliberate modeling assumptions rather than observational data.

```
Business User
  |
  v
Excel upload to S3 (INTFC_USR_INPUT_MODELLO_TE)
  |
  v (Trigger)
PRD_DP L0 L0_INTFC_USR_INPUT_MODELLO_TE_TD      -- landing
  |
  v
PRD_DP L1 L1_INTFC_USR_INPUT_MODELLO_TE_IMT_SP()
  |
  v
PRD_DP L1 L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD  -- historized
  |
  v
Consumed by L2 procedures (energy consumption analytics)
```

Listing 4.30: Manual input workflow – INTFC_USR_INPUT_MODELLO_TE

Implementation logic. The L1 procedure validates structural consistency (schema, mandatory fields) and historizes each upload by timestamp. No overwrite is allowed at L1, ensuring that changes in modeling assumptions are fully auditable over time. L2 consumption procedures explicitly reference the L1 historized table, guaranteeing deterministic recomputation for any given scenario and reference year.

INTFC_USR_CONTROL_PANEL_MODELLO_TE – Control parameters

The control panel interface centralizes global parameters used across KPI computations, such as unit conversion factors, physical constants, and utilization coefficients. These parameters act as shared model configuration rather than scenario-specific inputs.

```
Business User
```

```
  |
  v
Excel upload to S3 (INTFC_USR_CONTROL_PANEL_MODELLO_TE)
  |
  v (Trigger)
PRD_DP L0 L0_INTFC_USR_CONTROL_PANEL_MODELLO_TE_TD
  |
  v
PRD_DP L1 L1_INTFC_USR_CONTROL_PANEL_MODELLO_TE_CPM_SP()
  |
  v
PRD_DP L1 L1_INTFC_USR_CONTROL_PANEL_MODELLO_TE_CPM_TD
  |
  v
Consumed by L2 procedures (energy consumption analytics)
```

Listing 4.31: Control panel – INTFC_USR_CONTROL_PANEL_MODELLO_TE

Implementation logic. The L1 historized table stores parameter versions over time, allowing the reconstruction of historical KPI outputs under the exact parameter set used at execution time. L2 procedures read the relevant parameter snapshot implicitly through execution date logic, without embedding hard-coded constants.

INTFC_USR_PROFILI_CARICO_MODELLO_TE – Load profiles

Purpose and scope. This interface manages hourly load curves for 20 predefined user profiles. These curves are required for peak power estimation and are therefore consumed by L2 power analytics rather than consumption KPIs.

```
Business User
  |
  v
Excel upload to S3 (INTFC_USR_PROFILI_CARICO_MODELLO_TE)
  |
  v (Trigger)
PRD_DP L0 L0_INTFC_USR_PROFILI_CARICO_MODELLO_TE_TD
  |
  v
PRD_DP L1 L1_INTFC_USR_PROFILI_CARICO_MODELLO_TE_CMT_SP()
  |
```

```
  v
PRD_DP L1 L1_INTFC_USR_PROFILI_CARICO_MODELLO_TE_CMT_TD
  |
  v
Consumed by L2 procedures (power analytics)
```

Listing 4.32: Load profiles – INTFC_USR_PROFILI_CARICO_MODELLO_TE

Implementation logic. Load profiles are historized to allow recalculation of peak power KPIs under different temporal assumptions. L2 power procedures explicitly join these profiles with installed capacity data (e.g., CURIT-derived KPIs), preserving deterministic outcomes for any execution year.

RQ linkage and PoC relevance

The manual input interfaces represent a complementary validation layer within the PoC, demonstrating how expert-driven and non-system-generated data can be incorporated into the analytical architecture without compromising governance or reproducibility.

- **RQ1 (data accessibility and integration)**: Validated by integrating non-system-generated data—such as configuration parameters, temporal patterns, and expert-defined assumptions—into certified analytical pipelines. These inputs are ingested into controlled L0/L1 layers and subsequently consumed by L2 procedures, demonstrating the platform's ability to harmonize operational and expert-driven data sources within a unified architecture.
- **RQ2 (governance, maintainability, and reproducibility)**: Demonstrated through strict historization of manual inputs, transactional persistence, and the externalization of constants from procedural logic. By separating configuration data from executable procedures, the architecture enhances maintainability, traceability, and controlled evolution of analytical assumptions over time.
- **RQ3 (analytical scalability and platform suitability)**: Validated by enabling expert-defined temporal patterns and scenario parameters to be integrated into the analytical layer in a structured and reproducible manner. This design supports the execution of descriptive and simulation-oriented analytical workloads without introducing fragility or manual overrides into certified L2 entities.

4.3.5. Final consumption orchestration procedure (Modello TE)

This subsection describes the final orchestration procedure used to generate the certified consumption output of the Modello TE. Unlike the upstream KPI-producing workflows (SIDES, CURIT, CENED), which are scheduled and generate dedicated L2 KPI entities, the final Modello TE consumption computation is executed manually by business users and updates a single consolidated L2 table.

Purpose and output contract

The main procedure `L2_MODELLO_TE_CONSUMI_SP_FINAL()` acts as an orchestrator that executes, in sequence, two computational steps:

- `L2_MODELLO_TE_CONSUMI_SP_STEP_1()` (Part 1),
- `L2_MODELLO_TE_CONSUMI_SP_STEP_2()` (Part 2).

Both steps update the same certified analytical entity:

```
PRD_DP L2 L2_F_MODELLO_TE_TD_TEM
```

Listing 4.33: Final Modello TE consumption output entity

The table `L2_F_MODELLO_TE_TD_TEM` represents the consolidated and business-oriented consumption output of the PoC (i.e., the "single version" used by BI and strategic analyses).

Operational flow (manual execution and sequencing)

The final procedure is typically executed manually by business users (commonly in February/March, after upstream annual KPIs are available and business inputs are aligned). Its operational role is to ensure a controlled sequencing of the two computational phases and to provide a single execution point for the end-to-end consumption calculation.

```
Business User (manual execution)
  |
  v
PRD_DP L2 L2_MODELLO_TE_CONSUMI_SP_FINAL()
  |
  +--> PRD_DP L2 L2_MODELLO_TE_CONSUMI_SP_STEP_1()
  |
  +--> PRD_DP L2 L2_MODELLO_TE_CONSUMI_SP_STEP_2()
  |
  v
```

```
PRD_DP L2 L2_F_MODELLO_TE_TD_TEM   (updated)
```

Listing 4.34: Final consumption orchestration flow

This sequencing is a key implementation mechanism to maintain determinism and to isolate intermediate logic into well-defined computational blocks.

STEP_1: consumption computation (Part 1)

Purpose. `L2_MODELLO_TE_CONSUMI_SP_STEP_1()` computes the first set of consumption KPIs and updates the target table. The execution time is approximately ~40 minutes, reflecting the breadth of cross-domain joins and aggregations performed.

Sections computed. The procedure covers the following logical "sheets" / sections:

- `FATTORE_DI_CAMBIAMENTO`
- `C_RISC-D` (Domestic heating consumption)
- `C_RISC-ND` (Non-domestic heating consumption)
- `C_ACS` (Domestic hot water consumption)
- `C_E-D` (Domestic electricity consumption)
- `C_PP` (Small processes consumption)
- `C_MPUBBLICA` (Public mobility consumption)

Input datasets (cross-domain integration). The step integrates multiple certified L2 KPI entities and historized manual inputs:

```
PRD_DP L2 L2_F_CONSUMI_ELE_MODELLO_TE_TD_MTE
PRD_DP L2 L2_F_MODELLO_TE_CONSUMI_GAS_TD_SAG
PRD_DP L2 L2_F_MODELLO_TE_CONSUMI_TRL_TD_TLR
PRD_DP L2 L2_F_POTENZA_IMPIANTI_CURIT_TD_PZA
PRD_DP L2 L2_F_CERTIFICAZIONE_ENERGETICA_EDIFICI_TD_CEN
PRD_DP L1 L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD
PRD_DP L1 L1_INTFC_USR_CONTROL_PANEL_MODELLO_TE_CPM_TD
```

Listing 4.35: STEP_1 inputs (certified L2 KPIs + historized L1 manual tables)

This is the point where the PoC most clearly demonstrates the methodological integration logic: upstream KPIs are already standardized and certified in L2, while manual business inputs are historized in L1 and injected as controlled parameters into the consumption model.

Output behavior. The procedure updates `L2_F_MODELLO_TE_TD_TEM`. As the final table is overwritten with the new calculation, the output is not historized within the table itself.

STEP_2: consumption computation (Part 2)

Purpose. `L2_MODELLO_TE_CONSUMI_SP_STEP_2()` completes the consumption computation by processing the remaining logical sections and producing the final aggregated synthesis. The execution time is also approximately ~40 minutes.

Sections computed. STEP_2 covers the following sections:

- `C_E-ND` (Non-domestic electricity consumption)
- `C_NUOVE_URBANIZZAZIONI` (New urbanization consumption)
- `C_CUCINA` (Cooking consumption)
- `C_MPRIVATA` (Private mobility consumption)
- `C_AUTOPRODUZIONE` (Self-production consumption)
- `SINTESI_CONSUMI` (Aggregated consumption synthesis)

Input datasets. STEP_2 uses the same certified L2 KPIs and historized L1 manual tables as STEP_1, ensuring consistent parametrization and cross-domain alignment across both phases.

Output behavior. The procedure updates `L2_F_MODELLO_TE_TD_TEM`, completing the final consumption dataset used for PoC validation and downstream analysis.

Critical point: overwrite behavior and reproducibility implications

A critical characteristic of the final orchestration is that `L2_F_MODELLO_TE_TD_TEM` is overwritten when `L2_MODELLO_TE_CONSUMI_SP_FINAL()` is executed. Consequently, previous outputs cannot be retrieved directly from the table.

This limitation does not invalidate the PoC, as reproducibility is maintained through:

- historized upstream inputs in L1 (manual inputs and control parameters),
- certified upstream KPI tables in L2 (SIDES, CURIT, CENED outputs),

- deterministic computation logic (consistent sequencing and input contracts).

However, the overwrite behavior represents a governance trade-off: the final table provides a single certified state for consumption analytics, but does not preserve time-versioned results in-place. In the PoC, this is treated as a scoped limitation rather than a redesign opportunity, consistent with the constraint of not altering the production architecture.

RQ linkage and PoC relevance

- **RQ1 (data accessibility and integration)**: validated through the consolidation of certified L2 KPIs (SIDES, CURIT, CENED) and historized manual inputs into a unified analytical entity.
- **RQ2 (governance, maintainability, reproducibility)**: validated through controlled manual execution, deterministic two-step sequencing, and reliance on historized upstream inputs despite overwrite behavior in the final table.
- **RQ3 (analytics and ML readiness)**: supported by producing a stable, business-oriented fact table that can be reused as a feature-ready dataset for future advanced analytics without implementing ML within the PoC.

4.3.6. Final power (peak load) procedure (Modello TE)

This subsection completes the description of the end-to-end analytical workflows by detailing the manual procedure used to compute power-related KPIs (i.e., electric grid peak loads) for the Modello TE. The procedure `L2_MODELLO_TE_POTENZE_SP_FINAL()` is conceptually parallel to `L2_MODELLO_TE_CONSUMI_SP_FINAL()` and is executed manually by business users, typically in February/March, after the final consumption outputs and load profiles are available.

Operational flow (manual execution and dependencies)

`L2_MODELLO_TE_POTENZE_SP_FINAL()` computes the KPIs for the *power* section of the model, focusing on peak loads and network capacity implications. The procedure updates the certified L2 output table:

```
PRD_DP L2 L2_F_POTENZE_MODELLO_TE_TD_PTE
```

Listing 4.36: Final Modello TE power output entity

The power workflow consumes the certified output of the consumption workflow and combines it with hourly load profiles historized in L1. Operationally, the procedure is executed manually by business users once the consumption calculation is finalized.

```
Business User (manual execution)
  |
  v
PRD_DP L2 L2_MODELLO_TE_POTENZE_SP_FINAL()
  |
  +--> reads: PRD_DP L2 L2_F_MODELLO_TE_TD_TEM
  |
  +--> reads: PRD_DP L1 L1_INTFC_USR_PROFILI_CARICO_MODELLO_TE_CMT_TD
  |
  v
PRD_DP L2 L2_F_POTENZE_MODELLO_TE_TD_PTE   (updated)
```

Listing 4.37: Final power computation flow (consumption → load profiles → peaks)

Computed sections (analytical perimeter)

The procedure computes multiple sections, each corresponding to a specific peak-load component or transformation required for power analytics:

- P_CALORE (Power for heat-related uses)
- P_APPLIANCES (Appliances power)
- P_CUCINA (Cooking power)
- P_MPUBBLICA (Public mobility power)
- P_MPRIVATA (Private mobility power)
- P_AUTOPRODUZIONE (Self-production power)
- P_NUOVE_URBANIZZAZIONI (New urbanization power)
- P_PP_MT_BT (Small processes power, MT/BT)
- P_PP_AT (Small processes power, AT)
- CONSUMI_AT_MT_BT (Consumption by AT/MT/BT)
- CONSUMI_MT_BT (Consumption by MT/BT)
- PRELIEVI_AT_MT_BT_EVOLUT (Evolutive withdrawals, AT/MT/BT)
- PRELIEVI_MT_BT_EVOLUT (Evolutive withdrawals, MT/BT)
- PRELIEVI_AT_MT_BT_STATIC (Static withdrawals, AT/MT/BT)

- PRELIEVI_MT_BT_STATIC (Static withdrawals, MT/BT)

This decomposition mirrors the model's logic: each component is computed independently, then reconciled into a coherent peak-load representation by tension category.

Input datasets

The procedure uses two inputs, both governed and traceable in the platform:

```
PRD_DP L2 L2_F_MODELLO_TE_TD_TEM
PRD_DP L1 L1_INTFC_USR_PROFILI_CARICO_MODELLO_TE_CMT_TD
```

Listing 4.38: Inputs to the final power procedure

L2_F_MODELLO_TE_TD_TEM provides the consumption KPIs already consolidated and certified by the final consumption. L1_INTFC_USR_PROFILI_CARICO_MODELLO_TE_CMT_TD provides historized hourly load profiles (20 user profiles), ensuring that peak computation can be reproduced under the same temporal assumptions.

Core analytical logic The procedure operationalizes peak-load estimation through the following deterministic steps:

1. **Read certified consumption KPIs** from the final consumption table.
2. **Join with hourly load profiles** to distribute annual (or aggregated) consumption into an hourly shape.
3. **Compute peak power** for each profile and time slot (e.g., by identifying maximum hourly load within the period).
4. **Aggregate peak results** across tension categories (AT, MT, BT) and across the modeled components (mobility, appliances, cooking, etc.).

This design keeps the power KPIs strictly dependent on: (i) certified consumption outputs, and (ii) governed, historized temporal profiles—supporting reproducibility and analytical consistency (RQ2).

Critical point: overwrite behavior

Similarly to L2_F_MODELLO_TE_TD_TEM, the power output table is overwritten when the final power procedure is executed. As a result, previous power outputs are not recoverable directly from the L2 table.

RQ linkage

- **RQ1**: integrates certified consumption KPIs with historized manual load profiles into a business-oriented L2 peak-load entity.
- **RQ2**: demonstrates governed manual execution and deterministic peak computation based on traceable inputs, despite overwrite behavior in the final table.
- **RQ3**: produces structured peak indicators suitable for future advanced analytics and scenario stress-testing (ML excluded in the PoC).

4.4. Computation of Total Electricity Consumption and Sub-Components: Implementation within the Proof of Concept

This subsection illustrates, through a concrete application example, how the Modello TE computes total electricity consumption and its internal breakdown into sectoral and functional components. The objective is not to exhaustively report numerical results, but to demonstrate how the conceptual framework defined in section 4.3.5 is operationalized into structured, traceable computations.

The electricity domain is selected as a representative case, as it integrates multiple consumption drivers (residential, non-residential, mobility, services, and self-production) and combines certified KPIs with manual assumptions and efficiency-based technology splits.

The implementation is expressed in the Modello TE through a chain of intermediate KPIs written into the certified L2 fact table `L2_F_MODELLO_TE_TD_TEM`, and subsequently used for higher-level aggregations (e.g., total electric heating consumption).

The objective of reporting this computation is: (i) to show how certified and historized inputs are combined deterministically into a consumptions KPI (RQ1), and (ii) to illustrate how manual assumptions are operationalized through controlled parameter tables and historized inputs (RQ2). The resulting KPI structure also supports analytical reuse and future advanced use cases (RQ3), since all intermediate quantities remain explicitly materialized and auditable.

Scope and reference perimeter

The computation focuses on the annual total electricity consumption for a given reference year and municipality, decomposed into mutually exclusive sub-components. Each sub-component corresponds to a specific consumption logic and is computed independently before being reconciled into a coherent total.

The breakdown considered in the PoC includes the following items:

- Residential heating (domestic heating)
- Non-residential heating
- Domestic hot water (ACS)
- Domestic electricity for lighting and appliances
- Non-domestic electricity for lighting and cooling
- Cooking
- Private mobility
- Public mobility
- Productive processes
- Public and private illumination
- New urban developments
- Electricity production from photovoltaic systems

Each item is computed using a specific combination of upstream KPIs, technology assumptions, and manual parameters, as detailed in the following subsections.

Total Electricity Consumption (KPI 801)

The total electricity consumption is computed as an aggregate KPI (`ID_KPI = 801`) within the `SINTESI_CONSUMI` category. For each projection year Y, the total value corresponds to the sum of all electricity-related sectoral KPIs (IDs 802–813) calculated in the analytical layer.

Formally:

$$C_{\mathrm{EE,tot}}(Y) = \mathrm{KPI}_{801}(Y) = \sum_{k=802}^{813} \mathrm{KPI}_k(Y).$$

where:

- Y denotes the projection year,

- $\text{KPI}_k(Y)$ represents the electricity consumption of sector k,
- the interval $[802, 813]$ includes all modeled electricity end-use components (domestic, non-domestic, mobility, productive processes, etc.).

This formulation ensures full traceability, as the total electricity demand is strictly derived from already computed sectoral contributions without additional transformations or external inputs.

Conceptual aggregation logic

The total electricity consumption is obtained as the aggregation of its individual components, each of which is computed independently and reconciled at the final stage. This design ensures transparency and allows partial recalculation without affecting unrelated components.

Formally, the total electricity consumption can be expressed as the sum of the following categories:

- Heating-related electricity consumption
- Non-heating electricity consumption
- Mobility-related electricity consumption
- Net contribution from new developments and trends

This separation reflects both the analytical structure of the Modello TE and the organization of the final L2 consumption table. In the following sections, some of the processes of each category will be displayed to better understand the flow and the final computation.

4.4.1. Heating-related electricity consumption

Heating-related electricity consumption is computed by isolating the share of heating demand met through electric technologies. In the PoC, this component shows high-efficiency electric solutions in order to maintain the thesis's readability.

The following sub-components are computed independently:

- Domestic heating via high-efficiency heat pumps
- Non-domestic heating via high-efficiency heat pumps
- Domestic hot water (ACS) supplied by electric systems

Each sub-component is derived by combining:

- certified thermal demand KPIs,
- technology mix coefficients (e.g. efficiency class shares),
- conversion parameters from the control panel,
- manual scenario inputs where applicable.

Domestic heat pumps (high efficiency)

To maintain the thesis narrative style while providing concrete evidence, only the most relevant excerpts are reported below: (i) the injection of manual inputs into intermediate KPIs, and (ii) the final materialization of KPI 194 as a function of intermediate KPIs.

(1) Manual input injection: surface per housing unit (KPI 236) and factor for high energy class (KPI 224). KPI 236 is loaded from the historized manual input interface. This ensures that the parameter is traceable to an explicit business upload and remains reproducible for the reference year.

```
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
( CD_ANNO_CONSUNTIVATO, CD_ANNO_PROIEZIONE, ID_KPI, DS_KPI,
  DS_TIPOLOGIA_CONSUMO, DS_TIPOLOGIA_CONTRIBUTO, COMUNE, NM_VALORE,
   DS_TIPO_VALORE )
SELECT
  IMT.CD_ANNO_CONSUNTIVATO,
  IMT.CD_ANNO_PROIEZIONE,
  236,
  DS_INPUT_NAME,
  DS_TIPOLOGIA_CONSUMO,
  'Mq per U.A',
  COMUNE,
  VALORE,
  '#'
FROM L1.L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD IMT
WHERE CD_INPUT_NAME = 'MQ_UA_RD'
  AND DS_TIPOLOGIA_CONSUMO = 'C_RISC-D'
  AND IMT.CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1;
```

Listing 4.39: KPI 236 – Mq per unità abitativa for domestic heating (from historized manual inputs)

The impact factor KPI is also loaded from the historized manual input interface. This isolates scenario assumptions from procedural logic and supports governance.

```sql
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
( CD_ANNO_CONSUNTIVATO, CD_ANNO_PROIEZIONE, ID_KPI, DS_KPI,
  DS_TIPOLOGIA_CONSUMO, DS_TIPOLOGIA_CONTRIBUTO, COMUNE, NM_VALORE,
   DS_TIPO_VALORE )
SELECT
  IMT.CD_ANNO_CONSUNTIVATO,
  IMT.CD_ANNO_PROIEZIONE,
  224,
  DS_INPUT_NAME,
  'C_RISC-D',
  'Impatto sui consumi classe energetica',
  COMUNE,
  VALORE,
  '%'
FROM L1.L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD IMT
WHERE CD_INPUT_NAME = 'IMPATTO_CONSUMI_CE_RDA'
  AND DS_TIPOLOGIA_CONSUMO IN ('C_RISC-D','FATTORE_DI_CAMBIAMENTO')
  AND IMT.CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1;
```

Listing 4.40: KPI 224 – Impact factor for domestic heating consumption (high energy class)

(2) Derived intermediate KPI KPI 233 represents the *average consumption per housing unit* for electric heat pumps and is derived from: (i) a baseline domestic heating consumption KPI (computed upstream in the same L2 fact table), (ii) a scenario input factor sourced from the historized manual inputs, (iii) a normalization factor KPI, and (iv) a control-panel conversion parameter. This design makes explicit which parts of the KPI are driven by certified model quantities versus business-controlled assumptions.

```sql
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
( CD_ANNO_CONSUNTIVATO, CD_ANNO_PROIEZIONE, ID_KPI, DS_KPI,
  DS_TIPOLOGIA_CONSUMO, DS_TIPOLOGIA_CONTRIBUTO, COMUNE, NM_VALORE,
   DS_TIPO_VALORE )
SELECT
  YEAR(CURRENT_DATE()) - 1,
  YEAR(CURRENT_DATE()) - 1,
  233,
  'Consumo medio unita abitativa per Pompa di calore EE',
  'C_RISC-D',
  'Consumo medio U.A - Edificio classe energetica bassa',
  'MI',
  (
    (SELECT NM_VALORE
     FROM L2.L2_F_MODELLO_TE_TD_TEM
```

```
    WHERE ID_KPI = 230
      AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1
      AND CD_ANNO_PROIEZIONE   = YEAR(CURRENT_DATE) - 1)
   *
   (SELECT VALORE
    FROM L1.L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD
    WHERE ID = 139
      AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1
      AND CD_ANNO_PROIEZIONE   = YEAR(CURRENT_DATE) - 1)
   /
   (SELECT NM_VALORE
    FROM L2.L2_F_MODELLO_TE_TD_TEM
    WHERE ID_KPI = 235
      AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1
      AND CD_ANNO_PROIEZIONE   = YEAR(CURRENT_DATE) - 1)
   *
   (SELECT VALUE
    FROM L1.L1_INTFC_USR_CONTROL_PANEL_MODELLO_TE_CPM_TD
    WHERE CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1
      AND DS_INPUT = 'Conversione kWh/smc'
      AND COMUNE = 'MI')
 ),
 '#';
```

Listing 4.41: KPI 233 – Average unit consumption for electric heat pumps (excerpt)

Baseline average unit consumption KPI 230 (traditional boiler reference) and KPI 234 (district heating reference). KPI 230 is computed in L2 as a derived quantity and is used as an upstream reference in KPI 233. The excerpt below highlights the core structure: a ratio between an upstream energy quantity and an occupancy/input factor, with a controlled unit conversion applied via the control panel.

```
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
( CD_ANNO_CONSUNTIVATO, CD_ANNO_PROIEZIONE, ID_KPI, DS_KPI,
  DS_TIPOLOGIA_CONSUMO, DS_TIPOLOGIA_CONTRIBUTO, COMUNE, NM_VALORE,
   DS_TIPO_VALORE )
SELECT
  YEAR(CURRENT_DATE()) - 1,
  YEAR(CURRENT_DATE()) - 1,
  230,
  'Consumo medio unita abitativa per CALDAIA TRADIZIONALE',
  'C_RISC-D',
  'Consumo medio U.A - Edificio classe energetica bassa',
  'MI',
```

```
(
  (SELECT NM_VALORE
   FROM L2.L2_F_MODELLO_TE_TD_TEM
   WHERE ID_KPI = 234
     AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1
     AND CD_ANNO_PROIEZIONE   = YEAR(CURRENT_DATE) - 1)
  /
  (SELECT VALORE
   FROM L1.L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD
   WHERE ID = 139
     AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1
     AND CD_ANNO_PROIEZIONE   = YEAR(CURRENT_DATE) - 1)
  /
  (SELECT VALUE
   FROM L1.L1_INTFC_USR_CONTROL_PANEL_MODELLO_TE_CPM_TD
   WHERE CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1
     AND DS_INPUT = 'Conversione kWh/smc'
     AND COMUNE   = 'MI')
),
'#';
```

Listing 4.42: KPI 230 – Baseline average unit consumption (excerpt, showing core ratio and conversion)

KPI 234 is another upstream baseline consumption KPI used as a reference term in the chain. The value is derived from historized manual inputs, of the heating energy consumption by distric heating, which corresponds to 26 kWh for square meter, assuming a 3 meters height of room per living unit and normalized using coefficients already materialized as intermediate KPIs in the same L2 fact table derived from CENED for the efficency based on the share per efficiency of heating and its effect on the power require for the heating.

```
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
( CD_ANNO_CONSUNTIVATO, CD_ANNO_PROIEZIONE, ID_KPI, DS_KPI,
  DS_TIPOLOGIA_CONSUMO, DS_TIPOLOGIA_CONTRIBUTO, COMUNE, NM_VALORE,
   DS_TIPO_VALORE )
SELECT
  YEAR(CURRENT_DATE()) - 1,
  YEAR(CURRENT_DATE()) - 1,
  234,
  'Consumo medio unita abitativa per TELERISCALDAMENTO',
  'C_RISC-D',
  'Consumo medio U.A - Edificio classe energetica bassa',
  'MI',
```

```
(
  (SELECT VALORE
   FROM L1.L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD
   WHERE ID = 142
     AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1
     AND CD_ANNO_PROIEZIONE   = YEAR(CURRENT_DATE) - 1)
  /
  (
    (SELECT NM_VALORE FROM L2.L2_F_MODELLO_TE_TD_TEM WHERE ID_KPI = 221 AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1)
    * (SELECT NM_VALORE FROM L2.L2_F_MODELLO_TE_TD_TEM WHERE ID_KPI = 227 AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1)
    +
    (SELECT NM_VALORE FROM L2.L2_F_MODELLO_TE_TD_TEM WHERE ID_KPI = 222 AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1)
    * (SELECT NM_VALORE FROM L2.L2_F_MODELLO_TE_TD_TEM WHERE ID_KPI = 228 AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1)
    +
    (SELECT NM_VALORE FROM L2.L2_F_MODELLO_TE_TD_TEM WHERE ID_KPI = 223 AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1)
    * (SELECT NM_VALORE FROM L2.L2_F_MODELLO_TE_TD_TEM WHERE ID_KPI = 229 AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1)
   )
),
'#';
```

Listing 4.43: KPI 234 – Baseline average unit consumption (excerpt, showing normalization by intermediate KPIs)

(3) Final KPI materialization: domestic heat pump consumption (high efficiency, KPI 194). Once intermediate KPIs are materialized for the reference year, KPI 194 is computed by selecting the required values from `L2_F_MODELLO_TE_TD_TEM` and combining them deterministically.

```
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
( CD_ANNO_CONSUNTIVATO, CD_ANNO_PROIEZIONE, ID_KPI, DS_KPI,
  DS_TIPOLOGIA_CONSUMO, DS_TIPOLOGIA_CONTRIBUTO, COMUNE, NM_VALORE, DS_TIPO_VALORE )
SELECT
  YEAR(CURRENT_DATE()) - 1,
  YEAR(CURRENT_DATE()) - 1,
  194,
  'Consumi per riscaldamento domestico (SMC/U.A.) per Pompa di calore EE - Alta efficienza',
```

```
'C_RISC-D',
'Totale consumi',
'MI',
(
  (SELECT NM_VALORE FROM L2.L2_F_MODELLO_TE_TD_TEM
   WHERE ID_KPI = 224 AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) -
 1
      AND CD_ANNO_PROIEZIONE = YEAR(CURRENT_DATE) - 1)
  *
  (SELECT NM_VALORE FROM L2.L2_F_MODELLO_TE_TD_TEM
   WHERE ID_KPI = 233 AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) -
 1
      AND CD_ANNO_PROIEZIONE = YEAR(CURRENT_DATE) - 1)
  *
  (SELECT NM_VALORE FROM L2.L2_F_MODELLO_TE_TD_TEM
   WHERE ID_KPI = 236 AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) -
 1
      AND CD_ANNO_PROIEZIONE = YEAR(CURRENT_DATE) - 1)
),
'#';
```

Listing 4.44: KPI 194 – Domestic heating consumption for high-efficiency electric heat pumps (excerpt)

(4) Projection logic for future years (KPI 218 and KPI 212) While the previous items describe the computation of domestic heating electricity consumption for the reference year (t_0 = current year − 1), the model also supports multi-year projections. This is achieved by explicitly separating the *reference year* (`CD_ANNO_CONSUNTIVATO`) from the *projection year* (`CD_ANNO_PROIEZIONE`) and by iteratively materializing future KPI values.

Definition and role of KPI 218: For domestic heating, the key driver of future electricity demand from heat pumps is the penetration of high-efficiency heat pumps in residential units. This dynamic is captured by KPI 218:

% Pompa di calore in U.A. – Alta efficienza – Domestico

For the base year, KPI 218 is directly sourced from the energy certification domain (CENED), ensuring alignment with observed building stock characteristics. The KPI is imported as-is for the reference year and stored with coincident consuntivated and projection years:

```
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
( CD_ANNO_CONSUNTIVATO , CD_ANNO_PROIEZIONE , ID_KPI , DS_KPI ,
```

```
  DS_TIPOLOGIA_CONSUMO, DS_TIPOLOGIA_CONTRIBUTO, COMUNE, NM_VALORE,
   DS_TIPO_VALORE )
SELECT
  CEN.CD_ANNO_CONSUNTIVATO,
  CEN.CD_ANNO_CONSUNTIVATO,
  218,
  '% Pompa di calore in UA Alta efficienza - Domestico',
  DS_TIPOLOGIA_CONSUMO,
  'Mix classe energetica',
  'MI',
  NM_VALORE,
  '%'
FROM L2.L2_F_CERTIFICAZIONE_ENERGETICA_EDIFICI_TD_CEN CEN
WHERE CD_INPUT_NAME = 'MIX_CLASSE_ENERGIA_POMPA_CALORE_ALTA_EFF_D'
  AND DS_TIPOLOGIA_CONSUMO = 'C_RISC-D'
  AND COMUNE = 'MILANO'
  AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1;
```

Listing 4.45: KPI 218 – Initialization for reference year from CENED (excerpt)

This step anchors the projection on an empirically observed penetration level rather than on assumptions. For future years ($t > t_0$), KPI 218 is not extrapolated independently. Instead, it is computed as the *residual share* after accounting for competing residential heating technologies. In particular, the model assumes that the total residential heating mix sums to unity, and that the heat-pump share evolves as:

$$\mathrm{KPI}_{218}(t) = 1 - (\mathrm{KPI}_{219}(t) + \mathrm{KPI}_{220}(t))$$

where:

- KPI 219 represents the projected share of alternative residential heating technologies (e.g. traditional boilers),
- KPI 220 represents additional non-electric or legacy technologies.

This logic ensures internal consistency of the technology mix over time and avoids double counting.

```
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
( CD_ANNO_CONSUNTIVATO, CD_ANNO_PROIEZIONE, ID_KPI, DS_KPI,
  DS_TIPOLOGIA_CONSUMO, DS_TIPOLOGIA_CONTRIBUTO, COMUNE, NM_VALORE,
   DS_TIPO_VALORE )
SELECT
  YEAR(CURRENT_DATE) - 1,
```

```
  <projection_year>,
  218,
  '% Pompa di calore in UA Alta efficienza - Domestico',
  'C_RISC-D',
  'Mix classe energetica',
  'MI',
  (
    1
    - (SELECT NM_VALORE FROM L2.L2_F_MODELLO_TE_TD_TEM
        WHERE ID_KPI = 219
          AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1
          AND CD_ANNO_PROIEZIONE = <projection_year>)
    - (SELECT NM_VALORE FROM L2.L2_F_MODELLO_TE_TD_TEM
        WHERE ID_KPI = 220
          AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1
          AND CD_ANNO_PROIEZIONE = <projection_year>)
  ),
  '%';
```

Listing 4.46: KPI 218 – Projection for future years (excerpt)

Definition and role of KPI 212. The Modello TE explicitly supports multi-year projections through a structured representation of technology mix dynamics. This evolution is governed by KPI 212, which captures the annual change in the penetration of high-efficiency heat pumps in residential heating.

For the reference year, the technology mix is initialized using observed data derived from the energy certification domain (CENED).

For projection years ($t > t_0$), KPI 212 is applied iteratively to propagate the technology mix forward in time. Conceptually, the penetration of high-efficiency heat pumps is updated as a function of the previous year's mix and the change factor encoded in KPI 212.

This logic is implemented by maintaining a constant `CD_ANNO_CONSUNTIVATO` (reference year) while incrementing `CD_ANNO_PROIEZIONE`, thus ensuring that all future scenarios remain anchored to a single validated baseline.

```
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
( CD_ANNO_CONSUNTIVATO,
  CD_ANNO_PROIEZIONE,
  ID_KPI,
  DS_KPI,
  DS_TIPOLOGIA_CONSUMO,
  DS_TIPOLOGIA_CONTRIBUTO,
```

```
  COMUNE,
  NM_VALORE,
  DS_TIPO_VALORE )
SELECT
  YEAR(CURRENT_DATE) - 1,
  <projection_year>,
  212,
  'Variazione mix tecnologico - Pompa di calore EE',
  'C_RISC-D',
  'Fattore di cambiamento',
  'MI',
  <computed_delta>,
  '%';
```

Listing 4.47: KPI 212 – Technology mix evolution over projection years (excerpt)

Role in the overall consumption model. KPI 218 and KPI 212 does not directly represent energy consumption. Instead, they act as a *structural multiplier* that influences downstream KPIs, including:

- KPI 212 (share of heat pumps)
- KPI 218 (share of high-efficiency heat pumps)
- KPI 194 (domestic heating consumption per residential unit)
- KPI 183 (total domestic electric heating consumption)

By isolating technology diffusion from physical consumption drivers, the model enables scenario analysis where policy levers (e.g. incentives, regulatory constraints) can be translated into measurable impacts on electricity demand without altering the underlying data model.

Electricity consumption for domestic hot water (ACS) using electric heat pumps

The calculation is organized around KPI 351, which represents the total electricity consumption for ACS, and is decomposed into three main components:

1. the number of inhabitants served by electric heat pumps (KPI 363)
2. the average electricity consumption per person (KPI 369)
3. exogenous behavioural modifiers (e.g. smart working impact)

(1) Population served by electric heat pumps (KPI 363). KPI 363 estimates the number of inhabitants whose ACS demand is met by electric heat pumps. This value is derived by combining:

- installed ACS heat-pump capacity from the CURIT domain
- demographic parameters
- correction terms accounting for technology substitution dynamics

```
SELECT
  POTENZA_ACS_PDC_EE
  *
  ( PERCENTUALE_UTILIZZO - PERCENTUALE_SOSTITUZIONE )
FROM L2_F_POTENZA_IMPIANTI_CURIT_TD_PZA;
```

Listing 4.48: KPI 363 – Inhabitants using electric heat pumps for ACS (excerpt)

This approach ensures that the ACS population is physically constrained by installed system capacity rather than purely scenario-driven assumptions.

(2) Average electricity consumption per person (KPI 369). The per-capita electricity consumption for ACS is computed from first principles using thermodynamic relationships and technical efficiency parameters. Specifically, it depends on:

- daily hot water volume per person
- annual utilization days
- temperature differential
- specific heat of water
- conversion efficiencies of boilers and electric heat pumps

```
SELECT
  ( VOLUME_ACS
    * GIORNI_UTILIZZO
    * DELTA_T
    * CALORE_SPECIFICO )
  / RENDIMENTO_CALDAIA
  / RENDIMENTO_PDC_EE;
```

Listing 4.49: KPI 369 – Per-capita electricity consumption for ACS (excerpt)

All parameters are sourced from the control panel (L1) and business input interfaces, ensuring traceability and scenario configurability.

(3) Total ACS electricity consumption (KPI 351). For each projection year, total electricity consumption for ACS is computed as the product of the population served by electric heat pumps and the corresponding per-capita electricity consumption. An optional behavioural correction factor is applied to account for changes in daily usage patterns.

```
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
SELECT
  YEAR(CURRENT_DATE) - 1          AS CD_ANNO_CONSUNTIVATO,
  <projection_year>               AS CD_ANNO_PROIEZIONE,
  351                             AS ID_KPI,
  'Consumi totali EE per Acqua calda sanitaria',
  'C_ACS',
  'Totale consumi',
  'MI',
  ( KPI_363 * KPI_369 * (1 + SMART_WORKING_FACTOR) ),
  '#';
```

Listing 4.50: KPI 351 – Total electricity consumption for ACS (excerpt)

(4) Projection logic for future years The extension of ACS electricity consumption beyond the reference year follows a *fixed-baseline, forward-projection* approach. All projections are anchored to a reference year ($t_0 = YEAR(CURRENT\_DATE) - 1$), while the projection year (t) varies along the planning horizon.

Formally, for each future year $t \in [t_0, t_0 + N]$, the model:

1. preserves the historical reference year as the unique source of observed data,
2. applies time-dependent parameters (technology mix, behavioural modifiers),
3. recomputes the resulting consumption indicators for the projection year.

This design ensures temporal consistency while preventing re-baselining effects.

Iterative computation across the planning horizon. Future-year values are generated through an explicit iteration over the projection horizon. For each projection year, total ACS electricity consumption (KPI 351) is recomputed as:

$$\mathrm{KPI}_{351}(t) = \mathrm{KPI}_{363}(t) \times \mathrm{KPI}_{369}(t) \times (1 + \Delta_{\mathrm{behaviour}}(t))$$

where:

- $\mathrm{KPI}_{363}(t)$ reflects the evolution of the population served by electric heat pumps,
- $\mathrm{KPI}_{369}(t)$ reflects changes in per-capita efficiency and usage,
- $\Delta_{\mathrm{behaviour}}(t)$ captures exogenous behavioural impacts (e.g. smart working).

All time-varying terms are sourced from business-maintained input tables, ensuring that scenario evolution is explicit and auditable.

Technological transitions (e.g. increased penetration of electric heat pumps) and behavioral variables (e.g. increase in population) are not hard-coded as trends but are introduced through dedicated KPI inputs given manually representing changes in the technology mix and population over time. These parameters are applied at each projection step, allowing gradual substitution effects to propagate consistently across years.

4.4.2. Non-heating electricity consumption

Non-heating electricity consumption covers all uses not directly related to space heating or hot water production. These components are modeled separately to reflect their distinct demand drivers and policy levers.

The PoC includes the following elements:

- Domestic electricity for lighting, appliances, and cooling
- Non-domestic electricity for lighting and cooling
- Electricity for cooking
- Electricity for productive processes

Each element is computed using certified electricity consumption KPIs and adjusted through manual factors where required to reflect scenario-specific assumptions.

Non-heating electricity consumption: cooking (C_CUCINA)

This subsection documents the computation of non-heating electricity consumption associated with cooking (domain `C_CUCINA`). Consistently with the other modules of the Modello TE, the implementation materializes a final KPI (total consumption) and a set of intermediate KPIs capturing the underlying drivers (stock size, technology mix, and unit consumption).

The final output KPI for the reference year ($t_0 = \mathrm{YEAR(CURRENT\_DATE)} - 1$) is **KPI 581** (*Consumi totali EE per Cucina*), computed as the product of:

- the number of dwellings with cooking demand (KPI 582),
- the share associated with the considered technology mix (KPI 585),
- the average electricity consumption per dwelling (KPI 587).

(1) Dwelling stock with cooking demand (KPI 582). KPI 582 estimates the total number of units where cooking demand is applicable. This metric integrates residential stock with a weighted non-residential component.

In practice, the calculation relies on previously materialized stock KPIs and a specific scaling coefficient. The logic is derived from the following components:

- **KPI 200**: Residential base (calculated as the number of users per POD minus the number of buildings).
- **KPI 278**: Number of non-domestic users per POD.
- **KPI 583**: The share (percentage) of non-domestic users equipped with a stove.

These KPIs are extracted from the table `L2_F_CONSUMI_MENSILI_TD_SUM`. The values are populated following the execution of the processing procedure `L2_F_CONSUMI_ANNUALI_MODELLO_TE_SP_EGT`.

```
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
SELECT
  YEAR(CURRENT_DATE) - 1,
  YEAR(CURRENT_DATE) - 1,
  582,
  'Num. Unita abitative con Cucina (residenziale e non)',
  'C_CUCINA',
  '# Unita abitative',
  'MI',
  ( KPI_200 + (KPI_278 * KPI_583) ),
  '#';
```

Listing 4.51: KPI 582 – Dwellings with cooking demand (excerpt)

(2) Technology-mix share applied to cooking (KPI 585). KPI 585 represents the electricity-related share applied to cooking consumption. In the current implementation, the value is inherited from an upstream KPI already computed in the model (KPI 206), ensuring coherence across domains where the same electrification driver is reused. In this case, corresponds to the share of buildings that have an electric heat pump, which means it comes from CURIT.

```sql
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
SELECT
  YEAR(CURRENT_DATE) - 1,
  YEAR(CURRENT_DATE) - 1,
  585,
  '% unita immobiliari con pompa di calore ELE per riscaldamento
   residenziale',
  'C_CUCINA',
  'Mix impianti di generazione',
  'MI',
  ( KPI_206 ),
  '%';
```

Listing 4.52: KPI 585 – Technology mix share for electric cooking (excerpt)

(3) Average electricity consumption per dwelling (KPI 587). KPI 587 is computed from a deterministic physical usage model, based on controlled parameters stored in the historized control panel (`L1_INTFC_USR_CONTROL_PANEL_MODELLO_TE_CPM_TD`). The implementation combines: rated power, daily usage time, weekly usage frequency, and annual usage weeks.

```sql
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
SELECT
  YEAR(CURRENT_DATE) - 1,
  YEAR(CURRENT_DATE) - 1,
  587,
  'Consumo medio cucina EE',
  'C_CUCINA',
  'Consumo medio per U.A',
  'MI',
  (
    PIANO_COTTURA_KW
    * (MINUTI_GIORNO / 60)
    * GIORNI_SETTIMANA
    * SETTIMANE_ANNO
  ),
  '#';
```

Listing 4.53: KPI 587 – Average electricity consumption for cooking (excerpt)

(4) Total electricity consumption for cooking (KPI 581). KPI 581 is materialized in `L2_F_MODELLO_TE_TD_TEM` for the reference year and represents the total electricity demand attributable to cooking in the city.

```
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
SELECT
  YEAR(CURRENT_DATE) - 1,
  YEAR(CURRENT_DATE) - 1,
  581,
  'Consumi totali EE per Cucina',
  'C_CUCINA',
  'Totale consumi',
  'MI',
  ( KPI_582 * KPI_585 * KPI_587 ),
  '#';
```

Listing 4.54: KPI 581 – Total electricity consumption for cooking (excerpt)

(5) Projection logic for future years (KPI 282, KPI 585 and KPI 584) The model differentiates between the *base year* and the projection years. The KPI 282 evolves following the number of housing units with a kitchen that is projected using a demographic scaling rule based on the population of Milan (stored in the `FATTORE_DI_CAMBIAMENTO` interface). The recurrence is:

$$\mathrm{KPI}_{582}(Y) = \mathrm{KPI}_{582}(Y-1) \cdot \frac{\mathrm{Pop}(Y)}{\mathrm{Pop}(Y-1)}$$

Operationally, the model reads the previous-year KPI 582 and multiplies it by the ratio of population inputs between consecutive years. This ensures that the activity driver evolves consistently with the scenario population trajectory.

An additional parameter is the technology mix projection: electrification pathway (KPI 585 and KPI 584). The share of electric cooking (induction) is projected by linear interpolation between the base-year value and a target value specified for 2050. Let $Y^* = 2050$ be the target year, and let $KPI585_{\mathrm{ee}}(Y)$ denote the electric cooking share:

$$KPI_{585}(Y) = KPI_{585}(Y_0) + (KPI_{585}(Y^*) - KPI_{585}(Y_0)) \cdot \frac{Y - Y_0}{Y^* - Y_0}$$

Demographic scaling via population-driven evolution of the activity driver (KPI 582). **Technology transition** via a linear pathway to 2050 for the electric cooking share KPI 585.**Stable technical coefficients** by holding per-unit intensities constant KPI 587. And, **scenario modifiers** through explicit multipliers (e.g., smart working impacts).

Non-heating electricity consumption: domestic electricity, lighting and cooling (C_E-D)

This subsection documents the implementation of the **domestic non-heating electricity** module (`C_E-D`), covering *electricity uses, lighting, and air conditioning.* The module produces the final KPI 400 (*Consumi totali Elettrico Domestico*), and decomposes it into a structured set of intermediate KPIs that capture:

- the dwelling stock (KPI 401),
- the penetration of air conditioning (KPI 403 and KPI 402),
- the average consumption with and without cooling (KPI 405 and KPI 404),
- the marginal consumption attributable to cooling (KPI 406),
- adoption dynamics over time (interpolation to 2050).

(1) Domestic dwelling stock (KPI 401). The number of domestic dwellings is projected over the horizon by reusing an upstream stock KPI (**KPI 200**) already materialized in `L2_F_MODELLO_TE_TD_TEM`. For each projection year t, KPI 401 is copied into the domestic electricity domain (`C_E-D`).

```
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
SELECT
  YEAR(CURRENT_DATE) - 1  AS CD_ANNO_CONSUNTIVATO,
  <projection_year>       AS CD_ANNO_PROIEZIONE,
  401,
  'Num. Unita abitative uso domestico',
  'C_E-D',
  '# Unita abitative',
  'MI',
  ( SELECT NM_VALORE
    FROM L2.L2_F_MODELLO_TE_TD_TEM
    WHERE ID_KPI = 200
      AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1
      AND CD_ANNO_PROIEZIONE = <projection_year> ),
  '#';
```

Listing 4.55: KPI 401 – Domestic dwelling stock across projection years (excerpt)

(2) Average consumption attributable to air conditioning (KPI 406). The incremental average consumption of an air conditioner (per dwelling) is sourced from the manual input interface (`L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD`) for the reference year. This ensures that the cooling intensity remains business-controlled and scenario-dependent.

```
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
SELECT
  YEAR(CURRENT_DATE) - 1,
  YEAR(CURRENT_DATE) - 1,
  406,
  'Consumo medio condizionatore per uso domestico',
  'C_E-D',
  'Consumo medio per U.A.',
  'MI',
  ( SELECT VALORE
    FROM L1.L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD
    WHERE DS_TIPOLOGIA_CONSUMO = 'C_E-D'
      AND CD_INPUT_NAME = 'CONSUMO_MEDIO_CONDIZIONATORE_PER_USO_D'
      AND CD_ANNO_CONSUNTIVATO = YEAR(CURRENT_DATE) - 1
      AND CD_ANNO_PROIEZIONE = YEAR(CURRENT_DATE) - 1 ),
  '#';
```

Listing 4.56: KPI 406 – Average air-conditioning consumption (reference year, excerpt)

(3) Penetration of air conditioning (KPI 403) and complement share (KPI 402). The share of dwellings equipped with air conditioning (KPI 403) is defined at two anchor points: the reference year (t_0) and the target year 2050. Intermediate values are generated through a linear interpolation, explicitly encoded in the procedure. The complement share (KPI 402) is derived deterministically as $1 - \mathrm{KPI}_{403}$.

```
-- Reference year value
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
SELECT YEAR(CURRENT_DATE)-1, YEAR(CURRENT_DATE)-1, 403, '% UA uso
    domestico con condizionatore',
       'C_E-D', 'Mix impianti di generazione', 'MI', IMT.VALORE, '%'
FROM L1.L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD IMT
WHERE IMT.DS_TIPOLOGIA_CONSUMO='C_E-D'
  AND IMT.CD_INPUT_NAME='UA_USO_D_CONDIZIONATORE'
  AND IMT.CD_ANNO_PROIEZIONE=YEAR(CURRENT_DATE)-1;

-- 2050 value
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
SELECT YEAR(CURRENT_DATE)-1, 2050, 403, '% UA uso domestico con
    condizionatore',
       'C_E-D', 'Mix impianti di generazione', 'MI', IMT.VALORE, '%'
FROM L1.L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD IMT
WHERE IMT.DS_TIPOLOGIA_CONSUMO='C_E-D'
```

```
  AND IMT.CD_INPUT_NAME='UA_USO_D_CONDIZIONATORE'
  AND IMT.CD_ANNO_PROIEZIONE=2050;
```

Listing 4.57: KPI 403 – Air-conditioning penetration anchored to 2050 (excerpt)

```
-- For each year t between reference year and 2050
KPI_403(t) = KPI_403(t0) + ( KPI_403(2050) - KPI_403(t0) ) * (t - t0) /
   (2050 - t0);

-- Complement share
KPI_402(t) = 1 - KPI_403(t);
```

Listing 4.58: KPI 403 – Linear interpolation and KPI 402 complement (excerpt)

(4) Average dwelling consumption without air conditioning (KPI 404). KPI 404 estimates the baseline electricity consumption of a dwelling excluding cooling. The reference-year value is derived from observed SIDES electricity consumption (domestic energy distributed) normalized by the number of domestic users, using a conservative approach based on a multi-year maximum (e.g. last 3 years). The model then removes the air-conditioning component to avoid double-counting.

```
BASE_kWh_per_user = max_{y \in \{t0-1,t0-2,t0-3\}} ( EnergiaDistribuita(
   y) / NumUtenti(y) );

KPI_404(t0) =
  ( BASE_kWh_per_user * KPI_401(t0)
    - (KPI_401(t0) * KPI_403(t0) * KPI_406(t0)) )
  / KPI_401(t0);

-- Future years keep the reference-year baseline (ceteris paribus)
KPI_404(t) = KPI_404(t0);
```

Listing 4.59: KPI 404 – Baseline consumption without cooling (conceptual excerpt)

Average dwelling consumption with air conditioning (KPI 405). The average consumption for dwellings equipped with cooling is computed by adding the marginal cooling consumption (KPI 406) to the baseline consumption (KPI 404), thereby maintaining a clear separation between structural consumption and the cooling add-on.

```
KPI_405(t) = KPI_404(t) + KPI_406(t0);
```

Listing 4.60: KPI 405 – Consumption with air conditioning (excerpt)

(5) Total domestic electricity consumption (KPI 400). Finally, the total domestic electricity consumption is computed as the sum of two cohorts:

- dwellings with air conditioning, weighted by KPI 403 and KPI 405
- dwellings without air conditioning, weighted by KPI 402 and KPI 404

```
KPI_400(t) =
  KPI_401(t) * KPI_403(t) * KPI_405(t)  +
  KPI_401(t) * KPI_402(t) * KPI_404(t);
```

Listing 4.61: KPI 400 – Total domestic electricity consumption (excerpt)

(6) Projection logic for future years (KPI 401, KPI 403 and KPI 402) The structure follows a modular decomposition:

- Number of domestic housing units
- Technology penetration: share of dwellings equipped with air conditioning
- Intensity coefficients: per-unit electricity consumption (with and without AC)
- Aggregation: weighted sum across dwelling types.

Number of domestic housing units (KPI 401). For each projection year Y, the number of domestic housing units is taken from KPI 200 (already projected elsewhere in the model):

$$U(Y) = \mathrm{KPI}_{401}(Y) = \mathrm{KPI}_{200}(Y).$$

Thus, domestic electricity demand scales proportionally to the projected housing stock.

Number of dwellings with AC KPI 408 The number of dwellings equipped with AC is:

$$\mathrm{KPI}_{408}(Y) = \mathrm{KPI}_{401}(Y) \cdot s_{\mathrm{AC}}(Y),$$

where $s_{\mathrm{AC}}(Y)$ corresponds to the grow in houses with AC given by manual input.

Air conditioning penetration (KPI 403 and 402). The share of dwellings with air conditioning, $KPI_{403}(Y)$, is projected through linear interpolation between base-year penetration and the 2050 target:

$$KPI_{403}(Y) = KPI_{403}(Y_0) + (KPI_{403}(2050) - KPI_{403}(Y_0)) \cdot \frac{Y - Y_0}{2050 - Y_0}$$

The complementary share without air conditioning is:

$$KPI_{402}(Y) = 1 - KPI_{403}(Y)$$

Intensity coefficients (constant over time) KPI 404 and KPI 406. Two per-unit electricity intensities are defined in the base year and kept constant in future years. Thus, technological intensity is assumed structurally stable, and only adoption rates vary.

$$\mathrm{KPI}_{404/406}(Y) = \mathrm{KPI}_{404/406}(Y_0)$$

4.4.3. Electricity consumption for mobility

Electricity consumption for mobility is modeled separately for private and public transport, reflecting different adoption dynamics and infrastructure constraints.

The following components are considered:

- Private electric mobility
- Public electric mobility

These components are computed by combining:

- mobility-related demand drivers
- technology penetration assumptions
- energy intensity parameters

Public transportation (C_MPUBBLICA)

The final component included in the model is public mobility (`C_MPUBBLICA`). This module estimates the total electricity consumption associated with public transportation by explicitly separating: (i) surface public transport (tram/rail in surface), (ii) metro, and (iii) rubber-based transport (electric buses).

The implementation follows a KPI chain that combines scenario-driven inputs (fleet size, electrification rate, metro expansion) with engineering-style consumption factors (kWh/km, annual km).

(1) Surface public transport (KPI 451). The electricity consumption associated with surface public transport is read as an scenario input (`CE_TRASPORTO_P_SUPERFICIE`). This design reflects the fact that surface systems can include heterogeneous technologies (tram lines, surface rail) where a bottom-up reconstruction is not always available.

```
KPI_451(t) = INPUT('CE_TRASPORTO_P_SUPERFICIE', C_MPUBBLICA, base_year);
```

Listing 4.62: KPI 451 – Surface public transport electricity consumption (excerpt)

(2) Metro electricity consumption and network expansion (KPI 452). Metro consumption is implemented in two phases:

- **Baseline (final year)**: the model reads the metro electricity consumption from `CE_TRASPORTO_P_METRO`.
- **Future years**: the baseline is scaled using a linear interpolation of a scenario-driven *network growth factor* (`%_TPL_INCREMENTO_RETE`) defined for 2050.

For each year t, the scaling factor increases proportionally to the time distance from the base year:

$$\mathrm{KPI}_{452}(t) = \mathrm{KPI}_{452}(t_0) \cdot \left(1 + g_{2050} \cdot \frac{t - t_0}{2050 - t_0}\right)$$

where g_{2050} is the long-term network expansion assumption.

```
-- baseline
KPI_452(t0) = INPUT('CE_TRASPORTO_P_METRO', C_MPUBBLICA/
    FATTORE_DI_CAMBIAMENTO, base_year);

-- future years (linear ramp up to 2050)
KPI_452(t) = KPI_452(t0) * ( 1 + INPUT('%_TPL_INCREMENTO_RETE',
    C_MPUBBLICA, 2050)
                              * (t - t0) / (2050 - t0) );
```

Listing 4.63: KPI 452 – Metro electricity consumption with network expansion (excerpt)

(3) Rubber-based public transport consumption (KPI 453). The electricity consumption for public transport **on rubber** is computed bottom-up:

$$\mathrm{KPI}_{453}(t) = \mathrm{KPI}_{455}(t) \cdot \mathrm{KPI}_{456}(t) \cdot \mathrm{KPI}_{457}(t)$$

which corresponds to: number of electric buses × (kWh per km) × (km per bus).

```
KPI_453(t) = KPI_455(t) * KPI_456(t) * KPI_457(t);
```

Listing 4.64: KPI 453 – Electricity consumption for bus transport (excerpt)

Fleet size and electrification of buses (KPI 454–455). The model starts by reading the manual input table (L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD) for total number of buses (KPI 454). Then, the number of electric buses (KPI 455) is computed as the product of total buses and the scenario-defined electrification share (%_AUTOBUS_ELETTRICI).

```
-- KPI 454: total buses
KPI_454(t) = INPUT('TOT_NUM_AUTOBUS', C_MPUBBLICA, base_year);

-- KPI 455: electric buses
KPI_455(t) = KPI_454(t) * INPUT('%_AUTOBUS_ELETTRICI', C_MPUBBLICA/
   FATTORE_DI_CAMBIAMENTO, t);
```

Listing 4.65: KPI 454–455 – Total and electric bus fleet (excerpt)

Energy intensity and activity of electric buses (KPI 456–457). To translate the electric fleet into energy consumption, the model uses KPI 456, as electricity consumption per kilometer from (C_VEICOLO_ELETTRICO_AL_KM), and KPI 457, average kilometers traveled per bus per year (TOT_KM_MEDI_PERCORSI_VEICOLI_ELE).

Both KPIs are direct scenario inputs (read from the manual interface) and are replicated across projection years.

```
KPI_456(t) = INPUT('C_VEICOLO_ELETTRICO_AL_KM', C_MPUBBLICA, base_year);
KPI_457(t) = INPUT('TOT_KM_MEDI_PERCORSI_VEICOLI_ELE', C_MPUBBLICA,
   base_year);
```

Listing 4.66: KPI 456–457 – Bus consumption factors (excerpt)

(4) Total public mobility electricity consumption (KPI 450). Finally, total electricity consumption for public transportation is computed as the sum of the three components:

$$\mathrm{KPI}_{450}(t) = \mathrm{KPI}_{451}(t) + \mathrm{KPI}_{452}(t) + \mathrm{KPI}_{453}(t)$$

```
KPI_450(t) = KPI_451(t) + KPI_452(t) + KPI_453(t);
```

Listing 4.67: KPI 450 – Total electricity consumption for public mobility (excerpt)

C_MPUBBLICA ensures that the electrification of public mobility and the potential expansion of metro infrastructure are consistently reflected in the total electricity balance, while keeping the assumptions fully transparent and auditable.

(5) Projection logic for future years (KPI 450).

4.4.4. Contribution from new developments and trends

Two additional elements complete the electricity balance:

- Electricity consumption associated with new urban developments
- Electricity production from photovoltaic systems

New urban developments are treated as incremental demand, while photovoltaic production is modeled as a negative contribution to net electricity consumption. Both elements are parameterized through manual inputs and integrated consistently into the final aggregation.

New urban developments (C_NUOVE_URBANIZZAZIONI)

The model explicitly accounts for new urban developments (`C_NUOVE_URBANIZZAZIONI`). This component captures the incremental electricity demand induced by new residential, office/school, and commercial infrastructures, and links *power requests* to *peak power accumulation* and finally to *electricity consumption*.

Unlike existing building stock, this module is fully driven by **forward-looking inputs** provided by the business, making it particularly relevant for network planning and capacity adequacy analyses.

(1) Annual electricity capacity upgrade requests (KPI 555–557). The entry point of the module is the annual requested increase in contracted electric capacity (MW), disaggregated by sector:

- **KPI 555**: residential developments
- **KPI 556**: offices and schools
- **KPI 557**: commercial activities

For each projection year t, the requested capacity is read directly from the manual input interface (`L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD`). This ensures that assumptions on urban growth remain fully scenario-controlled and adjustable by the business.

```
INSERT INTO L2.L2_F_MODELLO_TE_TD_TEM
SELECT
  YEAR(CURRENT_DATE)-1,
  <year>,
  555, -- (or 556 / 557)
  'Richieste potenziamento EL Residenziale',
```

```
  'C_NUOVE_URBANIZZAZIONI',
  'POTENZA RICHIESTA ANNUA [MW]',
  'MI',
  ( SELECT VALORE
    FROM L1.L1_INTFC_USR_INPUT_MODELLO_TE_IMT_TD
    WHERE CD_INPUT_NAME = 'RIC_POTENZIAMENTO_EL_RES'
      AND CD_ANNO_PROIEZIONE = <year> ),
  '#';
```

Listing 4.68: KPI 555–557 – Annual requested electric capacity for new developments (excerpt)

(2) Peak-to-contracted power coefficients (KPI 553–554). Requested capacity does not directly translate into network peak load. To account for simultaneity and utilization effects, two scaling coefficients are introduced:

- **KPI 553**: ratio between peak power and contracted power for residential uses
- **KPI 554**: analogous ratio for non-residential uses

These coefficients are scenario inputs and remain fully exogenous to the model.

(3) Cumulative peak power from new developments (KPI 550–552). For each sector, the model computes the **cumulative peak power** by converting annual requested power into peak power using the appropriate coefficient and accumulating the result year by year.

For the residential sector, this is expressed as:

$$\mathrm{KPI}_{550}(t) = \mathrm{KPI}_{555}(t) \cdot \mathrm{KPI}_{553}(t) + \mathrm{KPI}_{550}(t-1)$$

Equivalent formulations apply to offices/schools (KPI 551) and commercial activities (KPI 552).

```
KPI_550(t) =  KPI_555(t) * KPI_553(t)  + IFNULL(KPI_550(t-1), 0);
```

Listing 4.69: KPI 550–552 – Cumulative peak power from new developments (excerpt)

This cumulative structure reflects the irreversible nature of infrastructure expansion: once installed, capacity contributes permanently to peak demand.

(4) Electricity consumption from new developments (KPI 547–549). The final step converts cumulative peak power into annual electricity consumption by applying sector-specific conversion coefficients:

- **KPI 547**: residential new developments
- **KPI 548**: offices and schools
- **KPI 549**: commercial activities

Each KPI is computed as:

$$\text{Consumption}(t) = \frac{\text{Cumulative Peak Power}(t)}{\text{Consumption Coefficient}}$$

where the denominator is a business-defined parameter (e.g. `COEF_STIMA_CONSUMI_RES`) stored in the manual input table.

```
KPI_547(t) =  KPI_550(t)  / COEF_STIMA_CONSUMI_RES(t);
```

Listing 4.70: KPI 547–549 – Electricity consumption from new developments (excerpt)

5 | Results and Discussion

This chapter presents and discusses the results obtained from the implementation and proof of concept described in Chapter 4, evaluating them against the methodological framework and criteria defined in Chapter 3. The objective of the results analysis is to assess whether the proposed Snowflake-based data architecture and analytical workflows effectively address the research questions and design objectives of the thesis.

5.1. Evaluation of the Snowflake-Based Architecture Against Identified Requirements

Requirements elicited in Phase 1 (RQ1.2) were mapped to the concrete architectural and operational elements implemented in Phase 2 (RQ2.1–RQ2.2). This mapping is formalised in a comprehensive traceability matrix (Table A.2), which serves as the primary evaluation artefact of this chapter. The matrix enables a qualitative assessment of requirement coverage using categories such as *Covered*, *Partially covered*, and *Gap*, reflecting the current maturity of the proof of concept.

The traceability matrix supports a systematic evaluation of how architectural design decisions, such as the L0–L1–L2 layering, historisation mechanisms, role-based access control, and workload separation translate into measurable improvements over the legacy setup. The evaluation emphasises architectural properties and operational behaviour, which are more relevant for long-term sustainability and extensibility in an energy utility context.

Table 5.1 reports an excerpt of the traceability matrix, focusing on a subset of requirements that are particularly relevant for energy demand analytics and that directly address the core dimensions of RQ3.1 and RQ3.2. The full matrix is presented in Table A.2.

Table 5.1: Excerpt of the traceability matrix for key requirements related to data integration, traceability, and scalability

ID	Requirement	Model architectural element(s)	Coverage	Notes
DI09	Preservation of raw data	L0 landing tables; L1 historised copies	Covered	Raw data retained unchanged for audit and reprocessing
GS01	Role-based access control	Viewer / Analyst / Pro profiles; L2 read-only schemas	Covered	Access profiles formally defined and enforced
PS01	Interactive queries on large history	L2 marts; materialised tables	Covered	Designed to support BI and planning workloads
PS02	Independent compute scaling	Snowflake compute–storage separation; concurrent warehouses	Partially	Technical capability exists; governance guardrails required
AU07	Parameterised scenario analysis	Manual interfaces (IMT/CPM/CMT); SP_FINAL procedures	Covered	Scenario parameters already business-driven

Overall, the traceability analysis shows that the majority of the requirements critical for electricity-demand analytics are either fully covered or partially covered by the implemented architecture. In particular, substantial improvements are observed with respect to data accessibility, cross-domain integration, reuse of analytical assets, and reproducibility when compared to the legacy setup, providing a positive answer to RQ3.1. With respect to RQ3.2, the evaluation highlights clear benefits in terms of scalability and performance, enabled by Snowflake's separation of storage and compute and by workload-oriented architectural design.

5.2. Results of the Electricity Demand Model

This section presents the quantitative outputs generated by the Modello TE under the baseline scenario. Unlike previous sections, which evaluated architectural and operational properties, the focus here is on the empirical evolution of electricity demand and its structural decomposition.

To maintain readability, Table 5.2 reports selected milestone years (first two and last three of the projection horizon). The complete time series for all sub-components (2025–2051) is provided in Appendix B.

Table 5.2: Summary of electricity consumption and sub-components (selected years)

Indicator	UM	2025	2026	2048	2049	2050
Total electricity consumption	GWh	6647.95	6622.11	8302.20	8318.39	8335.03
Domestic hot water	GWh	13.09	14.39	38.03	39.02	40.25
Cooking	GWh	17.16	24.11	187.84	195.59	203.36
Domestic lighting and appliances	GWh	1401.15	1416.25	1643.45	1652.12	1660.80
Non-domestic lighting and appliances	GWh	4202.59	4232.94	4558.66	4567.78	4576.87
Lighting	GWh	36.48	36.48	36.48	36.48	36.48
Private mobility	GWh	48.77	83.26	560.34	565.83	570.50
Public mobility	GWh	304.93	311.90	372.80	374.63	376.47
New urbanisations	GWh	0.00	80.08	800.80	800.80	800.80
Productive processes	GWh	313.21	324.72	311.42	306.67	301.73
PV production	GWh	-96.66	-113.87	-492.58	-509.80	-527.01
Domestic space heating	GWh	26.85	28.23	56.08	57.61	59.55
Non-domestic space heating	GWh	180.38	183.63	228.89	231.68	235.23

5.3. Energy System Transition Results (2025–2050)

This section presents the aggregated results of the scenario-based energy demand projections developed through the Modello TE orchestration layer. The figures illustrate the structural transformation of the urban energy system between 2025 and 2050, highlighting vector substitution dynamics, electrification trends, and the progressive decline of fossil-based energy carriers.

5.3.1. Total Energy Consumption Evolution

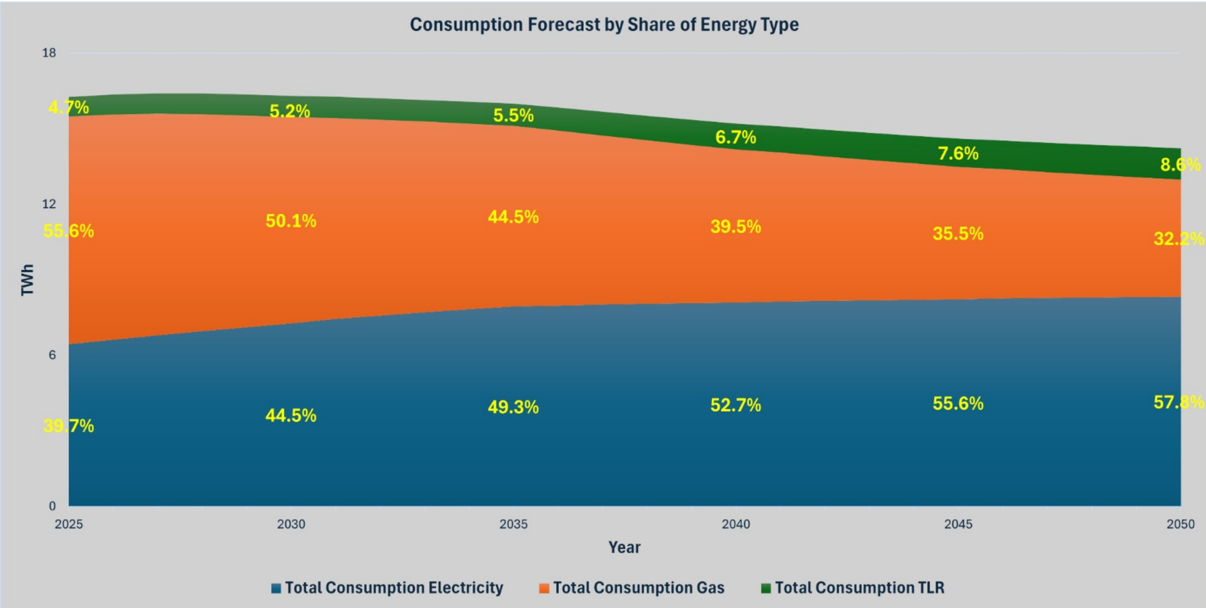

Figure 5.1: Total final energy consumption by energy carrier (TWh), 2025–2050.

Figure 5.1 illustrates the overall evolution of the urban energy mix. While total final energy demand exhibits a moderate decline over the period, the composition of energy carriers changes significantly. Electricity progressively increases its share, natural gas declines structurally, district heating expands steadily, and hydrogen emerges after 2030 as a complementary energy vector.

This pattern reflects a vector substitution dynamic rather than pure demand growth. The results are consistent with long-term decarbonisation trajectories in which electrification and network-based heating solutions progressively replace fossil-based consumption.

Between 2025 and 2050, total final energy consumption remains relatively stable in absolute terms, but its internal composition changes significantly:

- **Electricity:** increases from approximately 6.3 TWh to 8.3 TWh, corresponding to a **+32% increase**.
- **Gas:** decreases from roughly 8.9 TWh to 4.7 TWh, representing a **−47% reduction**.
- **District Heating (TLR):** increases from about 0.75 TWh to 1.24 TWh, corresponding to a **+65% growth**.

- **Hydrogen:** emerges from near-zero levels to approximately 0.20 TWh, representing structural diversification rather than substitution.

In relative terms, electricity's share increases from approximately **39.7% to 57.8%** (+18 percentage points), while gas declines from **55.6% to 32.2%** (–23 percentage points). This confirms a structural electrification pathway consistent with European decarbonization trajectories.

5.3.2. Electricity Demand Growth

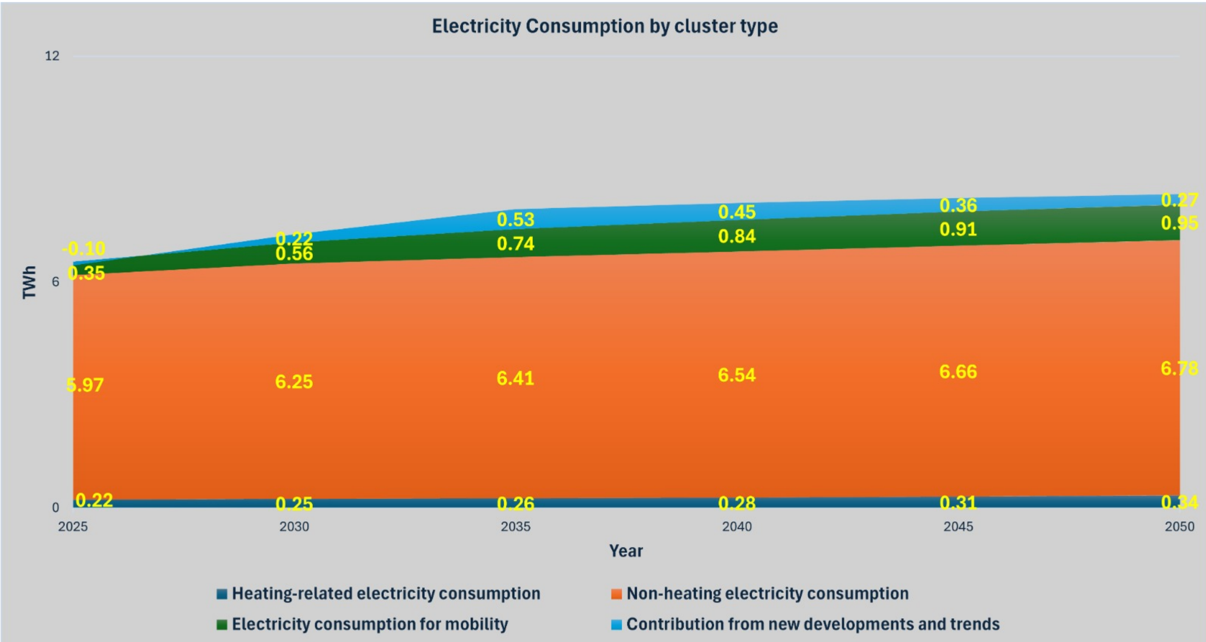

Figure 5.2: Total electricity consumption (TWh), 2025–2050.

As shown in Figure 5.2, electricity demand increases steadily across the projection horizon. The growth is driven primarily by electrification of end-uses, including heating systems, mobility, and new urban developments. Between 2025 and 2050, electricity demand increases by approximately one third, indicating a structural reorientation of the energy system toward electricity as the dominant carrier.

This trend confirms the modular decomposition demonstrated that mobility electrification and heating substitution are key contributors to long-term electricity demand growth. Electricity growth is not homogeneous across sectors. The decomposition highlights differentiated dynamics:

- **Non-heating electricity consumption:** increases from 5.97 TWh to 6.78 TWh

(+**14%**).

- **Heating-related electricity consumption:** rises from 0.22 TWh to 0.34 TWh (+**55%**).
- **Electric mobility:** grows from 0.35 TWh to 0.95 TWh (+**171%**)
- **New developments and emerging trends:** increase from - 0.10 TWh to 0.27 TWh.

The results indicate that electrification of heating and transport are the primary contributors to electricity demand growth, rather than traditional household or service consumption.

5.3.3. Natural Gas Phase-Out Dynamics

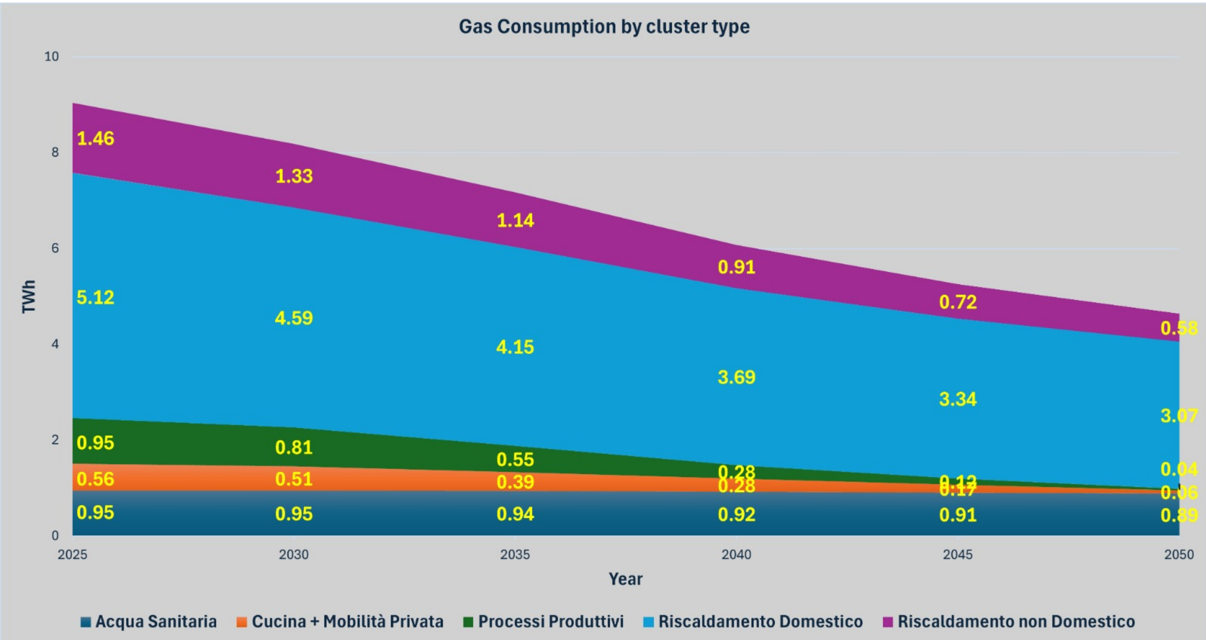

Figure 5.3: Total natural gas consumption (TWh), 2025–2050.

Figure 5.3 shows a pronounced decline in natural gas consumption over the projection period. Gas demand decreases progressively, reflecting policy-driven decarbonisation, electrification of heating systems, and efficiency improvements.

By 2050, gas consumption is reduced by approximately half compared to 2025 levels. This structural contraction confirms the expected long-term phase-out of fossil fuels within urban energy systems aligned with climate neutrality objectives. Gas consumption exhibits systematic contraction across all sectors:

- **Domestic heating:** declines from 5.12 TWh to 3.07 TWh (**–40%**).
- **Non-domestic heating:** decreases from 1.46 TWh to 0.58 TWh (**–60%**).
- **Productive processes:** drop from 0.95 TWh to 0.04 TWh (**–96%**).
- **Cooking + private mobility:** falls from 0.56 TWh to 0.06 TWh (**–89%**).
- **Sanitary hot water:** remains relatively stable, from 0.95 TWh to 0.89 TWh (**–6%**).

The near-elimination of gas in productive processes and mobility confirms strong substitution effects, while residual gas use is concentrated in thermal domestic applications.

5.3.4. District Heating Expansion

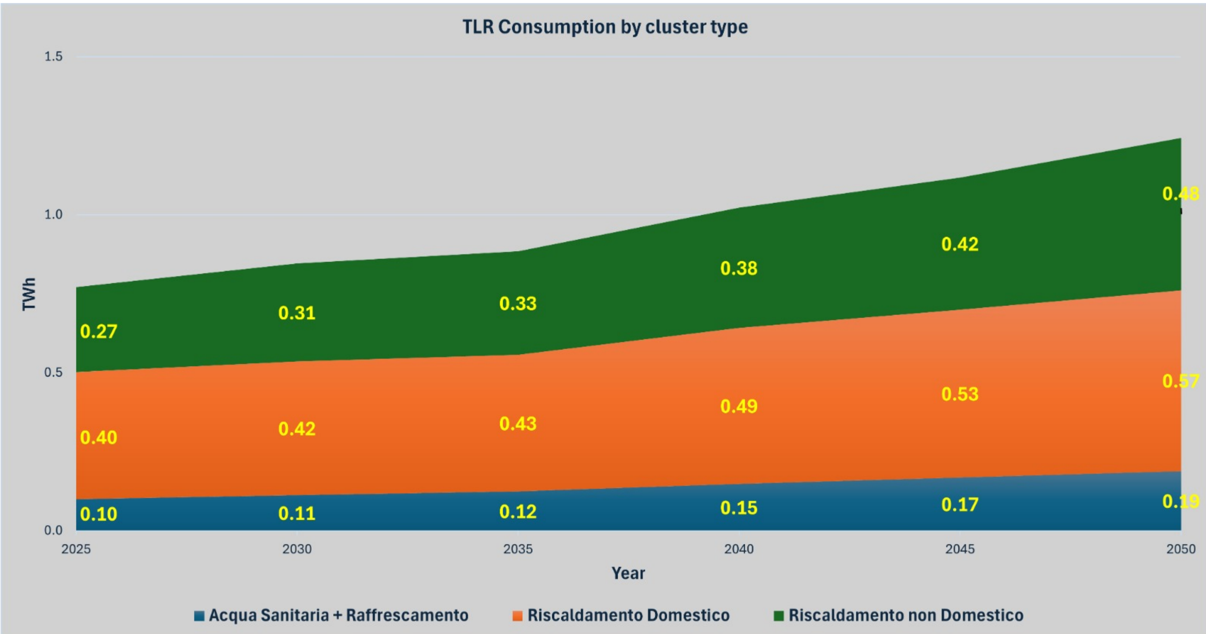

Figure 5.4: Total district heating (TLR) consumption (TWh), 2025–2050.

District heating exhibits steady expansion throughout the projection horizon (Figure 5.4). The increase reflects network-based decarbonisation strategies and gradual substitution of individual gas boilers with centralised heat supply solutions.

Although district heating does not reach the scale of electricity, its consistent growth suggests a complementary role within the broader energy transition framework. It contributes to system flexibility and facilitates integration of renewable and low-carbon heat sources. District heating shows steady expansion:

- **Domestic heating (TLR):** increases from 0.40 TWh to 0.57 TWh (+**43%**).
- **Non-domestic heating (TLR):** rises from 0.27 TWh to 0.48 TWh (+**78%**).
- **Sanitary hot water + cooling:** grows from 0.10 TWh to 0.19 TWh (+**90%**).

Although absolute values remain smaller than electricity, TLR expansion contributes to gas displacement and supports system-level decarbonization.

5.3.5. Interpretation: Structural Vector Substitution

Taken together, the four figures demonstrate that the projected energy transition is characterised by:

- A progressive electrification of final demand,
- A structural reduction of fossil gas consumption,
- Moderate but consistent expansion of district heating,
- Emergence of hydrogen as a supplementary energy carrier.

The results therefore indicate a rebalancing of the urban energy system rather than uniform growth. Electricity becomes the dominant vector by mid-century, while gas transitions toward a residual role. These dynamics are consistent with European climate-neutrality pathways and validate the planning-oriented design of the Modello TE as a scenario-support instrument rather than a black-box forecasting model.

6 | Conclusions and Future Work

6.1. Overview of Results and Link to Research Questions

The research objectives of the thesis are structured in four stages: diagnosis of the current state (Objective 1), architectural and operating-model design (Objective 2), evaluation of Snowflake's suitability for energy-demand analytics (Objective 3), and positioning of machine learning as a future development path (Objective 4). The present chapter focuses primarily on **Objective 3**, while explicitly drawing on the outcomes of Objectives 1 and 2 as contextual baselines.

Results related to Objective 1 (RQ1.1–RQ1.2) are not re-derived in this chapter, but are recalled as reference points for evaluation. In particular, the technical fragmentation of the legacy data landscape, limited data reuse, weak traceability, and the absence of standardized governance mechanisms—identified during the diagnostic phase—serve as the benchmark against which the Snowflake-based solution is assessed.

Results related to Objective 2 (RQ2.1–RQ2.2) are implicitly evaluated through the successful instantiation of the proposed architecture and operating model. The implementation choices regarding data modelling, ingestion patterns, layering, and operational workflows provide concrete evidence of the feasibility and coherence of the design decisions formulated in Chapter 2.

The core analytical focus of this chapter is therefore aligned with Objective 3, and in particular with the following research questions:

- **RQ3.1** To what extent does a Snowflake-based architecture improve data accessibility, reuse, scalability, maintainability, and reproducibility compared to the legacy setup, when applied to representative electricity-demand analytics?
- **RQ3.2** What are the main benefits and limitations observed in terms of performance, cost management, governance, and user experience when operating these workloads on Snowflake?

To address these questions, the evaluation is structured along the same dimensions defined in the methodological evaluation framework: data accessibility and integration, scalability and performance, and analytical usability. The analysis emphasizes architectural behavior, operational robustness, and the ability to support repeatable and extensible analytical workflows, key criteria for electricity-demand planning and regulatory reporting contexts.

Table 6.1: Alignment between research objectives, research questions, and results sections.

Objective	Research Questions	Results Section
Objective 1	RQ1.1 – RQ1.2	Validated through methodology Chapter 3, section 3.3, subsection 3.3.1 and described in table A.1
Objective 2	RQ2.1 – RQ2.2	Validated through implementation (Chapter 4)
Objective 3	RQ3.1 – RQ3.2	Sections 5.1

6.1.1. Modello TE as a Decision-Support Artifact

Beyond individual analytical computations, the *Modello TE* emerges from the proof of concept as a coherent decision-support artifact rather than as a black-box forecasting model. As implemented and described in Section 4.3, the Modello TE encapsulates the orchestration logic required to combine multiple analytical inputs into consistent electricity consumption and peak load projections.

The results show that the Modello TE functions primarily as a planning instrument. It enables structured scenario comparison by allowing planners to modify input parameters—such as policy assumptions, growth rates, or technology adoption trajectories—and observe their impact on final outputs without altering the underlying analytical code. This parameter-driven behaviour supports sensitivity analysis and facilitates alignment with regulatory and policy frameworks, which often require the evaluation of multiple predefined scenarios.

Overall, the results confirm that the proposed architecture and analytical workflows effectively support analytical usability and decision-making needs, providing a positive evaluation with respect to Methodology Section 5.1. The Modello TE, in particular, demonstrates how a Snowflake-based platform can serve as a robust foundation for policy-oriented electricity-demand planning, while remaining extensible for future analytical enhancements.

6.2. Positioning Against Existing Literature

Although this thesis does not aim to introduce new electricity-demand forecasting models, its results can be clearly positioned within—and differentiated from—the existing literature on energy system modelling, smart city data platforms, and modern cloud-based data architectures.

A substantial body of energy system modelling literature as seen in Chapter 1 focuses on scenario-based planning and long-term demand projections, often emphasizing methodological advances in modelling techniques, optimisation frameworks, or forecasting accuracy. These studies typically assume the availability of clean, integrated, and well-structured data, and devote limited attention to how such models are operationalised within real-world organisational and technological constraints. The results of this thesis align with the objectives of scenario-based planning, but shift the analytical focus from model formulation to the infrastructural conditions required to make such planning workflows reproducible, auditable, and scalable in practice.

Similarly, research on smart city data platforms frequently highlights the potential of integrated urban data ecosystems to support energy efficiency, mobility planning, and sustainability goals. While this literature underscores the importance of interoperability and data sharing, it often remains at a conceptual or architectural level, with limited empirical evidence of end-to-end operationalisation within regulated utility environments. The proof of concept presented in this thesis complements these contributions by demonstrating how a concrete, production-oriented data platform can support cross-domain energy analytics while respecting governance, traceability, and organisational boundaries.

With respect to modern data architectures, existing industry-oriented literature on cloud-native platforms—such as Snowflake-based lakehouse architectures—primarily discusses scalability, elasticity, and cost efficiency from a technical perspective. This thesis extends that discourse by evaluating such architectures through the lens of energy-demand analytics, emphasizing reproducibility, analytical transparency, and decision support rather than raw performance metrics.

The key distinction of this work can therefore be summarised as follows:

> Existing literature predominantly focuses on *models*. This thesis focuses on *making models operational at scale*.

By centring the analysis on data readiness, architectural design, and operational governance, the thesis contributes a complementary perspective to the literature—one that bridges the gap between theoretical modelling capabilities and their sustainable application in real-world energy planning contexts.

6.3. Limitations Revisited and Implications

The limitations identified in Section 6.3 are revisited here in light of the results obtained from the implementation and evaluation phases. Rather than restating methodological constraints, this section interprets their practical significance and discusses their implications for both utility operators and future research on data-driven energy planning.

A first limitation concerns the dependency of the results on data availability and quality. While the proposed architecture demonstrably improves integration, traceability, and reuse, its effectiveness ultimately depends on the completeness, consistency, and historical depth of the underlying source data. In the proof of concept, certain datasets required manual reconciliation or were only partially harmonised, reflecting common conditions in legacy utility environments. This limitation does not undermine the architectural validity of the solution, but highlights that data modernisation is an iterative process in which platform capabilities must be complemented by sustained data quality initiatives.

A second limitation relates to ingestion patterns and temporal resolution. The evaluation focused on batch-oriented and scheduled ingestion processes, which are representative of current planning and regulatory reporting needs. Real-time or near-real-time ingestion was not tested within the scope of this thesis. As a result, the performance characteristics of the platform under continuous streaming workloads remain unvalidated. This limitation is consistent with the methodological choice to prioritise planning-oriented analytics over operational monitoring, but it should be considered when extrapolating the results to use cases requiring real-time responsiveness.

Third, performance and scalability were evaluated primarily through architectural analysis and observed operational behaviour, rather than through systematic stress testing or benchmarking under extreme load conditions. The absence of such tests reflects the proof-of-concept nature of the implementation and the constraints of the case-study context, rather than a limitation of the underlying platform.

6.4. Future Work

The results of this thesis naturally open several directions for future research and practical development.

A first area for future work concerns empirical performance and cost evaluation. While this study assessed scalability and performance primarily through architectural analysis and observed behaviour, extended operational deployment would allow systematic benchmarking under high-load conditions and detailed cost-efficiency analysis across different workload configurations. Such studies would complement the qualitative findings presented here and provide additional guidance for large-scale utility adoption.

A second direction relates to ingestion patterns and temporal resolution. Extending the architecture to support real-time or hybrid batch–streaming data ingestion would enable investigation of use cases beyond planning and regulatory reporting, such as operational monitoring and near-real-time decision support. This would require evaluating how streaming architectures interact with the existing layering, governance, and historisation mechanisms.

A third and particularly relevant avenue for future work involves the integration of machine learning–based electricity-demand forecasting. As outlined in Objective 4, the proposed architecture provides many of the prerequisites for advanced analytics, including centralized feature availability, reproducible pipelines, and controlled execution environments. Future research could explore how tools such as Snowpark can be used to embed ML workflows within the platform, and how model lifecycle management, monitoring, and governance can be aligned with the requirements of regulated energy planning.

Finally, comparative studies across multiple utilities or regulatory contexts would help generalize the findings of this thesis. Applying the proposed framework to different organisational settings could shed light on how architectural principles must be adapted to varying data maturity levels, governance structures, and policy requirements.

In conclusion, this thesis demonstrates that data modernization is not merely a technical upgrade, but a strategic enabler of credible and scalable energy-demand analytics. By foregrounding architectural readiness and operational governance, it provides a foundation upon which advanced modelling and machine learning capabilities can be responsibly and effectively built in future work.

6.5. Conclusions

This thesis set out to investigate how a modern, Snowflake-based data architecture could support electricity-demand analytics and planning within a real-world utility context. Rather than focusing on the development or optimisation of individual forecasting models, the work addressed a more foundational question: how to design and operationalise a data and analytics platform capable of making energy-demand models usable, reproducible, and scalable in practice.

Through a structured design-science approach, the thesis first diagnosed the limitations of the existing data and analytics landscape, identifying fragmentation, limited reuse, weak traceability, and governance constraints as key barriers to effective electricity-demand forecasting and planning. Based on these findings, a Snowflake-centric architecture and operating model were designed and instantiated, using a layered L0–L1–L2 framework to separate raw ingestion, historised integration, and business-oriented analytics.

The proof of concept demonstrated that heterogeneous data sources related to energy consumption, installed power, building characteristics, and business inputs can be coherently integrated within a single platform. The resulting architecture improves data accessibility, reuse, and reproducibility, while enabling concurrent ingestion, transformation, and analytical workloads through compute isolation. Importantly, these properties emerge as systemic characteristics of the architecture rather than as outcomes of ad-hoc optimisation.

A central contribution of the thesis lies in the analytical workflows developed for electricity-demand planning. By decomposing total electricity consumption into transparent and modular sub-components and orchestrating them through the *Modello TE*, the solution supports scenario analysis, sensitivity assessment, and policy-aligned projections. This positions the platform as a decision-support system rather than as a black-box forecasting tool, aligning with European energy planning practices that prioritise transparency, auditability, and consistency across planning cycles.

The evaluation, grounded in a traceability-based framework, confirms that the proposed architecture satisfies the majority of functional and non-functional requirements identified during the diagnostic phase. Where gaps or partial coverage remain, these are primarily related to organisational governance and operational maturity rather than to technical limitations of the platform. Overall, the results provide a positive answer to the research questions associated with Objectives 2 and 3, demonstrating Snowflake's suitability as a foundation for scalable and governable electricity-demand analytics.

More broadly, the thesis contributes to the literature by shifting the analytical focus from modelling techniques to operationalisation. While existing studies largely concentrate on improving forecasting accuracy or algorithmic sophistication, this work highlights that such advances can only be sustainably leveraged when supported by robust data architecture, governance, and orchestration mechanisms.

Bibliography

[1] R. E. Al-Mamlook, M. Al Mamlook, and M. Almukhtar. Conventional models and artificial intelligence-based models for energy consumption forecasting: A review. *Renewable and Sustainable Energy Reviews*, 149:111347, 2021. doi: 10.1016/j.rser.2021.111347. URL `https://doi.org/10.1016/j.rser.2021.111347`.

[2] S. Ali, A. Ahmad, I. Khan, B. Javed, M. T. Khan, and M. M. Rahman. An insight of deep learning-based demand forecasting in smart grids: Methods, applications, and challenges. *Applied Energy*, 342:121083, 2023. doi: 10.1016/j.apenergy.2023.121083. URL `https://doi.org/10.1016/j.apenergy.2023.121083`.

[3] K. Amarasinghe, D. L. Marino, and M. Manic. Deep neural networks for energy load forecasting: Comparison and hybridization. *Energies*, 12:1587, 2019. doi: 10.3390/en12091587.

[4] V. Bianco, O. Manca, and S. Nardini. Electricity consumption forecasting in italy using linear regression models. *Energy*, 34(9):1413–1421, 2009. doi: 10.1016/j.energy.2009.06.034. URL `https://doi.org/10.1016/j.energy.2009.06.034`.

[5] J. d. J. Camacho, B. Aguirre, P. Ponce, B. Anthony, and A. Molina. Leveraging artificial intelligence to bolster the energy sector in smart cities: A literature review. *Energies*, 17(353), 2024. doi: 10.3390/en17020353.

[6] H. Chen, X. Cui, C. Chen, and D. Niu. Forecasting mechanism for energy transition in chinese cities based on configuration perspective and tcn-fecam-mtransformer. *Energy*, 329:136619, 2025. doi: 10.1016/j.energy.2025.136619.

[7] A. Gasparin, S. Lukovic, and C. Alippi. Deep learning for time series forecasting: The electric load case. *Energy*, 237:121–155, 2021. doi: 10.1016/j.energy.2021.121155.

[8] L. Hernandez, C. Baladrón, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, and J. Lloret. Short-term load forecasting for microgrids based on artificial neural networks. *Energies*, 6:1385–1408, 2013. doi: 10.3390/en6031385.

[9] G. Huang, Y. Tang, Y. Xu, Y. Su, B. Zhou, L. Zhao, J. Lin, and

Z. Ma. Research and application of power grid data integration technology based on middleware framework. pages 768 – 772. Institute of Electrical and Electronics Engineers Inc., 2024. ISBN 9798350375138. doi: 10.1109/ICEEPS62542.2024.10693033. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85207494910&doi=10.1109%2FICEEPS62542.2024.10693033&partnerID=40&md5=d6de7d7f4aa5aff5a24f6eb39b4c3933. Cited by: 0; Conference name: 3rd International Conference on Energy and Electrical Power Systems, ICEEPS 2024; Conference sponsors: IEEE; Conference location: Guangzhou; Conference code: 203117.

[10] S. Huang, Q. Hu, and Y. Li. Optimization and development methods for data asset management in power systems based on knowledge graphs. pages 411 – 416. Institute of Electrical and Electronics Engineers Inc., 2024. ISBN 9798331505905. doi: 10.1109/ISNEET64164.2024.10956034. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-105004649210&doi=10.1109%2FISNEET64164.2024.10956034&partnerID=40&md5=36015334e34eaadb5f48c837f22a94c9. Cited by: 0; Conference name: 5th International Symposium on New Energy and Electrical Technology, ISNEET 2024; Conference sponsors: IEEE; Conference location: Hangzhou; Conference code: 208357.

[11] B. Ibrahim and L. Rabelo. A deep learning approach for peak load forecasting: A case study on panama. *Energies*, 14(11):3039, 2021. doi: 10.3390/en14113039. URL https://doi.org/10.3390/en14113039.

[12] S. Jha and R. Nasr. Pioneering digital transformation of product and service delivery workflow: Global application with local customized solution for energy industry. 2025. doi: 10.2118/224347-MS. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-105006850977&doi=10.2118%2F224347-MS&partnerID=40&md5=1fdc39249087130d4a72ca2f4d3ede14. Cited by: 0; Conference name: 2025 SPE Oklahoma City Oil and Gas Symposium, OKOG 2025; Conference sponsors: ; Conference location: Oklahoma City; OK; Conference code: 208906.

[13] R. Kumar, N. Sharma, and P. Singh. Electrical load forecasting: A methodological overview. *Energy Reports*, 10:1552–1571, 2024. doi: 10.1016/j.egyr.2024.04.012. URL https://doi.org/10.1016/j.egyr.2024.04.012.

[14] F. Laborie, O. C. Røed, G. Engdahl, and A. Camp. Extracting value from data using an industrial data platform to provide a foundational digital twin. volume 2019-May. Offshore Technology Conference service@otcnet.org,

2019. ISBN 9781613990834; 9781615679713; 9781613990780; 9781613998526; 9781510824294; 9781613990797; 9781613990704; 9781613997871; 9781613990681; 9781613997079. doi: 10.4043/29576-ms. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85066633373&doi=10.4043%2F29576-ms&partnerID=40&md5=3e2bba4b9c03ae552c2f32b42c57b445. Cited by: 10; Conference name: Offshore Technology Conference 2019, OTC 2019; Conference sponsors: ; Conference location: Houston; TX; Conference code: 148087.

[15] M. A. Mat Daut, M. Y. Hassan, H. Abdullah, H. Abdul Rahman, M. P. Abdulla, and F. Hussin. Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review. *Renewable and Sustainable Energy Reviews*, 70:1108–1118, 2017. doi: 10.1016/j.rser.2016.12.015. URL https://doi.org/10.1016/j.rser.2016.12.015.

[16] K. Misiurek, T. Olkuski, and J. Zyśk. Review of methods and models for forecasting electricity consumption. *Energies*, 18(4032), 2025. doi: 10.3390/en18154032.

[17] S. Mukherjee and A. Das. Industrial data operations: Empowering the energy industry with smarter data management solutions. 2025. doi: 10.2118/224586-MS. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-105007000402&doi=10.2118%2F224586-MS&partnerID=40&md5=b91c9fa480a215d04e083a4f2cef1f76. Cited by: 1; Conference name: 2025 SPE Gas and Oil Technology Conference, GOTECH 2025; Conference sponsors: ; Conference location: Dubai City; Conference code: 208905.

[18] I. Rojek, D. Mikołajewski, K. Galas, and A. Piszcz. Advanced deep learning algorithms for energy optimization of smart cities. *Energies*, 18(407):1–19, 2025. doi: 10.3390/en18020407. URL https://doi.org/10.3390/en18020407.

[19] A. Salim, A. Giannousakis, and P. Fragkos. Decarbonisation scenarios for the european residential sector. *Energy Policy*, 180:113551, 2025. doi: 10.1016/j.enpol.2024.113551.

[20] Snowflake Inc. Snowflake key concepts and architecture, 2025. URL https://docs.snowflake.com/en/user-guide/intro-key-concepts. Accessed: 2025-11-23.

[21] Softweb Solutions. Exploring snowflake's multi-cluster shared data architecture, May 2024. URL https://www.softwebsolutions.com/resources/benefits-of-multi-cluster-warehouse-in-snowflake/. Accessed: 2025-11-23.

[22] C. Tudor, R. Sova, P. Stamatiou, V. Vlachos, and P. Polychronidou. Future-proofing EU-27 energy policies with AI: Analyzing and forecasting fossil fuel trends. *Electron-*

ics, 14(3):631, 2025. doi: 10.3390/electronics14030631. URL https://doi.org/10.3390/electronics14030631. Academic Editor: Pasquale De Falco.

[23] P. O. Ugbehe, O. E. Diemuodeke, and D. O. Aikhuele. Electricity demand forecasting methodologies and applications: a review. *Sustainable Energy Research*, 12(19), 2025. doi: 10.1186/s40807-025-00149-z.

[24] Z. Wang and F. Ren. Developing a decision support system for sustainable urban planning using machine learning-based scenario modeling. *Scientific Reports*, 15 (13210), 2025. doi: 10.1038/s41598-025-90057-5. URL https://doi.org/10.1038/s41598-025-90057-5.

[25] Z. Wang, T. Hong, H. Li, and M. A. Piette. Predicting city-scale daily electricity consumption using data-driven models. Technical report, Lawrence Berkeley National Laboratory, 2021.

[26] G. Yıldız Doğan, A. Aksoy, and N. Öztürk. A hybrid deep learning model to estimate the future electricity demand of sustainable cities. *Sustainability*, 16(15):6503, 2024. doi: 10.3390/su16156503. URL https://doi.org/10.3390/su16156503.

A | Appendix A

Table A.1: Consolidated requirements catalogue for the Snowflake-based data platform

ID	Category	Description	Source	Priority	Related RQ
DI01	Data & integration	Energy consumption data (electricity, gas, district heating) must be available in a single, harmonized schema across all points of delivery, independent of the originating source system.	System documentation; Source database SIDES	High	RQ1.1, RQ1.2
DI02	Data & integration	Each consumption record must be associated with a unique and stable identifier for the point of delivery (POD), enabling consistent joins across datasets.	Manual adjustment of number of PODs	High	RQ1.1
DI03	Data & integration	Time-series data must use a standardized temporal resolution, for this forecast we would use daily consumption given by POD	Aggregation issues simplifying in residential and non-residential	High	RQ1.2
DI04	Data & integration	Historical energy-demand data must be retained and accessible for a sufficient time horizon, to support trend analysis and future forecasting activities.	Planning and analytics requirements	High	RQ1.2

ID	Category	Description	Source	Priority	Related RQ
DI05	Data & integration	Missing, measurements must be explicitly represented and traceable.	Observed data-quality issues	Medium	RQ1.1
DI06	Data & integration	External data relevant to demand analysis (e.g. factor changes, restructuring rate, calendar effects) must be integrated centrally and aligned temporally with consumption data.	Manual enrichment in current workflows via operator or API	High	RQ1.2
DI07	Data & integration	Data transformations applied during integration must be documented and reproducible to ensure consistent results	Analyst and governance feedback	Medium	RQ3.1
DI08	Data & integration	Data ingestion processes must support updates, allowing new data to be added without reprocessing full historical datasets.	Constraints due to manual entries	Medium	RQ1.2
DI09	Data & integration	Raw source data must be preserved in its original form alongside curated and transformed versions to ensure traceability and reprocessing capability.	Audit and reproducibility requirements	Medium	RQ1.2
GS01	Governance & security	Access to energy consumption data must be controlled through role-based access mechanisms.	Internal access practices	High	RQ1.2
GS02	Governance & security	Personal and sensitive customer information must support fine-grained protection mechanisms, such as column-level masking or equivalent controls.	Data-protection and privacy guidelines	High	RQ1.2

ID	Category	Description	Source	Priority	Related RQ
GS03	Governance & security	Clear data ownership for each dataset, including responsibility for quality, documentation and lifecycle management.	Organization data-management gaps	High	RQ1.2
GS04	Governance & security	Changes to data structures and transformations must be traceable, allowing reconstruction of when, why and by whom a change was introduced.	Auditability requirements	Medium	RQ3.1
GS05	Governance & security	Data classification rules must be defined and enforced to distinguish between public, internal and restricted datasets.	Information-security standards	Medium	RQ1.2
GS06	Governance & security	The lifecycle of datasets (creation, modification, deprecation and deletion) must be governed through explicit rules and documented processes.	Data-management best practices	Medium	RQ1.2
GS07	Governance & security	Governance mechanisms must allow controlled data sharing with external stakeholders (e.g. regulators, consultants) without duplicating datasets.	Regulatory reporting needs	Low	RQ1.2
GS08	Governance & security	Security controls must ensure that data is protected during ingestion, storage and access.	IT security requirements	High	RQ1.2
PS01	Performance & scalability	The platform must support interactive analytical queries over large volumes of historical energy-demand data without requiring manual aggregation.	Performance issues in current reports	High	RQ1.1, RQ3.1

ID	Category	Description	Source	Priority	Related RQ
PS02	Performance & scalability	The platform must be able to scale compute resources independently to support concurrent analytical workloads from multiple user groups.	Concurrency limitations in legacy systems	High	RQ1.2
PS03	Performance & scalability	Performance must remain stable as data volume increases over time, without requiring structural redesign of data models or pipelines.	Expected growth in metering data	High	RQ1.2
PS04	Performance & scalability	The platform must support predictable query performance for recurring analytical workloads, such as periodic reporting and monitoring tasks.	Operational reporting requirements	Medium	RQ3.1
PS05	Performance & scalability	The platform must support isolation between workloads to prevent performance degradation caused by competing analytical processes.	Shared-system bottlenecks	Medium	RQ3.1
PS06	Performance & scalability	The platform must support isolation between workloads to prevent performance degradation caused by competing analytical processes.	Shared-system bottlenecks	Medium	RQ3.1
AU01	Analytics & usability	Analysts must be able to access curated energy-demand datasets using SQL or Python queries languages without requiring changes to ingestion or transformation pipelines.	Analyst feedback; current workflow analysis	High	RQ3.1

ID	Category	Description	Source	Priority	Related RQ
AU02	Analytics & usability	The platform must support creation of reusable analytical views for common aggregations (e.g. daily or monthly demand by area) to avoid duplication of logic across analyses.	Reporting and BI requirements	High	RQ1.2, RQ3.1
AU03	Analytics & usability	Analytical results must be reproducible over time, enabling the same query to return consistent results for the same input parameters and data version.	Audit and validation needs	Medium	RQ3.1
AU04	Analytics & usability	The platform must enable exploratory data analysis over large time-series datasets without requiring data export to local tools.	Current dependence on spreadsheets	Medium	RQ1.1
AU05	Analytics & usability	Documentation and metadata describing datasets, metrics and assumptions must be easily accessible to analysts to reduce onboarding time.	Knowledge-transfer gaps	Medium	RQ1.2
AU06	Analytics & usability	Analytical workflows must support parameterisation (e.g. time window, geography, customer segment) to enable flexible scenario and sensitivity analysis.	Planning and scenario analysis needs	Medium	RQ3.1

Table A.2: Traceability matrix: Requirements mapped to Implementation architectural elements

ID	Requirement	Model architectural element(s)	Coverage	Notes / actions
DI01	Harmonised energy consumption schema across all PODs	L2_F_CONSUMI_MENSILI_TD_SUM; SIDES integration	Partially	Harmonisation exists but must be validated across electricity, gas, district heating
DI02	Unique and stable POD identifier	SIDES POD model; L1 historised tables; L2 joins	Partially	Manual reconciliation indicates need for canonical POD definition
DI03	Standardised daily time resolution for forecasting	L2 KPI aggregation logic	Covered	Daily aggregation enforced by model logic
DI04	Multi-year historical retention	L1 historisation (dt_ini_val / dt_fin_val); validity snapshots	Covered	Explicitly supported by L1 design
DI05	Explicit representation of missing values	L1/L2 transformation procedures	Gap	Define quality flags (missing / estimated / calculated) in L2 outputs

ID	Requirement (Phase 1)	Phase 2 architectural element(s)	Coverage	Notes / actions
DI06	Central integration of external drivers	Manual interfaces (IMT/CPM); CURIT; CENED API	Covered	Business inputs and external sources already integrated
DI07	Documented and reproducible transformations	Stored procedures; Informatica catalog (L2 entities)	Partially	Procedures exist; improve documentation/versioning of logic
DI08	Incremental updates without full reloads	Scheduled ingestion from SIDES; Tasks; L2 refresh logic	Covered	Hourly/monthly incremental ingestion already implemented
DI09	Preservation of raw data	L0 landing tables; L1 historised copies	Covered	L0/L1 layering explicitly preserves raw data
GS01	Role-based access control	Viewer / Analyst / Pro profiles; L2 read-only; Datalab schemas	Covered	Profiles formally defined and enforced
GS02	Protection of sensitive data	Role separation; restricted raw access	Partially	Explicit masking policy should be formalised
GS03	Clear data ownership and stewardship	Informatica catalog (owner, field descriptions)	Covered	Mandatory for L2 publication

ID	Requirement (Phase 1)	Phase 2 architectural element(s)	Coverage	Notes / actions
GS04	Traceability of changes	Stored procedures; controlled L2 publication	Gap	Add formal change log / versioning metadata
GS05	Dataset classification rules	Role model; catalog metadata	Gap	Classification attribute (internal/restricted) not explicit
GS06	Governed dataset lifecycle	Official L2 layer; controlled publication	Partially	Deprecation and retirement rules need definition
GS07	Controlled external data sharing	Qlik exports; governed L2 datasets	Gap	Define formal sharing workflow and approval process
GS08	Security during ingestion and access	Snowflake platform baseline; access profiles	Covered	Assumed within corporate security standards
PS01	Interactive queries on large history	L2 marts; materialised tables for large volumes	Covered	Explicitly designed for BI workloads
PS02	Independent compute scaling	Snowflake lakehouse; concurrent user profiles	Partially	Technical capability exists; governance guardrails needed

ID	Requirement (Phase 1)	Phase 2 architectural element(s)	Coverage	Notes / actions
PS03	Stable performance as data grows	Layering + marts + materialisation strategy	Partially	Growth thresholds should be defined
PS04	Predictable performance for recurring jobs	Scheduled Tasks (annual/monthly KPIs)	Covered	Cron-based orchestration already defined
PS05	Workload isolation	Viewer via Qlik; Analyst L2; Pro Datalab	Partially	Formalise workload separation rules
PS06	Ingestion scalability	SIDES ingestion; L0/L1 separation	Covered	Ingestion already decoupled from analytics
AU01	SQL/Python access to curated data	Analyst access to Snowflake L2	Covered	Explicitly allowed by role model
AU02	Reusable analytical views	L2 business entities; Qlik mappings	Covered	L2 designed as reusable business layer
AU03	Reproducible analytical results	L1 historisation; deterministic procedures	Partially	Add run metadata (execution date, parameters)
AU04	Exploratory analysis without exports	Qlik Sense; Snowflake direct access	Covered	No Excel dependency required
AU05	Accessible documentation	Informatica catalog	Covered	Mandatory for L2 entities

ID	Requirement (Phase 1)	Phase 2 architectural element(s)	Coverage	Notes / actions
AU06	Foundation for future ML	L0/L1/L2 layering; Pro access	Partially	ML explicitly deferred; architecture is ready
AU07	Parameterised scenario analysis	Manual interfaces (IMT/CPM/CMT); SP_FINAL steps	Covered	Scenario parameters already business-driven

B | Appendix B

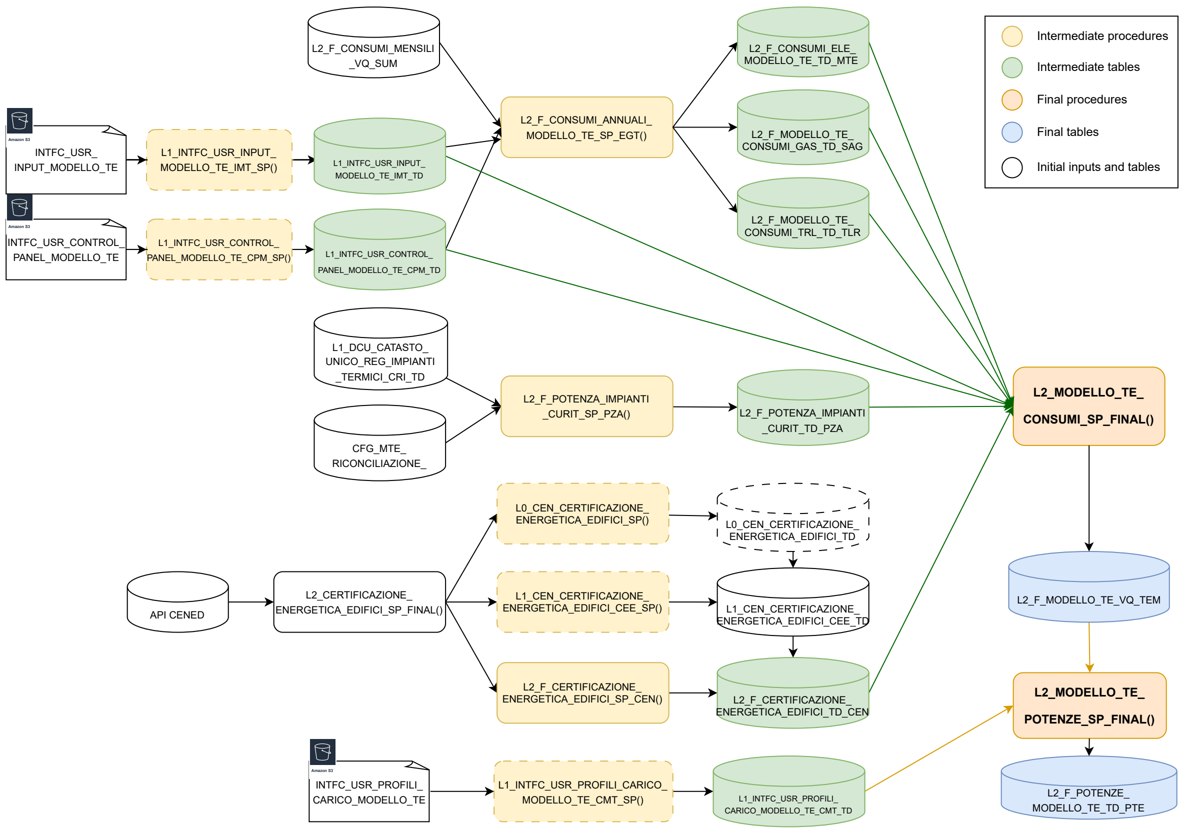

Figure B.1: Procedural and data-layer architecture of the MODELLO_TE domain in Snowflake.

Table B.1: Appendix excerpt: Electricity consumption results (selected years, 2025–2050).

Indicator	UM	2025	2027	2029	2037	2039	2041	2043	2049	2050
Total electricity consumption	GWh	6447.95	6795.95	7117.54	8023.87	8077.90	8142.25	8190.85	8318.39	8335.03
Domestic hot water	GWh	13.09	15.62	18.04	27.55	29.25	31.20	33.11	39.02	40.25
Cooking	GWh	17.16	31.15	45.42	104.18	118.98	134.17	149.25	195.59	203.36
Domestic lighting and appliances	GWh	1401.15	1430.64	1457.31	1543.98	1560.44	1580.31	1597.27	1652.12	1660.80
Non-domestic lighting and appliances	GWh	4202.59	4260.92	4310.10	4442.52	4459.64	4486.21	4504.22	4567.78	4576.87
Lighting	GWh	36.48	36.48	36.48	36.48	36.48	36.48	36.48	36.48	36.48
Private mobility	GWh	48.77	116.84	181.37	431.36	465.79	494.80	518.87	565.83	570.50
Public mobility	GWh	304.93	318.87	332.81	352.62	356.29	359.96	363.63	374.63	376.47
New urbanisations	GWh	0.00	160.16	320.32	800.80	800.80	800.80	800.80	800.80	800.80
Productive processes	GWh	313.21	340.21	357.70	349.75	342.75	336.25	330.93	306.67	301.73
PV production	GWh	-96.66	-131.09	-165.52	-303.23	-337.66	-372.09	-406.51	-509.80	-527.01
Domestic space heating	GWh	26.85	29.70	32.65	39.93	42.33	45.37	48.38	57.61	59.55
Non-domestic space heating	GWh	180.38	186.44	190.84	197.92	202.81	208.77	214.42	231.68	235.23

Listings

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Description
SIDES	Energy consumption management system for collecting, storing, and analyzing electricity, gas, and district heating data.
CURIT	Regional registry of heating systems containing technical and operational information on thermal plants and boilers.
CENED	Database managing Energy Performance Certificates (EPCs) for residential and non-residential buildings.
SAP	Enterprise resource planning system supporting finance, procurement, and asset management processes.
GIS	System for managing, analyzing, and visualizing georeferenced and spatial data.

Acknowledgements

Expreso mi profundo agradecimiento a mi madre Adiela, por darme la vida, la libertad para vivirla y por ser esa espectadora incondicional en primera fila, que anhela mi felicidad y éxito. A mi familia, aunque lejos de casa, juntos hemos llevado al mundo la esencia de Colombia y el legado de los Franco. A mi novia Nataly, por ser mi hogar lejos de casa, un refugio lleno de amor, risas y sueños compartidos. A mis amigos, que a pesar de la distancia han sido cómplices leales a través de los años. A cada persona que he conocido, pues hasta las conexiones más sutiles alteran nuestro rumbo. Finalmente, gracias a Dios por cada bendición y por permitirme reconocer Su presencia en todo lo anterior.