POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Executive Summary of the Thesis

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

Laurea Magistrale in Engineering Management

Authors: Ojeda Nicolas

Professor: Davide Chiaroni

Co-advisor: Dr. Luca Pedicone, Dr Eugenio Mazza

Academic year: 2025–2026

---

1. Introduction

The accelerating energy transition and the increasing complexity of urban energy systems require modern, scalable, and analytically robust data infrastructures. Traditional legacy data environments, characterized by fragmented architectures, manual processes, and limited scalability, are no longer sufficient to support advanced analytics, forecasting, and policy-driven decarbonization strategies.

This thesis addresses this gap by the implementation of a Snowflake-based Intelligent Data Platform to support long-term energy transition planning within a regulated utility context. The research adopts an exploratory single-case study approach and develops a governed and reproducible architecture based on a structured L0–L1–L2 framework.

2. Context and Literature Review

2.1. The Data Management Challenge

Smart utilities generate massive volumes of heterogeneous data from smart meters, IoT sensors, and SCADA systems. The literature identifies "data silos" and "information asymmetry" as primary barriers to efficiency. Traditional databases struggle to handle the variety and velocity of this data, leading to fragmented operations where critical decision-making information is locked within specific departments [1][2].

2.2. Cloud Platform Architecture

To overcome these limitations, the industry is shifting toward cloud-native architectures that decouple storage from compute. The literature highlights "Industrial Data Operations" (DataOps) and the creation of "foundational digital twins"—systems that organize raw signals around equipment and topology rather than rigid tables—as best practices. This approach supports the creation of semantic layers that standardize definitions across business domains, facilitating the transition from descriptive reporting to predictive analytics [3].

2.3. Forecasting Evolution

While statistical models (e.g., ARIMA) remain useful for stable, aggregated baselines, the literature suggests a pivot toward Machine Learning (ML) and Deep Learning (DL) to manage the non-linearity introduced by Electric Vehicles (EVs) and heat pumps [4][5]. However, a prerequisite for deploying these advanced models is a robust data infrastructure that guarantees data quality and reproducibility [6][7].

3. Methodology and Architecture Design

The proposed architecture ensures traceability, scalability, and governance by organizing data flows into three logical layers instantiated in the Snowflake environment.

3.1. The L0–L1–L2 Framework

- **L0 – Landing Layer:** Serves as the raw ingestion layer, preserving the original structure and semantics of source systems (e.g., API payloads, CSV uploads). This guarantees data immutability and full reprocessability.
- **L1 – Integration Layer:** Performs historization, validation, and harmonization of data across a unified temporal dimension. Time-travel capabilities allow analysts to reconstruct past scenarios by querying data as it existed at a given timestamp.
- **L2 – Business Layer:** Exposes certified *Facts* and *Dimensions*. It hosts the domain logic of the *Modello TE*, producing validated KPIs under strict governance rules to ensure a "single version of the truth."

3.2. Data Integration and Deterministic KPI Generation

The Proof of Concept integrates four heterogeneous but complementary data domains into a governed analytical pipeline: **SIDES** (energy consumption metering), **CURIT** (heating system registry), **CENED** (energy performance certifications via API), and **manual business inputs** (expert-defined scenario parameters historized in L1).

The architecture harmonizes them within the L0–L1–L2 framework and transforms them into certified KPIs through deterministic procedures materialized in the L2 fact table.

Three core transformation logics operationalize this integration:

- **Consumption Workflow (SIDES):** Monthly electricity, gas, and district heating data are filtered (e.g., residential vs. non-residential, voltage levels) and aggregated into standardized annual KPIs, ensuring consistent energy vector comparability.
- **Installed Power Workflow (CURIT):** Heterogeneous technology labels from the regional registry are reconciled into standardized Modello TE categories. This enables coherent calculation of installed power shares by technology and building type.
- **Certification Workflow (CENED):** API-derived building certification data are processed to compute the municipal efficiency mix (e.g., high-efficiency heat pumps vs. conventional boilers), anchoring scenario modeling to empirically observed baselines.

Manual business inputs are injected at the integration layer to govern scenario parameters (e.g., renovation rates, impact coefficients), ensuring that all analytical outputs remain traceable and reproducible.

The full orchestration of these data sources and transformation workflows within the Snowflake architecture is illustrated in Figure 3 (Appendix).

4. Implementation of the Modello TE

To demonstrate how the *Modello TE* transforms complex data transformations within the Snowflake L0–L1–L2 framework, this section details the deterministic calculation logic for one component for reading porpouses. The item is the heating-related electricity consumption, specifically focusing on high-efficiency domestic electric heat pumps.

4.1. The "Modello TE" Calculation Logic

The computation is a structured and deterministic chain of intermediate Key Performance Indicators (KPIs) materialized in the certified L2 fact table `L2_F_MODELLO_TE_TD_TEM`. This approach combines thermal demand baselines, technology mix coefficients, conversion parameters, and historized manual scenario inputs.

4.1.1 Manual Input Injection

The calculation begins by integrating expert-defined parameters historized in the L1 layer to

ensure governance and traceability of scenario assumptions.
Two primary manual inputs are injected into the model for the reference year:

- **KPI 236**: Square meters per housing unit (surface area).
- **KPI 224**: Impact factor on consumption for high energy class buildings.

These factors are stored with explicit timestamps in L1, ensuring that all assumptions are fully auditable.

4.1.2 Derived Intermediate Baseline Consumption

To establish a baseline consumption level, the model computes **KPI 233**, representing the average heating consumption per housing unit for electric heat pumps.
KPI 233 is computed as:

$$\mathrm{KPI}_{233} = \frac{\mathrm{KPI}_{230} \cdot \alpha_{139}}{\mathrm{KPI}_{235}} \cdot \gamma_{\mathrm{kWh/smc}} \tag{1}$$

Where:

- KPI_{230} = Heating consumption of traditional boiler (10 smc) - Sourced from L2, derived from district heating consumption.
- KPI_{235} = Performance Coefficient for lower class efficiency (3) - Manual input coefficient stored in L2 from L1.
- α_{139} = Efficiency (85%) - Manual input coefficient stored in L1.
- $\gamma_{\mathrm{kWh/smc}}$ = Unit conversion factor (10.69 kWh/smc) - Manual input coefficient stored in L1.

By formalizing the SQL logic into Equation (1), the deterministic nature of the Modello TE becomes explicit, demonstrating that each KPI materialized in L2 is a traceable mathematical transformation of upstream quantities and controlled parameters. This design explicitly distinguishes between:

- Operationally certified physical quantities (SIDES/CURIT/CENED),
- Business-controlled scenario assumptions (manual inputs).

4.1.3 Final KPI Materialization for the Reference Year

Once intermediate KPIs are materialized, the domestic heating electricity consumption per residential unit for high-efficiency electric heat pumps is computed as **KPI 194**.
The deterministic logic combines:

$$\mathrm{KPI}_{194} = \mathrm{KPI}_{224} \times \mathrm{KPI}_{233} \times \mathrm{KPI}_{236} \tag{2}$$

Where:

- KPI_{224} = High-efficiency impact factor (28%)
- KPI_{233} = Average heat-pump heating consumption
- KPI_{236} = Average dwelling surface area (88m2)

The resulting value is written into the certified L2 table under transactional control, ensuring consistency and atomicity of updates.

4.1.4 Projection Logic for Future Years (2025–2050)

The *Modello TE* explicitly separates the reference year, denote as t_0, corresponding to the current year minus one. From projection years in order to forecast the long-term impact of heat pump adoption.
The drivers of future electricity demand is the penetration rate of heat pumps in residential units, captured by KPI 212 and the change in high-efficiency heat-pumps over time in residential units captured by KPI 218.

Initialization (Base Year). For the reference year, KPI_{212} is obtained using the information stored in L2 (CURIT) and KPI_{218} is anchored to empirically observed data ingested from the regional energy certification API (CENED).

Future Projections ($t > t_0$). The propagation logic for KPI_{212} is materialized for each projection year through a deterministic transformation of upstream certified KPIs in the L2 layer.
For $\mathbf{KPI}_{212}(t)$ can be formalized as:

$$\mathrm{KPI}_{212}(t) = \mathrm{KPI}_{206}(t) \times \mathrm{KPI}_{200}(t) \tag{3}$$

Where:

- $\mathrm{KPI}_{212}(t)$ = Number of residential units equipped with electric heat pumps.
- $\mathrm{KPI}_{206}(t)$ = Share of residential units adopting electric heat pumps.

- $\text{KPI}_{200}(t)$ = Total number of residential units.

For $\textbf{KPI}_{218}$, the penetration rate is not extrapolated independently. It uses the penetration year and the previous year:

$$\text{KPI}_{218}(t) = \text{KPI}_{218}(t-1) + \text{TR}_{218}(t) \tag{4}$$

Where:

- $\text{KPI}_{218}(t)$ = Share of high-efficiency electric heat pumps at year t
- $\text{TR}_{218}(t)$ = Annual technological restructuration aiming for 100% by 2025

However, to ensure consistency of the total technology mix (sum equal to one), KPI 218 is simultaneously constrained by:

$$\text{KPI}_{218}(t) = 1 - (\text{KPI}_{219}(t) + \text{KPI}_{220}(t)) \tag{5}$$

Where:

- $\text{KPI}_{219}(t)$ = Share of medium efficiency heat-pumps
- $\text{KPI}_{220}(t)$ = Share of low efficiency heat-pumps

4.2. Separation of Physical Drivers and Diffusion Dynamics

A key architectural principle of the model is the explicit separation between:

- **Physical consumption drivers** (KPI 194),
- **Technology diffusion dynamics** (KPI 218, KPI 212).

This separation enables planners to simulate "what-if" scenarios by modifying diffusion parameters without altering the consumption model.

As a result, the total high-efficiency heat-pumps for domestic electric heating consumption evolves according to:

$$= \text{KPI}_{194} \times \text{KPI}_{218}(t) \times \text{KPI}_{212}(t)$$

4.3. Architectural Evaluation

The Snowflake-based architecture was evaluated using a traceability matrix to map each requirement to its corresponding architectural or operational implementation element, enabling a qualitative assessment of coverage.

The evaluation confirms that the majority of critical requirements for energy demand analytics are addressed by the implemented L0–L1–L2 framework in Snowflake. In particular:

- **Data traceability:** Raw data retention in L0 and historisation in L1 ensure auditability of all analytical outputs.
- **Governance and access control:** Role-based access profiles and controlled L2 schemas enforce secure and consistent data usage.
- **Integration and analytical readiness:** Cross-domain harmonization of operational, registry, certification, and scenario data enables unified KPI generation.
- **Scalability:** Snowflake's separation of storage and compute allows concurrent analytical workloads without impacting operational processes.

Overall, the evaluation demonstrates that the architecture substantially improves these elements compared to the legacy setup.

4.4. Energy Transition Forecasts (2025–2050)

Regarding the calculation of the heating-pump consumption computed in the implemnetation section 4. We obtained the following results at table 1:

KPI	Description	2025	2050
236	Avg. surface (m^2/unit)	88	88
224	High Efficiency (%)	28.04	28.04
233	HP base cons. (kWh)	30.62	30.62
194	HP High-eff cons. (kWh/unit)	755.87	755.87
200	House units (n.)	766635	834910
206	HP share (%)	2.94	9.43
212	HP house units (n.)	22577	78782
218	High-eff HP share (%)	59.28	100
–	**High-eff HP heating demand (GWh)**	10.12	59.55

Table 1: Evolution of high-efficiency residential heat pump KPIs (2025–2050).

4.4.1 Electricity Demand Growth

Total electricity demand increases from 6.45 TWh in 2025 to 8.34 TWh in 2050, representing a growth of approximately 29%. The expansion is primarily driven by electrification trends, particularly private mobility and new urban developments. Another remark is photovoltaic self-consumption partially mitigates net growth as see in table 2 and figure 1.

Sector	2025	2050
Domestic Hot Water	13.09	203.36
Domestic Appliances	1401.15	1660.80
Non-Domestic Appliances	4202.59	4576.87
Public Lighting	36.48	36.48
Private Mobility	48.77	570.50
Public Mobility	304.93	376.47
New Urban Developments	0.00	800.80
Industrial Processes	313.21	301.73
PV Production (Self-cons.)	-96.66	-527.01
Domestic Heating	26.85	59.55
Non-Domestic Heating	180.38	235.23
Total Demand	**6447.95**	**8335.03**

Table 2: Electricity Demand by Sector (GWh)

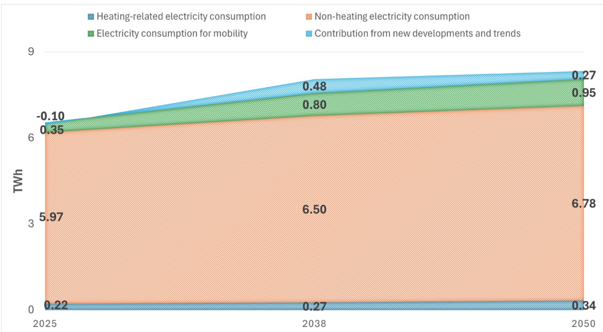

Figure 1: Total electricity consumption (TWh), 2025–2050.

4.4.2 Total Energy Consumption Evolution

As shown in Table 3 and Figure 2, the urban energy mix undergoes a structural transformation between 2025 and 2050. Electricity demand increases, while natural gas consumption declines sharply by nearly 49%. In contrast, district heating expands by over 60%, reflecting network development and fuel substitution dynamics. The relative share of electricity grows from 39.7% to 57.8% of total final energy consumption, confirming a progressive electrification of end uses. Overall, the results highlight a clear vector substitution process, with electricity emerging as the dominant energy carrier by 2050.

Energy Vector	2025	2050	Change
Electricity	6447.95	8335.03	+29.27%
Natural Gas	9042.63	4642.36	-48.67%
District Heating	771.25	1243.39	+61.21%

Table 3: Total Final Energy Consumption by Vector (GWh) (2025–2050)

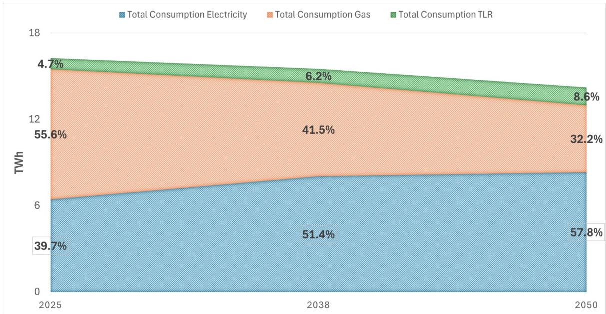

Figure 2: Total final energy consumption by energy carrier (TWh), 2025–2050.

4.5. Governance and Reproducibility

Although L2 output tables are overwritten to provide the latest forecast, full reproducibility is guaranteed because manual inputs, operational data, and transformation logic are historized in L1. Any past forecast can be deterministically reconstructed.

5. Conclusions

5.1. Strategic Contribution

This thesis demonstrates that data modernization is a strategic enabler of energy transition planning. The research addresses a fundamental challenge: establishing a governed, scalable, and reproducible data foundation capable of sustaining long-term electricity-demand analytics within a regulated utility environment.

5.1.1 Architectural Impact

The implementation confirms measurable improvements in data accessibility, cross-domain integration, and analytical reuse. Compute–storage separation enables workload isolation and scalability without compromising operational stability. Importantly, reproducibility is embedded structurally through historization and governed parameter management, ensuring that forecasts can be reconstructed deterministically over time.

5.1.2 Decision-Support Enablement

The *Modello TE* operationalizes this architecture into a transparent decision-support system. Since it operates as a transparent "white box" planning tool, enabling scenario testing and immediate impact visualization.

5.2. Limitations and Strategic Implications

Results are grounded in A2A, one company context and may not generalise directly to all energy companies.
The findings must be interpreted within the context of data maturity and its effectiveness remains dependent on the completeness source data. Data modernization therefore requires sustained governance and quality initiatives alongside platform deployment.
Furthermore, the evaluation focused on batch-oriented planning workflows rather than real-time streaming scenarios, and performance was assessed through architectural behavior and not stress testing.
These constraints reflect the proof-of-concept scope of the study.

5.3. Positioning and Forward Outlook

While existing literature predominantly concentrates on improving forecasting techniques, this thesis focuses on making such models operational at scale. The contribution lies in bridging the gap between modelling theory and company execution, specifically in Milan, Italy.
The results indicate that the architecture provides a robust foundation for future extensions, including the integration of machine learning workflows within a controlled and auditable environment. In this sense, the transition to Snowflake is not simply a technical migration, but a structural upgrade that enhances strategic readiness for the evolving energy transition landscape.

6. Acknowledgements

Agradezco 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 novia Nataly, por ser mi hogar lejos de casa, un refugio lleno de amor, risas y sueños compartidos. A mi familia, aunque lejos de casa, juntos hemos llevado al mundo la esencia de Colombia y el legado de los Franco.

References

[1] Shuai Huang, Qian Hu, and Yubao Li. "Optimization and Development Methods for Data Asset Management in Power Systems Based on Knowledge Graphs". English. In: 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. Institute of Electrical and Electronics Engineers Inc., 2024, pp. 411–416. 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.

[2] Guanbiao Huang et al. "Research and Application of Power Grid Data Integration Technology Based on Middleware Framework". English. In: 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. Institute of Electrical and Electronics Engineers Inc., 2024, pp. 768–772. 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.

[3] Francois Laborie et al. "Extracting value from data using an industrial data platform to provide a foundational digital twin". English. In: vol. 2019-May. Cited by: 10; Conference name: Offshore Technology Conference 2019, OTC

2019; Conference sponsors: ; Conference location: Houston; TX; Conference code: 148087. 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.

[4] Prosper O. Ugbehe, Ogheneruona E. Diemuodeke, and Daniel O. Aikhuele. "Electricity demand forecasting methodologies and applications: a review". In: *Sustainable Energy Research* 12.19 (2025). DOI: 10.1186/s40807-025-00149-z.

[5] Mohammad Azhar Mat Daut et al. "Building Electrical Energy Consumption Forecasting Analysis Using Conventional and Artificial Intelligence Methods: A Review". In: *Renewable and Sustainable Energy Reviews* 70 (2017), pp. 1108–1118. DOI: 10.1016/j.rser.2016.12.015. URL: https://doi.org/10.1016/j.rser.2016.12.015.

[6] Rakan E. Al-Mamlook, Mohamed Al Mamlook, and Mahmood Almukhtar. "Conventional Models and Artificial Intelligence-Based Models for Energy Consumption Forecasting: A Review". In: *Renewable and Sustainable Energy Reviews* 149 (2021), p. 111347. DOI: 10.1016/j.rser.2021.111347. URL: https://doi.org/10.1016/j.rser.2021.111347.

[7] Vincenzo Bianco, Oronzio Manca, and Sergio Nardini. "Electricity Consumption Forecasting in Italy Using Linear Regression Models". In: *Energy* 34.9 (2009), pp. 1413–1421. DOI: 10.1016/j.energy.2009.06.034. URL: https://doi.org/10.1016/j.energy.2009.06.034.

7. Appendix

∞

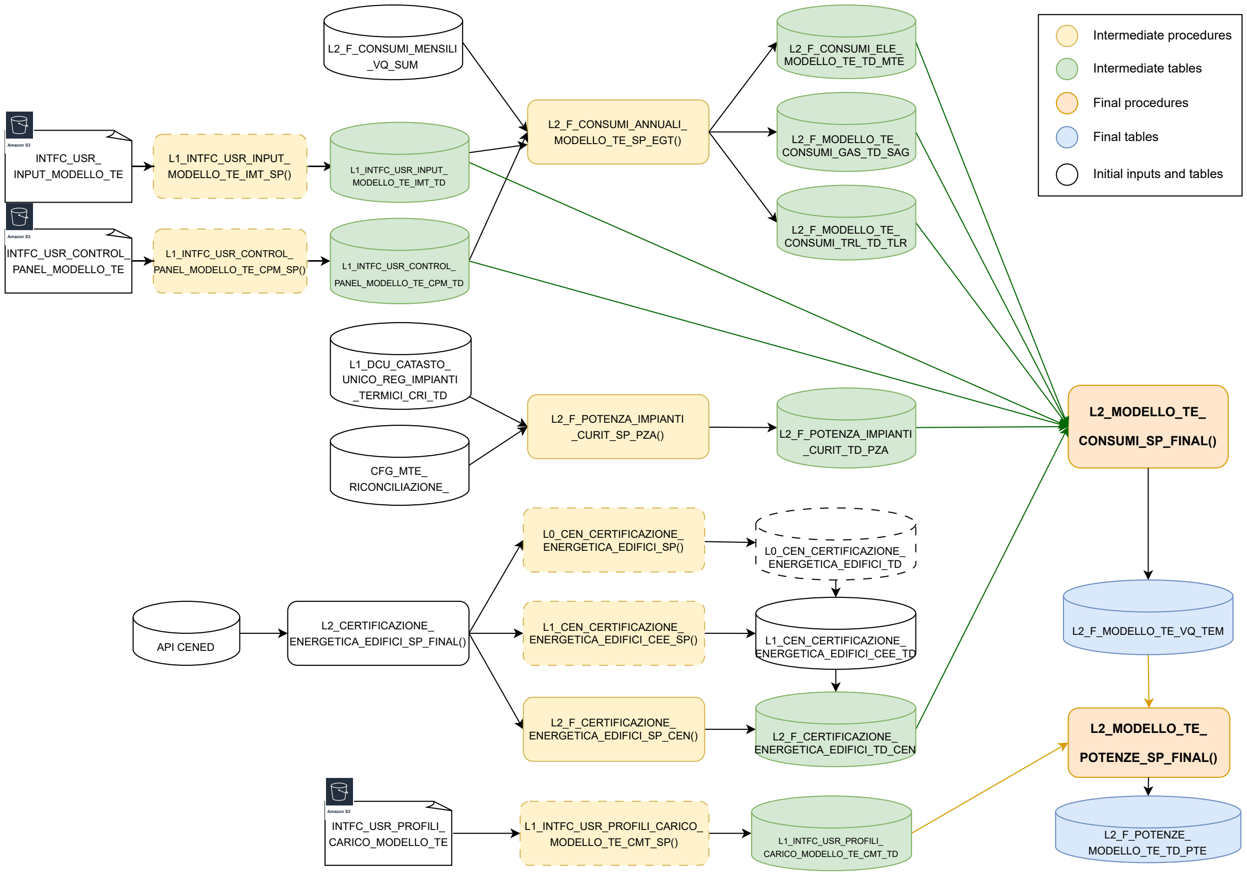

Figure 3: Procedural and data-layer architecture of the MODELLO_TE domain in Snowflake.