🎓 Capstone Project: Smart Energy Grid – Load Forecasting & Consumption Optimization

Business Scenario: Building a Secure Data Pipeline with AWS, Snowflake, Informatica Intelligent Cloud Services, and Tableau for Smart Energy Grid – Load Forecasting & Consumption Optimization

Project Overview: General

The goal of this capstone project is to design, develop, and deploy a secure data pipeline that extracts, transforms, and loads (ETL) smart meter and grid telemetry data from AWS S3 into Snowflake using Snowpipe and Snowpark (Python). The project also includes securing data using Data Masking, setting up AWS SQS for automated event notifications, and leveraging Informatica Intelligent Cloud Services (IICS) for advanced data integration. Additionally, Tableau will be used to create meaningful visualizations and insights from the data.

As a Snowflake Developer in a National Power Utility, the objective is to analyze and generate reports for better decision-making using Smart Energy data. This involves extracting, transforming, and loading data into a centralized data warehouse using IICS Mapping Tasks with Data Integration Services

Tools Utilized:

- Snowflake : For data loading and unloading and for Storage Integration using Snowpipe
- IICS : For creating Data Integration Tasks to move data from Snowflake to IICS.
- Tableau : For Report Development and visualization for getting business Insights.

Objective:

- Design and implement a robust data integration pipeline using Snowflake, IICS and Tableau to facilitate seamless data movement and transformation.
- Develop interactive dashboards and reports to visualize insights derived from Smart Meter and Grid data.
- Ensure scalability, security, versioning, availability, and reliability in handling data integration tasks through Tableau.
- Optimize data models and configurations in Tableau for enhanced performance and resource utilization.
- Document the entire data integration and visualization process for future reference and analysis in a PowerPoint template shared by LTIMindtree SPOC on Day 3.

Assumptions:

- The project team consists of 4-5 associates proficient in Snowflake, IICS and Tableau and other required tools.
- Access to necessary Tableau environments and datasets is provided.

• Moderate to advance knowledge of data integration concepts and data visualization techniques is assumed for all team members.

Expectations:

• Each use case must be error free and bug free.
• Design test cases wherever possible and prepare test case passing reports.
• Implement appropriate validation wherever necessary.
• Adoption of Agile practice in the end-end lifecycle.
• Identify unique prepositions for the project and highlight it in the presentation.
• Implementation of non-functional parameters like Security, Availability, Scalability is mandatory.
• Very careful design of appropriate Data models with Normalization must be done and
• presented.
• Design all the required Snowflake, IICS Tasks and Tableau Visuals with Error-Free with Robust design to handle Errors in Data Pipeline
• Identify the Unique Selling Proposition (USP) for Energy Analytics
• To follow AGILE practices and create Scrum Board and Burn Down Chart etc.
• Project completion within 4 days, followed by 2 additional days for enhancements, presentations, and further extensions.
• Documentation of all project components, including code, configurations, and insights.
• Regular communication and collaboration among team members, SMEs, and stakeholders.
• Adherence to best practices for data integration, visualization, and optimization.

Use Cases:

1. Data Integration: Designing and implementing data integration workflows in Tableau
   to extract, transform, and load Smart Energy data.
2. Data Visualization: Creating interactive dashboards and reports in Tableau to visualize Smart Meter and Grid data to analyze trends and patterns.
3. Scalability and Reliability: Ensuring scalability and reliability in handling data integration tasks through Tableau.
4. Optimization: Optimizing data models and configurations in Tableau for improved performance and resource utilization.
5. Security Measures: Implementing security measures in Tableau to protect sensitive data and ensure compliance with regulations
   Submission Guidelines:
   • Regular updates on project progress and challenges.
   • Documentation of Tableau reports, dashboards, and insights.
   • Presentation of key findings and recommendations.
   • Demonstration of data integration workflows and visualization techniques.
   • Collaborative review and feedback sessions.

🎓 Smart Energy Grid – Load Forecasting & Consumption Optimization:

Technical Specification

🏢 Case Study Context:

You are part of the Grid Analytics Division at EnerNex Utilities, operating across multiple regions with millions of smart meters and dozens of substations. With operations spanning urban and rural zones, the company manages a complex network of feeders, transformers, and renewable generation assets.

Business leadership is seeking to implement a centralized Energy Intelligence Platform to monitor real-time consumption, track substation load, assess renewable contribution, and improve planning. You are tasked with building this platform using:

- Snowflake for modern cloud data warehousing and automation
- IICS and Snowpipe for ingesting structured and semi-structured data
- Tableau for powerful, interactive visual analytics

🎯 Business Objectives

- Unify energy datasets from diverse systems
- Create a robust, automated data pipeline with incremental updates
- Enable secure sharing of reports with regional operations teams
- Apply enterprise-grade governance (RLS, masking, tags)
- Build executive-ready Tableau dashboards for monitoring and decision-making

📁 Input Datasets (All Provided as .csv):

Dataset Name	Description	Source
SmartMeter_Readings.csv	Hourly household consumption readings	Streaming to S3
Substation_Load.csv	Substation-wise transformer load and utilization	MySQL (via IICS)
Weather_Data.csv	Weather metrics (temperature, humidity, wind, irradiance)	S3 Bucket
Renewable_Generation.csv	Solar/wind farm generation summaries	Manual Upload
Customer_Profile.csv	Masked customer connection profile	Static Table
Region_Hierarchy.csv	Geographic mapping of country, state, zone	CSV Upload

Project Phases and Detailed Tasks

PHASE 1: 🔄 Data Acquisition and Loading

Task 1: Setup Snowflake Environment
- Create warehouse, database, schema
- Setup users, roles, and assign privileges
- Configure resource monitors for cost control

Task 2: Load Flat Files via SnowSQL
- Use PUT and COPY INTO for:
  - Renewable_Generation.csv
  - Region_Hierarchy.csv
  - Customer_Profile.csv
- Create matching staging tables with proper data types

Task 3: MySQL → Snowflake via IICS
- Use IICS to extract from substation_load (MySQL)
- Transform and load into stg_substation_load in Snowflake
- Schedule using a Taskflow

Task 4: Real-Time Meter Ingestion via Snowpipe
- Define External Stage pointing to S3 bucket
- Create File Format (CSV, detect date types)
- Create Pipe and enable Snowpipe auto-ingestion from SmartMeter_Readings.csv
- Test with 2–3 new meter drops into S3

PHASE 2: 📐 Data Modeling and Star Schema

Task 5: Build Dimension Tables
- DimCustomer – CustomerID, CustomerEmail, Phone
- DimRegion – Zone, Country, Region, RegionHierarchy
- DimWeather – WeatherDate, Temperature, Humidity, WindSpeed, SolarIndex

Task 6: Build Fact Tables
- FactConsumption: joins DimCustomer, DimRegion, DimWeather
- FactLoad: substation transformer load and utilization records
- FactGeneration: renewable generation (solar/wind) records

Task 7: Apply Surrogate Keys and Constraints
- Implement date dimensions, referential integrity
- Use MERGE statements for upserts

PHASE 3: ⚙️ Incremental Automation Using Streams, Tasks, and MVs

Task 8: Stream Setup
- Create STREAM objects on:
  - stg_substation_load
  - stg_smartmeter_readings
- Monitor changes using METADATA$ACTION

Task 9: Task Automation
- Create hourly TASK chains to:
  - Insert new readings to FactConsumption
  - Update latest transformer load in FactLoad
  - Log changes into audit_log table

Task 10: Materialized Views for Performance
- mv_load_forecast_accuracy: aggregated forecast vs actual by region
- mv_renewable_contribution: avg renewable share by day/region
- mv_overload_watchlist: highlight transformers over threshold

PHASE 4: 🔐 Security, Governance and Sharing

Task 11: Data Masking and RLS
- Apply MASKING POLICY to:
  - DimCustomer.customer_email
  - DimCustomer.phone
- Implement ROW ACCESS POLICY to restrict access to FactConsumption based on user's region

Task 12: Secure Data Sharing
- Create SECURE SHARE:
  - Includes only selected fields from mv_load_forecast_accuracy and mv_overload_watchlist
- Test access from reader account
- Document sharing process

PHASE 5: 📊 Visualization Using Tableau

Task 13: Data Connection
• Connect Tableau to Snowflake using live/extract mode
• Build calculated fields and data model within Tableau

Task 14: Design Dashboards
Dashboard Visuals & Features
Energy Flow Overview Sankey or flow chart: generation → substation → consumer
Forecast Accuracy Line + bar chart: forecast vs actual by region
Load Alert Heatmap: substation/transformer vs utilization threshold
Renewable Monitoring Scatter plot: day vs renewable share
Demand-Supply Match Gantt + stacked bar: region-wise fulfillment %

Task 15: Interactivity Features
• Region/Time filters
• Drilldowns (Country → State → Substation)
• Alerts for KPI deviations
• Parameter-based views (e.g., monthly vs weekly)

PHASE 6: 📝 Documentation and Submission

Task 16: Reporting
• Project architecture diagram (Snowflake + IICS + Tableau)
• ER Diagram of final schema
• Data flow pipelines
• Masking/RLS/Share configuration proofs
• SnowSQL scripts, Tableau workbook, CSVs

Deliverables

Type	Description
SQL Scripts	DDL, Stream, Task, Policies
Data Files	Cleaned .csv source files
Tableau Dashboards	.twbx or PDF exports
Documentation	Word or PDF project report
Screenshot Proofs	For shares, security, ingestion, etc.