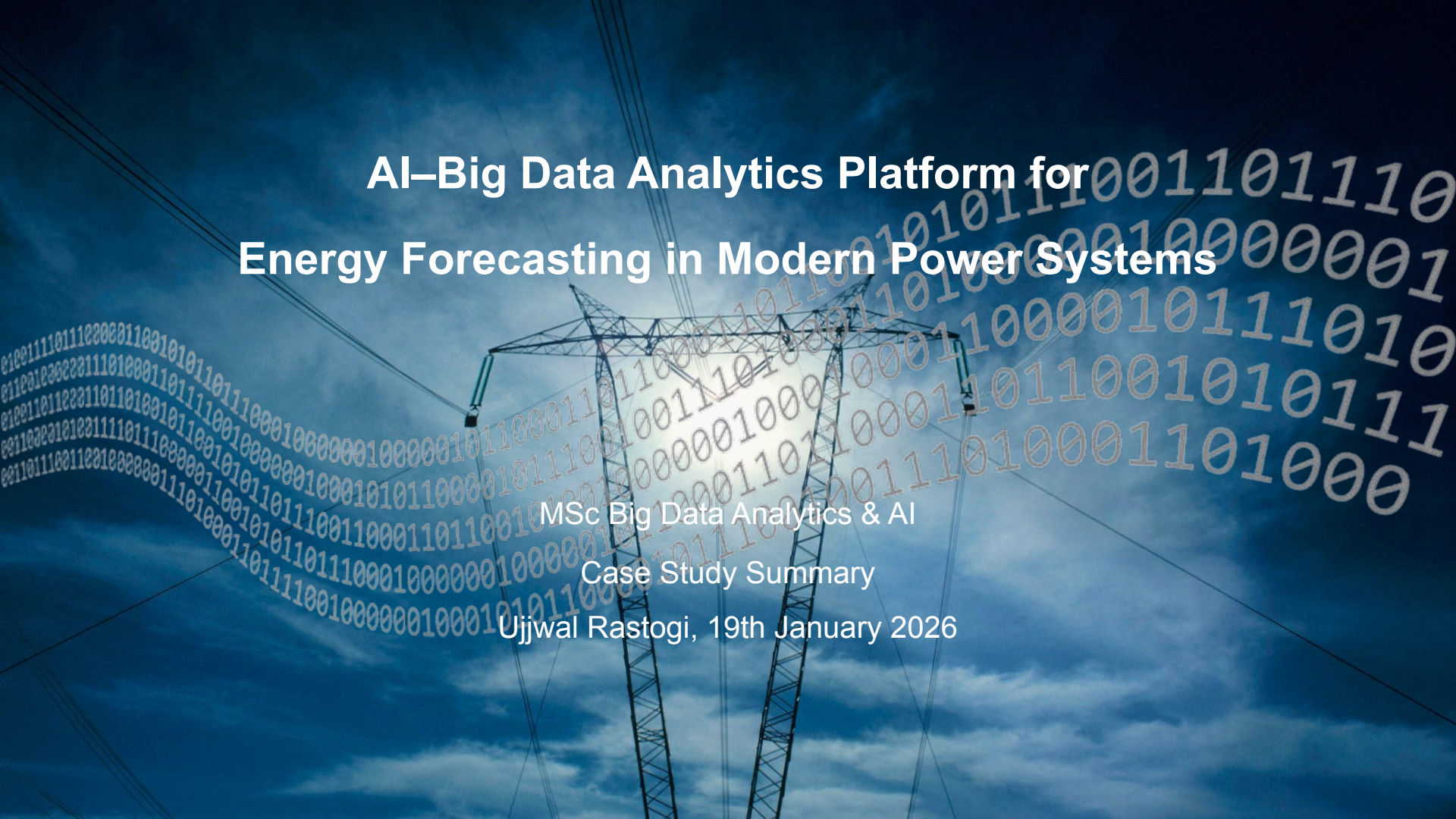


AI–Big Data Analytics Platform for Energy Forecasting in Modern Power Systems

MSc Big Data Analytics & AI

Case Study Summary

Ujjwal Rastogi, 19th January 2026

Introduction

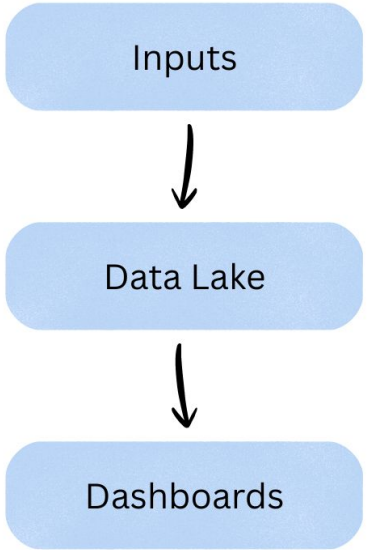
- Modern power systems generate massive volumes of data (smart meters, sensors).
- Big Data technologies enable efficient storage, processing, and advanced analytics.
- Focus: Energy/price forecasting to support planning, risk mitigation, and decision-making.

Energy Forecasting Challenges

- Electricity prices are volatile and influenced by:
  - Demand
  - Generation mix
  - Fuel prices
  - Grid congestion
- Accurate forecasts support operational planning, market participation, risk mitigation.

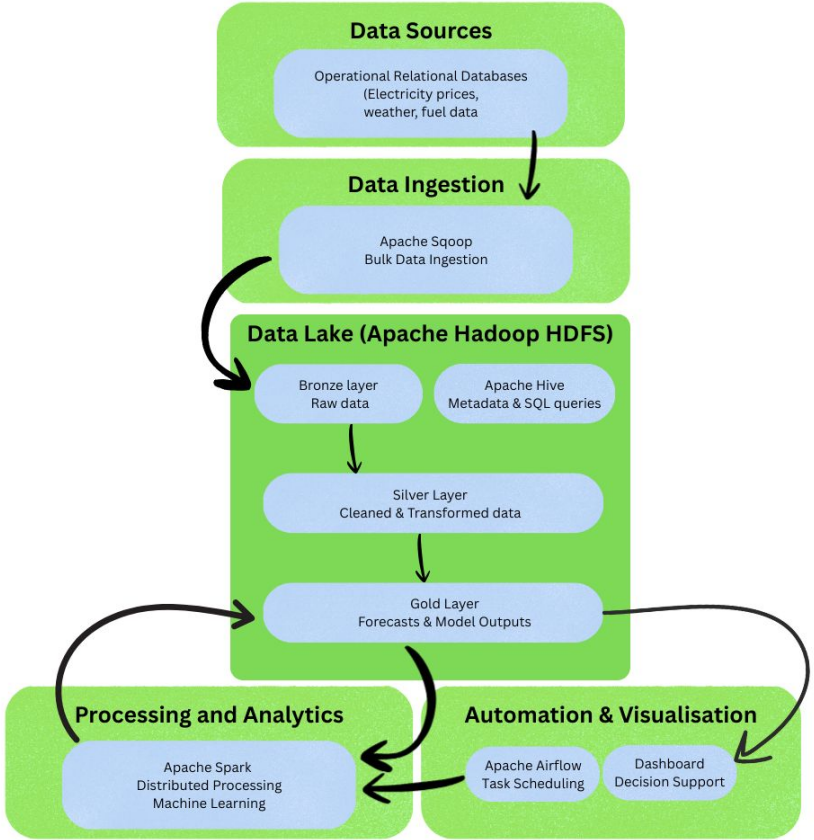
Case Study Overview

- Platform enables automated execution of forecasting models on historical & operational data.
- Open-source, on-premises →cost-effective and secure.
- Data lake handles structured, semi-structured, unstructured data.
- Dashboards visualize results for decision-making.



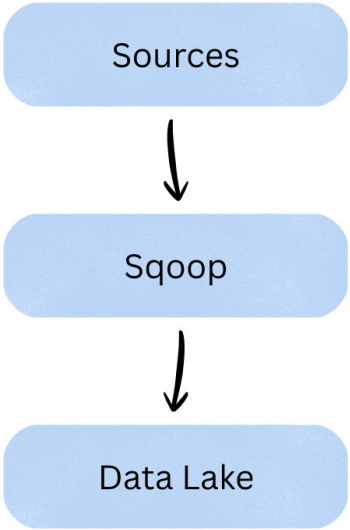
Big Data Architecture

- **Layered Data Lake Design:**
  - Bronze: Raw data
  - Silver: Cleaned & transformed data
  - Gold: Analytics-ready outputs & ML results
- Technologies: Hadoop/HDFS, Hive, Spark, Airflow



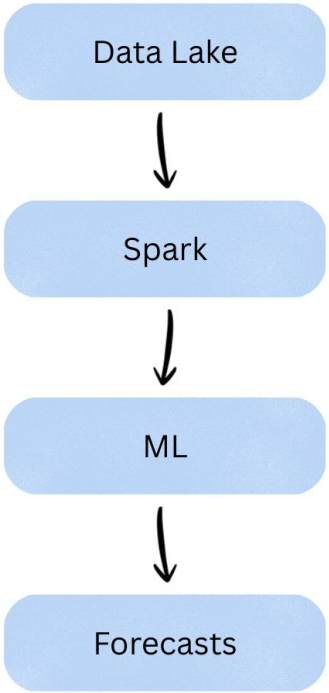
Data Sources & Ingestion

- Data sources: Historical prices, market components, weather, fuel prices
- **Apache Sqoop** → efficient, parallel ingestion into Data Lake
- Ensures datasets are consistently integrated for analytics



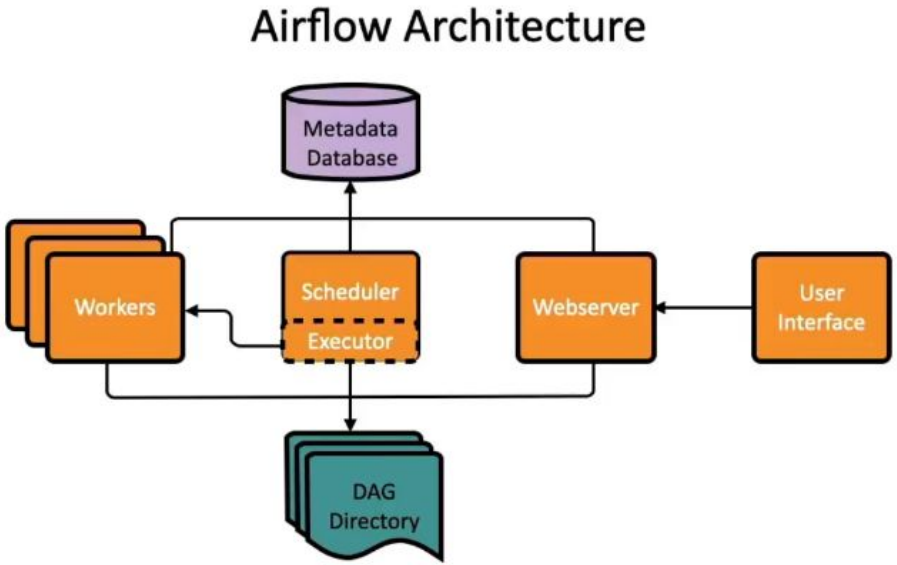
Processing & Analytics

- **Apache Spark** handles:
  - Large-scale transformations
  - Feature engineering
  - Execution of ML models
- Parallel execution allows region-specific forecasts
- ML models integrated into automated pipeline



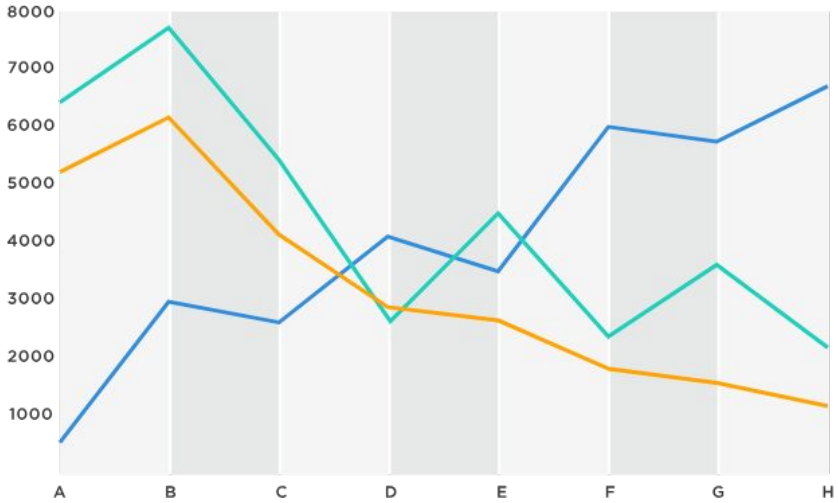
Automation & Orchestration

- **Apache Airflow** manages:
  - Data ingestion
  - Preprocessing
  - Model training
  - Forecast generation
- Scheduled forecasts: 24h, 72h, 120h intervals
- Supports near real-time analytics



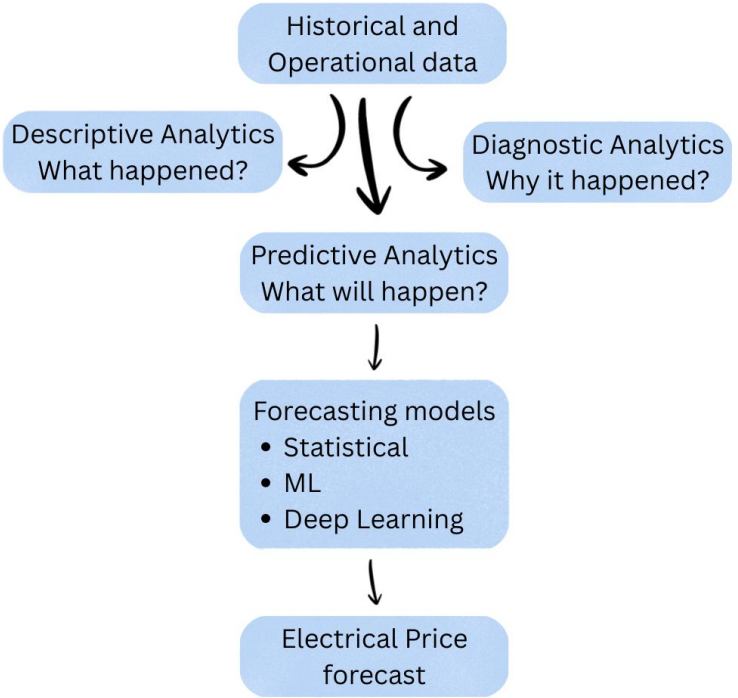
Visualization & Decision Support

- Forecasts stored in **Gold Layer**
- Dashboards compare actual vs predicted prices
- Insights accessible and actionable for stakeholders



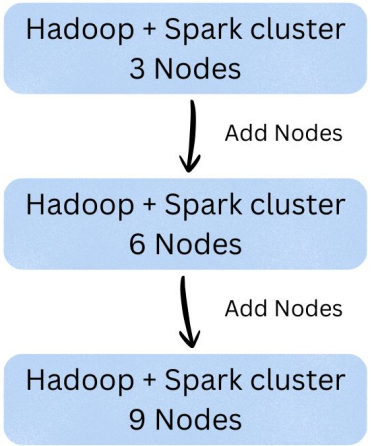
Analytics & Machine Learning

- Supports:
  - Descriptive & diagnostic analytics
  - Predictive analytics (focus)
- Models evaluated: Statistical, ML, Deep Learning
- Focus: Operationalization within Big Data platform, not algorithmic novelty



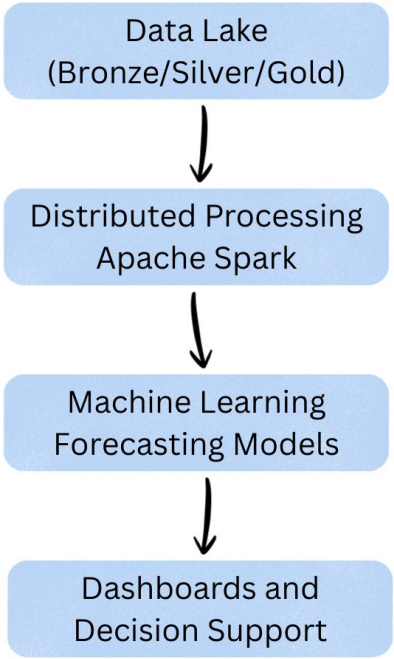
System Scalability & Design Choices

- Open-source, on-premises → lower costs & infrastructure control
- Horizontal scalability: add nodes to Hadoop/Spark clusters
- Flexible architecture for future use cases:
  - Load forecasting
  - Renewable generation forecasting
  - Asset health monitoring



Conclusion

- Layered architecture + distributed processing + ML –> scalable, automated, cost-effective analytics
- Architecture easy to extend
- ML integration shows Big Data delivering operational value
- Platform: practical approach to Big Data analytics in energy systems



Thank you