CAPSTONE MARKETS

[LinkedIn]

Table of Contents

Introduction

In today’s fast-paced financial markets, data-driven decision-making is no longer optional — it’s essential. Our team set out to build a fully online, cloud-native trading pipeline that integrates real-time data, advanced analytics, and automated reporting. Powered by Azure services, this system is designed to be scalable, resilient, and intelligent.

Mission Objectives

My project is guided by four core objectives:

1. **Deliver Real-Time News and Sentiment Reports** Provide traders with curated market news and sentiment analysis to support timely, informed decisions.
2. **Operate Fully Online** Run all operations via scalable, cloud-based platforms for seamless global access and performance tracking.
3. **Integrate Advanced Analytics and Technology** Leverage real-time data and machine learning to support smarter trading strategies.
4. **Drive Continuous Innovation** Explore new models, data sources, and algorithmic techniques to stay ahead in dynamic markets.

Data Sources

Source	Type	Structure	**Mode**
Real-Time Market Data	Streaming	Semi-Structured	Real-Time
Historical Market Data	Batch	Structured	Scheduled
News & Sentiment Data	Mixed	Unstructured/Semi-Structured	Streaming + Batch
Trade Execution Logs	Batch	Structured	Scheduled
Regulatory Data	Batch	Semi-Structured	Scheduled

Data Sinks

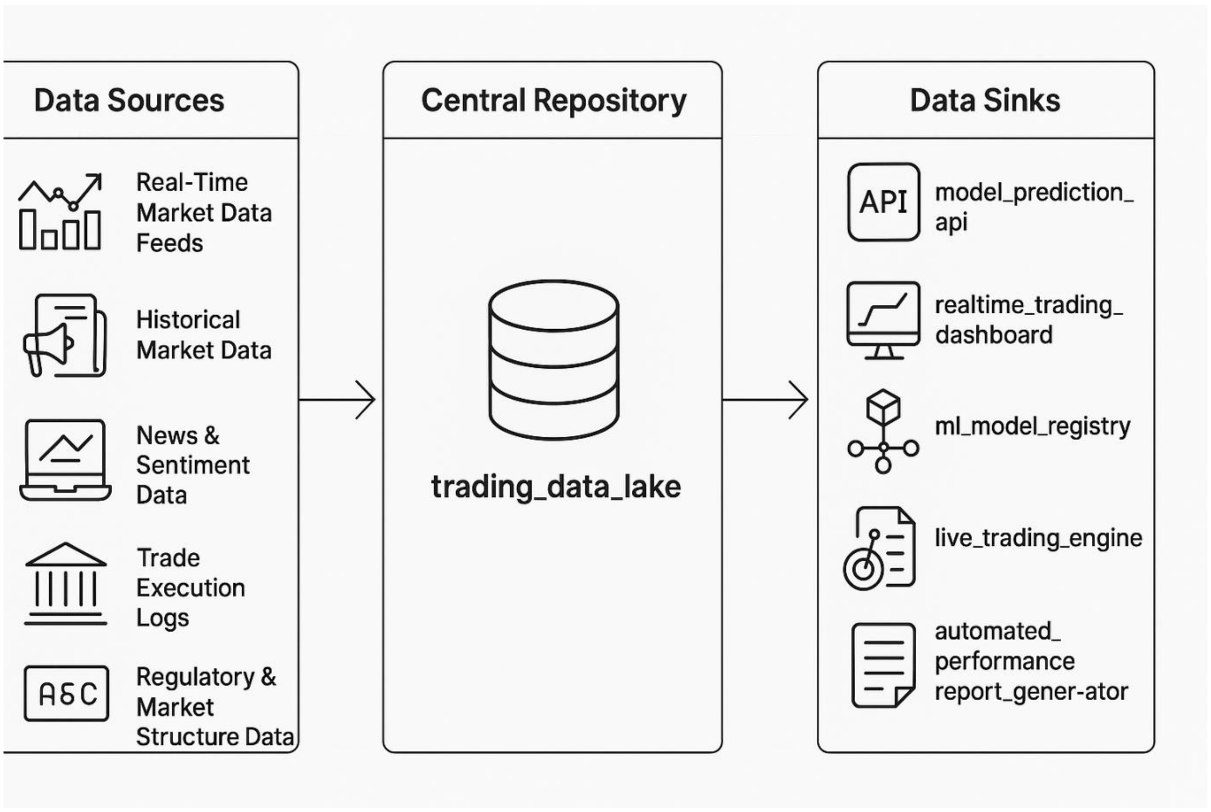
My processed data flows into five key destinations:

- **Model Prediction API**: Serves real-time model outputs from the silver layer
- **Power BI – Real-Time Dashboard**: Visualizes live trades and market signals
- **Power BI – Live Trading Engine**: Connects predictions to trade execution
- **Power BI – Performance Report Generator**: Automates reporting on model and trade performance

Architecture Evolution

Architecture Evolution - **Initial Architecture**:

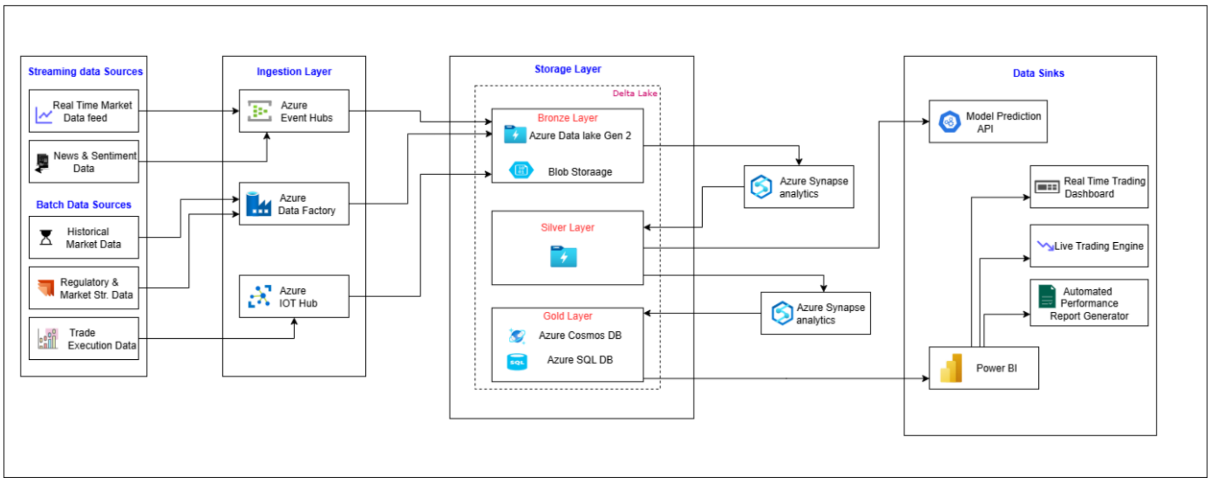
Basic Data Sources and Sinks



LinkedIn

Architecture Evolution - **Final Architecture**

Fully integrated pipeline with analytics



LinkedIn

Pipeline Flow

My final pipeline architecture is designed to handle diverse data sources, transform them efficiently, and deliver actionable insights in real time. It follows a modular, layered approach that ensures scalability, fault tolerance, and seamless integration across services.

End-to-End Data Journey

1. **Ingestion Layer** Data enters the system through Azure Data Factory (for batch sources) and Azure Event Hub (for streaming sources). These services provide reliable, scalable intake mechanisms for market data, sentiment feeds, trade logs, and regulatory inputs.

2. **Bronze Layer – Raw Storage** All ingested data is stored in Azure Data Lake Storage Gen2. This layer retains raw, unprocessed data for traceability, reprocessing, and auditability.

3. **Silver Layer – Cleaned & Enriched Data** Using Azure Synapse Analytics, data is cleaned, normalized, and enriched. This includes parsing sentiment scores, standardizing trade logs, and applying business rules. The silver layer serves as the foundation for model consumption and reporting.

4. **Gold Layer – Curated Outputs** Final datasets are aggregated and formatted for consumption. These are optimized for performance and used by downstream sinks such as dashboards, APIs, and reports.

5. **Consumption Layer (Sinks)**
    - **Model Prediction API**: Serves real-time model outputs for decision-making
    - **Power BI Dashboards**: Visualize live trading metrics and market signals
    - **Live Trading Engine**: Executes trades based on model predictions
    - **Automated Report Generator**: Produces performance summaries and analytics.

Design Principles

• **Modularity**: Each layer is decoupled, allowing independent scaling and maintenance

• **Observability**: Integrated monitoring and logging at every stage

• **Resilience**: Built-in retry logic, fallback mechanisms, and failure detection

• **Security**: Role-based access, encryption, and compliance-ready storage