

Gold_finance_ETL: Automating Financial Data Pipelines

Welcome to our presentation on Gold_finance_ETL, a robust data engineering project that automates financial time series data processing. We'll explore how this pipeline transforms raw financial data into structured insights for business stakeholders and analysts.

IN **Alpha Team**



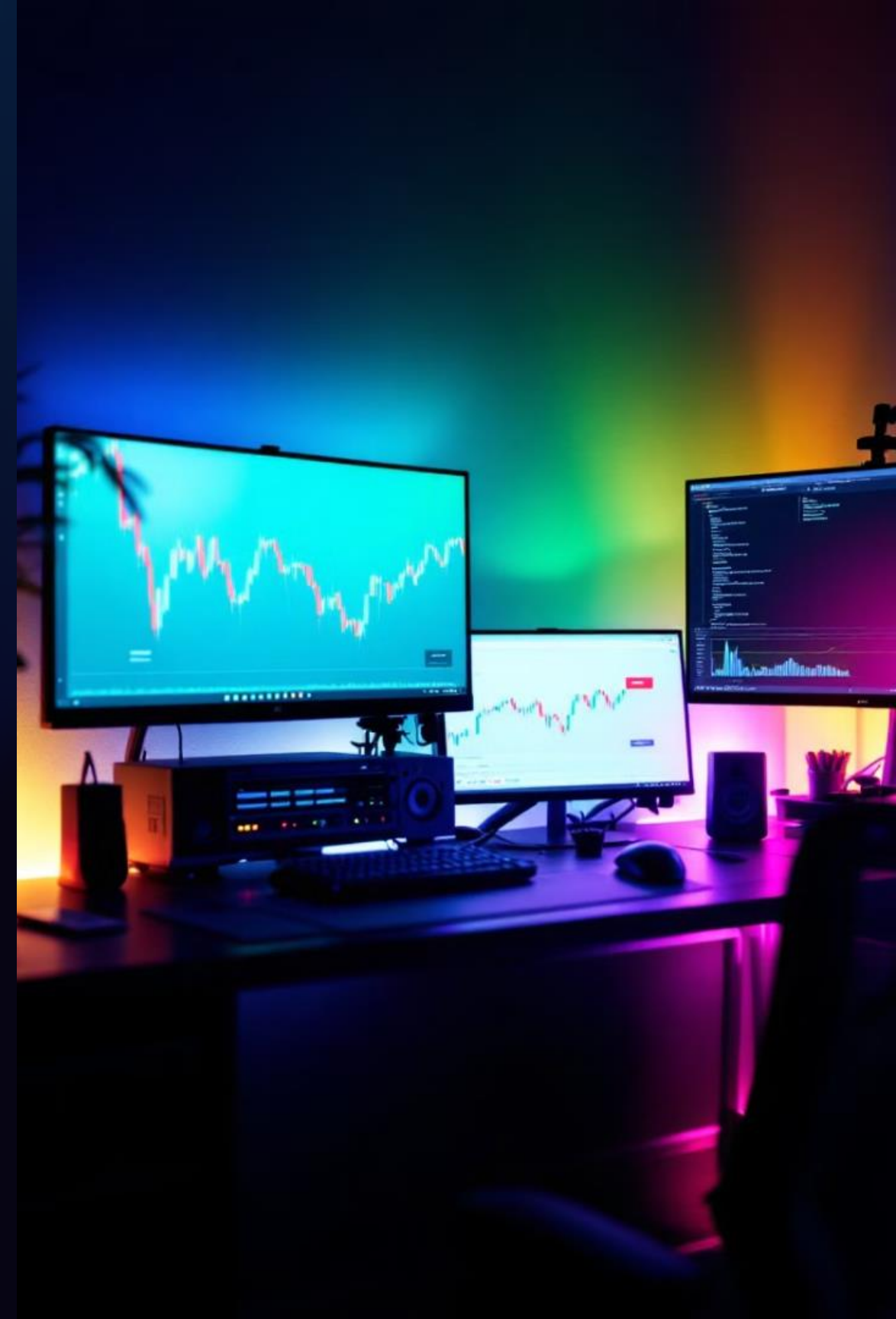
Overview: What is Gold_finance_ETL?

○ Automated ETL Pipeline

Extracts, transforms, and loads financial data from Alpha Vantage API daily.

○ Advanced Technologies

Utilizes PySpark, PostgreSQL, and Apache Airflow for efficient data processing.



Business Problem and Solution

Challenges

- High volume of rapidly changing financial data
- Unstructured formats requiring significant cleaning
- Time-intensive manual workflows prone to errors

Gold_finance_ETL Solution

- Fully automated workflows for daily data integration and analysis
- Aggregated metrics for informed decision-making

Data Source: Alpha Vantage API



Comprehensive Data

Provides daily prices and trading volumes for multiple financial symbols.



Daily Updates

Ensures accurate and timely data for trend analysis.

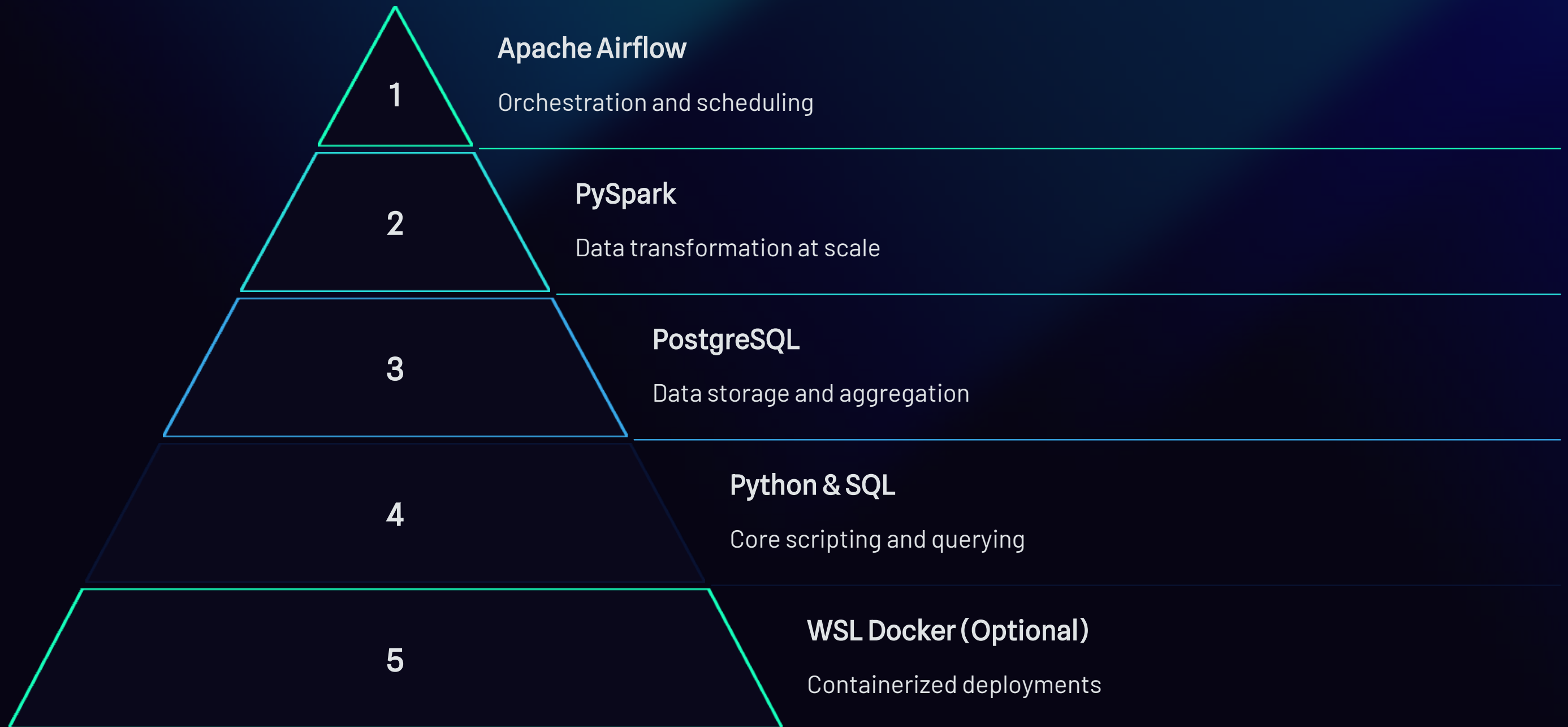


Industry Standard

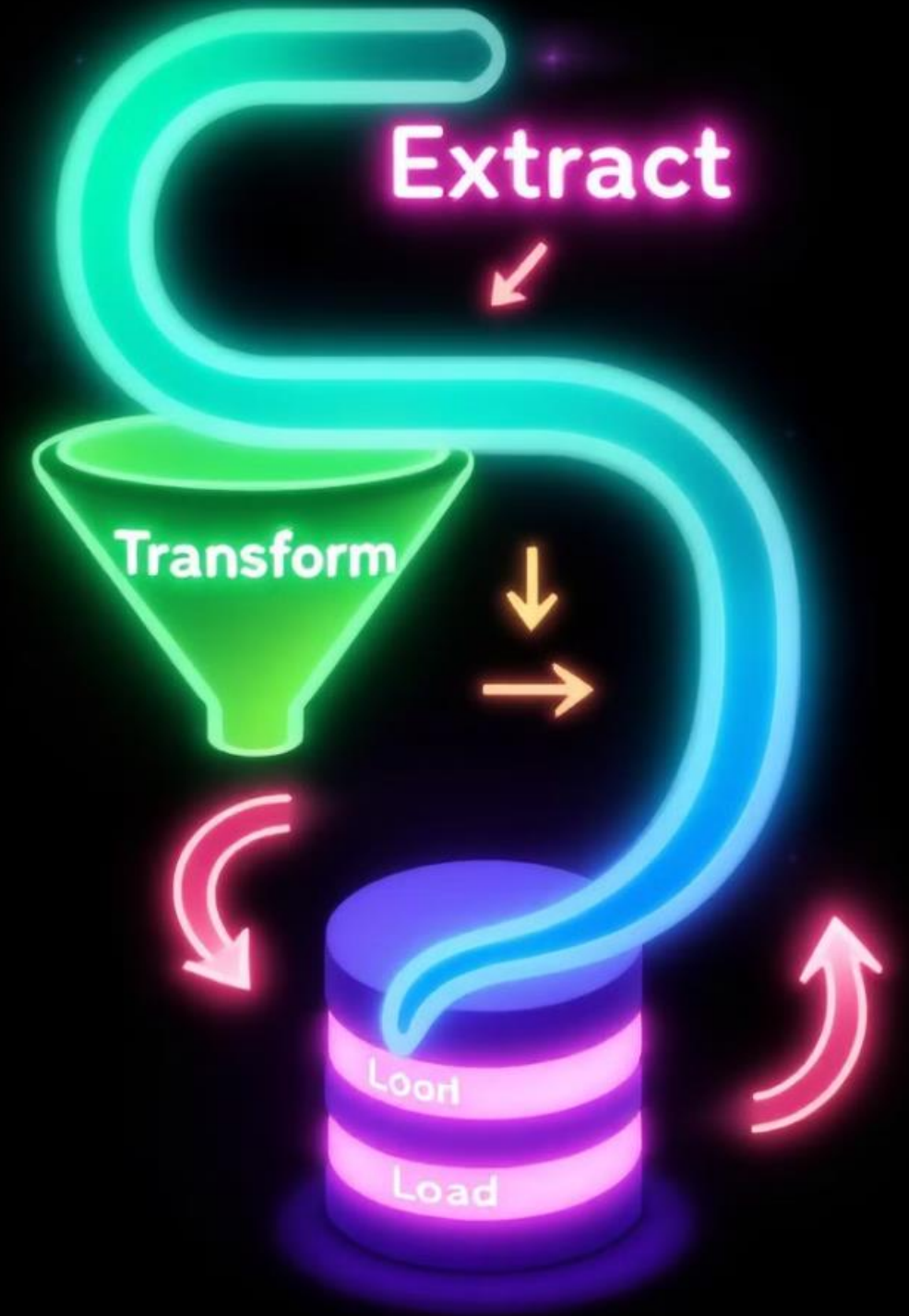
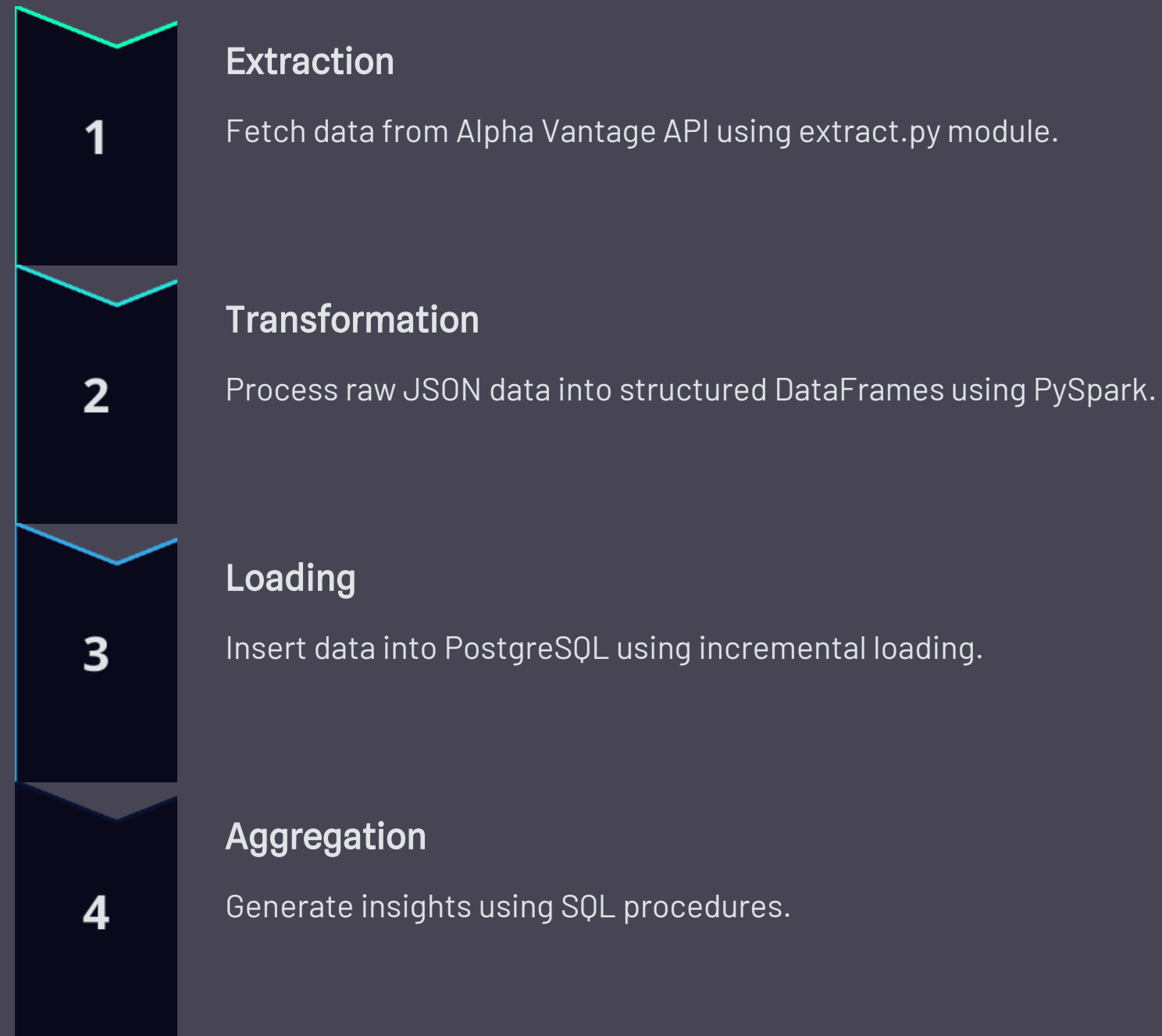
Widely used and trusted in the financial industry.



Technologies Used



Pipeline Workflow





Key Features and Benefits

Incremental Loading

Optimizes resource usage by processing only new data daily.

Automated Orchestration

Airflow DAG ensures seamless daily runs and failure recovery.

Scalable Design

Ready for future expansion and additional data sources.

Actionable Insights

Provides ready-to-use KPIs for data-driven decision-making.



Future Enhancements

1

Dashboard Integration

Connect PostgreSQL to BI tools like Tableau or Power BI.

2

Real-Time Processing

Incorporate Kafka for streaming data ingestion.

3

Additional Data Sources

Expand to other APIs for broader analysis.

4

Machine Learning

Predict trends using historical data.