Data Sources

Data Sources

Alpha Vantage — API Connections

YAHOO! FINANCE — Web Scraping

yahoo finance API — API Connections

Nasdaq Data Link — API Connections

U.S. SECURITIES AND EXCHANGE COMMISSION — XBRL 10K

Orchestration

Python Runtime

Durable Orchestration Function Application

(HTTP, Timer, Queue, Blob Triggers)

Func App Storage Account

- Application Insights
- Logs
- Diagnostic settings

Webhook

Simple Queue Service Message

Blob Storage Staging

Parallel Orchestration

Data Transforms

Data Quality Control

Validation & Quality Checks

Data Definitions — Data Retention — Normalized Tables

Blob Storage Parquet

DomGuardianAI

DGAI

Durable Orchestration

Parallel Execution

Azure OpenAI

OpenAI

Parquet Storage

SQL

Data Preparation

Cleaning — Normalization — Encoding
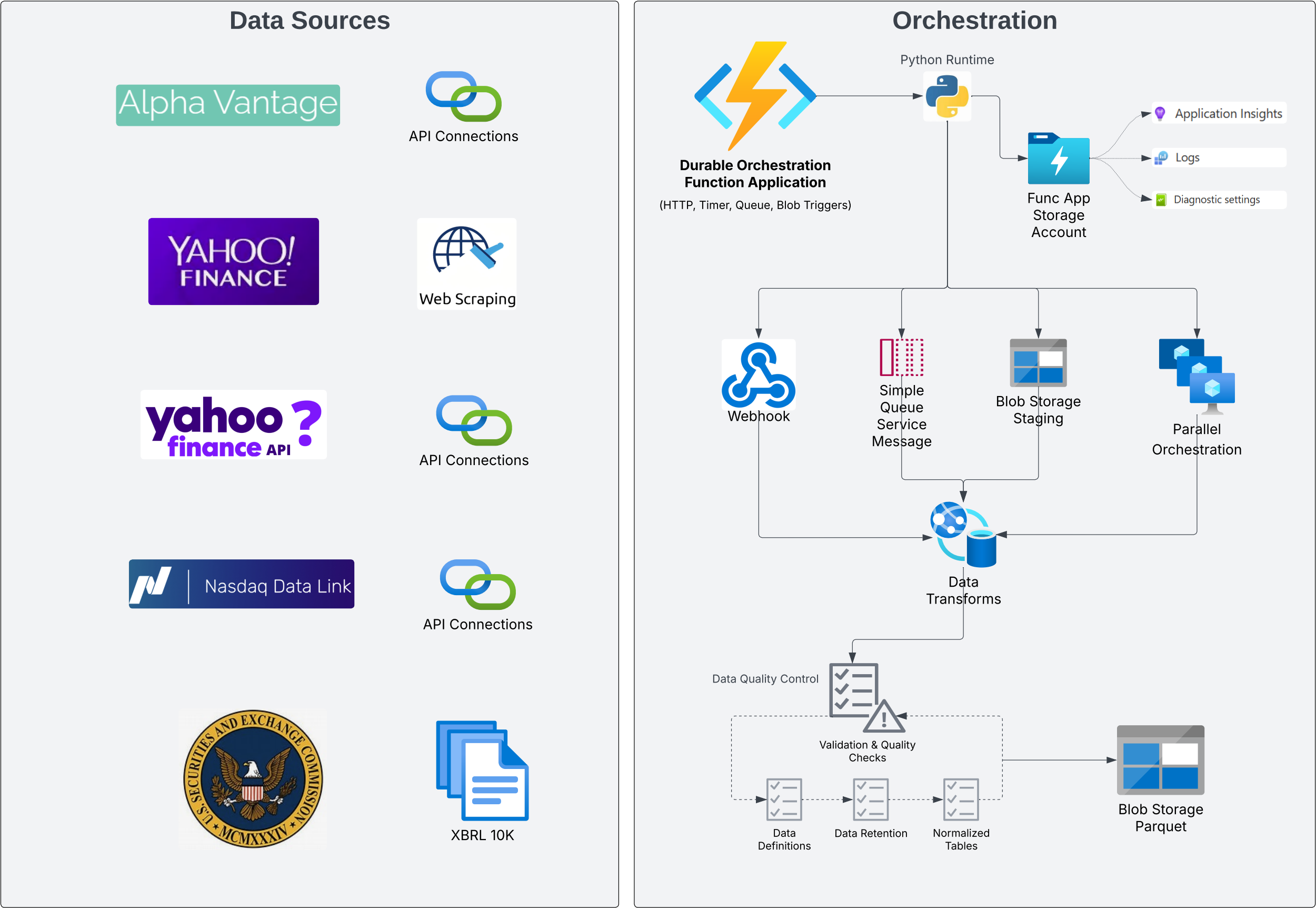
LSTM — Logistic Regression — Random Forrest Classifier — Neuron Networks — Tensor Flow

Data Sources & Orchestration

Data Lakehouse

Data Lake
Raw Storage layer of structured & unstructured data

Raw Storage — Staging Storage — Curated Storage

Data Warehouse

ALDS Gen2 storing parquet files, reading files over SQL server in further processing, transformations and analysis.

Azure Synapse On-demand Server

Python External Scripts

Storage Integration

SQL Databases & Marts

SQL

pyobdc

- News Feed
- XBRL 10K
- XBRL 8K
- Ticker Data
- NASDAQ Data
- Technical Analysis
- Data Warehouse

External Tables

Business Intelligence & Analytics

- SQL Server SSMS
- Visual Studio
- Power BI
- Python
- Azure Data Studio

Parallel Orchestration

Durable Orchestration

Parquet Storage

SQL

HTTP Trigger

Alpaca