[Data Pipeline] — Processing XML Type = CFE

process xml files with cfe format coming from azure blob storage, application connects into a specific container in a scheduled basis to capture incoming files and mark for processing, data is transferred to apache kafka, processed and augmented on ksaldb to proper fit into yugabytedb using postgres api



Data Lake

Repository of Raw Data Without Schema Enforcement





Process Files Incrementally

- Base ~ 01379480000106
- Container ~ imais
- XML Model Type ~ CFE

Application Ingestion Layer

Kafka Producer API



Decouple XML Files into Different Events Apply Business Logic

Validate Container Understanding XML Model Decouple Tree Structure Apply Business Logic

Schedule ~ Every 15 Minutes

Enterprise Data Hub [EDH]

Apache Kafka [Strimzi]



Get Events from Application and Ingest into Topics in a Parallel Manner

Split XML Parent & Child Ingest using JSON Format Using KEY – infCFE [1d]

Real-Time Ingestion

Processing Layer KSQLDB



Apply ETL Processing Logic to Combine Streams into a Single Output

Join Different Topics Apply Transformations Select Columns Output into a New Topic

Output Topic ~ ksqldb-stream-nfe-avro

Real-Time Processing

OSS Data Serving

YugaByteDB



Receive Output Topic in Real-Time and Insert into a Table

Insert Data into a Table Apply View to Encapsulate Logic Auxiliary Tables from SAS

Table ~ enriched nfe View ~ vw_enriched_products_compare_market

Real-Time Ingestion

Data Visualization

Metabase



Access View from YugaByteDB for Data Analysis

Real-Time Analysis



Continuous Delivery ArgoCD - GitOps



Logging FileBeat, ElasticSearch & Kibana



Monitoring Prometheus, Alert Manager & Grafana