SIGMAX – DATAFLOW

Data from Edge to Enterprise
THE APACHE STACK FOR DATAFLOW: EDGE TO ENTERPRISE

SigmaX Stack Features

- Arrow
- Arrow Flight
- Arrow Dive
- Data Format and Ingest
- Pub / Sub Message System
- Distributed SQL Query
- HDFS filesystem Support

Development support by:

SigmaX.ai
ACCELERATE
**THE DATAFLOW PROBLEM**

- Multiple Copies of Data
- AVRO Data Serialization / Deserialization
- Different In-memory and Durable Data Formats
- Exaggerated cache misses
- Forced to choose between fast read speeds or fast write speeds

**Apache Arrow**
- All Systems use the same memory format
- No overhead for cross system communication
- Many more target languages (R, Python, Etc.)

**Heterogeneous Computing:**
- use the best processor type for the job
  - Arrow from Ingest
  - Arrow from Storage
  - Enhanced Ingest Pre-analytics

**Cache performance:** Go direct to data with offsets. Performance benefits of columnar format
5G Example Challenge:
- Data ingest to match very high capacity, very low latency 5G network with 100% open source software eliminating vendor lock-in

SigmaX Stack:
- Hardware Accelerated Apache Arrow Record conversion
  - Supports Heterogenous AI+ML processing
  - Increases system efficiency
- Pulsar Innate Processing capability vs. other modern streaming solutions
  - Lower Latency
  - Larger Message Size
  - Better Throughput
  - Data Replication
  - Streams and Topics
  - Tiered Storage support
  - Decentralized
- Pulsar Optane enhancement (10x latency reduction, throughput boost)
**Arrow Flight:**
- Middle Step to verify and validate

**Why Flight?**
- Support for direct data ingest via FPGA smartNIC network ports (Latency)
  - Direct data streaming access to data at ingest points to client applications.
    - Lowest Latency
    - Highest Throughput
  
  - Apache Pulsar dataflow is still available and archives data as it ages.
  
  - Direct access to Heterogenous Processing resources
Direct Access Edge Solution:

- Support for Disaggregated Flight Servers from Ingest Elements
  - Many, many FPGA Ingest points talking to few Flight Servers
  - FPGA Ingest becomes an easily accessible resource for Client Applications
- Containerized delivery of Edge Ingest Solution
  - OpenShift or Kubernetes
- DevOps & DevSecOps friendly
- User applications reside at the Edge, in the Fog, Enterprise or Cloud
- User applications can authenticate and gain direct access to data being ingested at the edge
- User applications have streams and topics data streaming support (for post pre-analysis data)
- User applications have access to all historic data via Pulsar / tiered storage support.
THANKS!!!

Tom Paulick (443) 717-2522
tom@sigmax.ai
www.sigmax.ai