Real-Time Smart Data Lake for Accelerated In-Place Analytics of Scientific Data

Donpaul C. Stephens
donpaul@airmettle.com
Founder, AirMettle, Inc.
+1-646-872-2124

HDF5 User Group Meeting
August 2023
AirMettle has developed a real-time smart data lake solution that simplifies big data analytics and accelerates processing by an order of magnitude, or more.

It is implemented in the data lake storage layer and performs basic analytics tasks that:

• Reduce network traffic - Extract only what is needed before returning
• Improve data freshness - All data can be rapidly queried
• Enable real-time operation - Sub-second queries on GByte objects from storage
AirMettle’s Mission

AirMettle has developed a real-time smart data lake solution that simplifies big data analytics and accelerates processing by an order of magnitude, or more.

It is implemented in the data lake storage layer and performs basic analytics tasks that:

• Reduce network traffic - Extract only what is needed before returning
• Improve data freshness - All data can be rapidly queried
• Enable real-time operation - Sub-second queries on GByte objects from storage

https://techpartnerships.noaa.gov/sbir/awards/
Traditional Data Lake

Object Storage

Objects are internally partitioned
For storage in parallel

Data Lake
Traditional Data Lake:
Data generally arrives semi-structured

Comes from Everywhere

Object Storage

Objects are internally partitioned
For storage in parallel

Primarily
Semi-structured data

Data Lake
Traditional Data Lake:
Data must be moved to gain value from it

Analyzed
In Islands

Comes from Everywhere

Applications retrieve full objects*
To their own (small) clusters
for processing

Object Storage

Objects are internally partitioned
For storage in parallel

Primarily
Semi-structured data

Data Lake

*Applications retrieve full objects to their own (small) clusters for processing.
Smart Data Lake: Get only what is needed 100x Faster
Delegate simple processing to storage thereby eliminating the need to move 90%+ of the data

Analyzed Faster
In Islands

Comes from Everywhere

Retrieve what they need in an Immediately usable form

Objects are internally partitioned For storage & processing in parallel

Primarily Semi-structured data

Data Lake
Accelerated analytics of classic tabular data

Security Information & Event Management

- Scan historical data to diagnose current events
  - Determine how many records might be relevant before retrieving any

Natural Language Processing

- Search for key-words
  - Gather statistics of usage
  - Extract text if required for further analysis
Accelerated analytics of classic tabular data (S3 Select API)

Security Information & Event Management

- Scan historical data to diagnose current events
  - Determine how many records might be relevant before retrieving any

Natural Language Processing

- Search for key-words
  - Gather statistics of usage
  - Extract text if required for further analysis

Validated with Wisconsin & MIT

Star Schema Benchmark
Utilized 223 Select queries to Object Storage

Unprecedented speed of analysis: Directly from storage

No data warehouse required

AirMettle: Select 100x FASTER

Unpreceded speed of analysis: Directly from storage

No data warehouse required

Copyright © 2023 AirMettle, Inc.
AirMettle Accelerates

5x Acceleration on Complete Queries, today… just by using a different storage

S3 Select API enables comparison vs. major cloud’s object storage

Star Schema Benchmark, Scale Factor 1 with 1 object per table
Resiliency 101: How do storage solutions protect data?

RAID:

Data protection algorithms designed for HDD

Erasure Coding:
What that means for data reliably placed in storage: First 4 devices shown...

**Simple Table:**

<table>
<thead>
<tr>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSD</td>
<td>SSD</td>
<td>SSD</td>
<td>SSD</td>
</tr>
</tbody>
</table>

Bytes of data divided evenly across SSDs!

Data protection and streaming performance!

---

**Supports data protection algorithms designed for HDD!**
What that means for data reliably placed in storage: First 4 devices shown...
AirMettle: Data partitioning for processing AND protecting data

- Data is unchanged for client
- Each internal component can be processed in parallel

AirMettle internal metadata enables parallel in-storage analytics

Not to scale! Meta-data typically <0.1% of data

Object’s own metadata

AirMettle Patented IP
AirMettle: Data partitioning for processing AND protecting data

- Data is unchanged for client
- Each internal component can be processed in parallel

AirMettle protection algorithms are designed for our non-uniform data segments

AirMettle Patented IP
AirMettle Accelerates Multi-dimensional data

- Complex data format!
- Even the coordinates are multi-dimensional

- Scientific analysis also requires:
  - Sampling of different subsets for each query
  - Gathering statistical properties
• Complex data format!
• Even the coordinates are multi-dimensional

• Scientific analysis also requires:
  - Sampling of different subsets for each query
  - Gathering statistical properties

• Data is stored in partitions, based on semi-structural boundaries & size of segment
  - Enables distributed in-place parallel processing
  - Validated in Phase I
  - Queries are more complex than events w/ SQL
    - 2-3 Stage queries typical, but each stage can be executed in parallel
- Selection returned as “smaller” NetCDF4
- Partitioned tensor data can be efficiently scanned in parallel
- Multi-stage processing:
  - Check coordinates
    - Bounding box & potential mask
    - Identifies components holding relevant data
  - Scan / Filter desired data
  - Integrate result for return
Initial Support: Selection & Simple Aggregations via REST APIs

• Challenges included:
  - User-transparent partitioning
    - Internal data placement was more complex than originally anticipated
    - Typical internal overhead <0.3%
      - Dwarfed by data protection overheads
      - Erasure coding typically 20% to 35%
  - HDF5 does not have a re-entrant library
    - Developed support for concurrent analytics
    - Enables massive parallelism required for a shared storage service
  - Query engine required extensions for multi-stage execution & tensors

• Selection returned as “smaller” NetCDF4
• Partitioned tensor data can be efficiently scanned in parallel
• Multi-stage processing:
  - Check coordinates
    - Bounding box & potential mask
    - Identifies components holding relevant data
  - Scan / Filter desired data
  - Integrate result for return
• Very common operation to convert from higher-resolution to lower-resolution.
  - When combining data from different sources, the final result can only be the lowest common resolution among the different sources.
• Project formally commenced Aug 1\textsuperscript{st}
  - Developing parallel re-gridding on internal partitions

**Coming Summer ’24**
- Multi-dimensional in-place analytics

• Plan to explore alternative aggregation functions (Min/Max/Std. Dev) to enable faster analysis
  - Potential storm fronts, etc.
Roadmap: Hierarchical support & User Defined Functions! AI in-place

HDF5 (e.g. Satellite data) stored in native form

Satellite data

HDF5 object is internally partitioned

Data stored across cluster

Processed On-demand (In parallel)

Select Region

Select Model

User Defined AI

AI models stored as separate objects

Results returned as single object for queried region
Acknowledgement

This work was supported with government support under
- 2135007 awarded by the National Science Foundation
- NA22OAR0210591 awarded by the National Oceanic and Atmospheric Administration
- NA23OAR0210342 awarded by the National Oceanic and Atmospheric Administration
Unified Data Infrastructure (2.0)

**Sources**
- OLTP Databases via CDC
- ERP (Oracle, Salesforce, NetSuite, etc.)
- Operational Apps (Salesforce, Workday, Zendesk)
- Event Collectors (Segment, Snowflake)
- Logs
- 3rd Party APIs (e.g., Stripe)
- File and Object Storage

**Ingestion and Transport**
- Data Replication (Fivetran, Stitch, Matillion, Airbyte)
- Workflow Manager (Airflow, Apache/Amazon, Prefect, Elements’ Dagster)
- Event Streaming (Confluent Kafka, AWS Kinesis, Pulse, Upstream)

**Storage**
- Data Warehouse (Snowflake, BigQuery, Redshift)
- Data Lake (Delta, Tafile/Orchard, Hudi)
- Data Platform (S3, GCS, AAS, HDFS)

**Query and Processing**
- Spark Platform (Databricks, Amazon EMR)
- SQL Query Engine (Starburst/Pressly, Trino, Hive, Dremio, Apache/Google/Oracle)
- DS/ML Platforms (Pandas, Dask, Apache/Amazon, Ray, PyTorch, PySpark, ...

**Transformation**
- Metrics Layer (Grafana, Kibana, Transform, Supergen, BIC)
- Data Modeling (SAS, Looker)
- Workflow Manager (Airflow/Amazon, Prefect, Elements’ Dagster)

**Analysis and Output**
- Dashboards (Looker, Tableau, Tableau Prep, Power BI, ThoughtSpot)
- Embedded Analytics (Snowflake, Looker, SaaS)
- Augmented Analytics (Thoughtspot, Outlier, Anaconda, SaaS)
- Data Workspace (Excel, Jupyter, Dask)
- DS/ML Tooling (Dataiku, Segment, DataRobot, ...
- App Frameworks (Streamlit, Rasa)
- Custom Applications

**Real-time Analytics Database**
- Imply/Druid, ClickHouse, Firebird, Redshift

**Data Discovery**
- Algunst, Databricks, Atlan, Atlassian

**Data Governance**
- Collecta

**Data Observability**
- Mentor-Carta, BigQuery, Superconductive, Great Expectations, AccelData

**Entitlements & Security**
- Privunda, Immunize

Source: https://future.a16z.com/emerging-architectures-modern-data-infrastructure/
Holy data copies Batman!
Contemporary Unified Data Infrastructure (3.0)

- Eliminate Data Warehouse for simple queries
- Real-time analytics from semi-structured data

I want speed, not expensive copies!