



Flash Memory Summit

100x Faster Analytics from Reliable Computational Storage

By: Donpaul Stephens

Founder & CEO,  AirMettle



AirMettle's Mission

- AirMettle is developing a real-time smart data lake solution that simplifies big data analytics and accelerates processing by an order of magnitude, or more.
- It is delivered as an intelligent data lake storage service which performs basic analytics tasks that:
 - Reduce network traffic
 - Improve data freshness,
 - Enable real-time operation.



AirMettle's Mission



Flash Memory Summit

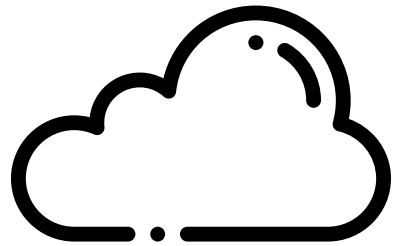
- AirMettle is developing a real-time smart data lake solution that simplifies big data analytics and accelerates processing by an order of magnitude, or more.
- It is delivered as an intelligent data lake storage service which performs basic analytics tasks that:
 - Reduce network traffic
 - Improve data freshness,
 - Enable real-time operation.

Honored to receive awards



Approaches to enhance modern application performance

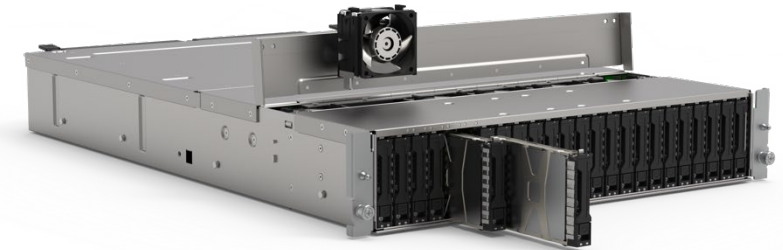
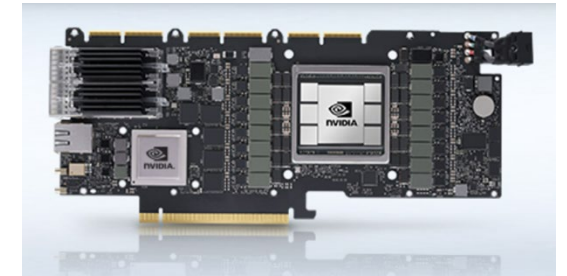
Software APIs to enable serverless compute



S3 Select
Rekognition

A service which end-clients can use...

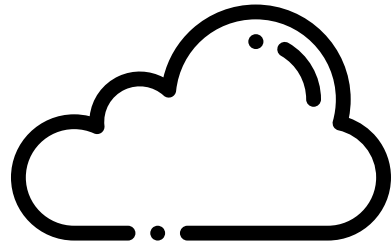
Processing
in-storage hardware



or components for building systems?

Approaches to enhance modern application performance

Software APIs to enable serverless compute



S3 Select
Rekognition

Software/Hardware
Co-design

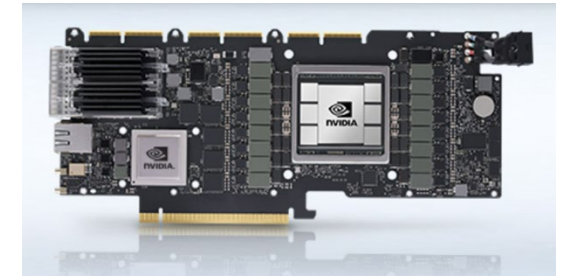


Real-Time Smart Data Lake

Software-defined
Patent-pending

Unlock the potential of storage to dramatically accelerate services

Processing
in-storage hardware



Use **existing APIs** to accelerate queries **without changing top-level software**,
Further accelerate with optimized hardware when available

Traditional Data Lake



Objects are internally partitioned
For storage in parallel

Data Lake

Traditional Data Lake:

Data generally arrives semi-structured

Comes
from
Everywhere



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



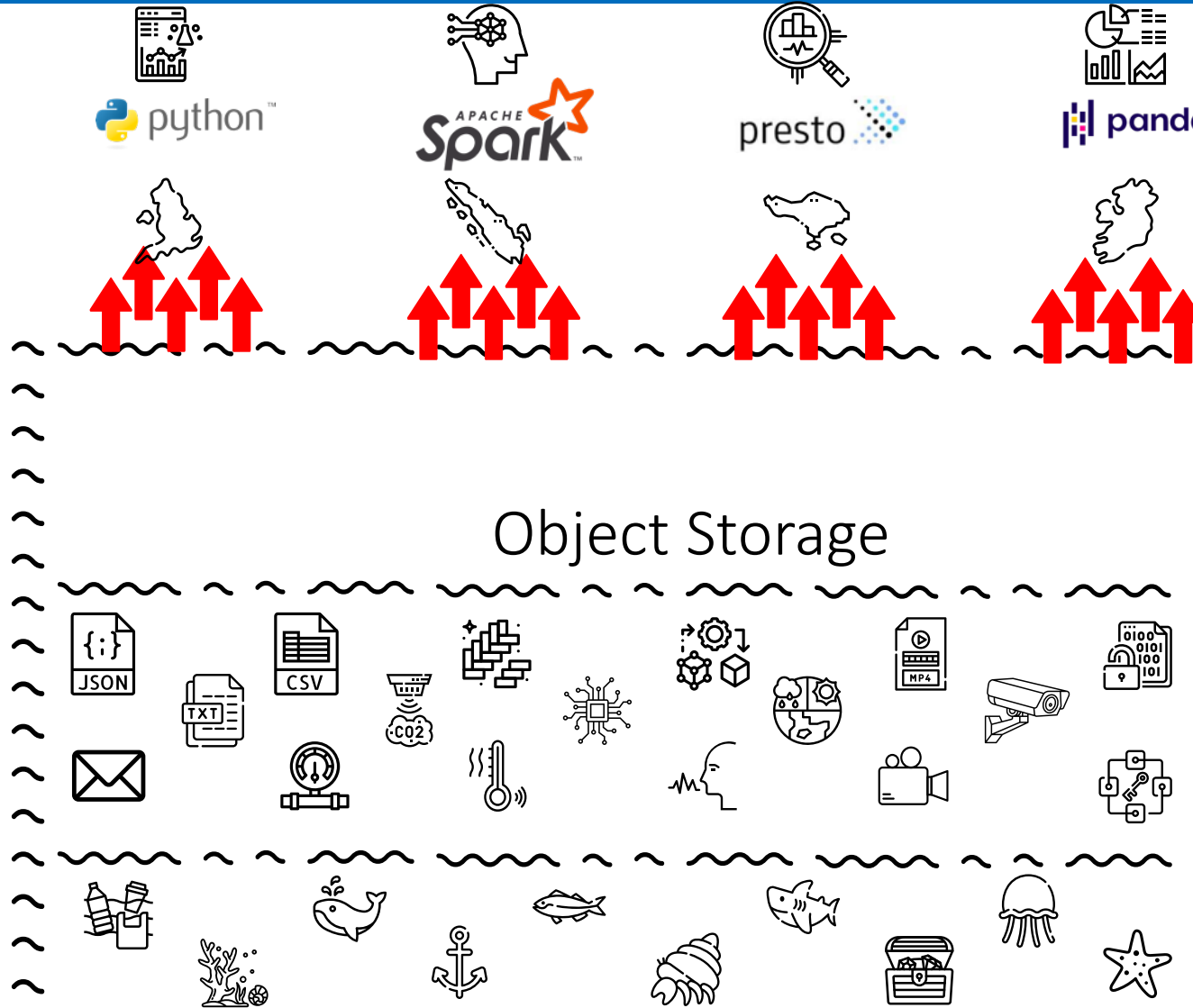
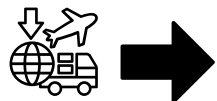
Flash Memory Summit

Analyzed
In Islands



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

Comes
from
Everywhere



Objects are internally partitioned
For storage in parallel

Primarily
Semi-structured data

Data Lake

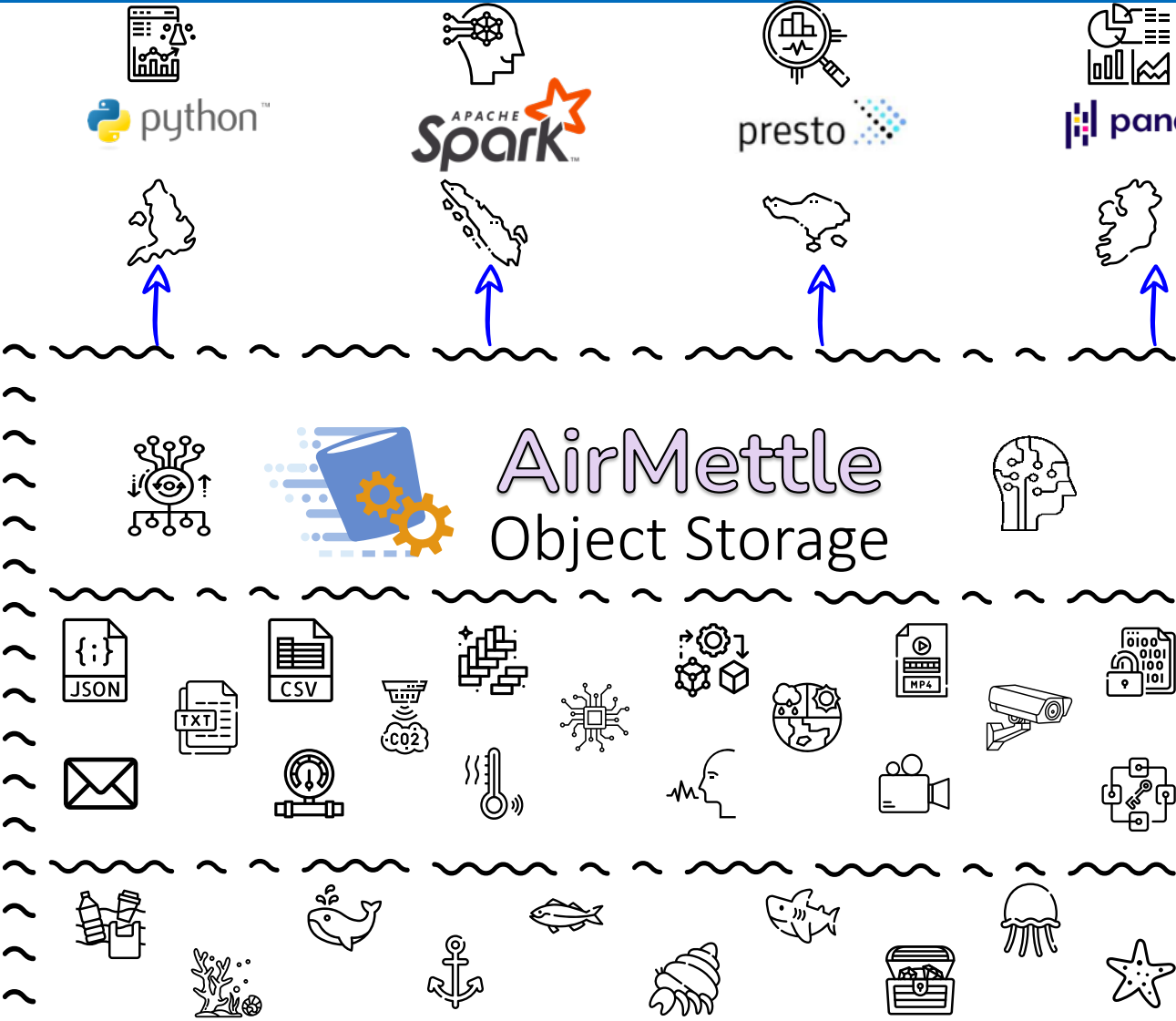
Smart Data Lake:

Load the desired subset or result, eliminate the need to move 90%+ of the data

Analyzed **Faster**
In Islands



Comes from
Everywhere



Retrieve what is needed in an
Immediately usable form

Objects are internally partitioned
For storage & **processing** in parallel

Primarily
Semi-structured data

Data Lake

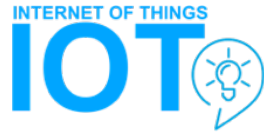
Accelerated analytics of classic tabular data

Industry Standard, Object Storage Select API



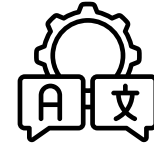
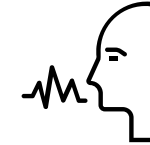
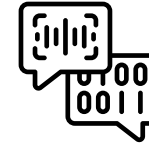
Flash Memory Summit

Security Information & Event Management



- Scan historical data to diagnose current events

Natural Language Processing



- Find text with key-words

Validated with

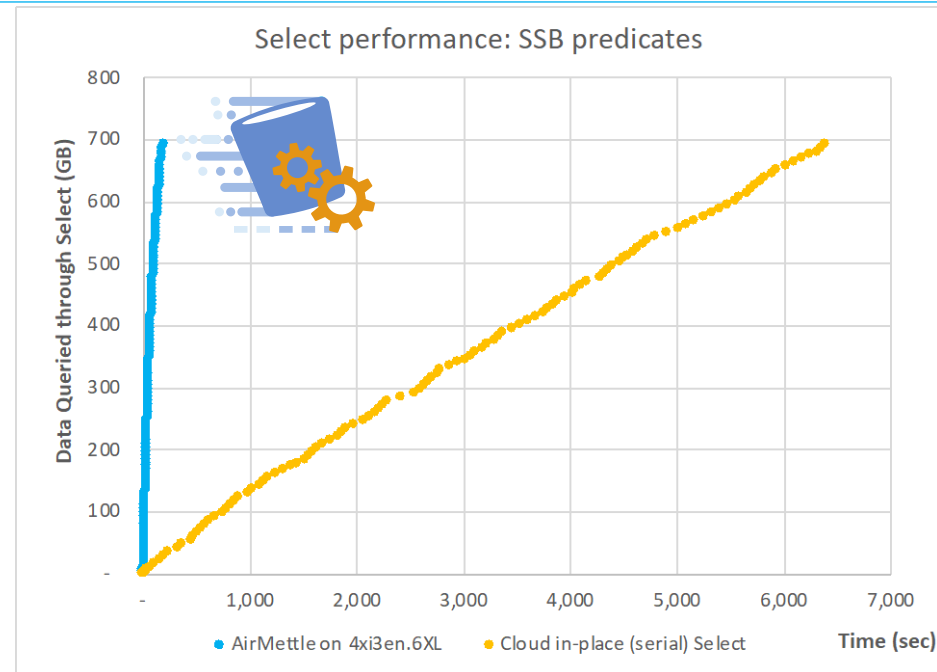


&



Star Schema Benchmark

Utilized 223 Select queries to Object Storage:



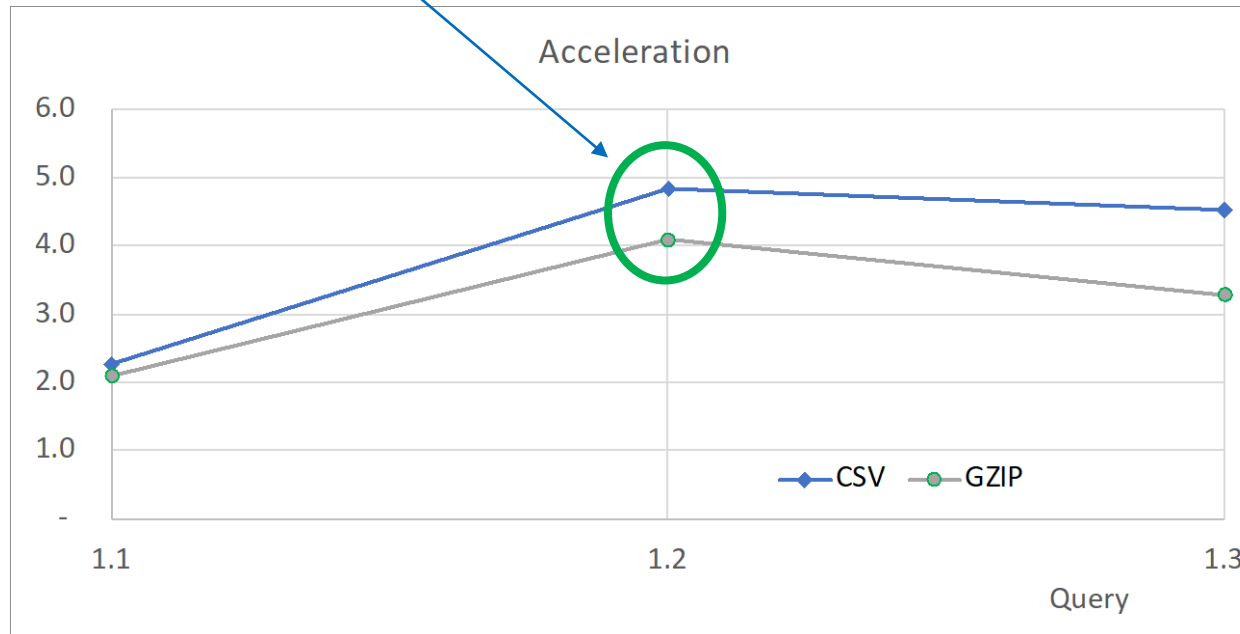
100 X faster

Under a minute vs. 1 hour 45min

Unprecedented speed of analysis: Directly from storage

No data warehouse required

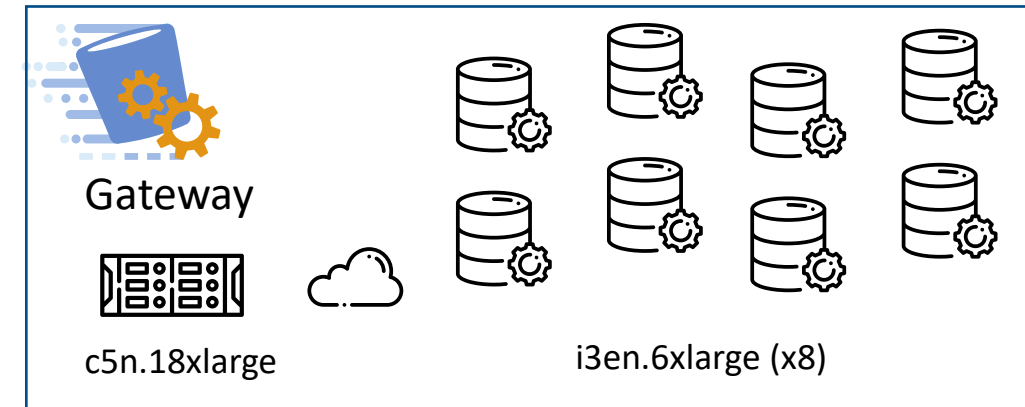
5x Acceleration on Complete Queries,
today... just by using a different storage



Star Schema Benchmark, Scale Factor 1 with 1 object per table



c5n.18xlarge



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

Accelerated video analytics

Private Sector applications

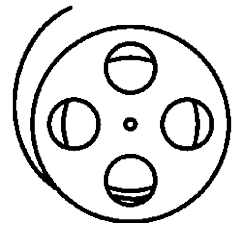
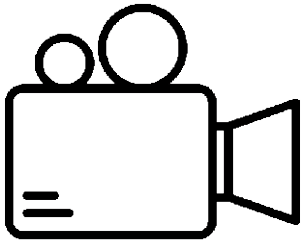
Director of Fantasy Adventure show

Problem:

Make sure current-day items (e.g. coffee cup)

Do not appear on screen

Movie object



17:40 → 17:41 in Seconds!

Public Sector applications

Validate alibi:

Find (innocent) suspect from video footage

Suspect says they walk alone in the mornings...
between their apartment and the park...



Find (most likely) matches...

Clear innocent people quickly, make justice more efficient

Summary



Flash Memory Summit

- The future belongs to data lake technologies that will provide efficient, lower-cost object storage for semi-structured data from many sources.



- **AirMettle unlocks the power of computation within the storage**

- Built to leverage Flash to its true potential
- Basic queries run 100x faster than existing cloud solutions
- Complete queries through higher-level frameworks 5x Faster
 - Same APIs, just a more intelligent storage service

- **Coming soon:**

- Video A simple line-art icon of a video camera.

- GPU optimized storage A line-art icon of a hand flexing a bicep with a lightning bolt above it, symbolizing power or speed.
(DPU seminar)

Thank you

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



AirMettle