Bringing Intelligence to Storage

In-Situ Processing

August 2017

Making Sense Out of Data

The World's Most Valuable Resources are No Longer Oil, But Data

…"As devices from watches to cars connect to the internet, the volume is increasing: some estimate that a self-driving car will generate 100 gigabytes per second. Meanwhile, AI techniques such as machine learning extract more value from data."

The Economist – May 6th 2017

This photograph was produced by Agência Brasil, a public Brazilian news agency. Their website states: (English translation: All content on this website is published under the Creative Commons Attribution 3.0 Brazil License unless specified otherwise and content replicated from other sources.) https://commons.wikimedia.org/wiki/File:Oil_platform_P-51_(Brazil).jpg

This image or file is a work of a Defense Advanced Research Projects Agency (DARPA), an agency of the United States Department of Defense, employee, taken or made as part of that person's official duties. As a work of the U.S. federal government, the image is in the public domain. https://commons.wikimedia.org/wiki/File:DARPA_Big_Data.jpg

Boundary Conditions

Big Data
3 V's

Volume	**Velocity**	**Variety**
Zettabytes	Streaming	Unstructured
Terabytes	Batch	Structured

AND

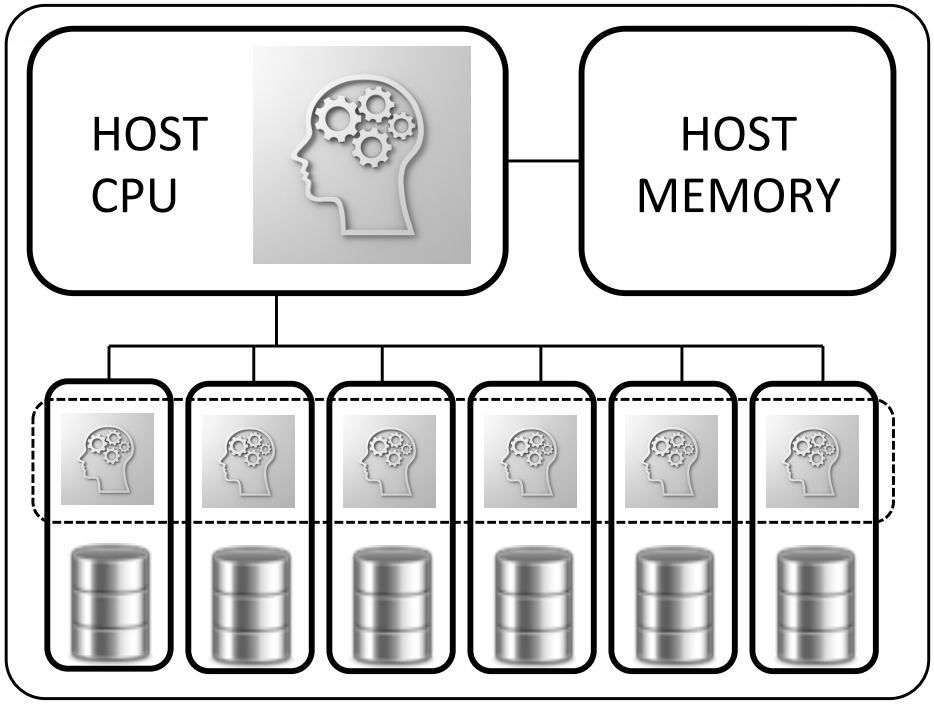
The Cost of Moving Data

The Cost of Moving Data

"Moving Computation is Cheaper than Moving Data"

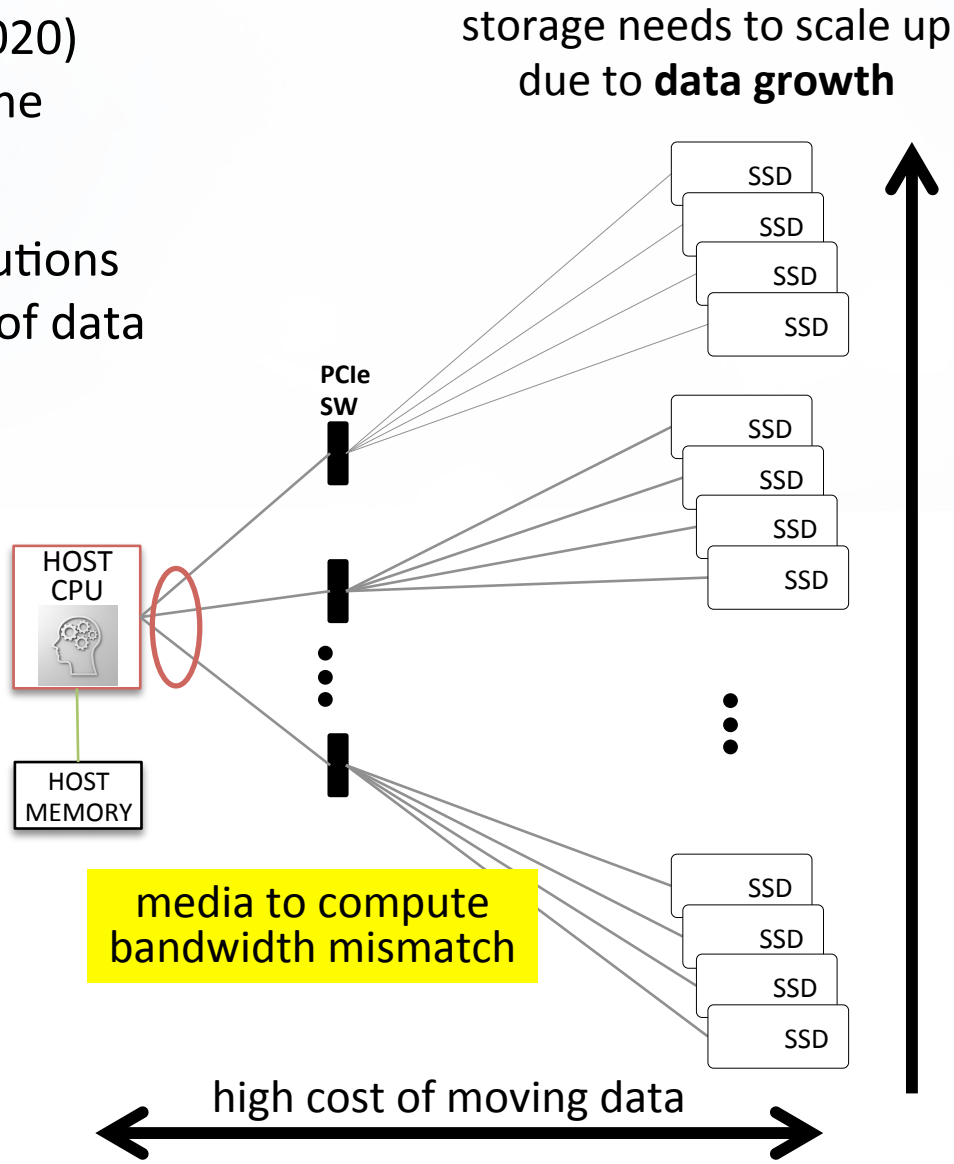
…A computation requested by an application is much more efficient if it is executed near the data it operates on*. This is especially true when the size of the data set is huge.* ***This minimizes network congestion and increases the overall throughput of the system****…*

HDFS Architecture Guide *(https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html)*

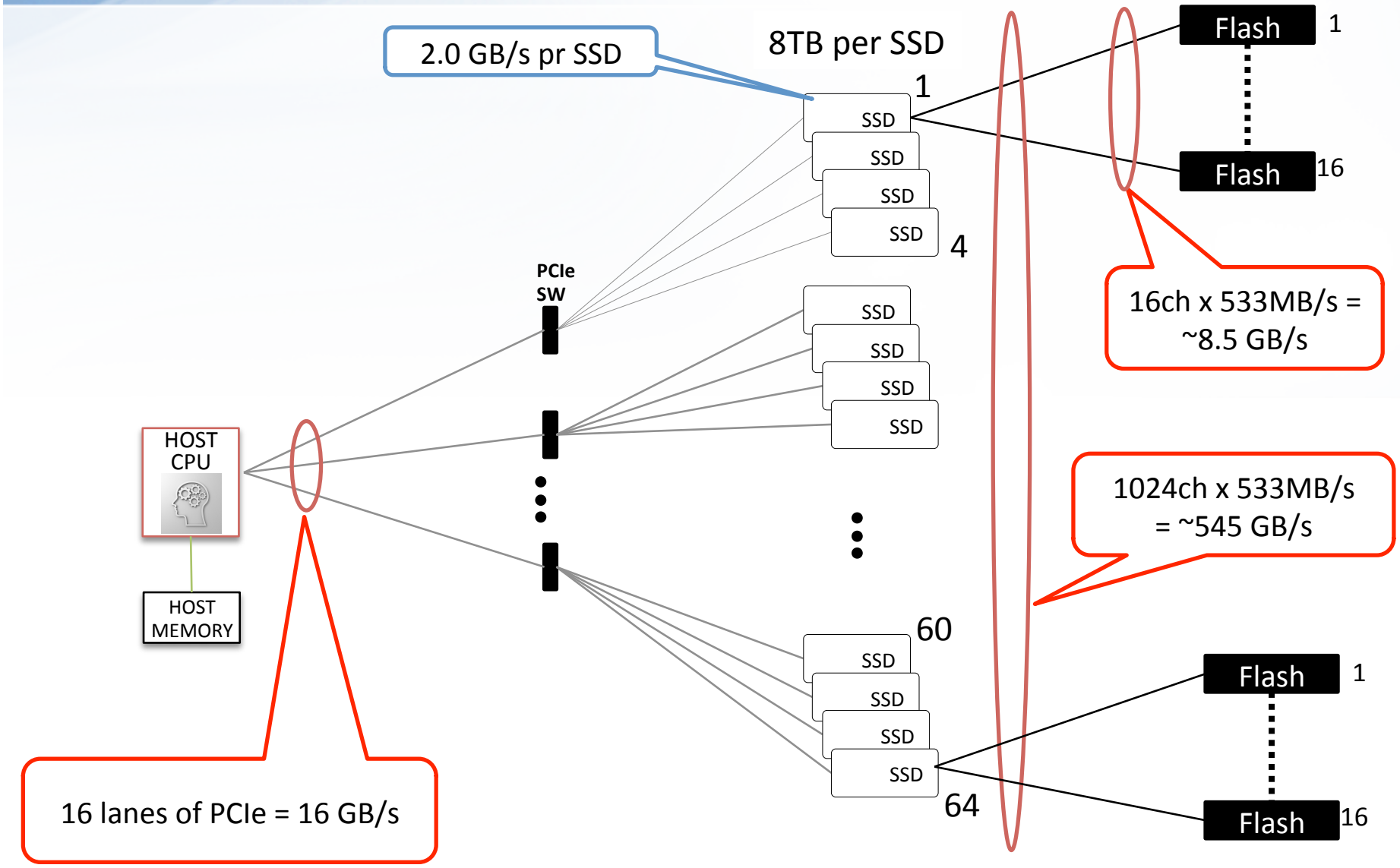


Compute to Bandwidth Mismatch

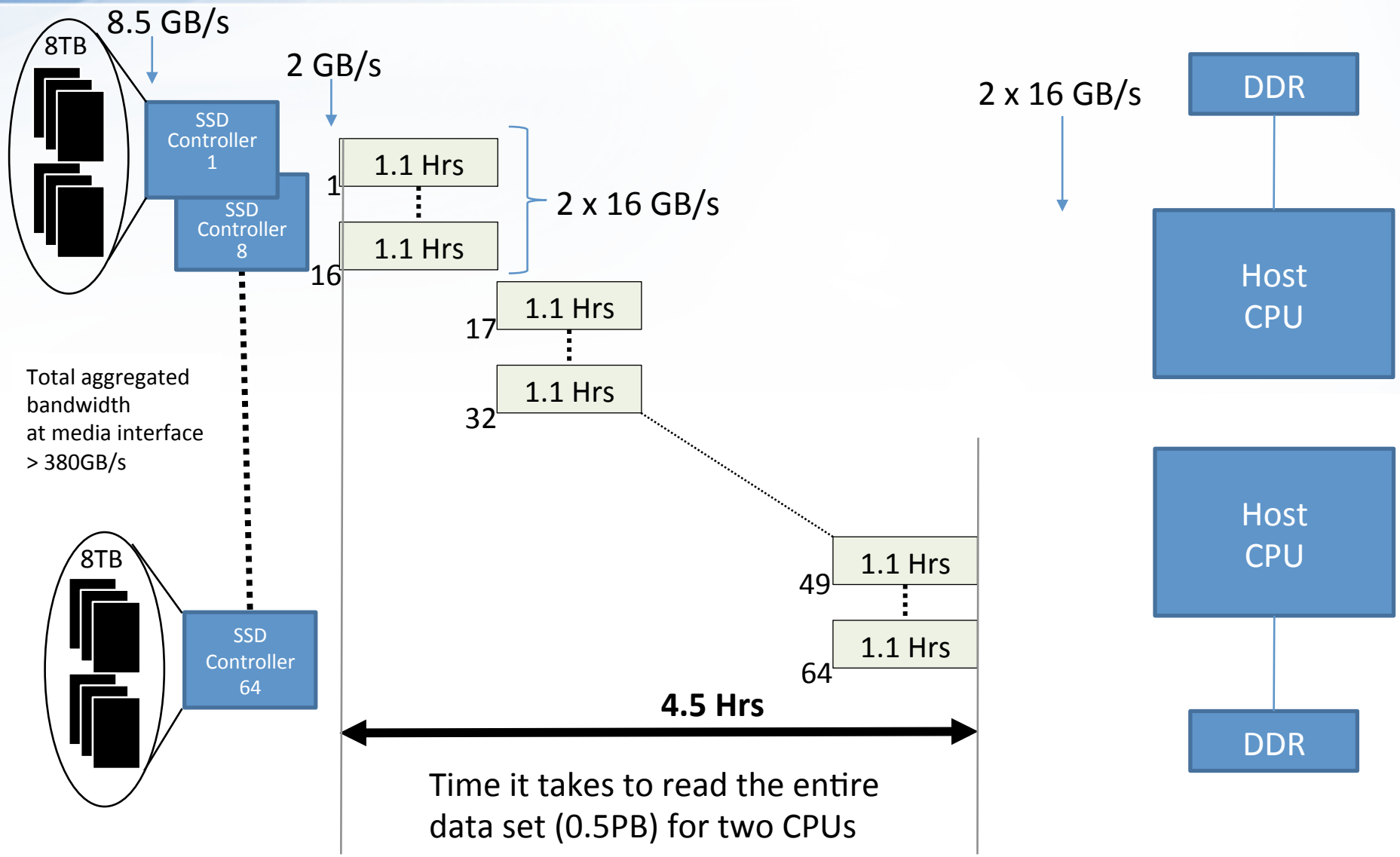
- **Exponential** data growth (44ZB by 2020)
- Infrastructure cannot keep up with the **volume** and **velocity**
- Need **ultra high capacity** storage solutions that can help reduce the movement of data and power consumption.



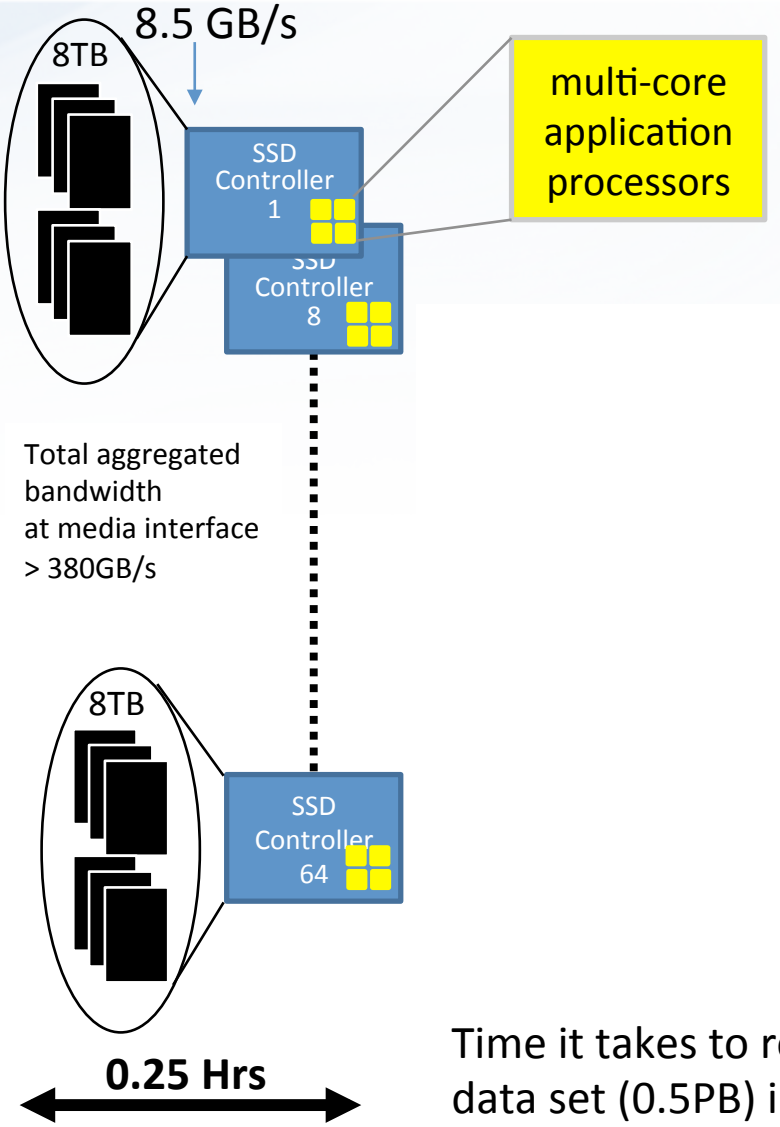
Open Compute Use Case



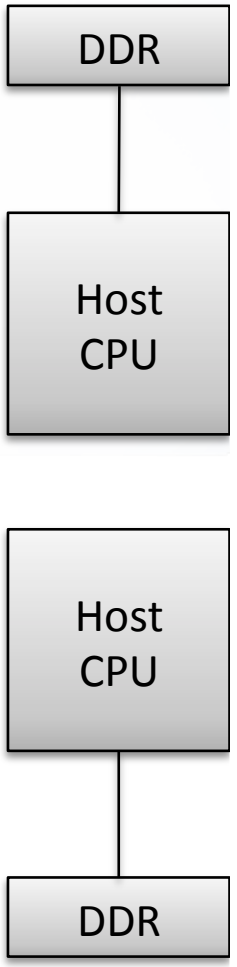
Data Ingest Scenario



Data Ingestion with In-Situ Processing



Time it takes to read the entire data set (0.5PB) into SSDs



What is In-Situ Processing?

"In-Situ": *in-place (Latin)*

In-Storage Computation Platform

Enabling distributed processing

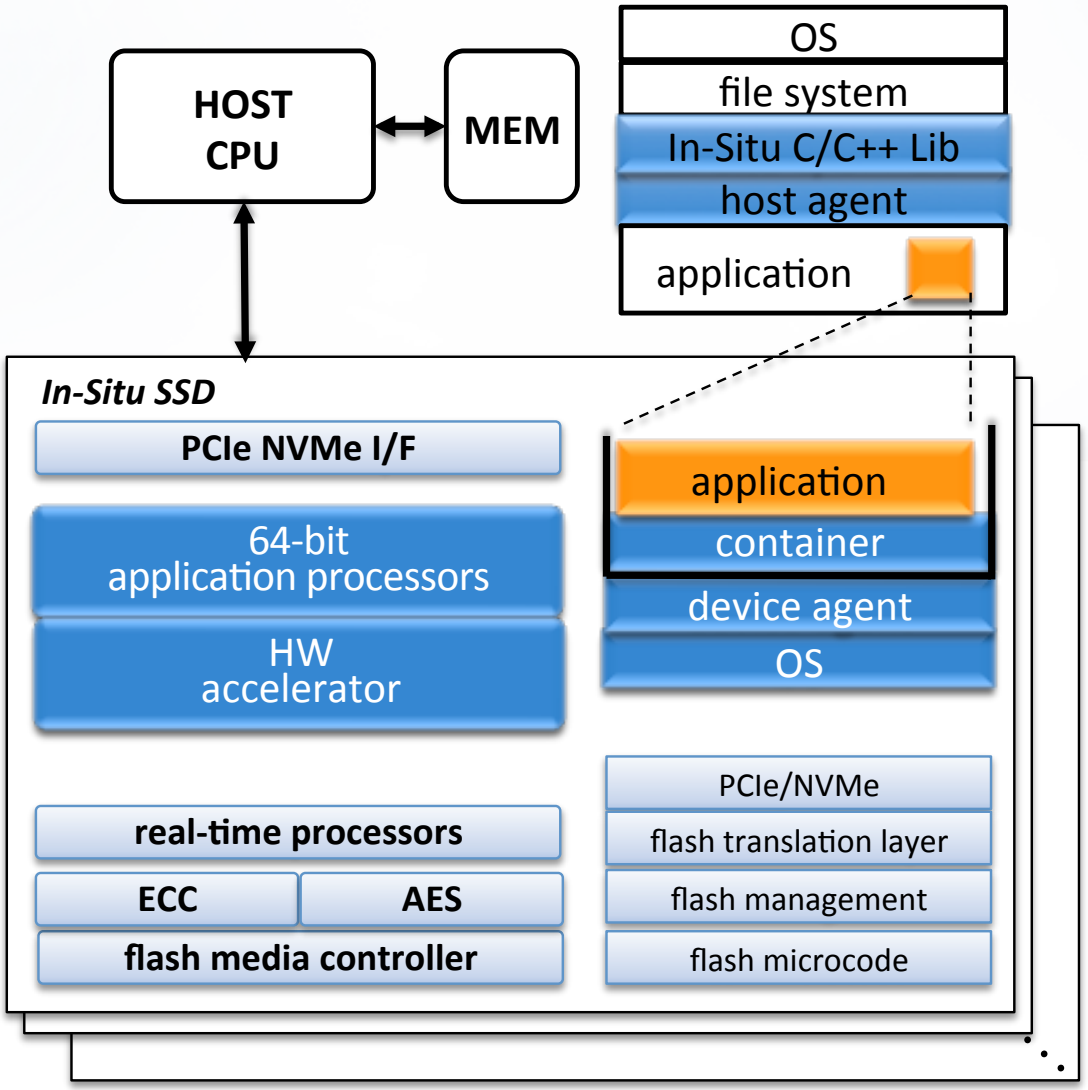
Ideal for **IO intensive** applications

Implements the "move compute to data" paradigm

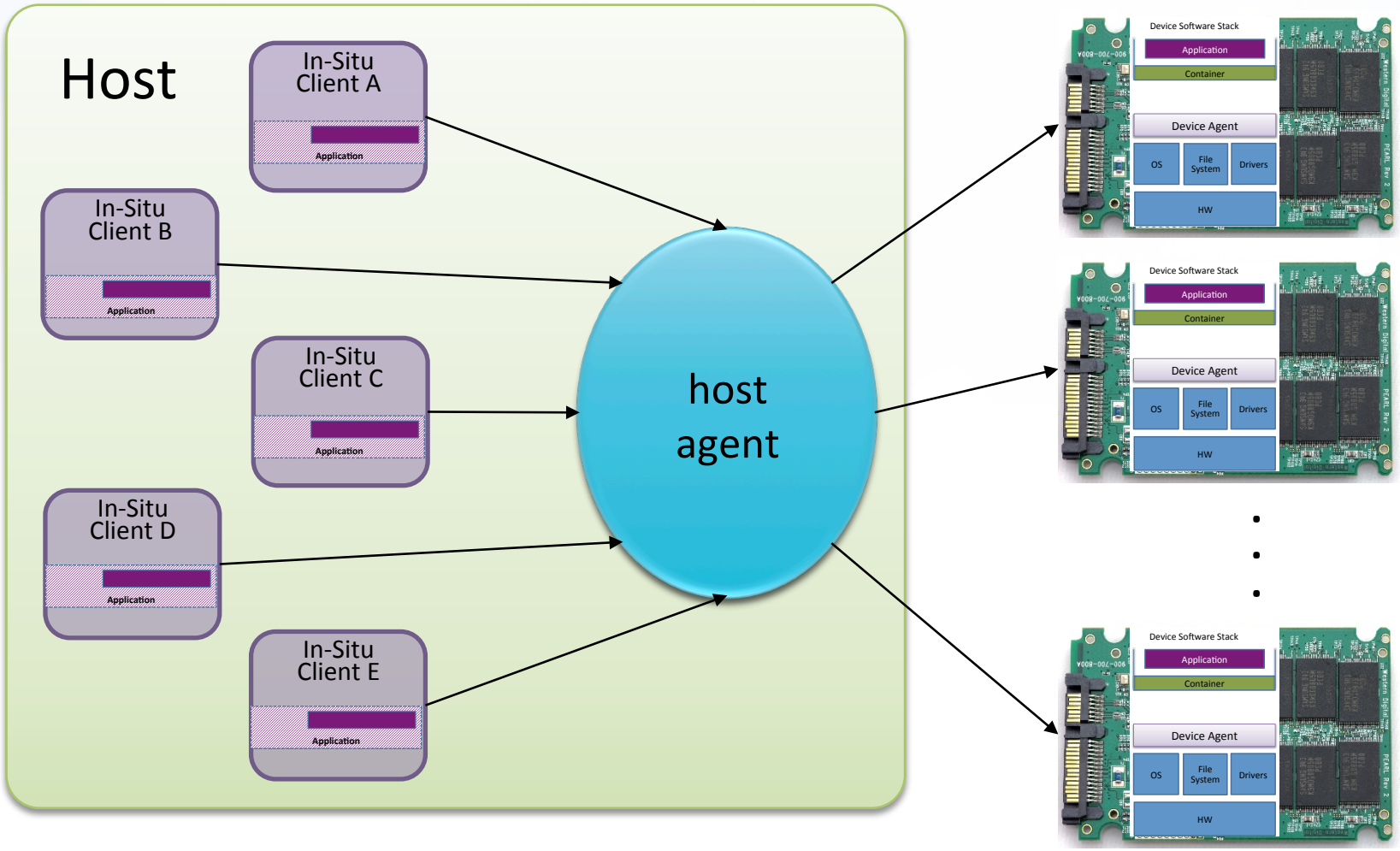
What is In-Situ Processing?

- ✧ Dedicated compute resources
- ✧ HW acceleration for data analytics
- ✧ No impact on host read/write
- ✧ No impact on NVMe driver

- ✧ Standard protocols
- ✧ Seamless programming model
- ✧ Scalable
- ✧ Multi-tenancy



Scalability and multi-tenancy

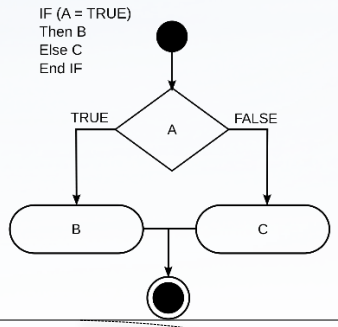


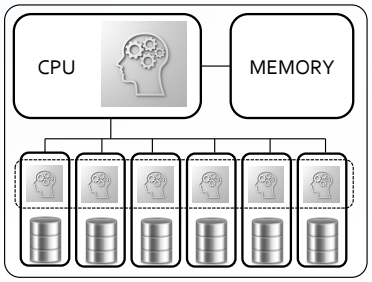
In-Situ Use Cases

filtering, search, grep

> = ≤

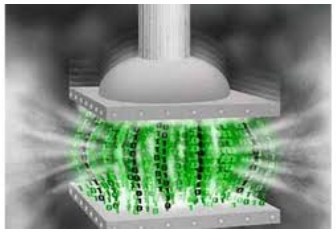
< ≠ ≥





ideal for IO intensive workloads
seamless
versatile
scalable

(de)compression off-load



GZIP, LZW, BZIP…

security

encryption & authentication (data/video streaming)

Use Cases Examples

> Shell scripts

> Video analytics

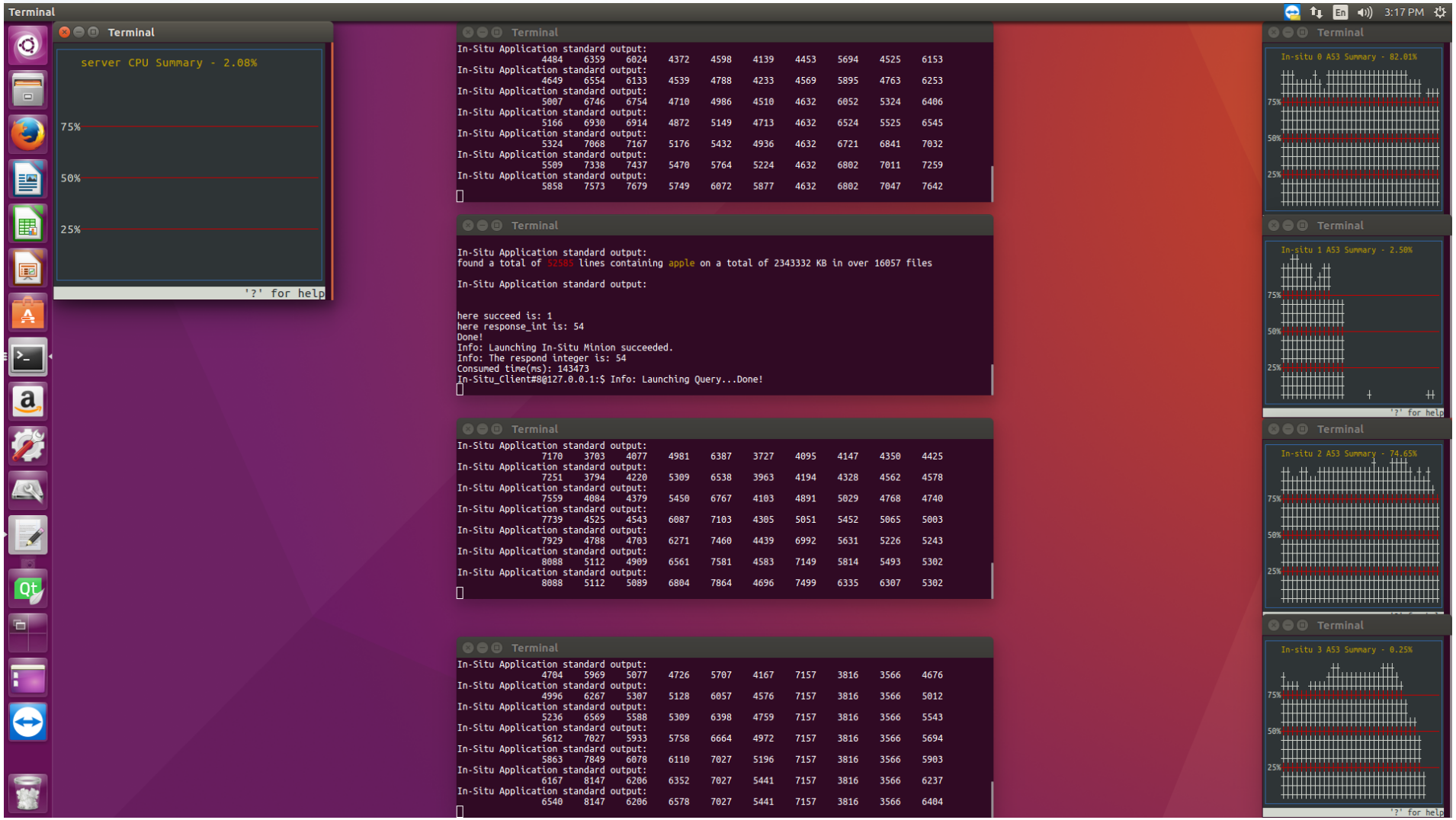
Distributed *grep*

Shell scripts of reasonable complexity have been tested on the In-Situ Development platform.

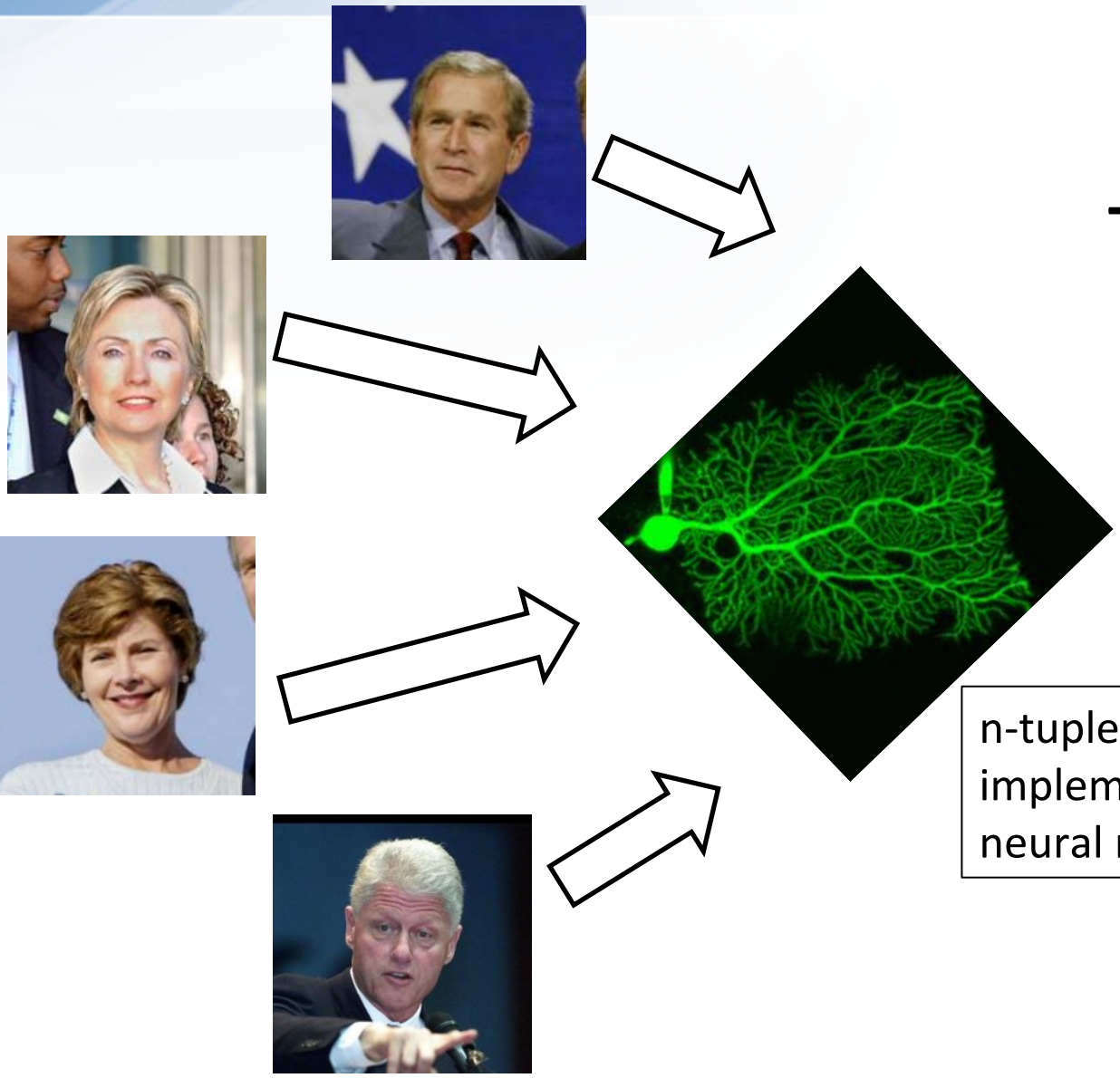
```
> for name in `find /mnt/catalina0 -name *.zip` ; do unzip -p
$name; done | grep -ci " orange ";
```

- recursively searches for compressed files in a directory
- decompresses the file
- searches for a keyword
- counts the number of times the keyword was found

Distributed *grep*



Facial Recognition Using Neural Networks

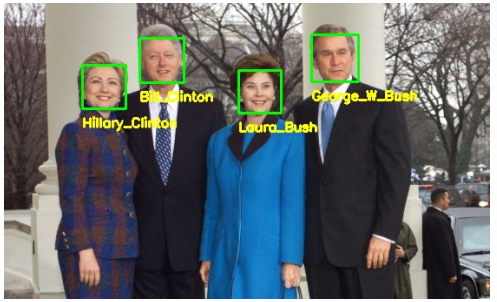


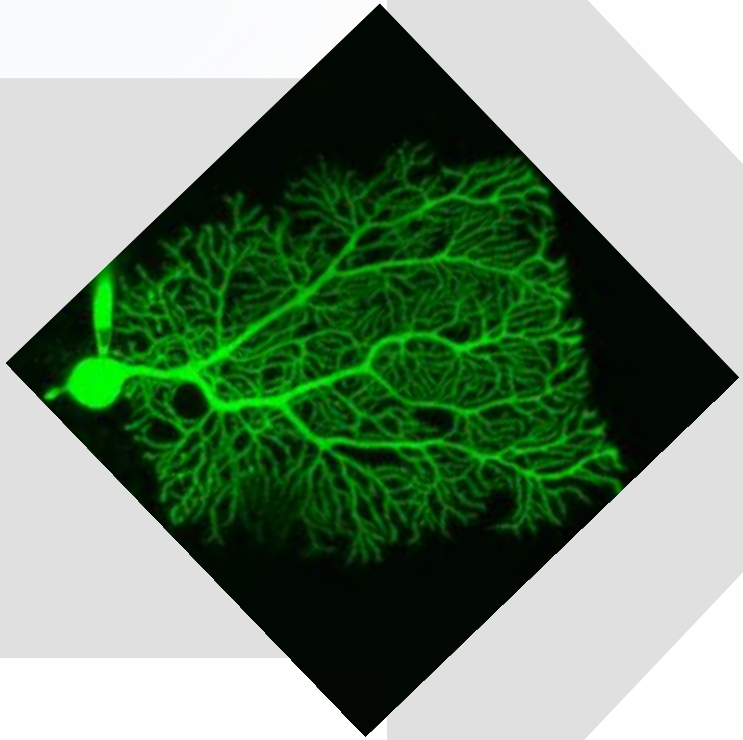
Training phase

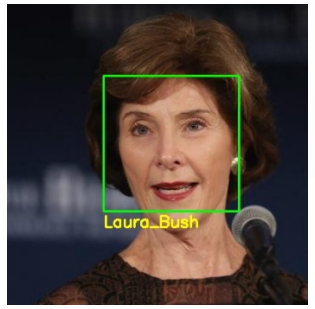
n-tuple classifier
implemented as a weightless
neural network

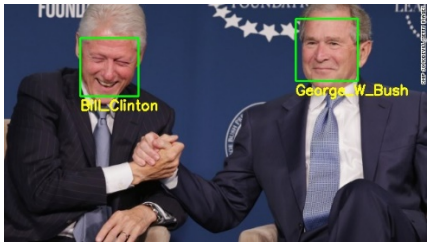
Facial Recognition Using Neural Networks

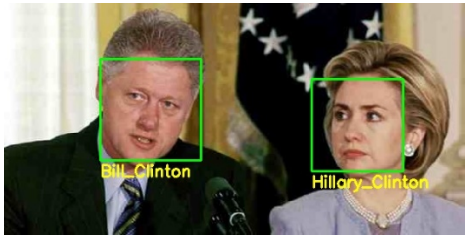
Inference phase



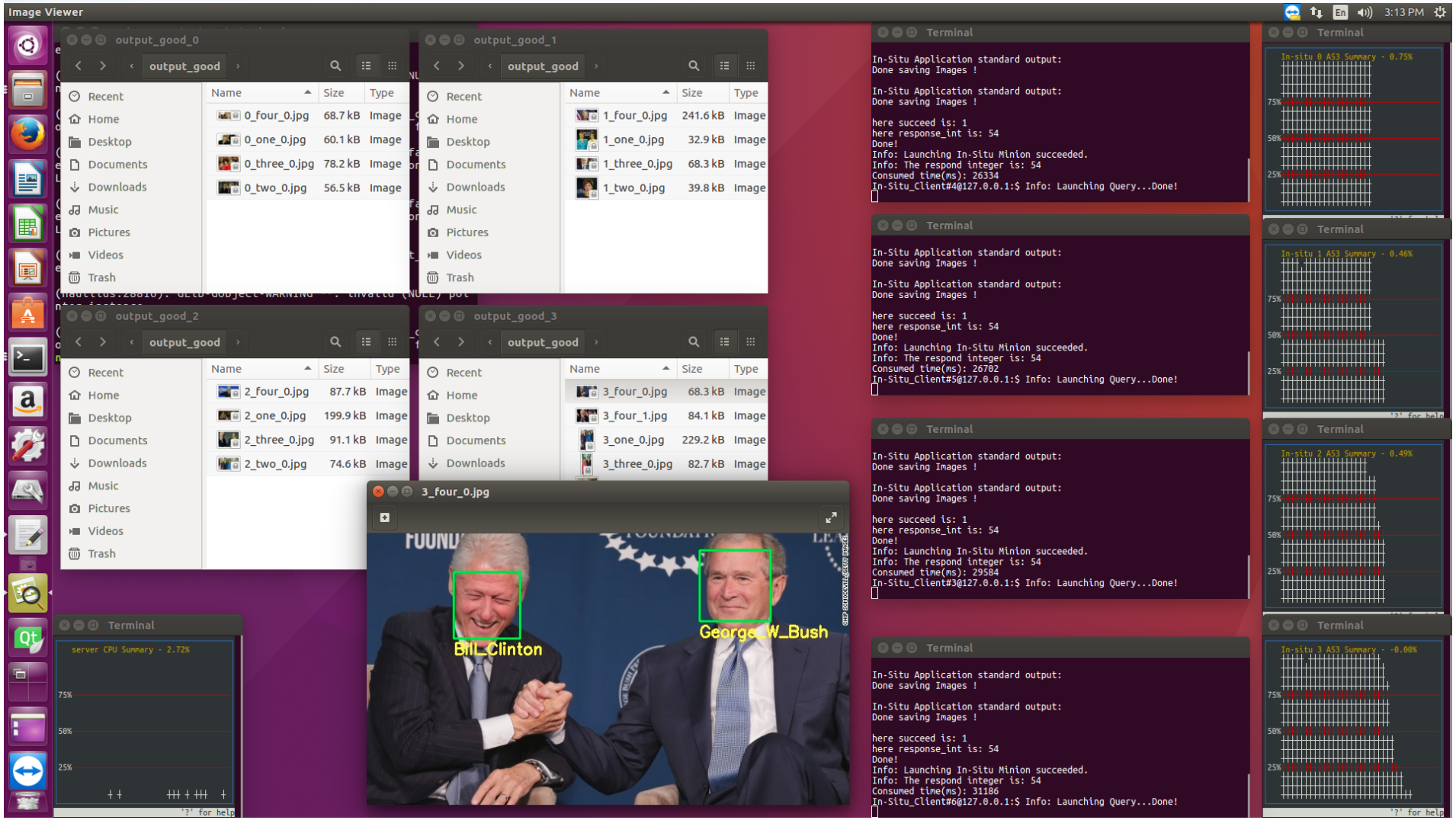




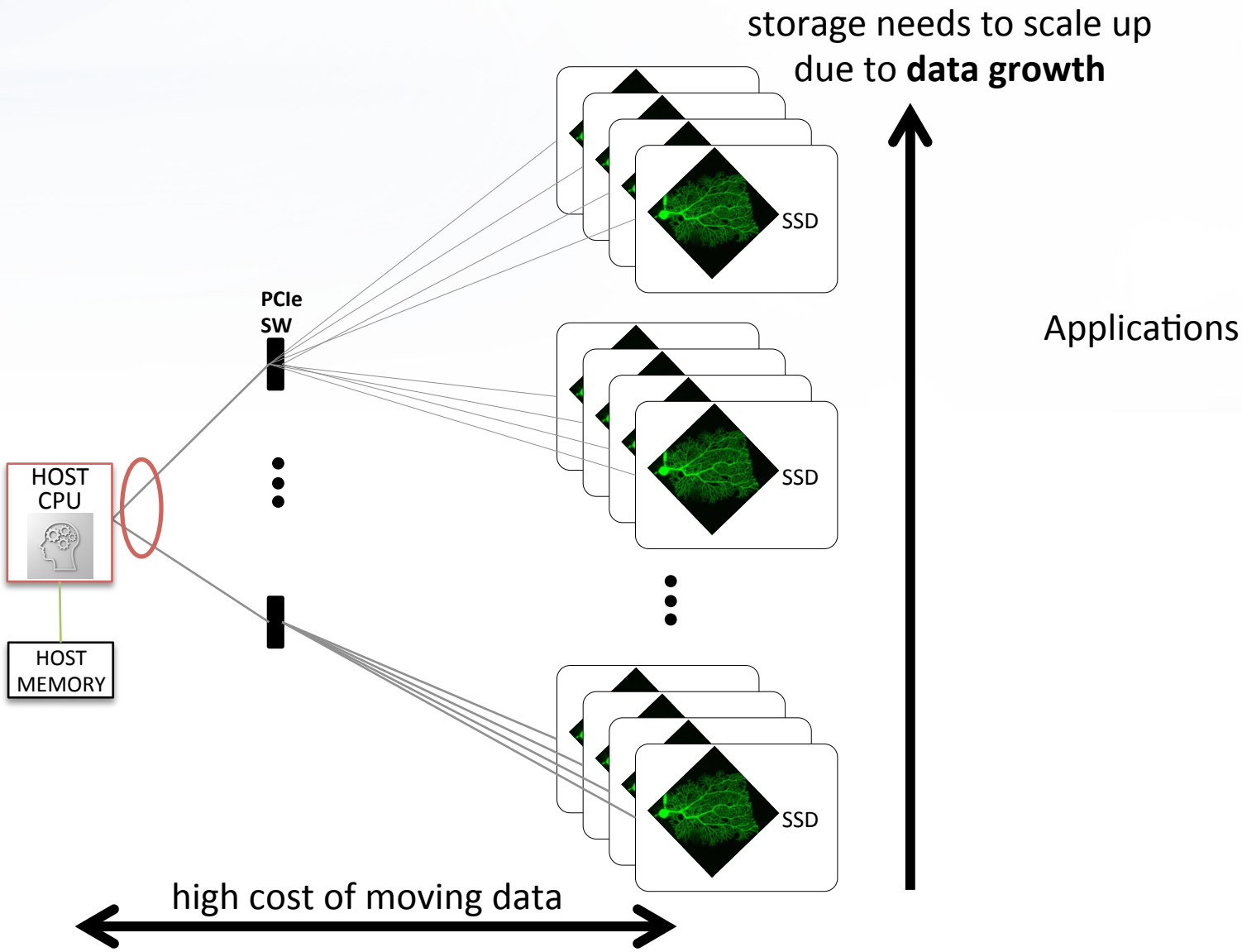




Facial Recognition Using Neural Networks

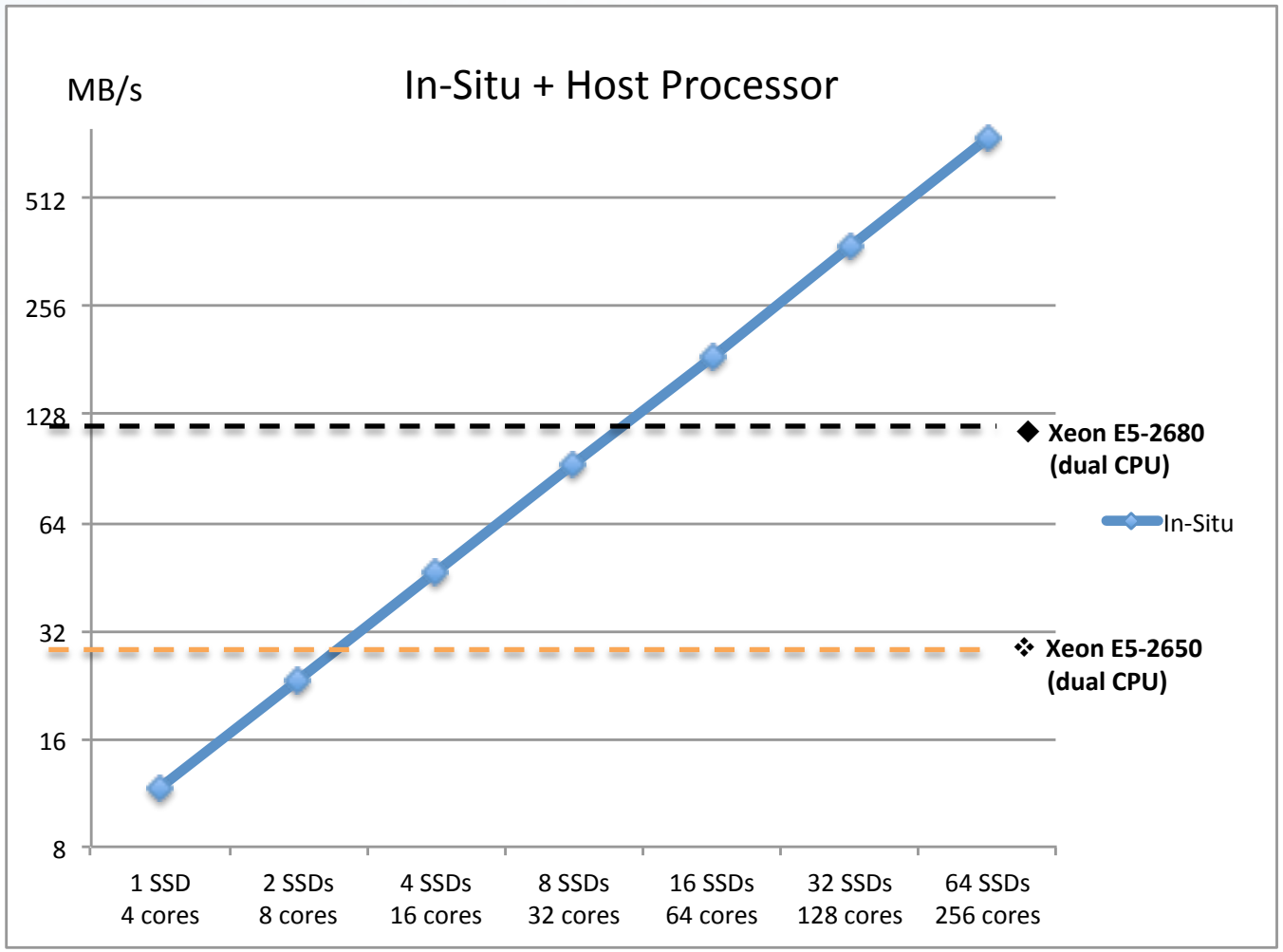


Distributed Processing with Embedded AI



Performance

Example: distributed BZIP2 Compression

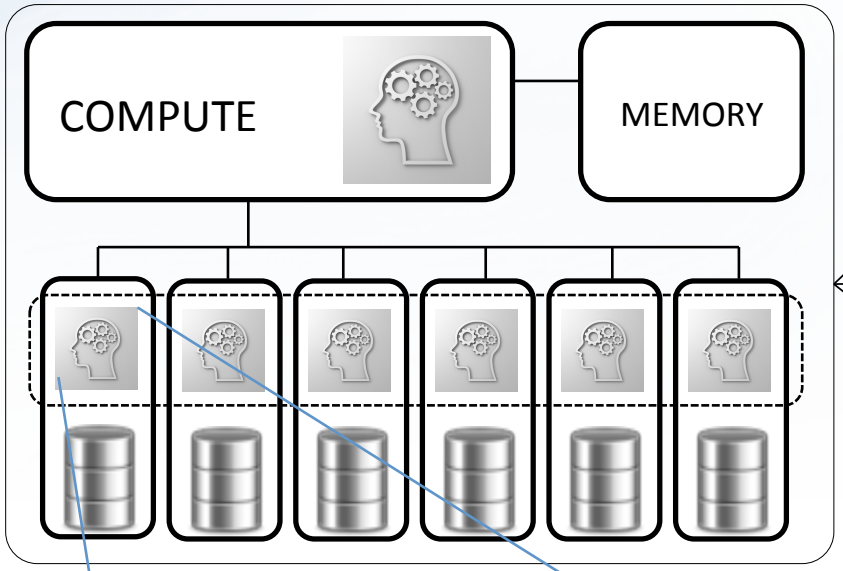


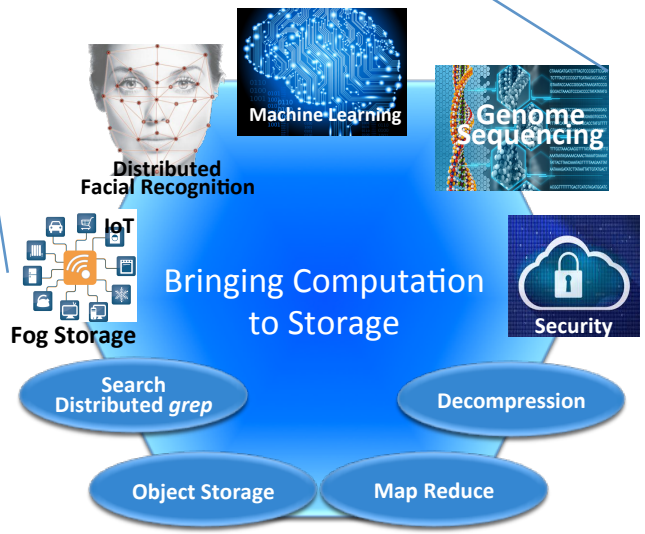
Conclusion

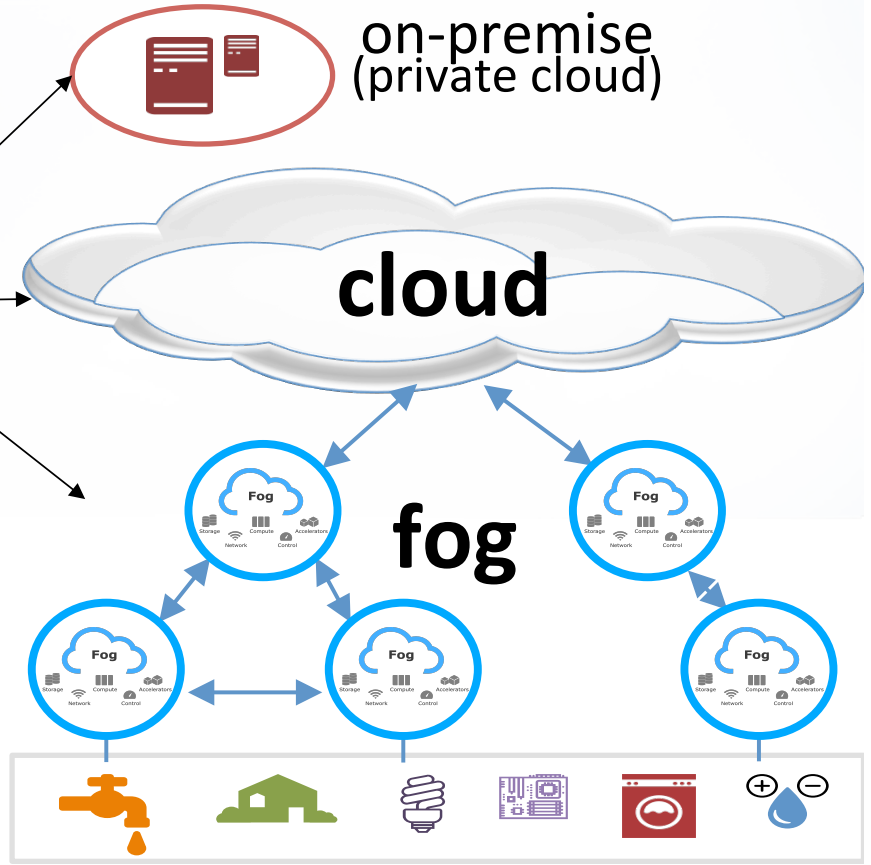
~~In-Situ vs. Host Processor~~

In-Situ + Host Processor

Conclusion







Moving computation closer to data

- Enables distributed processing
- Processing scales with storage capacity
- Optimized for IO intensive applications

Thank You