



Enabling Data Analytics in a SSD

A Converged Platform

Vladimir Alves, PhD

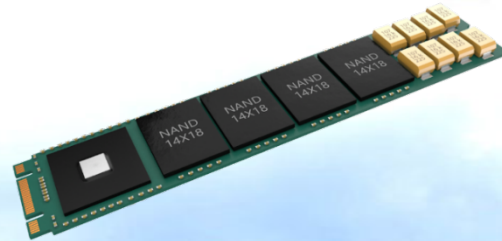
CTO – NxGn Data

The Evolution of Storage



Since 1956 storage devices have been performing the same basic functions: read and write and sometimes fail...

Mine Big Data in Storage





Computational Storage

A Converged Platform that takes compute to data



High performance storage



Computational capabilities



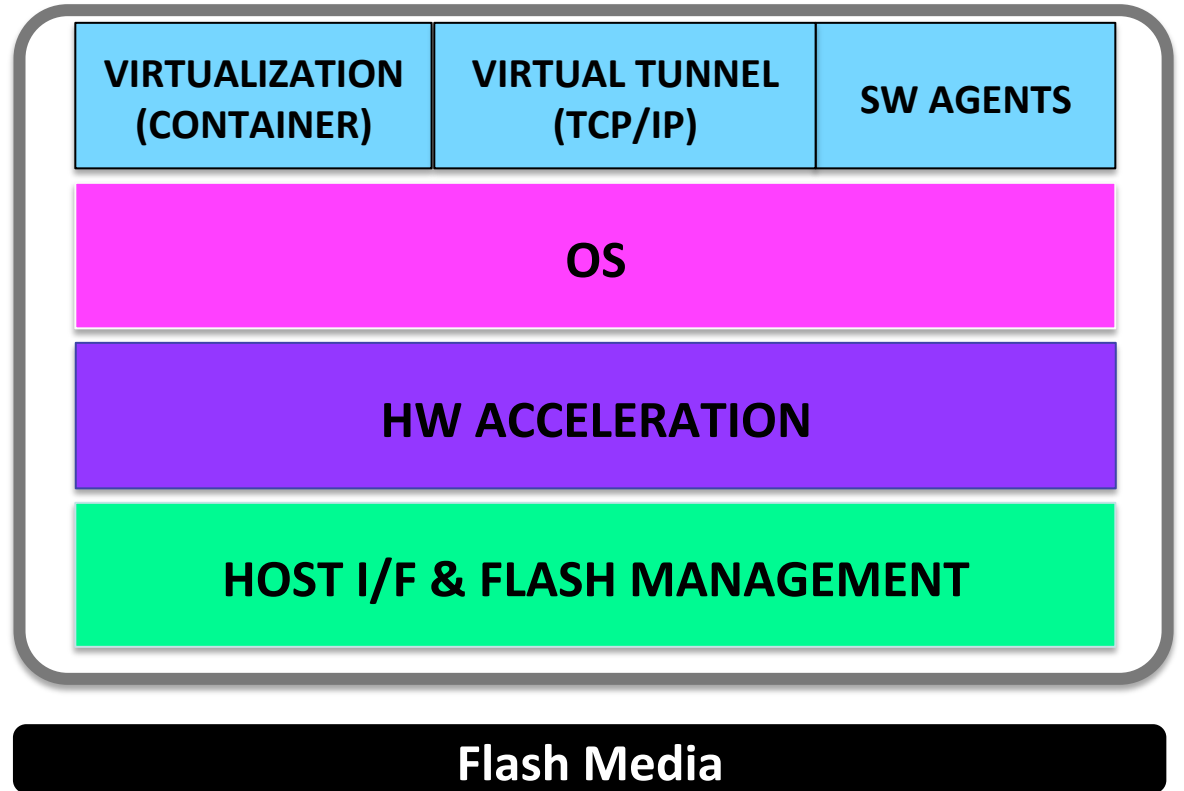
Virtualization

A Paradigm Shift in Storage



In-Situ processing

Flash Memory Summit 2015
Santa Clara, CA



A Paradigm Shift in Storage



In-Situ processing

In-Situ Processing

- **~10x** acceleration in data analytics
- **Virtually** eliminate network & interface traffic
- **90%** reduction in energy consumption
- **75%** cost reduction

A Paradigm Shift in Storage



In-Situ processing

Distributed Grep on Amazon S3/ CloudFiles

- Low-cost *search/scan* feature in cloud-storage

In-situ Cloud-enabled MapReduce

- Low-cost MapReduce embedded into cloud-storage

Swift/Ceph Storage Node

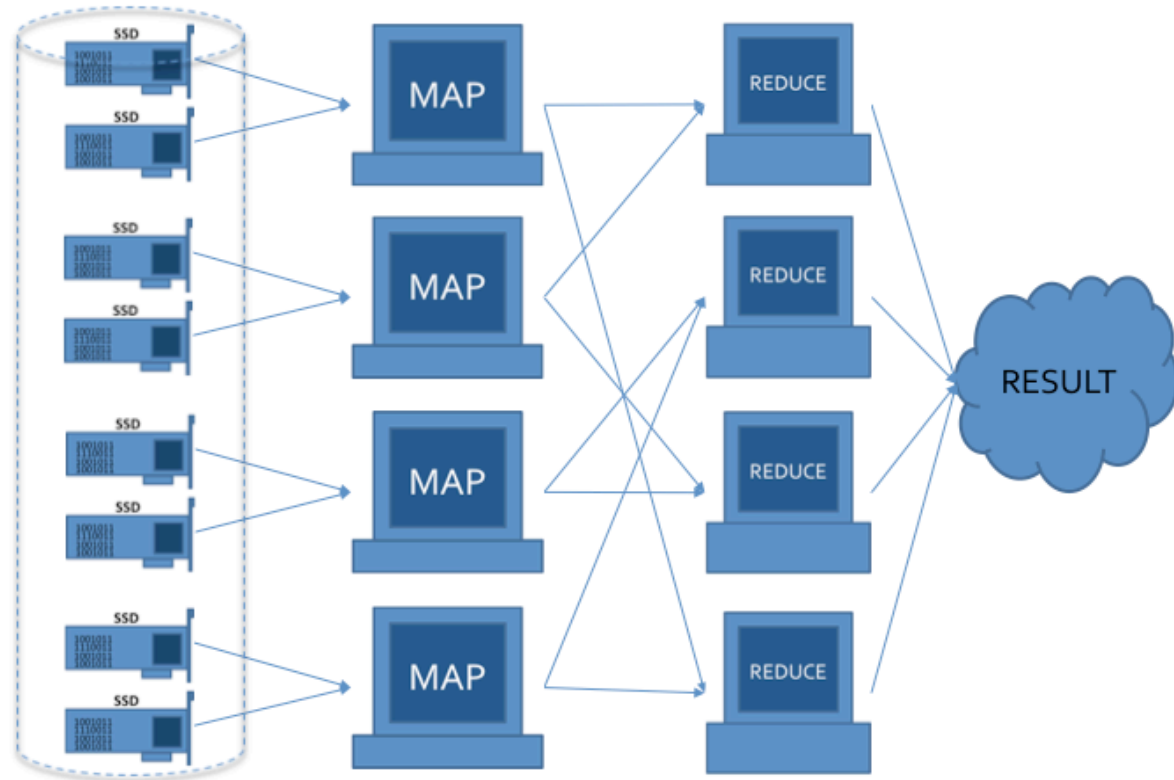
- Lightweight, low-cost Storage Node for scalability

Making Hadoop Fly in Clouds

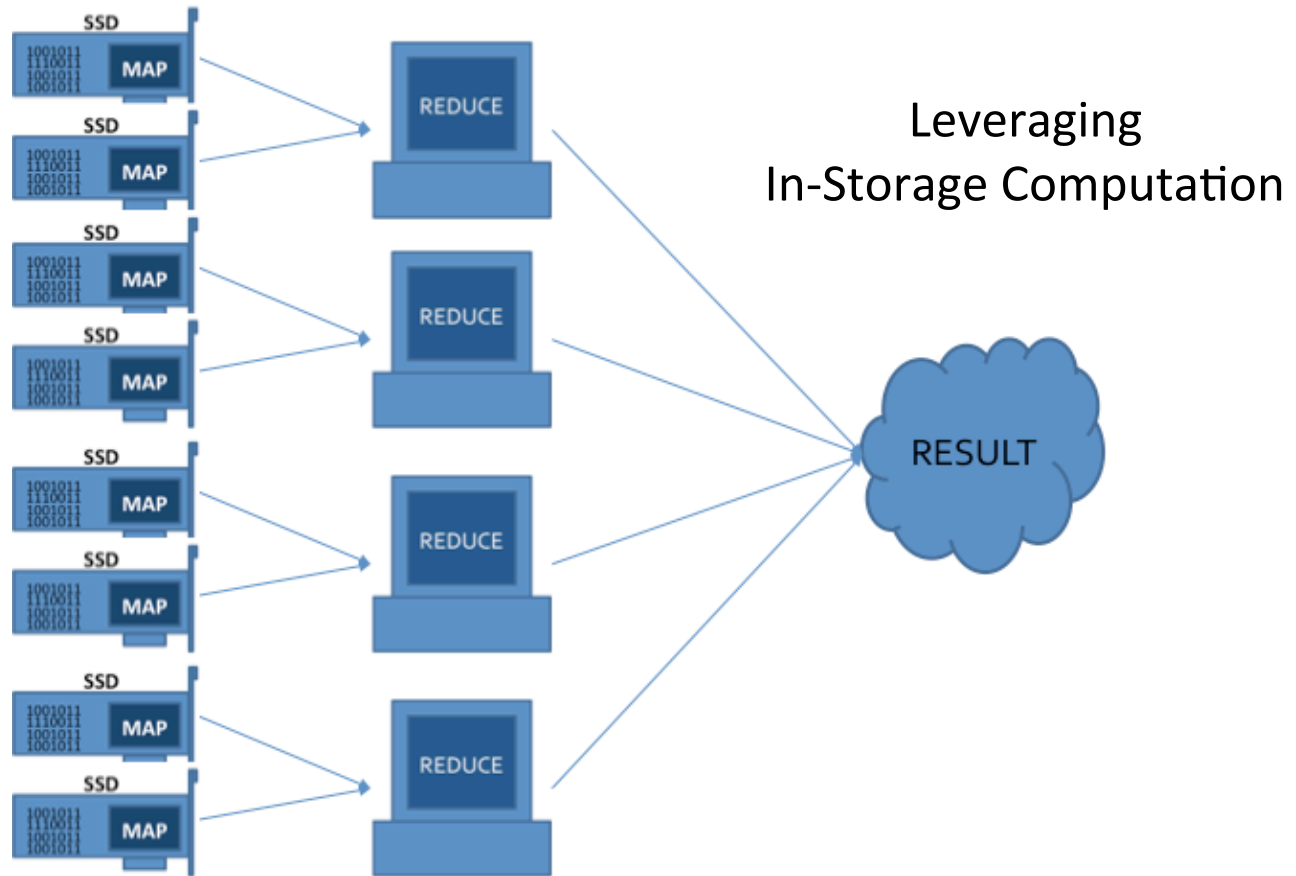
- *Smarter* cloud storage for performance optimization

And more...

System Use Case: MapReduce



System Use Case: MapReduce

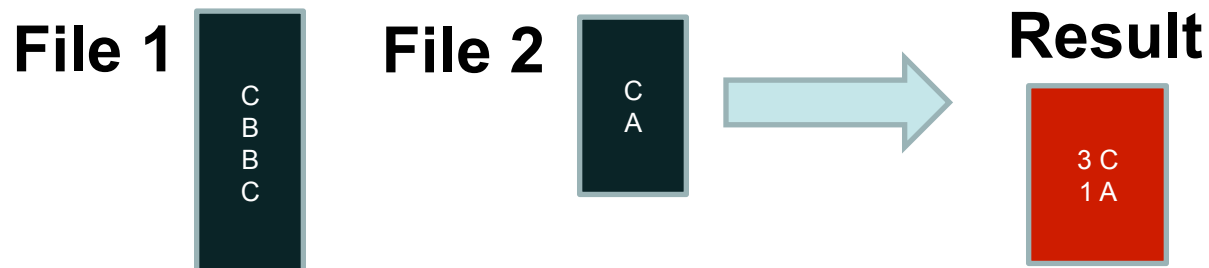


System Use Case: Distributed Grep

Common Unix command for search:

```
grep -Eh <regex> <inDir>/ * | sort | uniq -c | sort -nr
```

- counts lines in all files in <inDir> that match <regex> and displays the counts in descending order





System Use Case: Distributed Grep

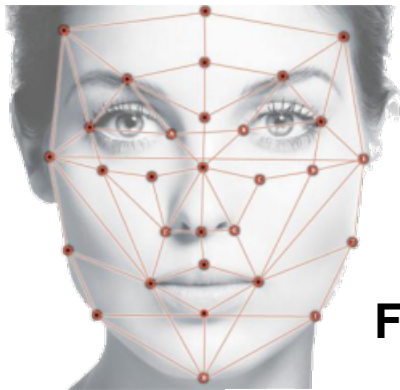
Map function in this case:

- input is (file offset, line)
- output is either:
 1. an empty list [] (the line does not match)
 2. a key-value pair [(line, 1)] (if it matches)

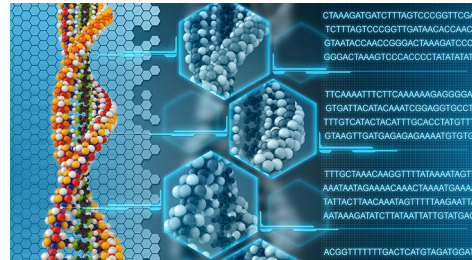
Reduce function in this case:

- input is (line, [1, 1, ...])
- output is (line, n) where n is the number of 1s in the list.

Future Applications



Facial Recognition



Genome Sequencing

**High Performance
Scientific Computing**



IoT



Thank you

NxGn Data