HPDA:
a Storage perspectives

GIULIANO TAFFONI, SARA BERTOCCO

IVOA 2018 – 31 MAY 2018 – VICTORIA - CA

Horizon 2020

What is HPDA?

The ability of increasingly powerful HPC systems to **run data-intensive problems** at larger scale, at higher resolution, and with more elements (e.g., inclusion of the carbon cycle in climate ensemble models)

The proliferation of **larger, more complex scientific instruments** and sensor networks, from "smart" power grids to the Large Hadron Collider and Square Kilometer Array.

The growth of stochastic **modeling, parametric modeling** and other iterative problem-solving methods, whose cumulative results produce large data volumes.

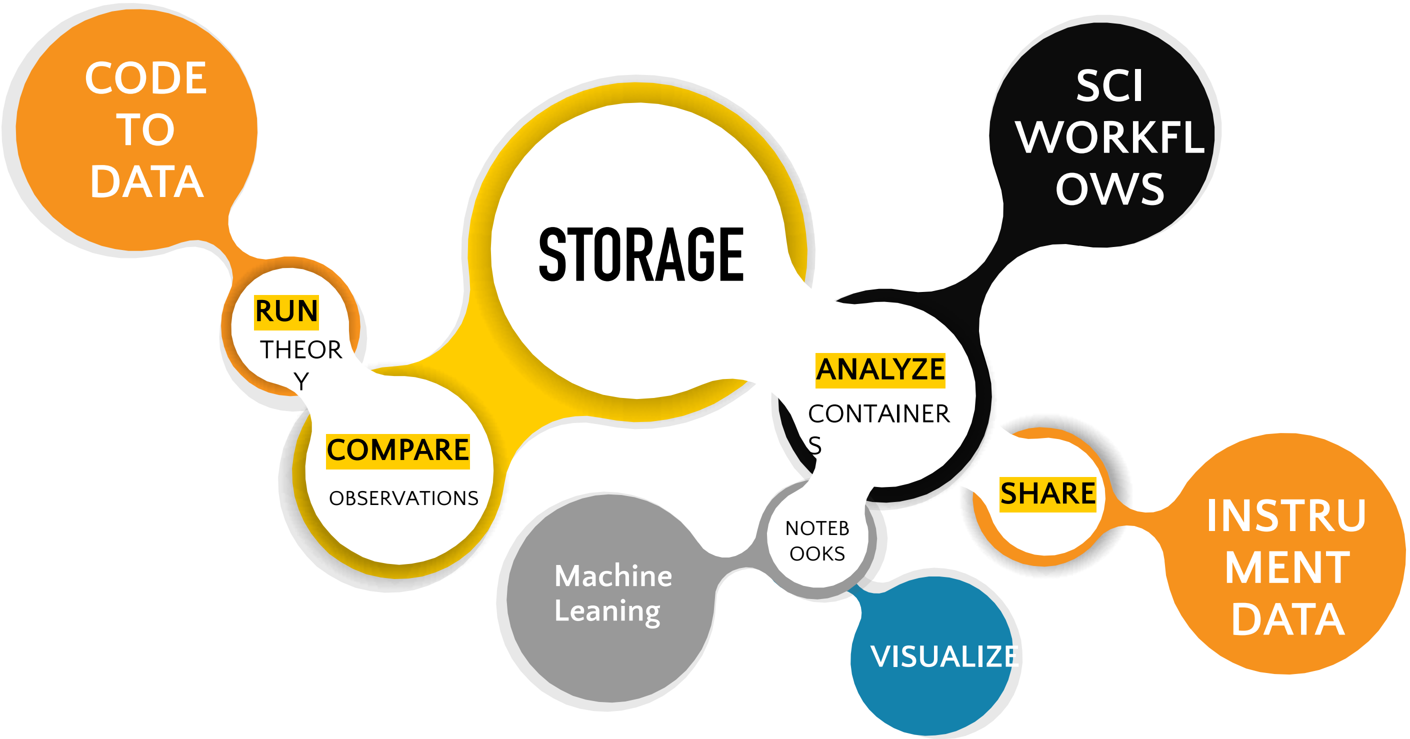
The availability of newer **advanced analytics methods and tools**: MapReduce/Hadoop, graph analytics (NVIDIA IndeX), semantic analysis, knowledge discovery algorithms (IBM Watson), COMPS and pyCOMS, and more

The escalating need to perform advanced analytics in **near-real time**—a need that is causing a new wave of commercial firms to adopt HPC for the first time

HPDA is…


BIG
DATA
STORAGE
ANALYTICS
COMPLEX
SOCIAL
RELATIONAL
LARGER
CONTINUES
CITATION
TENS
SET
USE
BIOLOGICAL
PROCESSING
CAPACITY
HUNDREDS
NETWORKS
DATABASES
RECORDS
SEARCH
DIFFICULTY
ALSO
TOOLS
DISK
TARGET
APPLIED
SENSOR
MAY
SHARED
EXAMPLES
EVERY
DEFINITION
CURRENT
MOVING
WITHIN
PRACTITIONERS
CAPTURE
BUSINESS
INTERNET
DISTRIBUTED
SETS
GENOMICS
MANAGEMENT
TERABYTES
PETABYTES
INCLUDE
SYSTEMS
TOLERABLE
SIZE
PARALLEL
SAN
GROW

Why are we talking about storage?



Storage challenges…

exabyte-scale storage needs in the coming decade, with many **projects** generating and processing hundreds of terabytes of data today

observational and simulation data in workflows that **require data to be co-located for effective analysis.**

Data management needs to extend **beyond computing centers to the wide-area network**

This includes expanding capabilities around **metadata storage, searching and querying, and event triggering.**

Complex workflows that must **connect and integrate simulation and analysis**.

Apply **new forms of data analysis and analytics**, including machine learning, to effectively process the massive amounts of data.

What do we require to storage?

Reliability and manageability

Capacity and durability

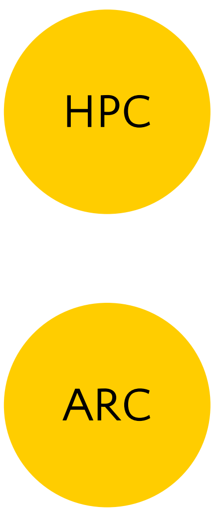
Optimized for performance (low latency and high throughput)

Scalability: grow over time to allow for external project to add additional space

Ability to ingest and store data from remote instruments

Access controls for “publishing” and sharing

State of the Art

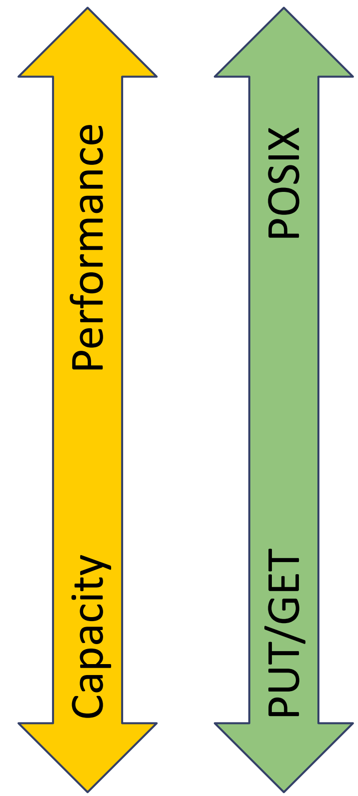


Temporary storage, which contains data being actively used by simulation and data analysis applications over the course of hours to days.

Campaign storage, which contains data being actively used by larger workflows and science projects over the course of weeks to months.

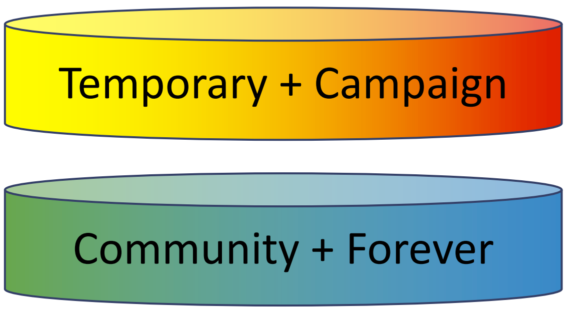
Community storage, which contains larger datasets that are shared among different projects within a scientific community over the course of years.

Forever storage, which contains high-value or irreplaceable datasets indefinitely



Lawrence Berkeley National Laboratory

What is the perspective



Magnetic disks from capacity+bandwidth to only capacity.

NVM for HPC and HPDA.

Tapes unlikely to be the optimal long-term solution (> 2025).

Magnetic disks from capacity+bandwidth to only capacity.

Increasing difficulty in scaling POSIX-based parallel filesystems to extreme scales.

Object stores, eschew POSIX I/O semantics in favour of stateless put/get operations and immutable data objects (DAOS, Ceph, Swift, etc)

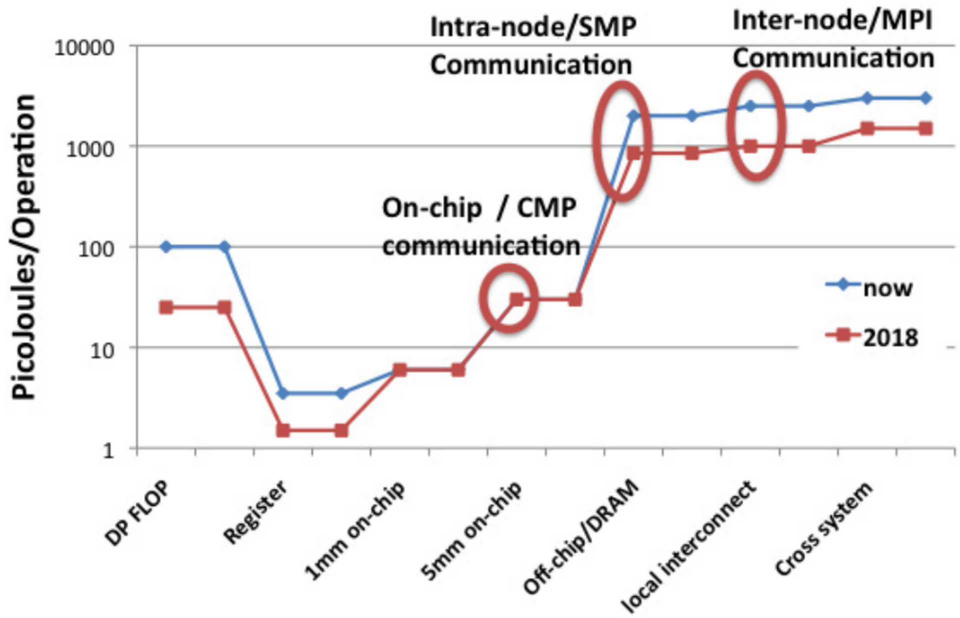
POSIX moves from a native interface to a middleware layer

Emergence of other I/O middleware packages like HDF5 and ADIOS.

Minimize data movement

Move your code close to the data when possible (Tempory+Campaign)

Move your storage close to high throughput WAN



Cloud IO ⇔ HPC IO

Many of emerging data workloads are driven by **machine learning and other data analytics techniques** that rely on workflow frameworks (e.g. Apache Spark), analytics packages (e.g., Caffe and TensorFlow), and domain-specific libraries that traditionally have **not been used in HPC**.

Many of these emerging applications areas are associated with **observational and experimental facilities** that are already generating large volumes of data

New approach to storage

Object Storage

What is an Object Storage

Most common cloud storage

Manage data as object (unstructured)

Keys maps data to blocks

Cost effective compared with other solutions

Scalable and redundant

High throughput (but not low latency) with multiple async requests

HTTP range requests

CEPH, SWIFT, AMAZON S3, etc…

Working with Object Storage

Python APIs and RESTful API to work with notebooks

```
[1] from swiftclient import client swift =
client.Connection(authurl=url, user=username, key=password,
tenant_name=project_name, auth_version='2')
[2]container_name=»ivoa"
[3] swift.get_container(container_name)
```

Containers and Objects

Organize data in containers and objects: logical and practical view (e.g. to instruct your code to treat data in different way)

Note: Object storage is just one component of a storage middleware

A view on the future

Increasing difficulty in **scaling POSIX-based** parallel file systems to extreme scales.

New software interfaces are a requirement to make optimal use of emerging low-latency storage hardware.

Object stores, initially driven by the extreme-scale I/O needs of cloud providers, eschew POSIX I/O semantics in favour of **stateless put/get operations and immutable data objects**.

POSIX will survive

Because these new **non-POSIX interfaces** are optimized for performance over usability, though, **I/O middleware** will become more important to bridge the semantic gap between the I/O operations that scientific applications demand and the I/O operations supported by the underlying storage system.

POSIX interface will be implemented as **middleware atop a native object** interface rather than being the lowest-level user interface to storage.

I/O middleware packages: HDF5 or ADIOS

Challenges for middleware

Ability to **efficiently index, search, and describe datasets** is in charge of middleware

Improve capacity of Object Storage providing capacities that are now in charge of applications or services.

ADD metadata: contains predefined information about a data object, such as name, ownership, timestamps, etc., as well as unlimited number of user-defined tags as rich metadata

A VO solution

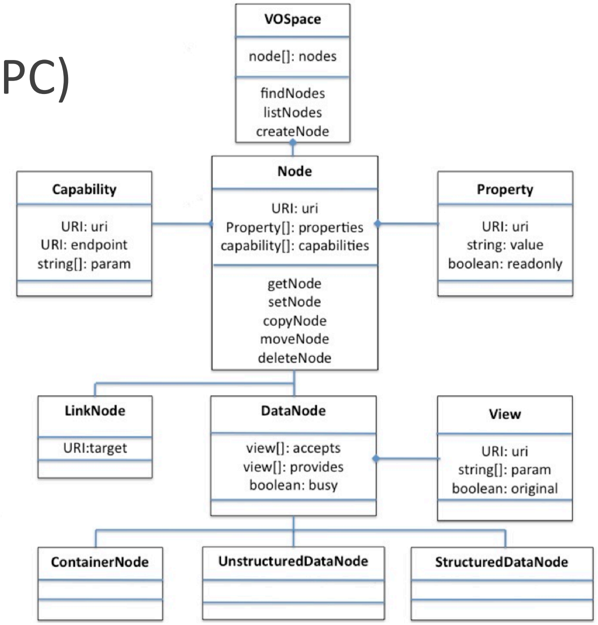
VOSpace is a data management solution and also an Object storage that answer to a number of open questions from HPDA.

We need to improve metadata management (HPC)

- Reduce latency (e.g. SoMeta)

We need to improve transfer service

- Increase throughput
- multiple layers of storage,
- Asynchronous I/O



Conclusions

The hierarchical file system of today will only be one of a number of views through which users can interact with their data. Alternate views of data, searchable by user-defined attributes associated with data, are a feature of today's cloud-based storage that will find their way into the HPC space.

Search and discovery based on user-defined metadata will be better integrated directly into the storage system.

Connect users with new technologies, and ensure technologies meet user and operational requirements.