Geographically Distributed Software Defined Storage (proposal)

Sergey Khoruzhnikov[1], *Vladimir Grudinin*[1], *Oleg Sadov*[1], ***Andrey Shevel***[1,2], *Arsen Kairkanov*[1], *Oleg Lazo*[1], *Anatoly Oreshkin*[2]

1 ITMO University, S.Petersburg (Russia)

2 National Research Centre "Kurchatov Institute"
PETERSBURG NUCLEAR PHYSICS INSTITUTE

The 7th International Conference "Distributed Computing and Grid-technologies in Science and Education" 4-9 Jul 2016, Dubna, Russia

Presenter: Andrey Y Shevel

Outline

- **Sources of Big Data.**
- **Storage.**
- **Software Defined Storage.**
- **Geographically Distributed Software Defined Storage (GDSDS).**

Scientific sources of Big Data

- **Scientific experimental installations**
  - **http://www.iter.org - International Thermonuclear Experimental Reactor** *(coming)*
    - **~1 PB/year**
  - http://www.lsst.org **- Large Synoptic Survey Telescope**
    - **~10 PB/year**
  - http://www.cern.ch — CERN, http://www.fair-center.eu - FAIR, http://www.cta-observatory.org - CTA — The Cherenkov Telescope Array
    - **~20+PB/year (each site)**
  - **https://www.skatelescope.org/ - Square Kilometre Array**
    - **~300-1500 PB/year**
- **Marginal remark: total volume of data in the World grows two times an year, i.e. around 75% of data were written last two years.**

Storage

- All types of storage are distributed (depends on the scale of distribution: among disk drives, servers in Data Center, or amongst Data Centers (large RTT => 10 msec).
- Several of storage systems for science are already running and proposed.
- Many commercial companies suggest distributed data storage solutions: Google (Mesa: GeoReplicated, Near RealTime, Scalable Data Warehousing), Dropbox, Box, Adrive, Amazon, DDN Storage, …
- Which are appropriate solutions for globally distributed data storage in scientific research and education ?
    - Obviously we need for *software defined* solutions.

Main features of SDS

Software Defined Storage should include:

- Automation – Simplified management that reduces the cost of maintaining the storage Infrastructure.
- Standard Interfaces – APIs for the management, provisioning and maintenance of storage devices and services.
- Virtualized Data Path – Block, File and Object interfaces that support applications written to these interfaces.
- Scalability – Seamless ability to scale the storage infrastructure without disruption to availability or performance.

Technical details of GDSDS

- Important features:
  - Data storing and Data transfer
    - Reliability: *data replication, erasure coding*.
    - Reduce the volume: *Data compression*.
    - Security: *Data encryption, ACL*.
  - GDSDS Web portal and GDSDS CLI.
  - Network architecture.
  - Caching, Tiering.
  - Automatic storage deployment by user request.

Network aspects on GDSDS

- First of all we have to keep in mind the CAP theorem:
  - Theoretically NOT possible to guarantee all below requirements at the same time.
    - Consistency
    - Availability
    - Partitioning

Similar developments

- Project OsiRIS at University of Michigan - https://indico.cern.ch/event/466991/contributions/1143627/

Basic assumptions on GDSDS

- It is assumed
  - GDSDS consists of several groups of storage servers located in geographically different regions.
  - Group of servers are connected by a number of parallel virtual data links.
    - Data links might have different features: speed, price, encryption type (including *quantum encryption), etc*.
  - Data links has to be configured with SDN.
  - Client can ask to perform a number of operations:
    - Create, Delete, Replicate, Migrate, etc an instance of Virtual Storage allocated on GDSDS. The instance might be created with different SLA
    - Write/Read data to/from the instance of Virtual Storage.

Examples for SLA

- Data Encryption (with specific type of Encryption).
- Data Compression.
- On one specific DC or on many DCs with specific types of Data Links.
- Type of backend.

Allocation of instance of Virtual Storage

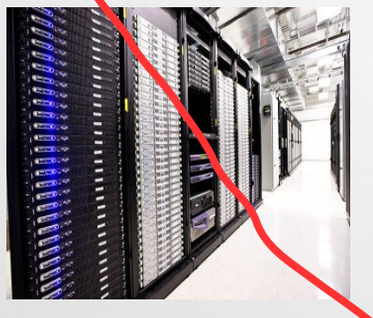
Clients-1

Clients-2

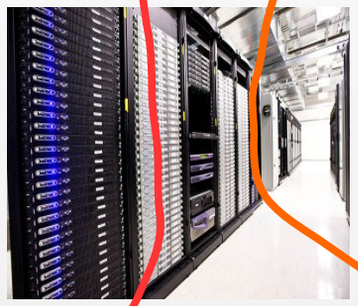
client=VM or mobil device

Virtual Storage-1

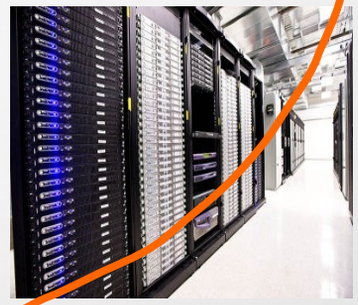
Virtual Storage-2



DC-1



DC-2



DC-3

Development process consideration

- During implementation GDSDS the project working repository is strongly required (any update or/and installation must be done only with this working repository).
- Update the working repository is separate specific activity.
- Several existing SDS systems might be considered as backend: SWIFT, CEPH, EOS(?), existing cloud storage is possible as well.
    - CEPH is under testing now as backend for the proposal.
    - Tech implementation with docker (under intensive testing now).

Who are developers

- We (volunteers from PNPI and ITMO) are positioning the proposal as open source development.
- It is supposed that any person with interest to such the topics could take participation:
    - Researchers, students.
- Yes, we are looking for support.

Time for questions & suggestions!

References

- Jakob Blomer // Survey of distributed file system technology // ACAT 2014, Prague (in references) Also *iopscience.iop.org/article/10.1088/1742-6596/664/4/042004/pdf*
- Why so Sirius? Ceph backed storage at the RAL Tier-1.
  - https://indico.cern.ch/event/466991/contributions/2136880/contribution.pdf
- Analysis of Six Distributed File Systems – HAL-Inria - *https://hal.inria.fr/hal-00789086/file/a_survey_of_dfs.pdf*
- https://en.wikipedia.org/wiki/Comparison_of_distributed_file_systems
- XtreemFS is a fault-tolerant distributed file system for all storage needs http://www.xtreemfs.org/
- Software Defined Storage LizardFS is a distributed, scalable, fault-tolerant and highly available file system - https://lizardfs.com/about-lizardfs/

Big Data

- Big Data have the set of features: *triple V*
  - Data flow Velocity
  - Data Volume
  - Data Variety