Networking Research and Education

Greek Research and Technology Network S.A.
GRNET

(Big) Data Management

EGI Federated Cloud - Face to Face Workshop
Amsterdam, 2015-01-21

Christos KK Loverdos

loverdos@grnet.gr

The goal is to present part of the Data landscape in order to initiate a discussion about existing and needed use-cases

The use-cases should drive design and service provisioning from cloud sites

Big Data is very important for the scale we are concerned with. We could start with a definition …

Big Data Definition (Attempts)

- The 3 Vs [http://www.gartner.com/newsroom/id/1731916]
  - Volume, Velocity, Variety
- The 4 Vs [http://www.ibmbigdatahub.com/infographic/four-vs-big-data]
  - Volume, Velocity, Variety, Veracity
- The 4 Vs + 1 C [http://www.sas.com/en_us/insights/big-data/what-is-big-data.html]
  - Volume, Velocity, Variety, Variability, Complexity
- Wikipedia
  - So large and so complex
  - Difficult to process with traditional tools

… but probably it is most important to describe a few concerns that will help us understand the landscape and needs

(The technologies mentioned and their categorisation are indicative.)

~~Big~~ Data Management Concerns

- Schema
- Storage
- Access
- Computation

Data Schema

- Relational
- Non-Relational/Semi-Structured
  - JSON
  - Thrift
  - Avro
  - Protobuf
  - Parquet (columnar)

Data Storage

- Object/Blob Store vs File system
  - NFS
  - HDFS
  - CephFS
  - GlusterFS
  - OpenStack Swift
- NoSQL
  - Cassandra
  - HBase

Data access

- Local mounts
- POSIX semantics
- REST API
  - CDMI
- Interoperability
  - De-facto standards
  - Committee Standards

Computation

- Apache Hadoop
- Apache Hive
- Apache Spark
- Apache Storm
- IPython
    - Prediction: Notebooks are the future

Computation II

- Data-flow pipelines
- Analytics
- Machine Learning
- Graph processing
- Stream processing
- Real-time vs Batch-oriented

Looking ahead

- There is no single best way
  - Slightly different needs call for diff. solutions
- Identify the champions from the use-cases
- Give room to emerging technologies
- It’s all about services
  - A scientific group has requirements
  - A service provider can fulfil their needs

So, step-up and present your use-cases. Service providers will follow and propose solutions.

Thank you !