



# Experiment transfers dashboard on Hadoop

Luca Magnoni  
IT-SDC

22 October 2014



# Experiment transfers dashboards



FAX Monitoring ( RC.50 )  
 Latest statistics update: 2014-10-22T09:00:00.860723

TRANSFER PLOTS (2014-10-22 05:10 to 2014-10-22 09:10 UTC) SLIDING PLOT: GROUPING ▾ TYPE ▾ SERIES ▾ SIZE ▾ STYLE ▾ BIN: SIZE ▾ FORMAT ▾ STEP ▾

▼ Summary

**Common options**

Interval  
Last 4 hours

User protocol  
XRootD

Access type  
Remote access

Transfer VS Reading  
Reading  
Copy

**Client - Server options**

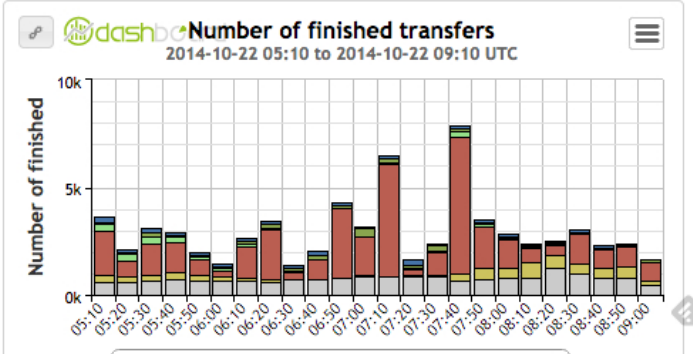
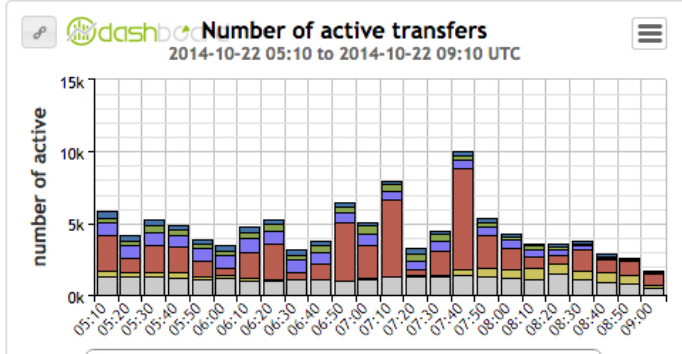
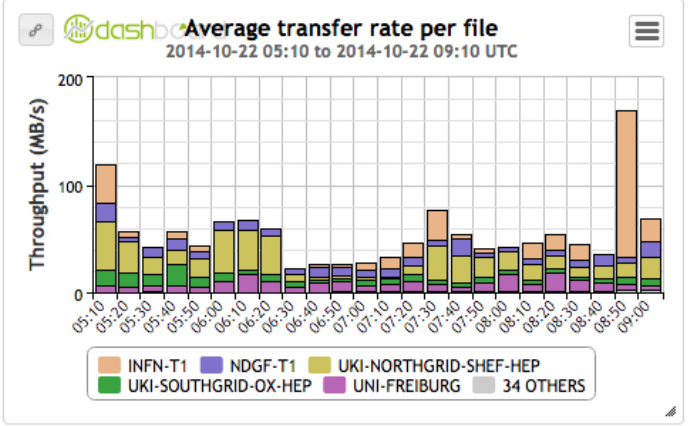
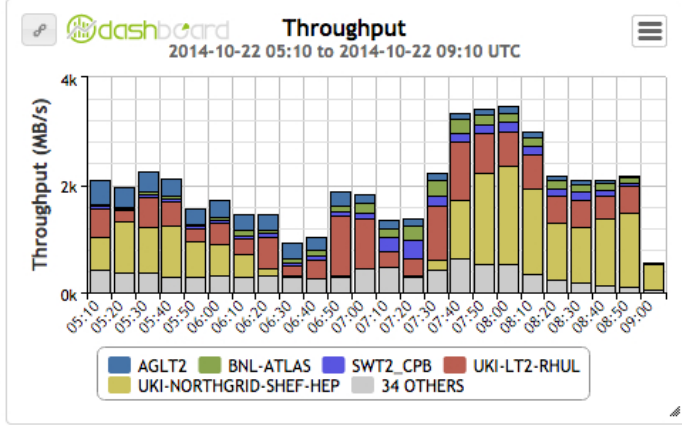
Server:  
Grouping: SITE

Clients:  
Grouping: SITE

Users:

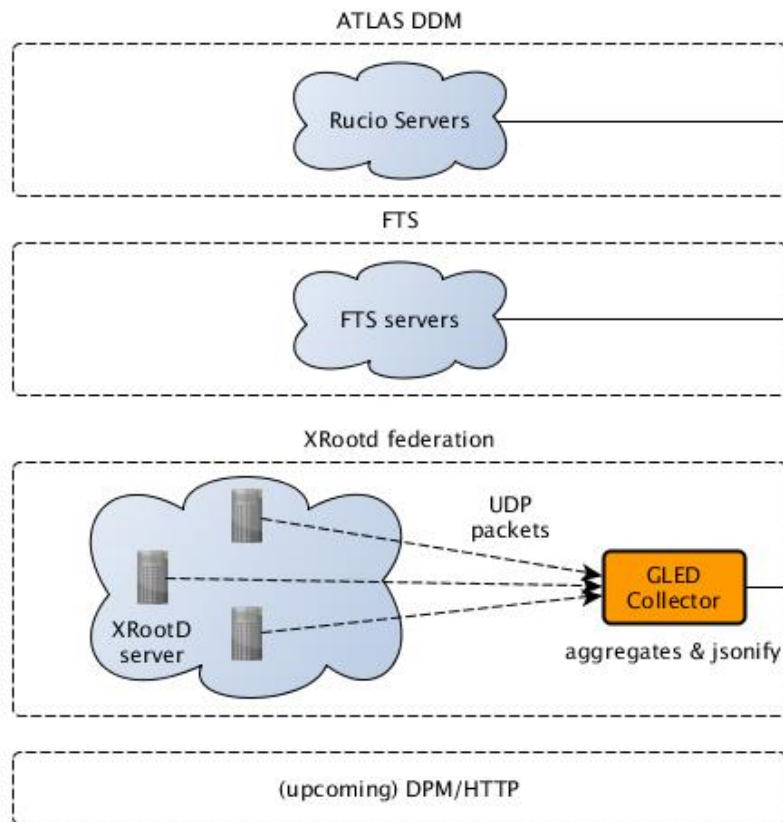
- Interval
- Access type
- Transfer VS Reading
- Sources
- Destinations

Matrix Transfer Plots Access Plots Site Statistics Custom Ranking Plots Site History Access Pattern MAP (Experimental)

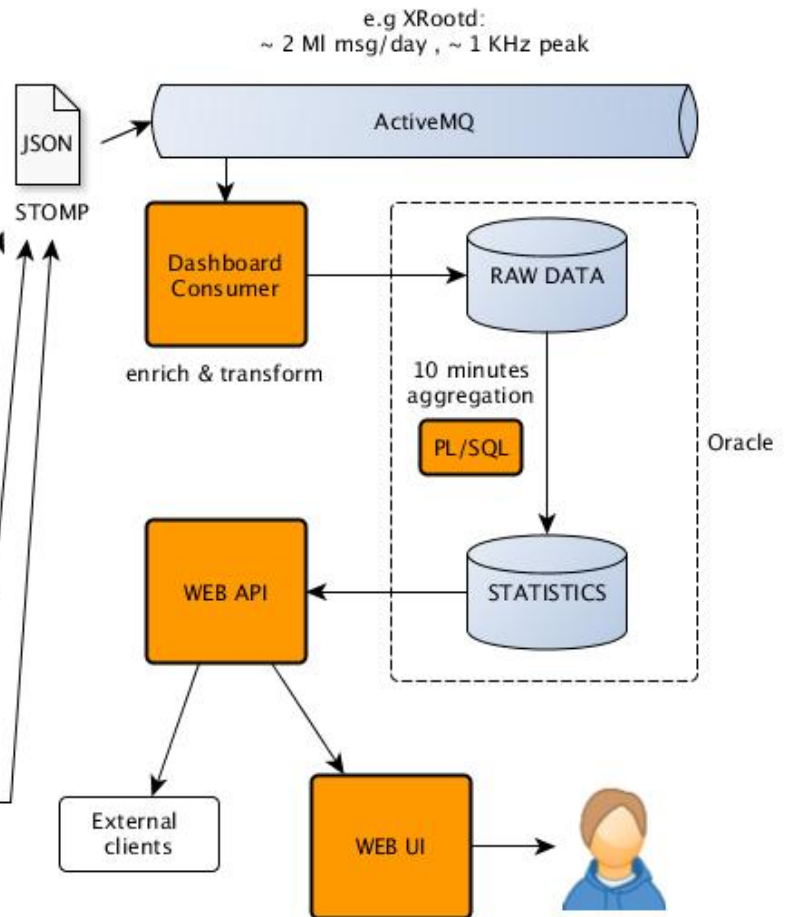


# Today's architecture

## Many data sources

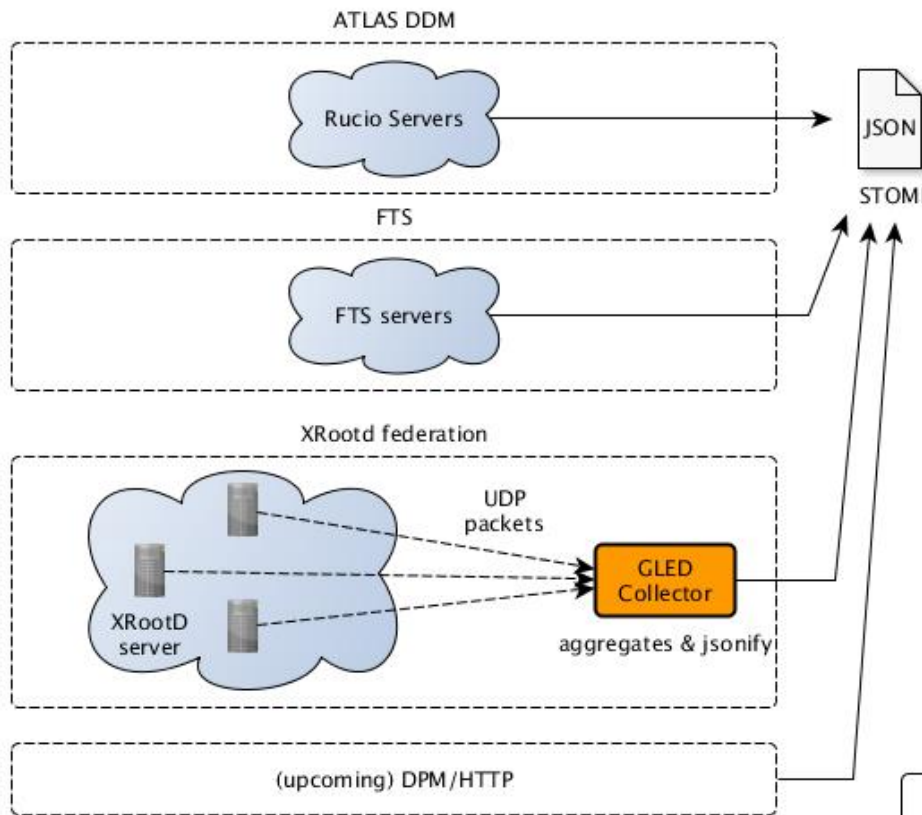


## Common Architecture

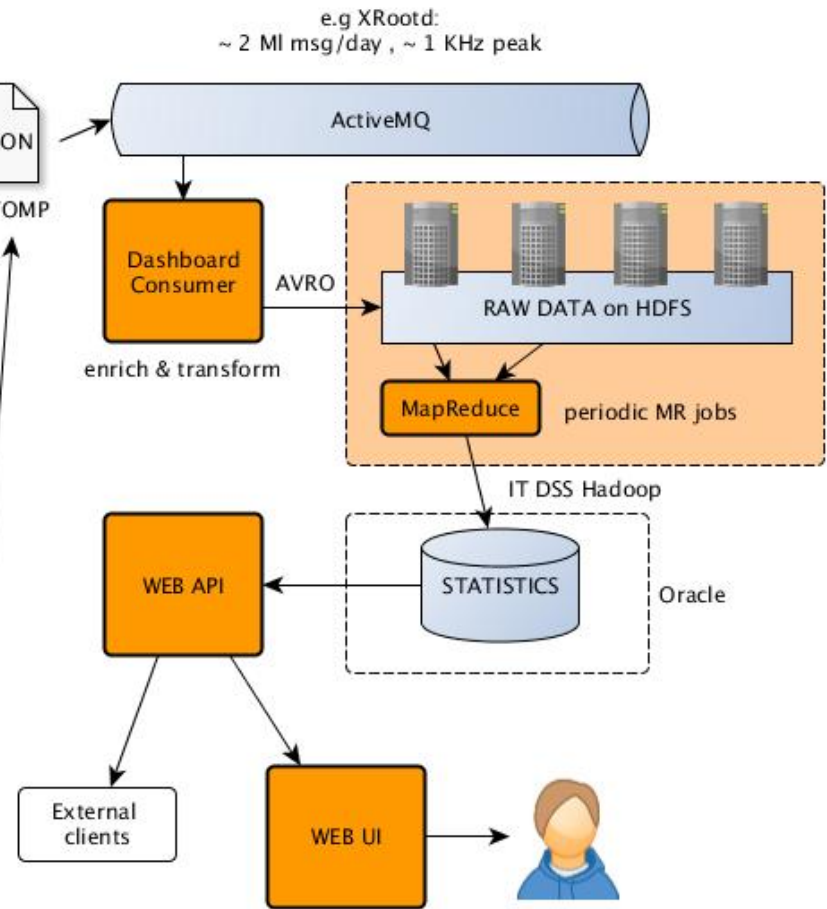


# Architecture evolution on Hadoop

## Many data sources



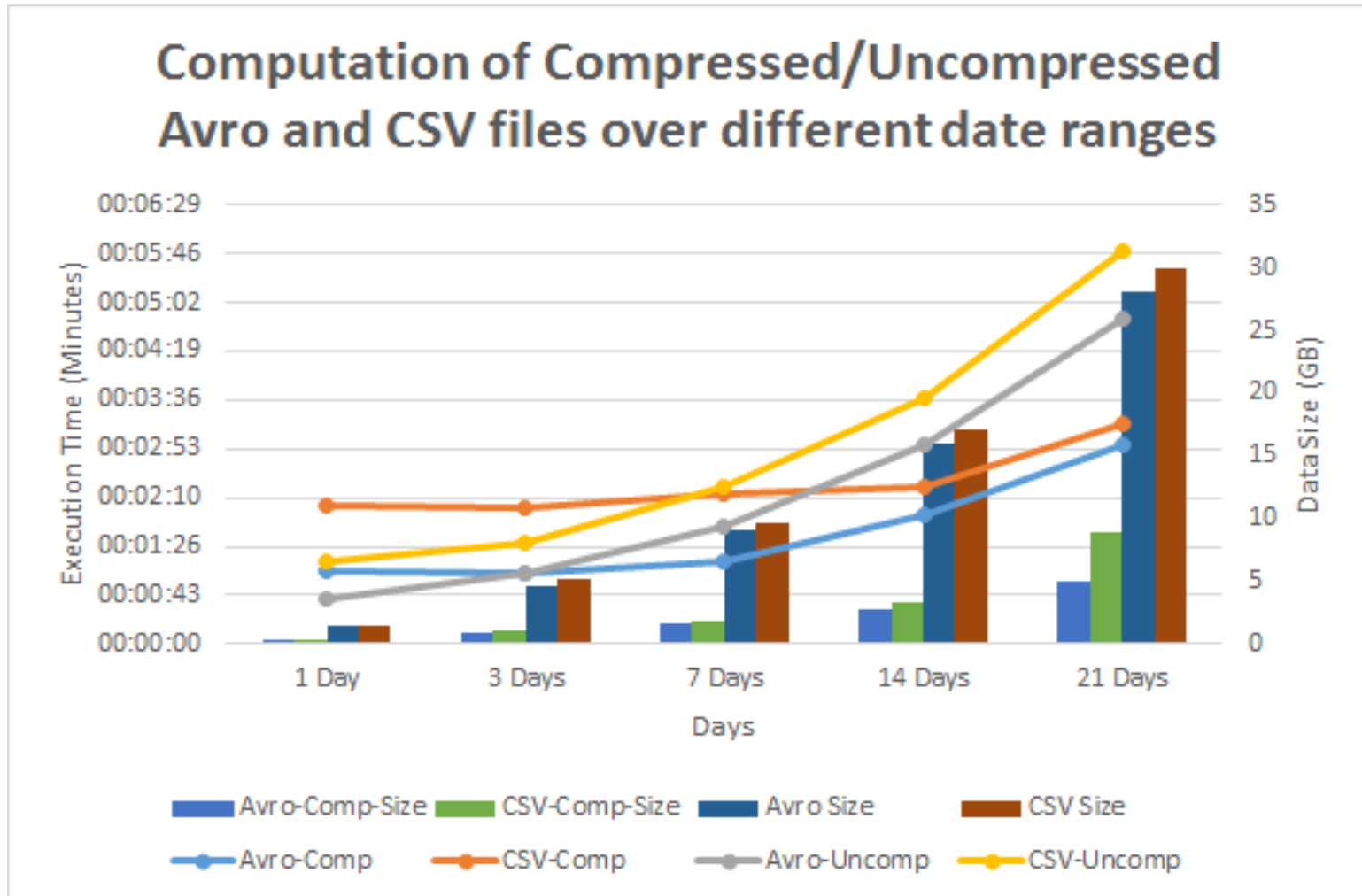
## Common Architecture



# Some numbers

- XRootD dashboard as first use case
  - ~ 150 GB/month (all VO)
- Data format optimization
  - CSV and AVRO evaluation
  - AVRO compressed (~ 30 GB/month)
- HDFS partitioning: daily file with date directory structure
- Periodic MR job access ~ 1 GB per run

# Some (preliminary) result (FAX/EOS)



# Summary

- Hadoop/MapReduce natural fit for computation of dashboard statistics
  - Simpler code
  - Faster
- We are starting migration to the IT-DSS Hadoop cluster for production set-up
- Other use case (e.g. ATLAS) can profit from the same infrastructure