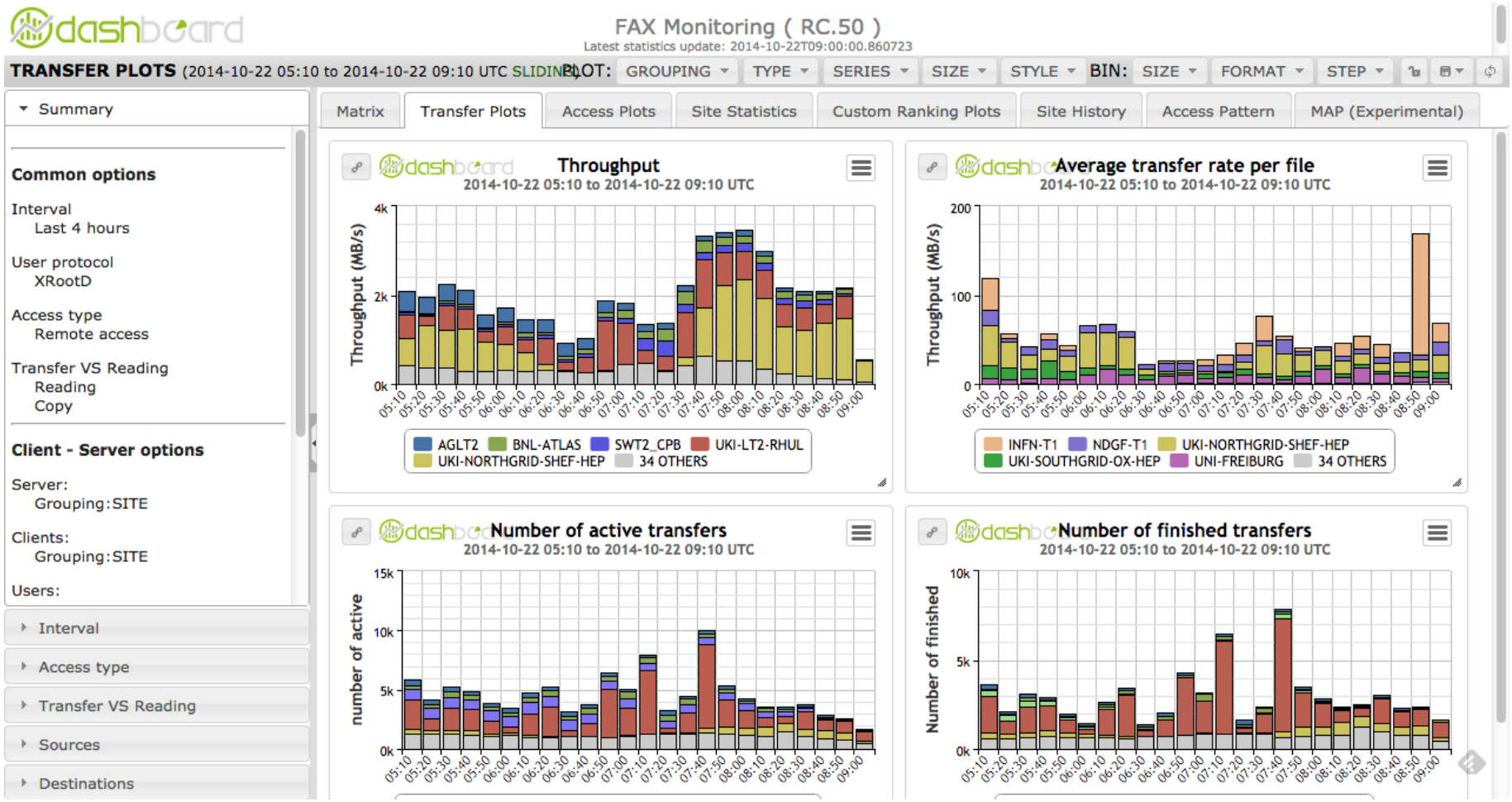
Experiment transfers dashboard on Hadoop

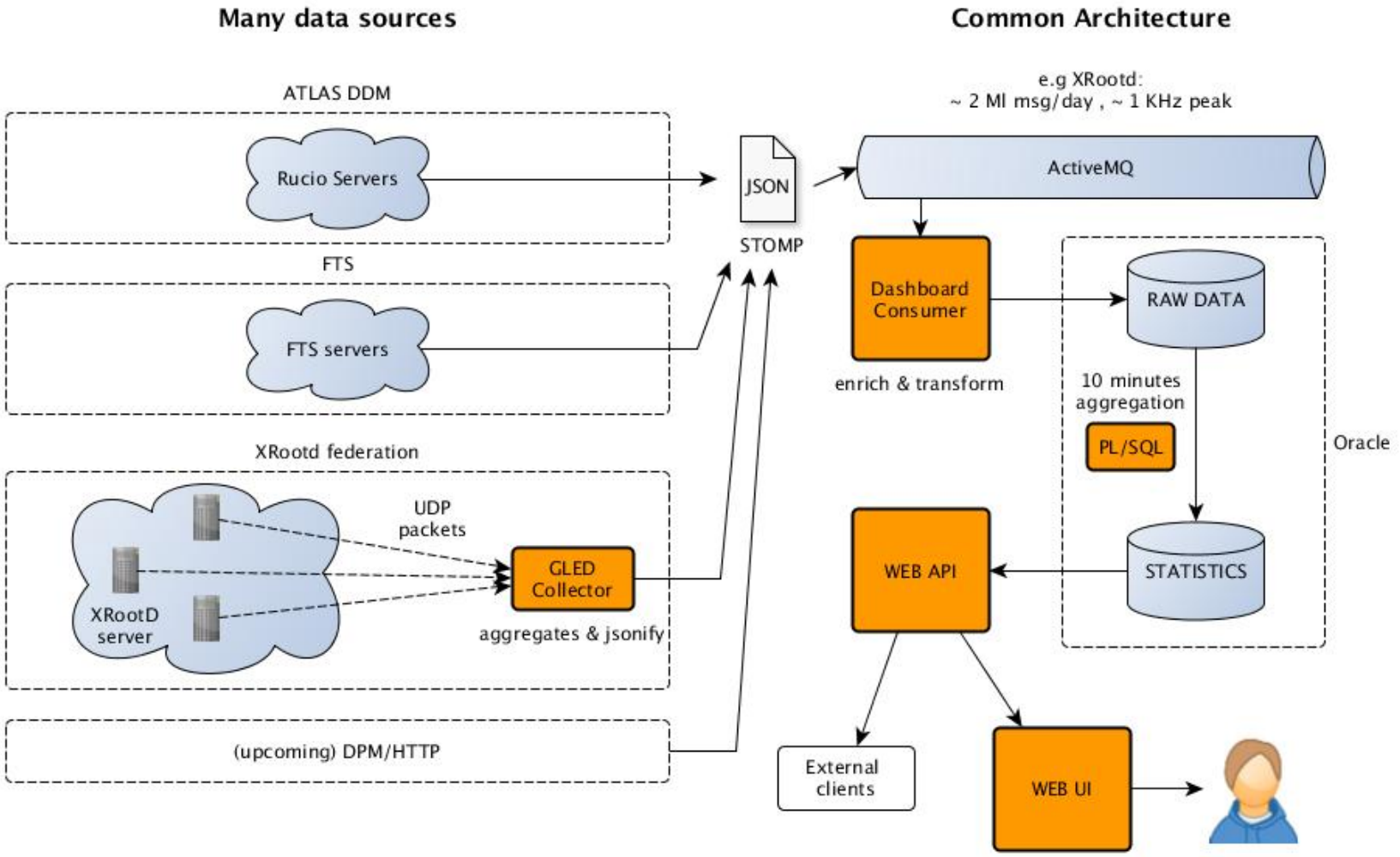
Luca Magnoni
IT-SDC

22 October 2014

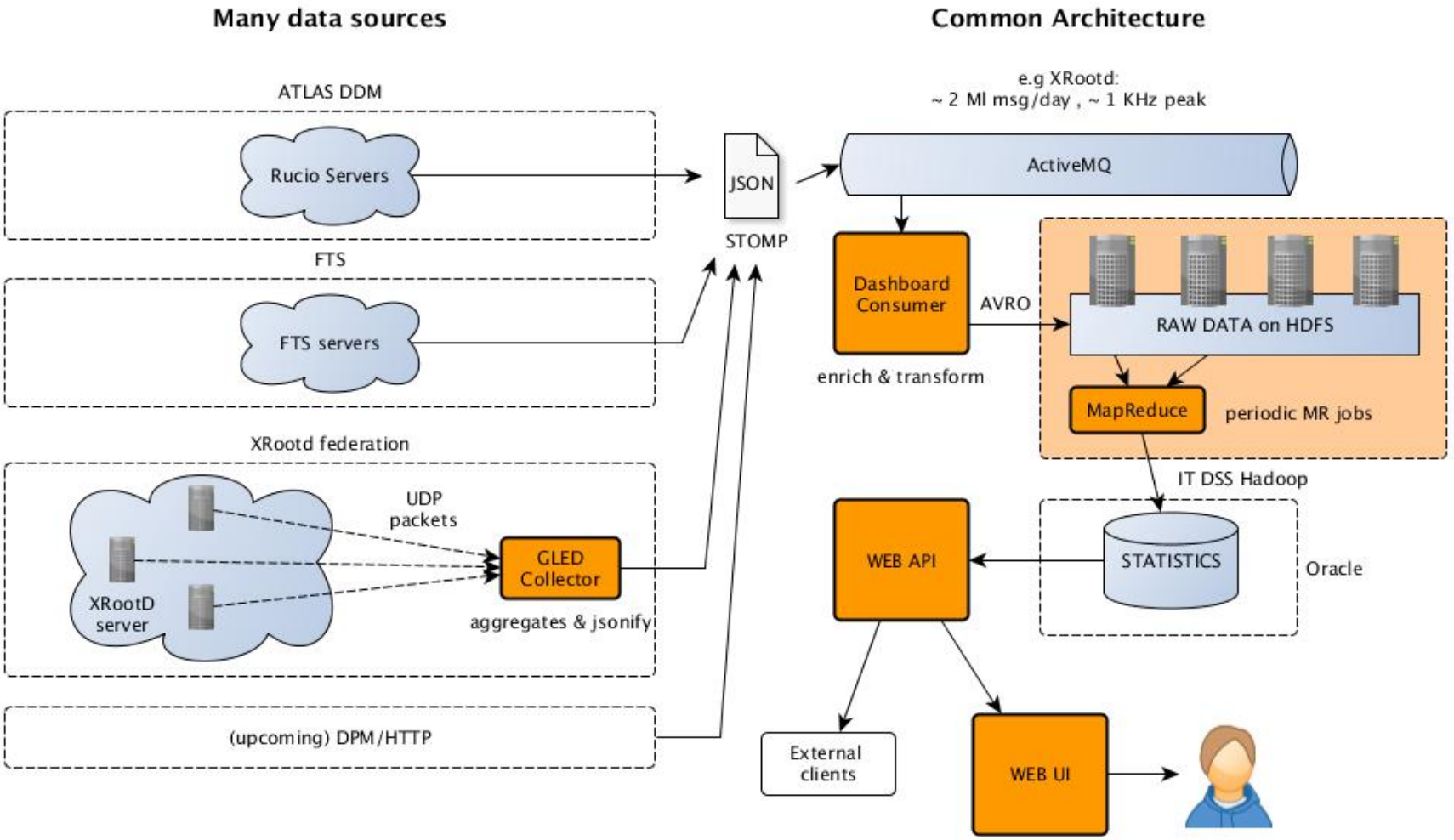
Experiment transfers dashboards



Today's architecture



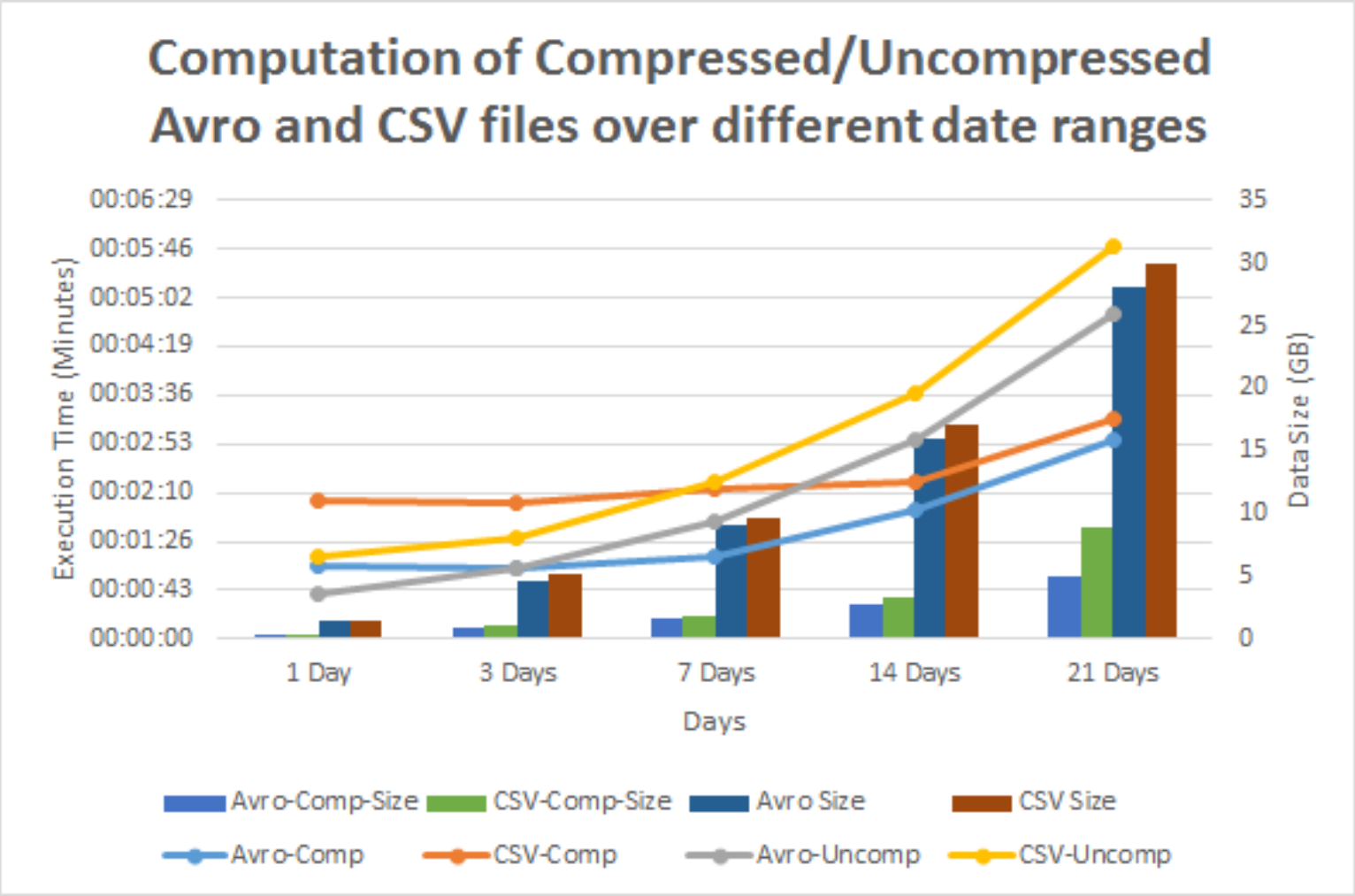
Architecture evolution on Hadoop



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