WLCG Workshop 2017

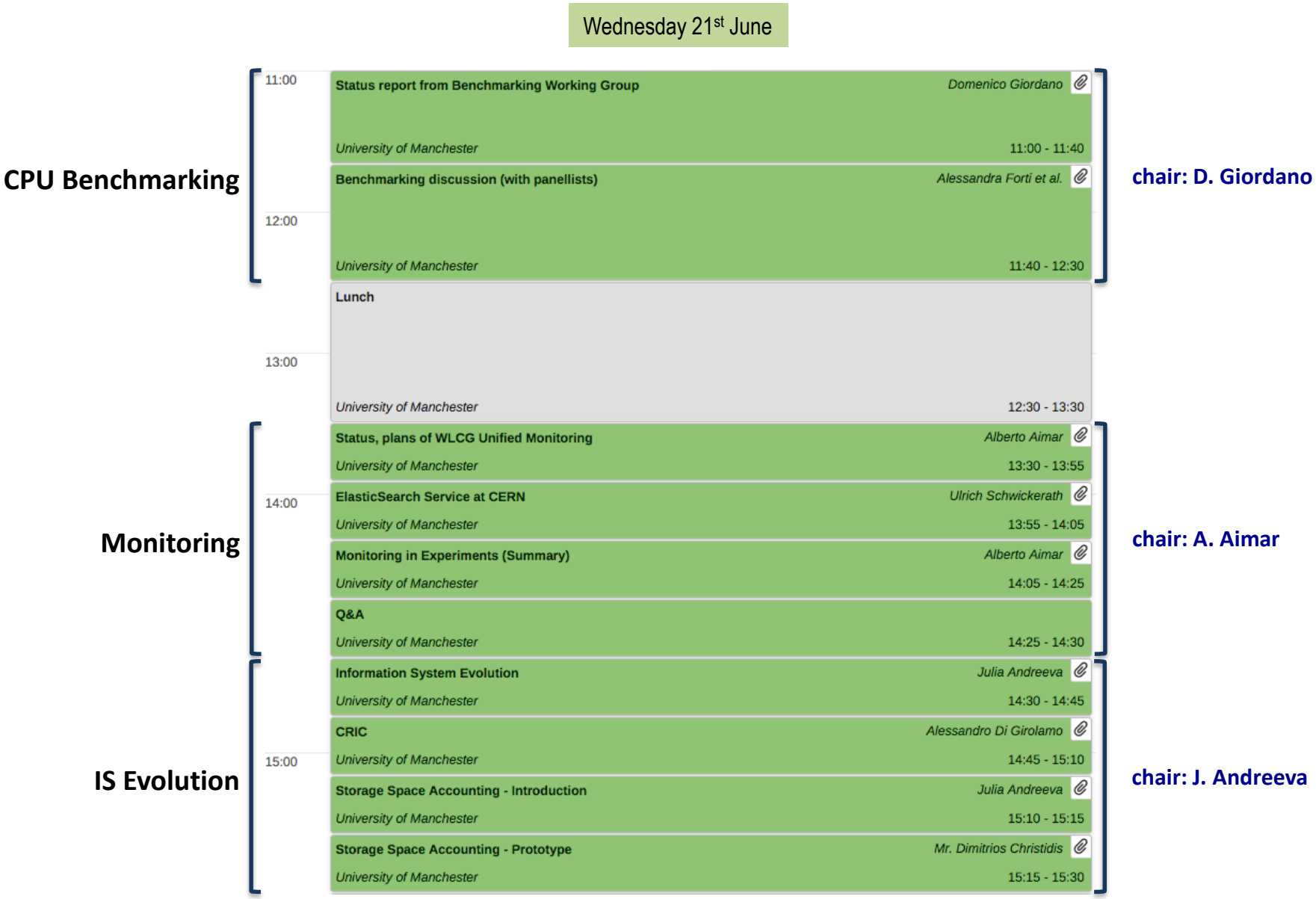
[Manchester]

Operations Session Summary

J. Flix

D. Giordano, A. Aimar, J. Andreeva

Contents of the Ops session



Benchmarking: Plan of the session

D. Giordano

- Summary of the **HEPiX Benchmarking WG activity**
  - → Mandate, recent activities, foreseen plans
- Discussion, animated by a panel with experiments' and sites' representatives
- Objective
  - → Discuss the discrepancy among HS06 and HEP workloads
    - Among the studied benchmarks, is there a valid substitute of HS06?
  - → Clarify the opinion of the Experiments about current fast benchmarks

HEPiX Benchmarking Working Group
hepix-cpu-benchmark@hepix.org
https://twiki.cern.ch/twiki/bin/view/HEPIX/CpuBenchmark

Benchmarking: DB12 and HS06

D. Giordano

- **DB12** doesn't show the stability and characteristics to probe all components of the CPU potentially used by HEP workloads
  - → Limited instruction mix; does not stress the memory subsystem
  - → Adoption for procurement would represent significant risk
  - → DB12 is still attractive for fast benchmark in jobs

- **HS06**
  - → Preliminary study still shows good agreement among HS06 and CMS MC ttbar, when server fully loaded
  - → Passive benchmarking
  - → Discrepancies among HS06 and ATLAS reco jobs are within 10%
  - → Need to better understand the reasons of the discrepancies for LHCb and ALICE

Benchmarking: next steps

D. Giordano

- **SPEC2017** is now available: will start testing it

- Work in progress to setup a *testbed* for the HS06 successor
  - → Support from the Experiments is mandatory here
  - → Proposal of evaluating a HEP benchmark suite, built with HEP workloads, received positive feedback
  - → The overhead of maintaining the benchmark software and migrating to platforms should be also considered

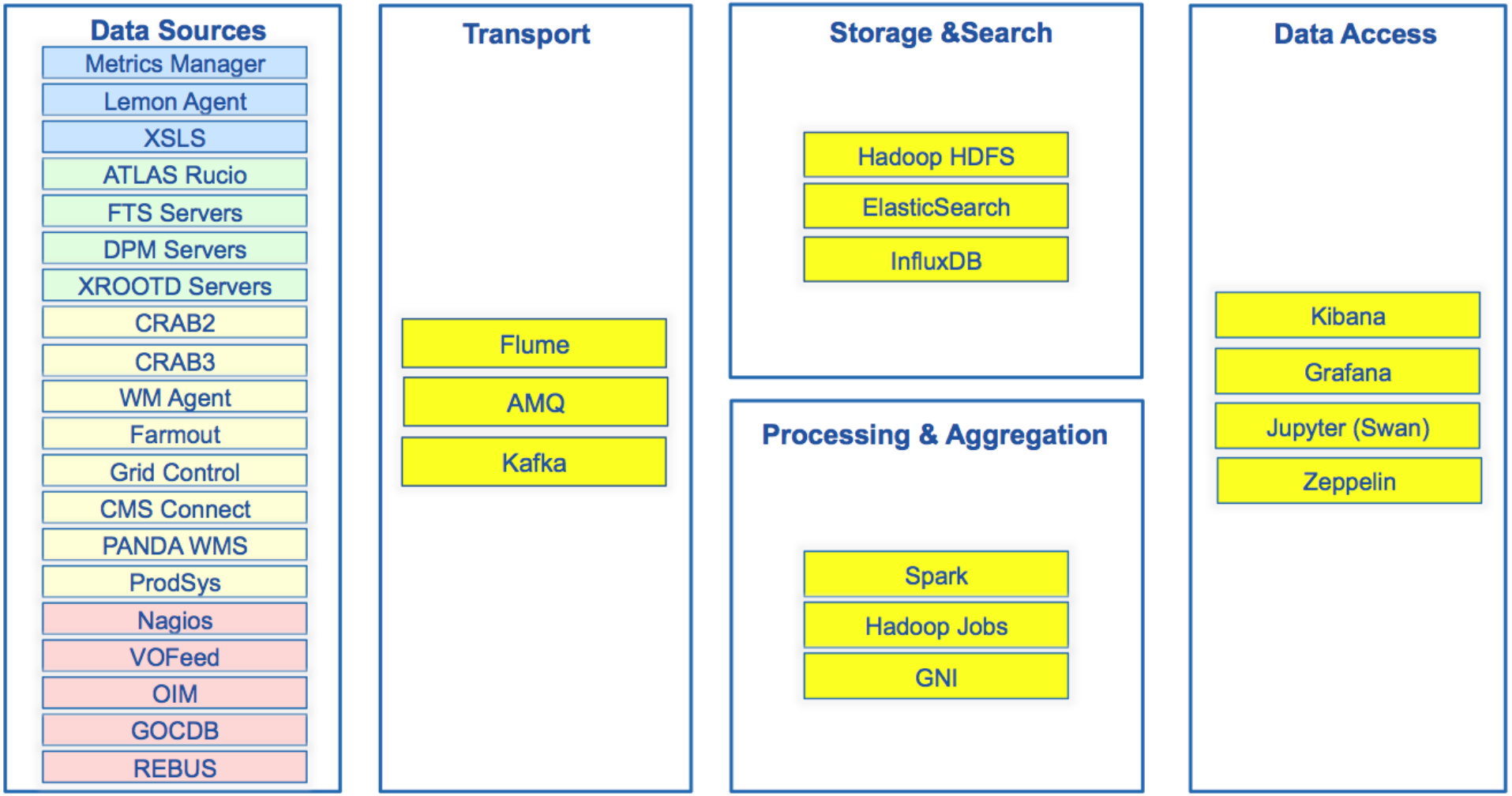
Survey & Discussion

- Alessandra Forti (Atlas experiment and site repres.)
- Andrew McNab (LHCb and site repres.)
- Manfred Alef (WG chair and site repres.)
- Pepe Flix (CMS experiment and site repres.)
- Latchezar Betev & Costin Grigoras (ALICE experiment repres.)

panellist discussion
http://cern.ch/go/mTL6

Monitoring: WLCG Unified Monitoring

A. Aimar



* Slides: http://cern.ch/go/6hMV

Monitoring: WLCG Unified Monitoring

A. Aimar

- Current Data and Services

WLCG Data in new MONIT	
	FTS
	XROOTD
	XROOTD ALICE
	DDM RUCIO
	DDM TRACES
	DDM ACCOUNTING
	PHEDEX
	ATLAS JM – PANDA/PRODSYS
	CMS JM - HT CONDOR
	SAM3 ETF
	AGIS
	VOFEED
	REBUS
	GOCDB
	OIM

Services Proposed

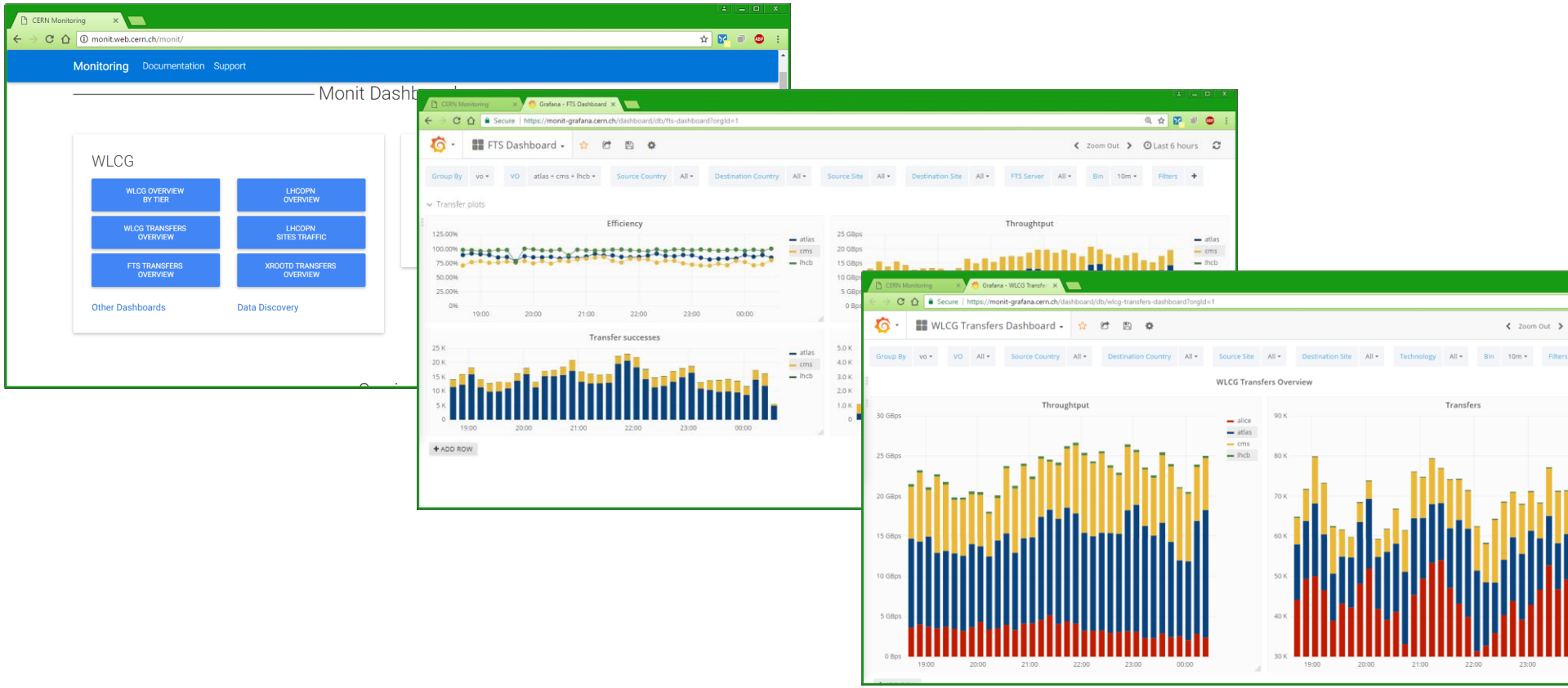
- Monitor, collect, visualize, process, aggregate, alarm
- Infrastructure operations and scale
- Helping and supporting
  - Interfacing new data sources
  - Developing custom processing, aggregations, alarms
  - Building dashboards and reports

Monitoring: WLCG Unified Monitoring

A. Aimar

- MONIT Portal (monit.web.cern.ch)

→ Single entry point for the MONIT dashboards and reports. Direct links to dashboards and reports. Sections for WLCG and Experiments



Monitoring: WLCG Unified Monitoring

A. Aimar

- Status:
  - → Data already in MONIT. Processing completed. Initial dashboards are there.
  - → Ready to provide training, on dashboards and reports
  - → Work with experts to check and develop together final new dashboards
  - → Share with final users the new solutions (shifters, sites, managers, experts, etc.)

- Now help from Experiments to complete migration of dashboards/reports
  - → Develop final use cases (shifters, experts, managers, site managers, etc.)
  - → Verify data quality and plots
  - → Extend and create new dashboards
  - → Spread the word, training, presentation in experiments, etc

Monitoring: new ElasticSearch @CERN

A. Aimar

- Setup a centralised Elasticsearch service and replace existing clusters

- Consolidation: centralised management, share resource, standard hardware, virtualisation

- Expectations: special requirements, privacy & security, performance, scalability

- Solution
  - → Share resources where possible, consolidate smaller use cases
  - → Use dedicated clusters where needed: special networking requirements (eg. Technical network (TN), high demand use cases (eg. CERN IT monitoring), dedicated clusters for ALICE, ATLAS, CMS, LHCb

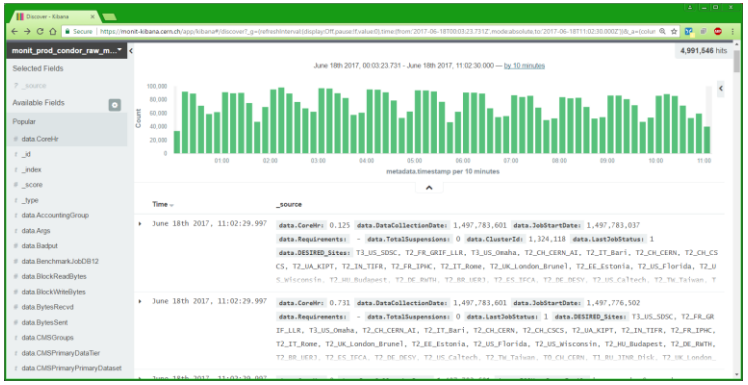
* Slides: http://cern.ch/go/N6b9

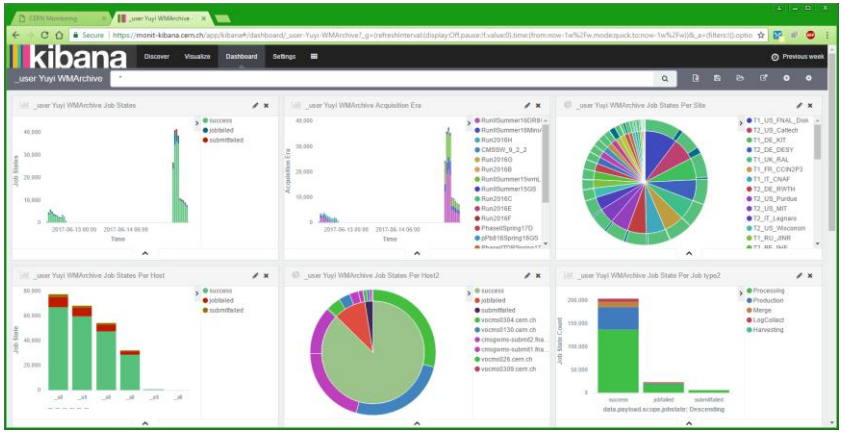
Monitoring: new ElasticSearch @CERN

A. Aimar

- ~20 Clusters up and running
  - → O(40) use cases (entry-points). Migrating to ES 5.X

- Access control (ACL) implementation
  - → Privacy and security requirements. Needed for efficient consolidation of resources
  - → Commercial plugins: Tested XPACK and SearchGuard. Concerns about costs and performance
  - → Implemented model (ES 5.x only) – Pure OpenSource solution (Apache 2, GPL V3) – Index-level security





* See **BACKUP** slides for a summary of the current status of the monitoring for each of the LHC experiments

Monitoring: Summary

A. Aimar

- All looking at external technologies instead of fully in-house solutions
  - → ElasticSearch, Kibana, Kafka, Messaging, Spark, HDFS
  - → less custom-made less control, but huge communities and free improvements

- ATLAS and CMS
  - → have their infrastructure for deeper analytics studies
  - → rely on MONIT for monitoring, and as migration from WLCG dashboards
  - → use MONIT curated data with aggregation, enrichment, processing, buffering, dashboards, etc.

- ALICE and LHCb
  - → run their own infrastructure
  - → based on central IT services (ES, HDFS)
  - → could share (some) data in MONIT, if needed

IS Evolution: CRIC

J. Andreeva

- Currently WLCG IS consists of multiple information sources, some of them (BDII) are fully distributed

- Difficult to integrate and validate data and to debug problems

- No clear and common way for description of non-traditional/non-grid resources (commercial clouds, HPC, etc…)

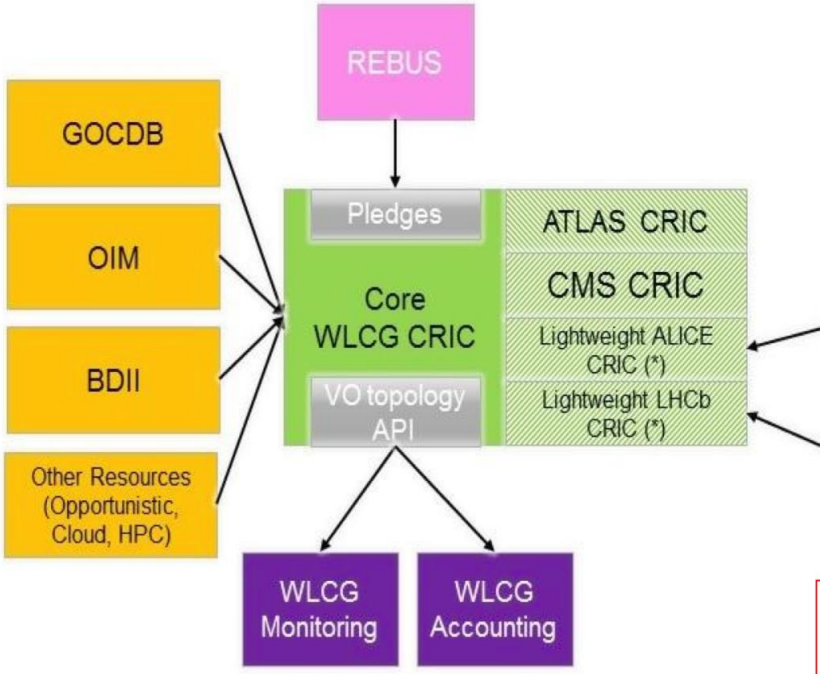
- A (new) single entry point for the topology and configuration information under development
  - → **Computing Resource Information Catalog (CRIC)** aims to "glue together" physical resources of the WLCG infrastructure and experiment frameworks
  - → CRIC is the evolution of Atlas Grid Information System (AGIS)

* Slides: http://cern.ch/go/NQ9R

IS Evolution: CRIC

J. Andreeva

Plugin based: Core and Experiments



(*) Maintained by WLCG to store very simple experiment topology information (i.e. experi

Core CRIC

- *Resources "Provided by"*
- Single entry point for WLCG topology and service configuration
- Consumes information from available information sources

Experiments CRICs

- *Resources "Used by"*
- Describes experiment topology
- Uses core CRIC and adds extra info needed by experiment operations and workflows

IS Evolution: Storage space accounting

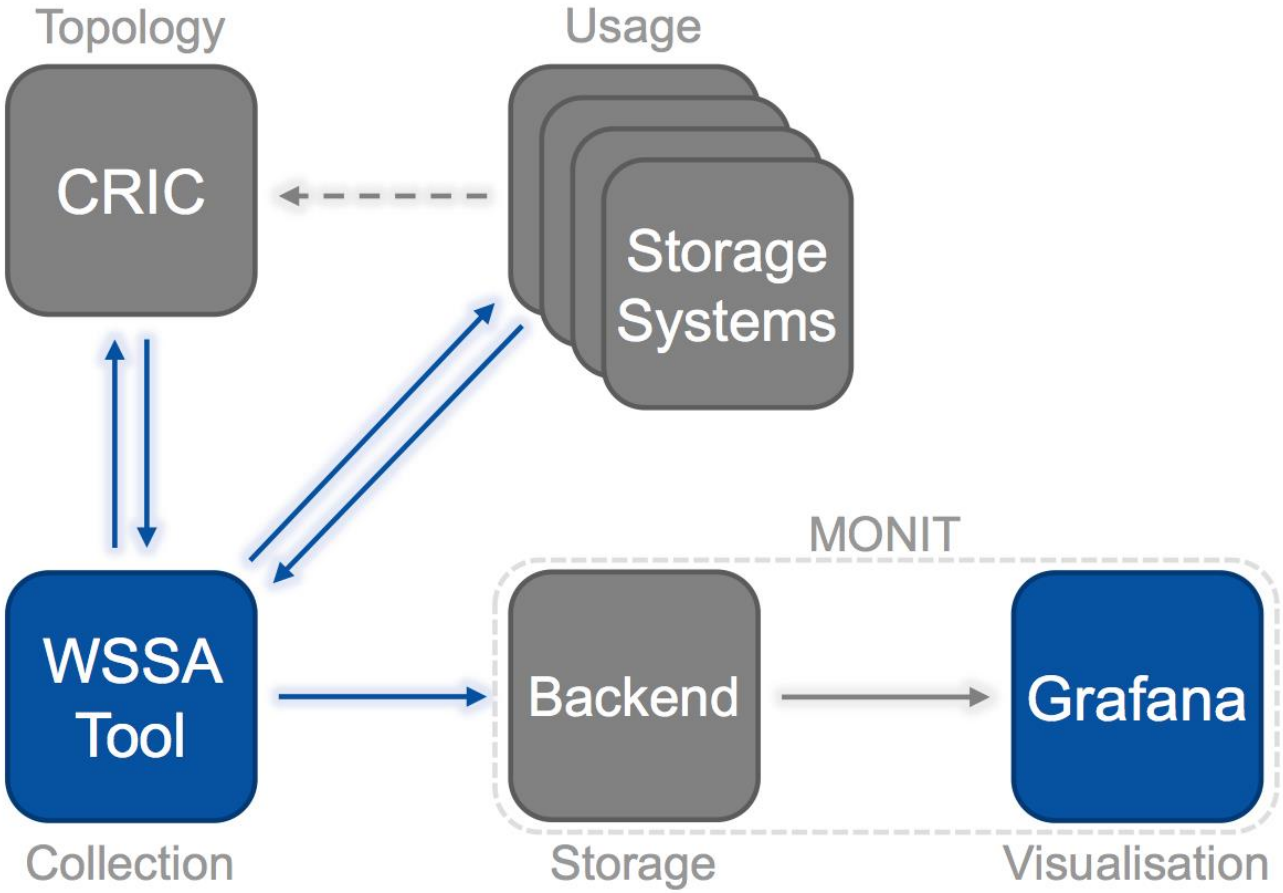
J. Andreeva

- **Absence** of functional storage space accounting in WLCG
  - → In-house experiment solutions exist
    - → mostly based on SRM which is not a mandatory service any more
- WLCG Storage space accounting to provide high level overview of the free and used space with the granularity of VO shares @sites
  - → with possible split of those shares by areas for some dedicated usage
  - → It should work in the SRM-less world.
  - →The queries for used/free spaces are required both for accounting and computing operations - Should be enabled for at least one non-SRM protocol setup
- Storage Resource Reporting proposal:
  - → integrating CRIC (topology) and MONIT (data flow)
  - → Start prototyping from the existing experiment setups

* Slides: http://cern.ch/go/F6fJ / http://cern.ch/go/HF9v

IS Evolution: Storage space accounting

J. Andreeva



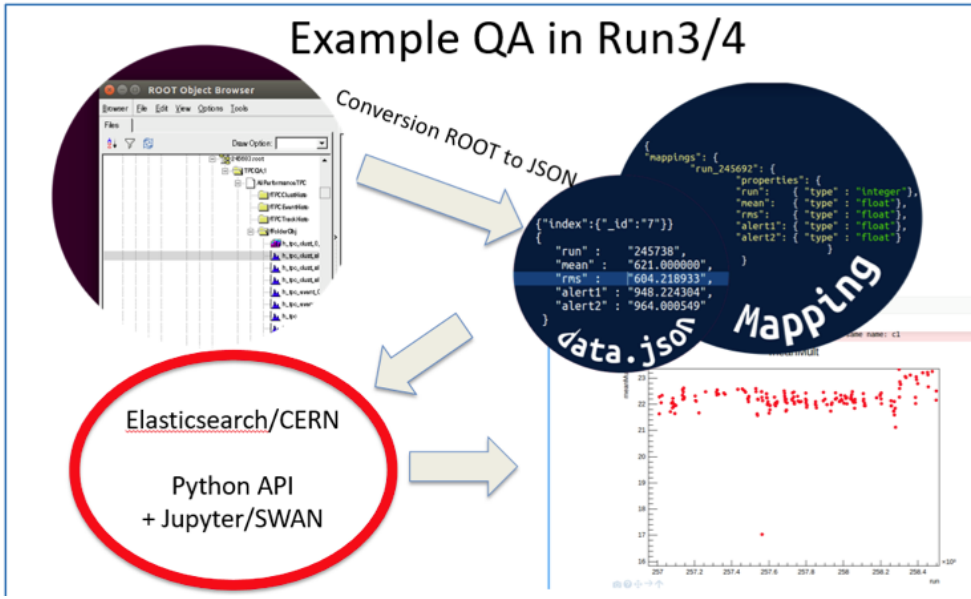
BACKUP SLIDES

Monitoring for the LHC Experiments

ALICE Run 3/4

Online/Offline (O2) in Run3/4

- Parallel synchronous and asynchronous data processing (ALFA framework with FairMQ)
- QA repository/database
  - NoSQL database (**Elasticsearch, Cassandra**…or others) for metadata information
  - **EOS** file system for storing QA objects (ROOT histograms and trees)
- Data aggregation and enrichment
  - Custom solution based on **ROOT**
  - Apache **Kafka** is considered
- Interactive QA analysis
  - **SWAN/Jupyter** Notebook
  - Custom database queries using Python
  - Custom visualization tools based on JSROOT



WLCG
Worldwide LHC Computing Grid

Monitoring for the LHC Experiments

ATLAS

ATLAS Monitoring Baseline technologies

- ElasticSearch: Dedicated instance by CERN IT (IT ES). Instance by Univ. of Chicago. As part of MONIT, curated data
- Notebooks: Zeppelin instance, part of MONIT, and dedicated to ADC
- Hadoop: HDFS raw and aggregated data. Preparation of data for ElasticSearch ingestion. Machine learning

ATLAS and MONIT

- ATLAS ADC Analytics and Monitoring Launched Working Groups to check and improve, implement the MONIT dashboards and reports
- DDM and DDM Accounting: Initial dashboards from MONIT on going. Need checking and improvements by experts
- Job Monitoring, Sites monitoring: data in MONIT, initial dashboards prepared

ElasticSearch @ Uchicago

- Hardware-based: 8 nodes total (each 8 core, 64GB RAM), 5 data nodes (3x1 TB SSD each)
- Contents: 15'000'000'000 docs. Clients: Kibana,Notebooks, Embedded visualisations, Cron jobs

ATLAS and Hadoop

- use the analytics cluster at CERN. 40 nodes, 2TB RAM total, 2PB storage total

ElasticSearch and Hadoop are both critical and used in production

- Most of the tools/capabilities are available. Interplay between tools not ideal and efficient, lots of custom-made duct-tape

Monitoring for the LHC Experiments

CMS

In process of migrating to use CERN's standard MONIT framework

CMS monitoring covers many kinds of resources:

- CMSWeb – front-end tool for displaying monitoring results
- Data Popularity – dataset usage

- SpaceMon – monitoring disk usage across CMS grid
- Site Status Board – shows status of CMS grid sites

- CRAB/ASO operations – monitors files transferred for user grid jobs
- HTCondor job monitoring

- WMAgent – production job submission and reporting
- WMArchive – grid job reports

WLCG
Worldwide LHC Computing Grid

Monitoring for the LHC Experiments

LHCb

DIRAC Monitoring

- Based on DIRAC framework and designed for:
  - real time monitoring (WMS jobs, DIRAC components, etc.)
  - managing semi-structured data (JSON)
  - efficient data storage, data analysis and retrieval
  - provide high quality reports
- Technologies used:
  - **Elasticsearch** distributed search and analytic engine
  - DIRAC Graph library for creating high quality plots
  - DIRAC web framework (front end) for the plots
  - Messaging Queue systems: **AMQ** (via stomp) and **RabbitMQ**
  - **Kibana** flexible analytic and visualization framework

Pilot NagiosProbes for SAM/ETF

- Output of the probes is retrieved by a script to SAM/ETF
- Once in production, VM-based resources will immediately appearing in WLCG SAM dashboards / reports