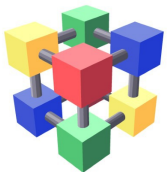# WLCG workshop 2015: run2 and beyond

Alessandra Forti

On behalf of WLCG Ops Coordination

GDB

12 May 2015

Layout

- In one slide
- Increased requirements
- Blurring site boundaries
- Opportunistic resources
- Multicore
- Reconstruction problem
- Federated storage
- Storage middleware
- CEPH/SSD/caches
- Service maintainance

- Support
- A "simple Tier2"
- A "simpler Tier2"
- Monitoring
- The operational costs of protecting pages
- More of a community
- Manpower worries
- Conlcusions
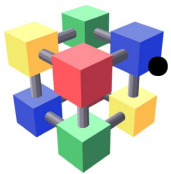
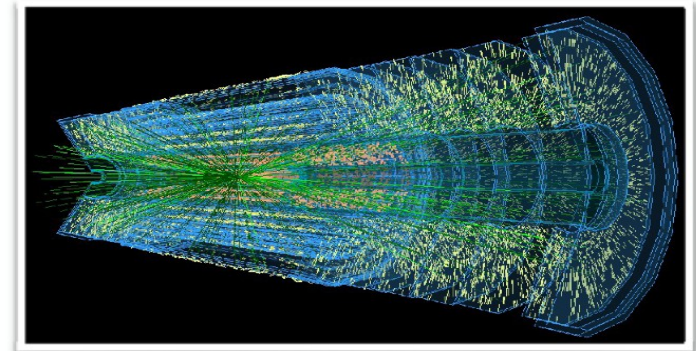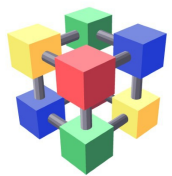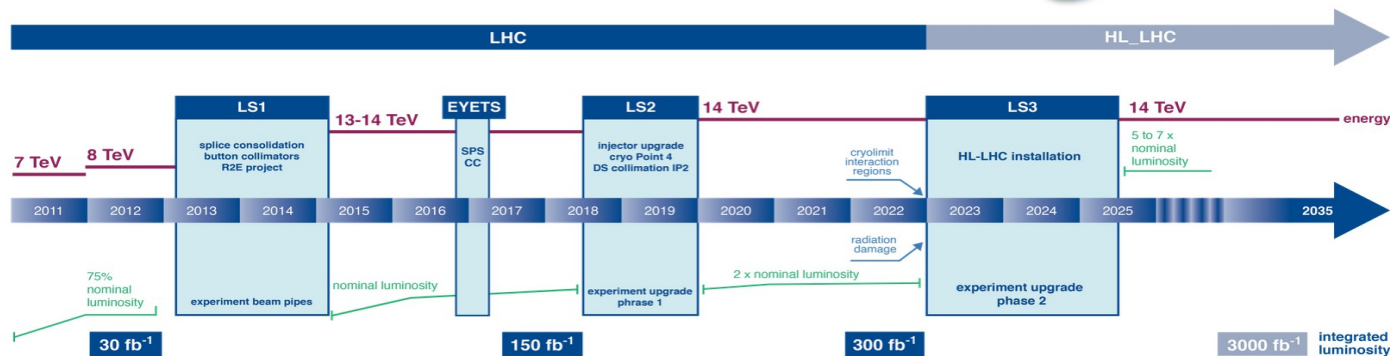# Workshop in one slide

- Readiness for run2

- HL-LHC: beyond run3

  - ATLAS+CMS >2022

  - LHCb ~2018

- Technology and HW evolution

- Budget constraints

run2

HL-LHC

LHC / HL-LHC Plan

LHC/ HL-LHC Plan (last update 03.03.2015)
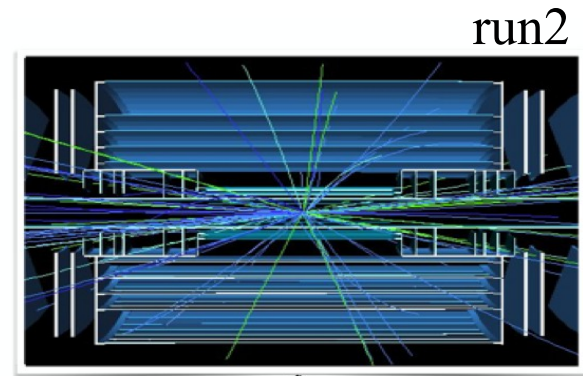
Increased requirements

- Increased data x10

  - ATLAS+CMS: 300 fb-1 → 3000/ fb-1

  - LHCb: 5 fb-1 → 50fb-1

  - Alice: x100 more data

  - Optimization of the code
    - Matching applications with HW

  - Adopting new coding paradigms

  - Doing more reprocessing
    - Keeping less data

  - Doing full reco in HLT
    - LHCb and Alice triggerless

CMS

Detector	HLT output rate (kHz)	Total
Phase 1	1	3
Phase-II (140)	5	65
Phase-II (200)	7.5	200

Scale of computing resource needs relative to Run 2 including the increase in projected HLT output rate

ATLAS

|	HLT Output	Events per year	RAW per Event	RAW data per year
Run1	600Hz	3.6B	0.7MB	2.5PB
Run2	1kHz	5B	1.0MB	6PB
Run3	1kHz	5B	1.2MB	7.2PB
Run4	5kHz	25B	2.5MB	75PB

Blurring site boundaries

- MONARC model is dead
  - Rigid hierarchy of sites
  - Fixed data distribution model
- All experiments are using sites in a flattened or service oriented topology
  - Most grid sites can run most applications
  - **Tier2 reliability** and network improvements
- Increased flexibility of experiment frameworks
  - Storage less sites for example
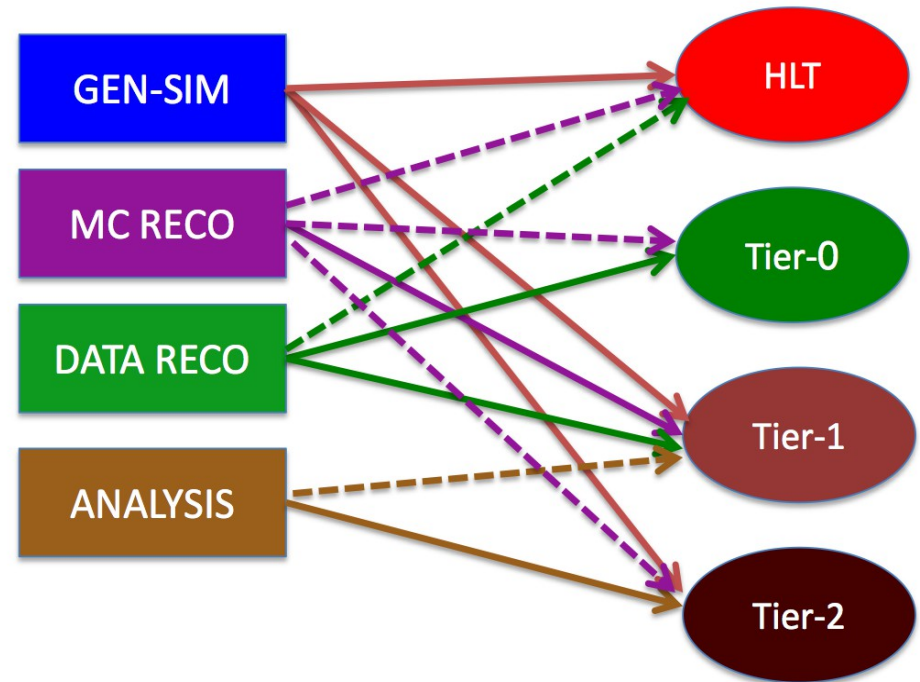
- Cannot really blur the lines
  - T1s have a different SLA and tape
  - Funding agencies wouldn't understand a chaotic model

Opportunistic resources

- Effort to exploit whatever resources they can with varying degrees of commitment
  - HLT farms
  - HPC
  - Commercial clouds
    - In particular USCMS and USATLAS
  - Volunteer computing
    - BOINC
- Mostly doing event generation
  - ATLAS worked on processing 1 event at the time can do also reconstruction
    - Solution that could be adopted by other experiments

Multicore

- Multicore spear headed by ATLAS and CMS
  - LHCb started to look into it for users
- Originally to reduce the memory footprint for run2
  - Enough for run2 but not for HL-LHC
- Now seen as first step towards more advanced parallelisation of the code to exploit many cores architectures
  - CMS and LHCb have both R&D effort
  - ATLAS not much effort so far
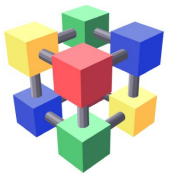
Reconstruction problem

- How to
  - Reconstruct HL-LHC events with 200 pileup (events)
  - Maintaining the physics performance
  - With no prospect to increase resources
- CMS and ATLAS share this problem
  - Suggestion to work on common code and algorithms to avoid duplication
- Working on common solutions at all level becoming a kind of mantra
  - Funding agencies not looking favourably at duplication of effort for the sake of it
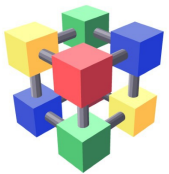
Federated storage

- Effort dedicated by the experiments to data remote access, or Federated storage

  - ATLAS+CMS+Alice using xrootd

  - LHCb : Gaudi for input files, Http for consistency checksxs

- Not everyone in the experiments expresses total enthusiasm for all data being accessed everywhere but

  - Failover mechanism

  - Diskless sites

- Robustness still to be demonstrated

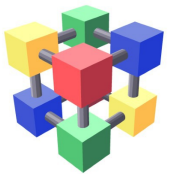 - Even if monitoring shows a steady increase in xrootd traffic

Storage middleware

- All agree there is a zoo of protocols

- All experiments started to look into abandoning SRM

  - GFAL2 and FTS3 help handling the different protocols

  - Still several protocols and different experiments different phylosophies

    - LHCb working towards WEBdav

    - ATLAS using still gridftp, xrootd, httpd depending on the action

    - ALICE and CMS working towards xrootd only solution

- Sites expressed a preference for abandoning HEP produced software and adopt industry standard

  - i.e. httpd

- WLCG storage working group should continue the work started few months ago
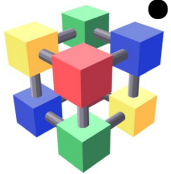
CEPH/SSD/caches

- Proposals to use sites more and more as caches

  - Equipped with smaller fast storage

- Or in any case have caches in front of slower bigger storage to better support diskless sites

  - A big number of sites are looking into CEPH as a storage technology. Among many pros, it enables SSD caching out of the box. This could be a game changer for more efficient WAN transfers. So supporting CEPH as an SE technology could be very beneficial."

- SSD becoming more affordable

Service maintenance
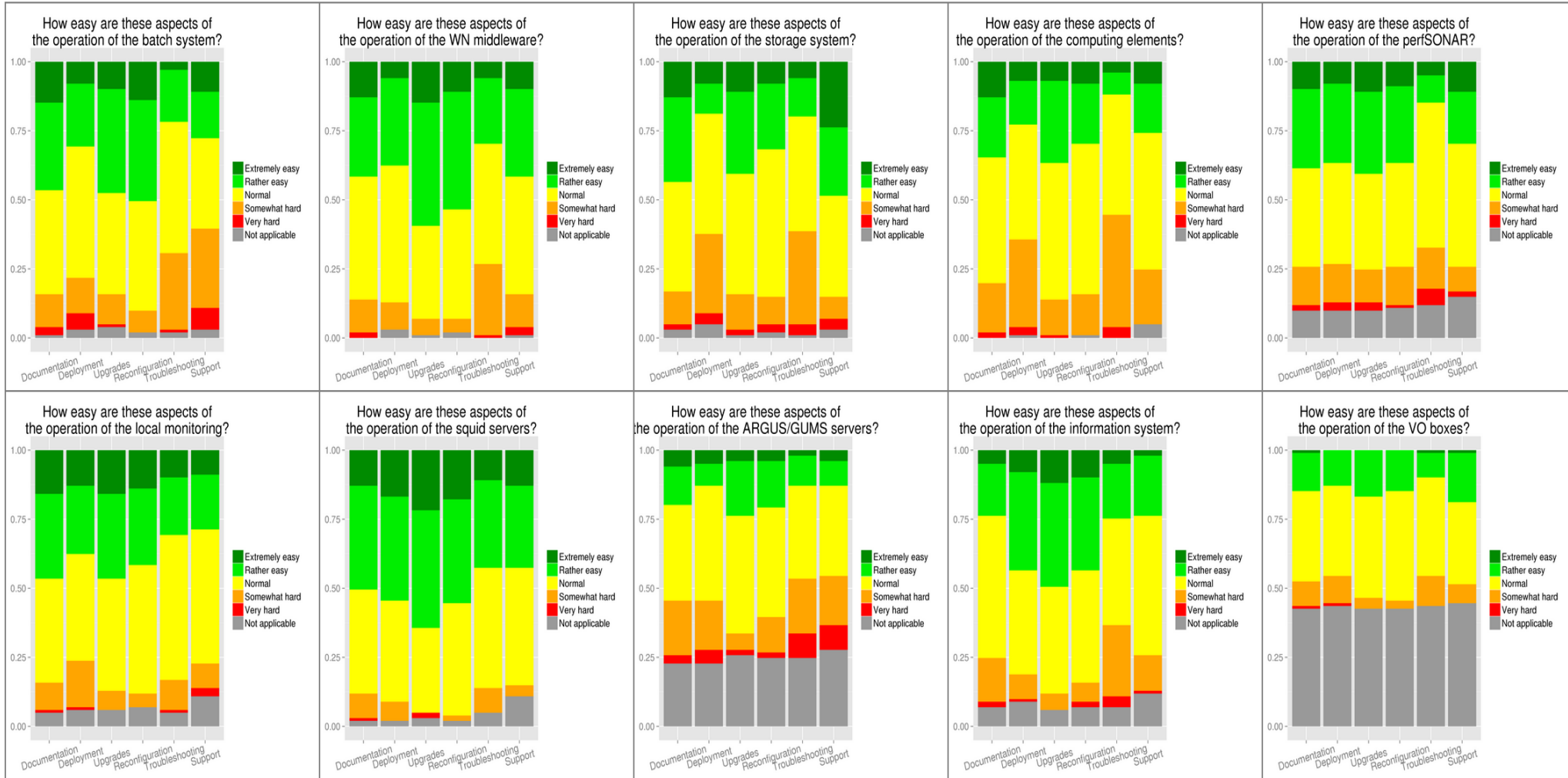
Support

- Peaks of negativity for deployment and troubleshooting services in particular SEs and CEs

  - Poor error reporting, poor documentation, lack of log parsing tools.

- YAIM/EGI/EMI future is unclear.

  - YAIM created problems at few more sophisticated sites but many others are still using it underneath puppet when available.

- The number of services not really required for operations should be reduced.

  - WLCG services should be simpler and fewer

  - Particular grudges with glexec and perfsonar

- The lack of proper support for ARGUS and other tools (Torque/Maui) is a concern

- Time spent managing hardware, producing user documentation should also be considered a WLCG activity.
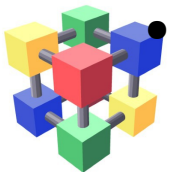
A "simple Tier2"

- There is this new concept of a "simple Tier2" but unless we revolutionise our sites there is very little that can be removed. Some suggestions

  - APEL box has been mentioned by several sites as a burden but it is not strictly necessary

    - ARC-CEs (and OSG) publish directly into APEL not clear why all the other CREAM sites couldn't do the same

  - Alternative CEs

    - ARC-CE is considered simpler and more robust.

  - Push new sites towards Htcondor as community is building up

    - Alternatively SLURM or UGE or latest LSF (latest ver support cgroups but cost)

  - Re-evaluate the need of an heavy weight BDII, mostly used for service discovery and getting few unreliable numbers in Rebus.

    - Are all the lines it publishes really needed?

    - Service in itself is light weight **if something fills the values for you. And YAIM is fading.**

 - Keep up the work to reduce the number of storage protocols

A "simpler Tier2"

- Experiments would prefer to handle smaller number of bigger well integrated sites
  - Relation with experiments also manpower demanding
  - Requires still expertise
- Smaller sites with little manpower would be better off with the cloud model
  - Batch system-less and disk-less (or at any rate small cache storage)
  - Attempt at preserving the hardware and some leverage while consolidating the manpower
    - VAC model (proposed by the UK)
- Funding agencies pushing towards cloudification.....

Monitoring

- Still conflicting messages between the necessity to monitor the whole experiment chain and having the capability to isolate the "site functionality" problems properly for availability calculations.

  - Sites and experiments upset by descrapancies both ways false positives and false negatives

    - Should publicize SAM integration in local nagios more

    - Local monitoring essential to catch problems before jobs do.

      - Wild variety of sites setup

- New site models will require a review of what we consider available
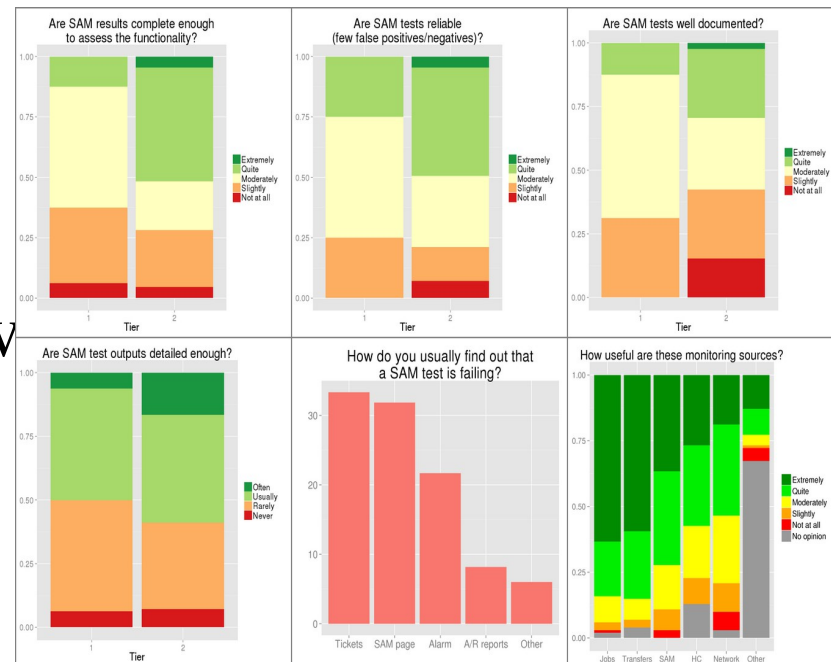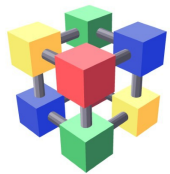
The operational cost of protecting pages

- One of the recurrent requests talking about troubleshooting, documentation and monitoring is that quite a lot of information is nowadays protected and not searchable.

  - The experiments twikies and JIRA tickets are protected too and let's be honest the twiki/JIRA search functionality ain't no google.... and "I found this savannah ticket that says..." ain't no more.

  - On the monitoring side it's practically impossible to access any meaningful monitoring unless one belongs to the experiment in question

    - ATLAS – the only experiment keeping the monitoring open is now also closing one of the most important monitoring services.

- We should be aware that all this has a dramatic operational cost

  - If it was bad before with distributed documentation now it is worst because google has lost its effectiveness and site administrators are forced to open CERN accounts to access information.
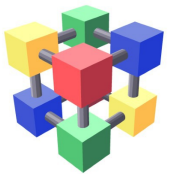
More of a community

- No central body doing the packaging and testing anymore

    - Fragmentation of software provisioning

    - Interaction with Pts (Product Teams)

    - Differentiation of configuration tools

    - Lack of central testing

- WLCG Operations Coordinations, TF and WG replacing this but with no paid manpower

    - Need stronger participation from Tier2 sites

    - More agile documentation

Manpower worries

- Budget constraints doesn't hit only the hardware
  - hit mostly the manpower
- Next few years will be of big changes in the infrastructure: technology, hardware, software, middleware
- Worries about the loss of expertise repeated from several people
  - Can live with lower level of hardware cannot live without the person who runs the infrastructure and supports the experiments
    - A lot of work done by system administrators often in cooperation with experiments in adapting to new technologies

Conclusions

- Medium term
    - Facilitate the information exchange in the community
    - Improve quality, availability and access of documentation
    - Reduce and simplify the services deployed at sites
    - Simplify the process of bringing new resources into production
    - Further improve monitoring and adapt it to new scenarios

- HL-LHC scale
    - Match applications and processing workflows to expected resource availability
    - Work on common solutions at all levels
    - Extend exploitation of opportunistic resources
    - Further encourage adoption of industry standards

Some keywords

- Increase in resource requirements not balanced by increase in funding!

  - Run2 trend increased requirements/less funding will continue

- Evolution:

  - Technology, HW → experiments software, frameworks

- Flexibility: differentiate resource usage

- Sharing expertise

  - Between experiments

- Sharing expertise&resources

  - With other sciences

- Build a community

- Budget constraints&manpower