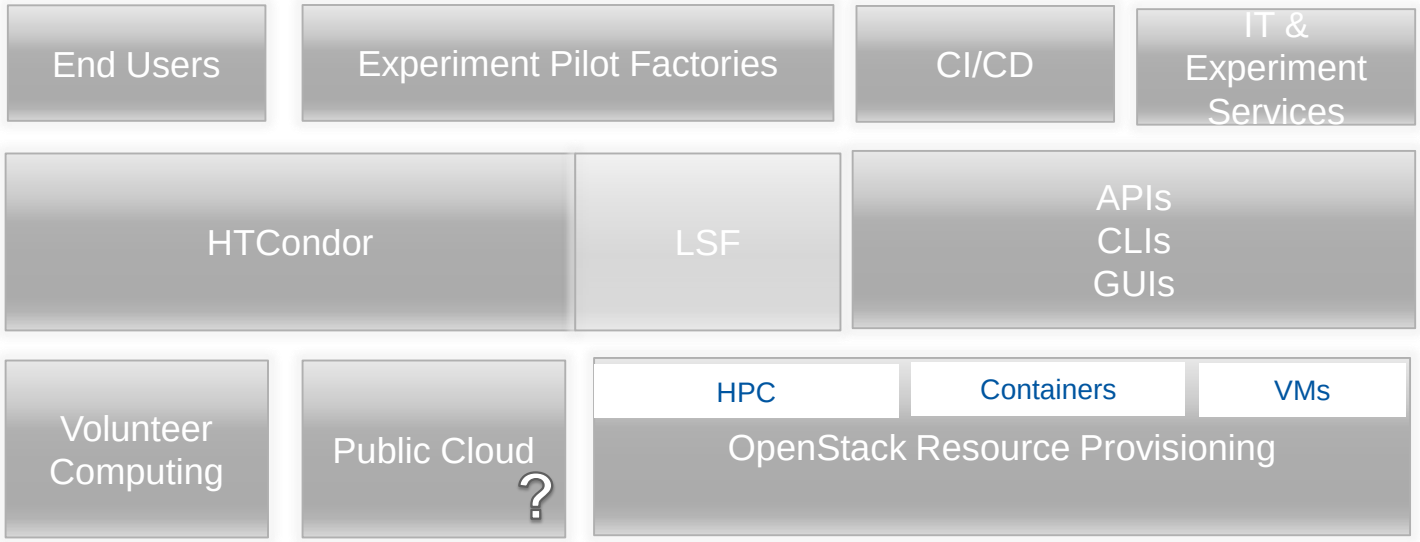
Compute services at CERN

Gavin McCance

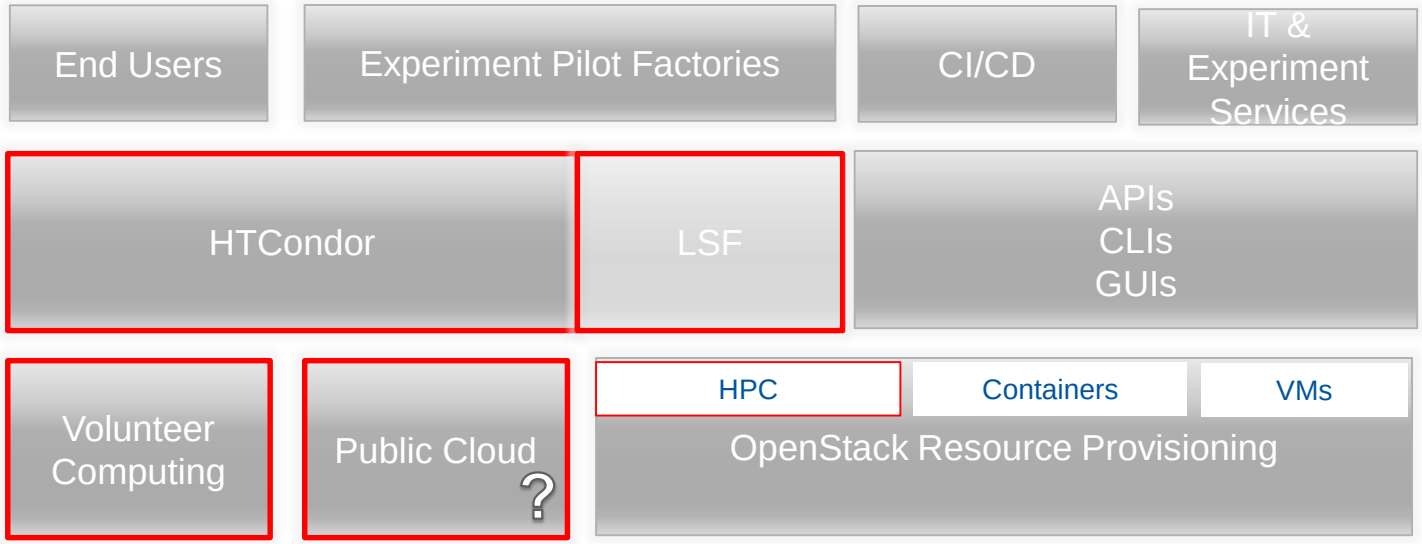
Tier-0 compute services 2017

- Universal resource provisioning layer for bare metal, containers and VMs
- HTCondor as the single end user interface
- Continue investing in automation and other communities for scaling with fixed staff
- Self service for end users within the policies and allocations

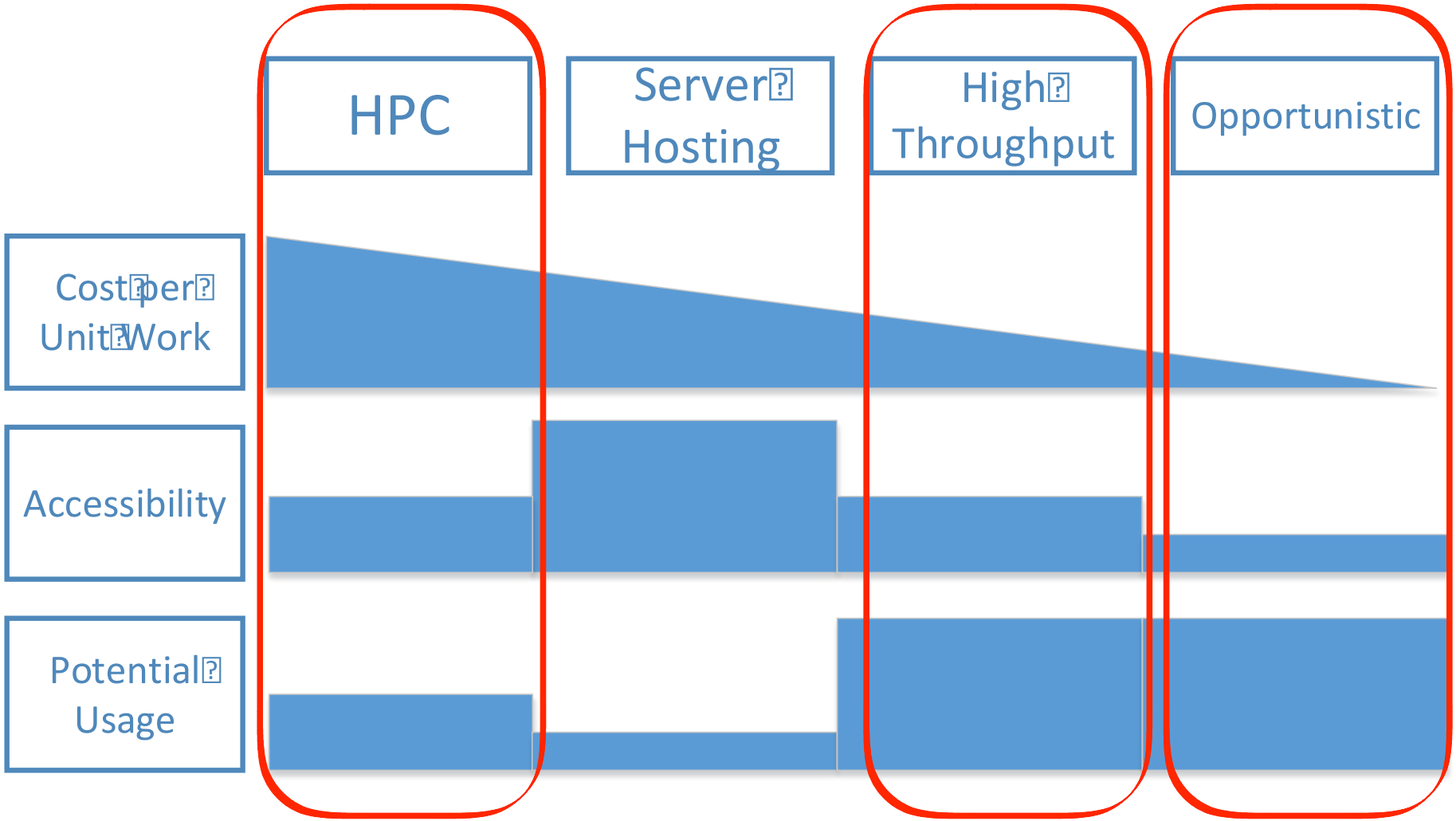


Tier-0 compute services 2017

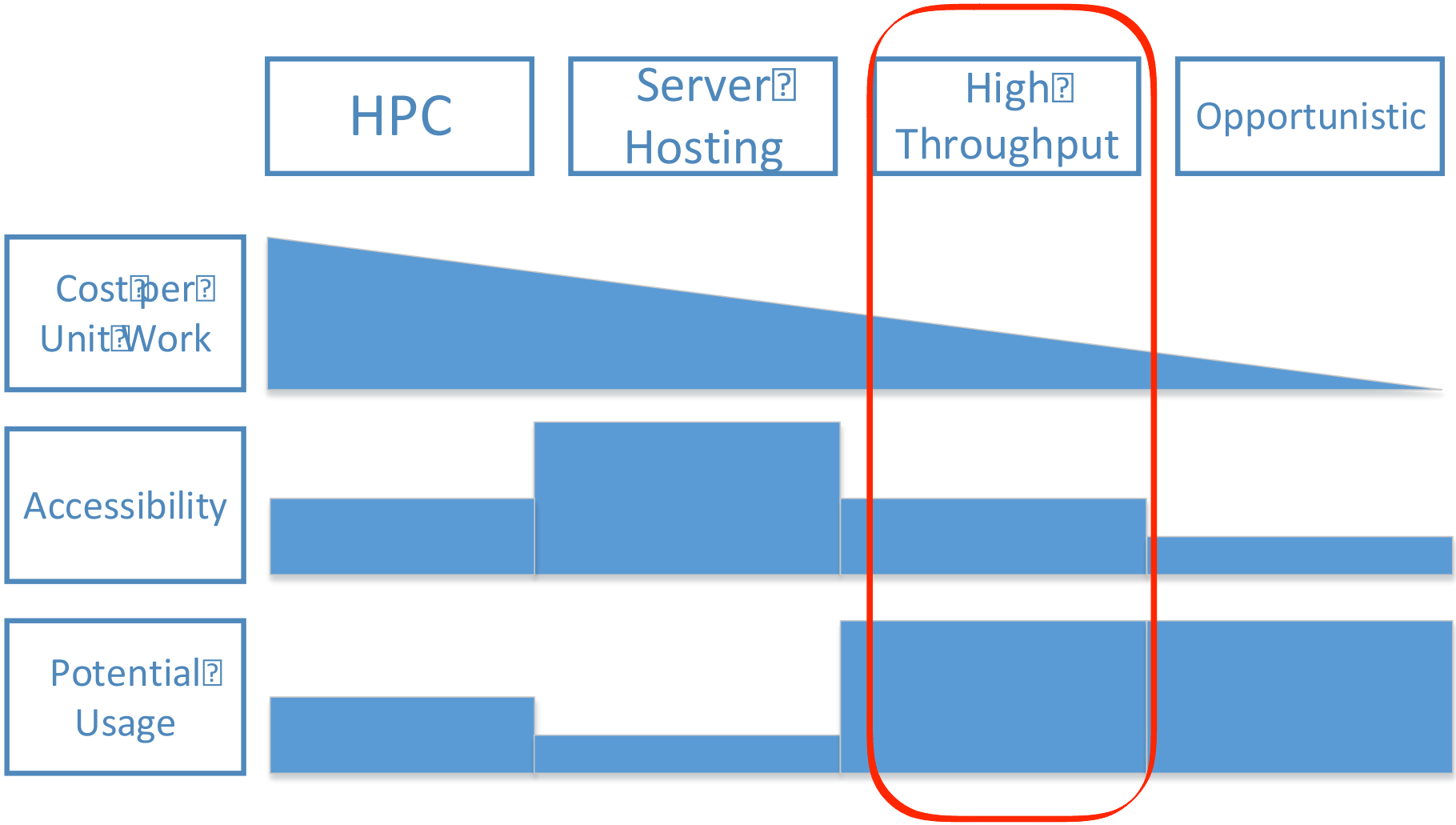
- Universal resource provisioning layer for bare metal, containers and VMs
- HTCondor as the single end user interface
- Continue investing in automation and other communities for scaling with fixed staff
- Self service for end users within the policies and allocations



Compute Workloads



Compute Workloads



High-throughput batch service

- 113k CPU cores and increasing
  - LHC: Prompt Tier-0 calibration / hot events
  - LHC: Tier-0 bulk reconstruction
  - LHC: Our share of WLCG Grid quota
  - All other CERN experiments (e.g. Compass)
  - Associated experiments (e.g. AMS)
  - Various local CERN groups in EP (ATLAS, LHCb, etc)
  - Typically delivering around 500k jobs per day

High-throughput batch service

- Batch service balances the fair-share across all competing applications according to CERN resource policies
- Users interaction pattern: “submit a job, sits in queue, runs, get result back”

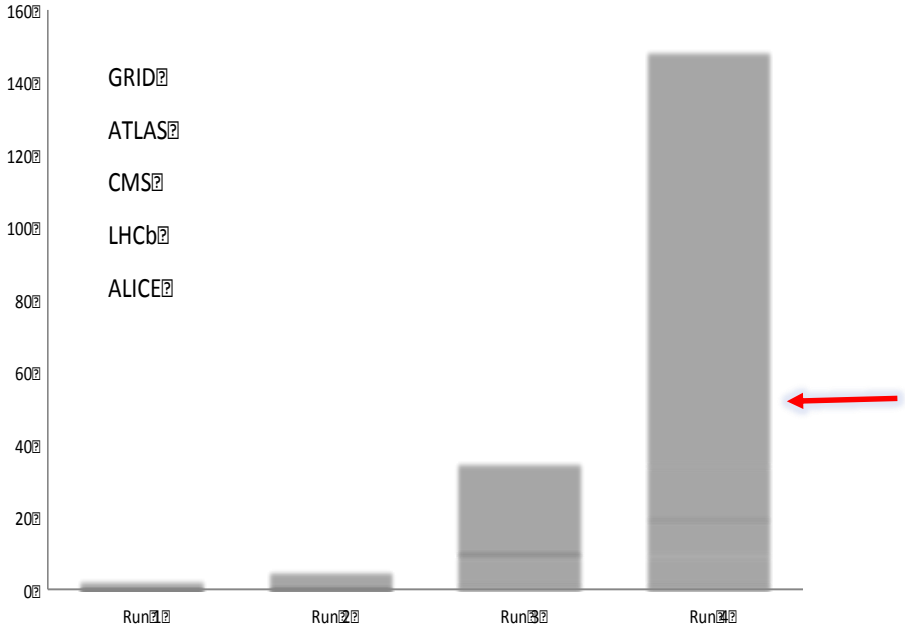
- Currently migrating from proprietary product (LSF) to Open Source HTCondor
- We also run misc Griddy services (Compute Element and Argus) interfacing WLCG Grid to local site resources

Batch HTC challenges and plans

- Scaling the service for Run3 and beyond
- Understanding and improving overall CPU efficiency (currently 60%)
- Moving 1300 users to Condor (~end of Run2) and retiring LSF

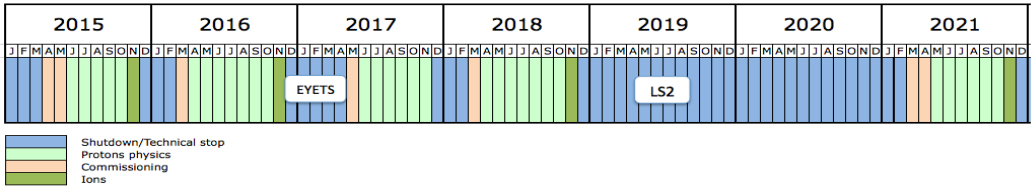
- Expanding into the external cloud / hosting…

Computing scale challenge



- Efficiency studies…!
- Expand to public cloud?
- Volunteer computing?

The outline LHC schedule out to 2035 presented by Frederick Bordry to the SPC and FC June 2015 can be found here

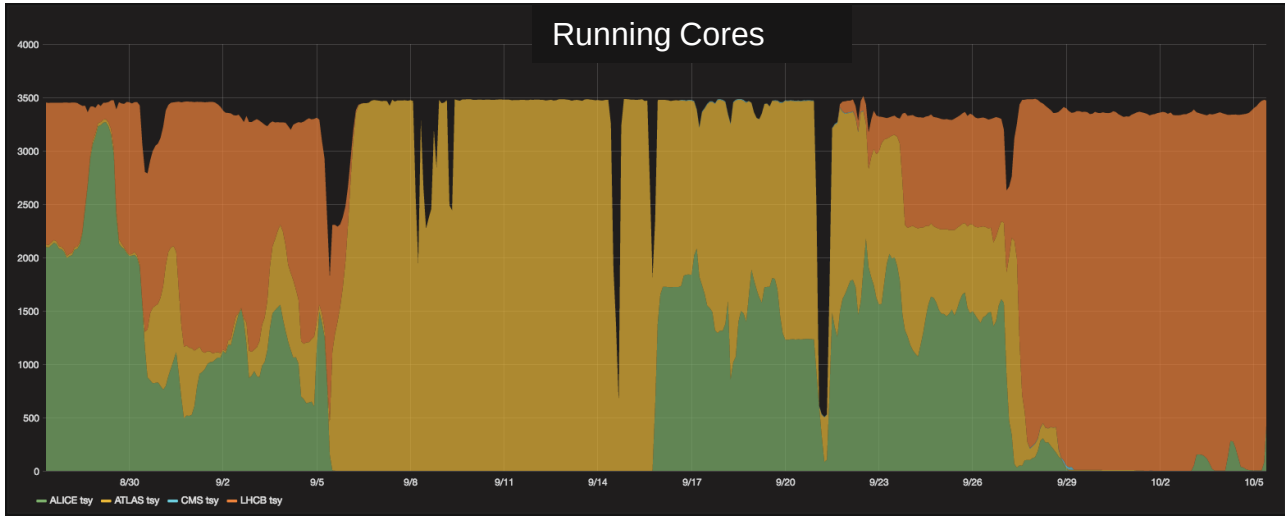


To the public cloud and beyond..

- Investigating extending HTCondor pool to include external cloud resources
- IBM Softlayer -> T-Systems -> HNScience Cloud
- Public procurement challenge
- From technical HTCondor point of view, adding external nodes was quite easy
- More work involved from puppet infrastructure, monitoring, cloud APIs, and coordination with experiments
  - Standard tooling (cf. interfaces) critical

Recent activity: T-Systems

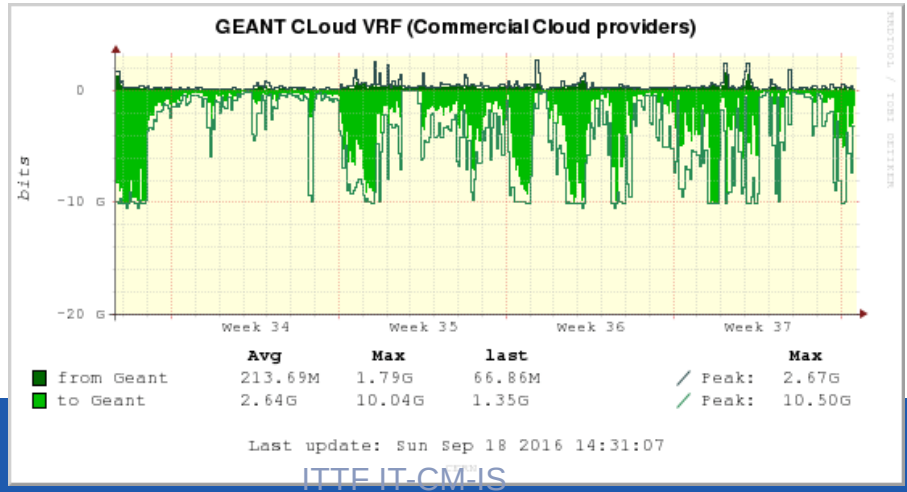
- Batch resources fully loaded
  - shared among VOs



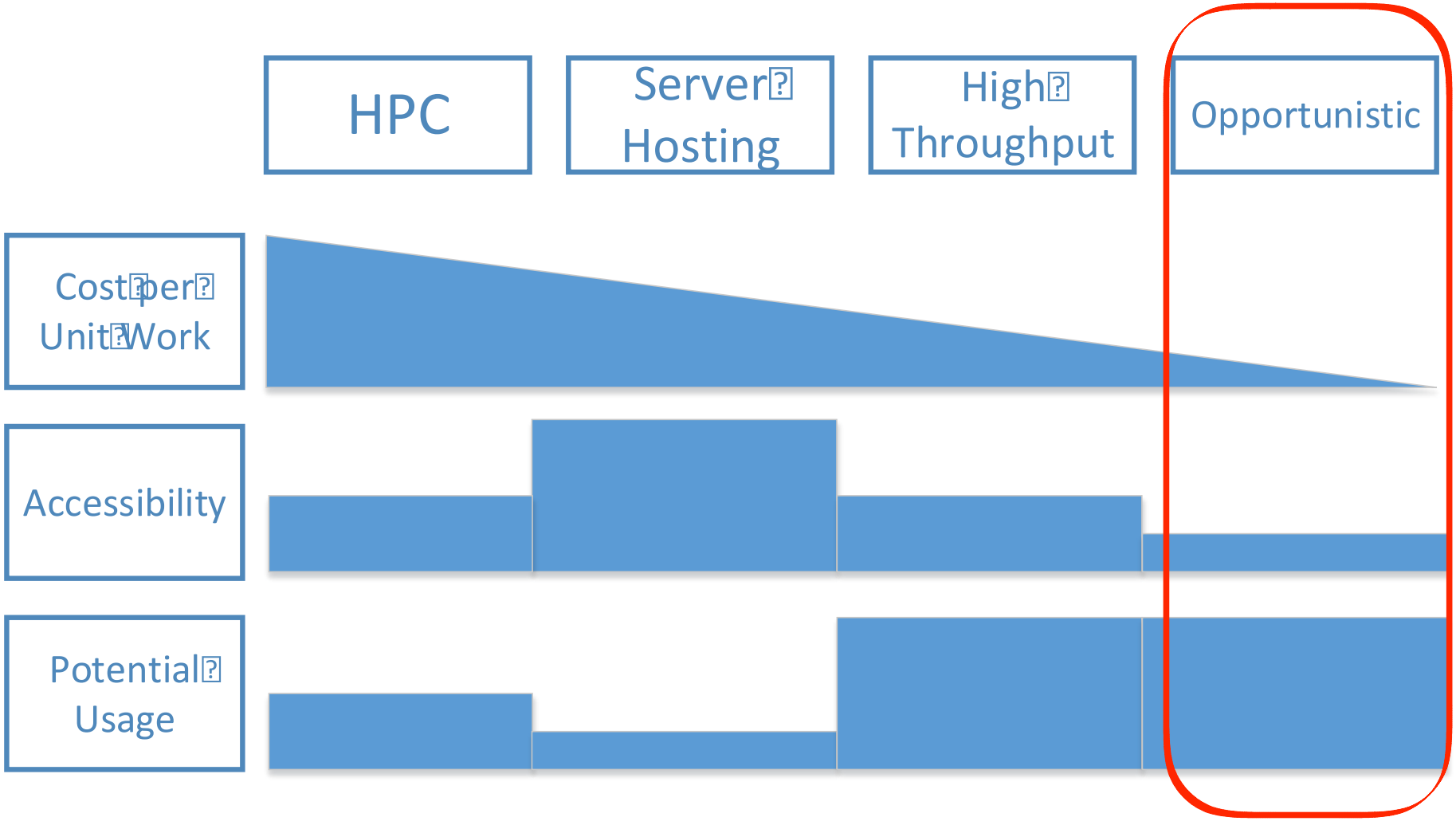
- Mixture of “CPU-intensive” and “network-intensive” tasks
  - MC workloads tend to dominate: easier to manage?

	Max	Avg
LHCB tsy	99.05	85.04
ALICE tsy	93.83	75.98
ATLAS tsy	100.00	64.13

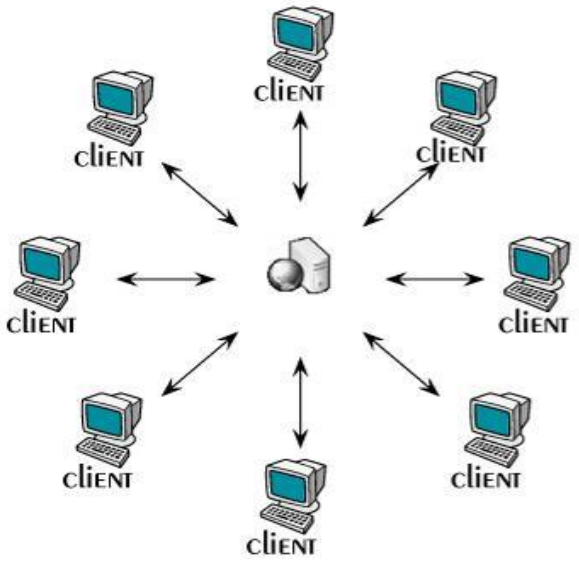
- WAN largely used
  - Sometimes even saturated



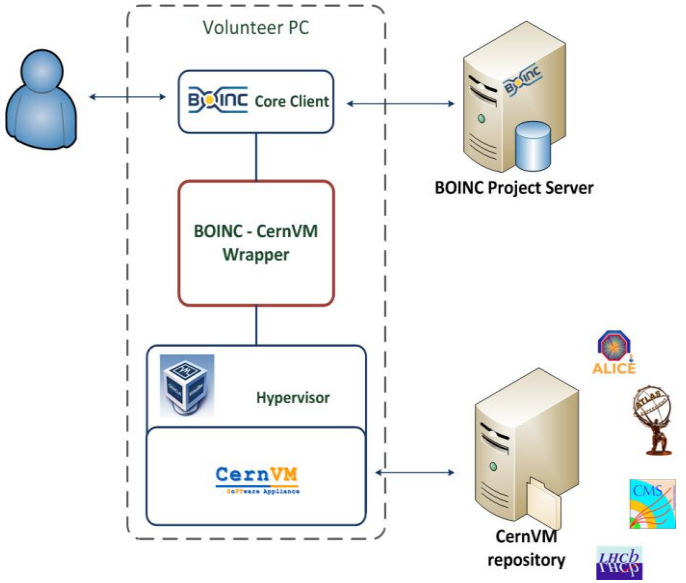
Compute Workloads



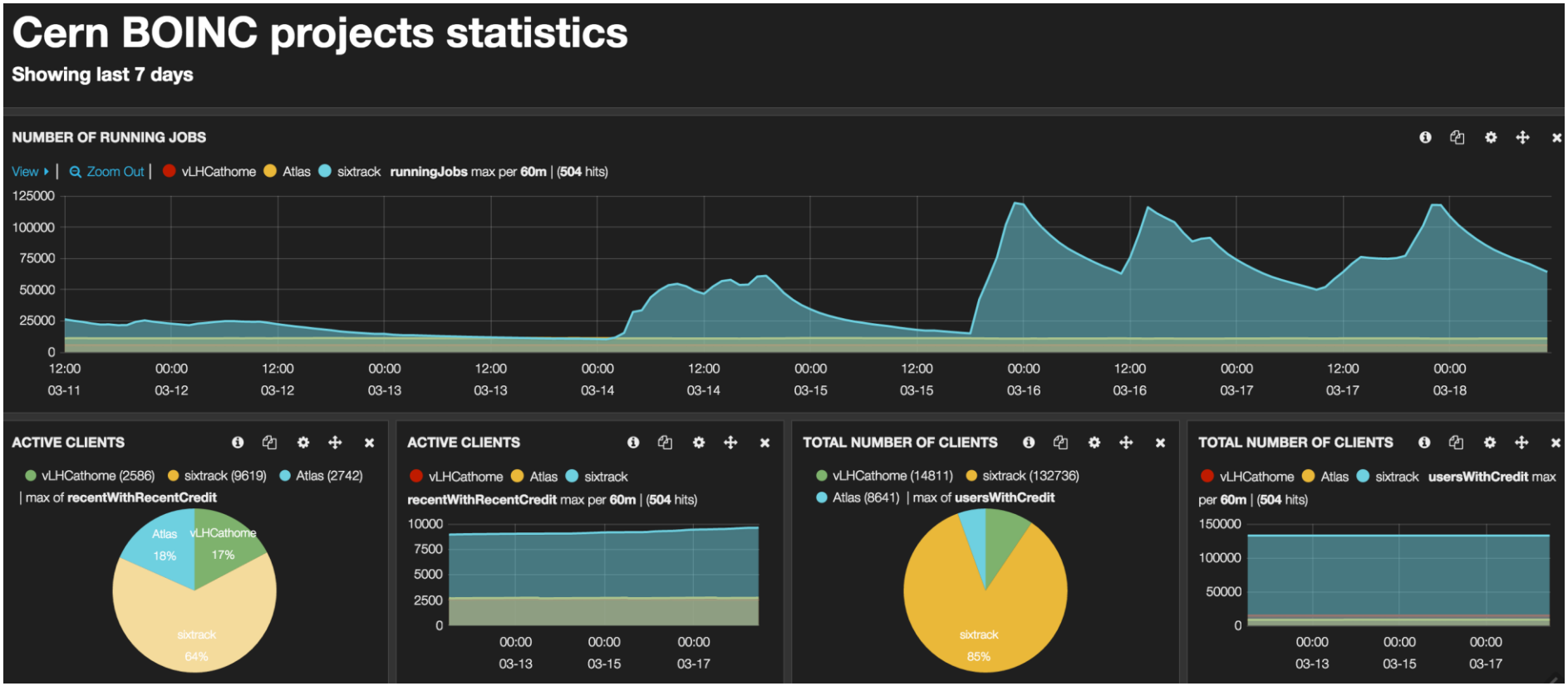
LHC@home



- Scavenged resources
  - Allows us to get additional computing resources
  - Volunteers (e.g. home PCs)
  - Institute desktops
  - Supercomputing backfill
  - Small farms with easy deployment!
- Unpredictable but significant resources
- Target CPU bound simulations (not data intensive)
  - Over 50% of LHC compute is simulation!
- Outreach benefits!



Lots of potential resources…



Opportunistic challenges and plans

- Resources free but ops cost is not
- Reduce ops overhead - leverage same technologies as we're using to expand into the cloud (e.g. Condor)
- Ramping up capacity and ensure the BOINC infrastructure can scale

- Continue to work with experiments to find workload that's suitable for this use-case

Compute Workloads