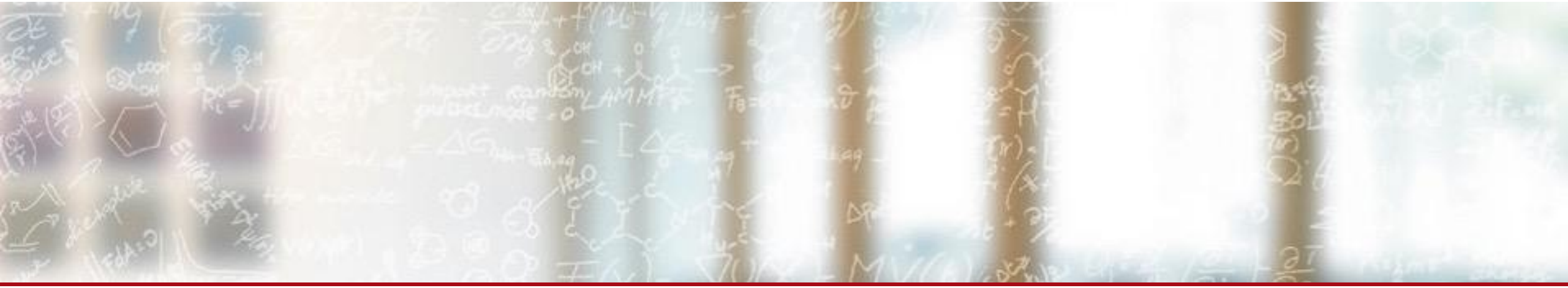
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich



Tier-0 spill-over status

CHIPP-CSCS face to face meeting

Pablo Fernandez, Gianfranco Sciacca, Miguel Gila

21st June, 2018

LHC Tier-0 spillover

- Goal: elastic provisioning of Tier-0 & HLT activities
  - Support computational peaks / support spill-over / on-demand for the Tier-0 at CERN
  - Open the way to new balancing between Tier-0, Tier-1 and Tier-2 activities
  - Evaluate solution and interaction in preparation for Run 3 (in 2021+)

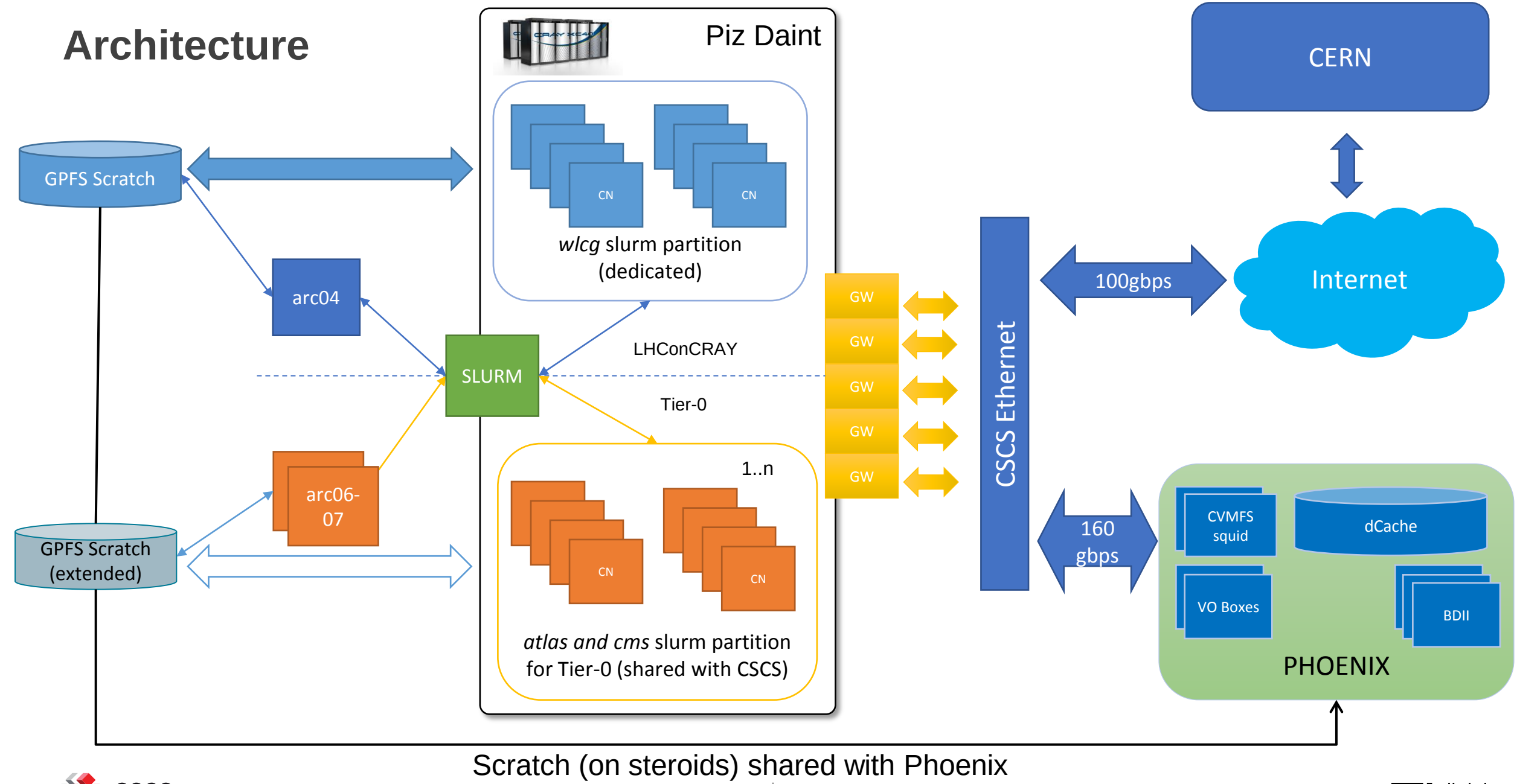
LHC Tier-0 spillover implementation

- ATLAS and CMS have selected two workflows and tests are underway

- Extra resources have been made available
  - Up to 150 nodes on Piz Daint (3x of what is currently being used)
  - 1 PB of storage on dCache (800 TB ATLAS, 200 TB CMS)
  - More DVS Servers (5 → 7 → 13 → 19)
  - 4 new ARC servers (2 ARC endpoints, 2 data staging)

- Dynamic allocation of resources, Slurm takes care of everything:
  - Mount/unmount CVMFS on demand (4GB cache)
  - Start/Stop node monitoring
  - Dedicated queue overlapping with regular MC nodes
  - Static Swap (evaluating usage)

Architecture



ATLAS workflow

- Bursty on-demand processing on “physics_Main stream”
  - Reconstruction of a full run: 50-60 TB of input RAW data, 35-40 TB of output data
- Steady processing on “physics_BphysLS stream, O(10%) of physics_Main rate”
  - 5-6 TB of input RAW, 2 TB of output data
- Data pre-staged to a dedicated area in dCache
  - Stage-in from dCache and stage-out to CERN performed asynchronously by ARC

Current status

- Facing several complex integration issues on both sides
  - Shortage of memory
  - ARC instabilities
  - SLURM scheduling
  - ATLAS optimised job definition and job brokering
- Some progress, but not quite there yet, still commissioning the 10% stream workflow

CMS workflow

- Regular Tier-0 workload is more of a background task than a spill-over
  - In this mode, the idea is to lower the priority against ATLAS (to allow for their peaks)
- Most interesting “spill-over” task is the b-parking stream
  - But being more critical, CMS needs to make sure the Tier-0 simulations work smoothly first
- Data to be read directly from CSCS compute nodes from CERN EOS
  - Potentially doable, below ½ of CSCS internet connection

Current status

- CMS is running on 2 nodes
- Testing successful since 20.06.2018, evaluating how to approach scale-up

Conclusion

- ATLAS and CMS are exploiting additional resources on Piz Daint, with the idea to assess feasibility to do further Tier-0 activities after the big LHC shutdown
- Up to 150 nodes on Piz Daint will be used, which is equivalent to what the Tier-2 will use in 2019
  - Hence it's the perfect test for what's coming!
- The dependencies with the Tier-2 have been kept low, but not null
  - dCache is shared, no way around that, but extra space has been granted (1 PB)
  - Scratch is shared, but greatly enhanced (SSD layer put in front)

- Gianfranco and Pablo are paying attention to minimizing the impact on the Tier-2