PNNL-SA-127779

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Belle II Distributed Data Management and Networking

VIKAS BANSAL, MALACHI SCHRAM, ANTONIO LEDESMA

Pacific Northwest National Laboratory, Richland, WA

DPF ‘17, Fermilab, August 3, 2017

Belle II Collaboration

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

c.f. CERN Greybook July 2017
ATLAS: 39 countries, 217 inst., 7783 members
CMS: 49 countries, 208 inst., 6217 members
ALICE: 41 countries, 167 inst., 2799 members
LHCb: 17 countries, 74 inst., 1494 members

Computing requirement on par with LHC Run I

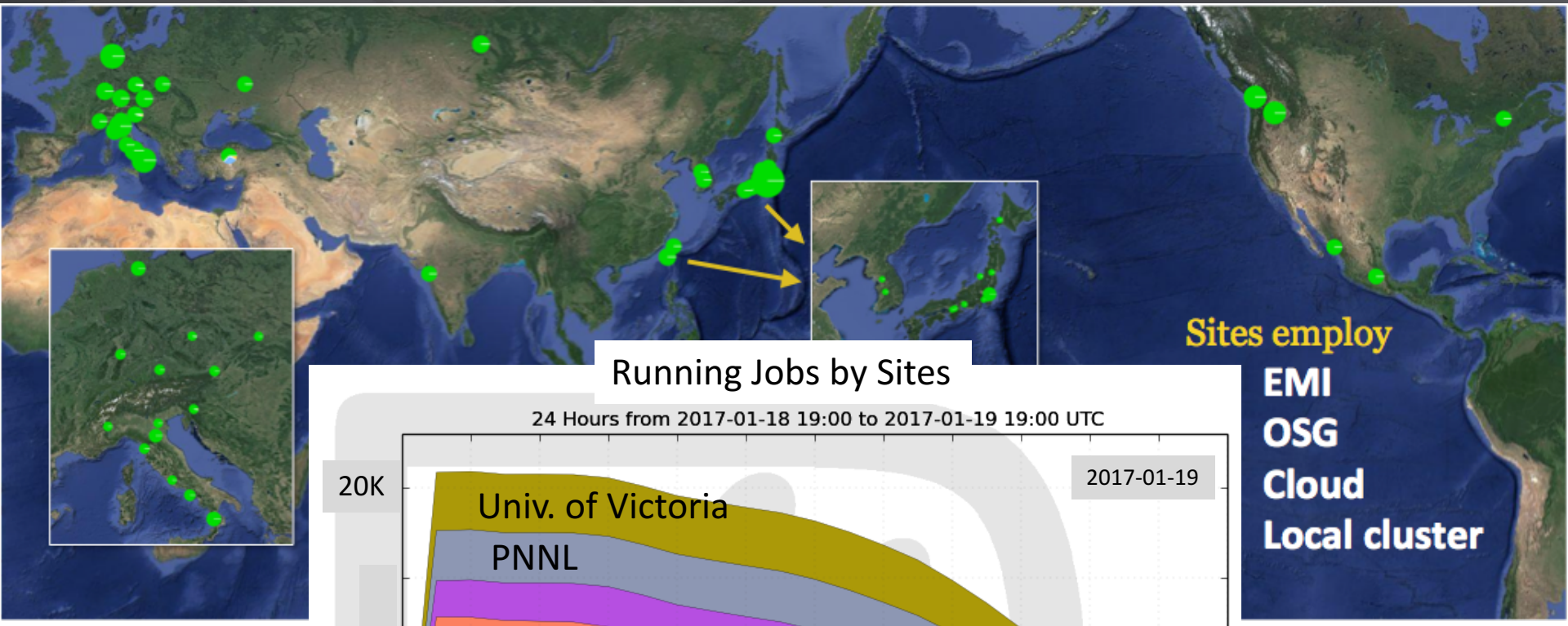
- The Belle II experiment is for Super B factory at KEK in Japan
- Complementary physics to the LHC based on precision measurements from high-intensity beams
- Total integrated luminosity : 50 ab^{-1}
- Collisions start in Early 2018
- Similar data rate as from LHC Run I

Raw Data : 100 KB / event
Detector mDST : 5 KB / event
MC mDST : 6 KB / event
Reconstruction : 20 HEPSPEC *s / event
MC : 60 HEPSPEC *s / event

	2017	2018	2019
Disk [TB]	3000	4500	11000
Tape [TB]	2000	2200	6500
CPU [KHEPSPEC06]	210	400	480

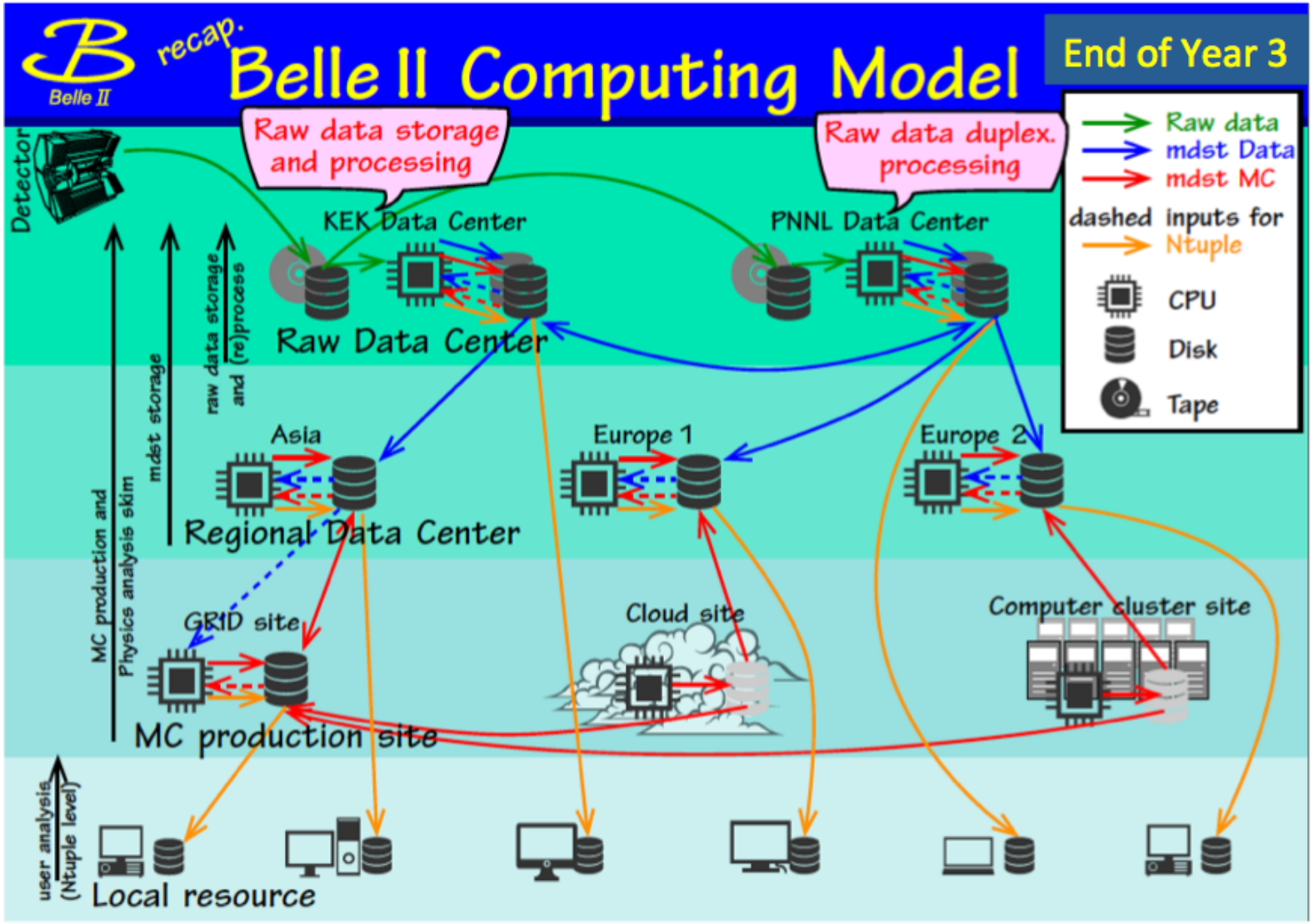
Belle II computing sites across the globe

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle *Since 1965*



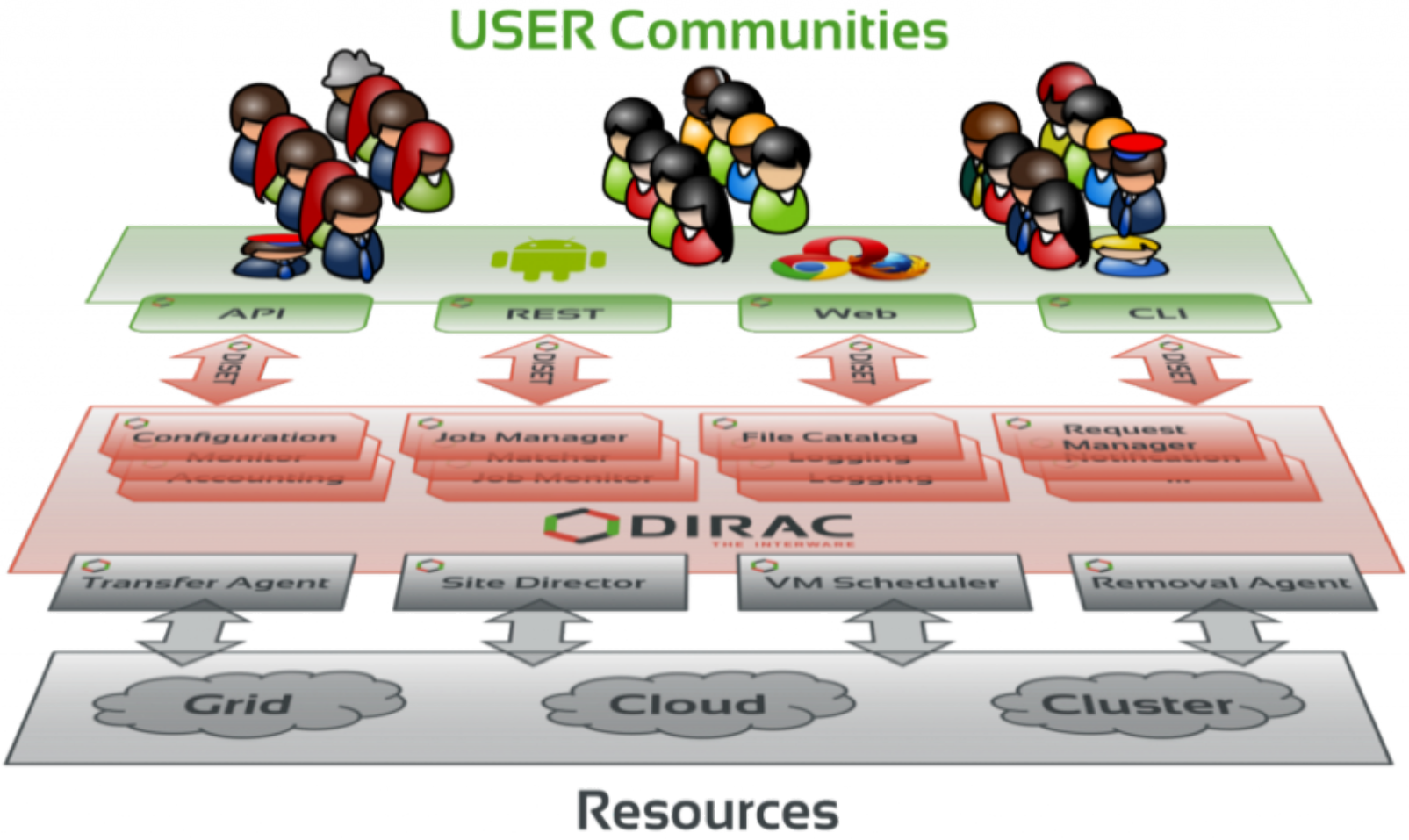
Belle II computing model until Year 3

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965



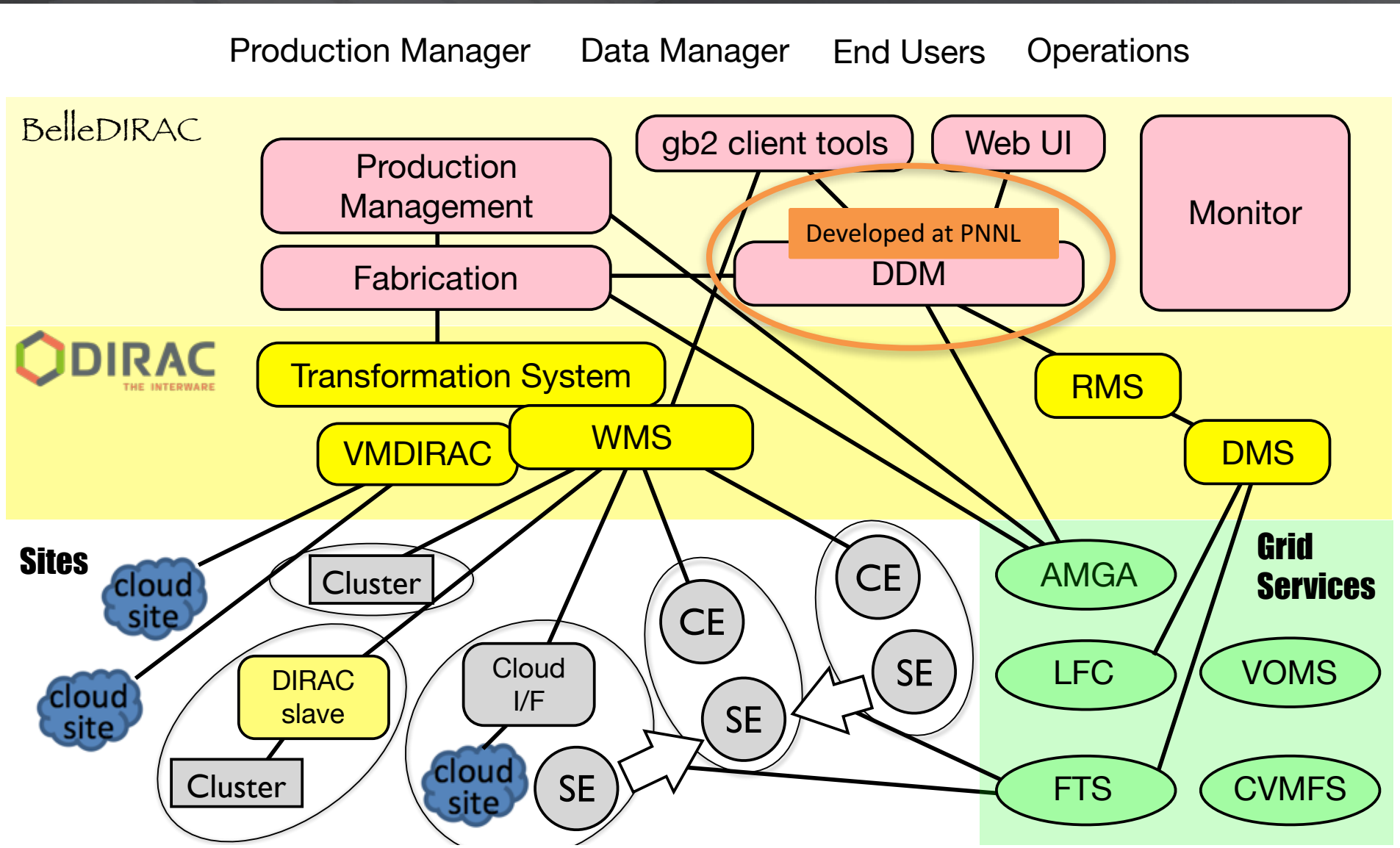
Belle II distributed computing software

- DIRAC (Distributed Infrastructure with Remote Agent Control) as the solution of choice
- Successfully used by LHCb. Extended for Belle II.



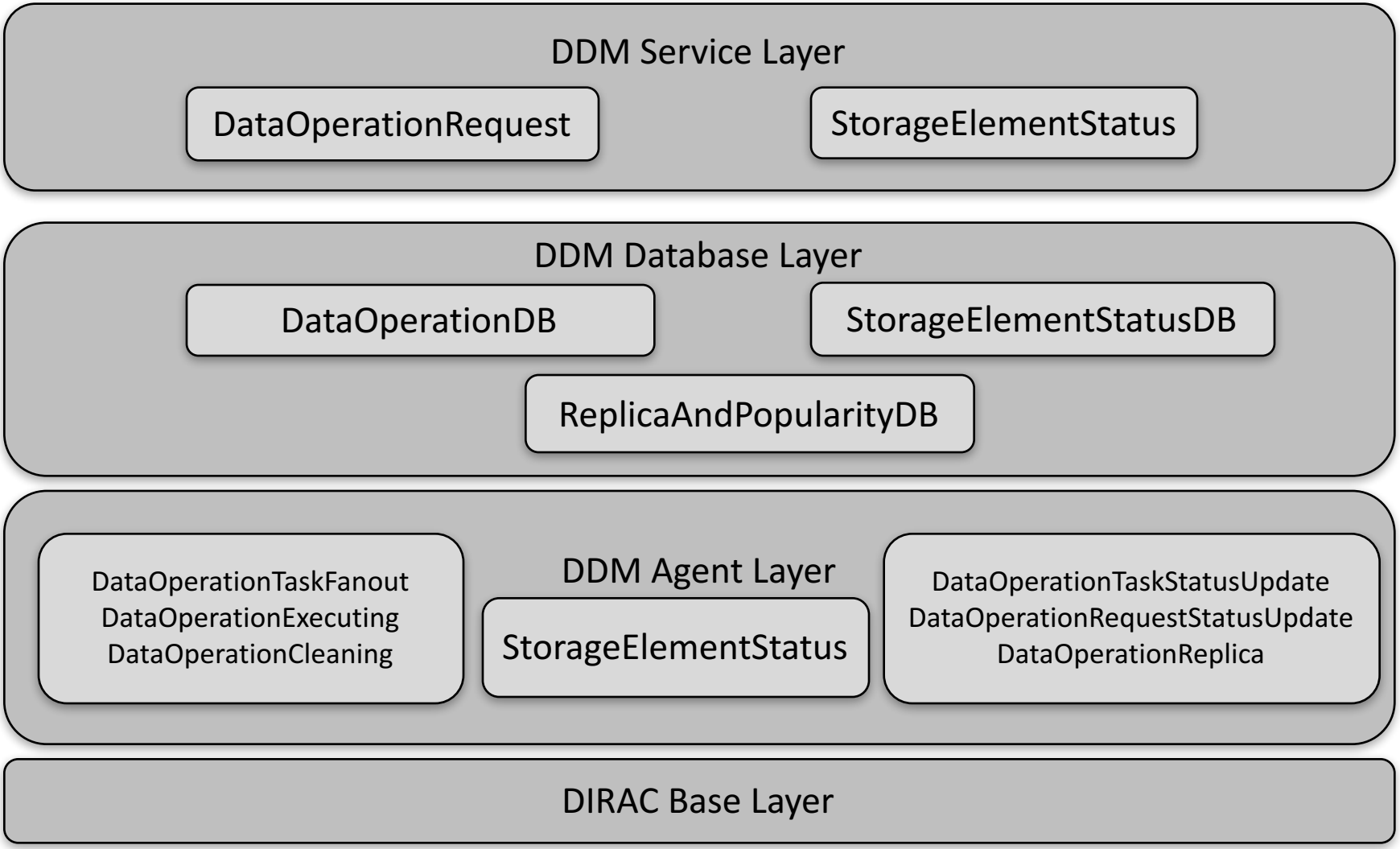
Belle II distributed computing layers

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

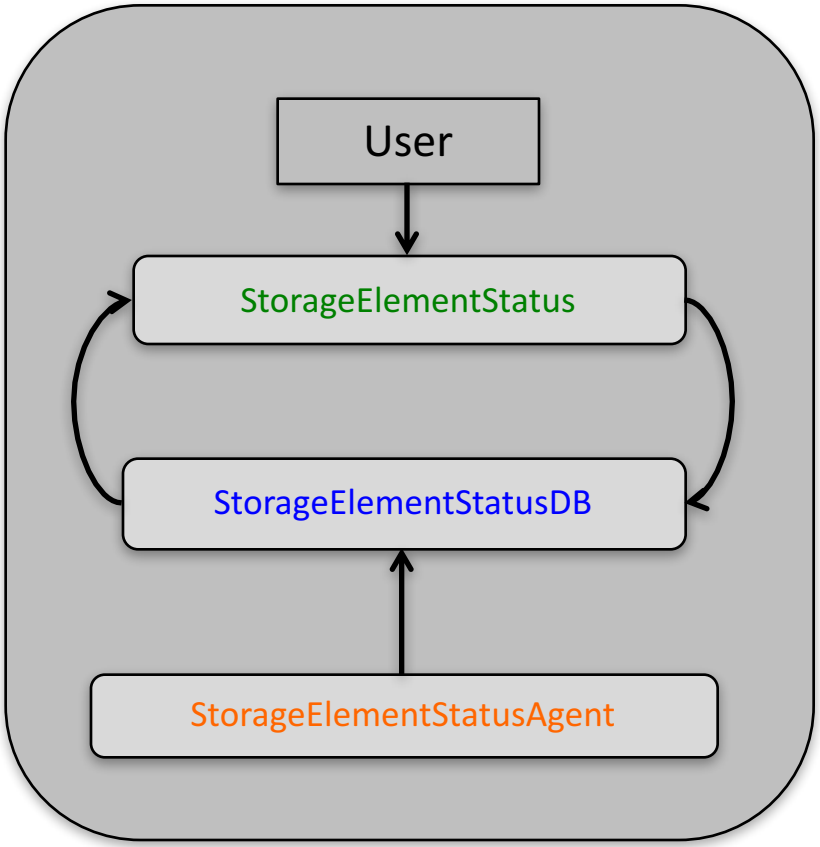


Distributed data management system overview

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965



StorageElement Status Workflow



Purpose:

- Provide near real time storage elements information

Current implementations:

① StorageElementStatus provide API layer for internal and external users.

② StorageElementStatusAgentDB provide persistified information

③ StorageElementStatusAgent :

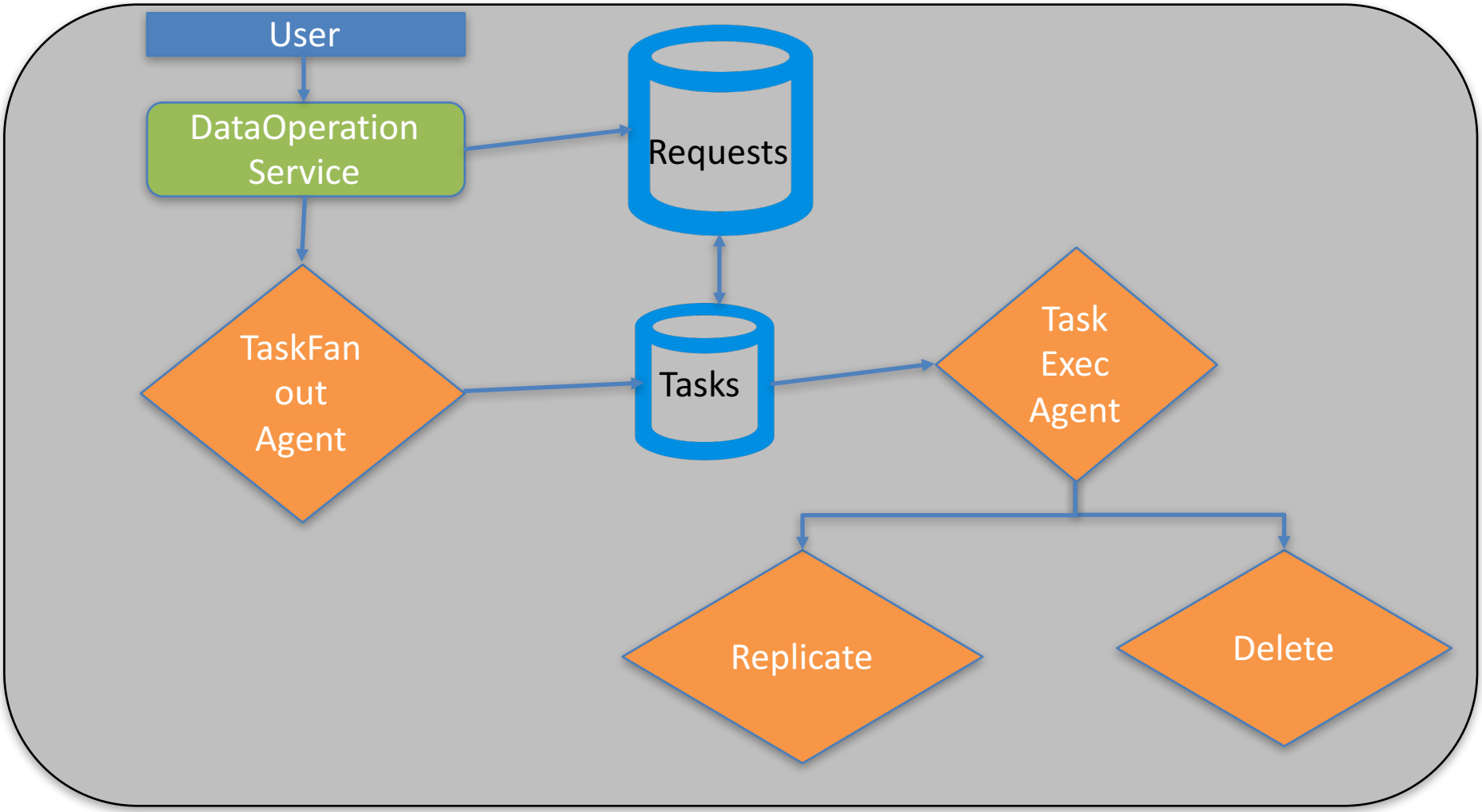
- Available space
- Access rights (w/r) at file and director level
- ADLER32 Checksum

Extra features:

- Include full ACL

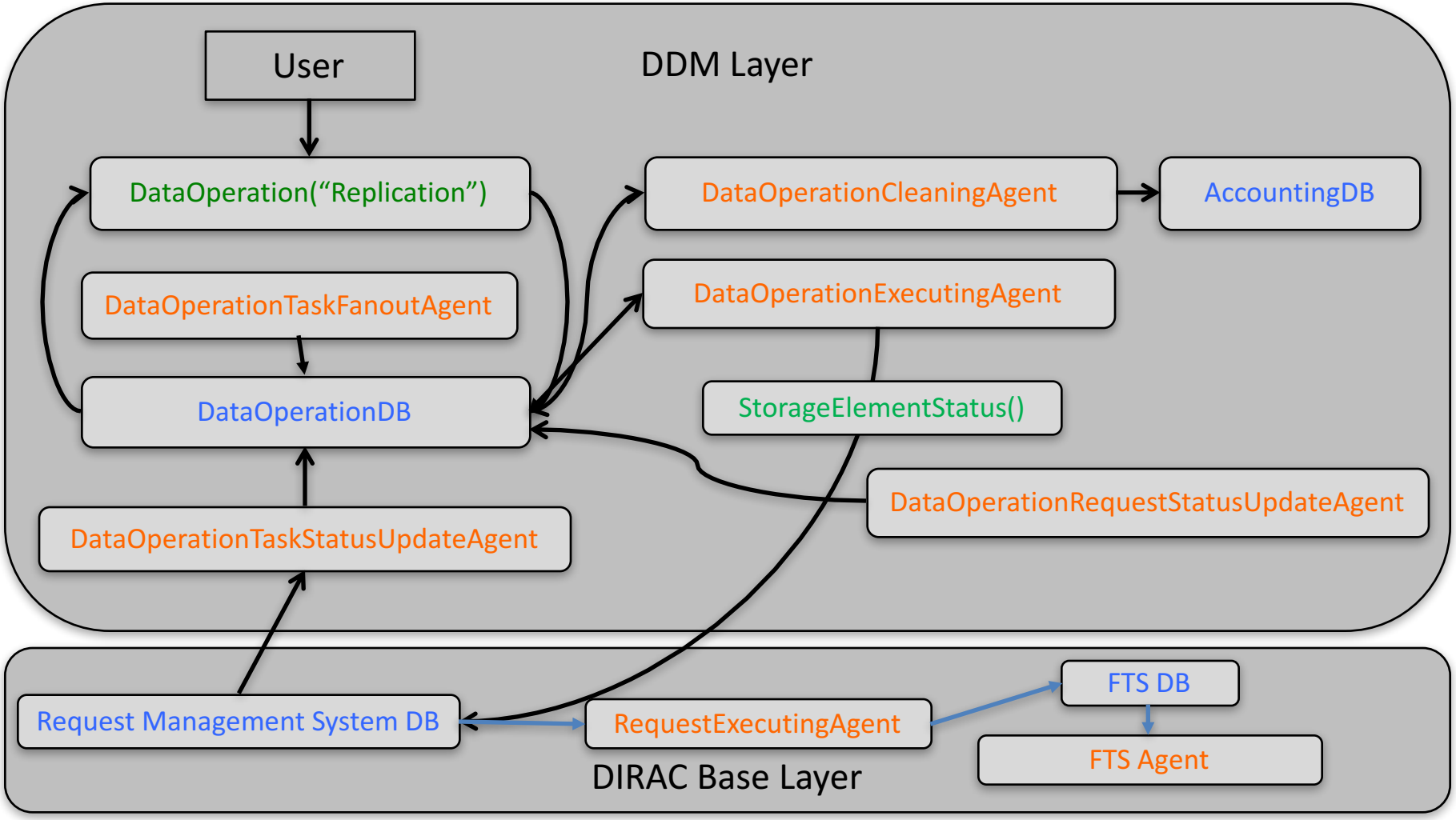
DataOperation : Workflow Overview

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

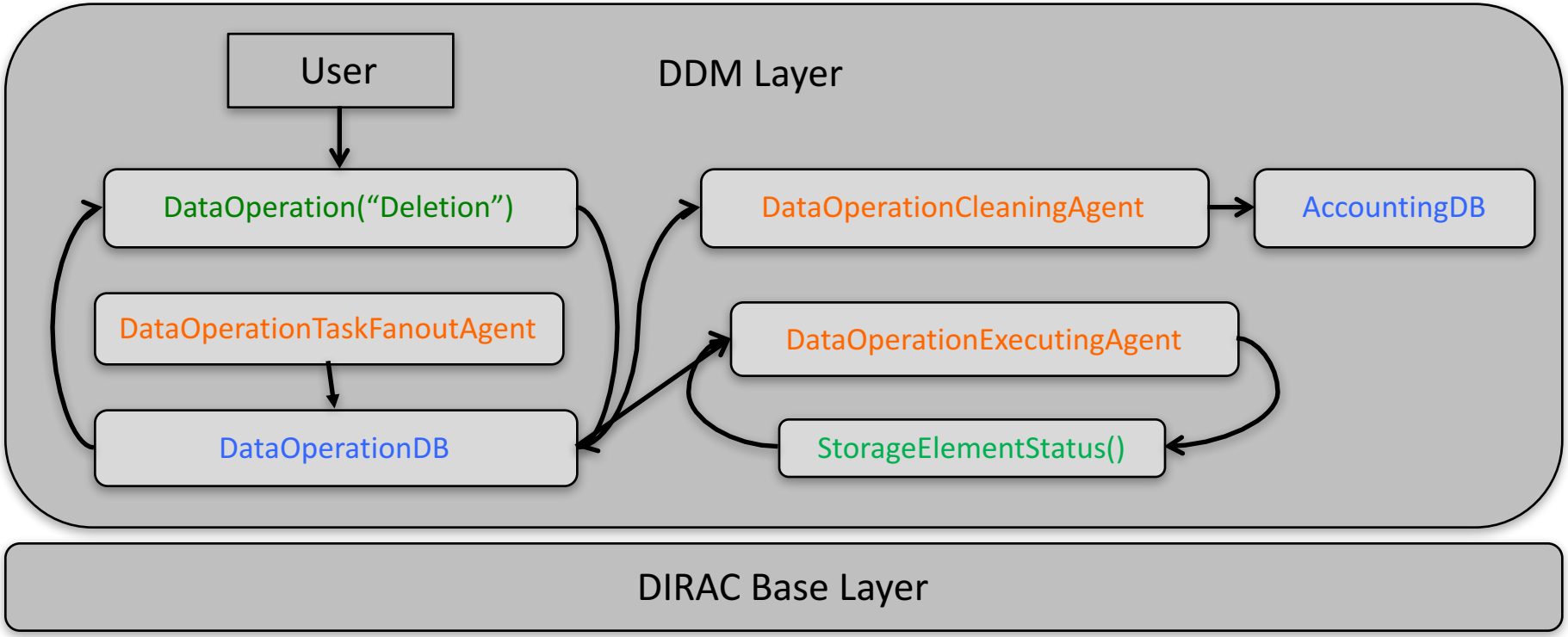


DataOperation : Replication Workflow

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965



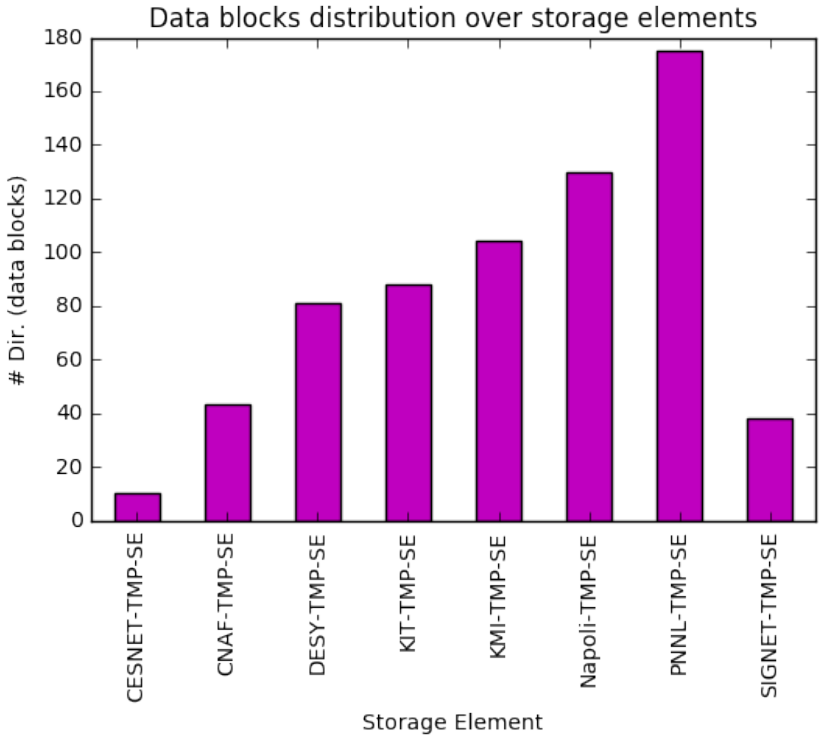
DataOperation : Deletion Workflow



- DDM by design does not delegate deletion of files to base DIRAC to avoid race conditions on data files
  - In addition Belle II utilizes AMGA (metadata service) that is not available in base DIRAC
- Data operation executing agent only acts on one operation on physical file at any given time

Distributed data management system in production

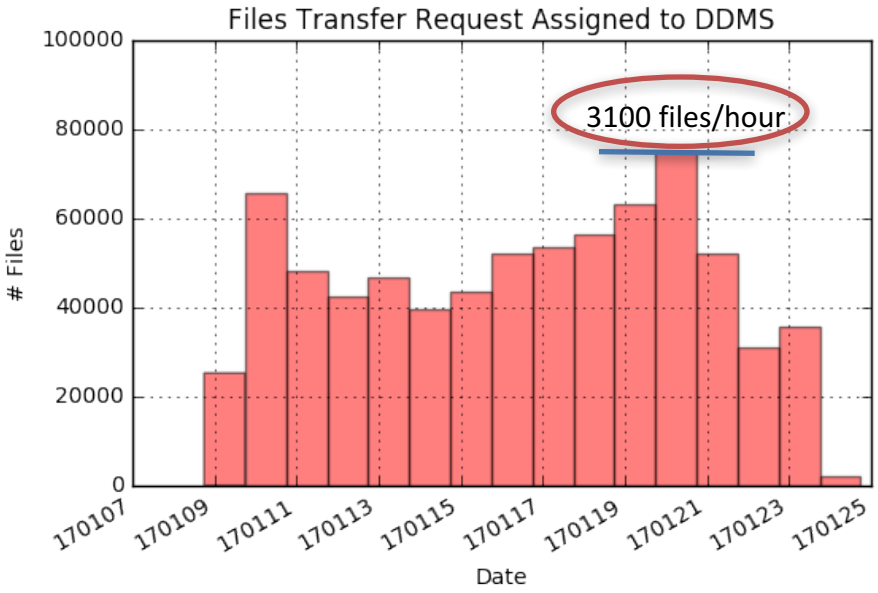
- Distributed Data Management System (DDMS) is successfully used in Monte Carlo (MC) samples production
- Production system hand overs data files to DDMS
  - DDMS decides to transfer the files/data to specific sites as per policy.

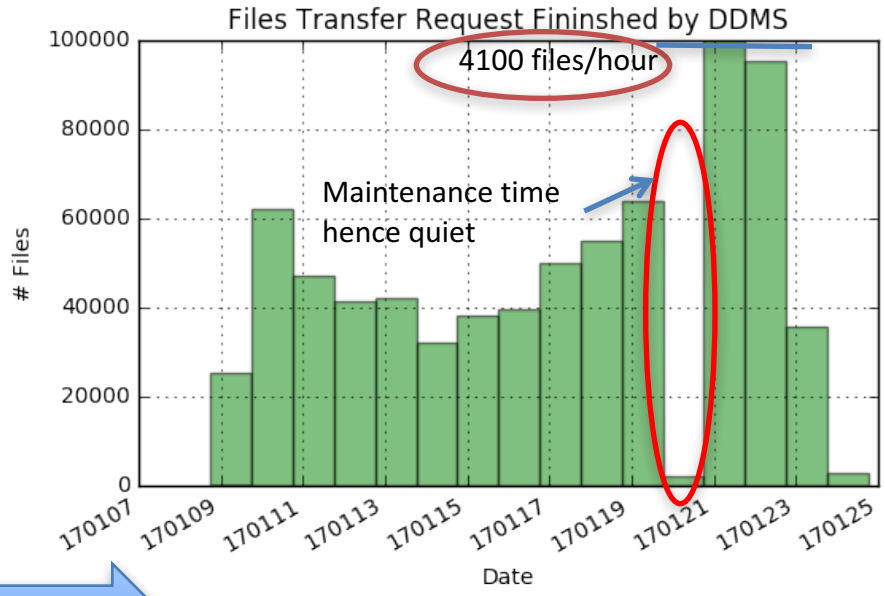


Distributed data management system in production

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Production system hand overs data files to DDMS
  - DDMS check for site health before starting any transfers. Replicates through DIRAC using File Transfer Server and reports back finished transfers

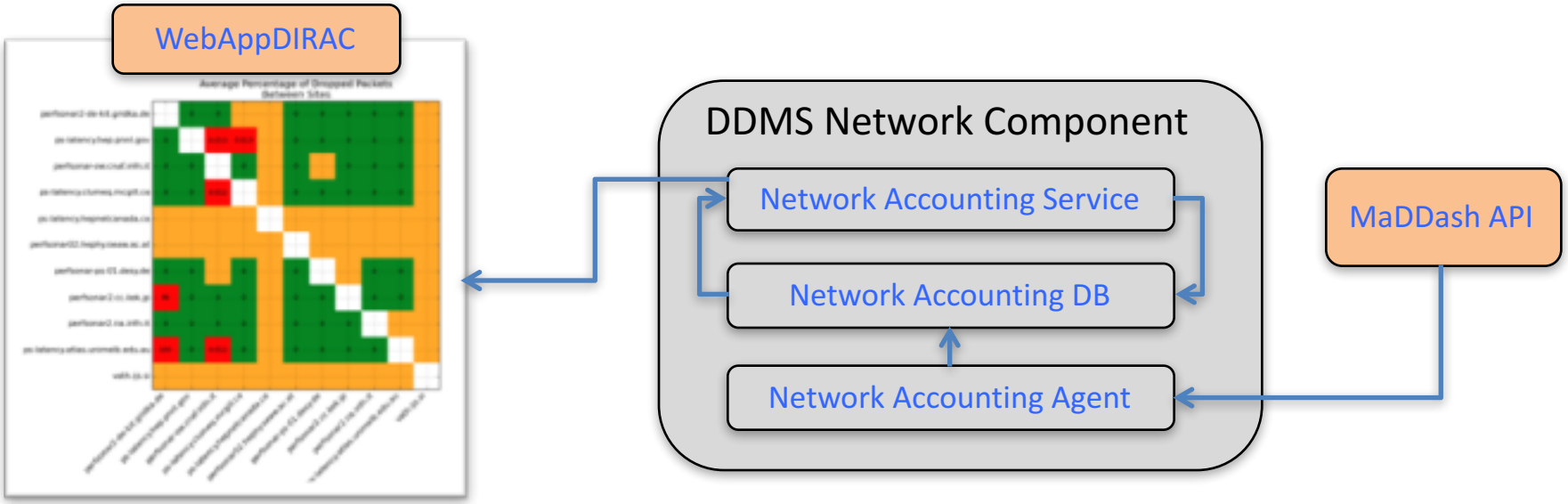




Belle II wide area networking

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

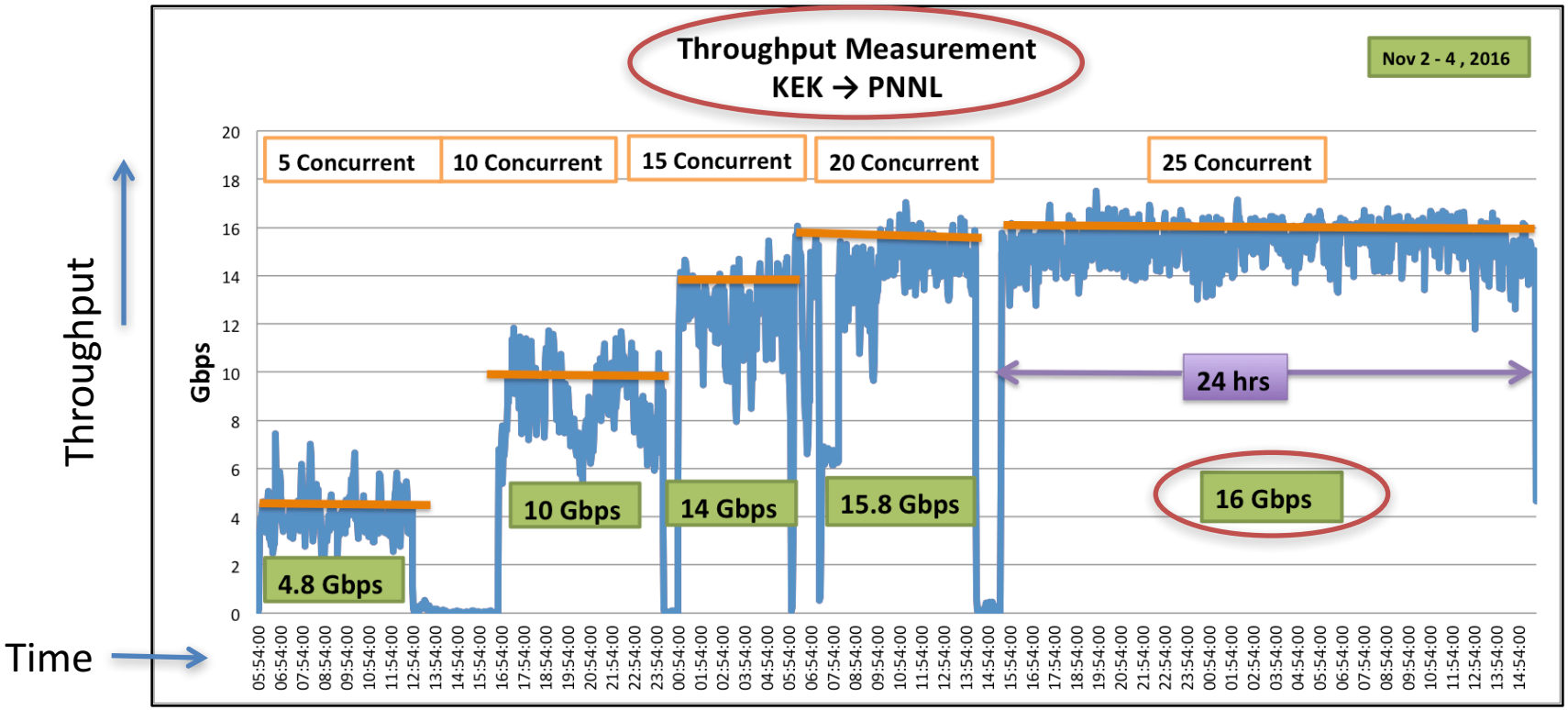
- Powered by National Research and Education Networks (NREN), e.g. ESnet
- Dedicated servers set up at Belle II sites for network monitoring
- PNNL leads the effort in setting up Belle II Monitoring Dashboard (MaDDash) mesh to measure sites-pair wise bandwidth and latency
- Network info is part of DDMS inside DIRAC
- Automate notification to sites with network problems



WAN data challenge KEK ↔ PNNL

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Estimated network bandwidth for peek outgoing traffic from KEK : 9 Gbps
- KEK outgoing traffic measured at 16 Gbps
- DDM can be made to asses various network paths for optimal data transfers



Summary and Plan Moving Forward

- Belle II computing needs on par with LHC Run I
- Belle II distributed computing software is written in DIRAC framework
- Distributed data management system handles data transfers and deletion and is currently deployed in production mode
- We are actively working on next version of DDM that will directly schedule transfers to FTS to avoid DIRAC's RMS layer
- We foresee to add network health and monitoring in distributed data management system

Detailed Deletion Workflow

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965