# Belle II computing

Ikuo UEDA (KEK IPNS)

The 40th Anniversary Symposium of the US-Japan Science and Technology Cooperation Program in High Energy Physics

April 15, 2019 @ University of Hawaii

JFY 2012 - 2015 Japan:KEK + US:PNNL Project Completed

Establishment of a remote data center for acceleration of Belle II data center

Belle IIデータ再プロセス高速化のためのリモート・データセンターの設立

JFY 2016 - 2018 Japan:KEK + US:PNNL→BNL(since 2018) Project Completed

Development of a scalable and automatized production system for the Belle II experiment (US side research title : 2016)

Automatized Production System for Belle II (resarch title was renamed : 2017 or later : funded by the Japan side only)

Belle II実験における拡張性を考慮した自動化プロダクション・システムの開発

JFY 2019 (- 2020) Japan:KEK + US:PNNL Project Applied

Hiding Data Access Times in HEP Distributed Workflow

*JFY 2012 - 2015* *Japan:KEK + US:PNNL* *Project Completed*

*Establishment of a remote data center for acceleration of Belle II data center*

*Belle II*データ再プロセス高速化のためのリモート・データセンターの設立

*JFY 2016 - 2018* *Japan:KEK + US:PNNL→BNL(since 2018)* *Project Completed*

*Development of a scalable and automatized production system for the Belle II experiment* *(US side research title : 2016)*

*Automatized Production System for Belle II* *(resarch title was renamed : 2017 or later : funded by the Japan side only)*

*Belle II*実験における拡張性を考慮した自動化プロダクション・システムの開発



*JFY 2019 (- 2020)* *Japan:KEK + US:PNNL* *Project Applied*

*Hiding Data Access Times in HEP Distributed Workflow*

# Purposes

*JFY 2012 - 2015*

## *Establishment of a remote data center for acceleration of Belle II data center*

*Goal : Acceleration of the speed of the Belle II data reprocessing*
*by establishing the remote data center in U.S.A.*

*to trigger the Belle II computing activity in U.S.A.*
*to let the KEK computing resource concentrate on RAW data process*
*to reduce the risk of data loss in unexpected contingency*
*to develop human resources for computing and middleware*

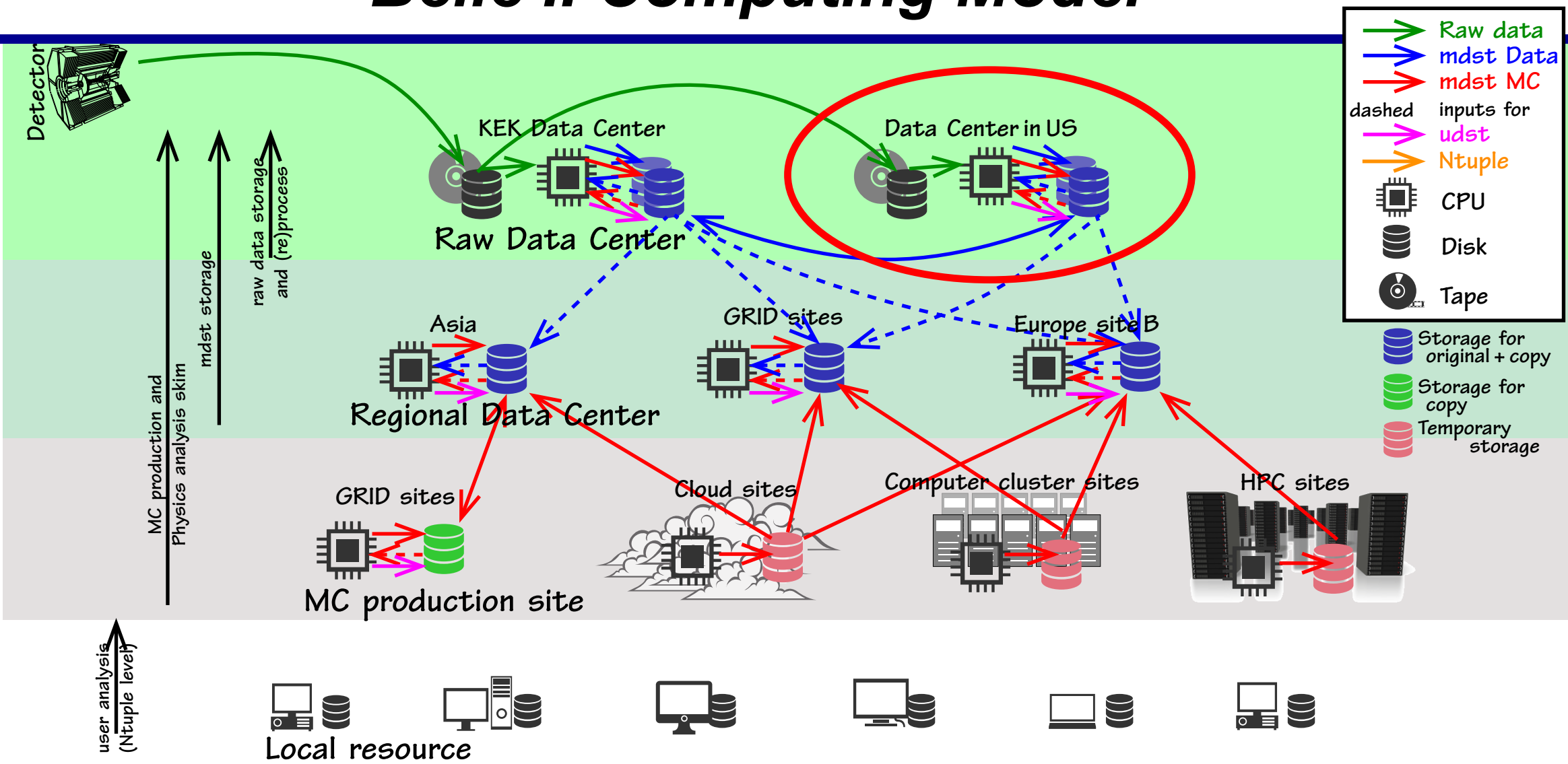
*JFY 2016 - 2018*

## *Automatized Production System for Belle II*

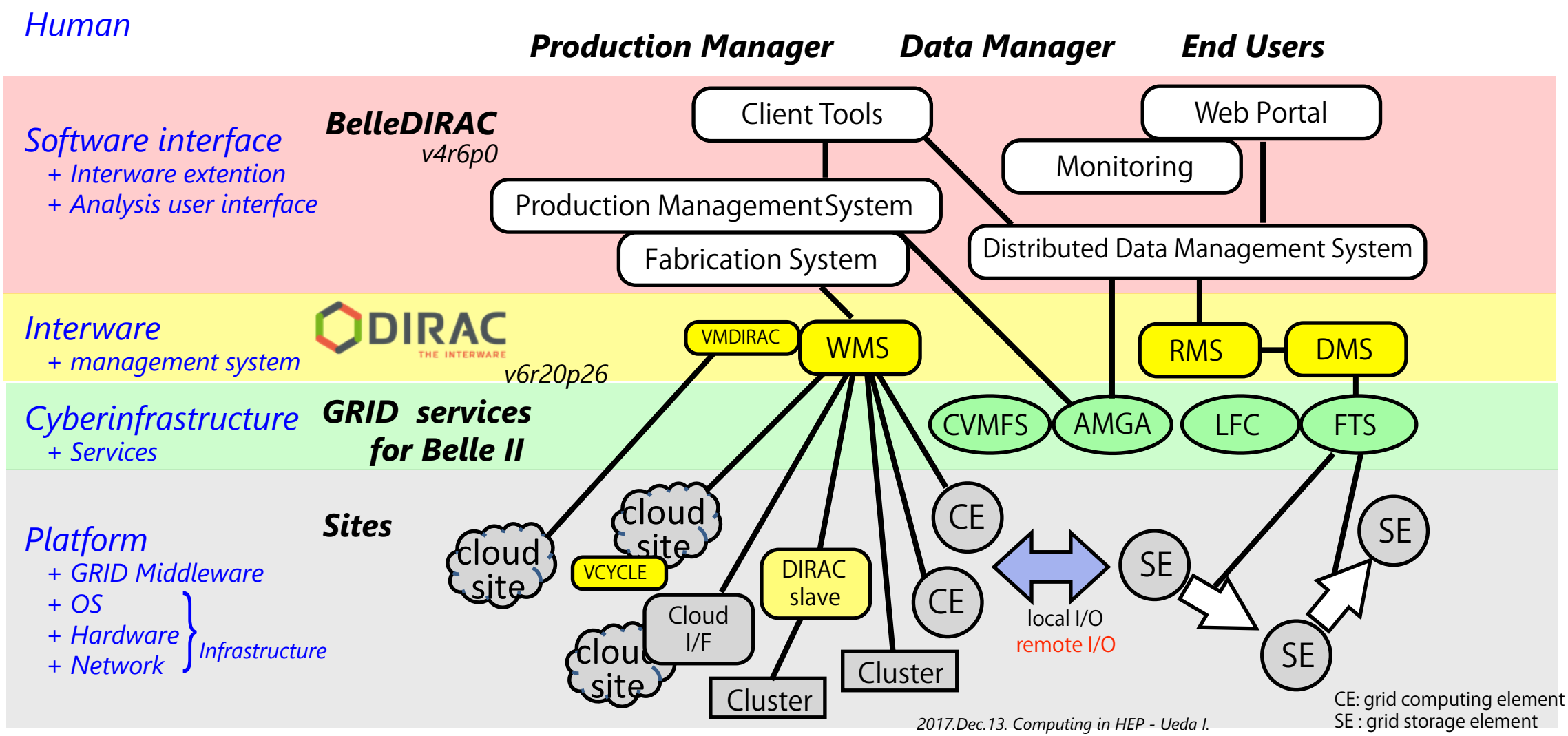
*Goal : Integration of the scalable and automatized production system to the Belle II experiment*

*to reduce the burden on expert time and chance of human errors*
*to control complicated and different types of jobs smoothly and effectively*
*to deliver physics data to users as soon as data-taking finishes*

# Belle II Computing Model



# Belle II Distributed Computing Structure



# Automatized Production System


KEK+PNNL(BNL)

**Production manager (human)**
define the project for:
- raw data process
- simulation
- user analysis

**Project management system**
create a project for:
- raw data process
- simulation
- user analysis

KEK

KEK

**Fabrication system**
control jobs for:
- raw data process
- simulation
- user analysis

**Distributed data management system (DDM)**
control the data management
- bulk replication
- bulk deletion

**Data quality system**
verify that outputs can be used in physics analysis
- feedback to production manager (human)

PNNL(BNL)

Nagoya+Niigata

KEK

**Verification system**
verify that tasks are correctly finished
- feedback to Fabrication & Distributed data management

**Monitoring system**
check the jobs/network status
- feedback to Fabrication & Distributed data management
- sending problematic status to GOCDB

*Different types of production*

*MC production (w/ or w/o BG)*
*Skim production*
*RAW data process*

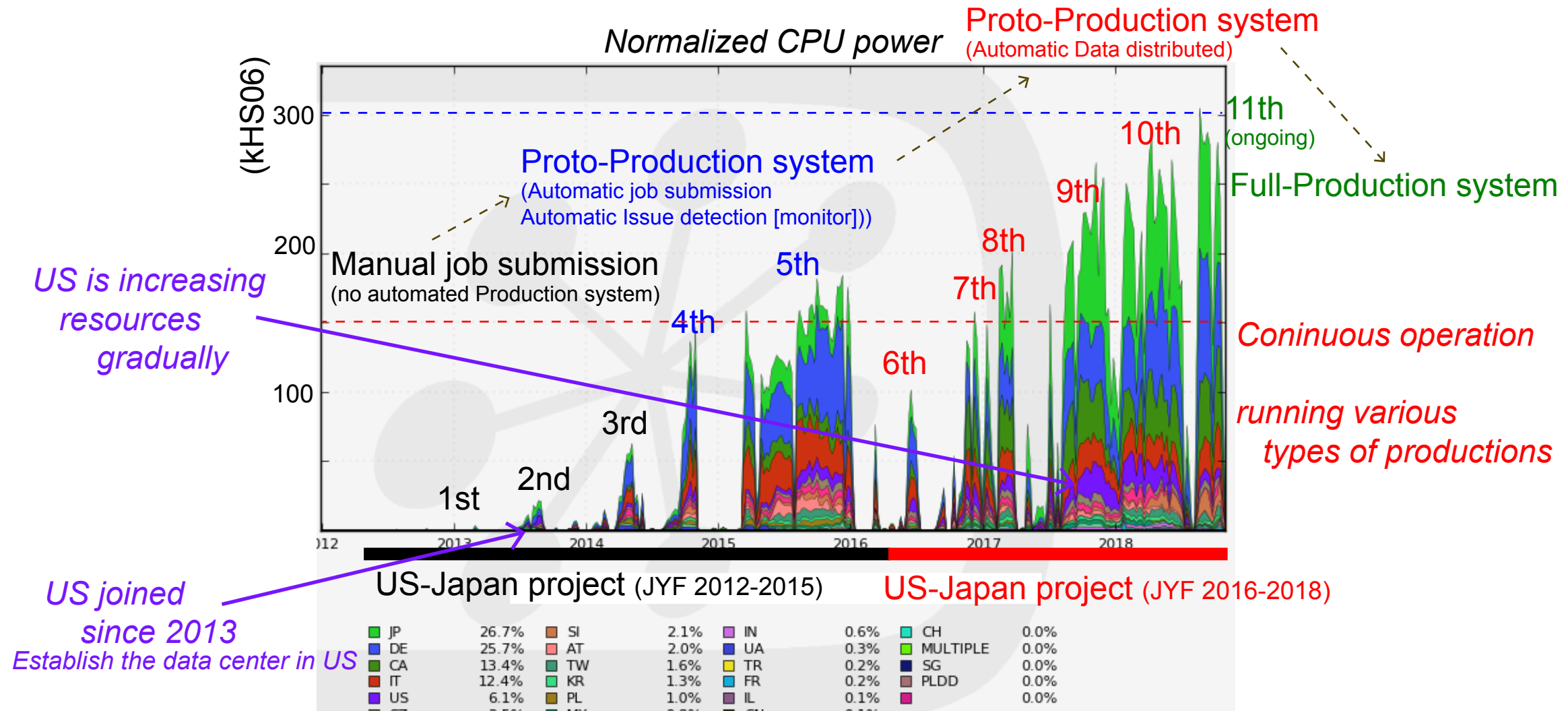
*Huge variety of modes*

*BB, udsc, signal, background*
*many physics skims*

*Complicated data management over world-distributed sites*

*Reduce human error and perform effective operation*

# Research Highlight : One page summary



JFY 2012 - 2015 Japan:KEK + US:PNNL Project Completed

Establishment of a remote data center for acceleration of Belle II data center

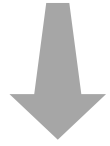
Belle IIデータ再プロセス高速化のためのリモート・データセンターの設立

JFY 2016 - 2018 Japan:KEK + US:PNNL→BNL(since 2018) Project Completed

Development of a scalable and automatized production system for the Belle II experiment (US side research title : 2016)

Automatized Production System for Belle II (resarch title was renamed : 2017 or later : funded by the Japan side only)

Belle II実験における拡張性を考慮した自動化プロダクション・システムの開発



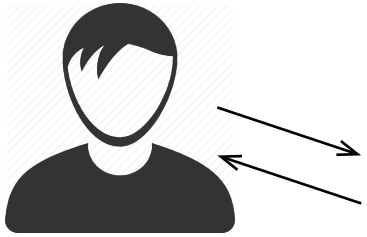
JFY 2019 (- 2020) Japan:KEK + US:PNNL Project Applied

Hiding Data Access Times in HEP Distributed Workflow

# MC production jobs

*"Jobs go to Data"*

↳ *requires storage at each site*




*MC job w/ BG*

*Data distributed before submitting jobs*

*BG files*


*computing site with Storage*

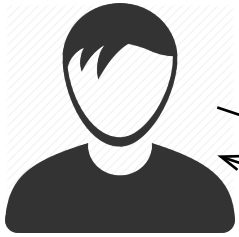
*Jobs access data on local storage*

*Issue : Inefficient use of compute resources without local storage*

# MC production jobs

"Jobs go to Data" ⟶ "Remote Data Access"

↳ enables conribution of compute resources
w/o local GRID storage




MC job
w/ BG

Storage-server-less
computing site

necessary BG files
from other sites

BG files


computing site with Storage

Issue : Time consumed in Remote Accesses

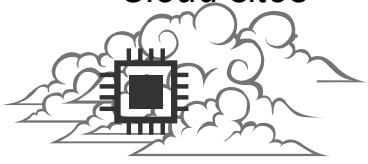
# Belle II computing sites

GRID sites

KEK, BNL, DESY, GridKA, KISTI, CNAF, many European sites

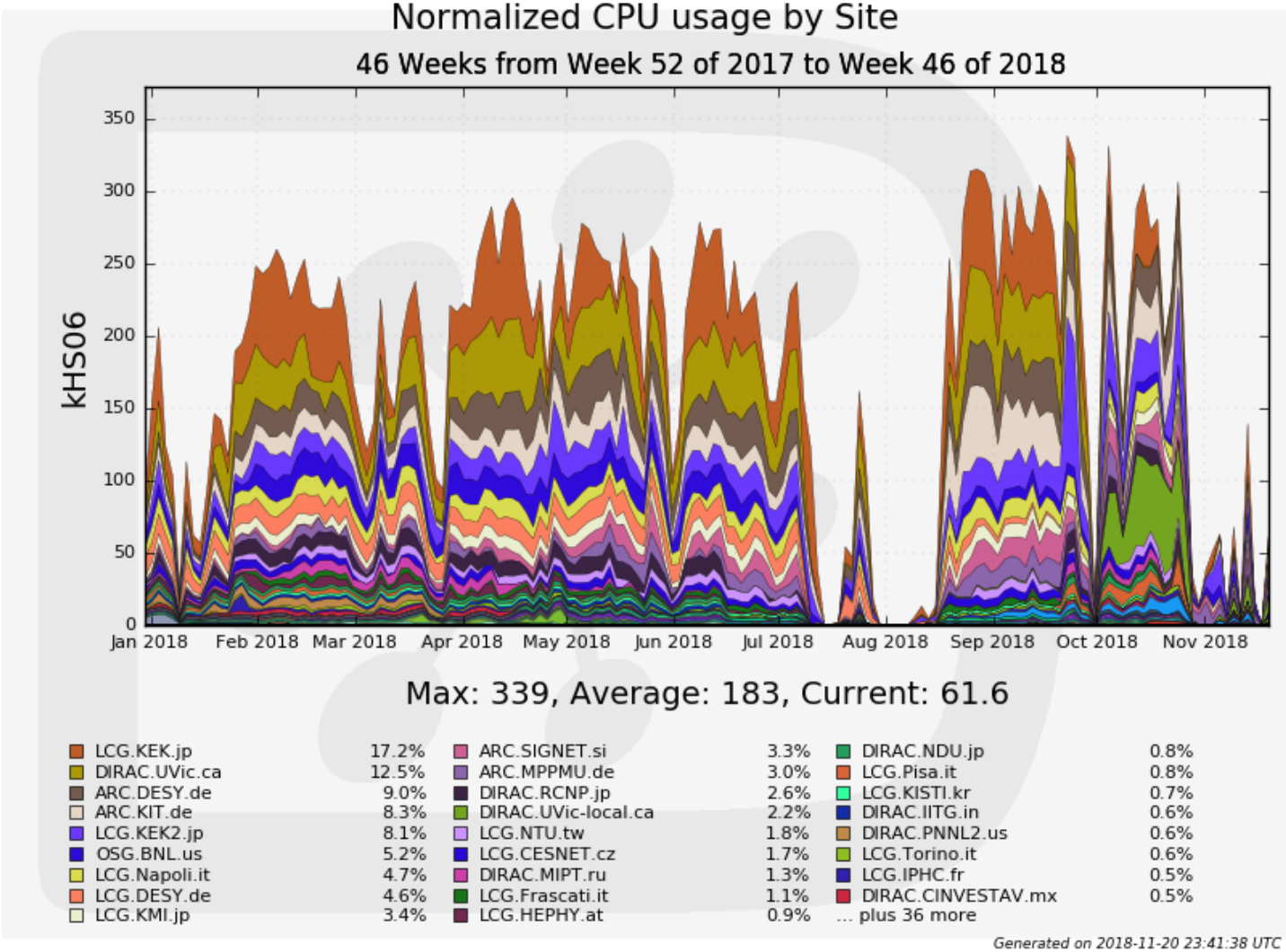
~30 sites : ~75%

Cloud sites

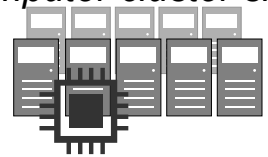


Univ. of Victoria, Univ. of Melbourne

several sites : ~15%




Computer cluster sites



Many Universities in Japan, Korea, India, China, Russia, Mexico,

~25 sites : ~10%

*Large contribution from Compute Resources w/o local GRID storage*

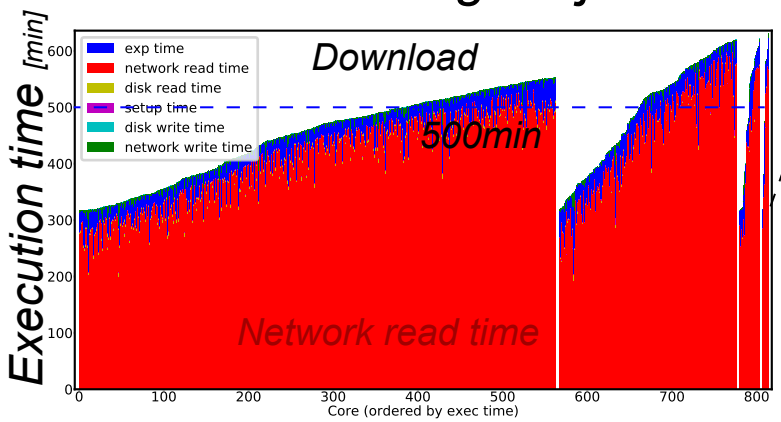
# Remote Data Access

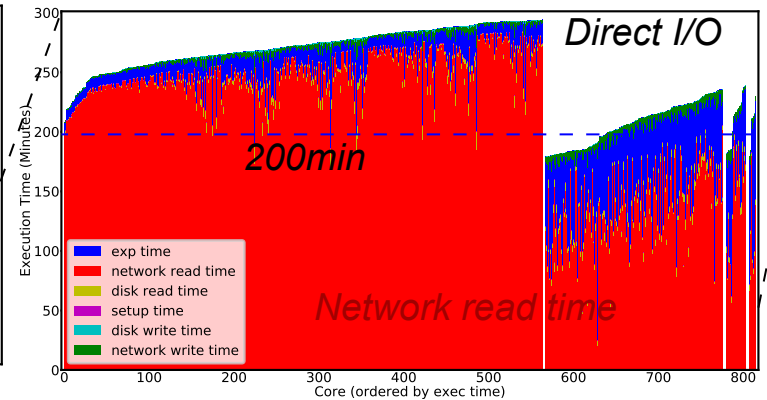
- *Download*
  - ⇨ *copying whole files unnecessarily*
  - ⇨ *CPU idle during download*

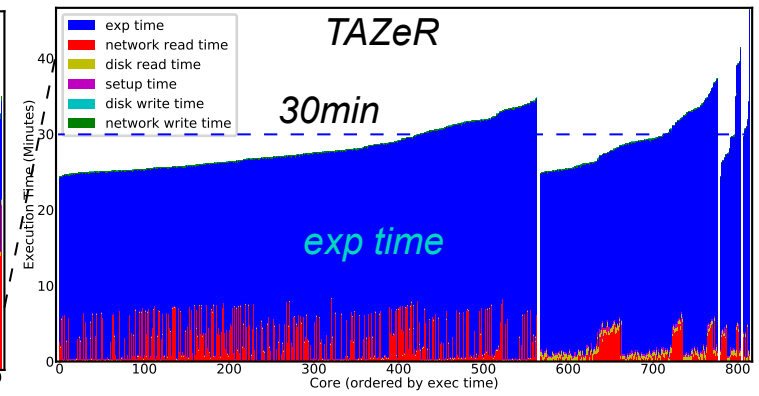
- *Direct I/O (e.g. xrootd)* → *existing remote I/O technique*
  - ⇨ *chaotic remote accesses can be inefficient*

- *Organized Streaming (TAZeR : Transparet Asynchronous Zero-copy Remote I/O)*
  - ✓ *I/O optimization with pre-fetching to memory*
  - ✓ *Intelligent job scedhuling*

*support from DOE as a part of "Integrated End-to-end Performance Prediction and Diagonosis"*



# Applying TAZeR to Belle II

*TAZeR*

✓ *Hiding netowkr and I/O latencies with*

- *I/O optimization*
- *Intelligent job scheduling*

*Belle II*

✓ *Efficient use of compute resources without local GRID storage*

*Proposed project : "Hiding Data Access Times in HEP Distributed Workflow"*

*To icrease throughput of Belle II Monte Carlo simulations*
*To identify the conditions under which TAZeR improves HEP workflow*