

Belle II Data Production

Jake Bennett

The University of Mississippi

Belle II Summer School - July 2019



THE UNIVERSITY of
MISSISSIPPI



Belle II Data Production

- **Primary goal: ensure smooth and timely production of data and MC samples for physics analysis and other studies**
- **Basic philosophy: take a global view of data flow to identify missing or problematic functionality and performance**
 - Close collaboration with computing, software, physics groups, etc.
 - Contributions to software validation, data quality monitoring, distributed computing, etc.
 - Allow flexibility for changing conditions
- **Primary tasks:**
 - Calibration and alignment (<https://confluence.desy.de/display/BI/Data+Production+Calibration>)
 - RAW data (re)processing (<https://confluence.desy.de/display/BI/Phase+3+data>)
 - MC production (<https://confluence.desy.de/display/BI/Data+Production+MC12>)
 - Analysis skimming (<https://confluence.desy.de/display/BI/Skimming+Homepage>)
- **Other tasks**
 - Prepare beam background samples (<https://confluence.desy.de/display/BI/Beam+background+samples>)
 - Quality assurance monitoring (<https://mirabelle.belle2.org/home/display/>)
 - Software validation (<https://confluence.desy.de/display/BI/Data+Production+Analysis+Validation>)

Coordinator: @Jake Bennett

Skim manager: @Racha Cheaib

Background manager: @Marko Staric

Calibration software manager: @David Dossett (sw liaison)

Calibration manager: @Umberto Tamponi

Calibration skim manager: @Trabelsi Karim

Validation manager: @Sviatoslav Bilokin

Data processing manager: @Stefano Lacaprara

MC processing manager: @Alessandro Gaz

Data production liaisons

(responsibilities of the data production liaisons can be found [here](#))

Group	Liaison
Semileptonic & Missing Energy Decay	@Mario Merola
Radiative & Electroweak Penguin	@Saurabh Sandilya
Time Dependent CP Violation	@Sviatoslav Bilokin
Hadronic B to Charmless	@Ilya Komarov
Hadronic B to Charm	@Nibedita Dash
Bottomonium	@Suxian Li
Charmonium	@Sen Jia , @Yubo Li
Charm	@Yeqi Chen
Tau & Low Multiplicity	@Tomoyuki Konno

First things first

- The [data production page](#) is a one-stop shop for all of the data processing and MC production details for Belle II

Find details about the data

Links to DP meetings

Details about calibration work

Availability of official MC

and much more!

Confluence

Spaces

People

Create

...

Search

?

9+

Data production WebHome

Phase 3 data

Experiment 3

Experiment 5+6 - full dress re

Data Production Meetings

Data Production Leadership

Data Production Liaisons

Data Production Shifts

Data Production Calibration

Data Production Global Cosmi

Data Production Global Cosmi

Data Production MC7

Data Production MC8

Data Production MC9

Data Production MC10

Data Production MC11

Data Production MC12

Data Production Rehearsal

Data Production Schedule

Instructions for gbasf2 analysi

Skimming Homepage

Data Production Analysis Valic

Production validation

Dashboard / Belle II / Belle II Internal WebHome

Edit

Save for later

Watching

Share

...

Data production WebHome

Jake Bennett posted on 20. Sep. 2016 03:43h - last edited by Jake Bennett on 09. Jun. 2019 21:58h

Coordinator: @Jake Bennett

Skim manager: @Racha Cheaib

Background manager: @Marko Staric

Calibration software manager: @David Dossett (sw liaison)

Calibration manager: @Umberto Tamponi

HLT skim manager: @Trabelsi Karim

Validation manager: @Sviatoslav Bilokin

Data processing manager: @Stefano Lacaprara

MC processing manager: @Alessandro Gaz

DP leadership responsibilities are listed here.

Mailing list: dataprod@belle2.org

Introduction

The data production scheme for Belle II requires significant coordination of tasks within a well defined time frame. The role of the data

Data production liaisons

(responsibilities of the data production liaisons can be found here)

Group	Liaison
Semileptonic & Missing Energy Decay	@Mario Merola
Radiative & Electroweak Penguin	@Borys Knysh
Time Dependent CP Violation	@Sviatoslav Bilokin
Hadronic B to Charmless	@Ilya Komarov
Hadronic B to Charm	@Nibedita Dash
Bottomonium	@Suxian Li
Charmonium	@Sen Jia , @Yubo Li
Charm	@Yeqi Chen
Tau & Low Multiplicity	@Tomoyuki Konno

Contact details for DP leadership

Physics working group liaisons (collect MC production requests, perform validation tasks, etc.)

<https://confluence.desy.de/display/BI/Data+Production+WebHome>

First things first

- The [data production page](#) is a one-stop shop for all of the data processing and MC production details for Belle II

Many more useful links!

Bi-weekly group meetings
(detailed minutes on each page)

Weekly announcement
(register on dataprod@belle2.org)

Requests/concerns welcome
(contact dataprod@belle2.org
or create a [JIRA ticket](#))

MC productions

- [MC12: January 2019](#)
- [MC11: August 2018-January 2019](#)
- [MC10: Jan-July 2018](#)
 - [Data Production Shifts](#)
- [MC9: July-Dec 2017](#)
 - [Shift calendar](#)
- [MC8: Feb-June 2017](#)
 - [Shift calendar](#)
 - [Samples for analysis use](#)
- [MC7: Nov 2016-Jan 2017](#)
 - [Shift calendar](#)
 - [Samples for analysis use](#)
- [Previous MC Campaigns](#)

Shift Manuals

- [Standard MC shift manual](#)
- [Expert MC shift manual](#)

Skims

- [Skimming Homepage](#)

Data

Phase 3 data:

- Experiment 7+: [status](#), run summary for [exp7](#), [exp8](#)
- Experiment 5-6 (Full dress rehearsal): [status](#)

Phase 2 data:

- Experiment 3 (Phase 2 collisions): [status](#), [run summary](#), [calibration](#), [skims](#)

Other details

- [Experiment numbering](#), [Description of the Run Summary](#)
- [HLT Skims](#)
- [Use of Releases](#)

Validation

- ["Physics modes" Validation](#)
- [Monitoring \(MiraBelle\)](#)

Calibration

- [Prompt Calibration](#)
- Preparation of calibration constants for each reprocessing: [meetings](#), [indico](#) (used until end of May 2018)

First things first

- The [data production page](#) is a one-stop shop for all of the data processing and MC production details for Belle II

Phase 3 data

Jake Bennett posted on 11. Mar. 2019 18:44h - last edited by Jake Bennett on 10. Jun. 2019 00:40h

- [JIRA tickets to collect feedback](#)
- [Official Processing](#)
 - [Processing scheme](#)
 - [Official reprocessing details](#)
 - [Prompt processing JIRA ticket](#)
 - [Prompt reprocessing information sheet](#)
 - [RAW data availability](#)
- [Unofficial processing](#)
 - [PromptComicProc2 Unofficial COSMIC prompt processing of exp 7 with 31X](#)
 - [PromptProc2 Unofficial PHYSICS prompt processing of exp 7 with 31X](#)
 - [PromptProc0 OBSOLETE Unofficial prompt processing of exp 7](#)
- [Run information](#)

This page will be populated with phase 3 sample availability and other related details.

JIRA tickets to collect feedback

If you find an issue in the reprocessed data or want to inform the DP group about a bad run, please fill a

- Experiment 8: [BIIDP-1499](#) - Phase 3, experiment 8 feedback and processing [OPEN](#)
- Experiment 7: [BIIDP-1345](#) - Phase 3, experiment 7 feedback and processing [OPEN](#)

Official Processing

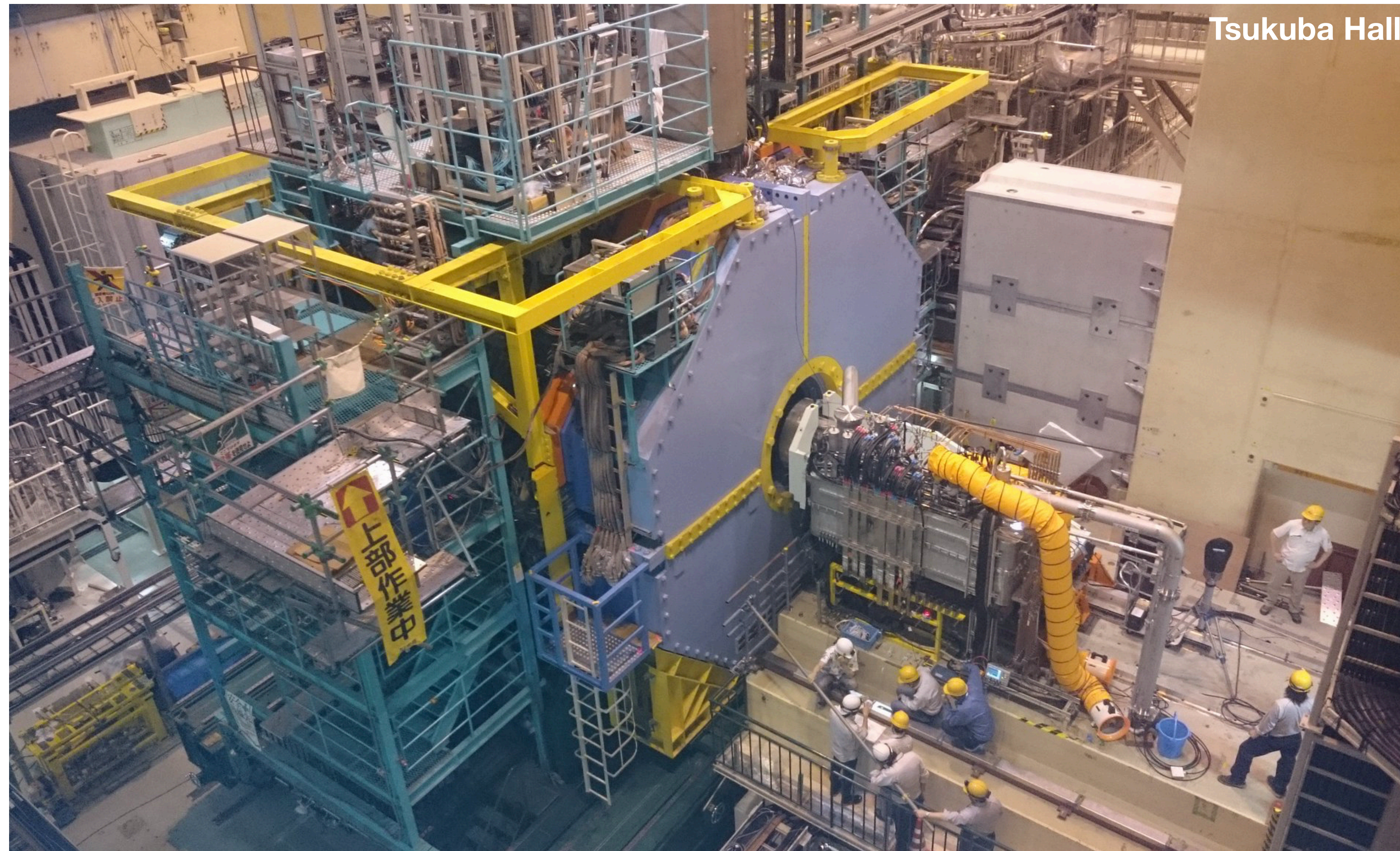
Processing scheme

Official data can only be processed when calibration is completed. For early phase 3, the data processing sc

1. Collect data until an amount sufficient for calibration is acquired (typically ~1 week for early running)
2. Define a "run bucket" that includes similar data of sufficient quantity for calibration
3. Perform calibrations (tracking and alignment first, then outer detectors, etc) - this step takes ~2-3 we
4. Collect calibration constants into an official data_reprocessing global tag
5. Process the data contained in the run bucket of interest

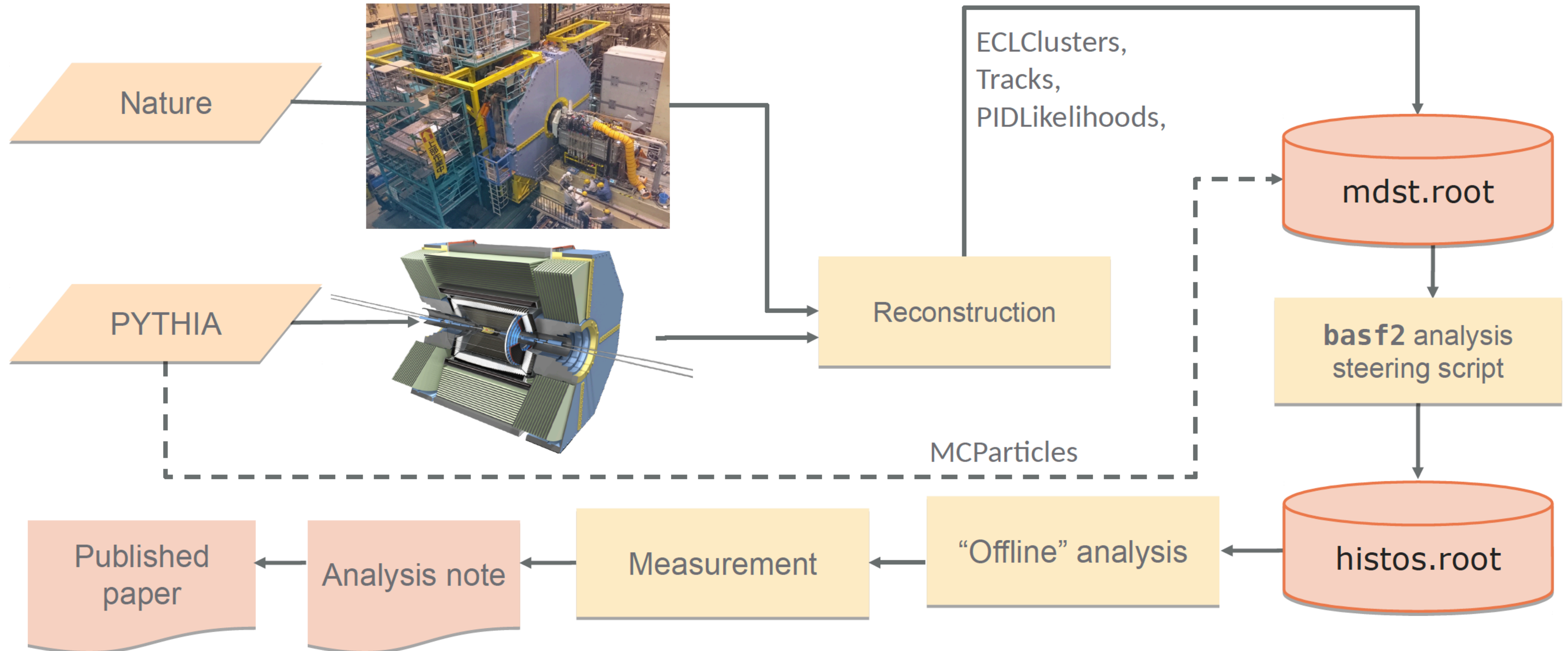
Processing Label	When	Exp #	Run Numbers	Integrated Luminosity	Magnetic Field Payload Rev	Global Tag (ID)	Release	Status	JIRA Ticket
proc9	June-July	3, 7, 8	Exp 3: 4S: 529-5613 Exp 7: 4S: 909-4120 Exp 8: 4S: 43-1022, 1036-1554 Continuum: 1703-1835 E scan: 1025-1031	Offline: Exp3: 4S: 509.9 ± 0.3 ± 3.6 pb ⁻¹ Exp7: N/A Exp8: N/A Online: Exp3: 4S: 0.504/fb Exp7: 4S: 0.689/fb Exp8: 4S: 1.949 /fb Continuum: 0.831 /fb E scan: 0.038 /fb	ON: (rev 8)	data_reprocessing_proc9 (ID = 654)	03-02-02	IN PROGRESS	BIIDP-1580 - Processing 9 - exp 3, run 509, all and exp 8, run 1835 OPEN
Reprocessed data paths: ALL (unskimmed data): DONE Exp3: /group/belle2/dataproduct/Data/release-03-02-02/DB00000654/proc9/e0003/4S/r<RUN>/all/mdst/sub00/ Exp7: /group/belle2/dataproduct/Data/release-03-02-02/DB00000654/proc9/e0007/4S/r<RUN>/all/mdst/sub00/ Exp8: /group/belle2/dataproduct/Data/release-03-02-02/DB00000654/proc9/e0008/<4S/Continuum/Scan>/r<RUN>/all/mdst/sub00/ SKIMS (details available are at HLT Skims) -) HLT SKIM = {hlt_mumu_2trk, hlt_hadron, hlt_bhabha, hlt_gamma_gamma}, TYPE = {cdst, mdst} DONE Exp3: /group/belle2/dataproduct/Data/release-03-02-02/DB00000654/proc9/e0003/4S/r<RUN>/skim/<HLT_SKIM>/<TYPE>/sub00/ Exp7: /group/belle2/dataproduct/Data/release-03-02-02/DB00000654/proc9/e0007/4S/r<RUN>/skim/<HLT_SKIM>/<TYPE>/sub00/ Exp8: /group/belle2/dataproduct/Data/release-03-02-02/DB00000654/proc9/e0008/4S/r<RUN>/skim/<HLT_SKIM>/<TYPE>/sub00/									

The big picture



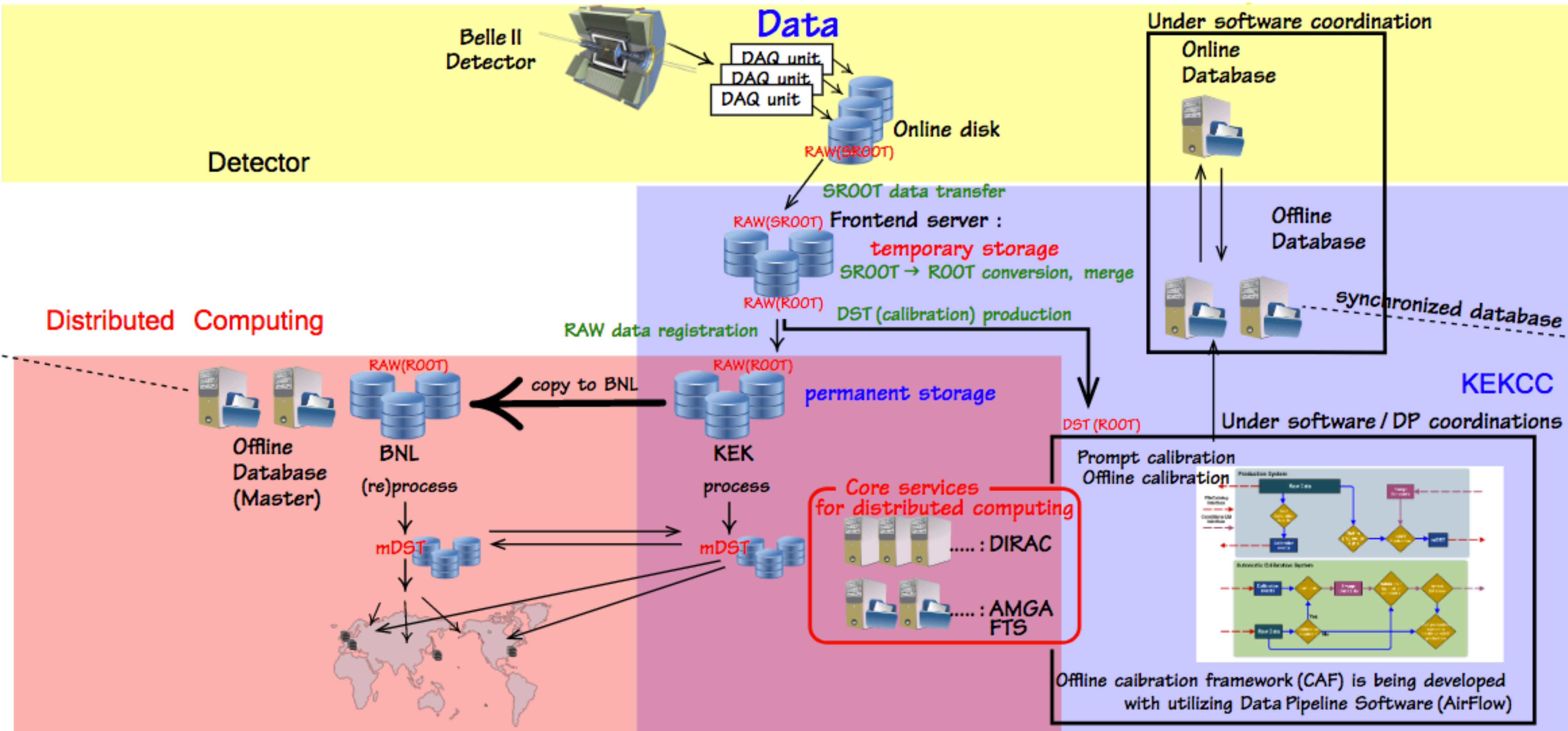
*shamelessly stolen borrowed from Sam Cunliffe's talk "Introduction to the analysis package" - Belle II SKW, 15.06.2018

The big picture



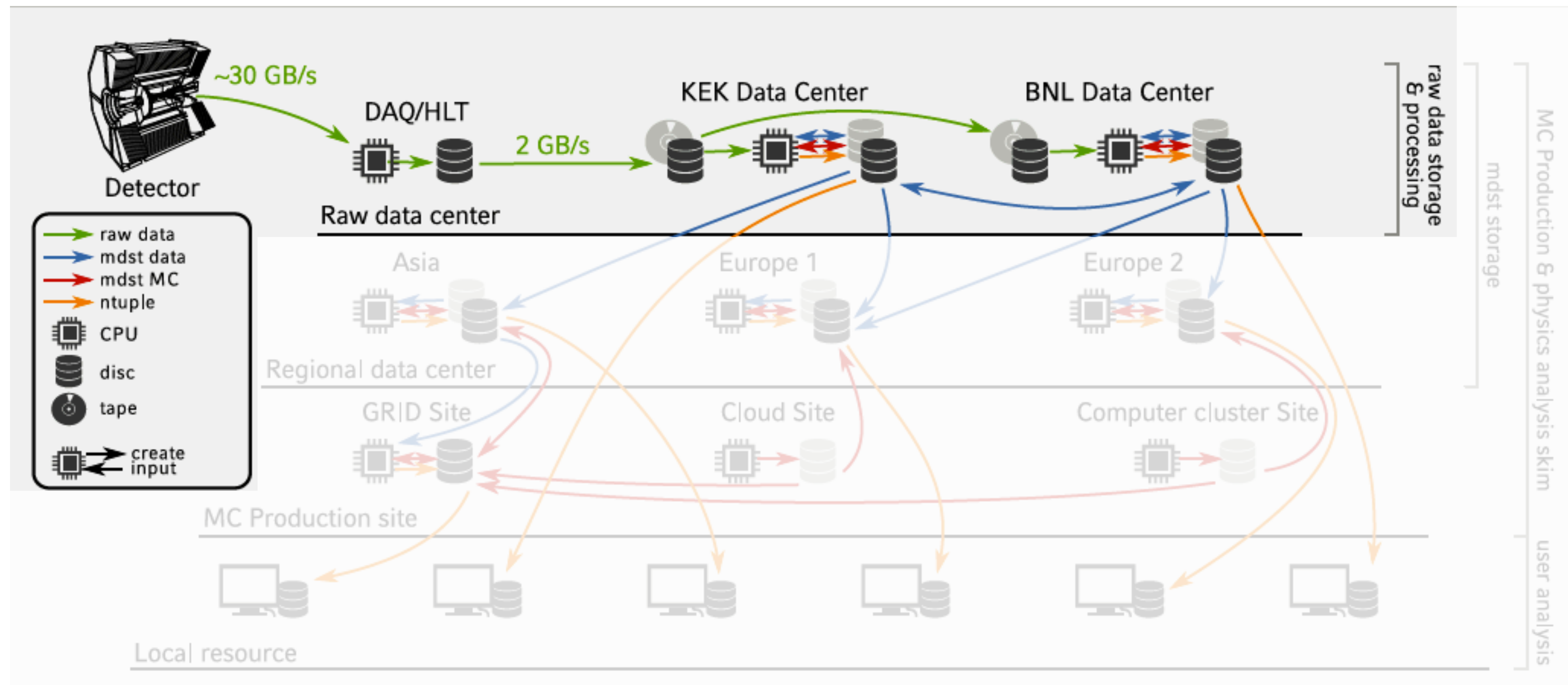
*shamelessly stolen borrowed from Sam Cunliffe's talk "Introduction to the analysis package" - Belle II SKW, 15.06.2018

Belle II data flow



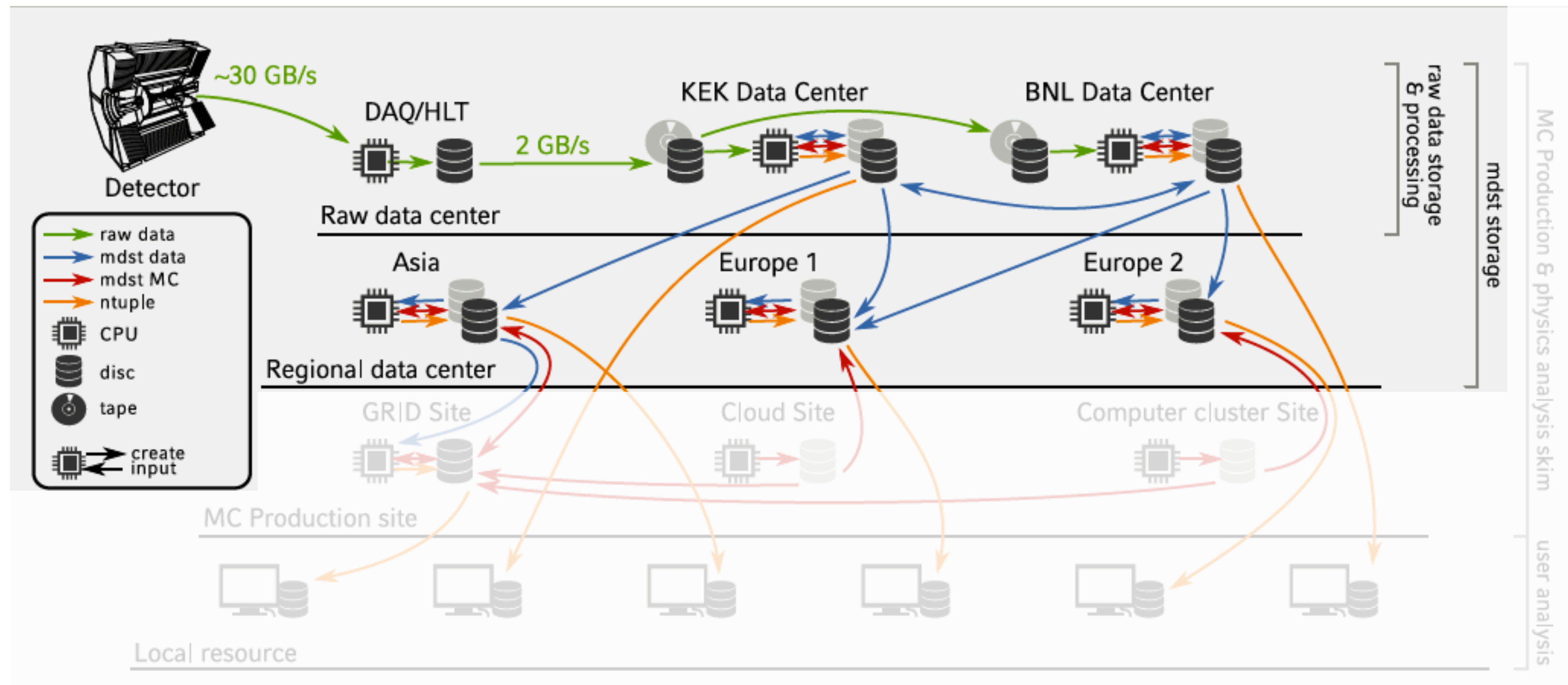
Belle II data flow

- Raw data storage and reprocessing at KEKCC and BNL - Raw data centers
- mDST storage on GRID storage elements (SE) - Regional data centers
- Skimming and analysis on GRID computing elements (CE) - MC production sites
- nTuple analysis on local resources - Local resources



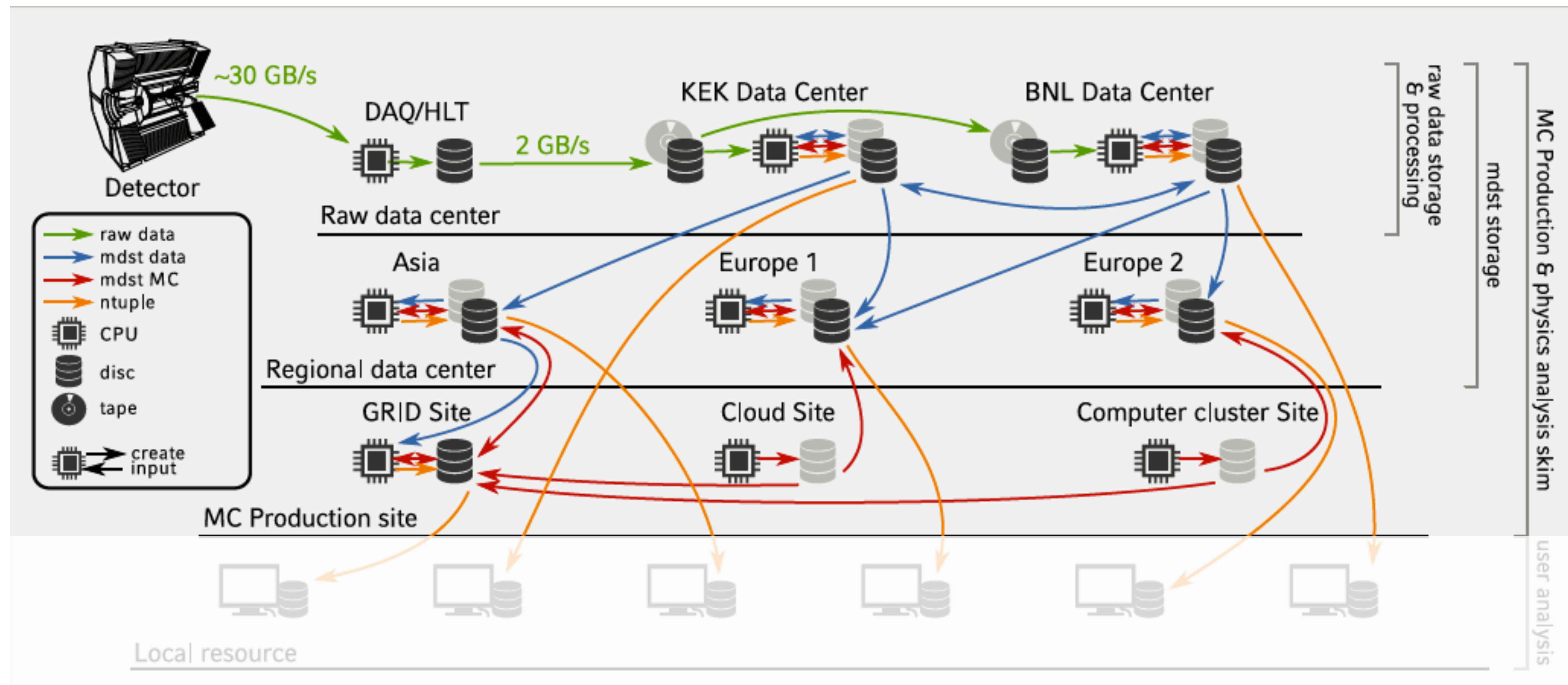
Belle II data flow

- Raw data storage and reprocessing at KEKCC and BNL - Raw data centers
- mDST storage on GRID storage elements (SE) - Regional data centers
- Skimming and analysis on GRID computing elements (CE) - MC production sites
- nTuple analysis on local resources - Local resources



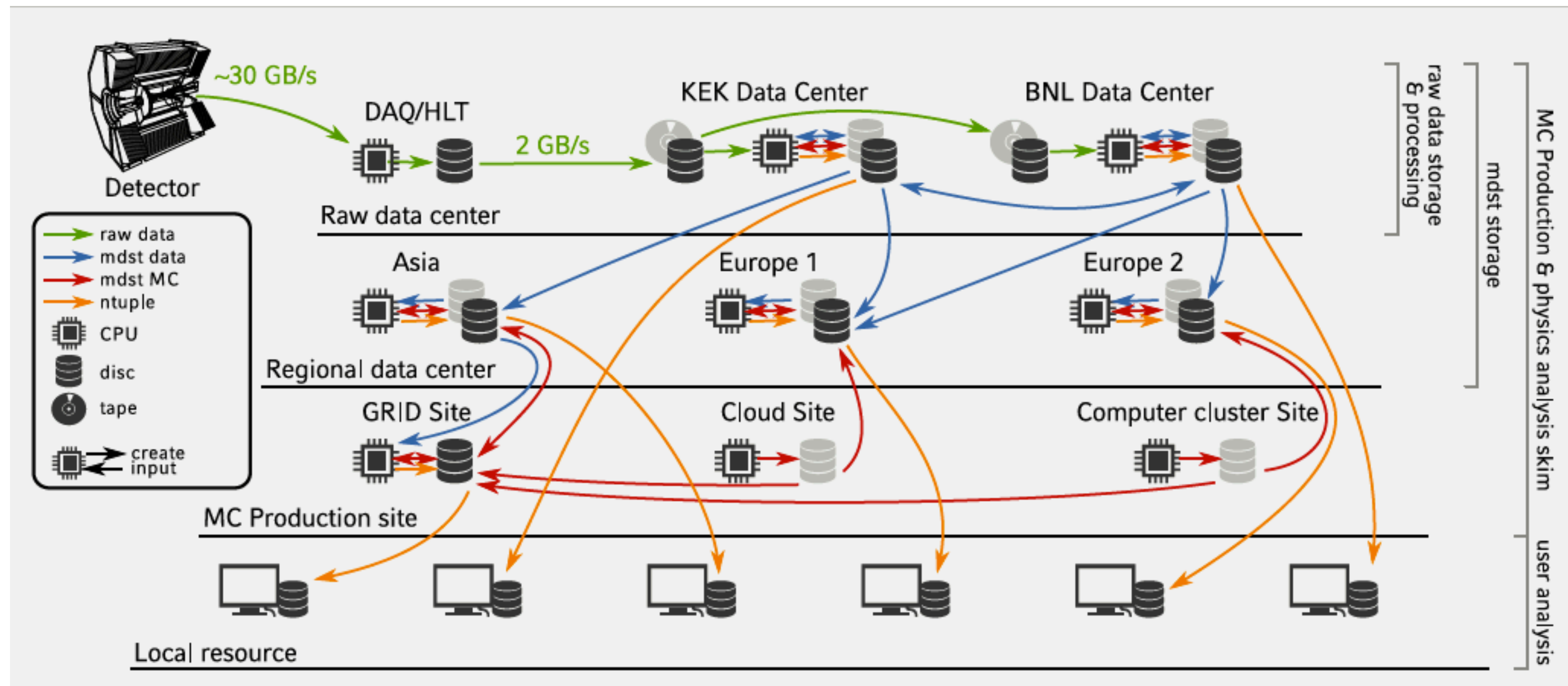
Belle II data flow

- Raw data storage and reprocessing at KEKCC and BNL - Raw data centers
- mDST storage on GRID storage elements (SE) - Regional data centers
- Skimming and analysis on GRID computing elements (CE) - MC production sites
- nTuple analysis on local resources - Local resources



Belle II data flow

- Raw data storage and reprocessing at KEKCC and BNL - Raw data centers
- mDST storage on GRID storage elements (SE) - Regional data centers
- Skimming and analysis on GRID computing elements (CE) - MC production sites
- nTuple analysis on local resources - Local resources



Brief interlude: file types

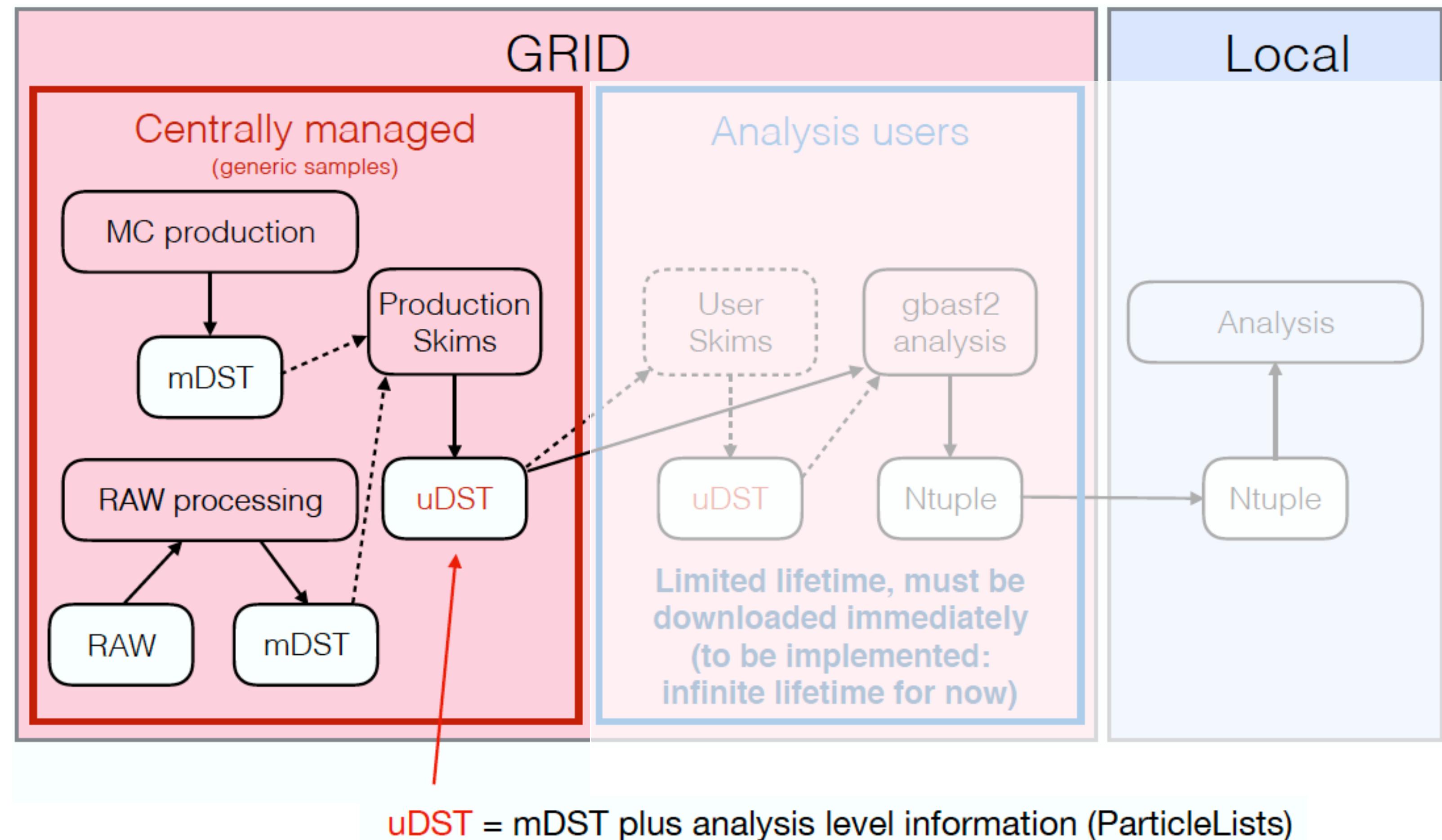
- In general, Belle II output is stored in ROOT files containing various subsets of dataobjects, dbobjects, nTuples, etc
- **RAW**: raw data containing detector information
- **DST**: data summary table
 - all available dataobjects (from reprocessing) are included
 - not generally produced or used as it contains everything (and is huge)
- **mDST***: [mini data summary table](#)
 - strictly controlled version of a DST file
 - only a subset of available processed dataobjects are included
 - sufficient information for analysis use
- **cDST**: [calibration data summary table](#)
 - mDST objects, plus additional dataobjects useful for calibration
- **uDST**: user data summary table
 - mDST objects, plus analysis objects (ParticleLists)
 - produced from skims - reduce time needed to run analysis jobs
 - **these are the samples you should be using for analysis!**



***mDST Storage: ~61 MB per additional kB**

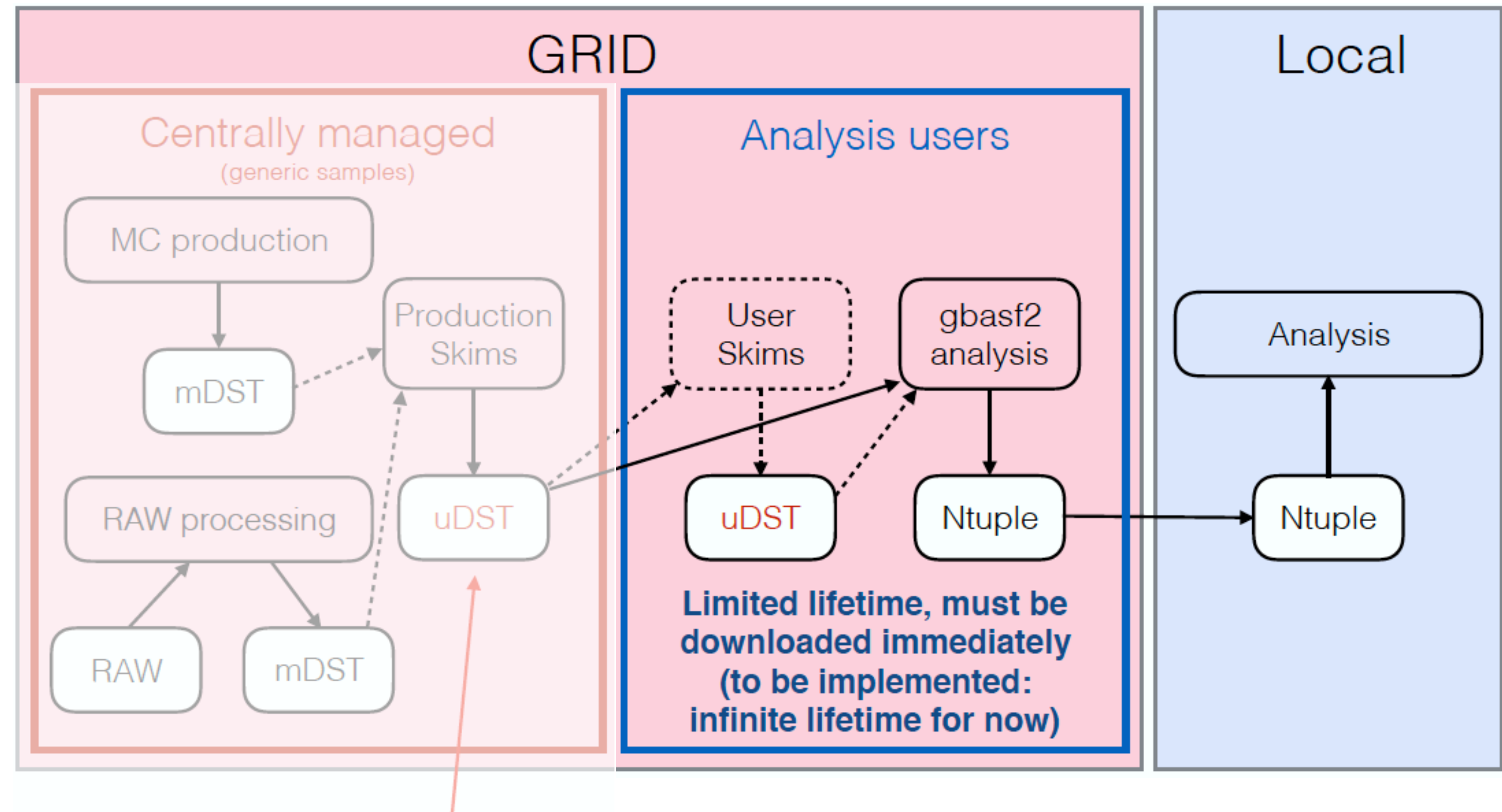
Data production/analysis scheme

- Simulation and processing for official data and MC samples is performed centrally
 - This is the primary purpose of the data production group
 - Though you are certainly free to produce test samples on your own - *don't use them for publishable physics analyses!*
 - Requests for MC samples can be made to the DP liaison in your physics working group
 - Skimming is also centrally performed - each physics working group has a skim liaison as well



Data production/analysis scheme

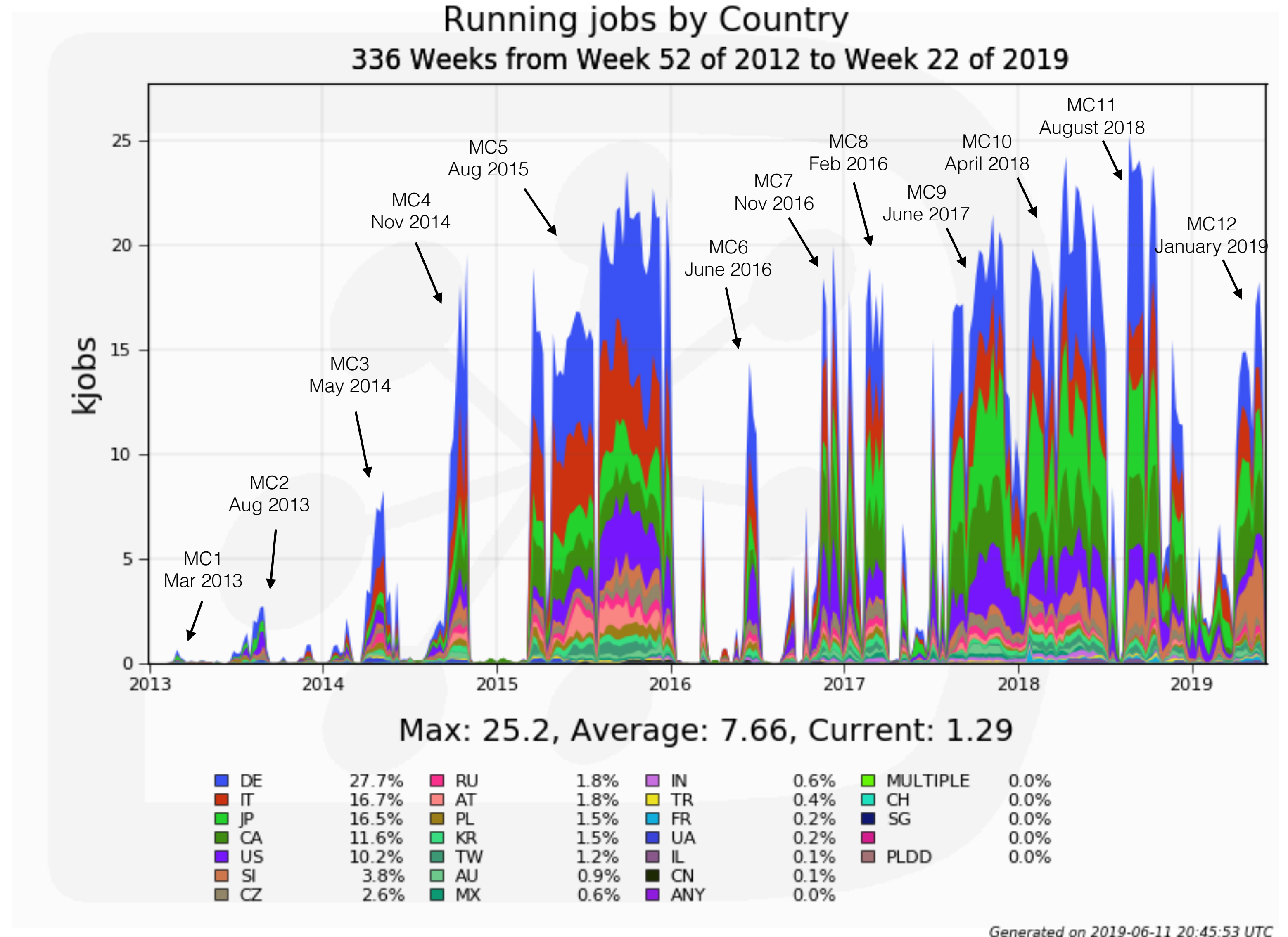
- The default place to run your analysis jobs is on the grid
 - After official skimming, uDST files are available for grid-based analysis
 - Users can either perform another skim on the official skim samples or run an analysis directly on the uDSTs
- Transfer your ntuple files to local resources for the rest of your analysis
 - Fitting, plotting, etc.
 - Files on the grid (will) have a limited lifetime, so ntuple files must be downloaded to local resources as quickly as possible!



uDST = mDST plus analysis level information (ParticleLists)

Official MC production

- Official MC production is done in “campaigns”
 - Roughly corresponding to a particular software release/GT
 - We are currently at the tail end of the twelfth official MC campaign
<https://confluence.desy.de/display/BI/Data+Production+MC12>
- Output files available only on the grid
 - Get your certificates ready!
 - Limited samples on kekcc for fast availability for FEI training, etc.



Twelfth official Belle II MC campaign

Data Production MC12

Jake Bennett posted on 18. Jan. 2019 21:19h – last edited by Alessandro Gaz on 10. Jun. 2019 10:53h

MC12 samples

VERY IMPORTANT: Always check the LPN of your files using 'gb2_ds_list' before submitting jobs to make sure the files exist where you expect them!

** The LPN for MC12 samples begins with '/belle/MC/release-03-00-00/DB00000487/MC12/' unless otherwise noted

*** For productions with both BGx1 and BGx0, the LPN is given for the BGx1 sample. The LPN for the BGx0 sample is identical except for the production ID.

Priority is noted as follows: 0 - trivial, 1 - minor, 2 - major, 3 - critical, 4 - blocker

Note that if you are working with files at KEKCC, the ghi disk system (anything starting /ghi or /hsm) will automatically purge files from the disk and move them to tape when there is not enough space available. You can check the files with ghils (<https://kekcc.kek.jp/service/kekcc/html/Eng/HSM20System.html#s49cb0a0>) and request that they be staged with hstage (<https://kekcc.kek.jp/service/kekcc/html/Eng/HSM20System.html#d100fc92>).

Key	Scripts committed, ready for production	Jobs submitted	Released for analysis use	BGx0 ready, BGx1 processing	Not produced or removed
-----	---	----------------	---------------------------	-----------------------------	-------------------------

- [Introduction](#)
- [MC12 samples](#)
 - [Phase III Y\(4S\) generic](#)
 - [Early phase 3 geometry](#)
 - [Early phase 3 geometry \(500 fb-1\)](#)
 - [Block1:](#)
 - [Block2:](#)
 - [Block3:](#)
 - [Block4:](#)
 - [Block5:](#)
 - [Phase III Y\(4S\) signal](#)
 - [DST files for LID performance](#)
 - [Tau and Low Multiplicity Decays \(Tomoyuki Konno \)](#)
 - [Semileptonic and Missing Energy Decays \(Mario Merola\)](#)
 - [Hadronic B to Charm \(Nibedita Dash \)](#)
 - [Hadronic B to Charmless \(Ilya Komarov \)](#)
 - [Time Dependent CP Violation \(Sviatoslav Bilokin\)](#)
 - [Radiative and Electroweak penguins \(Borys Knysh \)](#)
 - [Charm \(Yeqi Chen \)](#)

Semileptonic and Missing Energy Decays ([@Mario Merola](#))

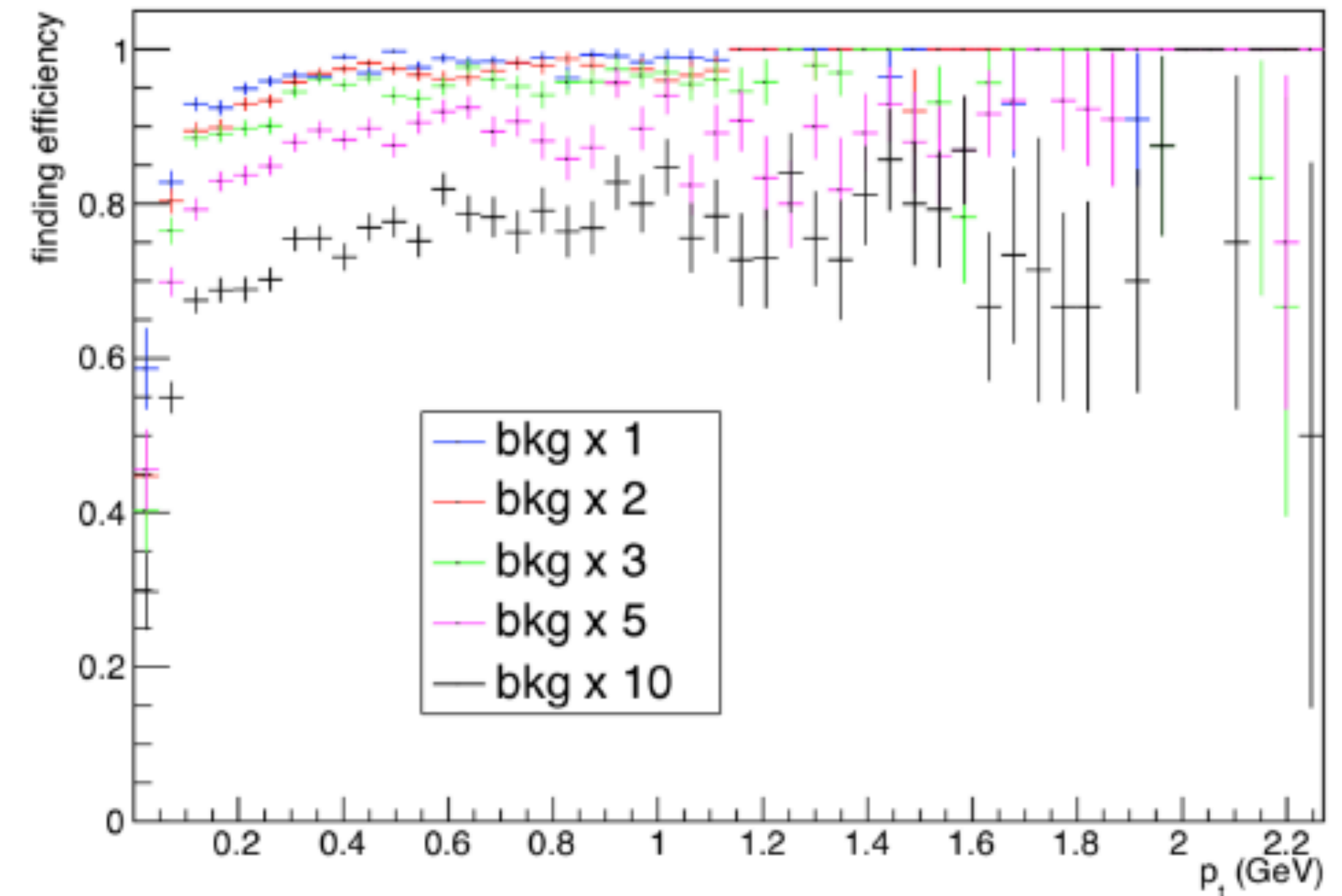
Early Phase 3 Y(4S) nominal with existing backgrounds

Priority	Signal	Event Type	Number of Events (10^6)	Ratio without/with background	Production ID without/with background	LPN***	Link to json file	JIRA ticket	Estimated size (GB)
2	Bp_DstTauNu	1264340000	50	0.2/0.8	7088/7207	/belle/MC/release-03-01-00/DB00000547/MC12b/prod00007207/s00/e1003/4S/r000000/1264340000/mdst/sub00		<input checked="" type="checkbox"/> BIIDP-1173 - SL & L request for MC12 <div>RESOLVED</div>	550
2	BO_XcTauNu	1193300007	100	0.2/0.8	7089/7208	/belle/MC/release-03-01-00/DB00000547/MC12b/prod00007208/s00/e1003/4S/r000000/1193300007/mdst/sub00		<input checked="" type="checkbox"/> BIIDP-1174 - SL & L request for MC12 <div>RESOLVED</div>	1100
2	Bp_XcElINu	1293710005	50	0.2/0.8	7094/7213	/belle/MC/release-03-01-00/DB00000547/MC12b/prod00007213/s00/e1003/4S/r000000/1293710005/mdst/sub00		<input checked="" type="checkbox"/> BIIDP-1179 - SL & L request for MC12 <div>RESOLVED</div>	550

More details available in the associated JIRA tickets

Run-dependent MC production

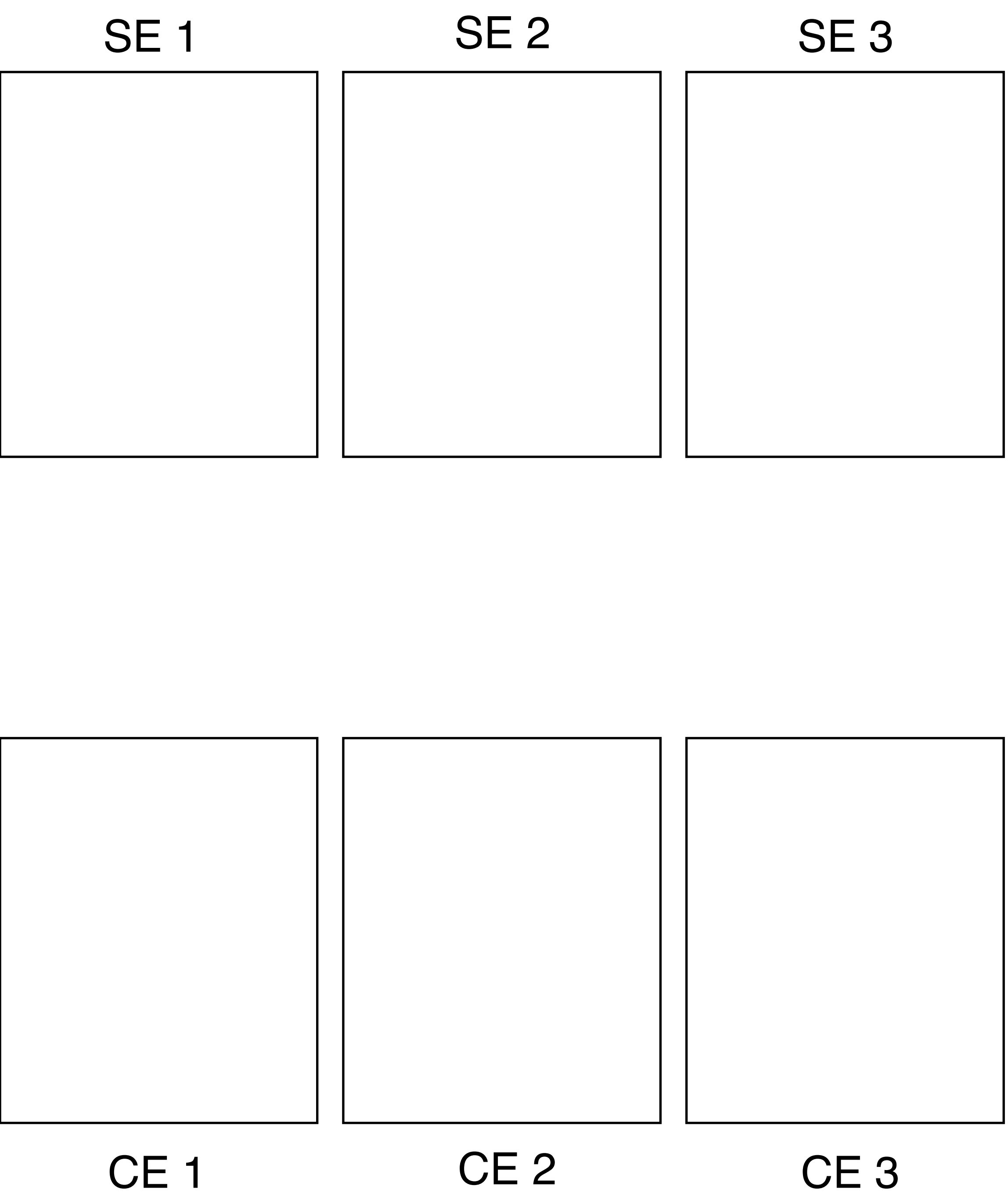
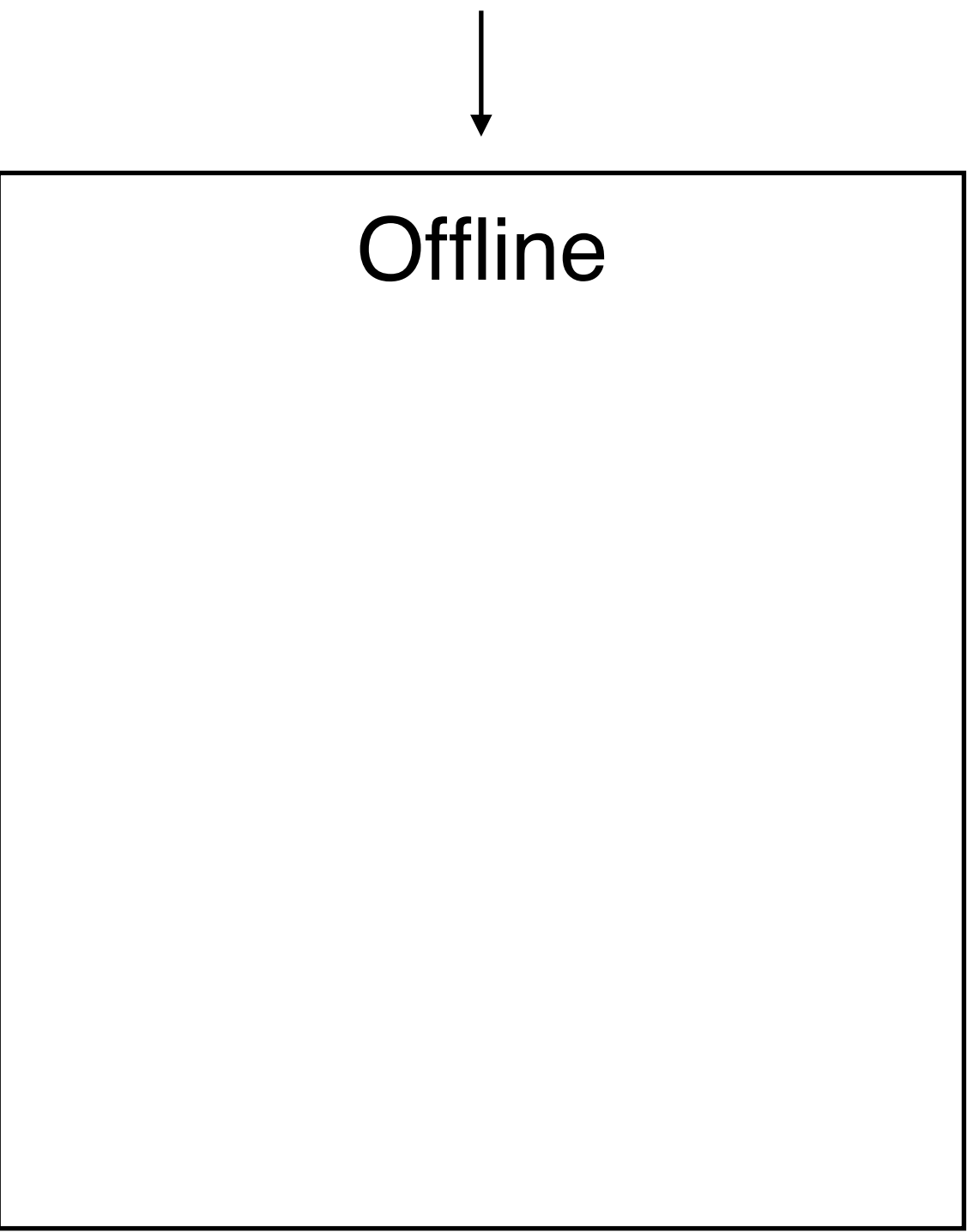
- Precision analyses require good correspondence between data and MC
- So far the official MC production has assumed ideal (and constant) detector performance, with backgrounds from simulation
- As we have now started taking data, we want the simulation to:
 - **Simulate realistic detector performance**
(variations over time, improvements in calibrations, etc.)
 - **Include realistic backgrounds**
(depend on collimator settings, beam conditions, luminosity, etc.)
- Beam backgrounds:
 - **can have a big effect on physics analyses**
(tracking, ECL resolution, PID, etc)
 - can **vary by run** due to luminosity dependence and changing conditions (**even within runs**)
- Random trigger beam background events should be used
 - Produce BG overlay files from random triggers - distribute on the grid
 - **Run dependent MC sample production in initial stages at kekcc**
 - <https://stash.desy.de/projects/B2P/repos/mc/pull-requests/128/overview>
 - Problem: background files are too big to fit on every grid storage element!

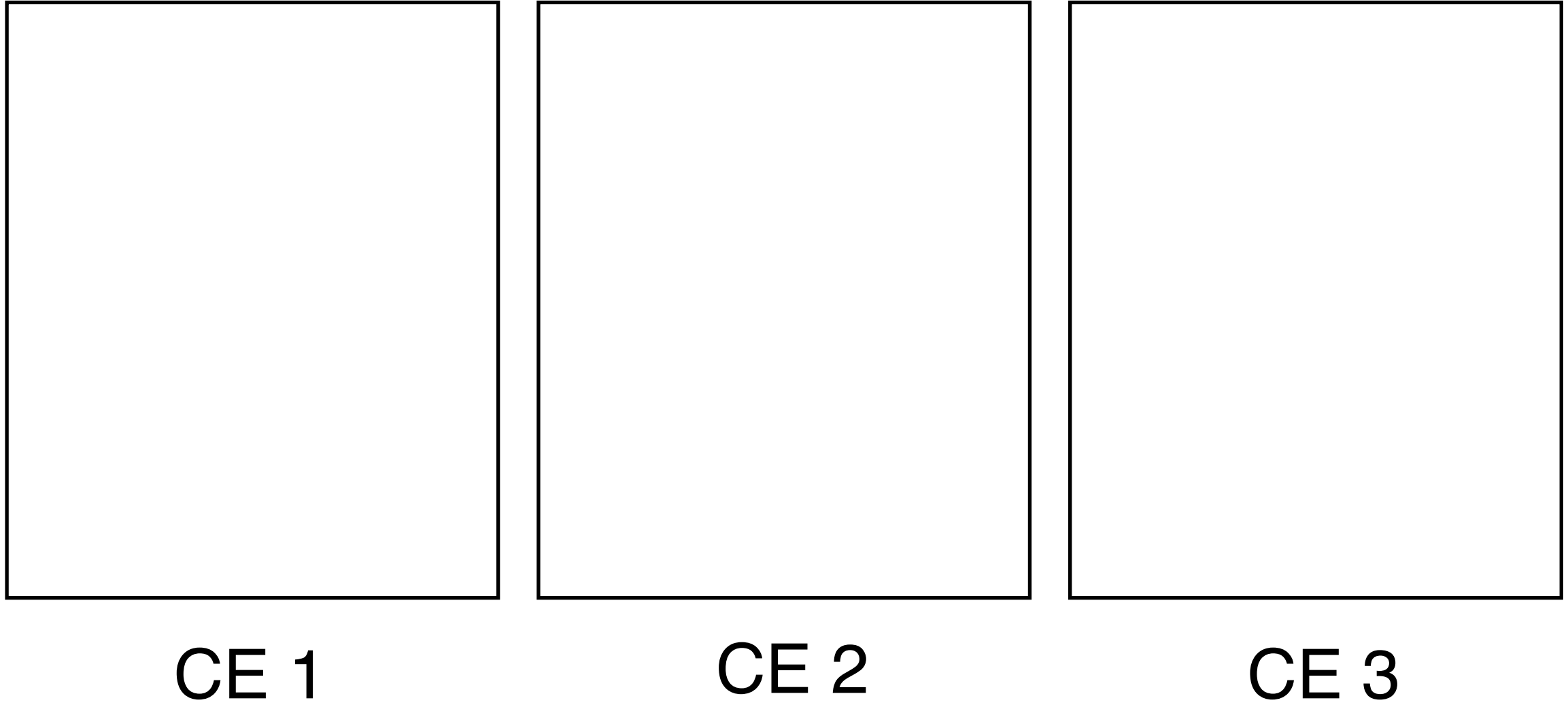
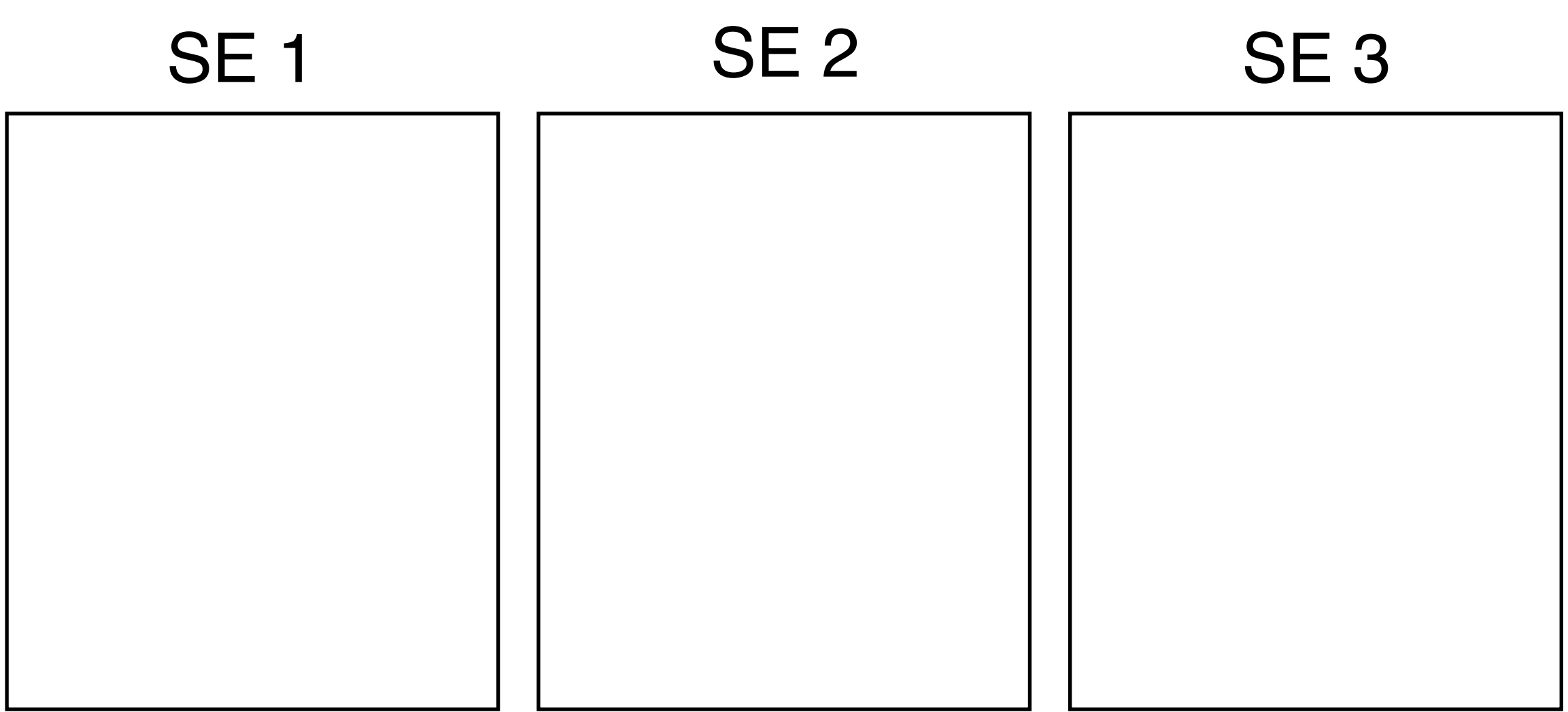
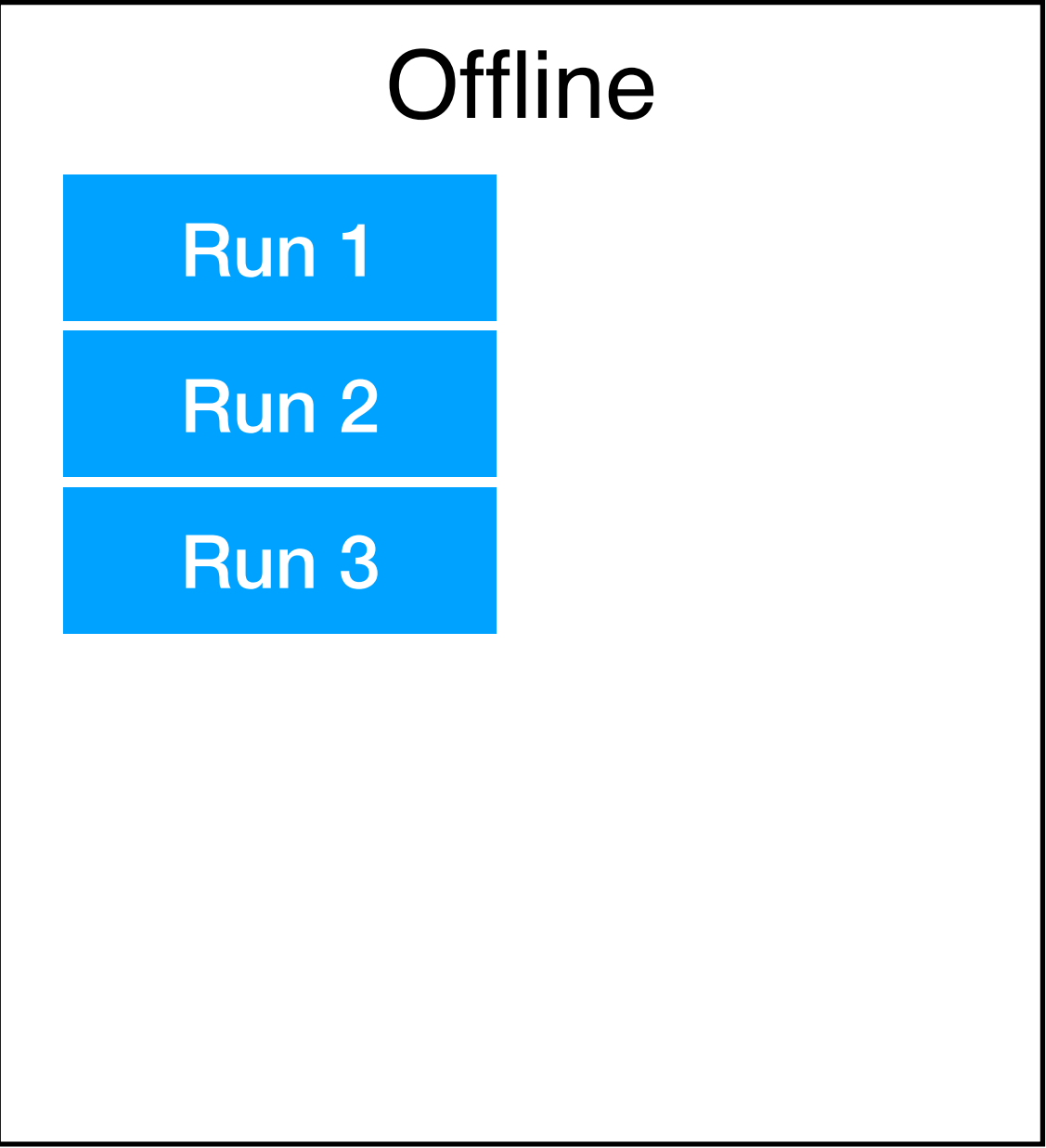


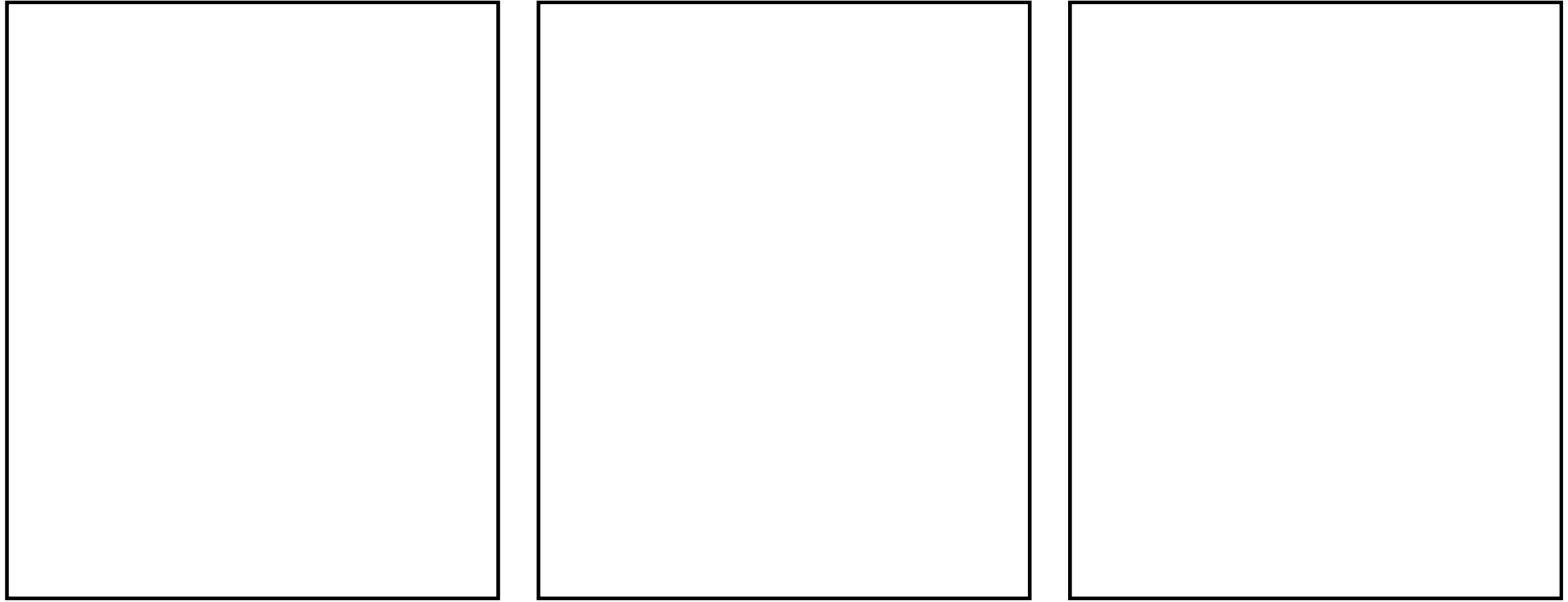
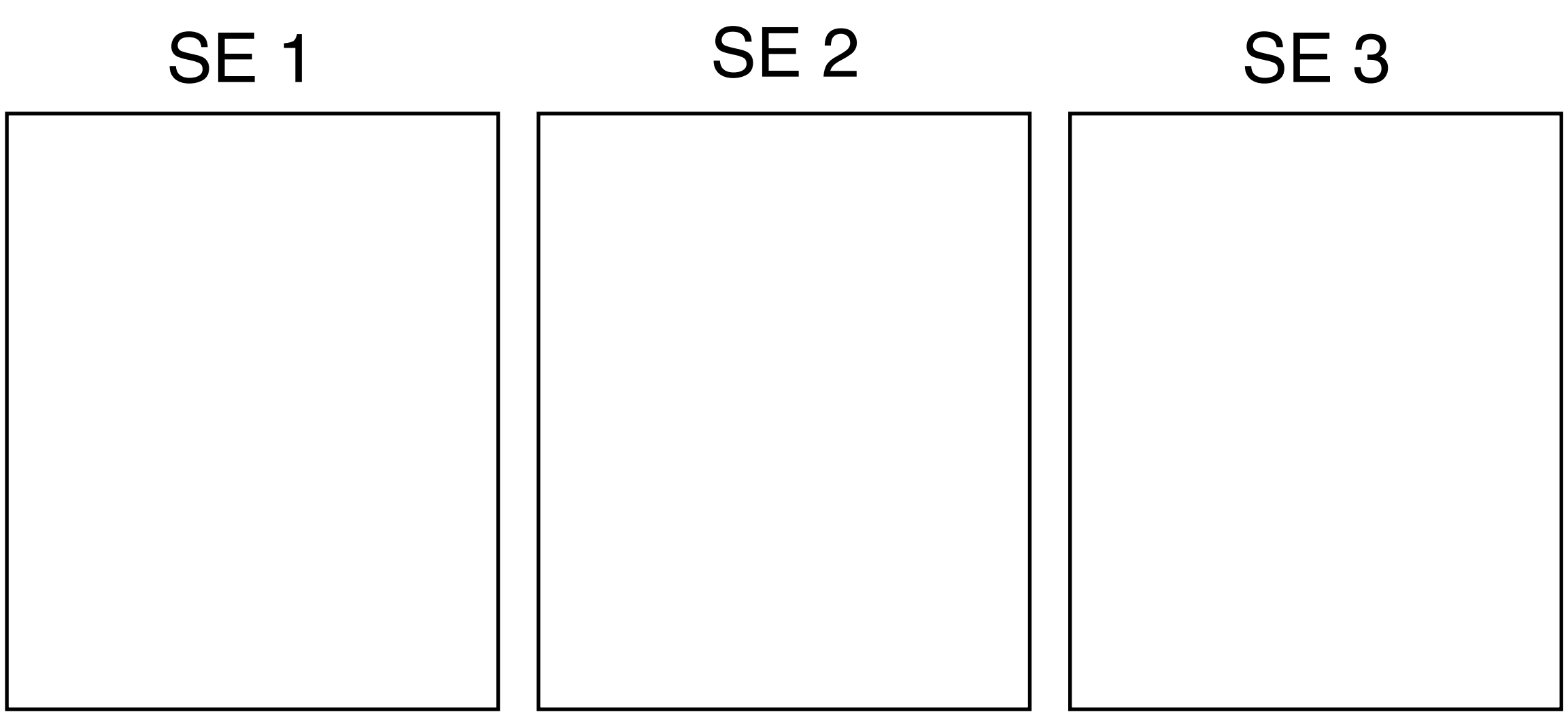
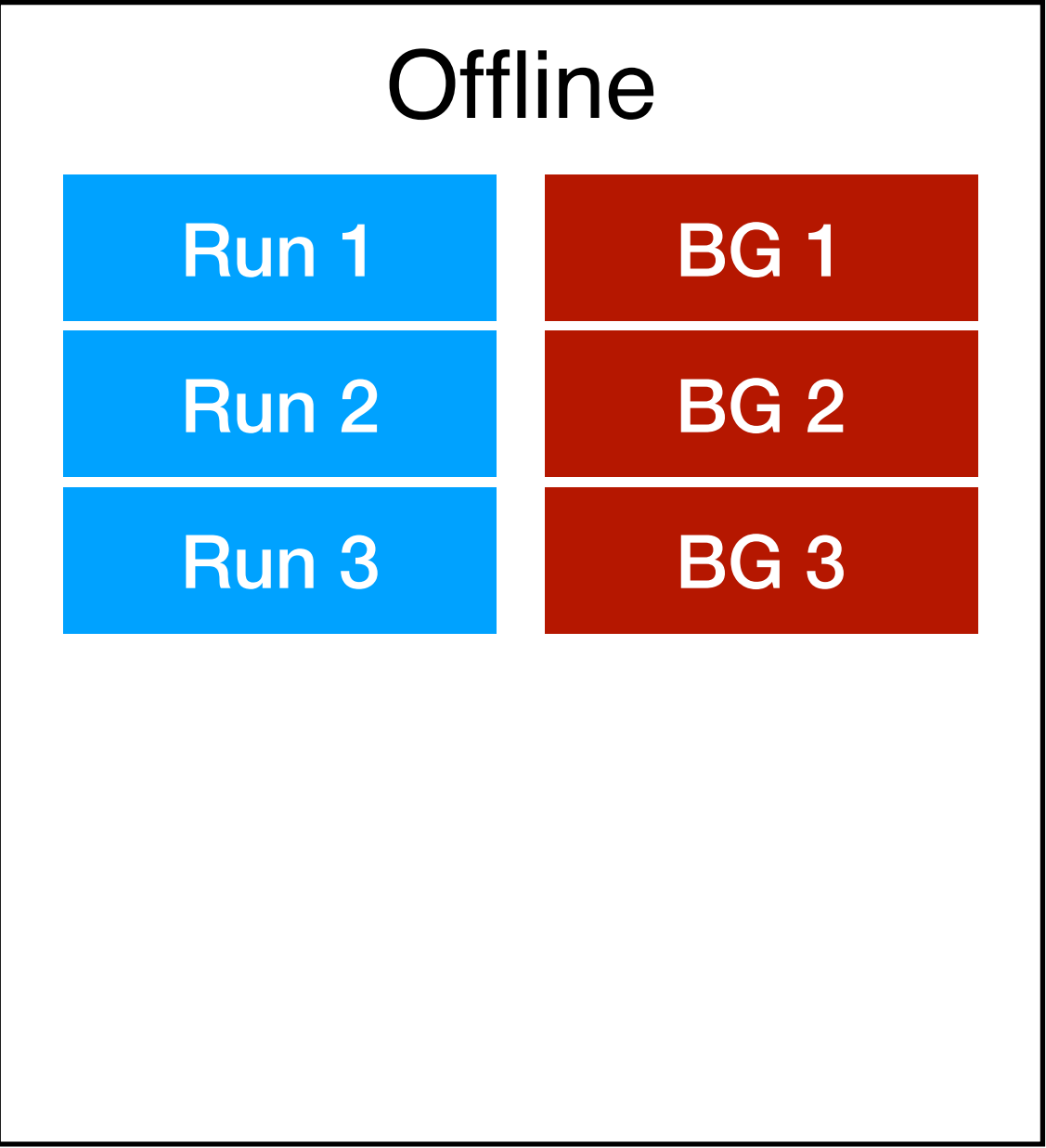
'BGOverlay.physics.0003.r06522.root' contains 1151 entries.

#:	Branch Name	Total		Per entry	
		Size	Compressed Size	Size	Compressed Size
1:	ARICHDigits	383.17K	81.60K		
2:	BKLMDigits	3.65M	488.99K		
3:	CDCHits	4.56M	1009.54K		
4:	ECLWaveforms	139.54M	123.68M		
5:	EKLMDigits	727.13K	102.96K		
6:	EventMetaData	163.56K	29.42K		
7:	PXDDigits	44.77K	13.04K		
8:	SVDShaperDigits	4.11M	1.01M		
9:	TOPDigits	3.15M	1.19M		

=====



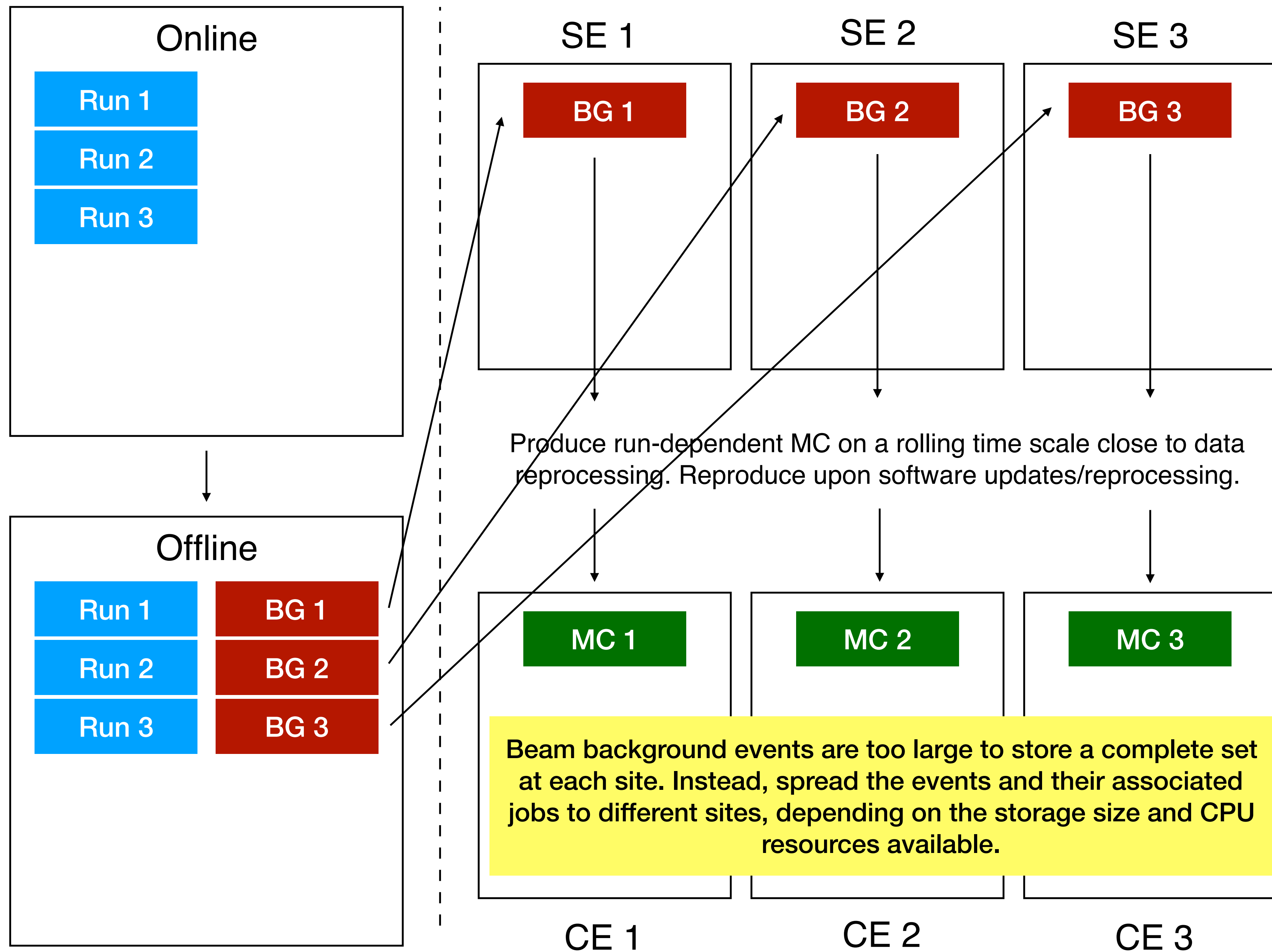


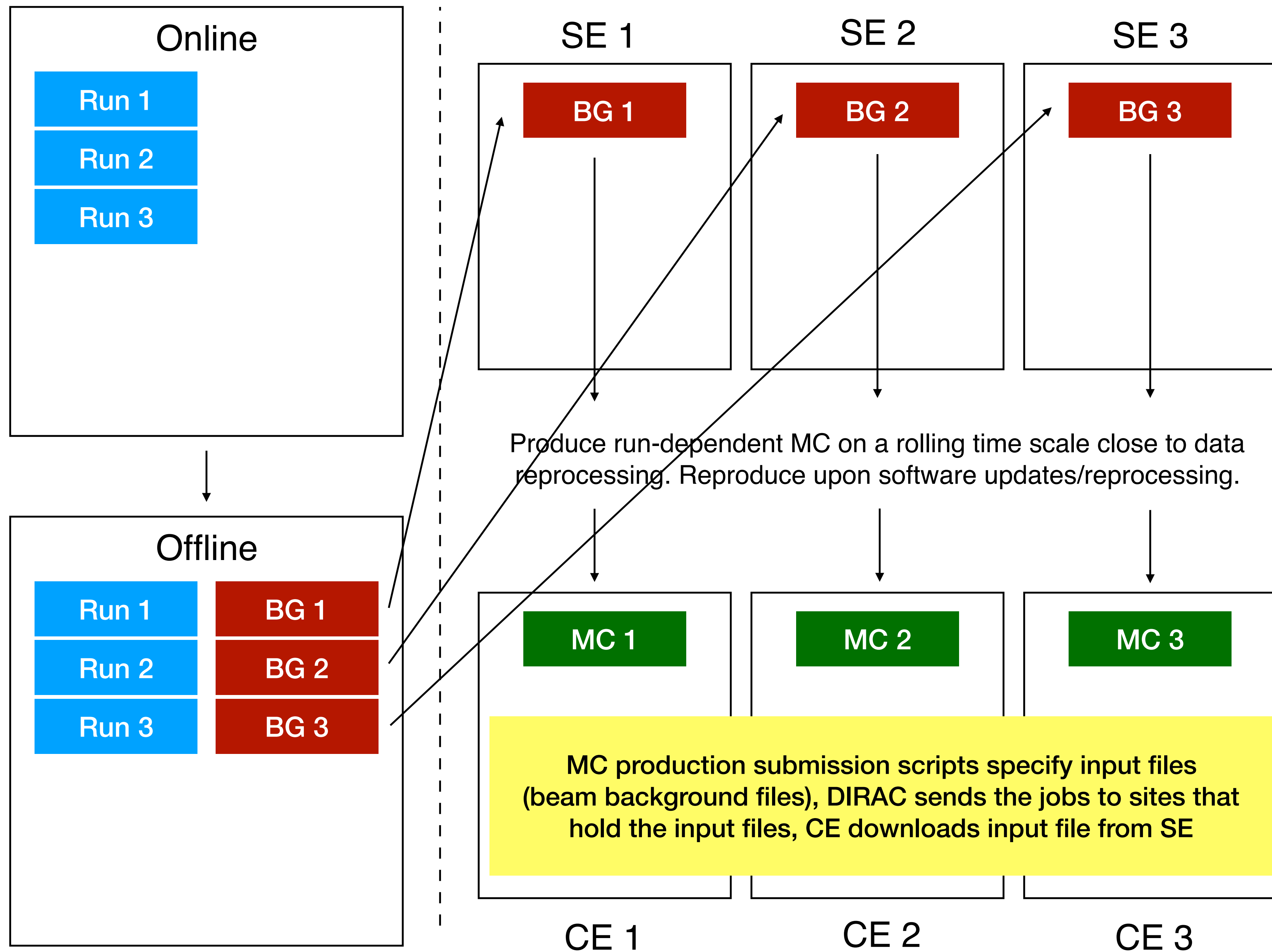


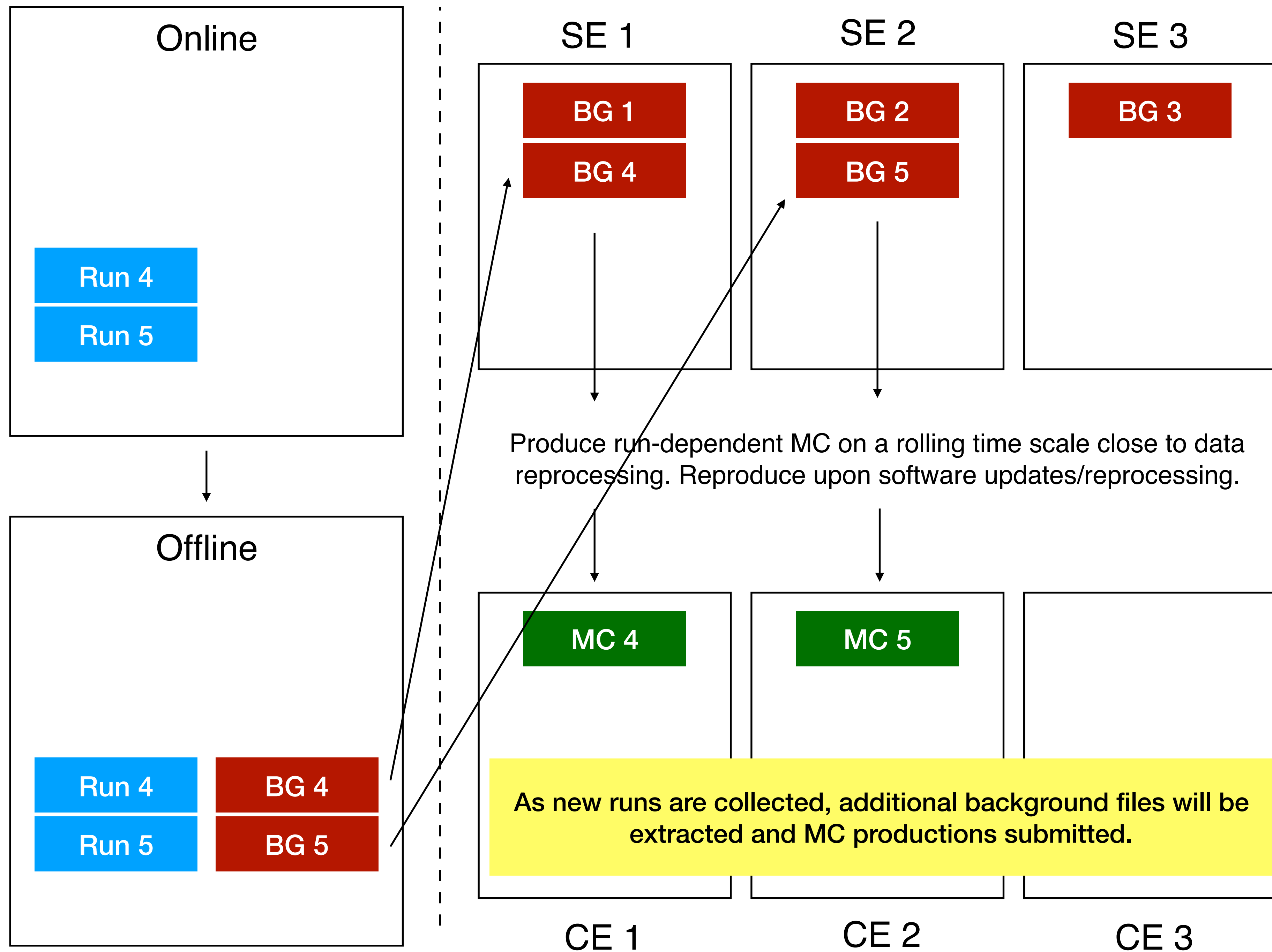
CE 1

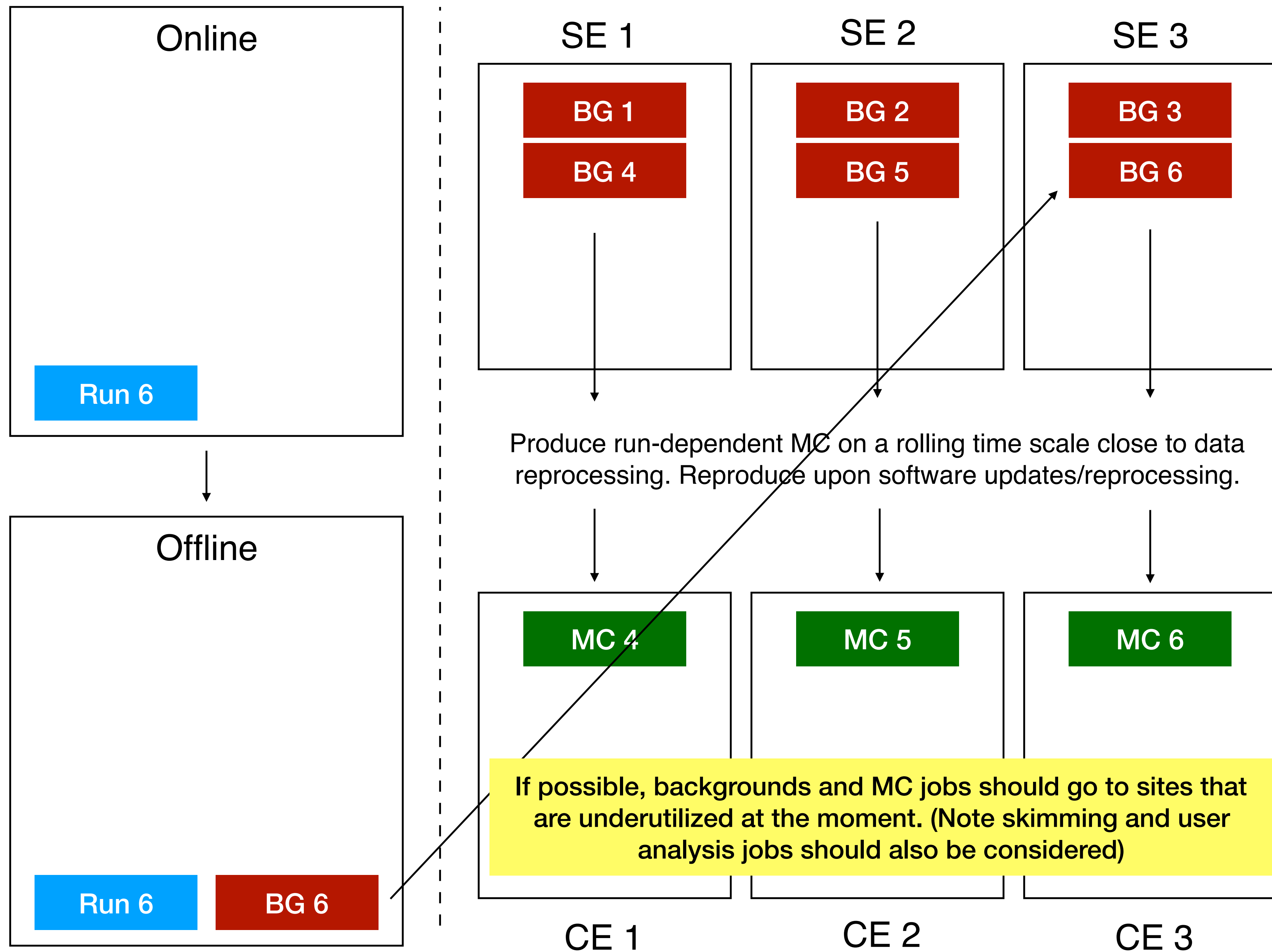
CE 2

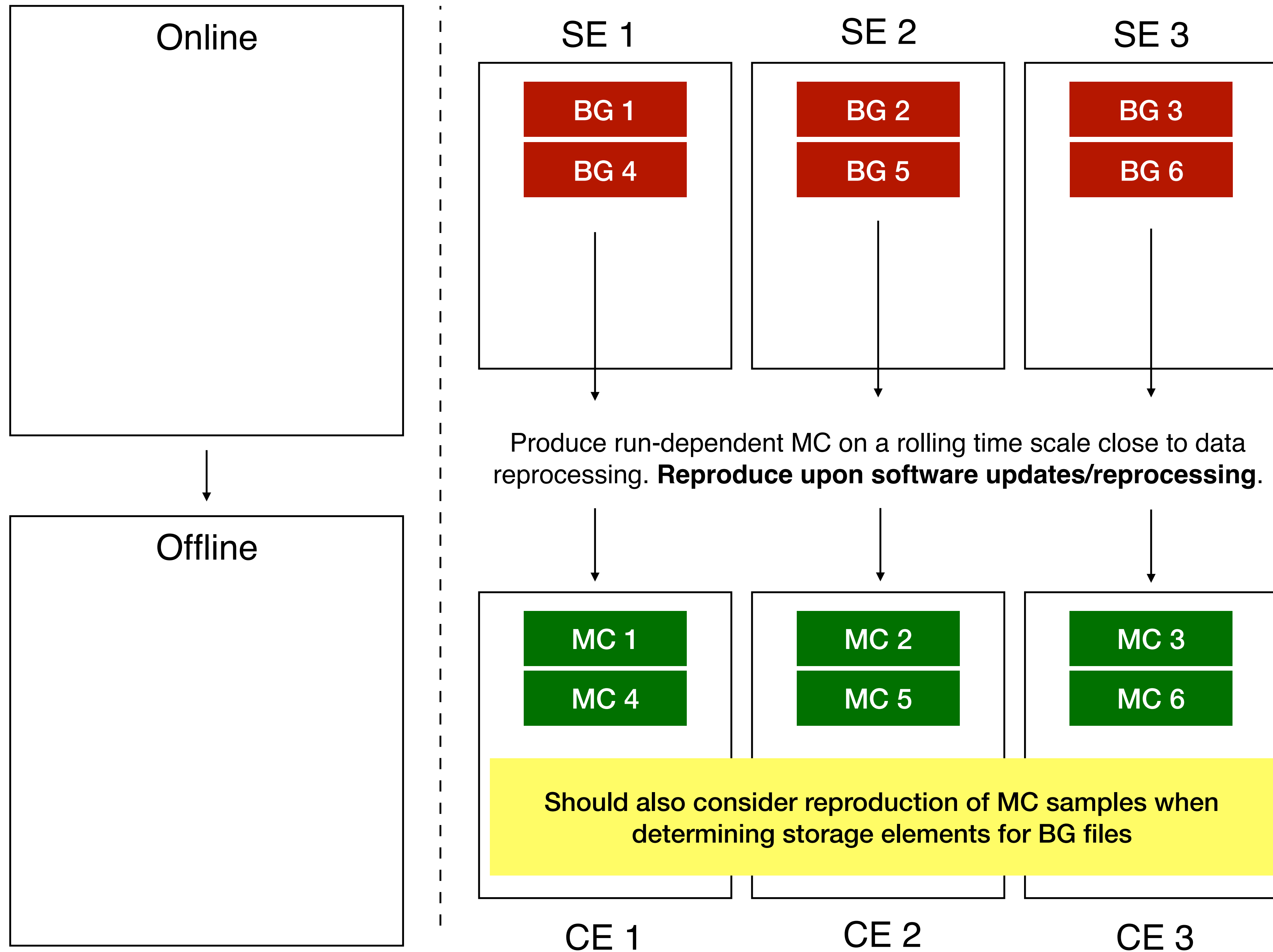
CE 3

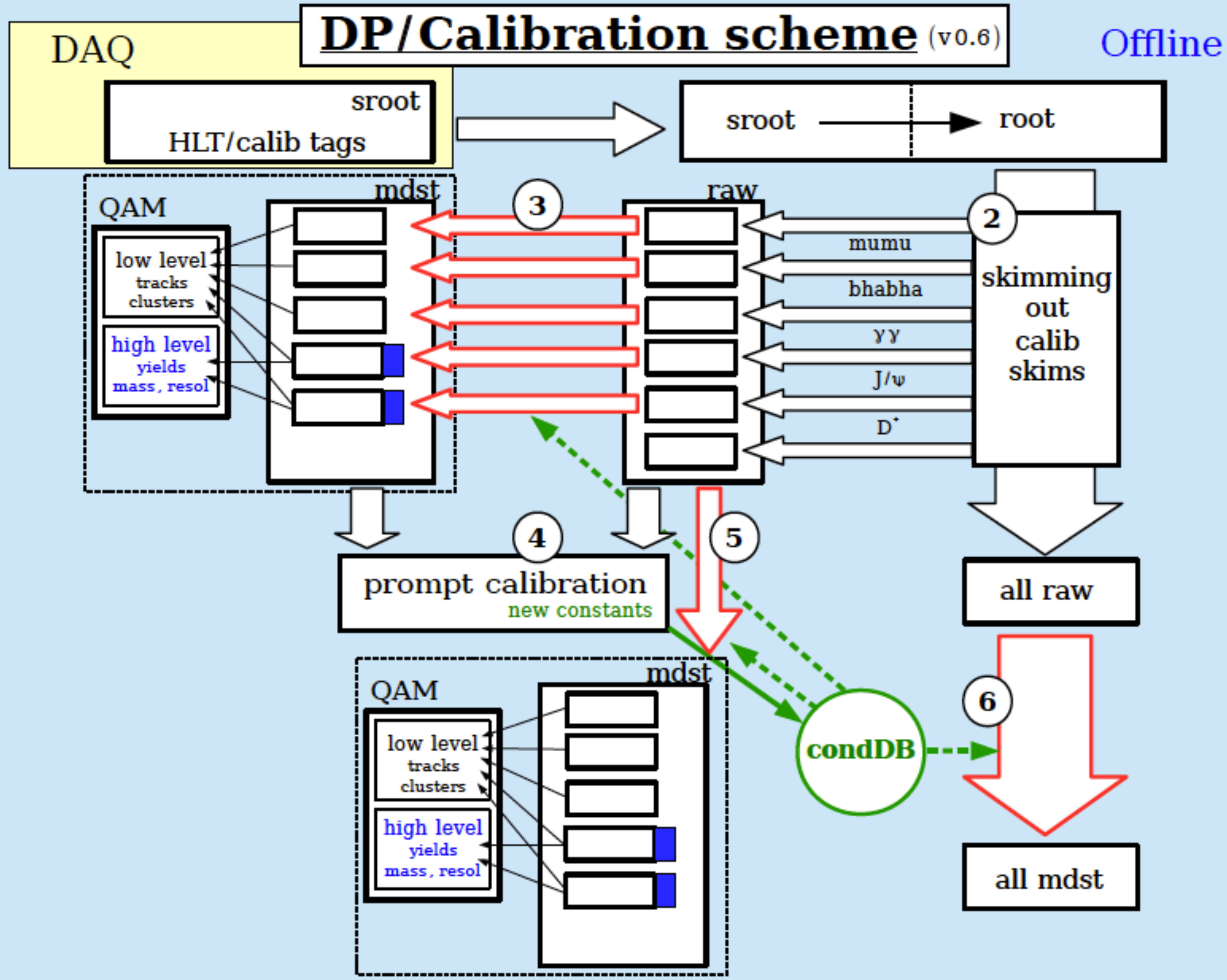






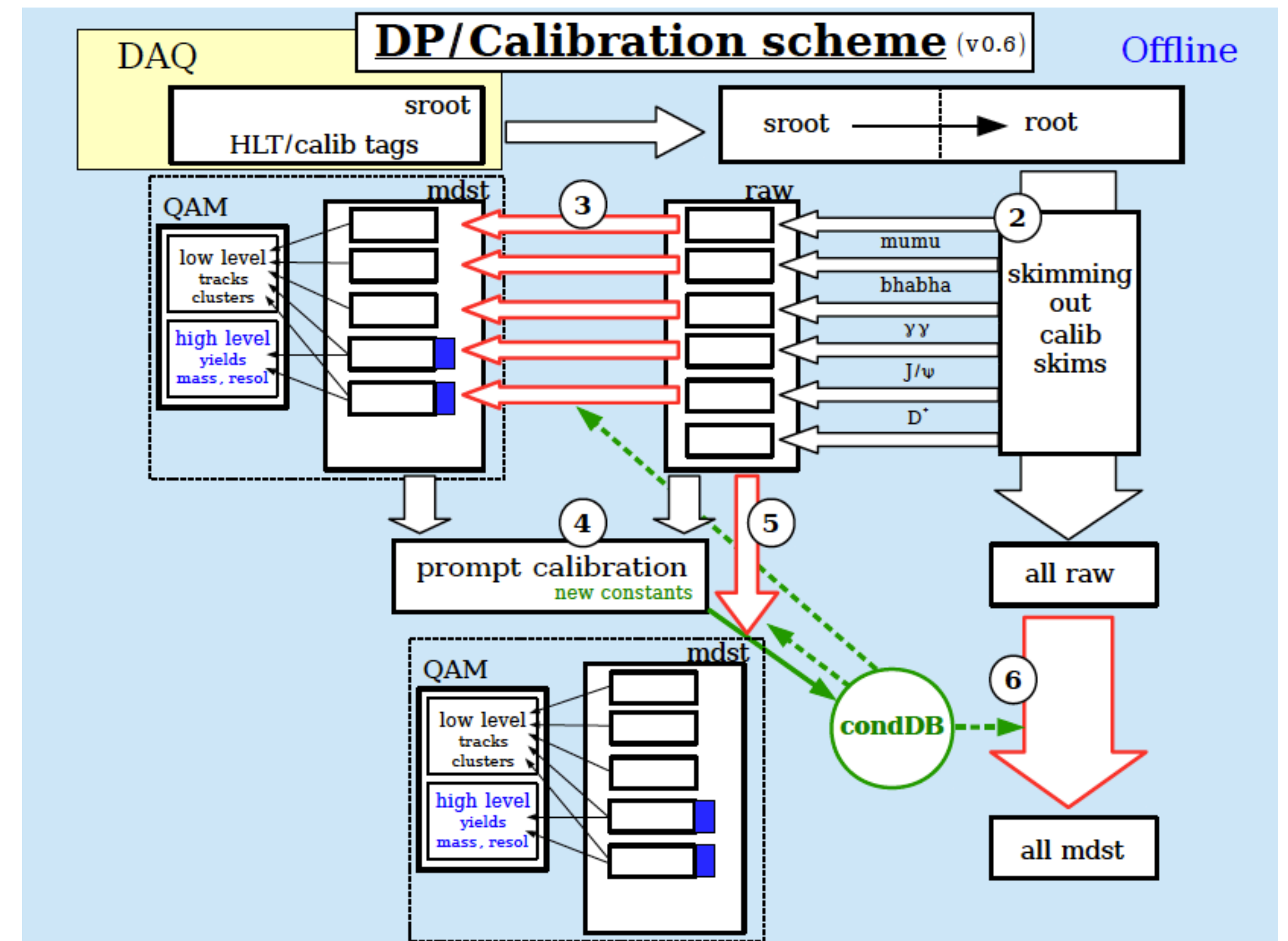






Full processing scheme

- **Preparations for (re)processing**
 - Prepare a global tag with the latest detector and calibration constants (coordinate with calibration group and GT manager)
 - Decide on and test a release (coordinate with release managers)
 - Determine the appropriate datasets (coordinate with computing coordinator)
- **(Re)processing**
 - Output in mDST/cDST format
 - Luminosity estimation
 - Calibration skims
- **Quality Assurance Monitoring (QAM)**
- **Physics skims**



Step 1: Prepare a global tag with the latest detector and calibration constants

- Various effects can cause bias or degradations
 - misalignment, time offset, channel-to-channel variations, weather, etc.
- Careful calibration is necessary to properly reconstruct the data

Calibration flow:

1. Local calibrations

2. CDC tracking

3. Alignment updates

4. Produce cDSTs

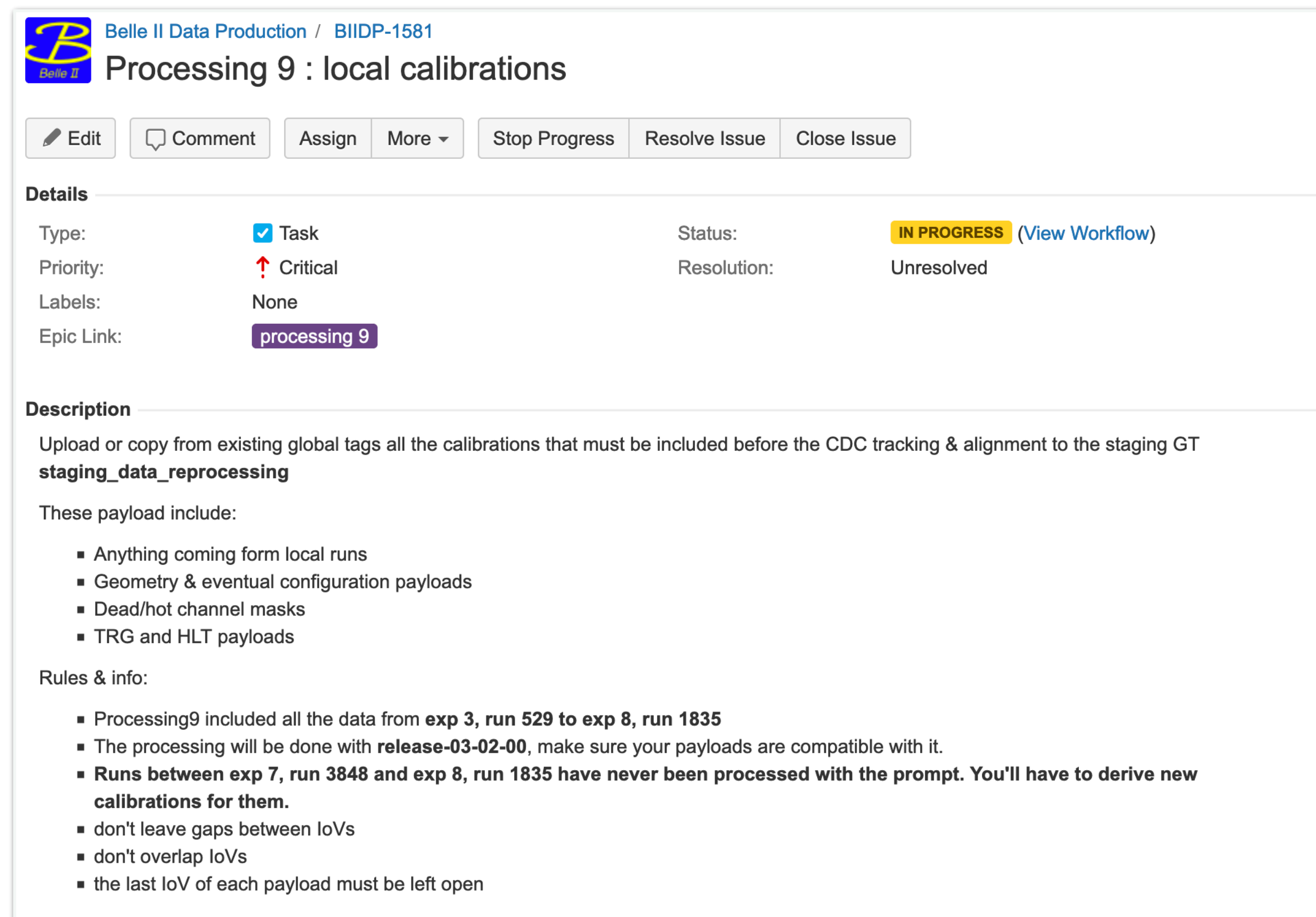
5. All other calibrations

6. Data processing

7. Validation

There is a hierarchy!


- None of these steps can be parallelized
- Each step requires the results of the previous one










The screenshot shows the Belle II Data Production (BIIDP) interface for task BIIDP-1581, titled "Processing 9 : local calibrations". The interface includes a header with the Belle II logo and navigation buttons (Edit, Comment, Assign, More, Stop Progress, Resolve Issue, Close Issue). The "Details" section shows the task is a "Task" with "Critical" priority, "None" labels, and an "Epic Link" of "processing 9". The status is "IN PROGRESS" (View Workflow) and the resolution is "Unresolved". The "Description" section explains the task's purpose: "Upload or copy from existing global tags all the calibrations that must be included before the CDC tracking & alignment to the staging GT staging_data_reprocessing". It lists the payload contents: "Anything coming from local runs", "Geometry & eventual configuration payloads", "Dead/hot channel masks", and "TRG and HLT payloads". The "Rules & info" section provides specific instructions: "Processing9 included all the data from exp 3, run 529 to exp 8, run 1835", "The processing will be done with release-03-02-00, make sure your payloads are compatible with it.", "Runs between exp 7, run 3848 and exp 8, run 1835 have never been processed with the prompt. You'll have to derive new calibrations for them.", "don't leave gaps between IoVs", "don't overlap IoVs", and "the last IoV of each payload must be left open".

Step 1: Prepare a global tag with the latest detector and calibration constants

- Various effects can cause bias or degradations
 - misalignment, time offset, channel-to-channel variations, weather, etc.
- Careful calibration is necessary to properly reconstruct the data


 Belle II Data Production / BIIDP-1581

Processing 9 : local calibrations


 Edit  Comment  Assign  More  Stop Progress  Resolve Issue  Close Issue


Details

Type: ☒ Task

Priority:  Critical

Labels: None

Epic Link: 

Status:  IN PROGRESS [\(View Workflow\)](#)

Resolution:

Description

Upload or copy from existing global tags all the calibrations that must be included before the CDC to **staging_data_reprocessing**

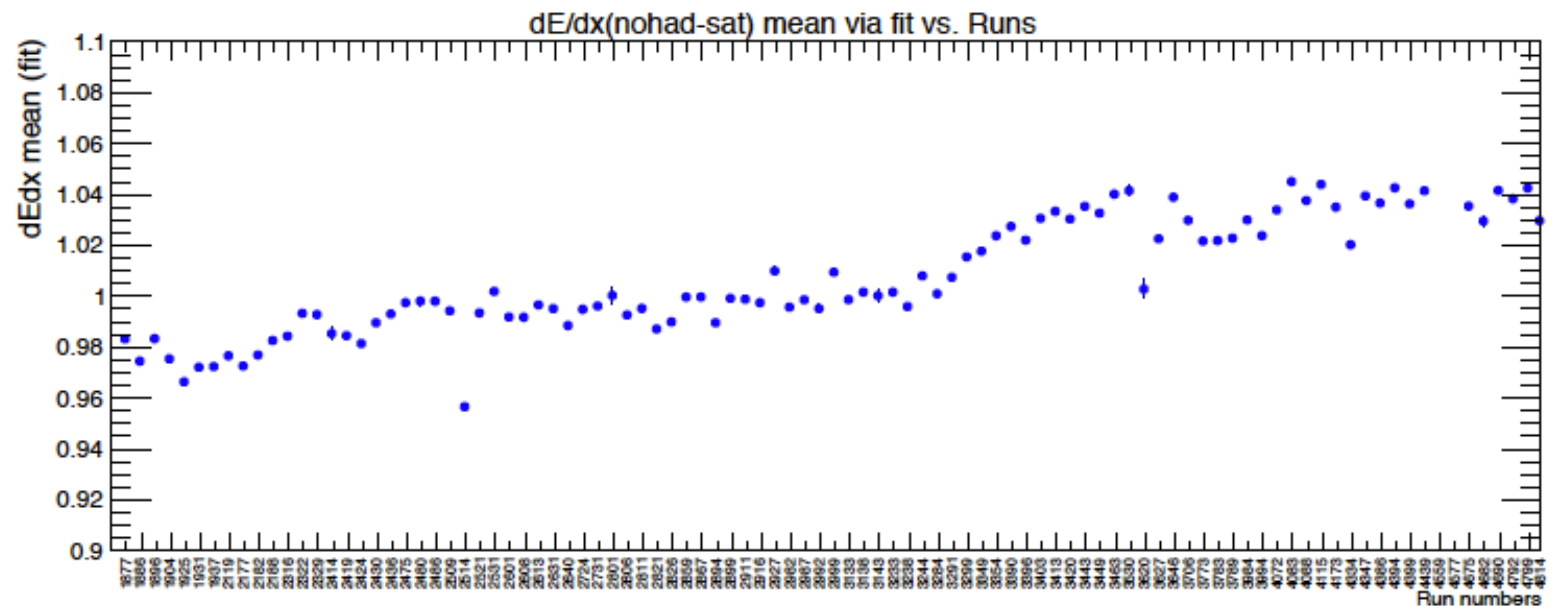
These payload include:

- Anything coming from local runs
- Geometry & eventual configuration payloads
- Dead/hot channel masks
- TRG and HLT payloads

Rules & info:

- Processing9 included all the data from **exp 3, run 529 to exp 8, run 1835**
- The processing will be done with **release-03-02-00**, make sure your payloads are compatible
- **Runs between exp 7, run 3848 and exp 8, run 1835 have never been processed with the calibrations for them.**
- don't leave gaps between IoVs
- don't overlap IoVs
- the last IoV of each payload must be left open

1 dE/dx resolution for bhabhas ~ 7%
2 variation versus run gain ~ 6%!



Step 1: Prepare a global tag with the latest detector and calibration constants

- A **Global Tag** is a collection of **payloads** that are used by the software to determine needed constants

LMU

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

What is a Global Tag?

global tag

payload #1

rev.1

rev.2

rev.1

payload #2

rev.1

rev.2

payload #3

rev.1

rev.3

IOV

time

payload

One atom of conditions data (e.g. BeamParameters).
In our case this is just a file.
They are identified by name and have different revisions

IOV

Short for “interval of validity”, the run interval for which the payload is valid.
Can be a fixed run range (closed) or starting at a given run (open)

global tag

Is an **immutable** set of payloads and their IOVs

Martin Ritter

32nd B2GM, 2019-02-07

Global Tag (GT) page

Marko Bracko posted on 18. Jul. 2017 17:53h – last edited by Sam Cunliffe on 24. May. 2019 11:09h

Which global tag should I use?

- In case of real data check this page. If you don't find the answer you are looking for, or you don't understand what you sh tags for data processing and analysis (see below).
- In case of MC ask the responsible person for the global tags for data processing, MC production, and analysis (see below)

What do I have to do when I want to update/add conditions?

Ask the responsible person (see below).

GT web browser

Questions? Write to potekhin@bnl.gov

Home

GlobalTag

Payload

Types of Payload

Global Tag Comparison

GlobalTag: 41 items found Click on items for more details

Name (can be partial): data_reproces:

Status: All

Type: All

Modified by:

items per page: 25

Submit

Name	ID	Default?	Description	Status	Type	Modified	Mod. by	Total Payloads
data_reprocessing_prompt_bucket6	618	✓	running GT for the first (prompt) reprocessing	PUBLISHED	RELEASE	05/25/2019 10:12 a.m.	tamponi	5387
data_reprocessing_prompt_2019-05-13	609	✓	running GT for the first (prompt) reprocessing	NEW	RELEASE	05/13/2019 3:15 p.m.	tamponi	3731
data_reprocessing_prompt_bucket6_cdst	607	✓	running GT for the first (prompt) reprocessing	PUBLISHED	RELEASE	05/12/2019 9:39 a.m.	tamponi	3731
data_reprocessing_prompt_bucket6_alignment	604	✓	running GT for the first (prompt) reprocessing	PUBLISHED	RELEASE	05/09/2019 9:37 a.m.	tamponi	3724
data_reprocessing_prompt_2019-05-09	601	✓	running GT for the first (prompt) reprocessing	NEW	RELEASE	05/08/2019 10:46 p.m.	tamponi	2883
data_reprocessing_prompt_2019-05-06	600	✓	running GT for the first (prompt) reprocessing	NEW	RELEASE	05/06/2019 8:19 a.m.	tamponi	2883
data_reprocessing_prompt_bucket5	599	✓	running GT for the first (prompt) reprocessing	PUBLISHED	RELEASE	05/04/2019 11:47 a.m.	tamponi	2883
data_reprocessing_prompt_bucket4b	598	✓	running GT for the first (prompt) reprocessing	PUBLISHED	RELEASE	05/02/2019 12:41 p.m.	tamponi	2054
data_reprocessing_prompt_2019-05-01	595	✓	running GT for the first (prompt)	NEW	RELEASE	05/01/2019 11:20 a.m.	tamponi	2054

<https://kds.kek.jp/indico/event/29835/session/27/material/1/0.pdf>

[https://confluence.desy.de/display/BI/Global+Tag+\(GT\)+page](https://confluence.desy.de/display/BI/Global+Tag+(GT)+page)

<https://blcondtest01.sdcc.bnl.gov/>

Step 1: Prepare a global tag with the latest detector and calibration constants

```
[jbennett@cw02 ~]$ basf2 --info
```

```

          eeeeeee
        eeeeeeeeeeeeeee
      eeeeeee eeeeeeeeeeeeeee
    eeeeeee eeeee eeeeeee
  eeeee eeeee eeeee eeeee
 eeeee eeeee eeeee eeeee
  eeee eeeeeeeeeeeeeeeee
    eeeeeeeeeeeeeee
      eeeeeeeeeeeeeee
        eeeee eeeee eeeeeee
    eeeee eeeee eeeeeee
  eeeee eeeee eeeee eeeee
 eeeee eeeee eeeee eeeee
  eeeee eeeee eeeee eeeeeee
    eeeeeeeeeeeeeeeee
      eeeeeeeeeeeeeee
        BBBB BBB 11 11 222222
       BB BB eeee 11 11 eeee 22 22
      BB BB ee ee 11 11 ee ee 22 22
     BBBB BBB eeeeeee 11 11 eeeeeee 22 22
    BB BB ee 11 11 ee 22 22
   BB BB ee ee 11 11 ee ee 22 22
  BBBB BBB eeeee 11 11 eeeee 222222

```

```

BASF2 (Belle Analysis Software Framework 2)
Copyright(C) 2010-2018 Belle II Collaboration
Release release-03-02-00
Version release-03-02-00

```

```

-----
BELLE2_RELEASE:      release-03-02-00
BELLE2_RELEASE_DIR:  /cvmfs/belle.cern.ch/sl6/releases/release-03-02-00
BELLE2_LOCAL_DIR:
BELLE2_SUBDIR:       Linux_x86_64/opt
BELLE2_EXTERNALS_VERSION: v01-07-01
BELLE2_ARCH:         Linux_x86_64
Default global tags:  release-03-02-00_rev2
Kernel version:      2.6.32-754.12.1.el6.x86_64
Python version:       3.6.6
ROOT version:         6.14/06

basf2 module directories:
  /gpfs/home/belle2/jbennett
  /cvmfs/belle.cern.ch/sl6/releases/release-03-02-00/modules/Linux_x86_64/opt
-----

```

```
[jbennett@cw02 ~]$ b2conditionsdb tag show release-03-02-00_rev2
[INFO] Getting info for global tag release-03-02-00_rev2
```

```

-----
id            621
-----
name          release-03-02-00_rev2
description    Software Development and Testing Tag\x0aThis global tag containing necessary payloads for testing of the software and\x0aexperiment
               independent MC. It cannot be used to analyze data and only contains\x0apayloads for the following intervals of validity:\x0a\x0a -
               exp 0: nominal Belle II Configuration with full PXD (full phase 3)\x0a - exp 1002: Phase 2 Configuration with minimal vertexing
               detectors and\x0a additional background detectors\x0a - exp 1003: Belle II Configuration with partial PXD (early phase
               3)\x0a\x0aPrevious global tag: release-03-02-00_rev1\x0aChanges are collected in Pull request #4182, see
               https://stash.desy.de/projects/B2/repos/software/pull-requests/4182
type          RELEASE
status        PUBLISHED
# payloads    313
created       2019-06-06 21:47:01 local time
modified      2019-06-07 16:15:27 local time
modified by   tuppr_ware

```

- There are many tools that can help you find global tag details
- The primary option should be b2conditionsdb-recommend

```
[jbennett@cw02 ~]$ b2conditionsdb-recommend --help
usage: b2conditionsdb-recommend [-h] [--input INPUT] [--analysis]
```

This tool determines to recommended set of global tags.

optional arguments:

```
-h, --help            show this help message and exit
--input INPUT, -i INPUT
                       input file that should be processed, set to "MC" for
                       generation of run-dependent Monte-Carlo
--analysis, -a _       include global tag for analysis tools
```


Step 2: (Re)processing

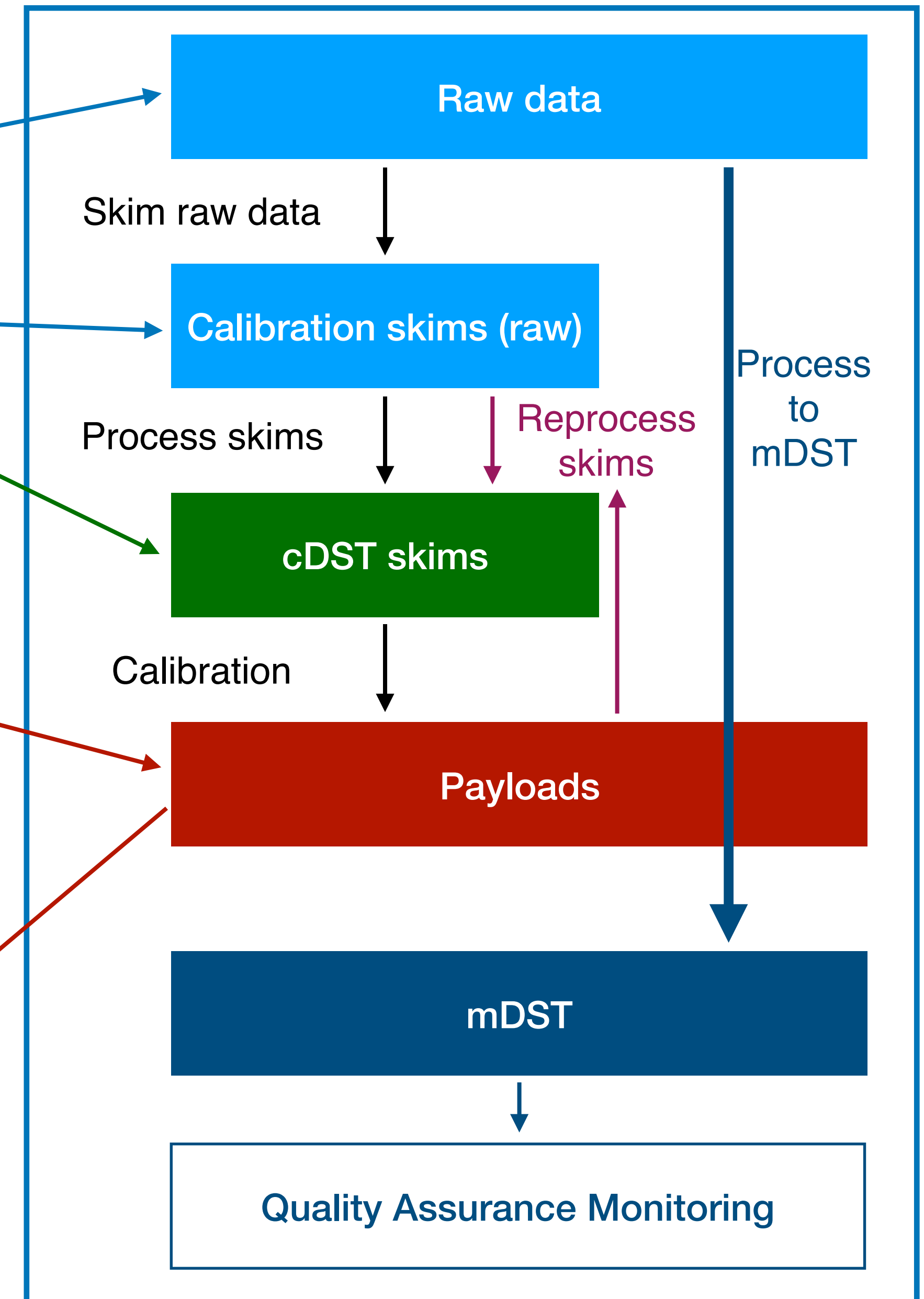
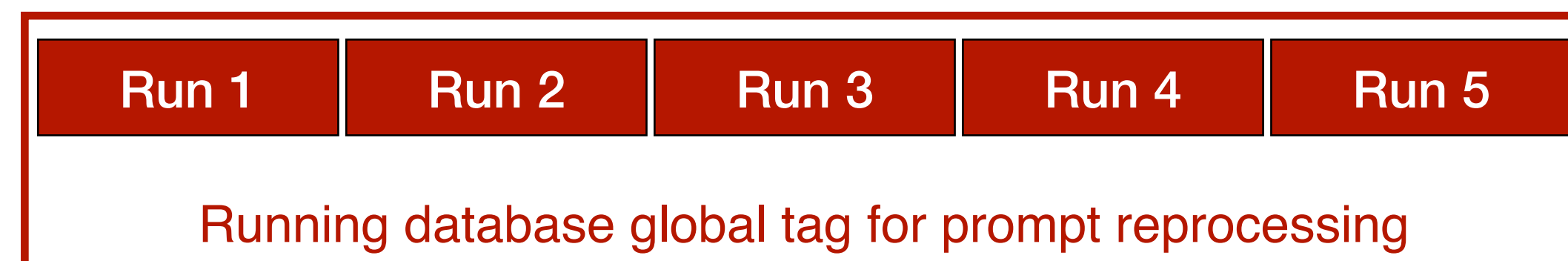
- Once the global tag is ready, the data can be officially processed to mDST format
- Requires:
 - Raw data available on the offline system (<https://agira.desy.de/browse/BIIDP-1497>) and staged to disk (<https://kekcc.kek.jp/service/kekcc/html/Eng/HSM20System.html>)
 - Definition of runs to be processed (all calibrations available) - called a run “bucket” or “proc”
 - Validated software release (<https://agira.desy.de/browse/BII-4880>)
 - Documentation!

Prompt reprocessing information sheet

Bucket Label	Exp #	Run Numbers	Run Type	Integrated Luminosity (pb ⁻¹) §	Magnetic Field Payload Rev	Global Tag	Global Tag ID	Reprocessing Status	JIRA Ticket
0	6	374-1052	Cosmic		B-OFF: 374-1035 (rev 21)	data_reprocessing_prompt_bucket0	550	<div>DONE</div>	<div><input checked="" type="checkbox"/> BIIDP-1225 - run bucket 0 - exp 6 run 374 to exp 6 run 1052</div> <div>CLOSED</div>
					QCS-OFF: 1036-1054 (rev 9)				
reprocessed data path: /ghi/fs01/belle2/bdata/Data/Cosmic/e0006/4S/GCR6a/release-03-00-03/DB00000550/r<RUN>/all/cdst/sub00/									
1	6	1053-1389	Cosmic		QCS-OFF: 1053-1389 (rev 9)	data_reprocessing_prompt_bucket1b	561	<div>DONE</div>	<div><input checked="" type="checkbox"/> BIIDP-1266 - run bucket 1 -</div>

Prompt reprocessing scheme

- ROOT formatted raw data on offline system, registered, and replicated to raw data processing centers
- Calibration skims from raw data processed to cDST at “calibration center”
- Prompt calibration and QAM run at “calibration center” (includes multiple reprocessing to cDST with updated tracking for dependent calibrations)
- Calibration constants added to offline global tag
- Latest runs reprocessed to mDST
- Requires that prompt calibration algorithms are ready for automation
- When offline calibrations and/or software updates are complete, or if significant changes to prompt calibration, reprocess all available data to mDST



But that's not all!

- For early phase 3, the data production group made many samples available ASAP after data taking
 - **Unofficial processing (user beware!)** - <https://agira.desy.de/browse/BIIDP-1533>
 - Raw data immediately processed to cDST for quick studies (R2 plots, luminosity measurements, detector studies, etc.)
 - Only performed on hlt skims
 - Uses processing conditions from latest official processing (no new calibrations = bad data!)
 - **Still processing everything** (eventually only for hlt skims)
 - **Still producing cDST for all events** (eventually only for calibration skims)
- Official (re)processing is also being performed on the grid (e.g. <https://agira.desy.de/browse/BIIDP-1574>)
 - This will be the official (and only) method to process raw data after early phase 3
 - Makes use of computing resources at KEK (Japan) and BNL (USA)
- Latest processing (proc9) includes up to experiment 8 run 1554 ($\sim 3/\text{fb} = 1/2$ of the existing data): read for analysis!
 - <https://confluence.desy.de/display/BI/Phase+3+data#Phase3data-Officialreprocessingdetails>
- Remainder of the data (bucket7) at the calibration step (<https://agira.desy.de/browse/BIIDP-1621>)

Quality Assurance Monitoring (QAM)

- Throughout the data processing chain, data quality monitoring is performed (well... it will be)
 - Check for problems with data corruption, calibration errors, unexpected behavior, etc.

- Important step to ensuring good quality physics data!**

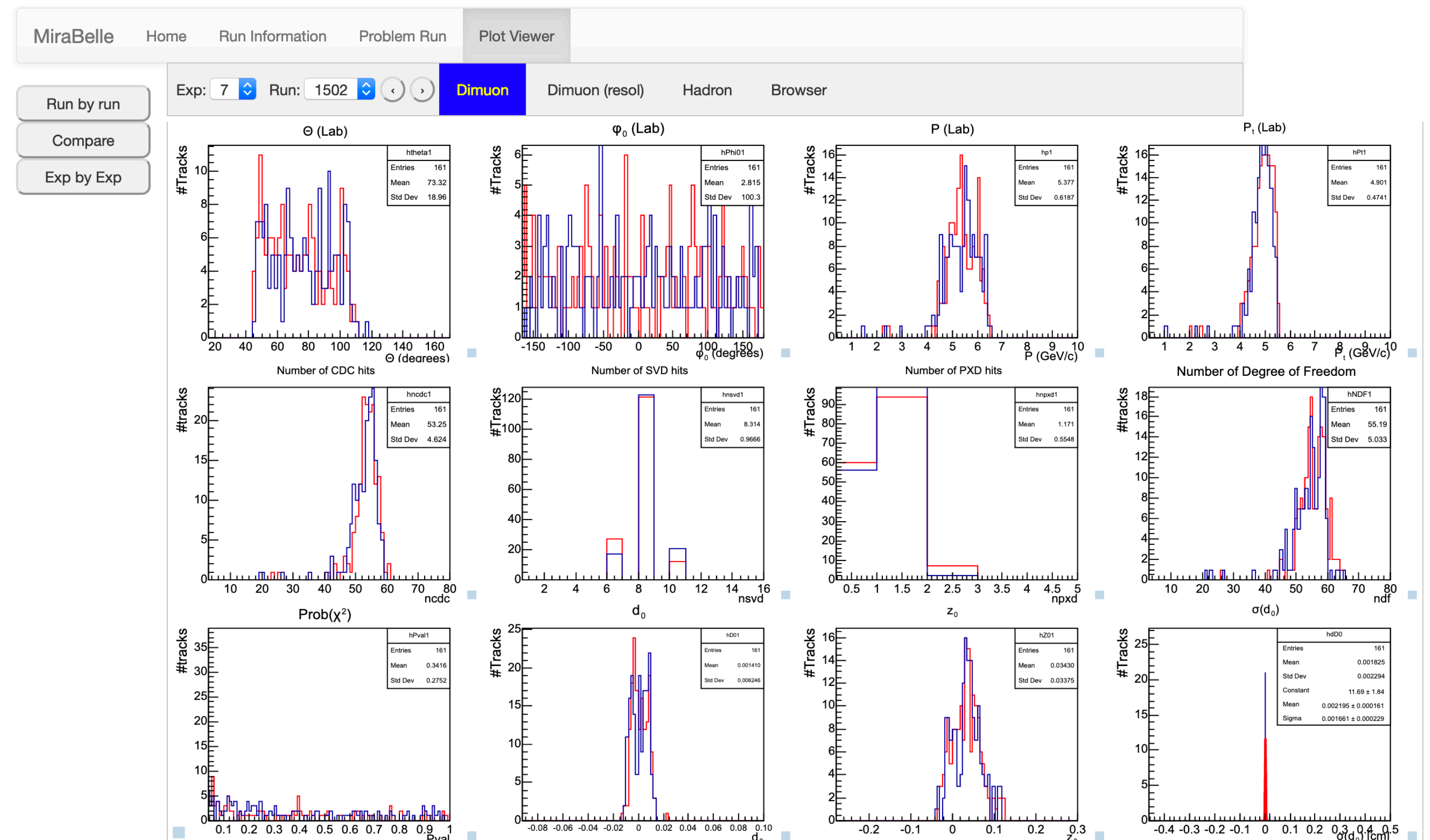
- MiraBelle: tool for official QAM after processing

- Review various distributions
- Compare different runs
- Flexible tool - still in development

- Important note: the tools and procedures are still being developed/exercised

- If you find issues, alert the DP team!

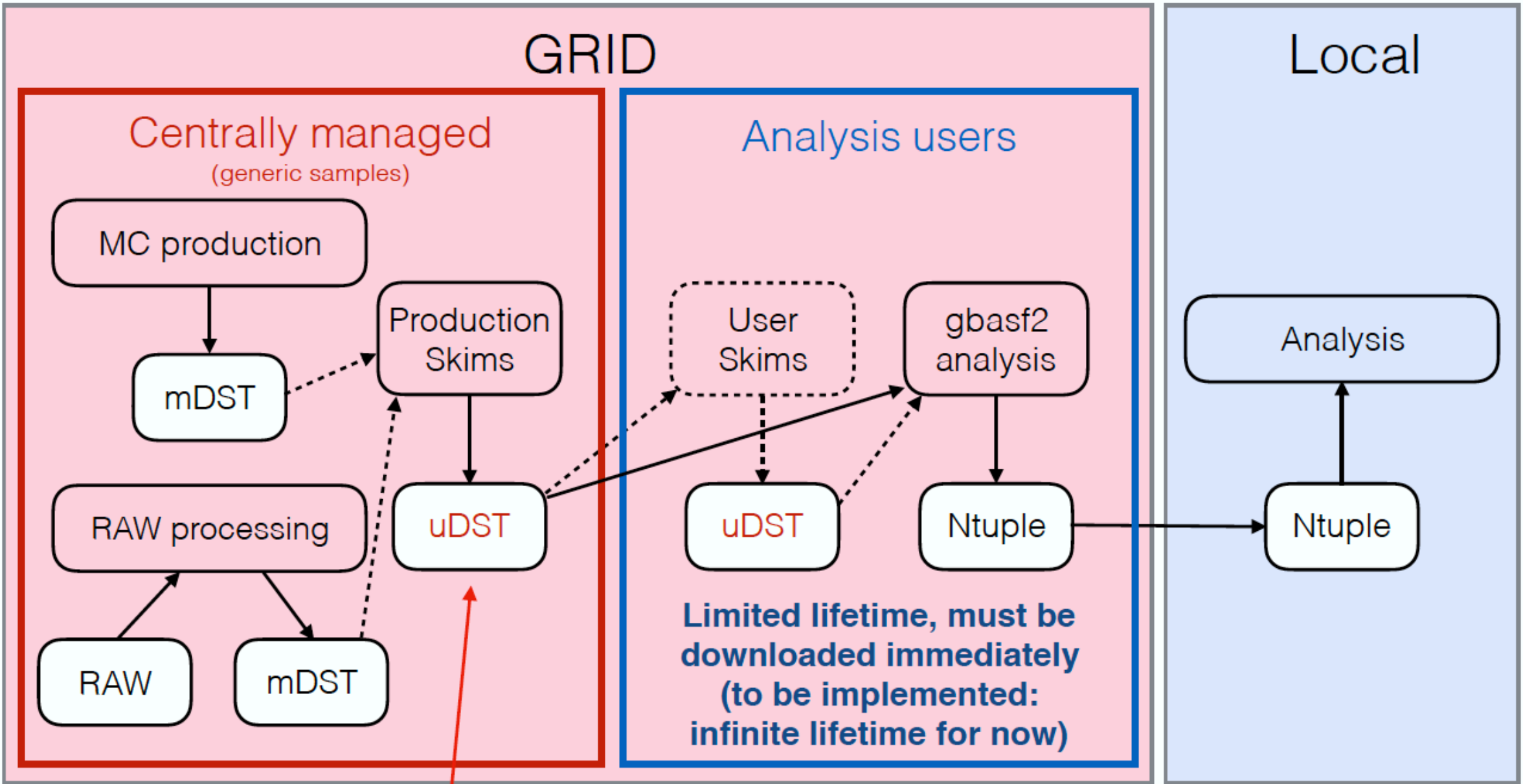
- [Make a JIRA ticket for BIIDP](#)
- Alert your working group liaison
- Ask [questions!](#)
- Join the DP effort (contact me!)



We're not done yet!

- **Keep in mind that you will not be able to touch mDST files for much longer!**
 - After processing, the mDST samples are skimmed to reduce the number of events and provide analysis dataobjects directly - reduces the CPU time for your jobs
 - Analysts should be using uDST files, not mDST!
 - **If your skim is not available, you won't be able to perform your analysis!**
 - Contact your skim liaison and/or start developing

The skim production manager is [@Racha Cheaib](#).



uDST = mDST plus analysis level information (ParticleLists)

Physics Working Group	Member	Status
BtoCharm WG	@Niharika Rout	Ph.D student, IITM
BtoCharmless WG	@Kim Hai Smith	Ph.D student
TCPV WG	@Reem Rasheed	Ph.D student
EWP WG	@Justin Tan	Ph.D. Student
Charm WG	@Guanda Gong	Ph.D. Student
Charmonium WG	@Sen Jia	Ph.D student
Bottomonium WG	@Sen Jia	Ph.D student
SL + Missing Energy WG	@Sophie Hollitt	Ph.D Student
	@Hannah Marie Wakeling	Ph.D Student
	@Philip Grace	Ph.D Student
Dark and Low Multiplicity WG	@Sam Cunliffe	Postdoctoral fellow
Tau WG	@Kenji Inami	Staff

Where can I find more information?

- <https://www.belle2.org/>
- chat: <https://chat.belle2.org>
- questions: <https://questions.belle2.org>
- calendar: <http://superb.kek.jp/meetings/calendar.html>
- glossary: <https://confluence.desy.de/display/BI/Main+Glossary>
- sympa (email lists): <https://lists.belle2.org/sympa/home>
- SpeakApp: <https://speakapp.link/>
 - <https://confluence.desy.de/display/BI/Main+AudioVideoConferencing>
- git/stash: <https://stash.desy.de/>
 - <https://confluence.desy.de/pages/viewpage.action?pageId=35832648>
 - <https://confluence.desy.de/pages/viewpage.action?pageId=35819226>
- JIRA: <https://agira.desy.de>
- basf2: <https://b2-master.belle2.org/software/development/sphinx/index.html>
- gbasf2: <https://confluence.desy.de/display/BI/Computing+GBasf2>
 - <https://confluence.desy.de/display/BI/Instructions+for+gbasf2+analysis>
- Documentation, training and outreach: <https://confluence.desy.de/display/BI/Documentation%2C+Training+and+Software+outreach>
- Data production: <https://confluence.desy.de/display/BI/Data+production+WebHome>
 - Phase 2 data: <https://confluence.desy.de/display/BI/Experiment+3>
 - Phase 3 data: <https://confluence.desy.de/display/BI/Phase+3+data>
- KEK user site: https://krs.kek.jp/uskek/ui/UI_00000E.do