



Belle II Data production

2024 Belle Analysis Workshop



Renu

on behalf of the Data Production group

Supported by US DOE funding

19th October, 2024 - 23th October, 2024

Belle II Data Production

- **Primary Goal:**

- Smooth, timely production of data and MC samples for physics analysis and other studies

- **Tasks:**

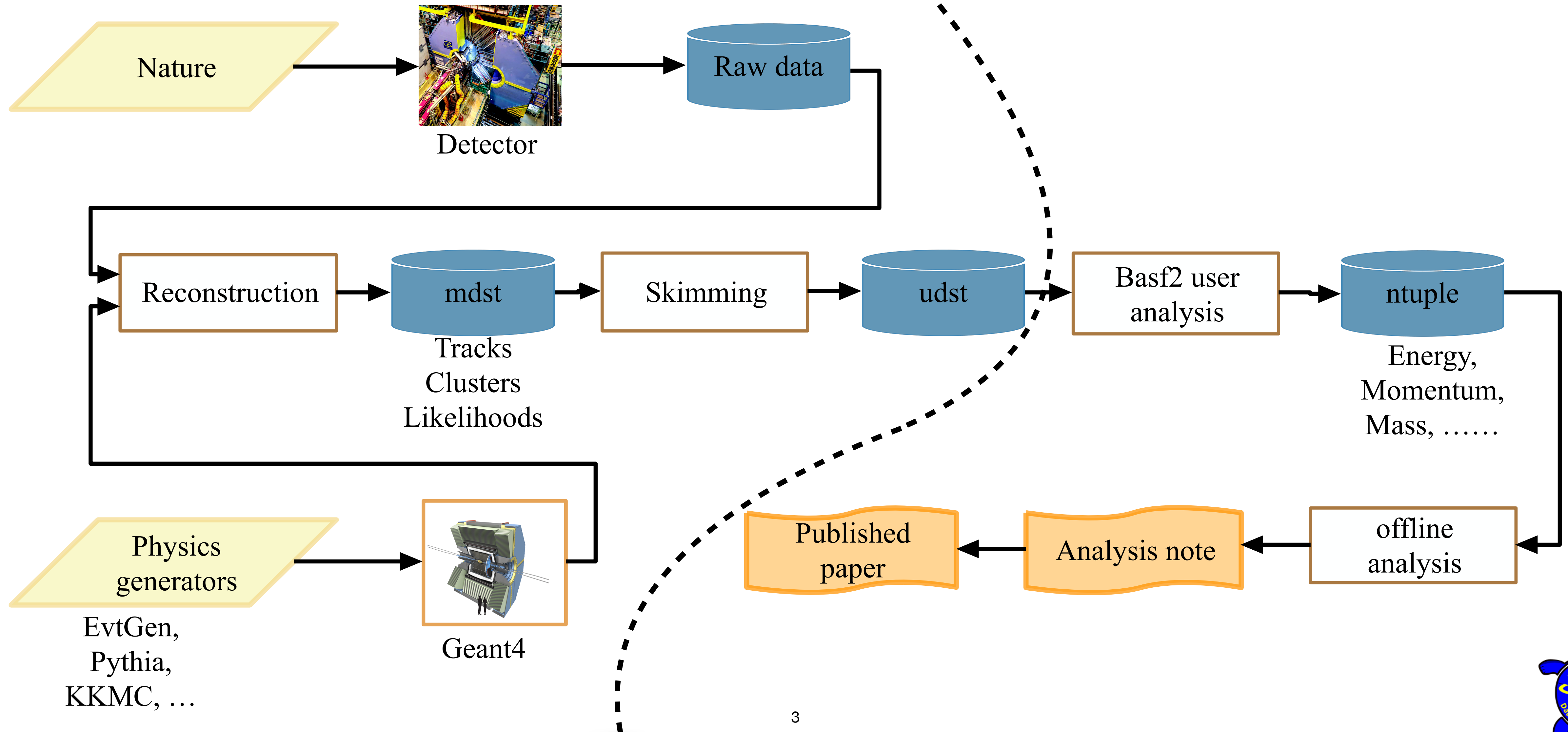
- Calibration and alignment (<https://confluence.desy.de/display/BI/Data+Production+Calibration>)
- RAW data (re)processing (<https://confluence.desy.de/display/BI/Phase+3+data>)
- MC production (<https://confluence.desy.de/display/BI/Data+Production+MC12>)
- Analysis skimming (<https://confluence.desy.de/display/BI/Skimming+Homepage>)



The Big Picture

Managed by collaboration

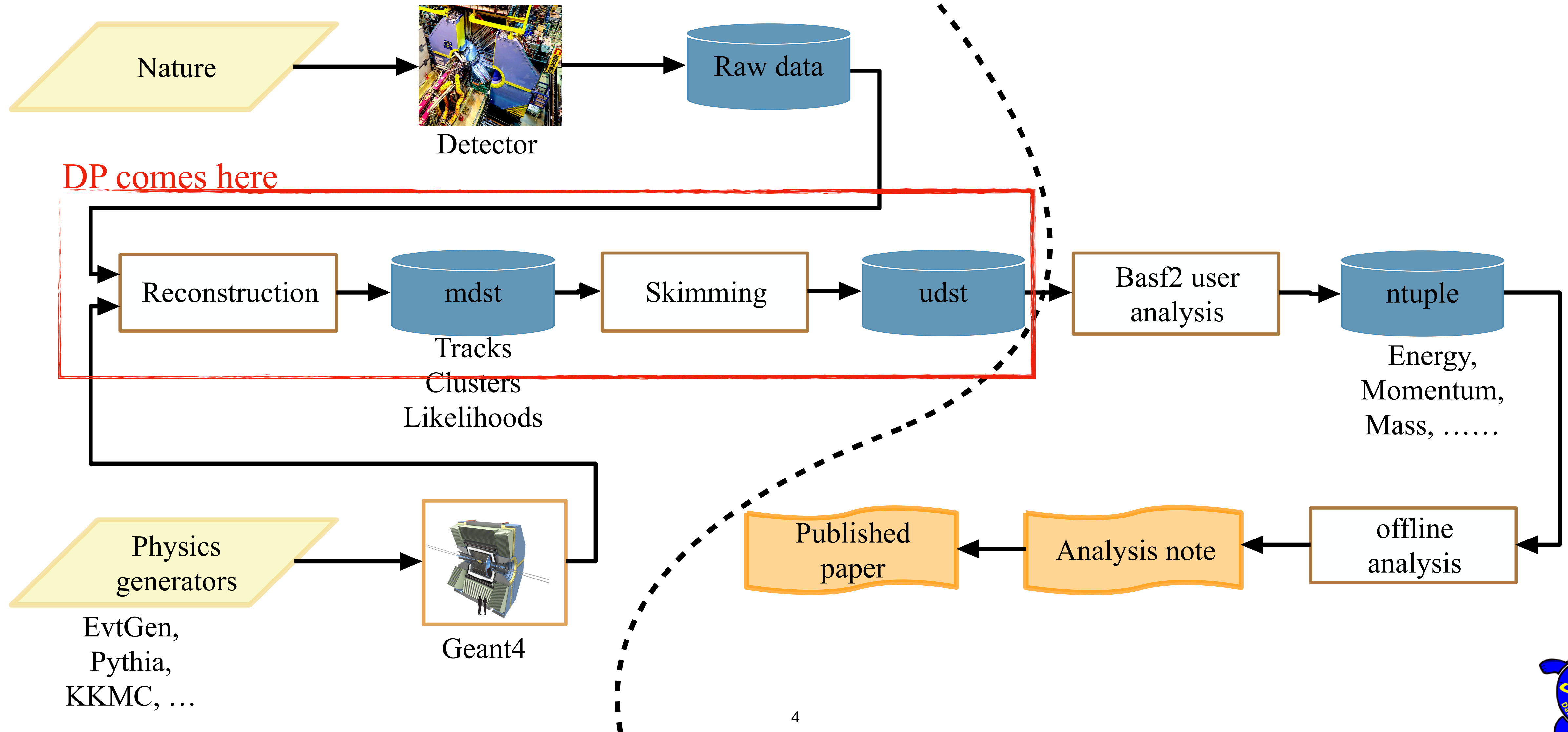
Managed by analyst



The Big Picture


Managed by collaboration

Managed by analyst



Data Confluence page

Data Production WebHome

Q🔔👤☰

[Space Shortcuts](#)
[Create Space Shortcuts Page ?](#)

[Recently Visited](#)
[Belle II Internal](#)
[Belle II](#)

[Belle II](#)

- ▼ Belle II
 - > B2TIP
 - Belle II Authorship
 - > Belle II Collaborative Services and Tools
 - > Belle II Group Pages
 - ▼ Belle II Internal
 - > Archive WebHome
 - > Computing Steering Group
 - > Computing WebHome
 - ▼ Data Production WebHome
 - > Collection summary
 - > Data main page
 - > Data Production Analysis
 - Validation_Draft
 - > Data Production Calibration main page
 - Data Production Operation WebPage
 - Data Production service Task list
 - > Data production status
 - Data Production Validation Page
 - > Data production WebHome - OLD
 - HLT skim expert page - NEW DRAFT
 - > MC main page
 - > Offline Luminosity Page
 - Public Datasets Task Force
 - Review of /dataprod disk at KEKCC
 - > Skim main page

🏠 / Belle II / Belle II Internal / Data Production WebHome

Data Production WebHome

Last modified by [Andreas Gellrich](#) on 2024/09/16 11:40

[Edit](#) [+ Create](#) [⋮](#)

Welcome to the **Data production** confluence page.
Here you will find all the **official information** about the available Data and MC samples.

[Data Production Status](#)

MC main page	Data main page	Skim main page
Luminosity main page	Validation page	Calibration page
DP repository	Special Processing page	Operation

Service tasks
Help Data Production in limited and well defined task to earn you authorship in Belle II.

- [Data Production service Task list](#)

Data production, what's new



Data Confluence page

Data Production WebHome

- Guidelines on Belle II Talks and Posters
- Life WebHome
- > Main WebHome
- > Operations WebHome
- > Organization WebHome
- > Physics Performance Webhome
- > Physics WebHome
- > Publications WebHome
- > Social WebHome
- > Software
- Sustainability
- > Belle II Membership Management System
- > Belle II Newcomers
- > Belle II Public
- > Belle II Registration and Account FAQ
- Belle II Support Contacts
- BelleII Masterclass exercises
- DESY Intranet Access
- DESY Multi-Factor Authentication
- DESY Requirements for Account Names and Passwords
- Grid User Certificates and VO membership
- Group meeting slides
- > How-to articles
- Initialization of loading the collector firmware
- > Meeting notes
- VO Membership

Contact details for DP leadership

Who's who and contacts

Coordinators: @Stefano Lacaprara @gdujany (deputy) - Former: @tamponi @jbennett
Skim manager: @scavino @vberta @shdelamo
Calibration manager: @Renu Garg, @tamponi (deputy)
Validation manager: @charris @pecker
Data processing manager: @pgironel @watanuki (deputy),
MC processing manager,: @Giovanni Gaudino @Gaurav Sharma (deputy)
MC signal: @hanx @liwenzhe
Operation manager: @gdujany

DP leadership responsibilities are listed here. (To be updated)

Meetings and Mailing list

Mailing list: dataprod@belle2.org

Meetings: [meetings page](#).

Minutes: https://hackmd.io/0oKpJkveRhWdJ5uQ6y3_g

(2023) <https://hackmd.io/vGeJNbMqSV6F01cY3bVZ0w>

(2022) <https://hackmd.io/dbskL9vDQjeQ1PXuu-Zzog>

Shift Manuals

- Standard DP shift manual
- DC Expert shift manual

Data production liaisons

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

Group	Liaison
Semileptonic and Missing Energy Decays	@Tommy Martinov @charris
Radiative & Electroweak Penguin	@iprudiev
Time Dependent CP Violation	@Noah Brenny
Hadronic B decays	@Xiaodong Shi
Quarkonium	@liyong71 @zhulin
Charm	@Jaeyoung Kim
Tau	@naveen10 @Kenji Inami
Dark-sector and low multiplicity	@Gaurav Sharma
Performance	@lekaiyao
Upgrade	@benni

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

Technical pages of the DP sub-groups

Here you can find information relevant for DP members. Nothing needed by analysis should end up here

Other resources

Check [this page](#) for the description of the data formats (RAW, mDST, cDST, uDST) and their usage



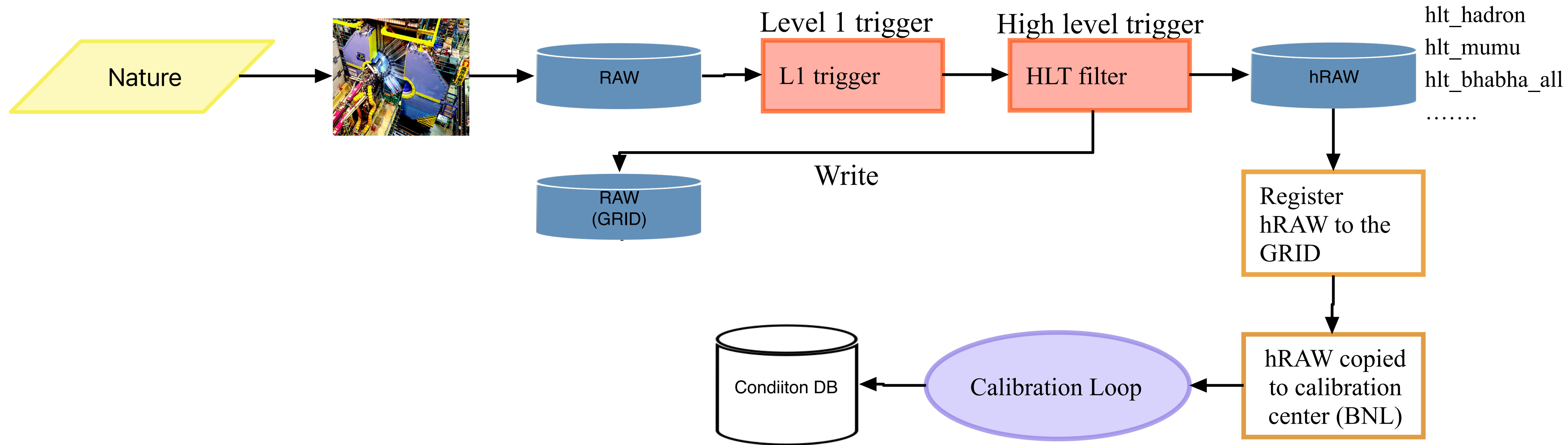
Data format

In general, Belle II output is stored in **ROOT** files containing various subsets of dataobjects, dbobjects, nTuples, etc

RAW:	un-processed, un-calibrated output of the detector
hRAW:	same as RAW, but only for events passing a given HLT filter or skim
DST	Data Summary Table - All available dataobjects are included (not generally used as it contains everything)
cDST:	calibration Data Summary Table - cDST contain RAW data and additional data-objects useful for calibration
mDST:	mini Data Summary Table - Controlled version of a DST. - only a subset of available processed data-objects are included - Flagged skim approach! - Use for most analysis (see below)
uDST:	user Data Summary Table - mDST objects plus analysis objects (e.g. particleLists) - produced from skims - reduce time needed to run analysis jobs - Samples created only for FEI based analysis!!



Data Flow



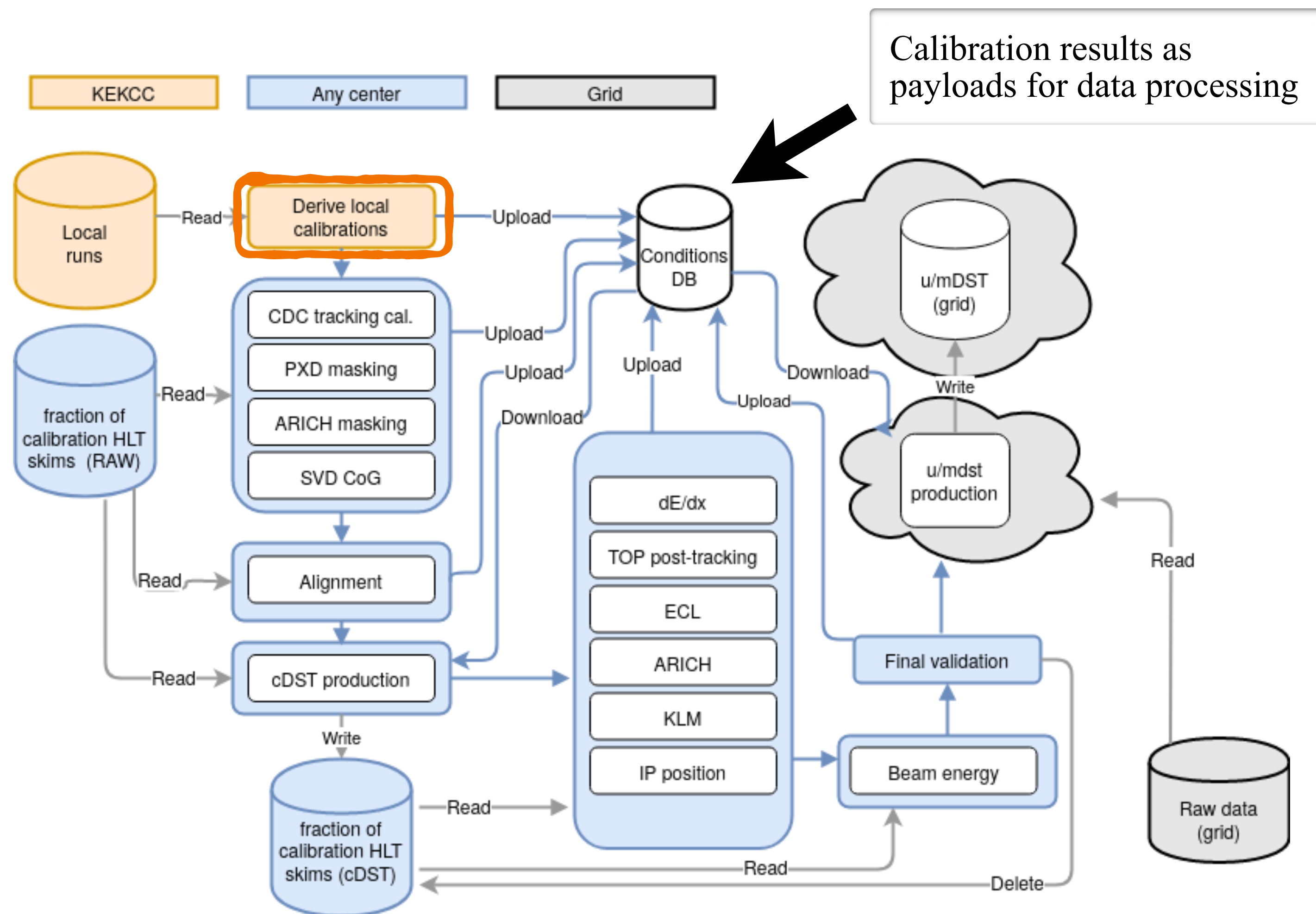
Calibration Loop



The **goal** of the calibration is to provide data usable for physics analysis

Full calibration loop divided into 5 steps:

- **Local calibration:** derived from local runs or DQM (e.g TOP laser calibration, SVD noise calibration)



Calibration results as payloads for data processing



Calibration Loop

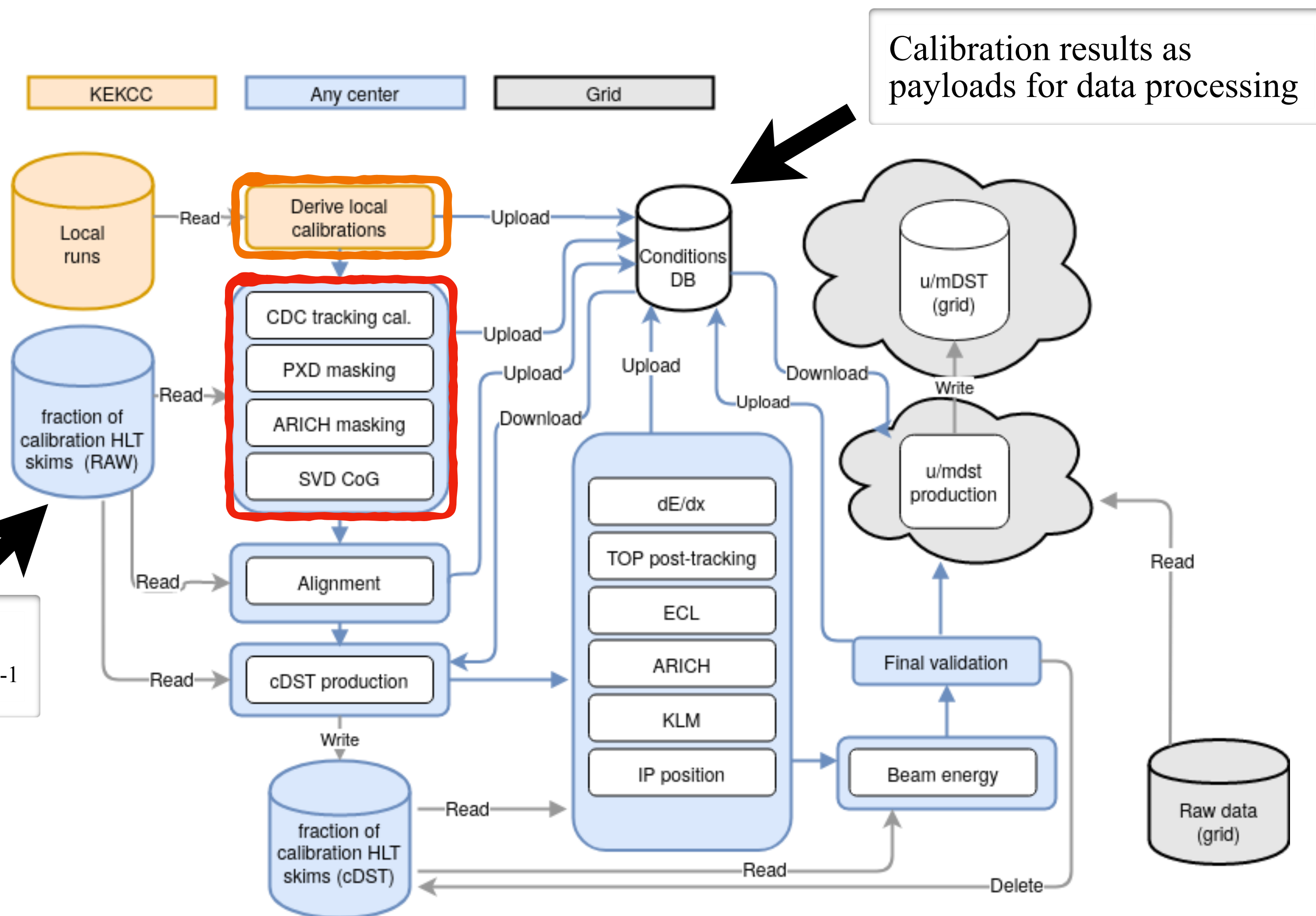


The **goal** of the calibration is to provide data usable for physics analysis

Full calibration loop divided into 5 steps:

- **Local calibration:** derived from local runs or DQM (e.g TOP laser calibration, SVD noise calibration)
- **Raw data based calibration:** must run on raw collision data (e.g channel masking, CDC tracking calibration)

Calibration skims (hRAW)
Adaptive pre-scaling to 12fb^{-1}



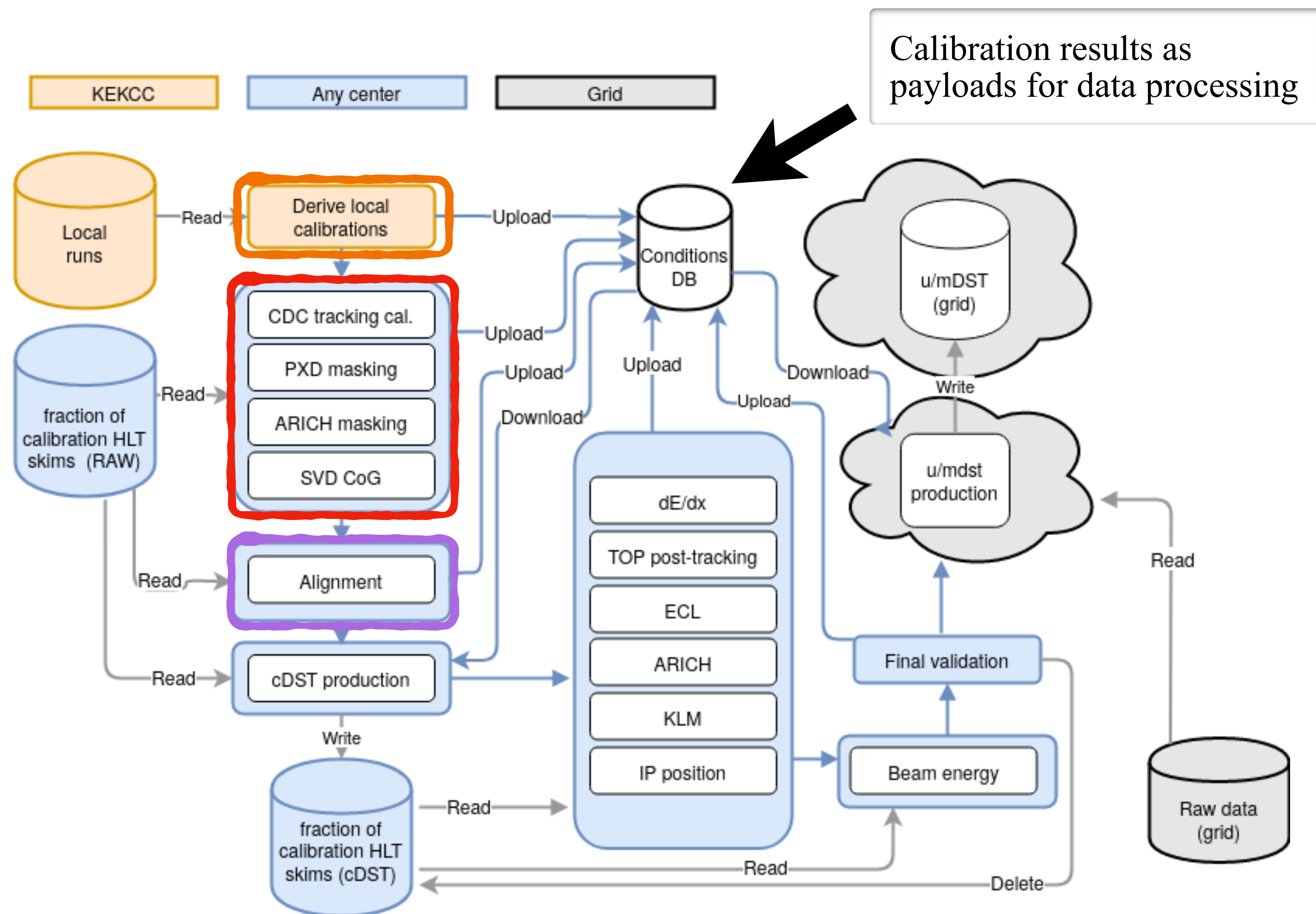
Calibration Loop



The **goal** of the calibration is to provide data usable for physics analysis

Full calibration loop divided into 5 steps:

- **Local calibration:** derived from local runs or DQM (e.g TOP laser calibration, SVD noise calibration)
- **Raw data based calibration:** must run on raw collision data (e.g channel masking, CDC tracking calibration)
- **Alignment:** requires raw collision data and both local and the other raw data based calibrations



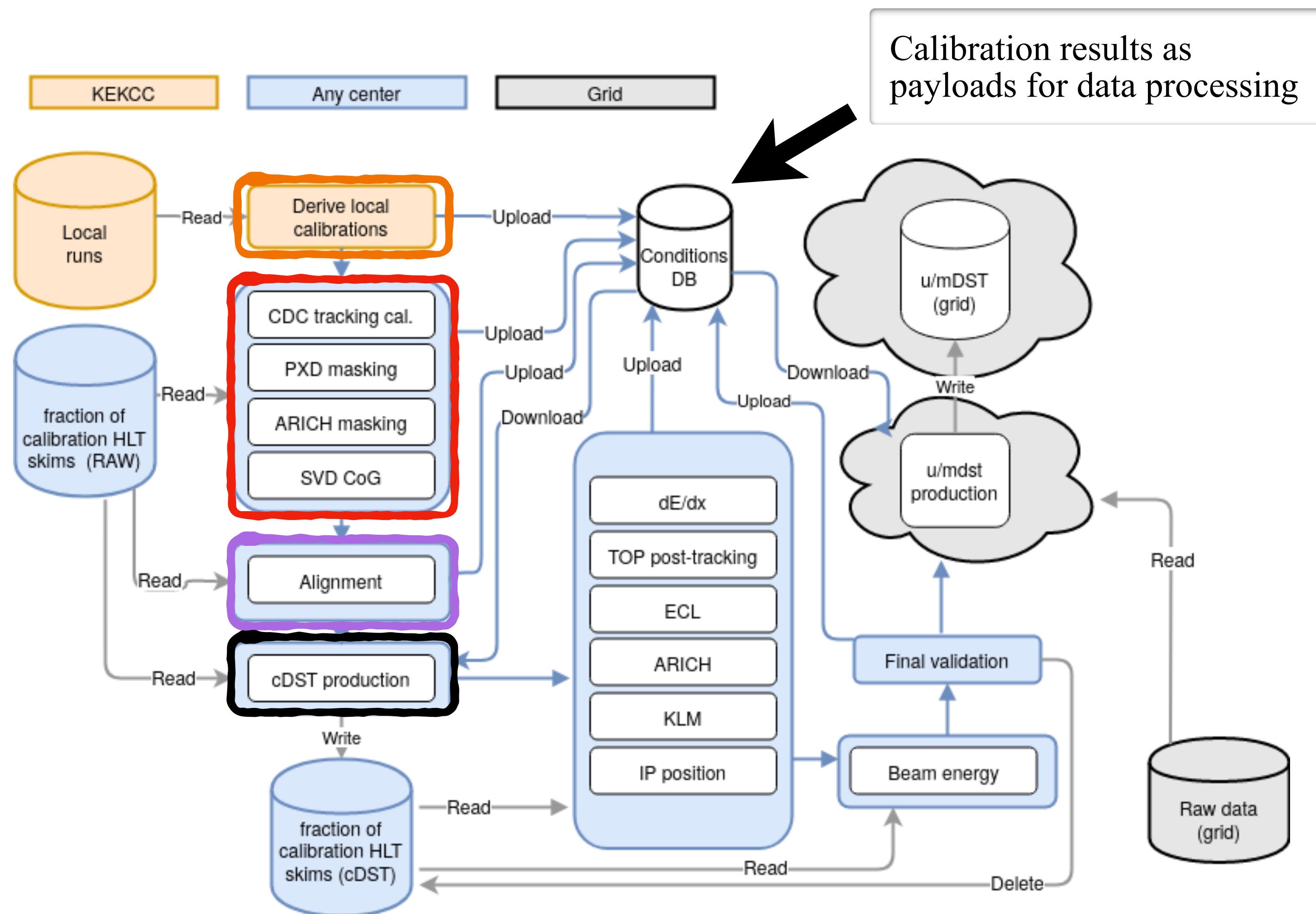
Calibration Loop



The **goal** of the calibration is to provide data usable for physics analysis

Full calibration loop divided into 5 steps:

- **Local calibration:** derived from local runs or DQM (e.g TOP laser calibration, SVD noise calibration)
- **Raw data based calibration:** must run on raw collision data (e.g channel masking, CDC tracking calibration)
- **Alignment:** requires raw collision data and both local and the other raw data based calibrations
- **cDST production**



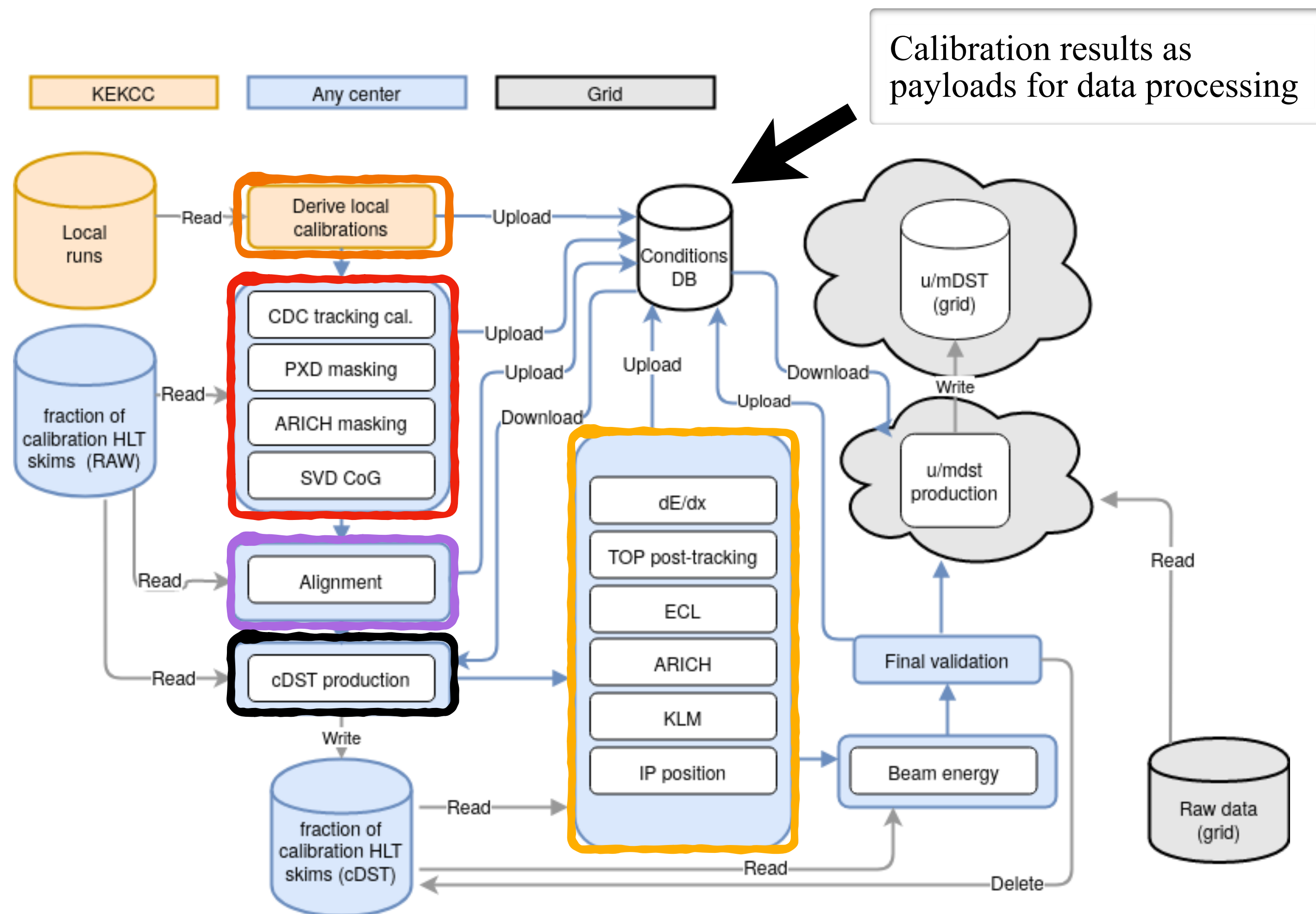
Calibration Loop



The **goal** of the calibration is to provide data usable for physics analysis

Full calibration loop divided into 5 steps:

- **Local calibration:** derived from local runs or DQM (e.g TOP laser calibration, SVD noise calibration)
- **Raw data based calibration:** must run on raw collision data (e.g channel masking, CDC tracking calibration)
- **Alignment:** requires raw collision data and both local and the other raw data based calibrations
- **cDST production**
- **Post-tracking calibration:** require good tracks, run on centrally produced cDST files (e.g CDC dE/dx)



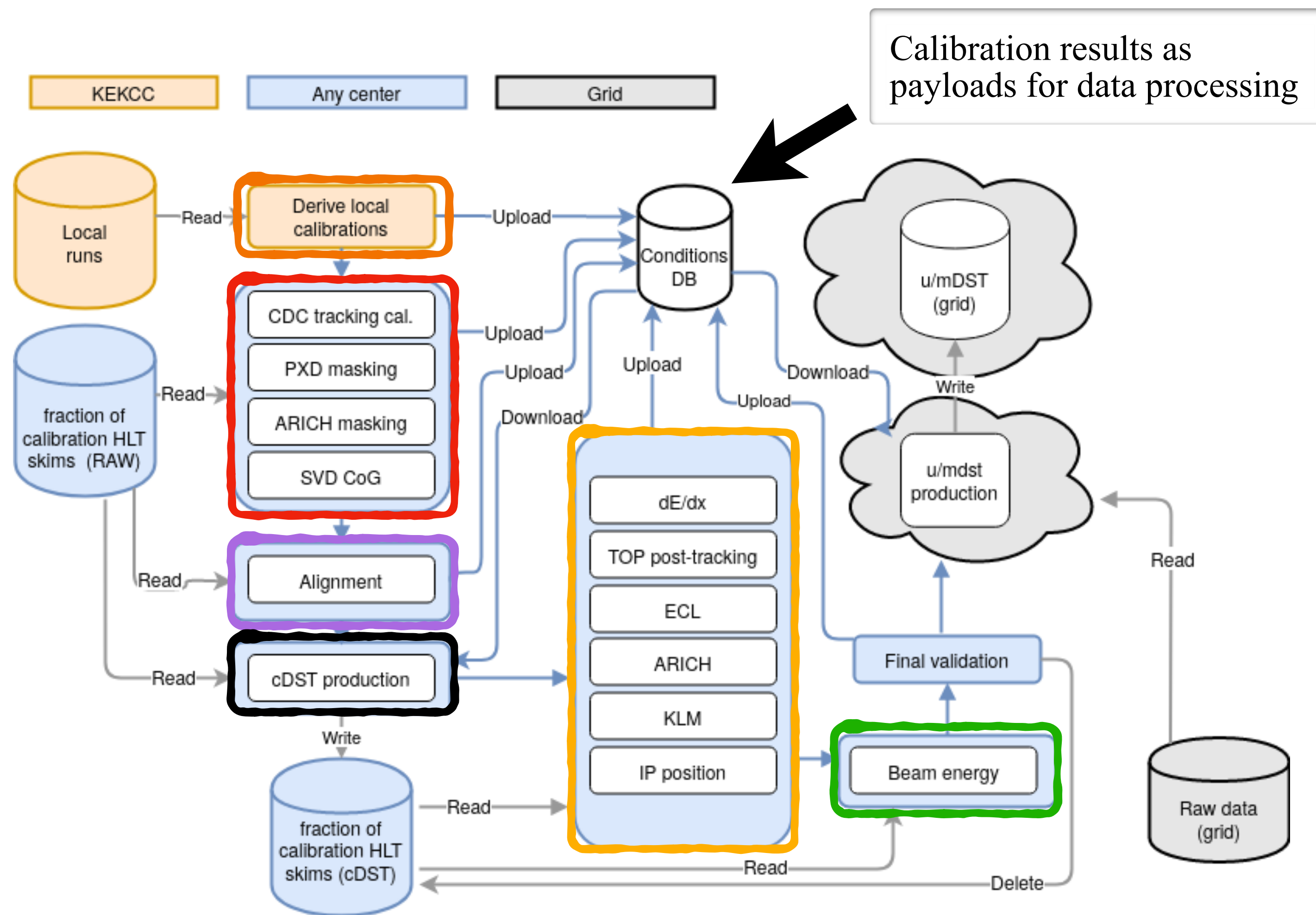
Calibration Loop



The **goal** of the calibration is to provide data usable for physics analysis

Full calibration loop divided into 5 steps:

- **Local calibration:** derived from local runs or DQM (e.g TOP laser calibration, SVD noise calibration)
- **Raw data based calibration:** must run on raw collision data (e.g channel masking, CDC tracking calibration)
- **Alignment:** requires raw collision data and both local and the other raw data based calibrations
- **cDST production**
- **Post-tracking calibration:** require good tracks, run on centrally produced cDST files (e.g CDC dE/dx)
- **Analysis based calibration:** rely on high quality data (e.g ECMS calibration)

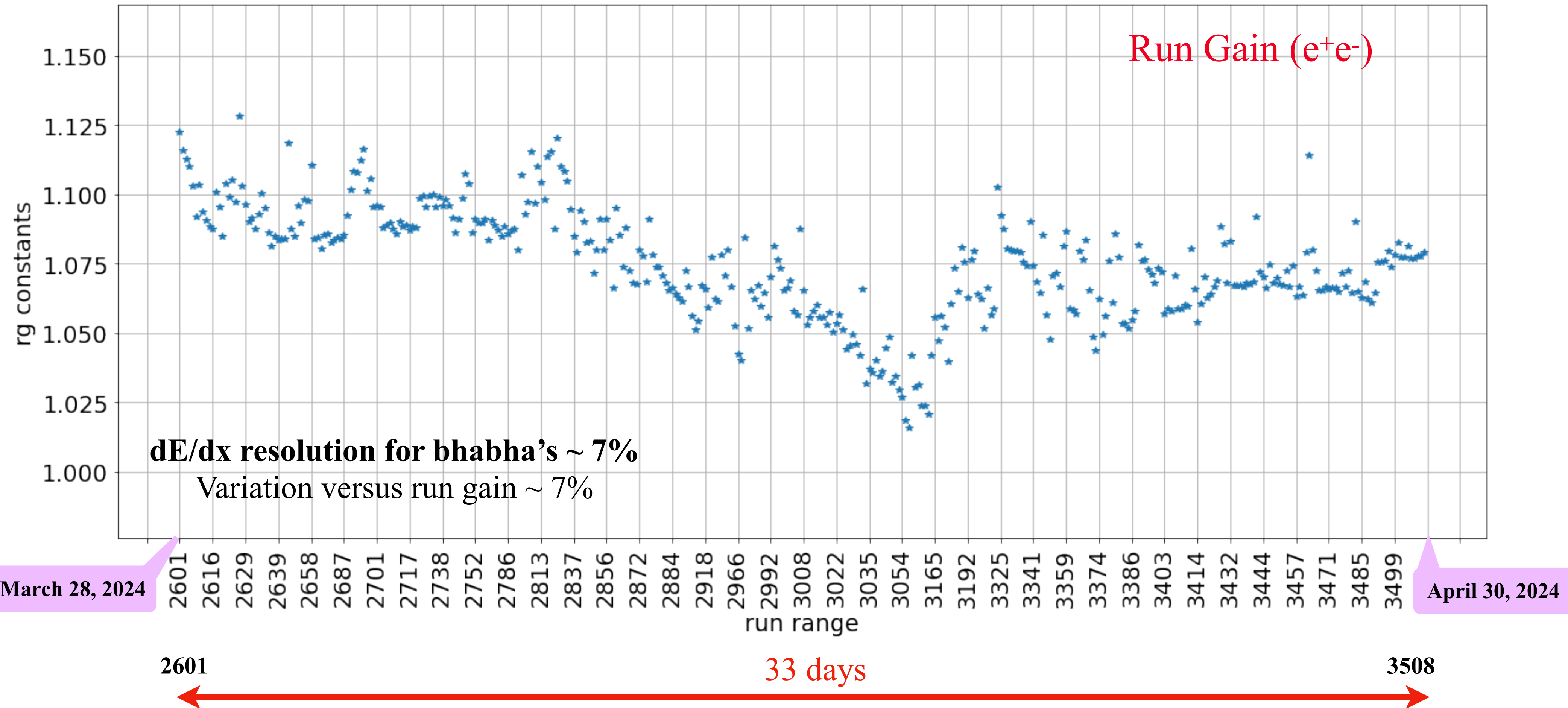


Calibration Example

Run gain is calculated as fit mean of dE/dx distribution from each run (bhabha tracks)



- **Run gain changes because instabilities of CDC**

- ▶ Gas composition (incl. water vapor)
- ▶ Temperature
- ▶ Pressure



Calibration Confluence page

Calibration page



Space Shortcuts

[Create Space Shortcuts Page ?](#)

Recently Visited

[Data Production Calibration main page](#)

Belle II

- ▼ Belle II
 - > B2TIP
 - Belle II Authorship
 - > Belle II Collaborative Services and Tools
 - > Belle II Group Pages
 - ▼ Belle II Internal
 - > Archive WebHome
 - > Computing Steering Group
 - > Computing WebHome
 - ▼ Data Production WebHome
 - > Collection summary
 - > Data main page
 - > Data Production Analysis Validation_Draft
 - ▼ Data Production Calibration main page
 - Proc16 logbook
 - Prompt Calibration bucket planning
 - **Prompt calibration bucket planning (pre-LS1)**

▼ / Belle II / Belle II Internal / Data Production WebHome / Data Production Calibration main page

Data Production Calibration main page

Last modified by [Stefano Lacaprra](#) on 2024/09/16 14:46

[Edit](#) [+ Create](#) [⋮](#)

- Important info
 - Calibration who's who
 - B2Notes, documentation, papers
 - Meetings, mailing lists, minutes, useful links
- General overview
 - The Belle II calibration strategy
 - Prompt calibration and reprocessing calibration
 - CAF and Airflow
- FOR EXPERTS ONLY
-
- Data size
- Implementation of the calibration loop
- Automatic calibration service
- Validation and sign-off
- Calibration FAQ
- Prompt calibration planning available here
 - [proc 13 re-calibration planning available here](#)
 - Older news:
 - [proc 11 instructions available here](#)
 - [proc 12 instructions available here](#)

Important info

Calibration who's who

Manager: [@Renu Garg](#)
Deputy: [@tamponi](#)

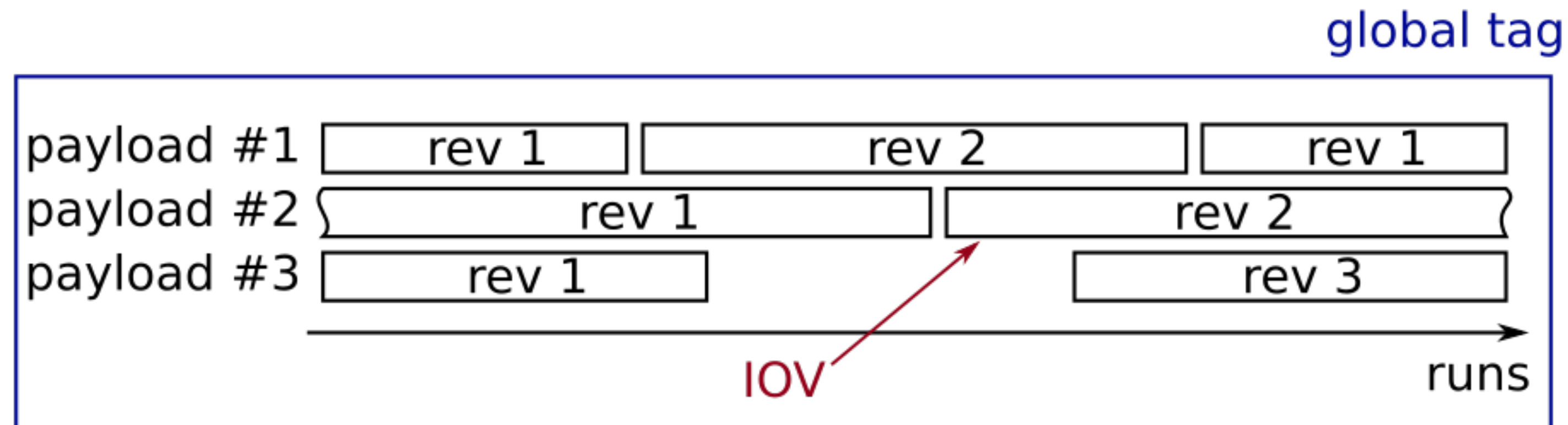
PXD expert: [@Maiko Takahashi](#)
SVD expert: [@qduiany](#)

Hot topics - What we have to work on :

1. Automate the calibration validation (CQM)
2. Enable calibrations on GRID



Global Tag



IoV

Interval of Validity

An experiment and run interval for which a payload is valid.

Payload

An atom of condition data (e.g CDCDedxRunGain)

Identified by name and revision number

ROOT format

Global tag

Collection of payloads and their IoVs

Has a unique name and a description

IoVs

- consists of four values:

$first_exp$, $first_run$, $final_exp$, $final_run$

e.g: 30, 24, 30, 140

Special case:

- $final_exp \geq 0 \ \&\& \ final_run < 0$:

valid for all runs in $final_exp$

e.g: 30, 0, 30, -1

- $final_exp < 0 \ \&\& \ final_run < 0$:

valid forever

e.g: 30, 0, -1, -1

Correct global tags are automatically selected during processing



Condition database: <https://cdbweb.sdcc.bnl.gov>

Questions? Write to [T1 CDB team](#)

[Home](#)
[GlobalTag](#)
[Payload](#)
[Types of Payload](#)
[Global Tag Comparison](#)

GlobalTag: 3194 items found Click on items for more details

ID: Name (can be partial): Status: Type: Modified by: items per page:

Name ▲	ID ▲	Default? ▲	Description ▲	Status ▲	Type ▲	Modified ▲	Mod. by ▲	Total Payloads ▲	Dist
mcrd_prompt_rel08	3308	⊖	Globaltag that contains the simulation payloads for MC16rd to supersede the ones in data_prompt_rel08 and online	TESTING	DEV	05/03/2024 10:12 a.m.	gdujany	0	0
mcrd_proc16	3307	⊖	Globaltag that contains the simulation payloads for MC16rd to supersede the ones in data_proc16 and online	TESTING	DEV	05/03/2024 8:28 a.m.	gdujany	1622	43
user_gaudino_mcrd_proc16	3306	⊖	staging MCrd GT for proc16 production	TESTING	DEV	05/03/2024 8:02 a.m.	gaudino	1622	43
temp_gdujany_beam_parameters_proc16_chunk1	3305	⊖	Beam parameters for proc16 chunk1	OPEN	DEV	04/29/2024 1:55 p.m.	gdujany	1566	1
neutrals_2024	3304	⊖	GT with photon energy bias correction variation and photon efficiency corrections obtained with preLS1-data and MC15rd	TESTING	DEV	04/26/2024 10:14 a.m.	eganiev	21	3
user_cwessel_ROISimulationParameters_update	3303	⊖	Special ROISimulationParameters for exp12 runs 1188 and 1189, c.f. gitlab issue 10473.	OPEN	DEV	04/25/2024 7:55 a.m.	cwessel	1	1
AIRFLOW_operation_staging_proc16_chunk2	3302	⊖	staging Gt for chunk2	OPEN	DEV	04/24/2024 2:37 p.m.	tamponi	0	0
user_takaham_pxd_exp1004_5deadL2modules	3301	⊖	user GT for PXD with PXDDeadPixelPar including 5 L2 modules which have been OFF since the beginning of Run2 physics run	TESTING	DEV	04/21/2024 12:27 p.m.	takaham	1	1
user_lacaprar_test3_software_trigger_cut	3300	⊖	A test GT to upload only software_trigger_cut from online	OPEN	DEV	04/19/2024 3:24 p.m.	lacaprar	108	108
user_lacaprar_test2_software_trigger_cut	3299	⊖	A test GT to upload only software_trigger_cut from online	OPEN	DEV	04/19/2024 3:23 p.m.	lacaprar	14	14
user_lacaprar_test_software_trigger_cut	3298	⊖	A test GT to upload only software_trigger_cut from online	TESTING	DEV	04/19/2024 3:11 p.m.	lacaprar	108	108
		⊖	Globaltag for storing payloads						



(Re)Processing scheme

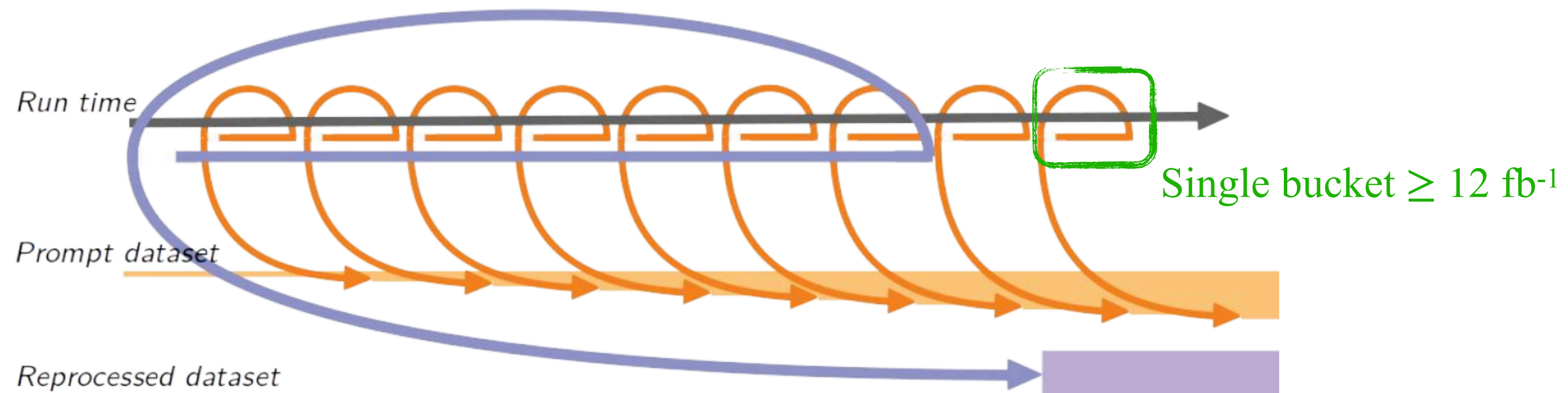
For any given data, two types of calibration and processing happens:

Prompt processing:

- Calibration @ BNL
- Minimum luminosity: 12 fb^{-1}
- Uses hRAW as input
- All calibrations included
- Terminology: **bucketXX**
- ~one bucket every 2 weeks of data taking
- Ideally already final calibration

Official reprocessing:

- Recalibration @ KEKCC
- To update older data with the latest reconstruction software
- Uses cDST as input
- Only calibration with expected improvement
- Terminology: **procXX**
- ~once/year until 2025, every other year starting from 2025



(Re)Processing scheme

For any given data, calibration and processing happens twice:

Prompt processing:

- Calibration @ BNL
- Minimum luminosity: 12 fb⁻¹
- Uses hR
- All calib
- Termino
- ~one bu
- Ideally a

Official reprocessing:

- Recalibration @ KEKCC

To update older data with the latest

Current data campaign:

Prompt processing:

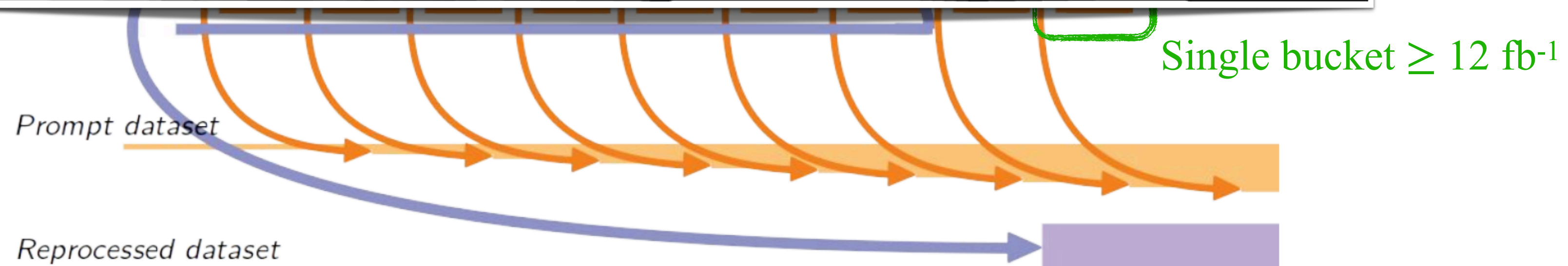
- **Bucket37-40**: experiment 30, 31, 32, 33
- Integrated luminosity: 103.63 fb⁻¹

Official processing: Done

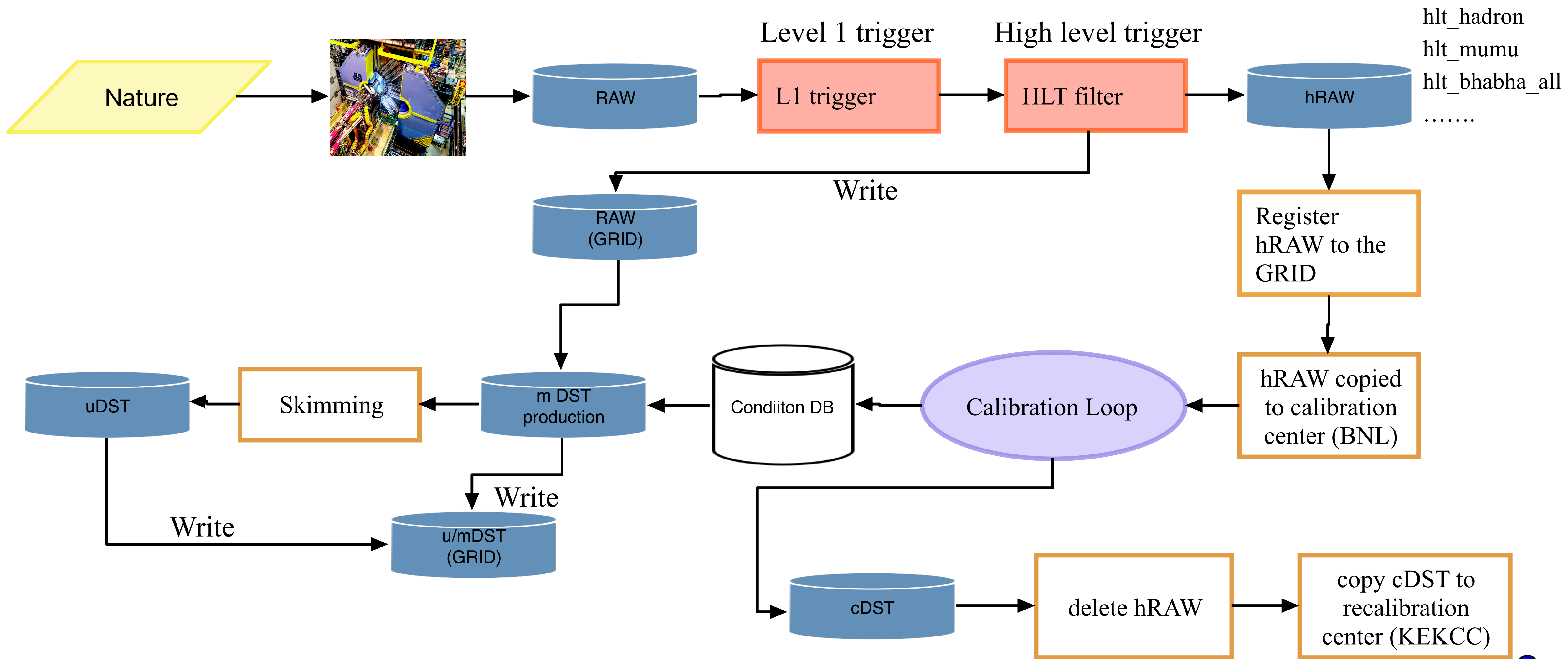
- **Proc16**: experiments 7-8, 10, 12, 14, 16-18, 20-22, 24-26
- Integrated luminosity: 442.8 fb⁻¹

improvement

per year starting



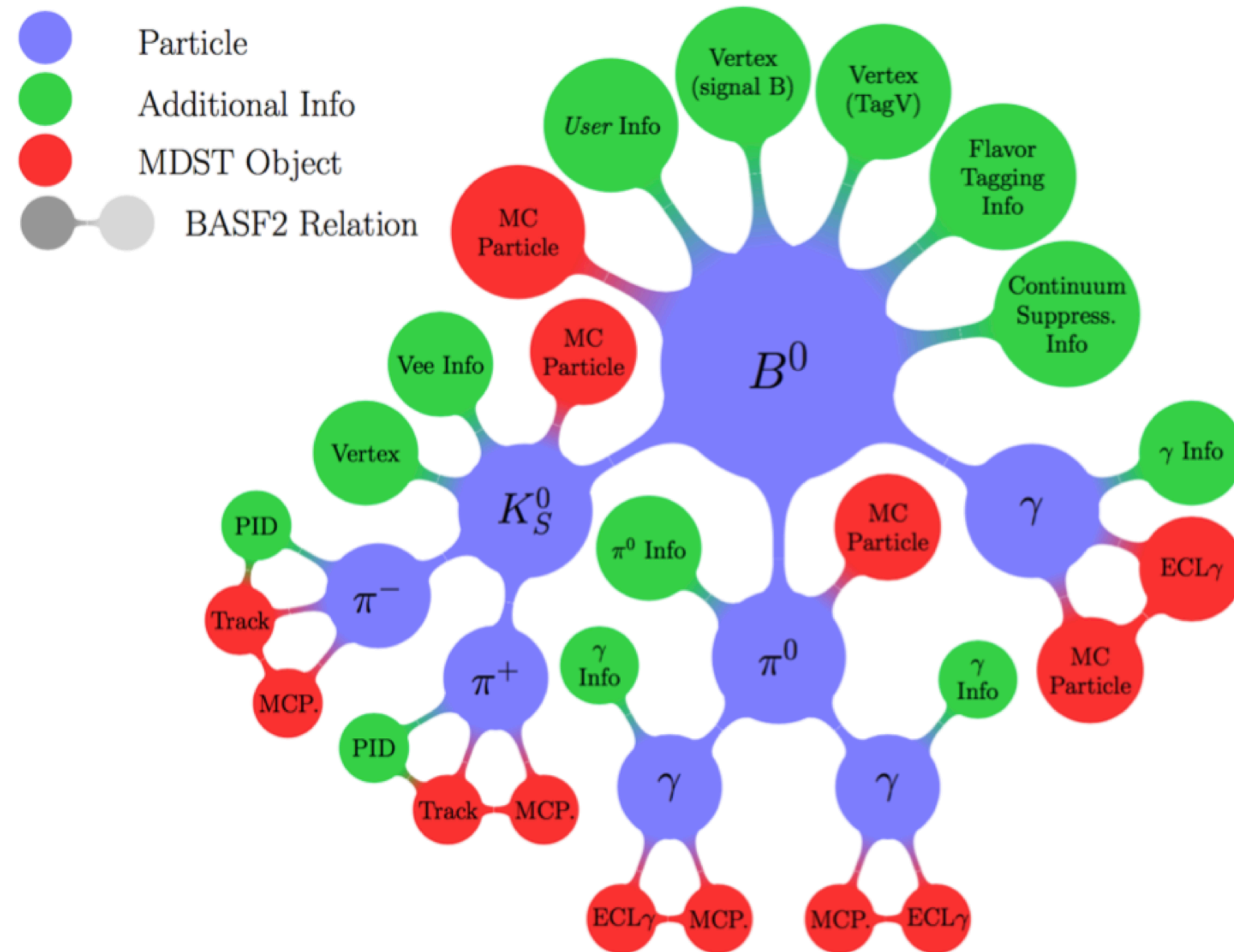
Data Flow



Analysis Skim

Skims are meant to provide analysis-oriented MC and data in reduced sizes

- Produced as uDST = mDST information + analysis-level information
 - Particle objects: which links particle hypotheses with tracks, neutral clusters, and particle identification information
 - Vertex fit results
 - Information from full B and D reconstruction, Continuum suppression and other complex algorithms
- That means more information in smaller files!
 - This allows for preprocessing that reduces the CPU requirements for the analysis
 - Your grid jobs on skimmed samples will finish much more quickly!



Use uDST for FEI based analysis!!



Flagged skim approach

Starting with proc16 and MC16 campaign a Flagged skim approach will be used

For more details:

[slides](#), [slides-1](#)

- **Merged** different skims into groups
- Add a **flag** to each event according to the skim, to identify the skim within the group
- Produce **mDST** output

- **Why?**

- Reduce the amount of work
- Exploit the large overlap between the skims
- Make submissions faster and lighter

- **How to use?**

- Need to add the flag selection when you make the ntuples

```
ma.inputMdstList(filelist=infile, path=path)
ma.applyEventCuts(
    'eventExtraInfo(passes_mySkim)==1',
    path=path)
```

Use flagged mDST for your analysis, if not FEI based!!

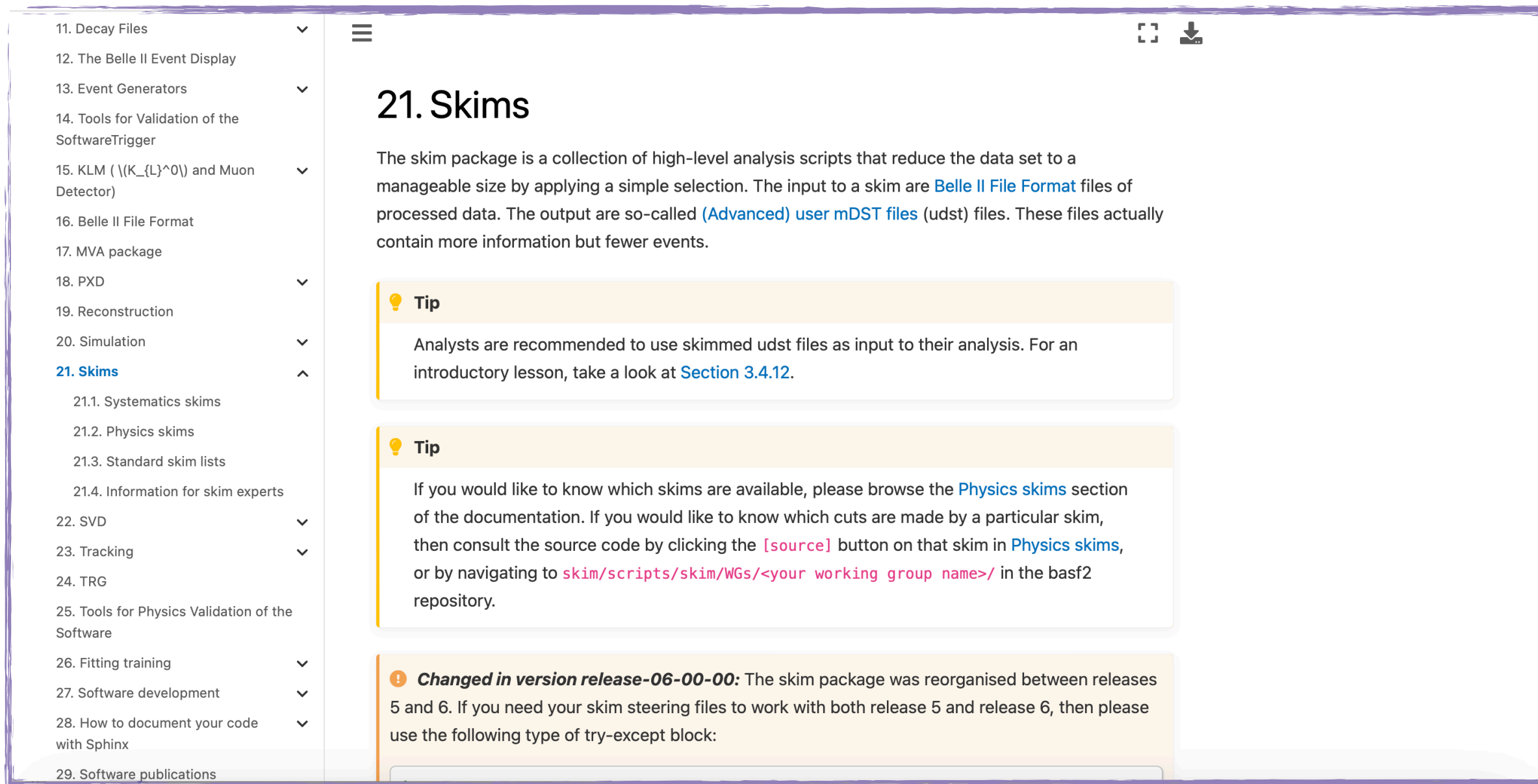


Understanding of Skimming

Two primary sources for information:

Documentation

- Good for general information and tutorials



The screenshot shows the Belle II Skims documentation page. The left sidebar contains a table of contents with sections 11 through 29. The main content area is titled "21. Skims" and contains the following text:

The skim package is a collection of high-level analysis scripts that reduce the data set to a manageable size by applying a simple selection. The input to a skim are [Belle II File Format](#) files of processed data. The output are so-called (Advanced) user mDST files (udst) files. These files actually contain more information but fewer events.

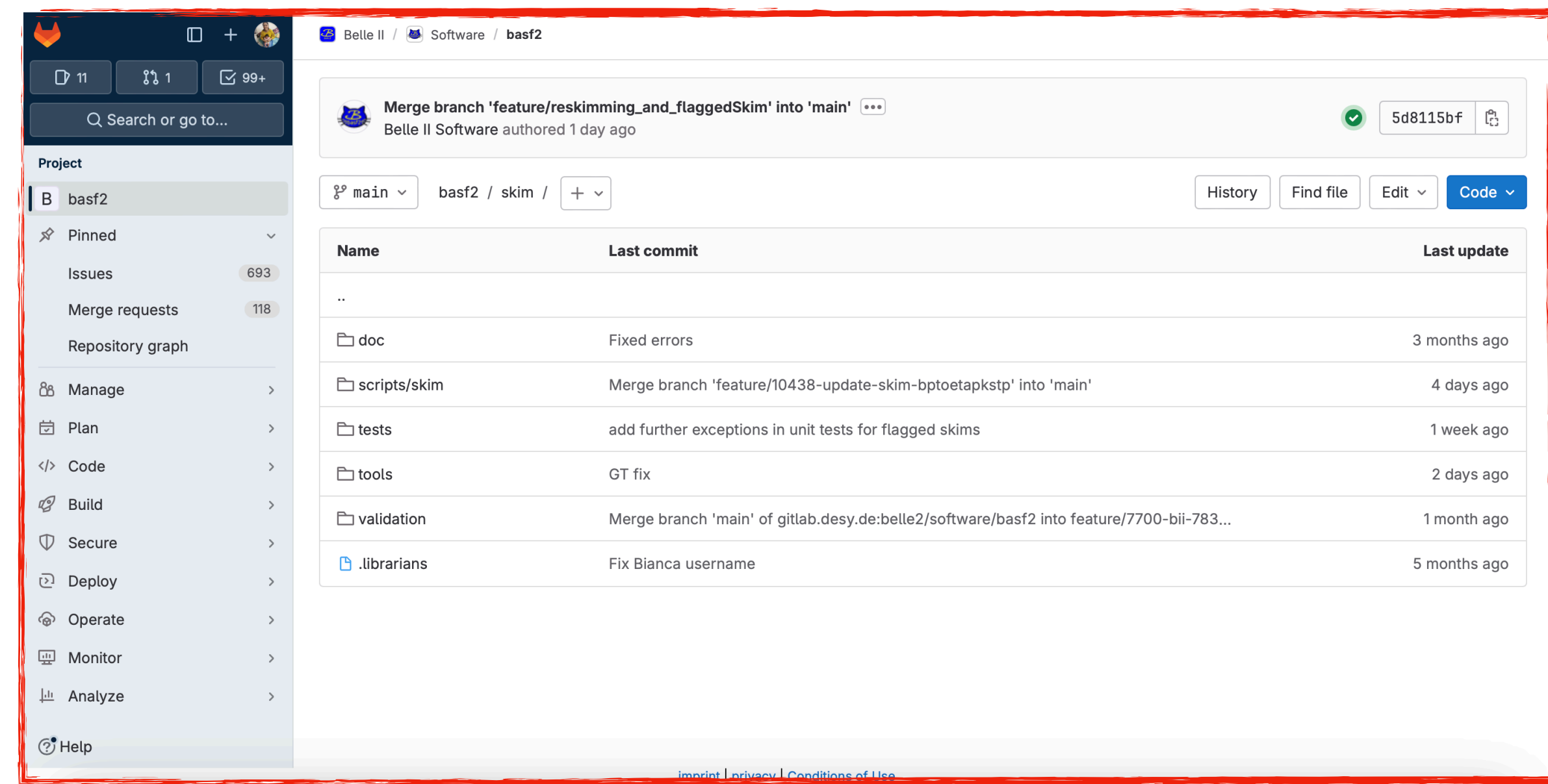
Tip
Analysts are recommended to use skimmed udst files as input to their analysis. For an introductory lesson, take a look at [Section 3.4.12](#).

Tip
If you would like to know which skims are available, please browse the [Physics skims](#) section of the documentation. If you would like to know which cuts are made by a particular skim, then consult the source code by clicking the [\[source\]](#) button on that skim in [Physics skims](#), or by navigating to `skim/scripts/skim/WGs/<your working group name>/` in the basf2 repository.

Changed in version release-06-00-00: The skim package was reorganised between releases 5 and 6. If you need your skim steering files to work with both release 5 and release 6, then please use the following type of try-except block:

Source code on gitlab

- Full information and always up to date



The screenshot shows the Belle II basf2 GitLab repository page. The top navigation bar includes the project name "Belle II / Software / basf2" and a search bar. The main content area displays a merge request for the branch "feature/reskimming_and_flaggedSkim" into "main", authored by Belle II Software 1 day ago. Below the merge request, there is a table of repository files and their last commit information:

Name	Last commit	Last update
..		
doc	Fixed errors	3 months ago
scripts/skim	Merge branch 'feature/10438-update-skim-bptoetapkstp' into 'main'	4 days ago
tests	add further exceptions in unit tests for flagged skims	1 week ago
tools	GT fix	2 days ago
validation	Merge branch 'main' of gitlab.desy.de:belle2/software/basf2 into feature/7700-bii-783...	1 month ago
.librarians	Fix Bianca username	5 months ago

It is important for analysts to understand the skim they are using to make sure that it is aligned with their analysis goals.



Finding your Skim

Three ways to find your skims:

1. Dataset searcher on DIRAC (web version)

- Provide a list of LFNs to use with gbasf2 analysis
- <https://dirac.cc.kek.jp:8443/DIRAC/>
- DIRAC apps → Dataset searcher

2. Dataset searcher via command line (gbasf2 environment)

- `source /cvmfs/belle.kek.jp/grid/gbasf2/pro/bashrc`
- `gb2_ds_search dataset --data_type data --campaign proc13 --exp_low 17 --exp_high 17 --general_skim hadron`

The screenshot shows the 'Dataset Searcher' web interface. It has a search form with the following fields and values:

- Data Type: MC Data
- Background level: BGx1 BGx0 Other
- Background level: [dropdown]
- Beam Energies: 4S
- Data Levels: [dropdown]
- Global Tags: [dropdown]
- Experiment High: [input]
- Run High: [input]
- General Skim Names: [dropdown]
- Campaigns: MC15rd_a
- Skim Types: [dropdown]
- Releases: [dropdown]
- Experiment Low: [input]
- Run Low: [input]
- MC Event Types: [dropdown]
- productionId: [input]

Below the search form, there is a table of LFNs:

LPN
/belle/MC/release-06-01-08/DB00002498/MC15rd_a/prod00027580/s00/e0024/4S/r00790/luubar/mdst
/belle/MC/release-06-01-08/DB00002498/MC15rd_a/prod00027580/s00/e0024/4S/r00791/luubar/mdst
/belle/MC/release-06-01-08/DB00002498/MC15rd_a/prod00027580/s00/e0024/4S/r00792/luubar/mdst

At the bottom, there are buttons for 'Dataset LFNs Metadata', 'Dataset Metadata', and 'Download .txt file'.

```
(base) renu@cw01:validation$ gb2_ds_search dataset --data_type data --campaign proc13 --exp_low 17 --exp_high 17 --general_skim hadron
142 matching datasets found:
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00102/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00103/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00104/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00105/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00106/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00107/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00108/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00112/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00116/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00117/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00118/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00119/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00121/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00125/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00132/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00145/hadron/mdst
/belle/Data/release-06-00-12/DB00002392/proc13/prod00026983/e0017/4S/r00146/hadron/mdst
```



Finding your Skim

3. Collections

- The easiest way to process data or MC as an analyst!
- Contains the full list of LPNs for a given dataset
- Ensures you use the correct files and don't miss any
- Available for skims (currently only by request...)

<https://gbasf2.belle2.org/collectionSearcher.html>
[Collection Summary](#), [Data collection](#), [MC collection](#)

The screenshot shows the 'Data main page' interface. On the left is a sidebar with navigation links like 'Collection summary', 'MC main page', and 'Data main page'. The main content area has a title 'Data main page' and a list of 'Conference collections' including 'Reprocessings' (with sub-items like proc16, proc13, proc12, proc11) and 'Prompt Processing'. Below this is a section for 'Conference collections' with a note about offline luminosity. At the bottom is a table with columns: Dataset description, Type, Beam energy / hit skims, Collections, Composition, Offline Luminosity (pb), and Notes.

Dataset description	Type	Beam energy / hit skims	Collections	Composition	Offline Luminosity (pb)	Notes
Moriond 2022 dataset	All	4S	/belle/collection/Data/Moriond2022_all_4S_v1	proc12 (exp7,8,10,12)+ prompt (exp 14,16,17,18)	189895.70±13.15	
		4S_offres	/belle/collection/Data/Moriond2022_all_4Soffres_v1	proc12 (exp8, 12) + prompt (exp 18)	17990.6 ± 3.6 (offres)	
	Hadron	4S	/belle/collection/Data/Moriond2022_hadron_4S_v1	proc12 (exp7,8,10,12)+ prompt (exp 14,16,17,18)	189273.70±13.15	



Skim Confluence page

Skim page

DESY

🏠 / Belle II / Belle II Internal / Data Production WebHome / Skim main page

Space Shortcuts

[Create Space Shortcuts Page ?](#)

Recently Visited

- [Data Production WebHome](#)
- [Belle II Internal](#)
- [Belle II](#)
- [Data Production Calibration main page](#)

Belle II

- ▾ Belle II
 - > B2TIP
 - Belle II Authorship
 - > Belle II Collaborative Services and Tools
 - > Belle II Group Pages
 - ▾ Belle II Internal
 - > Archive WebHome
 - > Computing Steering Group
 - > Computing WebHome
 - ▾ Data Production WebHome
 - > Collection summary
 - > Data main page
 - > Data Production Analysis
 - Validation_Draft
 - > Data Production Calibration main page
 - Data Production Operation WebPage
 - Data Production service Task list

Skim main page

Last modified by scavino on 2024/07/02 15:50

- [Welcome to the Skimming Confluence Page!](#)
 - [First things first. What are skims?](#)
 - [Types of skims and data types](#)
 - [More information](#)
- [Navigation of Skim Confluence Pages](#)
 - [Skim Information for Analysts](#)
 - [Skim Production Status](#)
 - [Skim Expert Page](#)
 - [Skim Meetings](#)
- [Mailing list](#)
- [Skimming Personnel](#)
- [Liaisons](#)

Welcome to the Skimming Confluence Page!

First things first. What are skims?

At Belle II, we collect a lot of data. As an analyst, when you run over this data, it is not feasible to use the entire dataset as input. It would simply take too long and use too many resources. For this reason, we must reduce the size of the dataset by applying a set of high-level cuts, keeping only events that are relevant to your analysis. You can then run your analysis on this "skimmed" dataset, significantly improving run times and resource consumption. This keeps you happy, and our friends at Distributed Computing happy. Win-win.

Types of skims and data types

General skims: all and hadron. These are high level skims, where "all" represents unskimmed data and "hadron" represents hlt_hadron skimmed data. Hadron is ~10% of all data, so it is preferred to use this. These are produced in **mdst** file format.

Analysis skims: As the name implies, these are analysis level skims, for which there are many (~80) and they are divided by working group. These are run on either all or hadron events, and can be tailored to suit particular analysis needs. These are produced in **udst** file format, meaning particle lists are created and saved in the output. It is strongly recommended for analysts to use these skims, and if there isn't one available that fits your needs, to contact your WG skim liaison and create a new one or modify an existing one.

More information

Mailing list

Join this mailing list by clicking the link and hitting subscribe. It is intended for skim experts, not analysts. All skim information relevant to analysts will be propagated by liaisons to their own WG mailing lists and at WG meetings, or to the physics mailing list if it is a large announcement.

dataprod-skim@belle2.org

You can also join the Data Production mailing list below for more general DP communications.

dataprod@belle2.org

Skimming Personnel

Role	Member
Coordination	@scavino
Development	@vberta
Operation	@shdelamo (beginning mid 2024)



MC production

Two types of MC production:

- **Signal MC:**

- Specific to your own analysis
- **Dec files:** specify your own dec file according to the dec file naming rules
- Contact the **Data production liaison** in your working group to get started!

- **Generic MC:**

- Produced automatically at every MC production campaigns
- They are just the typical processes which we expect to see at Belle II, such as:

$$e^+e^- \rightarrow \Upsilon(4S) \rightarrow B^+B^- \text{ (charged), } B^0B^0 \text{ (mixed)}$$

$$e^+e^- \rightarrow u\bar{u}, d\bar{d}, c\bar{c}, s\bar{s}$$

$$e^+e^- \rightarrow \gamma\gamma, e^+e^-, \mu^+\mu^-, \tau^+\tau^- \text{ (taupair)}$$

$$e^+e^- \rightarrow \ell\ell XX \text{ (} ee\pi\pi, eepp, \text{ etc.)}, hh\text{ISR (}\pi\pi\text{ISR, } KK\text{ISR, etc.)}$$

- Generated based on central decay file: [DECAY_BELLE2.DEC](#)

Data production liaisons

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

Group	Liaison
Semileptonic and Missing Energy Decays	@Tommy Martinov @charris
Radiative & Electroweak Penguin	@iprudiev
Time Dependent CP Violation	@Noah Brenny
Hadronic B decays	@Xiaodong Shi
Quarkonium	@liyang71 @zhulin
Charm	@Jaeyoung Kim
Tau	@naveen10 @Kenji Inami
Dark-sector and low multiplicity	@Gaurav Sharma
Performance	@lekaiyao
Upgrade	@benni



MC production campaign

Two types of MC production campaign:

Run-independent (RI)

- Easier to produce
- Use simulated background and static detector conditions
- Produced in predetermined luminosity
- Less accurate detector performance and beam backgrounds.
- Terminology: **MC16ri_X**

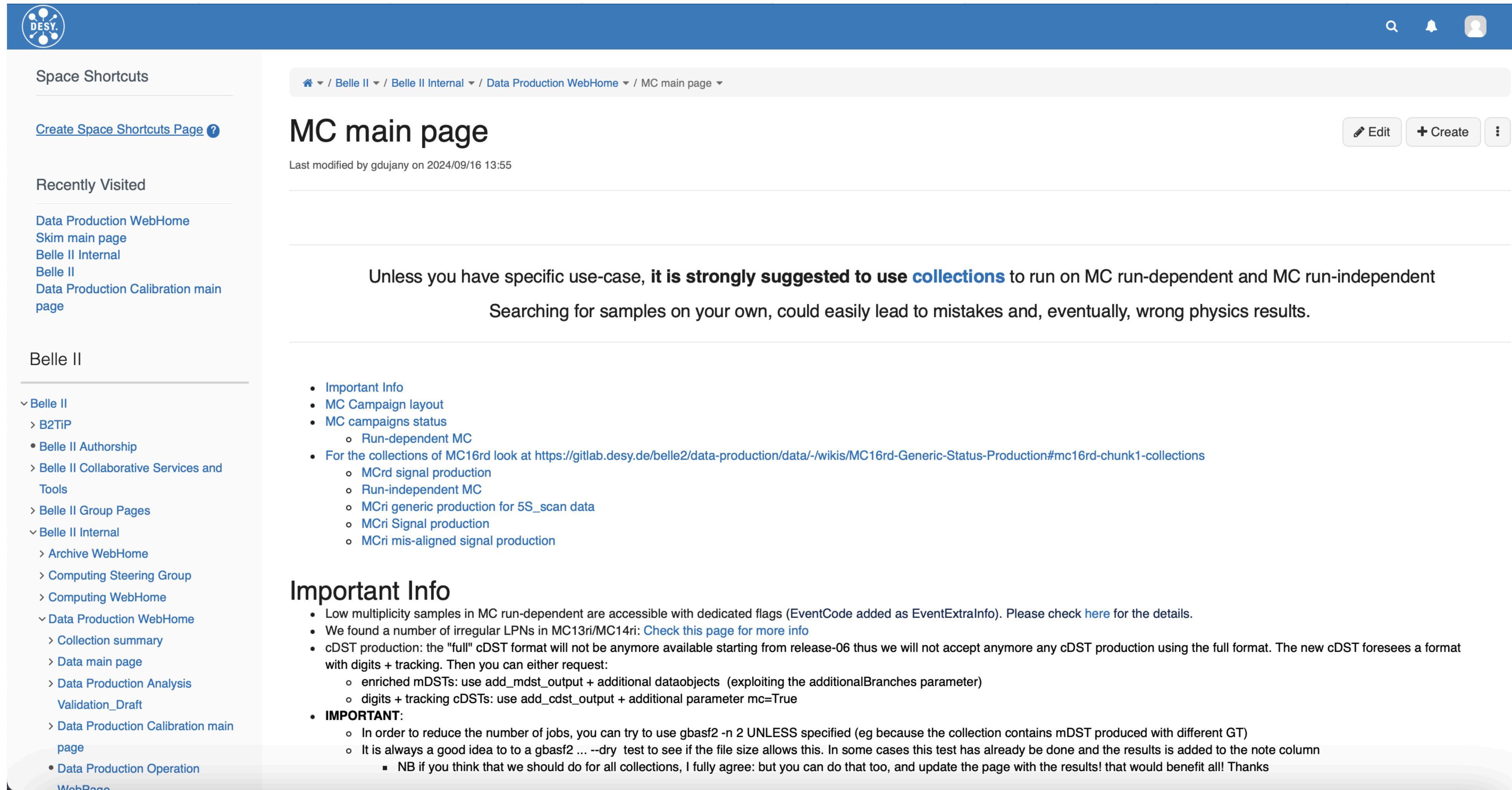
Run-dependent (RD)

- Difficult to produce
- Use random trigger events from data and real conditions
- Produced in streams (1 stream = luminosity of corresponding data)
- More accurate detector performance and beam backgrounds
- Terminology: **MC16rd_X**



MC Confluence page

MC page



The screenshot shows a Confluence page for 'MC main page' within the 'Belle II' space. The page header includes the DESY logo and navigation icons. The breadcrumb trail is: Home / Belle II / Belle II Internal / Data Production WebHome / MC main page. The page title is 'MC main page', last modified by gdujany on 2024/09/16 13:55. The main content area contains a warning: 'Unless you have specific use-case, it is strongly suggested to use **collections** to run on MC run-dependent and MC run-independent. Searching for samples on your own, could easily lead to mistakes and, eventually, wrong physics results.' Below this is a list of links: 'Important Info', 'MC Campaign layout', 'MC campaigns status' (with sub-links for 'Run-dependent MC'), and 'For the collections of MC16rd look at https://gitlab.desy.de/belle2/data-production/data/-/wikis/MC16rd-Generic-Status-Production#mc16rd-chunk1-collections' (with sub-links for 'MCrd signal production', 'Run-independent MC', 'MCri generic production for 5S_scan data', 'MCri Signal production', and 'MCri mis-aligned signal production'). A section titled 'Important Info' contains several bullet points regarding MC run-dependent samples, LPNs, cDST production changes, and a note about using 'gbasf2 -n 2' unless specified, with a request for updates to the page.

Space Shortcuts

[Create Space Shortcuts Page ?](#)

Recently Visited

- Data Production WebHome
- Skim main page
- Belle II Internal
- Belle II
- Data Production Calibration main page

Belle II

- ▼ Belle II
 - > B2TiP
 - Belle II Authorship
 - > Belle II Collaborative Services and Tools
 - > Belle II Group Pages
 - ▼ Belle II Internal
 - > Archive WebHome
 - > Computing Steering Group
 - > Computing WebHome
 - ▼ Data Production WebHome
 - > Collection summary
 - > Data main page
 - > Data Production Analysis
 - Validation_Draft
 - > Data Production Calibration main page
 - Data Production Operation
 - WebPage

Home / Belle II / Belle II Internal / Data Production WebHome / MC main page

MC main page

Last modified by gdujany on 2024/09/16 13:55

Unless you have specific use-case, it is strongly suggested to use **collections** to run on MC run-dependent and MC run-independent. Searching for samples on your own, could easily lead to mistakes and, eventually, wrong physics results.

- [Important Info](#)
- [MC Campaign layout](#)
- [MC campaigns status](#)
 - [Run-dependent MC](#)
- [For the collections of MC16rd look at https://gitlab.desy.de/belle2/data-production/data/-/wikis/MC16rd-Generic-Status-Production#mc16rd-chunk1-collections](https://gitlab.desy.de/belle2/data-production/data/-/wikis/MC16rd-Generic-Status-Production#mc16rd-chunk1-collections)
 - [MCrd signal production](#)
 - [Run-independent MC](#)
 - [MCri generic production for 5S_scan data](#)
 - [MCri Signal production](#)
 - [MCri mis-aligned signal production](#)

Important Info

- Low multiplicity samples in MC run-dependent are accessible with dedicated flags (EventCode added as EventExtraInfo). Please check [here](#) for the details.
- We found a number of irregular LPNs in MC13ri/MC14ri: [Check this page for more info](#)
- cDST production: the "full" cDST format will not be anymore available starting from release-06 thus we will not accept anymore any cDST production using the full format. The new cDST foresees a format with digits + tracking. Then you can either request:
 - enriched mDSTs: use `add_mdst_output` + additional dataobjects (exploiting the `additionalBranches` parameter)
 - digits + tracking cDSTs: use `add_cdst_output` + additional parameter `mc=True`
- **IMPORTANT:**
 - In order to reduce the number of jobs, you can try to use `gbasf2 -n 2` UNLESS specified (eg because the collection contains mDST produced with different GT)
 - It is always a good idea to to a `gbasf2 ... --dry` test to see if the file size allows this. In some cases this test has already be done and the results is added to the note column
 - NB if you think that we should do for all collections, I fully agree: but you can do that too, and update the page with the results! that would benefit all! Thanks



More great resources

- <https://www.belle2.org/>
- **chat:** <https://chat.belle2.org>
- **questions:** <https://questions.belle2.org>
- **sympa (email lists):** <https://lists.belle2.org/sympa/home>
- **Mailing list:** dataprod@belle2.org, dataprod-skim@belle2.org, software-calibration@belle2.org
- **Basf2 documentation (Sphinx):** <https://software.belle2.org/>
- **gbasf2:** <https://confluence.desy.de/display/BI/Computing+GBasf2>
- <https://confluence.desy.de/display/BI/Instructions+for+gbasf2+analysis>
- **Data production:** <https://confluence.desy.de/display/BI/Data+production+WebHome>
- **Gitlab:** <https://gitlab.desy.de/belle2/>
- **Experiment Numbering:** <https://confluence.desy.de/display/BI/Experiment+numbering>
- **Conditions Database:** <https://cdbweb.sdcc.bnl.gov/> (globaltag information)
- **DIRAC (for dataset searcher):** <https://dirac.cc.kek.jp:8443/DIRAC/>

