

What we have discussed at Belle II computing workshop

2011. 6. 9

Kihyeon Cho

Contents

- Draft for JKPS
- HLT and Prompt Reco
- Definition of basf
 - => Registration tool kit
- Snap shot vs. query
 - => User created data set

Draft for JKPS (Journal of the Korean Physical Society)

- https://hep.kisti.re.kr/cho/metadata_grid_cho.pdf
- Comments by June 25
- Submit it on June 26

The embedment of metadata system at grid farms at the Belle II experiment

S. Ahn, J.H. Kim, T. Huh, S. Hwang, K. Cho*, H. Jang, B. K. Kim, H. Yoon and J. Yu

Korea Institute of Science and Technology Information, Daejeon 305-806, Korea

Z. Drasal

Charles University, Prague 116 36, Czech

T. Hara, Y. Iida, R. Itoh, G. Iwai, N. Katayama, Y. Kawai, S. Nishida, T. Sasaki and Y.

Watase

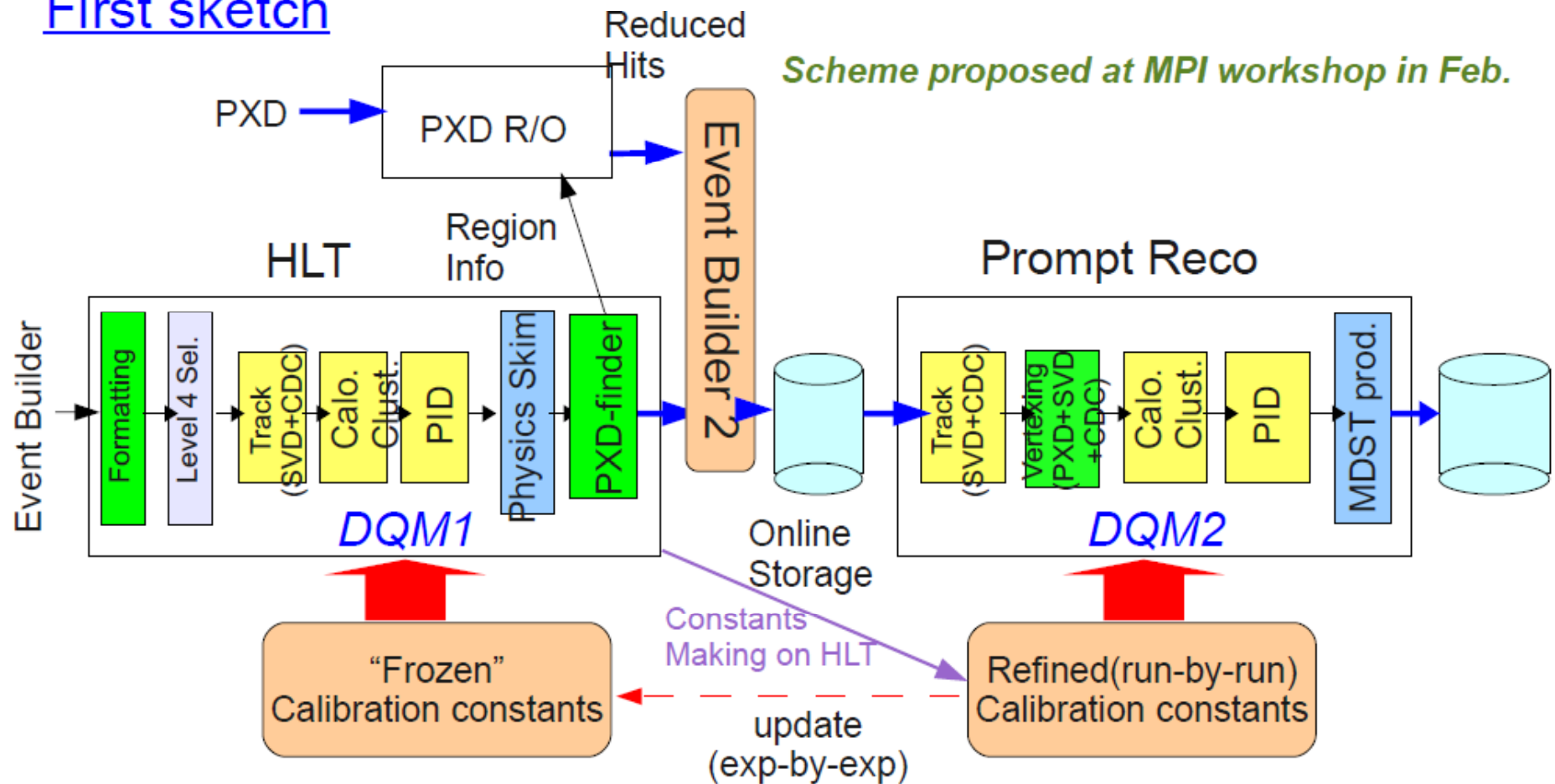
High Energy Accelerator Research Organization (KEK), Tsukuba 305-0801, Japan

HLT and Prompt Reco

- Talks by Itoh san
 - HLT + Express => real time within 48 hours
 - Prompt Reco => A few days after DAQ
(48hours ~ 7 days)
 - Reproductions => A few months after DAQ
- We need to make crontab:
 - HLT => 2 GB/sec registration
 - Tex file
 - Prompt Reco
 - Raw data file
 - Size is small

3. Implementation

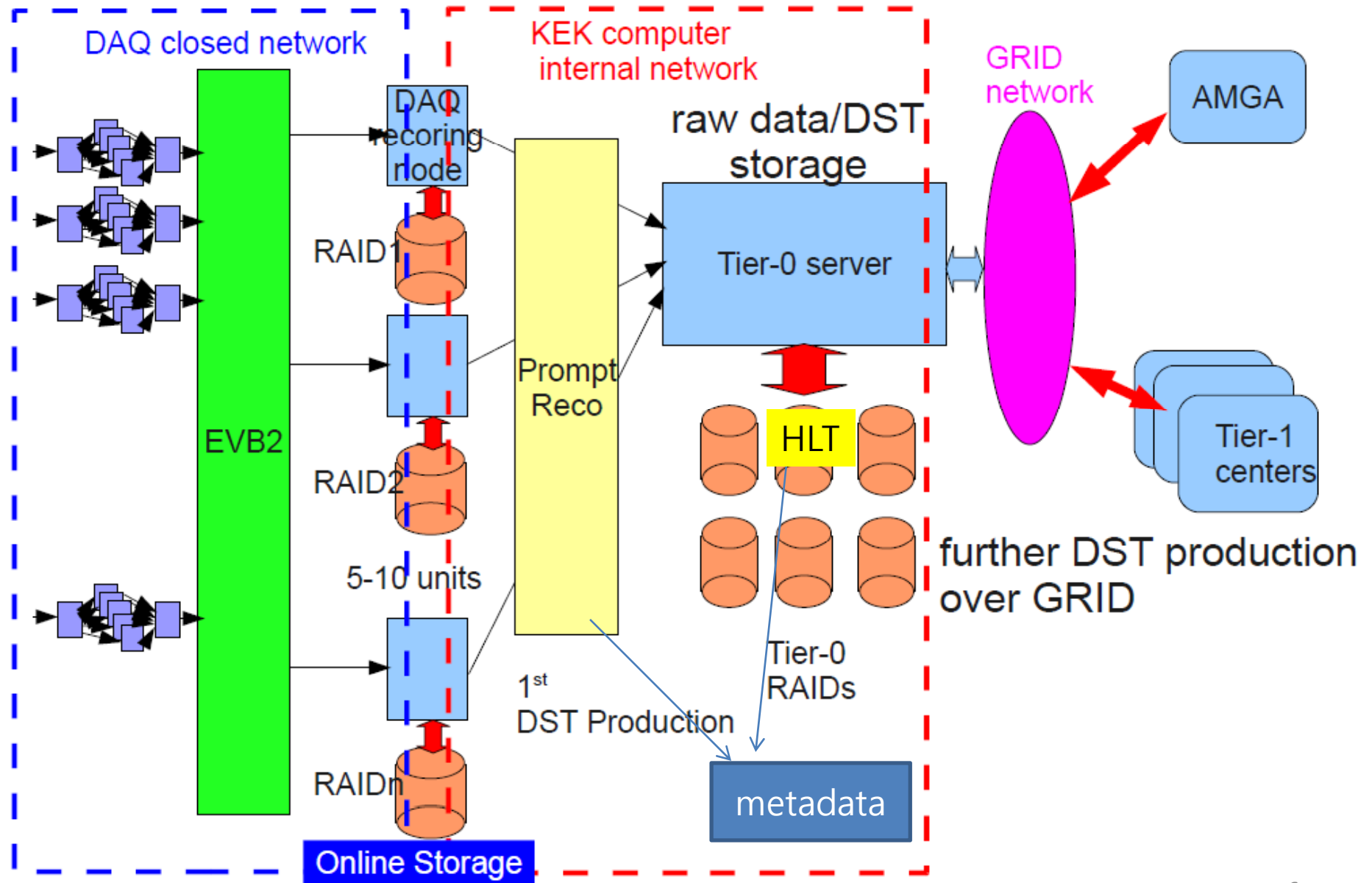
First sketch



* Idea is, **to run offline reconstruction instead of raw data copy right after a run is completed.**

* The refined constants are supposed to be made in HLT processing.

DAQ->Offline Data Flow

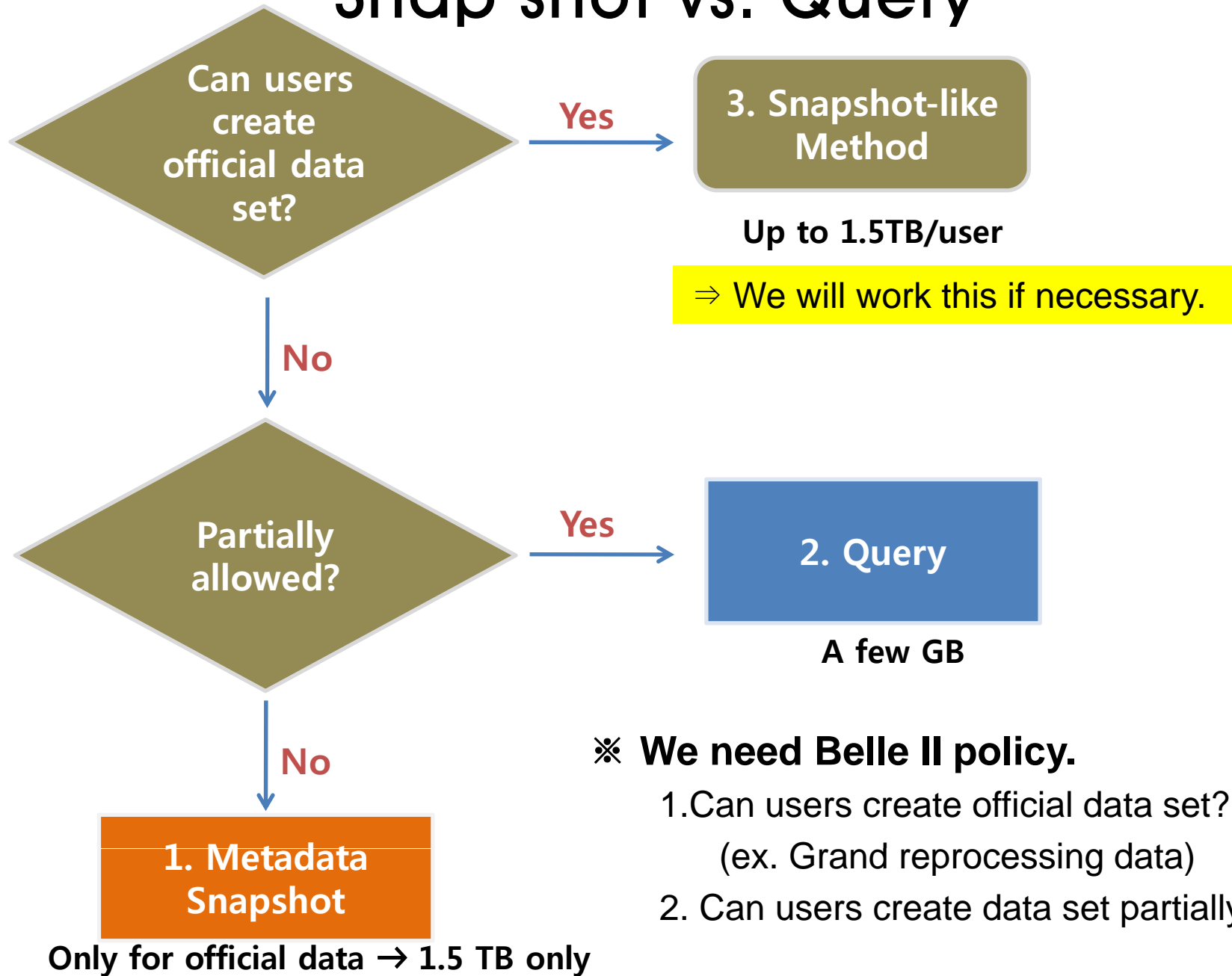


Definition of basf

- basf -> current Belle
- basf2 -> Geant4 + root + SVN+
- gbasf -> idea
- gbasf2 -> belle_amga_access(grid version)
dirac -> Belle VO (test)
Search, Data implement

=> Stand-alone data tool kit based on **basf2**
by Junghyun working with Milotz

Snap shot vs. Query



※ **We need Belle II policy.**

1. Can users create official data set?
(ex. Grand reprocessing data)
2. Can users create data set partially?

- For case 3, analysis group instead of users will decide data registration.
- We need to work on user-created data set.

=> Taesang Huh

Thank you.

Minutes of meetings

[Data Handling System : Kihyeon CHO]

- + Sunil Ahn will be at USC, USA for one year. Taesang Huh is taking over.
- + A test of metadata system on GRID farm is going on
- + AMGA package is ready now. It can be installed in Ljubljana machine.
- + To realize the analysis of Belle-I data on GRID system, standalone metadata registration tool is being prepared. Need to confirm which data should be used for this purpose. Plus, this has to be coordinated with Milosz.
- + The metadata system was discussed after the talk by Taesang Huh
- + Data management scheme for a service which handles the data copy from online RAID disk to the offline disk has to be considered. This is not related to the network configuration between the online and offline disks.

May 25

[Progress on Metadata System : Junghyun KIM] + [Study of User-Created Data Set : Taesang HUH]

+ Possible scenarios how to handle user-created dataset selections were shown.

1. Copy of the full metadata of each file in the dataset.
2. Just store the query
3. Store the result of the query, i.e. the list of files (snapshot) in a physical file and register the location in AMGA

+ The metadata size that was estimated for option 1 assumes that the full metadata of each file is copied, but it was pointed out that only an identifier is needed so that the data size is much smaller.

+ The issue with option 2 is that the result of the query may change in time since for example new files could be added.

+ An alternative to the physical file in option 3 is to store the snapshot in AMGA. Since only file IDs are needed in the snapshot there should be no issue with the metadata size.

+ Metadata should not be deleted. (physical files can be deleted)

+ Thomas should be contacted for basf2 installation issues.

+ Taesung Huh will provide the new AMGA Manager on the twiki