



Study of User-Created Datasets for Belle II

Taesang Huh, Sunil Ahn, Junhyun Kim

Supercomputing Center

KISTI (Korea Institute of Science and Technology Information)



Belle II Computing workshop 2011.05.25

INTRODUCTION

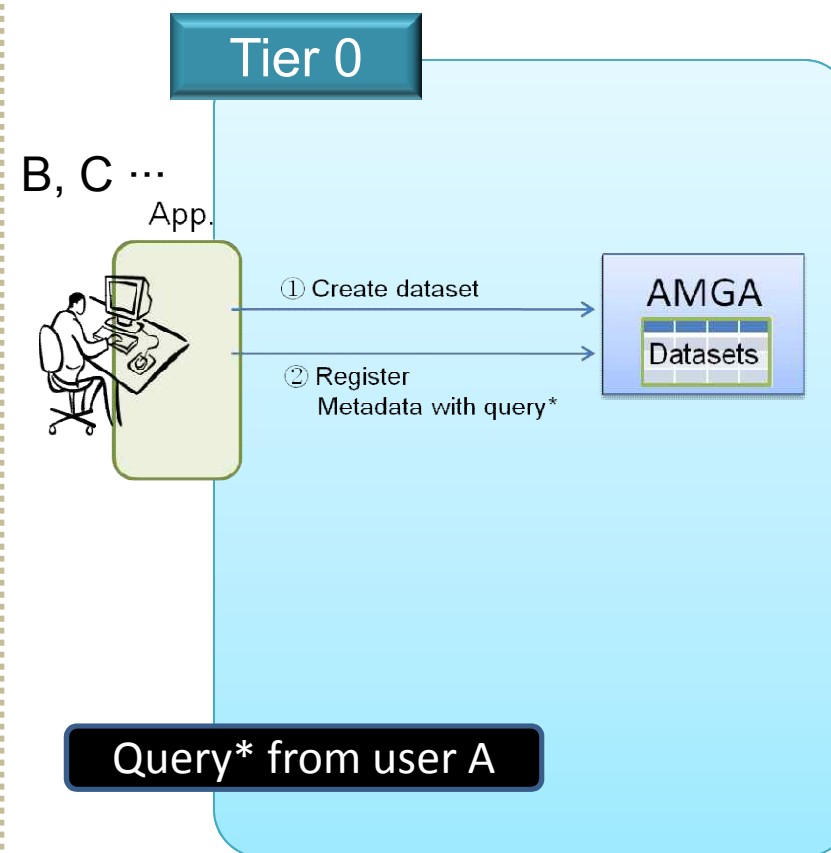
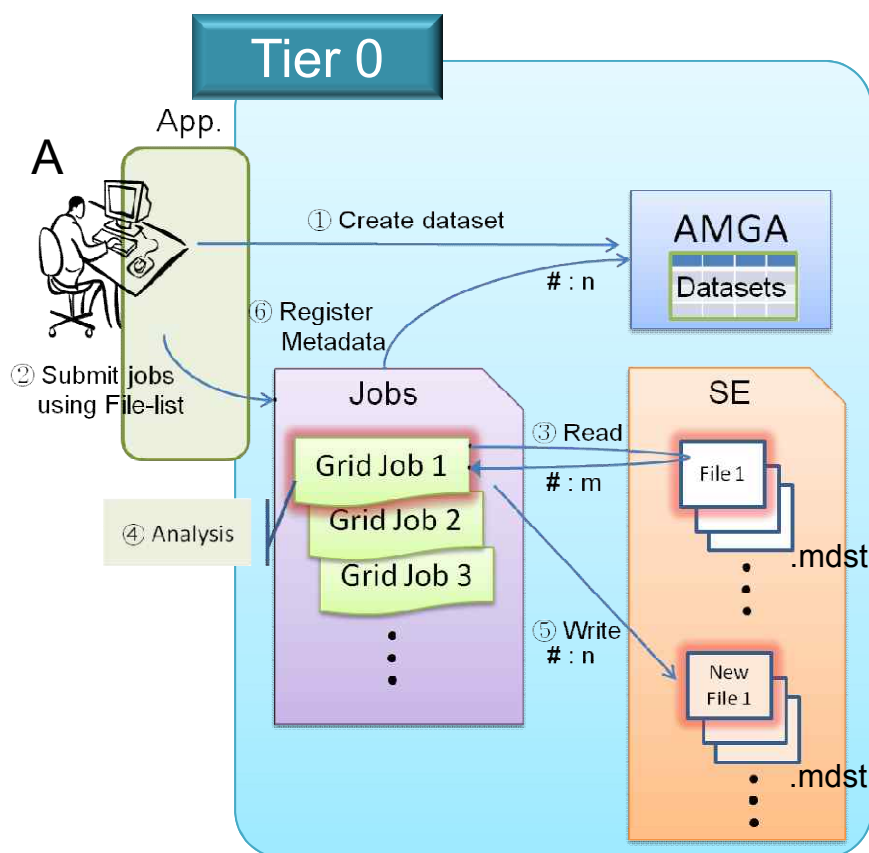
- Support of the logical grouping of files into user-named or defined "datasets"
- Propose schema for **User-Created Datasets(UCD)** considering of user scenario
- Plan to make new AMGA commands and API
- AMGA Manager for Belle II



Scenario I – Data Access for Official Data

- Make New UCD (1)
 - User : A
 - datasets

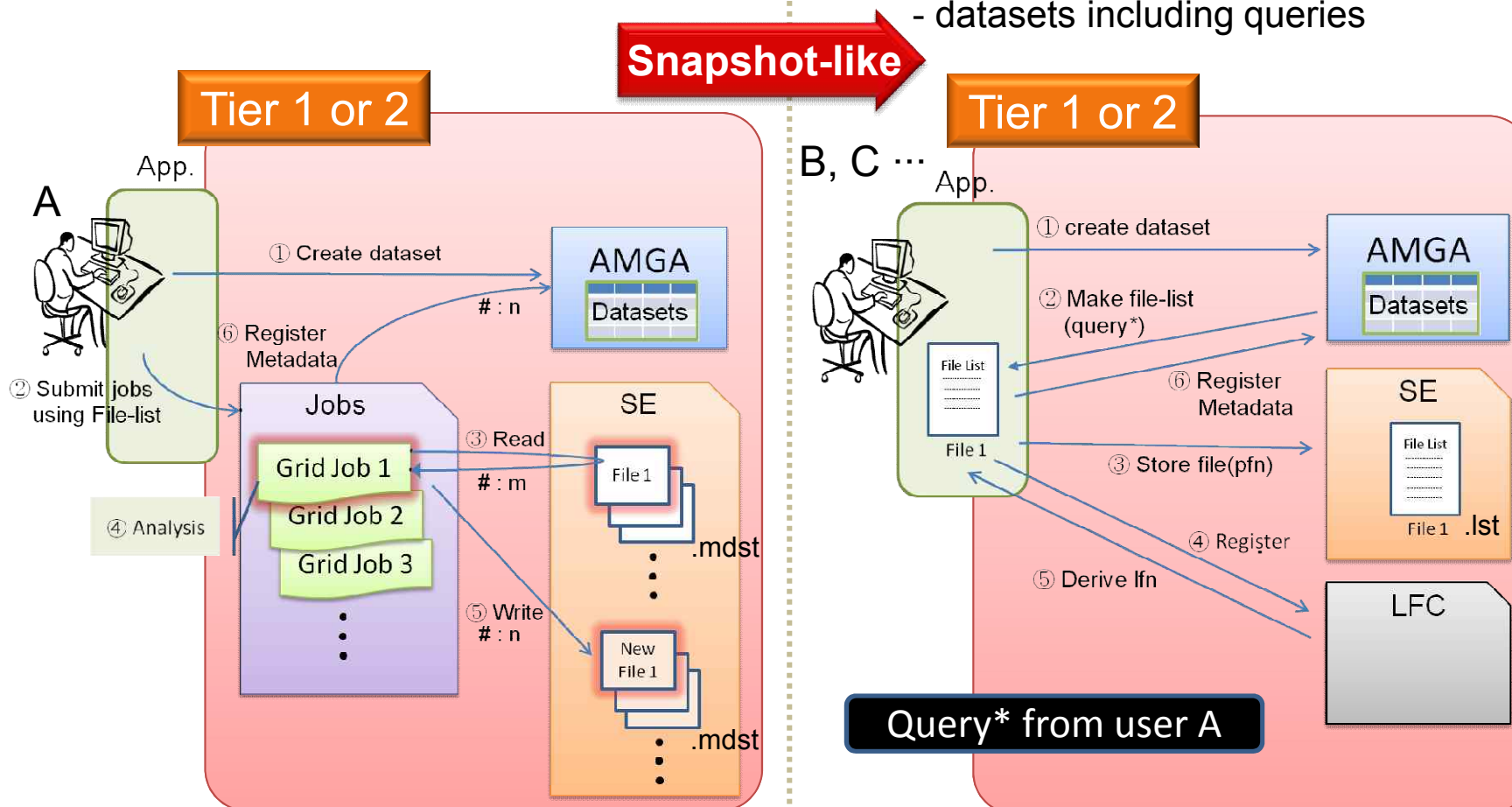
- Create metadata for queried results(2)
 - User : B, C ... (group members)
 - datasets including queries



Scenario II – Data Access for Non Official Data

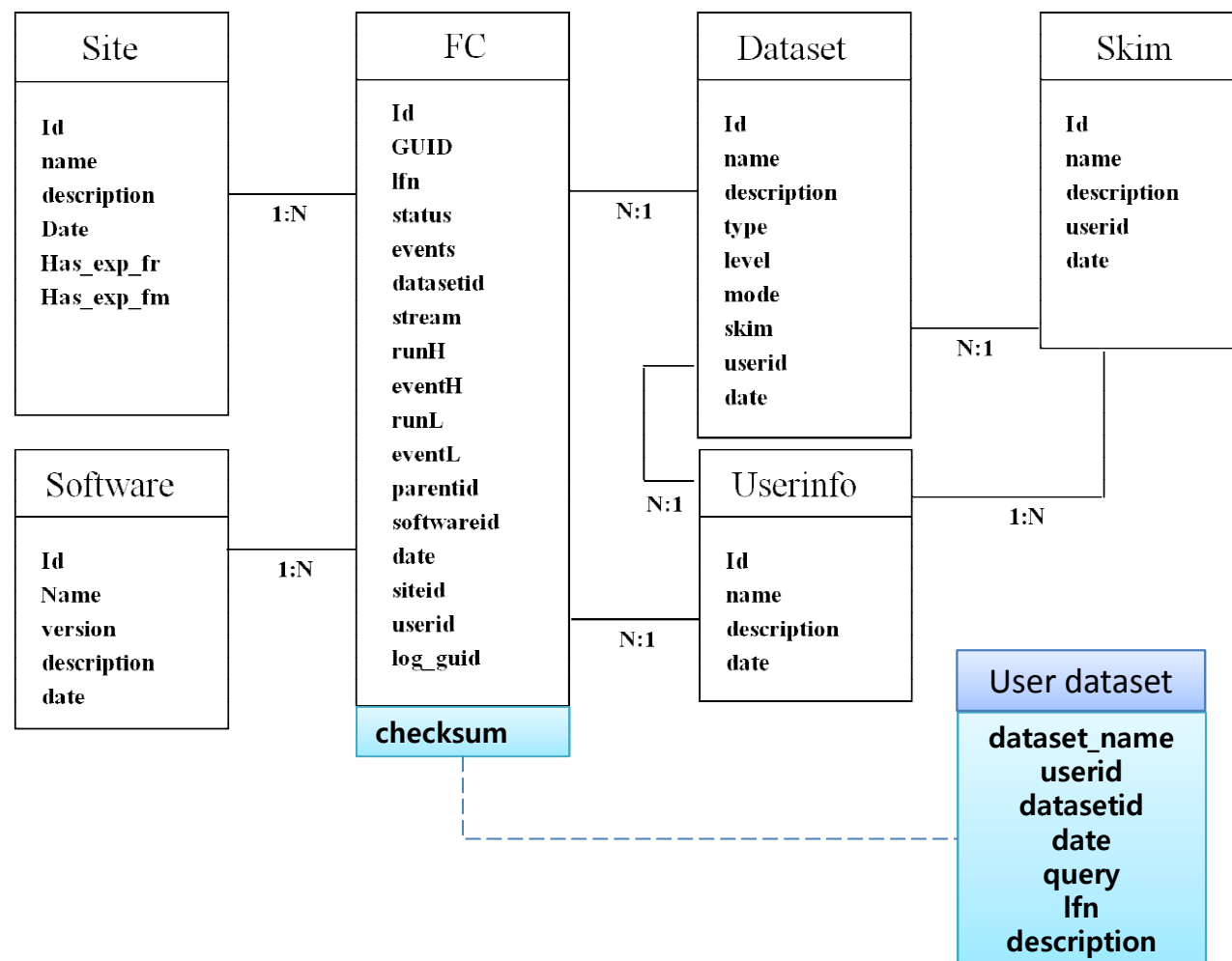
- Make new UCD(1)
 - User : A
 - datasets

- Retrieve Index file from UCD & Create metadata for Index files(2)
 - User : B, C ... (group members)
 - datasets including queries



Schema I – UCD schema

File-level Metadata Schema



- metadata for index file

Schema II

- Schema for Caching users' searching results

The schema of /belle2/User/[\$user_dataset]

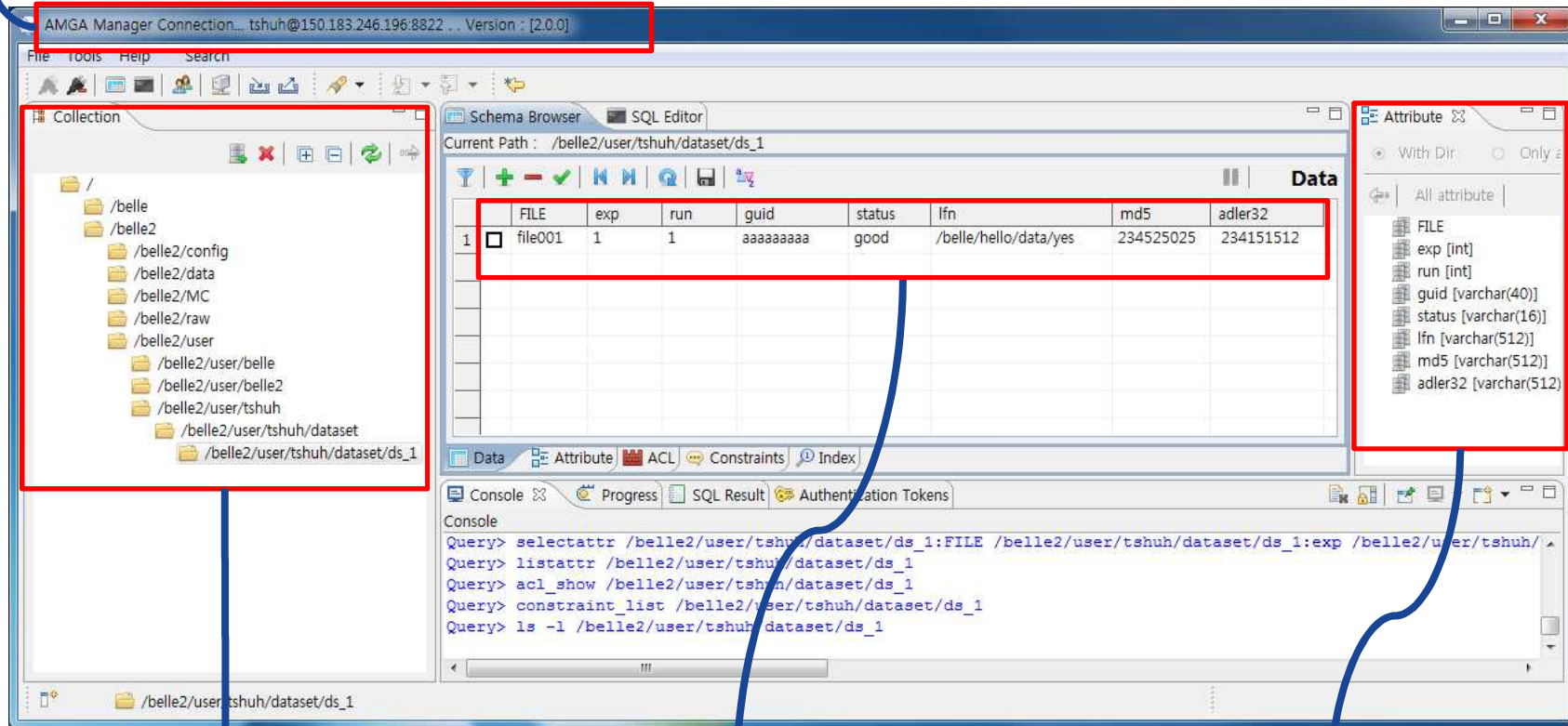
Attribute Name	Data Type in AMGA	Description
dataset_name	varchar(32)	unique dataset name
userid	smallint	ID of a user who create dataset
datasetid	smallint	dataset id
date	timestamp	date and time of query creation
query	varchar(512)	query for deriving a list of files
lfn	varchar(512)	logical file name for index file
description	varchar(1024)	description of user dataset

What will AMGA provide?

- Access AMGA
 - `belle_amga_access (...)` - now
 - `Gbasf2.py` - future
- Manipulate UCD
 - `belle_amga_create_userdataset`
 - `belle_amga_list_userdataset`
 - `belle_amga_remove_userdataset`
 - `belle_amga_add_userdata`
- various conditions in the same target dataset
 - `belle_amga_create_userview`
 - `belle_amga_list_userview`
 - `belle_amga_remove_userview`
 - `belle_amga_access_userview`

AMGA Manager

AMGA Manager Connection... tshuh@150.183.246.196:8822... Version : [2.0.0]



The screenshot shows the AMGA Manager interface with three main components highlighted by red boxes and blue arrows:

- Collection tree:** A hierarchical view of the dataset structure, including folders like /belle, /belle2, /belle2/config, /belle2/data, /belle2/MC, /belle2/raw, /belle2/user, and /belle2/user/tshuh/dataset.
- Sample Data:** A table showing a single data row with columns: FILE, exp, run, guid, status, lfn, md5, and Adler32. The row contains: file001, 1, 1, aaaaaaaaa, good, /belle/hello/data/yes, 234525025, 234151512.
- Attributes with data type:** A list of attributes and their data types: FILE, exp [int], run [int], guid [varchar(40)], status [varchar(16)], lfn [varchar(512)], md5 [varchar(512)], and Adler32 [varchar(512)].

The console at the bottom shows the following commands and their outputs:

```

Query> selectattr /belle2/user/tshuh/dataset/ds_1:FILE /belle2/user/tshuh/dataset/ds_1:exp /belle2/user/tshuh/
Query> listattr /belle2/user/tshuh/dataset/ds_1
Query> acl_show /belle2/user/tshuh/dataset/ds_1
Query> constraint_list /belle2/user/tshuh/dataset/ds_1
Query> ls -l /belle2/user/tshuh/dataset/ds_1
  
```

Collection tree

Sample Data

Attributes with data type

Contribution : AMGA Manager v1.1 for Belle II

Functions	AMGA Manager v1.1	AMGA Manager v1.0
DataGrid(SB)	<ul style="list-style-type: none"> • Streaming DataGrid • Play/Pause streaming data • Entries' limitation of an action ∴ possible to view a large scale of streaming metadata in real time 	<ul style="list-style-type: none"> • One time ∴ impossible to view a large scale of metadata in real time
Collection View	<ul style="list-style-type: none"> • Showing quickly collection view ∴ ls -s : get only sub-dir(AMGA ver >= 2.1.1) 	<ul style="list-style-type: none"> • Showing slowly collection view ∴ ls -l : get both of entries and sub-dir
	in the case of the target dir with a large scale of entries	
Federation	<ul style="list-style-type: none"> • Site Manager • Federation Manager 	Null

- AMGA Manager v1.0 : general-purpose
- AMGA Manager v1.1 : Reflecting Belle II requirements
(release and upload to twiki(Belle2) until early July)

Further Study

- Reflect user requirements in both Scenario and Schema
- Make new commands
(access, create/read/remove user created dataset, read, remove)
- AMGA API (insert metadata into AMGA at Basf2)
- Produce user guide line to make UCD



Thank you!

EMI is partially funded by the European Commission under Grant Agreement RI-261611