

# gbasf2 tutorial

Jake Bennett, Suravinda Kospalage

University of Mississippi

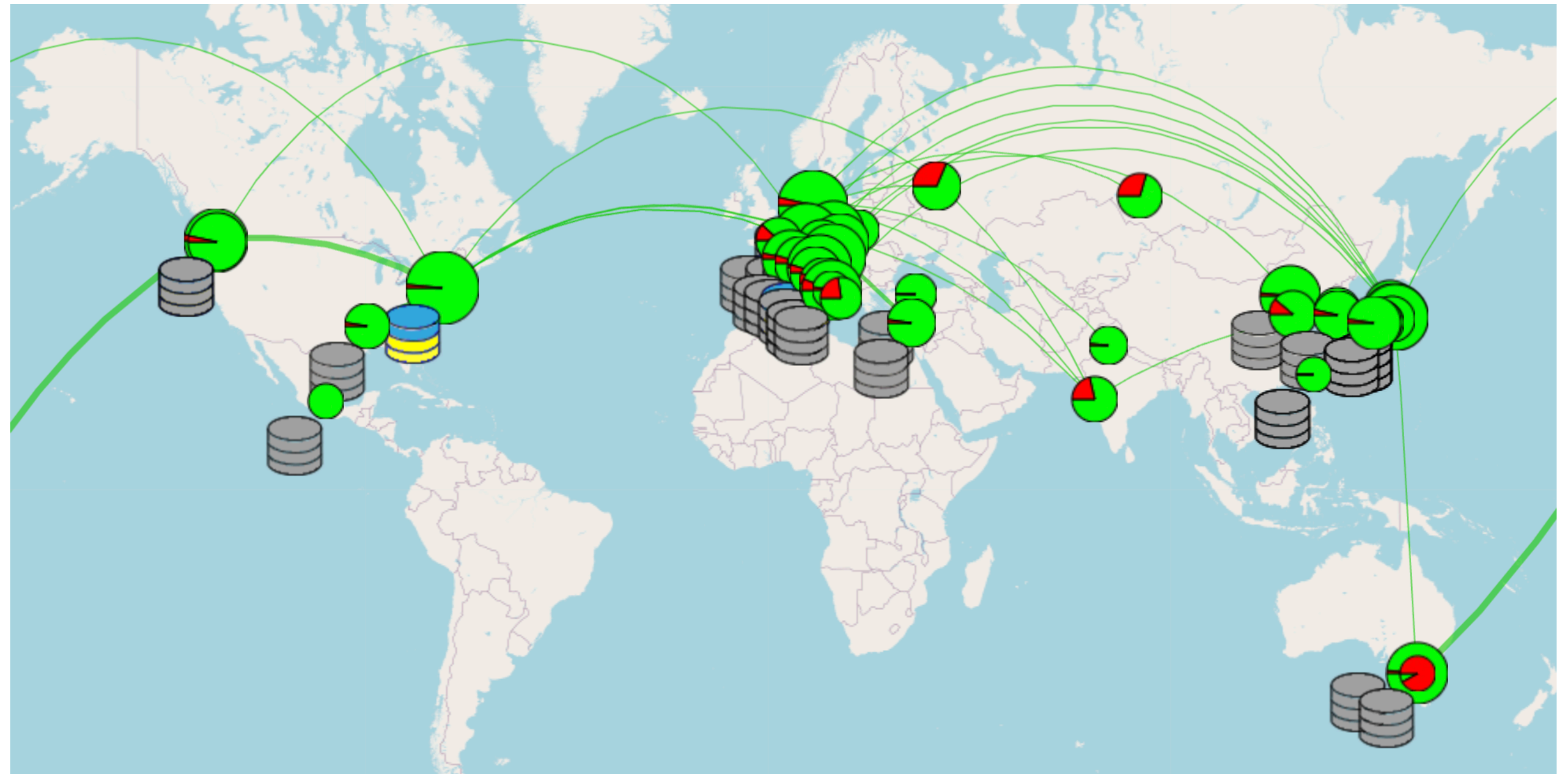
Belle II Summer Workshop, Duke University, July 25, 2023



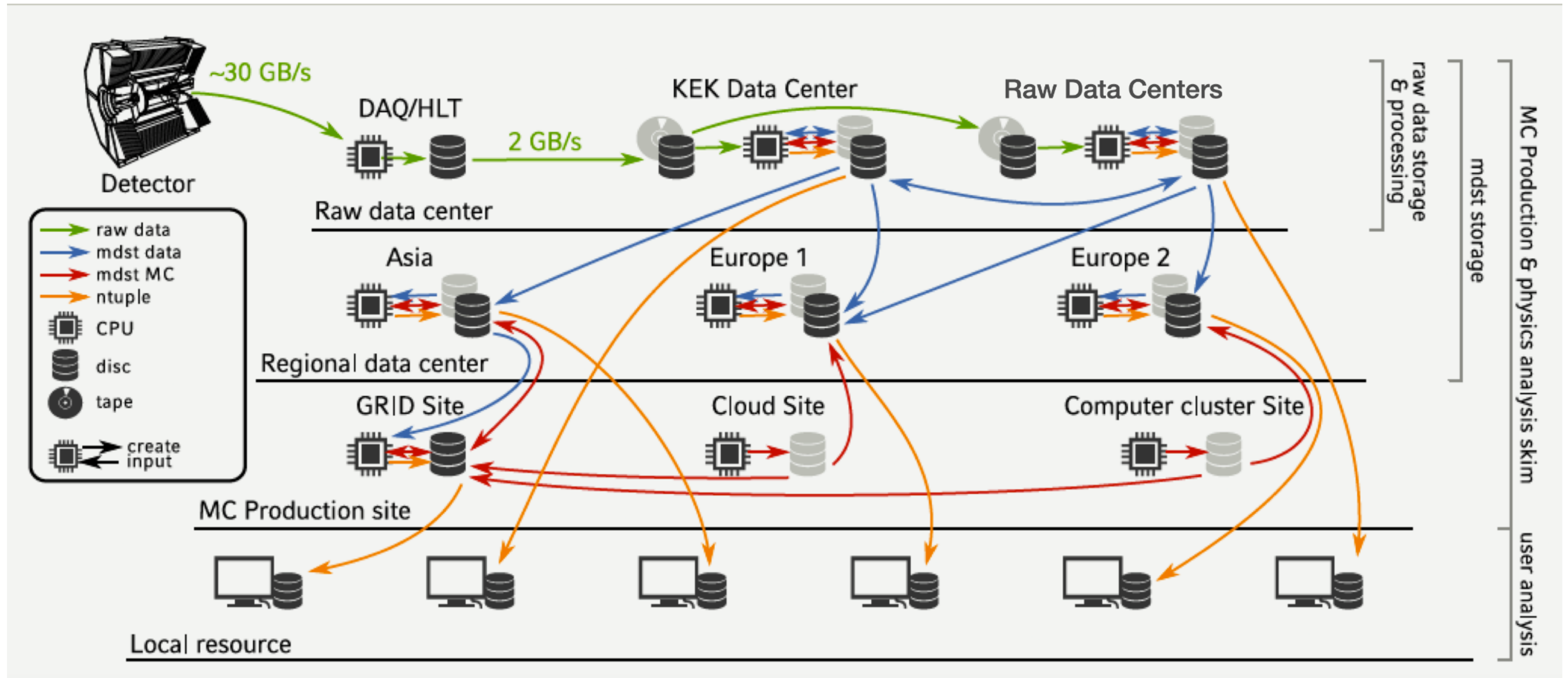
THE UNIVERSITY *of*  
**MISSISSIPPI**

# Brief introduction to Belle II Distributed Computing

- Belle II data processing and analysis requirements very large!
- Computing centers around the world contribute to Belle II tasks via the “grid”
- Grid-based tasks:
  - raw data processing
  - Monte Carlo production
  - analysis jobs
- **DIRAC**: “interware” connecting heterogeneous resources
- **BelleDIRAC**: experiment-specific extensions, tools
- **gbasf2**: grid-based user toolkit



# Brief introduction to Belle II Distributed Computing



# Prerequisites to running gbasf2

- Documentation and tutorials available in the [online textbook](#) and [gbasf2 documentation](#)

## Warning

Before getting started, make sure you understand the following:

- The GRID is NOT a local computing system like KEKCC.
- Once you submit jobs, they will be assigned to computing systems around the world.
- If your job is problematic, it will be distributed to the world and many sites will be affected.
- Remember, always test your jobs locally before submitting to the grid!

## Prerequisites:

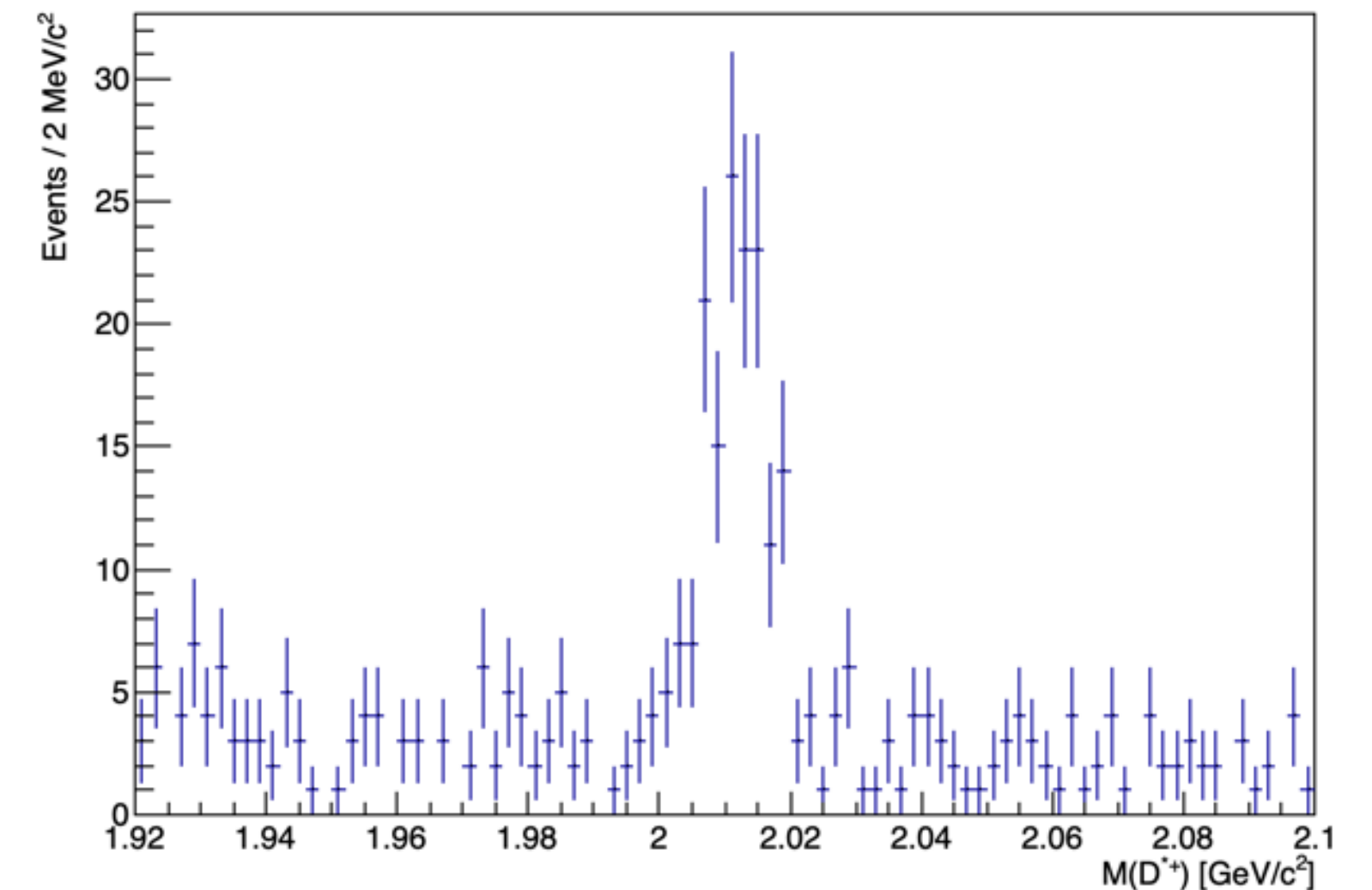
- See [Computing getting started](#).
- A system with SL6 or CentOS 7 (or access to KEKCC).
- A valid [grid certificate](#) installed in [~/.globus](#) and in a web browser.
- A working basf2 steering script (see the [First steering file](#) lesson)

It is required that you join the [comp users forum](#), where you can ask for help and receive announcements on releases and system issues.

Unfortunately, if you don't have a grid certificate, you will only be able to observe today

# Getting ready

- First rule of gbasf2: **test your script before submitting!**
- You should have a steering script from the basf2 tutorial
- If you need one, you can use [/home/belle2/jbennett/public/B2SW2023/gb2\\_example.py](/home/belle2/jbennett/public/B2SW2023/gb2_example.py) (can be downloaded or directly copied at KEKCC)
  - Reconstructs candidate events for  $D^{*+} \rightarrow [D^0 \rightarrow K^- \pi^+] \pi^+$
  - Input file: [/home/belle2/jbennett/public/B2SW2023/udst\\_000001\\_prod00028260\\_task114783000001.root](/home/belle2/jbennett/public/B2SW2023/udst_000001_prod00028260_task114783000001.root)



```
$ source /cvmfs/belle.cern.ch/tools/b2setup light-2305-korat
```

```
$ basf2 gb2_example.py
```

**Q: How do I find/decide which basf2 software release to use?**

# Getting ready to submit grid jobs

- Prerequisites
  - A system with SL6 or CentOS 7 (or access to KEKCC)
  - A valid [grid certificate](#) installed in `~/.globus` and in a web browser (to use DIRAC)

```
$ ls -l ~/.globus
```

```
-rw-r--r-- 1 jbenne b2_belle2 1966 Jan 12 2023 usercert.pem  
-rw----- 1 jbenne b2_belle2 1958 Jan 12 2023 userkey.pem
```

Your user key should only be readable by you!  
`'chmod 600 userkey.pem'`

- A working basf2 steering script (see the [First steering file](#) lesson)  
(or just use the one linked on the previous slide)

# Getting ready to submit grid jobs

- We will run gbasf2 in a new shell (note that basf2 and gbasf2 are NOT compatible)
- If your computing system has access to cvmfs (e.g. at KEKCC), the simplest way to use gbasf2 is via a central installation:

```
$ source /cvmfs/belle.kek.jp/grid/gbasf2/pro/setup.sh
```

- This sets all the necessary environment variables and initializes a grid proxy for you (you will be asked to enter your credentials)
  - Try executing 'gb2\_proxy\_info'
  - You can also pass arguments (e.g. to initialize a proxy with different permissions)
- If you need/want to install gbasf2, you can find instructions in the [online textbook](#)

Q: What did the source command do?

# Submit your first grid job!

- After you have set up gbasf2 (**and tested your script!**), you can submit your first grid job:

```
$ gbasf2 gb2_example.py -i /belle/Data/release-06-01-10/DB00002058/proc13/
prod00028260/e0012/4S/r04784/hadron/17240100/udst/sub00 -p testJob_B2SW23 -s
light-2305-korat
```

```
*****
***** Project summary *****
** Project name: testJob_B2SW23
** Dataset path: /belle/user/jbennett/testJob_B2SW23
** Steering file: gb2_example.py
** Job owner: jbennett @ belle (23:53:37)
** Preferred site / SE: None / None
** Input files for first job: LFN:/belle/Data/release-06-01-10/DB00002058/proc13/prod00028260/e0012/4S/r04784/
hadron/17240100/udst/sub00/udst_000001_prod00028260_task114783000001.root
** Number of input files: 1
** Number of jobs: 1
** Processed data (MB): 17
** Processed events: 2717 events
** Estimated CPU time per job: 46 min
*****
Are you sure to submit the project?
Please enter Y or N:
```



# I'm tired... you do some work now

- Look at the available gb2 tools (hint, ambiguous tab is your friend!)
  - Pick a group member to briefly discuss what you found for one tool

```
$ gb2<tab><tab>
gb2_check_release      gb2_ds_query_dataset  gb2_job_delete        gb2_project_analysis
gb2_diagnostic         gb2_ds_query_file     gb2_job_kill          gb2_project_summary
gb2_ds_collection      gb2_ds_quota          gb2_job_output        gb2_proxy_destroy
gb2_ds_count_events   gb2_ds_rep            gb2_job_parameters    gb2_proxy_info
gb2_ds_du              gb2_ds_rep_status     gb2_job_reschedule    gb2_proxy_init
gb2_ds_generate        gb2_ds_rm              gb2_job_status        gb2_se_list
gb2_ds_get             gb2_ds_rm_rep         gb2_list_destse       gb2set
gb2_ds_list            gb2_ds_search         gb2_list_site         gb2_update
gb2_ds_put             gb2_ds_siteForecast  gb2_prod_accounted
gb2_ds_query_datablock gb2_ds_sync           gb2_prod_dataset
```

# Monitoring jobs - command line tools

- Summary of all jobs/projects - good summary

```
$ gb2_project_summary
```

Project	Owner	Status	Done	Fail	Run	Wait	Submission Time(UTC)	Duration
testJob_B2SW23	jbennett	Good	1	0	0	0	2023-07-24 20:34:37	00:03:57

- More detailed status with 'gb2\_job\_status' using -p (project), -j (job ID), or -u (user)

```
$ gb2_job_status -p testJob_B2SW23
```

Job id	Status	MinorStatus	ApplicationStatus	Site
351459131	Done	Execution Complete	Done	LCG.CNAF.it

```
--- Summary of Selected Jobs ---
```

```
Completed:0 Deleted:0 Done:1 Failed:0 Killed:0 Running:0 Scouting:0 Stalled:0 Waiting:0
```

# Monitoring jobs - WebApp

The screenshot shows the Job Monitor web application interface. On the left is a navigation menu with 'Job Monitor' selected. The main area contains a 'Selectors' panel with filters for Site, Status, Minor Status, Application Status, Owner (set to 'jbennett'), OwnerGroup, Job Group, Job Type, and Time Span (set to 'Last Week'). Below the selectors are 'Submit', 'Reset', and 'Refresh' buttons. The main content area displays a table of jobs with the following data:

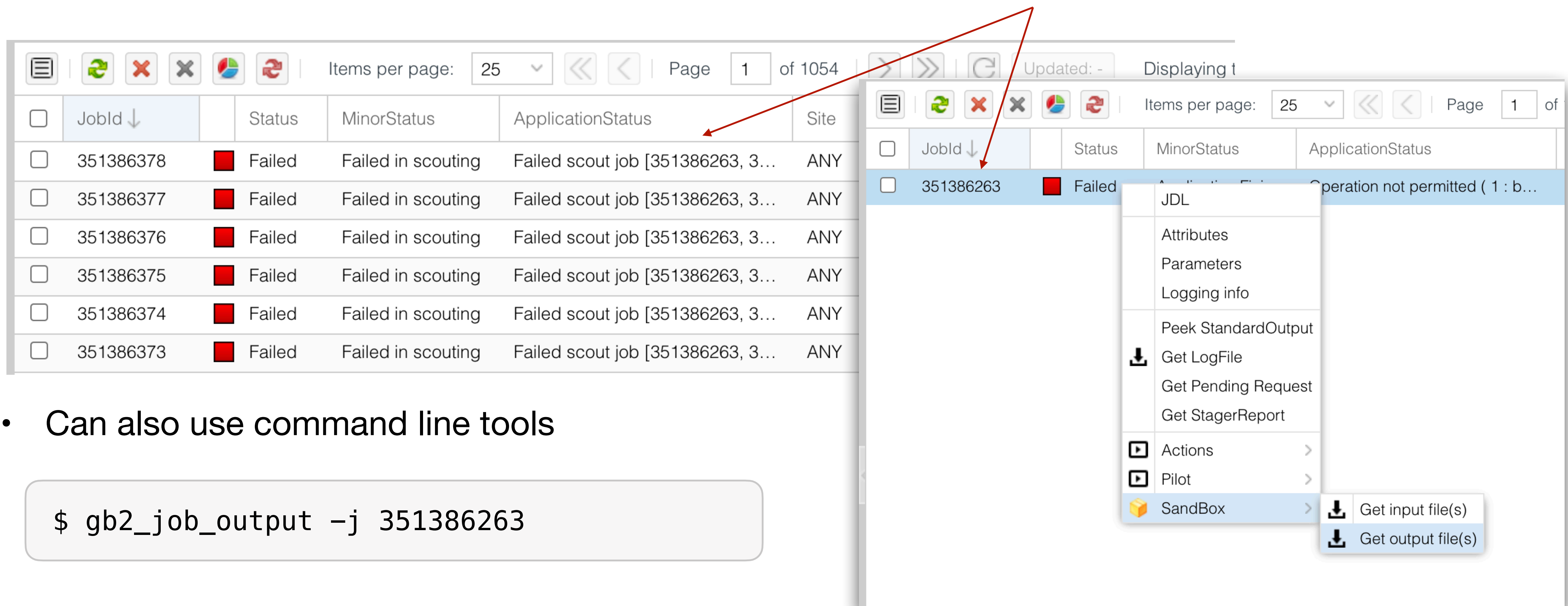
JobId	Status	MinorStatus	Application...	Site	JobName	LastUpdate[UTC]	Las
351459131	Done	Execution Complete	Done	LCG.CNAF.it	testJob_B2S...	2023-07-24 20:38:34	202

Annotations with red arrows point to specific elements in the table:

- 'Job status' points to the 'Done' status cell.
- 'More detailed status' points to the 'Execution Complete' minor status cell.
- 'Computing Site' points to the 'LCG.CNAF.it' site cell.
- 'Limit time span' points to the 'Last Week' dropdown in the Time Span selector.

# What to do if your jobs have failed

- If your jobs failed, most likely it was in scouting
  - Before your project runs, 10 scout jobs must complete (to avoid flooding the system with failures)
  - To understand what went wrong, find the output from the scout jobs



The screenshot shows a web-based job management interface. On the left, a table lists several failed jobs. The columns are JobId, Status, MinorStatus, ApplicationStatus, and Site. The first six rows show jobs with JobId 351386378 through 351386373, all with a status of 'Failed' and 'Failed in scouting' in the MinorStatus column. The ApplicationStatus column contains the text 'Failed scout job [351386263, 3...'. A red arrow points from the 'ApplicationStatus' column of the first row to a context menu on the right. The context menu is open for job 351386263, which has a status of 'Failed'. The menu items include: JDL, Attributes, Parameters, Logging info, Peek StandardOutput, Get LogFile, Get Pending Request, Get StagerReport, Actions, Pilot, and SandBox. The SandBox item is expanded, showing sub-items: Get input file(s) and Get output file(s).

JobId	Status	MinorStatus	ApplicationStatus	Site
351386378	Failed	Failed in scouting	Failed scout job [351386263, 3...	ANY
351386377	Failed	Failed in scouting	Failed scout job [351386263, 3...	ANY
351386376	Failed	Failed in scouting	Failed scout job [351386263, 3...	ANY
351386375	Failed	Failed in scouting	Failed scout job [351386263, 3...	ANY
351386374	Failed	Failed in scouting	Failed scout job [351386263, 3...	ANY
351386373	Failed	Failed in scouting	Failed scout job [351386263, 3...	ANY

```
$ gb2_job_output -j 351386263
```

- Can also use command line tools

```
$ gb2_job_output -j 351386263
```

# What to do if your jobs have failed

```
$ gb2_diagnostic --failed_job 348267037
```

```
INFO ##### General Info #####
INFO Timestamp: 2023-07-24 23:39:54.382192 UTC
INFO proxyInfo : OK
INFO DIRAC username: jbennett
INFO DIRAC groupname: belle
INFO User DN: /DC=org/DC=cilogon/C=US/O=Brookhaven National Laboratory/CN=Jake Bennett A6615086
INFO Rucio ping: {'version': '1.28.5'}
INFO #####
```

```
INFO
diagnostic of Failed Job .....
INFO ##### Job Summary #####
INFO Failed JobID: 348267037
INFO Minor status: Application Finished With Errors
INFO Application Status: Bad file descriptor ( 9 : basf2helper.py Exited With Status 9)
INFO #####
```

```
INFO
##### Logging Info #####
```

Source	Status	MinorStatus	ApplicationStatus	DateTime
JobManager	Submitting	Bulk transaction confirmation	Unknown	2023-07-11 22:08:55
JobPath	Checking	Scouting	Unknown	2023-07-11 23:12:21
Scouting	Scouting	Waiting for Scout Job Completion	Unknown	2023-07-11 23:12:21
ScoutingJobStatusAgent	Checking	Scouting	Unknown	2023-07-11 23:13:13
Scouting	Checking	JobSanity	Unknown	2023-07-11 23:26:35
JobSanity	Checking	InputData	Unknown	2023-07-11 23:26:35
InputData	Checking	JobScheduling	Unknown	2023-07-11 23:26:36
JobScheduling	Waiting	Pilot Agent Submission	Unknown	2023-07-11 23:26:36
JobAgent@LCG.KEK.jp	Waiting	Submitting To CE	Unknown	2023-07-24 22:31:56
JobAgent@LCG.KEK.jp	Waiting	Installing Software	Unknown	2023-07-24 22:31:56
JobAgent@LCG.KEK.jp	Waiting	Job Received by Agent	Unknown	2023-07-24 22:31:56
Matcher	Matched	Assigned	Unknown	2023-07-24 22:31:56
JobWrapper	Running	Job Initialization	Unknown	2023-07-24 22:32:00
JobWrapper	Running	Downloading InputSandbox	Unknown	2023-07-24 22:32:00
JobWrapper	Running	Input Data Resolution	Unknown	2023-07-24 22:32:01
JobWrapper	Running	Application	Unknown	2023-07-24 22:34:25
Job_348267037	Running	Application	Executing RunScriptStep1	2023-07-24 22:34:27
basf2	Running	Application	Running	2023-07-24 22:36:46
Job_348267037	Running	Application	basf2helper.py Exited With Status 9	2023-07-24 22:51:10
Job_348267037	Running	Application	Bad file descriptor ( 9 : basf2helper.py Exited With Status 9)	2023-07-24 22:51:11
JobWrapper	Completing	Application Finished With Errors	Bad file descriptor ( 9 : basf2helper.py Exited With Status 9)	2023-07-24 22:51:16
JobWrapper	Failed	Application Finished With Errors	Bad file descriptor ( 9 : basf2helper.py Exited With Status 9)	2023-07-24 22:51:16

Experts are here to help, but we need information (you should also try to understand and fix issues yourself, if you can)

Proxy information

Job Summary

Logging info

# What to do if your jobs have failed

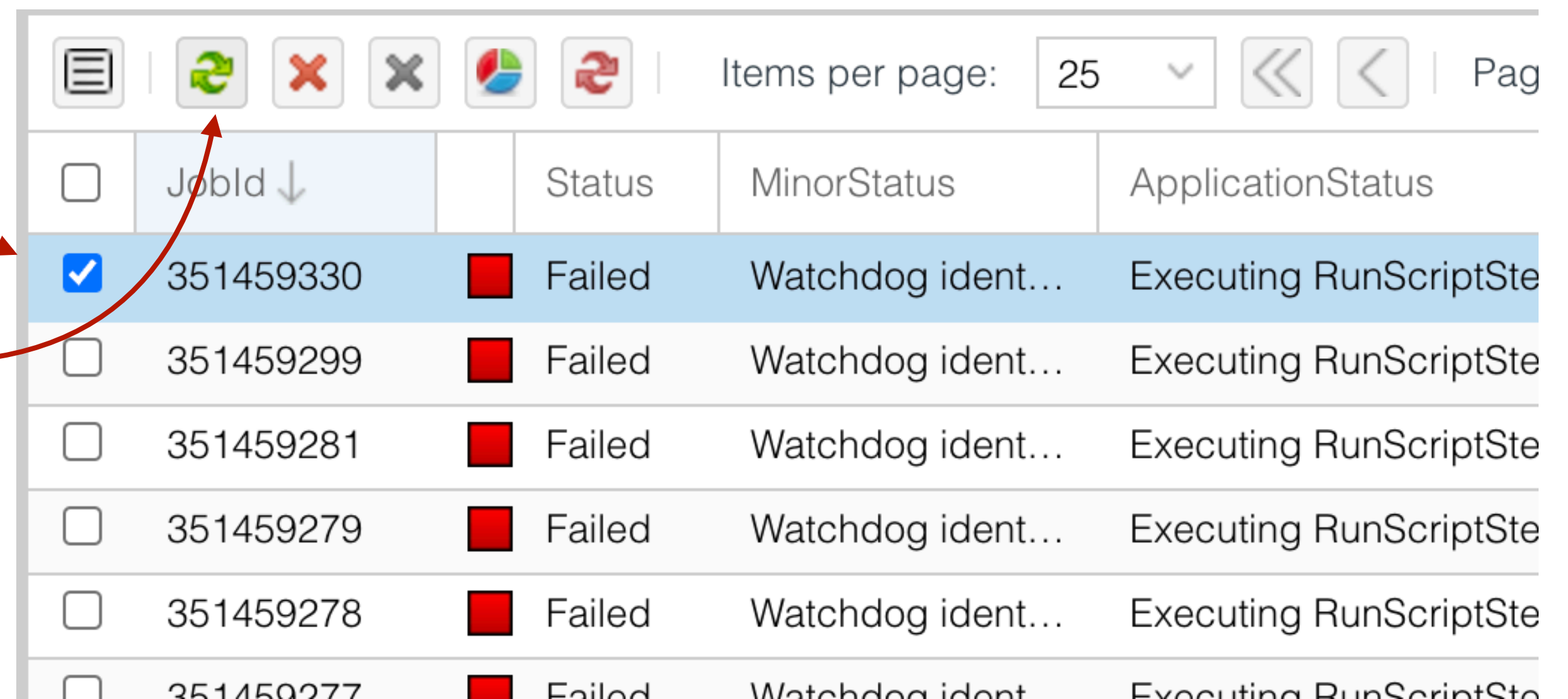
- Failed jobs can be rescheduled via the command line or DIRAC

```
$ gb2_job_reschedule ( [-j JOBID... | -p PROJECT...])
```

Select the jobs to be rescheduled

Click the button to reschedule

- Sometimes jobs hit the limit for rescheduling (e.g. site issues that cause repeated failures)
  - Tell the experts (include job numbers)



<input type="checkbox"/>	JobId ↓	Status	MinorStatus	ApplicationStatus
<input checked="" type="checkbox"/>	351459330	Failed	Watchdog ident...	Executing RunScriptSte
<input type="checkbox"/>	351459299	Failed	Watchdog ident...	Executing RunScriptSte
<input type="checkbox"/>	351459281	Failed	Watchdog ident...	Executing RunScriptSte
<input type="checkbox"/>	351459279	Failed	Watchdog ident...	Executing RunScriptSte
<input type="checkbox"/>	351459278	Failed	Watchdog ident...	Executing RunScriptSte
<input type="checkbox"/>	351459277	Failed	Watchdog ident...	Executing RunScriptSte

# After your jobs have completed

- Check the output
  - Note: gb2 tools will use your username/space by default

```
$ gb2_ds_list  
/belle/user/jbennett/testJob_B2SW23
```

```
$ gb2_ds_list testJob_B2SW23/sub00  
/belle/user/jbennett/testJob_B2SW23/sub00/ntuple_00000_job351459131_00.root
```

# After your jobs have completed

```
$ gb2_ds_get /belle/user/jbennett/testJob_B2SW23/sub00/ntuple_00000_job351459131_00.root
Files to download to /gpfs/home/belle2/jbennett/public/B2SW2023/testJob_B2SW23/sub00 : 1 file(s)
Do you want to download files:
Please type [Y] or [N]: Y
```

Download 1 files from SE

```
Trying to download davs://xfer-archive.cr.cnaf.infn.it:8443/webdav/belle/TMP/belle/user/
jbennett/testJob_B2SW23/sub00/ntuple_00000_job351459131_00.root to /gpfs/home/belle2/jbennett/
public/B2SW2023/testJob_B2SW23/sub00/ntuple_00000_job351459131_00.root
```

Successfully downloaded files:

```
/belle/user/jbennett/testJob_B2SW23/sub00/ntuple_00000_job351459131_00.root in /gpfs/home/
belle2/jbennett/public/B2SW2023/testJob_B2SW23/sub00
```

Failed files:



If your download failed, you can try again or ask the experts for help



# Pro-tip: watch for new features

```
$ gb2_ds_get -h
usage: gb2_ds_get.py [-h] [-v] [--usage] [-o local directory] [-u USER] [-r {MC,data,user}] [-l] [--new]
[-f] [--noSubDir] [-i file with lfns]
                    [--failed_lfns filename.txt]
                    dataset

positional arguments:
  dataset                specify dataset name

optional arguments:
  -h, --help            show this help message and exit
  -v, --verbose         increase verbosity (up to -vv)
  --usage              show detailed usage
  -o local directory, --output_dir local directory
                    path to local directory
  -u USER, --user USER specify user name
  -r {MC,data,user}, --subcate {MC,data,user}
                    specify a dataset category
  -l, --long           long listing (-ll: extra long)
  --new                Enables experimental feature(s) for the tool.
  -f, --force         skip confirmation
  --noSubDir          Avoid downloading of files in subdirectories of the given dataset
  -i file with lfns, --input_dslist file with lfns
                    Input file with list of LFNs to download
  --failed_lfns filename.txt
                    Set the name of the text file where failed LFNs will be stored
```

Enables multithreaded download



**--new** Enables experimental feature(s) for the tool.

# After you have downloaded your files, clean your grid space

```
$ gb2_ds_rm testJob_B2SW23  
LFN(s) to remove:
```

Dataset	Files
/belle/user/jbennett/testJob_B2SW23/sub00	1

Do you want to remove the following file(s):  
Please type [Y] or [N]: Y

LFN

Something weird happened

result

```
Failed  
/belle/user/jbennett/testJob_B2SW23/sub00 Failed to remove AMGA Directory /belle/user/  
jbennett/testJob_B2SW23/sub00  
/belle/user/jbennett/testJob_B2SW23 Failed to remove AMGA Directory /belle/user/  
jbennett/testJob_B2SW23
```

```
$ gb2_ds_list testJob_B2SW23  
No datasets
```

But it seems ok. If you are worried, contact the experts.

# Enough for the easy part

- Now it's your turn!

## 1. Find the input files most appropriate for your analysis

Dataset Searcher [Untitled 1] x

### Dataset Searcher via DIRAC

Metadata Searcher Tree Browser

Data Type:  MC  Data

Background level:  BGx1  BGx0  Other

Background level:  Campaigns:

Beam Energies:  Skim Types:

Data Levels:  Releases:

Global Tags:  Experiment Low:

Experiment High:  Run Low:

Run High:  MC Event Types:

General Skim Names:

LPN

Dataset LFNs Metadata  Dataset Metadata

# Enough for the easy part

- Now it's your turn!

## 1. Find the input files most appropriate for your analysis

### Dataset Searcher via command line

```
$ gb2_ds_search -h
usage: gb2_ds_search.py [-h] [-v] [--usage] {metadata,dataset,collection} ...

positional arguments:
  {metadata,dataset,collection}
    metadata           What kind of search is this?
                       Searches for dataset LPNs matching the given metadata constraints.
    dataset            Returns available metadata values.
    collection         "Returns a list of collections, or its content and metadata.
```

# Enough for the easy part

- Now it's your turn!

## 1. Find the input files most appropriate for your analysis

## Dataset Searcher via command line

```
$ gb2_ds_search dataset -h

optional arguments:
  -h, --help                show this help message and exit
  -o OUTPUT_FILE, --output_file OUTPUT_FILE
                           Output a text file containing all matching datasets.
  --campaign CAMPAIGN      The MC or Data production campaign name.
  --data_type DATA_TYPE
                           mc or data
  --data_level DATA_LEVEL
                           udst, mdst, etc
  --run_high RUN_HIGH      The highest allowed run number(INTEGER VALUE)(inclusive).
  --exp_high EXP_HIGH      The highest allowed Experiment number (INTEGER VALUE) (inclusive).
  --run_low RUN_LOW        The lowest allowed Run number (INTEGER VALUE) (inclusive).
  --exp_low EXP_LOW        The highest allowed Experiment number (INTEGER VALUE) (inclusive).
  --mc_event MC_EVENT      The MC event type ("uubar", "1110043100", etc) used for
  --skim_decay SKIM_DECAY
                           The skim type used to reconstruct and select events.
  --general_skim GENERAL_SKIM
                           The general skim name (all, hadron, etc)
  --beam_energy BEAM_ENERGY
                           4S, 5S, etc
  --global_tag GLOBAL_TAG
                           The global tag used to create the dataset.
  --release RELEASE        The basf2 release used to create the dataset.
  --bkg_level BKG_LEVEL
                           Background Level for MC
```

# Enough for the easy part

- Now it's your turn!

## 1. Find the input files most appropriate for your analysis

Dataset Searcher via command line

```
$ gb2_ds_search collection --list_all_collections /belle/collection/Data/*  
/belle/collection/Data/Moriond2022_all_4Soffres_v1  
/belle/collection/Data/Moriond2022_all_4S_v1  
/belle/collection/Data/Moriond2022_hadron_4Soffres_v1  
/belle/collection/Data/Moriond2022_hadron_4S_v1  
/belle/collection/Data/Moriond2023_4S_proc13_all_EMuonPlusMissEne_v1  
/belle/collection/Data/Moriond2023_4S_prompt_all_EMuonPlusMissEne_v1  
/belle/collection/Data/Moriond2023_prompt_all_4S_offres_v1  
/belle/collection/Data/Moriond2023_prompt_all_4S_v1  
/belle/collection/Data/Moriond2023_prompt_all_5Sscan_10657_v1  
/belle/collection/Data/Moriond2023_prompt_all_5Sscan_10706_v1  
/belle/collection/Data/Moriond2023_prompt_all_5Sscan_10751_v1  
/belle/collection/Data/Moriond2023_prompt_all_5Sscan_10810_v1
```

<https://confluence.desy.de/display/BI/Collection+summary>

# Enough for the easy part

- Now it's your turn!
  1. Find the input files most appropriate for your analysis
  2. Submit your jobs
  3. Monitor the output
  4. Download your samples
  5. Remove the output

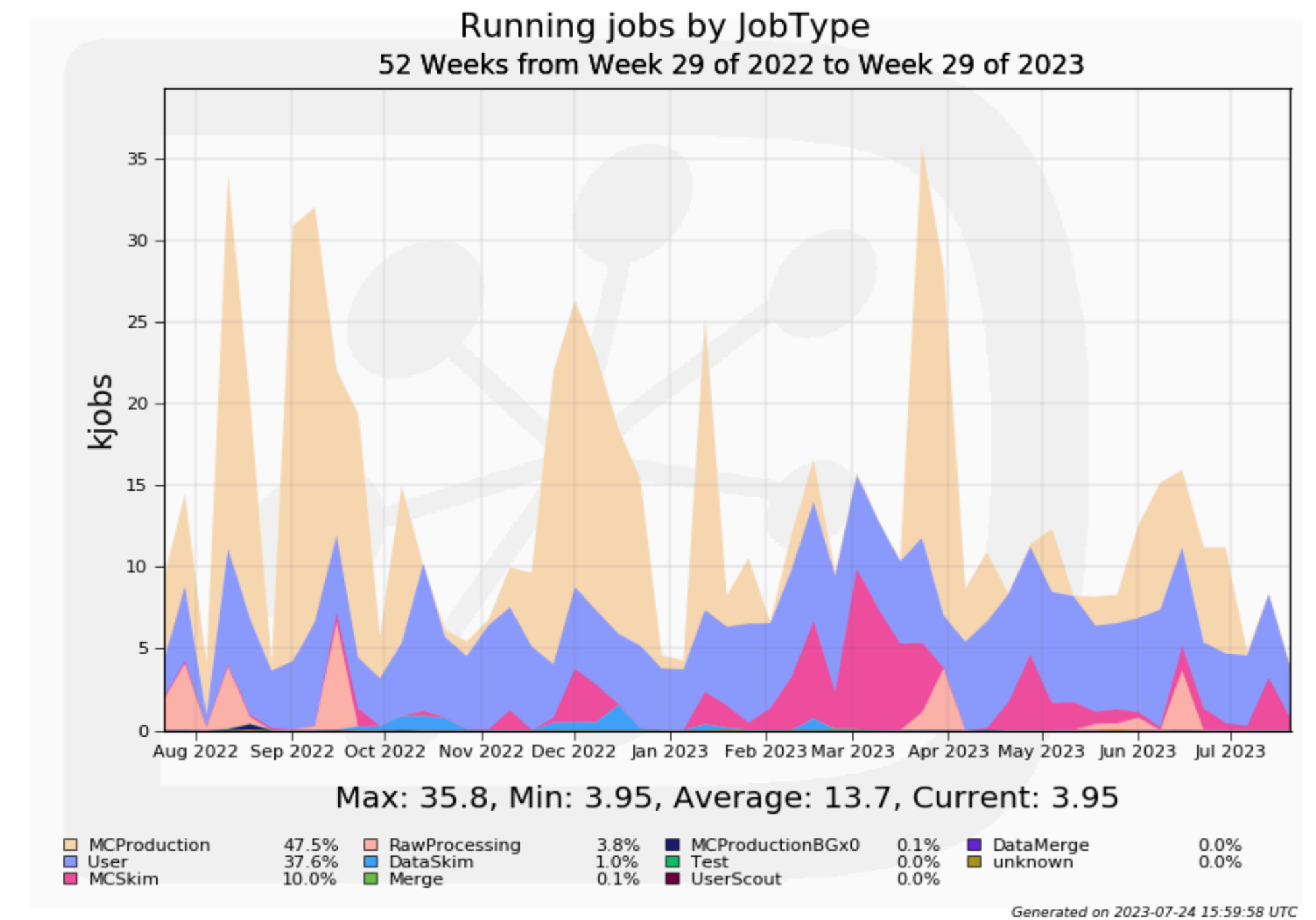


If possible, use skims! (If a skim is not available, develop one!)

<https://software.belle2.org/development/sphinx/skim/doc/04-experts.html#skim-registry>

# Final remarks

- Please be patient!
  - Job rates can be slow when there are many jobs
  - The experts are working hard on improvements
  - For faster service, provide more details!



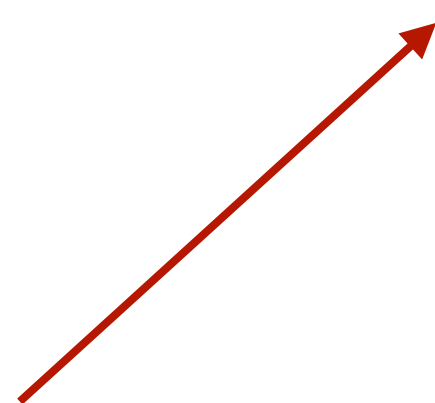
**WE NEED  
YOU**

- You can help us!
  - Provide feedback, requests, ideas
  - Take data production shifts
  - Train to become an expert shifter
  - Contribute to distributed computing development



# Useful links

- Documentation and tutorials available in the online textbook and gbasf2 documentation
- Make sure you join the comp-users-forum mailing list
  - Ask, read, or *answer* questions
  - Updates and announcements made here
- You can also ask and find answers to questions at
  - questions.belle2.org
  - gbasf2 FAQ page
  - gbasf2 confluence page
  - Computing glossary
- You can help by taking data production shifts, and train to become an expert shifter



## Shifts booking: 2023c SW and Data Production

SW QUALITY

DATA PRODUCTION

DATA PRODUCTION EXPERT

DP EXPERT SHADOW

Data Production quality assurance and control. Your duties are:

- Monitor the MC production system and activities
- Collect information and record them on the shift log
- Report issues to the operation experts
- Update the Operation Status summary page

Please consult the shifter manual:

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

for detailed info.

August 2023						
				04 Fri	05 Sat	06 Sun
				Not ..	Not ..	Not ..
				Not ..	Not ..	Not ..
				Not ..	Not ..	Not ..
07 Mon	08 Tue	09 Wed	10 Thu	11 Fri	12 Sat	13 Sun
Not ..	Not ..	OWL	OWL	Olga ..	Olga ..	Olga ..
Not ..	Not ..	Junxi..	Junxi..	DAY	DAY	DAY
Not ..	Not ..	SWING	SWING	SWING	SWING	SWING
14 Mon	15 Tue	16 Wed	17 Thu	18 Fri	19 Sat	20 Sun
OWL	OWL	OWL	OWL	Sever..	Sever..	Sever..
DAY	DAY	DAY	DAY	DAY	DAY	DAY
SWING	SWING	SWING	SWING	Gaura..	Gaura..	Gaura..
21 Mon	22 Tue	23 Wed	24 Thu	25 Fri	26 Sat	27 Sun
OWL	OWL	OWL	OWL	OWL	OWL	OWL
DAY	DAY	DAY	DAY	DAY	DAY	DAY
SWING	SWING	SWING	SWING	SWING	SWING	SWING
28 Mon	29 Tue	30 Wed	31 Thu			
OWL	OWL	OWL	OWL			
DAY	DAY	DAY	DAY			
SWING	SWING	SWING	SWING			