



gbasf2 tutorial

Quinn Campagna

University of Mississippi

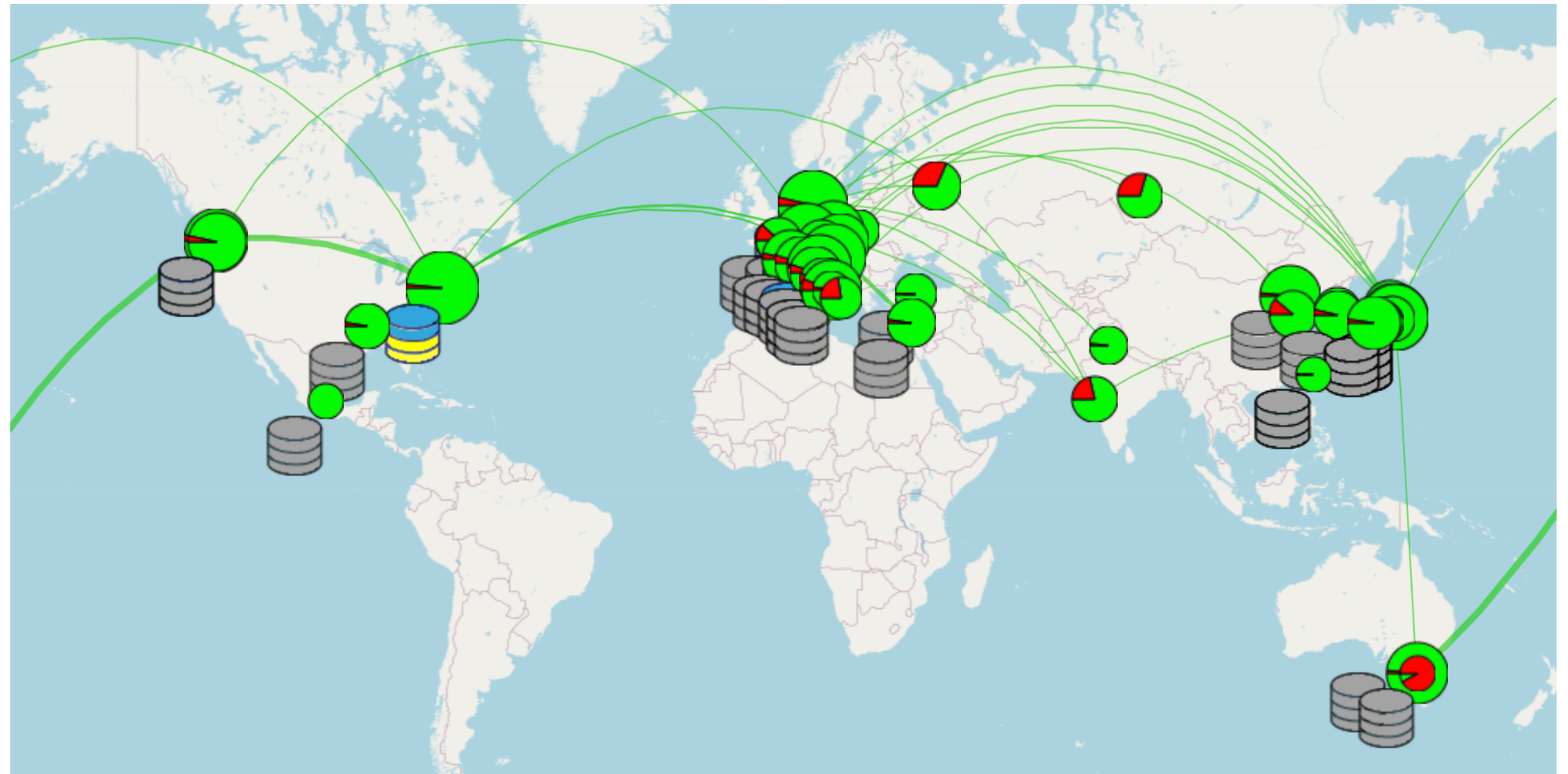
Belle II Summer Workshop, Virginia Tech, June 24, 2023



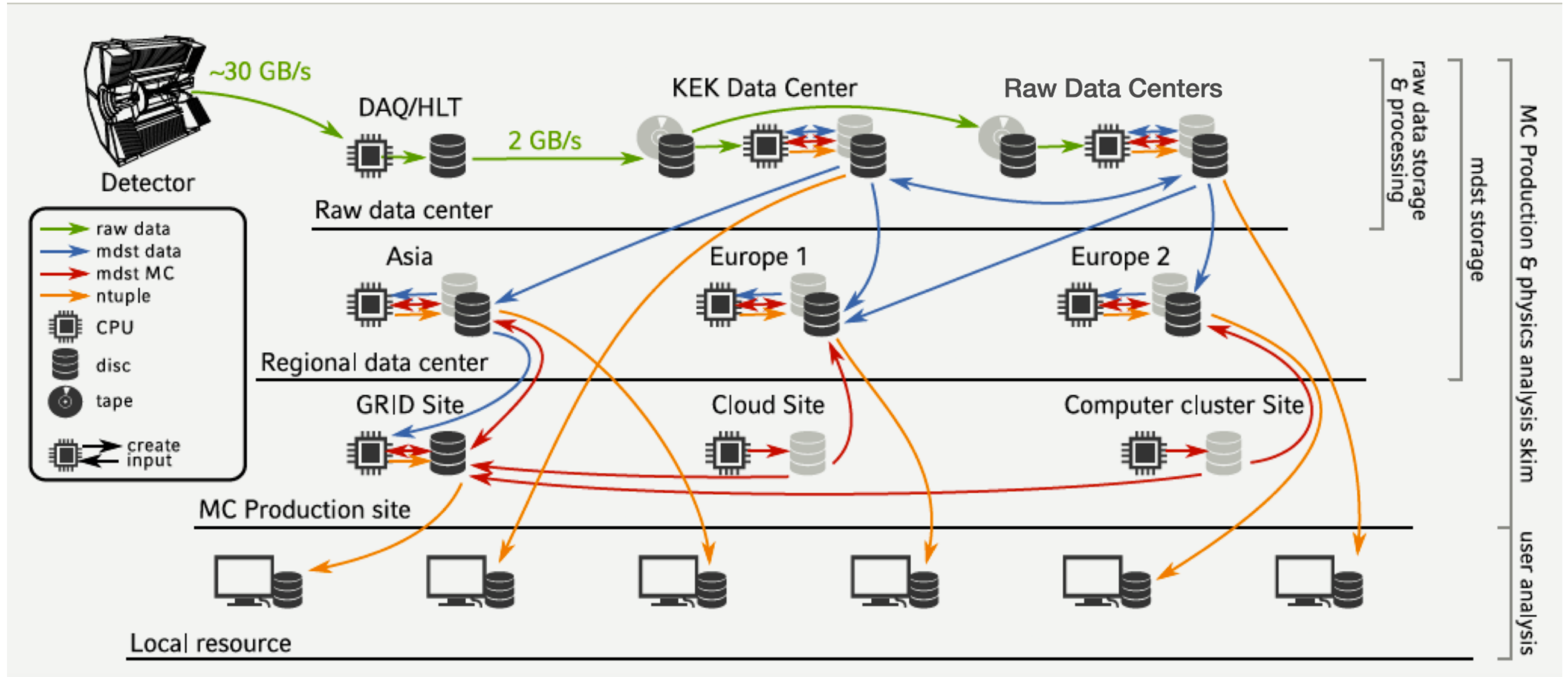
THE UNIVERSITY of
MISSISSIPPI

Belle II Distributed Computing

- Belle II data processing and analysis requires lots of computing power
- 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



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 EL9/Rocky9/AlmaLinux9 (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/qcampaign/public/B2SW2025/gb2_example.py (can be downloaded or directly copied at KEKCC)
 - Reconstructs candidate events for
$$B^0 \rightarrow [D^{*-} \rightarrow [D^0 \rightarrow K^- \pi^+, K^- K^+] \pi^-] \pi^+$$
 - Input file: /home/belle2/qcampaign/public/B2SW2025/mdst_000528_prod00049804_task290000530.root

CLINIC/PHARMACY NAME Belle II	
ADDRESS 1-1 Oho, Tsukuba, Ibaraki, Japan 305-0801	
PHONE NUMBER	
RX # 491584891	DATE June 24, 2025
PATIENT NAME Belle II user	
DIRECTIONS Test your steering script locally in basf2 before submitting to the grid	
DRUG NAME AND STRENGTH	
gbasf2	
QTY 100 mg	
PRESCRIBER Belle II Distributed Computing	

WARNINGS & PRECAUTIONS
Failure to test steering scripts can lead to:
Job Failure
Grid Slowdown
Problems at Computing Sites
Sad Computing Experts
Increased risk of cancer, as recognized by the state of California
Death

```
$ source /cvmfs/belle.cern.ch/tools/b2setup light-2505-deimos
```

```
$ 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 EL9/Rocky9/AlmaLinux9 (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 qcampaign b2_belle2 1606 Apr 25 07:27 usercert.pem  
-rw----- 1 qcampaign b2_belle2 1743 Apr 25 07:27 userkey.pem
```



Your user key should only be readable by you
If not, use '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/bashrc
```

- 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 probably shouldn't
 - If you REALLY need/want to, use the instructions [here](#)

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 project_name -s
light-2505-deimos -cputime 6000
```

```
*****
***** Project summary *****
** Project name: B2SW25_example_project
** Dataset path: /belle/user/qcampaign/B2SW25_example_project
** Steering file: gb2_example.py
** Job owner: qcampaign @ belle (23:52:02)
** 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: 6000 min
*****
Are you sure to submit the project?
Please enter Y or N:
```

Other gb2_tools

- 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_get             gb2_ds_quota           gb2_ds_siteForecast
gb2_job_reschedule     gb2_prod_dataset       gb2_proxy_destroy      gb2_ds_sync
gb2_diagnostic         gb2_ds_list            gb2_ds_rep             gb2_ds_sync
gb2_job_status         gb2_prod_releases      gb2_proxy_info         gb2_job_delete
gb2_ds_collection      gb2_ds_put             gb2_ds_rep_status      gb2_job_delete
gb2_list_destse        gb2_prod_status        gb2_proxy_init         gb2_job_kill
gb2_ds_count_events    gb2_ds_query_datablock gb2_ds_rm              gb2_job_kill
gb2_list_site          gb2_prod_summary       gb2_se_list            gb2_job_output
gb2_ds_du              gb2_ds_query_dataset   gb2_ds_rm_rep          gb2_job_output
gb2_prod_accounted     gb2_project_analysis   gb2_update             gb2_job_parameters
gb2_ds_generate        gb2_ds_query_file      gb2_ds_search          gb2_job_parameters
gb2_prod_campaigns     gb2_project_summary
```

Monitoring jobs - command line tools

- Summary of all jobs/projects - good summary

```
$ gb2_project_summary -u qcampaign
```

Project	Owner	Status	Done	Fail	Run	Wait	Submission Time(UTC)	Duration
b2luigi_test_137e344b2ff	qcampaign	Good	1	0	0	0	2025-06-13 21:20:20	00:15:27
Lc2SKs_chargedprompt97a31538dc7	qcampaign	Good	3522	0	0	0	2025-06-13 23:28:05	04:44:26
Lc2SKs_mixedprompt93d22595d27	qcampaign	Good	3461	0	0	0	2025-06-13 23:37:13	06:27:46
Lc2SKs_chargedproc1399229cbe84a	qcampaign	Good	5198	0	0	0	2025-06-13 23:48:42	09:14:21
Lc2SKs_ddbarproc1391ad4668a06	qcampaign	Good	4377	0	0	0	2025-06-14 00:00:14	11:52:14
Lc2SKs_ssbarprompt92502920793	qcampaign	Good	2658	0	0	0	2025-06-14 00:09:49	12:40:50
Lc2SKs_ccbar4S_offres9a71e236053	qcampaign	Good	766	0	0	0	2025-06-14 00:16:14	16:13:50
Lc2SKs_ccbar5S_scan9ce9233742b	qcampaign	Good	612	0	0	0	2025-06-14 00:22:34	17:13:24
Lc2SKs_ssbar4S_offres99ef5704b52	qcampaign	Good	355	0	0	0	2025-06-14 00:28:55	13:03:28
Lc2SKs_mixedproc139f14bb9ab9a	qcampaign	Good	5109	0	0	0	2025-06-14 00:41:30	16:14:11
Lc2SKs_ccbarprompt940ea43532d	qcampaign	Good	5534	0	0	0	2025-06-14 00:52:07	18:05:44
Lc2SKs_uubarprompt9f19c28a2d6	qcampaign	Good	5415	0	0	0	2025-06-14 01:03:40	18:57:13
Lc2SKs_uubar5S_scan90ccb598c91	qcampaign	Good	599	0	0	0	2025-06-14 01:09:53	24:16:19
Lc2SKs_uubar4S_offres9a0b7cadbe0	qcampaign	Good	751	0	0	0	2025-06-14 01:16:55	21:37:12
Lc2SKs_ddbarprompt999277f0ca9	qcampaign	Good	2688	0	0	0	2025-06-14 01:26:12	18:42:24
Lc2SKs_ssbarproc1391d46995131	qcampaign	Good	4333	0	0	0	2025-06-14 01:38:49	21:38:17
Lc2SKs_ddbar5S_scan9652011e4e6	qcampaign	Good	302	0	0	0	2025-06-14 01:44:56	23:34:02
Lc2SKs_ddbar4S_offres998286c3af3	qcampaign	Good	364	0	0	0	2025-06-14 01:51:13	24:09:16
Lc2SKs_uubarproc139808ab74ecd	qcampaign	Good	7059	0	0	0	2025-06-14 02:06:27	24:59:42
Lc2SKs_ssbar5S_scan94fbde32b68	qcampaign	Good	298	0	0	0	2025-06-14 02:12:30	25:25:10
Lc2SKs_ccbarproc1395b15b54934	qcampaign	Bad	7159	1	0	0	2025-06-14 02:27:18	27:15:37
Gbasf2Task_B2SW25b4feed6a98	qcampaign	Good	3	0	0	0	2025-06-19 23:16:26	00:07:54
B2SW25_example_project	qcampaign	Waiting	0	0	0	1	2025-06-23 03:03:03	00:00:00

Monitoring jobs - command line tools

- More detailed status with 'gb2_job_status' using -p (project), -j (job ID), or -u (user)

```
$ gb2_job_status -p B2SW25_example_project
Job id      Status    MinorStatus  ApplicationStatus      Site
=====
491451579   Running   Application  Running                 LCG.CNAF.it

--- Summary of Selected Jobs ---
Completed:0 Deleted:0 Done:0 Failed:0 Killed:0 Running:1 Scouting:0 Stalled:0 Waiting:0
```

Monitoring jobs - WebApp

Limit time span

Job status

More detailed status

Computing Site

Menu

Desktops&Applications

Tools

Applications

Accounting

Component History

Configuration Manager

Downtimes

File Catalog

Job Monitor

Job Summary

Pilot Monitor

Pilot Summary

Proxy Manager

Public State Manager

Registry Manager

Request Monitor

Resource Summary

Site Summary

Space Occupancy

System Administration

Transformation Monitor

BelleDIRAC Apps

Help

DIRAC

My Desktops

Shared

Settings

Job Monitor [Untitled 1]

Selectors

Site:

Status:

Minor Status:

Application Status:

Owner: qcampaign

OwnerGroup:

Job Group:

Job Type:

Time Span: For all time?

JobID(s):

Submit

Reset

Refresh

Items per page: 25

Page 1 of 2463

Updated: -

Displaying topics

☐	JobId ↓	Status	M...	Application...	Site	J...	LastUpdate[UTC]	LastSignOfLife[UTC]	Submis
☐	491451579	Done	E...	Done	LCG.CNAF.it	B...	2025-06-23 03:08:13	2025-06-23 03:08:12	2025-0
☐	490998405	Done	E...	Done	LCG.Pisa.it	G...	2025-06-19 23:23:19	2025-06-19 23:23:18	2025-0
☐	490998404	Done	E...	Done	LCG.Pisa.it	G...	2025-06-19 23:24:32	2025-06-19 23:24:31	2025-0
☐	490998403	Done	E...	Done	LCG.Napoli.it	G...	2025-06-19 23:25:09	2025-06-19 23:25:08	2025-0
☐	489288763	Done	E...	Done	OSG.BNL.us	L...	2025-06-14 16:18:28	2025-06-14 16:18:28	2025-0
☐	489288762	Done	E...	Done	LCG.KEK.jp	L...	2025-06-14 05:29:17	2025-06-14 05:29:18	2025-0
☐	489288761	Done	E...	Done	LCG.KEK.jp	L...	2025-06-14 05:22:11	2025-06-14 05:22:12	2025-0
☐	489288760	Done	E...	Done	OSG.BNL.us	L...	2025-06-14 19:47:19	2025-06-14 19:47:18	2025-0
☐	489288759	Done	E...	Done	OSG.BNL.us	L...	2025-06-14 22:33:44	2025-06-14 22:33:44	2025-0
☐	489288758	Done	E...	Done	OSG.BNL.us	L...	2025-06-15 05:46:45	2025-06-15 05:46:44	2025-0
☐	489288757	Done	E...	Done	OSG.BNL.us	L...	2025-06-15 09:18:13	2025-06-15 09:18:13	2025-0
☐	489288756	Done	E...	Done	LCG.KEK.jp	L...	2025-06-14 21:09:26	2025-06-14 21:09:26	2025-0
☐	489288755	Done	E...	Done	LCG.DESY...	L...	2025-06-15 02:40:25	2025-06-15 02:40:24	2025-0
☐	489288754	Done	E...	Done	OSG.BNL.us	L...	2025-06-15 02:18:16	2025-06-15 02:18:15	2025-0
☐	489288753	Done	E...	Done	OSG.BNL.us	L...	2025-06-15 04:18:29	2025-06-15 04:18:28	2025-0
☐	489288752	Done	E...	Done	LCG.KEK.jp	L...	2025-06-17 21:05:49	2025-06-17 21:05:49	2025-0
☐	489288751	Done	E...	Done	LCG.KEK.jp	L...	2025-06-17 20:48:05	2025-06-17 20:48:06	2025-0
☐	489288750	Done	E...	Done	LCG.KEK.jp	L...	2025-06-17 20:50:00	2025-06-17 20:50:00	2025-0
☐	489288749	Done	E...	Done	LCG.DESY...	L...	2025-06-14 10:50:56	2025-06-14 10:50:55	2025-0
☐	489288748	Done	E...	Done	LCG.DESY...	L...	2025-06-14 11:29:39	2025-06-14 11:29:38	2025-0
☐	489288747	Done	E...	Done	LCG.KEK.jp	L...	2025-06-17 21:16:36	2025-06-17 21:16:36	2025-0

What to do if your jobs have failed

- If your jobs failed, maybe 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

☐	JobId ↓	Status	M..	ApplicationStatus	Site	J...	LastUpdate
☐	491342581	Failed	F...	Failed scout job [4913404...	ANY	m...	2025-06-21
☐	491342580	Failed	F...	Failed scout job [4913404...	ANY	m...	2025-06-21
☐	491342579	Failed	F...	Failed scout job [4913404...	ANY	m...	2025-06-21
☐	491342578	Failed	F...	Failed scout job [4913404...	ANY	m...	2025-06-21
☐	491342577	Failed	F...	Failed scout job [4913404...	ANY	m...	2025-06-21
☐	491342576	Failed	F...	Failed scout job [4913404...	ANY	m...	2025-06-21
☐	491342575	Failed	F...	Failed scout job [4913404...	ANY	m...	2025-06-21

☐	JobId ↓	Status	M..	ApplicationStatus	Site
☐	491342581	Failed	F...	Failed scout job [4913404...	ANY

JDL

Attributes

Parameters

Logging info

Peek StandardOutput

Get LogFile

Get Pending Request

Get StagerReport

Actions >

Pilot >

SandBox >

Get input file(s)

Get output file(s)

- Can also use command line tools

```
$ gb2_job_output -j 491342581
```

What to do if your jobs have failed

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

```
$ gb2_diagnostic --failed_job 491340453
INFO ##### General Info #####
INFO Timestamp: 2025-06-23 03:17:46.475761 UTC
INFO proxyInfo : OK
INFO DIRAC username: qcampagn
INFO DIRAC groupname: belle_dcops
INFO User DN: /C=JP/O=KEK/OU=CRC/CN=CAMPAGNA Quinn
INFO Rucio ping: {'version': '35.6.0'}
INFO #####
```

Proxy information

```
INFO
diagnostic of Failed Job .....
INFO ##### Job Summary #####
INFO Failed JobID: 491340453
INFO Minor status: Application Finished With Errors
INFO Application Status: basf2helper.py Exited With Status 3
INFO #####
```

Job Summary

```
INFO
##### Logging Info #####
Source          Status    MinorStatus      ApplicationStatus      DateTime
=====
JobManager      Submitting Bulk transaction confirmation Unknown      2025-06-21 22:07:25
JobPath         Checking  Scouting          Unknown      2025-06-21 22:09:25
Scouting        Checking  JobSanity          Unknown      2025-06-21 22:09:25
JobSanity       Checking  InputData          Unknown      2025-06-21 22:09:25
InputData       Checking  JobScheduling      Unknown      2025-06-21 22:09:26
JobScheduling   Waiting   Pilot Agent Submission Unknown      2025-06-21 22:09:26
Matcher         Matched   Assigned           Unknown      2025-06-21 22:09:26
JobAgent@LCG.DESY.de Matched   Job Received by Agent Unknown      2025-06-21 22:09:27
JobAgent@LCG.DESY.de Matched   Installing Software Unknown      2025-06-21 22:09:30
JobAgent@LCG.DESY.de Matched   Submitting To CE   Unknown      2025-06-21 22:09:30
JobWrapper      Running   Job Initialization Unknown      2025-06-21 22:09:37
JobWrapper      Running   Downloading InputSandbox Unknown      2025-06-21 22:09:40
JobWrapper      Running   Input Data Resolution Unknown      2025-06-21 22:09:48
JobWrapper      Running   Application         Unknown      2025-06-21 22:10:08
Job_491340453   Running   Application         Executing RunScriptStep1 2025-06-21 22:10:13
basf2           Running   Application         Running      2025-06-21 22:10:20
Job_491340453   Running   Application         basf2helper.py Exited With Status 3 2025-06-21 22:10:32
Job_491340453   Running   Application         basf2helper.py Exited With Status 3 2025-06-21 22:10:33
JobWrapper      Completing Application Finished With Errors basf2helper.py Exited With Status 3 2025-06-21 22:10:40
JobWrapper      Failed   Application Finished With Errors basf2helper.py Exited With Status 3 2025-06-21 22:10:43
```

Logging info

```
INFO #####
INFO ##### job output #####
INFO ##### Error in std.out #####
ERROR 2025-06-21 22:10:31 UTC dirac-jobexec/BelleScript ERROR: 'basf2helper.py' Exited With Status 3

ERROR 2025-06-21 22:10:31 UTC dirac-jobexec/BelleScript ERROR: Non-zero status while executing /var/lib/condor/execute/dir_1036203/DIRAC_712c_n9jpilot/491340453/basf2helper.py
gb2_mcdslnu_epi_vtx_add-mcdslnu_epi_vtx_addv3_s.py light-2505-deimos mcdslnu_epi_vtx_addv3 exited with status 3

INFO ### Error in Script1_basf2helper.py.log ###
ERROR 2025-06-21 22:10:30 UTC Unknown ERROR: basf2 crashed (3)

ERROR ERROR: The execution of basf2 gb2_mcdslnu_epi_vtx_add-mcdslnu_epi_vtx_addv3_s.py -i mdst_000001_prod00047868_task11000001.root --job-information=basf2_output.json -n 10
failed.

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)

					Items per page: 25			Page 1
☐	JobId ↓		Status	M..	ApplicationStatus			
☒	491446403		Failed	A...	basf2helper.py Exited With St...			
☐	491440764		Failed	A...	basf2helper.py Exited With St...			
☐	491439005		Failed	A...	basf2helper.py Exited With St...			
☐	491438536		Failed	A...	Unknown			

After your jobs have completed

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

```
$ gb2_ds_list  
/belle/user/qcampaign/B2SW25_example_project
```

```
$ gb2_ds_list /belle/user/qcampaign/B2SW25_example_project/sub00  
/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_B0_00000_job491451579_00.root  
/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_D0_00000_job491451579_00.root  
/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_Dstar_00000_job491451579_00.root
```

After your jobs have completed

```
$ gb2_ds_get --new B2SW25_example_project
Do you want to download 3 files:
Please type [Y] or [N]: y
2025-06-23 03:27:26,897 INFO Processing 3 item(s) for input
2025-06-23 03:27:31,419 INFO Using 3 threads to download 3 files
2025-06-23 03:27:31,420 INFO Thread 0/3: Preparing download of user.qcampaign:/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_B0_00000_job491451579_00.root
2025-06-23 03:27:31,420 INFO Thread 1/3: Preparing download of user.qcampaign:/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_D0_00000_job491451579_00.root
2025-06-23 03:27:31,420 INFO Thread 0/3: Trying to download with davs and timeout of 360s from Napoli-TMP-SE: user.qcampaign:/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_B0_00000_job491451579_00.root
2025-06-23 03:27:31,420 INFO Thread 2/3: Preparing download of user.qcampaign:/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_Dstar_00000_job491451579_00.root
2025-06-23 03:27:31,420 INFO Thread 1/3: Trying to download with davs and timeout of 360s from Napoli-TMP-SE: user.qcampaign:/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_D0_00000_job491451579_00.root
2025-06-23 03:27:31,421 INFO Thread 2/3: Trying to download with davs and timeout of 360s from Napoli-TMP-SE: user.qcampaign:/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_Dstar_00000_job491451579_00.root
2025-06-23 03:27:31,444 INFO Thread 2/3: Using PFN: davs://dcache-belle-01.na.infn.it:443/dpm/na.infn.it/home/belle/TMP/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_Dstar_00000_job491451579_00.root
2025-06-23 03:27:31,444 INFO Thread 0/3: Using PFN: davs://dcache-belle-01.na.infn.it:443/dpm/na.infn.it/home/belle/TMP/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_B0_00000_job491451579_00.root
2025-06-23 03:27:31,445 INFO Thread 1/3: Using PFN: davs://dcache-belle-01.na.infn.it:443/dpm/na.infn.it/home/belle/TMP/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_D0_00000_job491451579_00.root
2025-06-23 03:27:35,880 INFO Thread 0/3: File user.qcampaign:/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_B0_00000_job491451579_00.root successfully downloaded. 20.835 kB in 3.46 seconds = 0.01 MBps
2025-06-23 03:27:35,901 INFO Thread 2/3: File user.qcampaign:/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_Dstar_00000_job491451579_00.root successfully downloaded. 34.355 kB in 3.74 seconds = 0.01 MBps
2025-06-23 03:27:36,420 INFO Thread 1/3: File user.qcampaign:/belle/user/qcampaign/B2SW25_example_project/sub00/basf2_example2_D0_00000_job491451579_00.root successfully downloaded. 256.566 kB in 4.25 seconds = 0.06 MBps

#####
Download Summary
-----
Total files: 3
Sucessfully Downloaded files: 3
Files already found locally: 0
Files that Failed to be downloaded: 0
Files with duplicate jobID (not Downloaded): 0
#####
```

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



After you have downloaded your files, clean your grid space

```
$ gb2_ds_rm B2SW25_example_project
LFN(s) to remove:

Dataset
=====
/belle/user/qcampaign/B2SW25_example_project/sub00  3
Do you want to remove the following file(s):
Please type [Y] or [N]: y
LFN
=====
Succeed
/belle/user/qcampaign/B2SW25_example_project/sub00  OK
/belle/user/qcampaign/B2SW25_example_project         OK
```

Finding files to run on

1. For most cases, it is the safest to use recommended collections from the Data Production group
 - For MC15 and 16, collections can be found [here](#)
 - If you want to run over the whole collection, just include its path as the input when submitting your project

```
$ gbasf2 gb2_example.py -i /belle/collection/path/to/collection -p project_name  
-s light-2505-deimos
```

- If you only want to use part of a collection, you can use gb2_ds_search and command line tools to pick out individual datasets from a collection

```
$ gb2_ds_search collection --list_datasets /belle/collection/MC/  
MC16rd_proc16_ddbar_Run1_4S_v1 | grep e0014
```

```
/belle/MC/release-08-00-10/DB00003307/MC16rd_proc16/prod00047191/s00/e0014/4S/  
r00000/ddbar/mdst
```

Finding files to run on



2. If you are an expert or are recommended to do so by an expert, you can use the dataset searcher
 - Can be done through the WebApp by choosing metadata values

The screenshot shows the "Dataset Searcher [Untitled 1]" web application. The interface is divided into a left sidebar and a main content area. The sidebar, titled "Menu", contains a "Desktops&Applications" section with a list of applications including Accounting, Activity Monitor, Component History, Configuration Manager, Downtimes, File Catalog, Job Monitor, Job Summary, Pilot Monitor, Pilot Summary, Proxy Manager, Public State Manager, Registry Manager, Request Monitor, Resource Summary, Site Summary, Space Occupancy, System Administration, Transformation Monitor, BelleDIRAC Apps, and Dataset Searcher (which is highlighted). The main content area has a title bar "Dataset Searcher" and a subtitle "Dataset Searcher via DIRAC" in red. Below the title bar, there are tabs for "Metadata Searcher" and "Tree Browser". The "Metadata Searcher" tab is active, showing a form with various search criteria: "Data Type" (radio buttons for MC and Data), "Background level" (radio buttons for BGx1, BGx0, and Other), "Background level:" (a dropdown menu), "Beam Energies:" (a dropdown menu), "Data Levels:" (a dropdown menu), "Global Tags:" (a dropdown menu), "Experiment High:" (a text input field), "Run High:" (a text input field), "General Skim Names:" (a dropdown menu), "Campaigns:" (a dropdown menu), "Skim Types:" (a dropdown menu), "Releases:" (a dropdown menu), "Experiment Low:" (a text input field), "Run Low:" (a text input field), and "MC Event Types:" (a dropdown menu). At the bottom of the form, there are three buttons: "Clear", "Search", and "Help". Below the form, there is a section labeled "LPN" with a table structure. At the very bottom, there are three status indicators: "Dataset LFNs Metadata" (green checkmark), "Dataset Metadata" (green checkmark), and "Download .txt file" (blue download icon).

Finding files to run on



2. If you are an expert or are recommended to do so by an expert, you can use the dataset searcher

- Can also be done using gb2_ds_search

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}
                                What kind of search is this?
  metadata                   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.
```

Finding files to run on

WARNING
PROCEED WITH
CAUTION

```
$ 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
```

Dataset Searcher via command line

Finding files to run on



Dataset Searcher via command line

```
$ gb2_ds_search collection --list_all_collections /belle/collection/Data/*  
/belle/collection/Data/proc13prompt_skim_15420100_v1  
/belle/collection/Data/proc13prompt_skim_15440100_v1  
/belle/collection/Data/Moriond2023_prompt_all_5Sscan_10657_v1  
/belle/collection/Data/Moriond2023_prompt_hadron_5Sscan_10657_v1  
/belle/collection/Data/Moriond2023_prompt_all_5Sscan_10706_v1  
/belle/collection/Data/Moriond2023_prompt_all_5Sscan_10751_v1  
/belle/collection/Data/proc13prompt_skim_16460200_v1
```

Workflow Management- b2luigi

- b2luigi- workflow management software based on a Spotify software, maintained by the collaboration
- Automates job submission, monitoring, and download
- Documentation found [here](#)
- Disclaimer: I only learned how to use this a couple weeks ago, I am definitely not an expert



Constructing a b2luigi script

- Basically a basf2 steering script with a wrapper to create a scheduling task
- We'll work our way through the example script found [here](#)
- First, import everything you would need for your steering script + b2luigi and Basf2PathTask

```
import basf2 as b2
import modularAnalysis as ma
import variables.utils as vu
import vertex as vx

import b2luigi
from b2luigi.basf2_helper import Basf2PathTask
```

Constructing a b2luigi script

- Create your analysis task
 - Can use `b2luigi.Parameter` (or `BoolParameter`) to create variables that can be given as parameters when calling the `AnalysisTask`
 - Must define what system you are using to submit the jobs (in our case `gbasf2`)
- Need to define properties that are required arguments for `gbasf2`
 - Can also define others, full list found [here](#)
- Also need to define where your output will go

```
class AnalysisTask(Basf2PathTask):
    treeFit = b2luigi.BoolParameter(default=True)
    event_type = b2luigi.Parameter(default="mumu")
    result_dir = "results/AnalysisGbasf2"

    batch_system = "gbasf2"

    @property
    def gbasf2_input_dslist(self):
        return "charged.txt"

    @property
    def gbasf2_project_name_prefix(self):
        # Keep projectname shorter than grid requirement
        projectname = "Gbasf2Task"
        return projectname[-21:]

    @property
    def gbasf2_release(self):
        return "light-2409-toyger"

    @property
    def gbasf2_download_logs(self):
        return False

    # Don't use `add_to_output` here, but define the output explicitly.
    def output(self):
        return {"ntuple_gbasf2.root": b2luigi.LocalTarget("ntuple_gbasf2.root")}
```

Constructing a b2luigi script

- Add your steering script to the AnalysisTask
 - Just like with regular gbasf2 submission, test it locally first
- Note: make sure you remove any lines from your path that executes the path (ie `b2.process(path=main)`), as this will trigger local execution instead of submitting to the grid

```
def create_path(self):
    main = b2.Path()

    particle = "mu" if str(self.event_type) == "mumu" else "e"

    ma.inputMdstList(
        environmentType="default",
        # Empty input file list, will be filled by the grid
        filelist=[],
        path=main,
    )

    # The rest of the steeringfile is identical to the local submission
    # with the difference of the correct outputfile name.

    # Fill final state particles
    ma.fillParticleList(decayString=f"{particle}+:fsp", cut="[abs(dr) < 2.0] and [abs(dz) < 4.0] and [kaonID > 0.2] and [Mbc > 5.24] and [abs(deltaE) < 0.2]")

    ma.fillParticleList(
        decayString="K+:fsp", cut="[abs(dr) < 2.0] and [abs(dz) < 4.0] and [kaonID > 0.2] and [Mbc > 5.24] and [abs(deltaE) < 0.2]"
    )

    # reconstruct J/psi -> l+ --
    ma.reconstructDecay(
        decayString=f"J/psi:ll -> {particle}+:fsp {particle}-:fsp", cut="[2.92 < M < 3.2] and [abs(deltaE) < 0.2]"
    )

    # reconstruct B+ -> J/psi K+
    ma.reconstructDecay(
        decayString="B+:jpsik -> J/psi:ll K+:fsp", cut="[Mbc > 5.24] and [abs(deltaE) < 0.2]"
    )

    # Run truth matching
    ma.matchMCTruth(list_name="B+:jpsik", path=main)

    # Perform tree fit if requested
    if self.treeFit:
        vx.treeFit(list_name="B+:jpsik", conf_level=0.00, massConstraint=["J/psi"], path=main)
```

```
# Run truth matching
ma.matchMCTruth(list_name="B+:jpsik", path=main)

# Perform tree fit if requested
if self.treeFit:
    vx.treeFit(list_name="B+:jpsik", conf_level=0.00, massConstraint=["J/psi"], path=main)

# Define variables to save
variables = ["Mbc", "deltaE", "chiProb", "isSignal"]
variables += vu.create_aliases_for_selected(
    list_of_variables=["InvM", "M"], decay_string="B+:jpsik -> ^J/psi:ll K+:fsp", path=main
)

# Write variables into ntuple
ma.variablesToNtuple(
    decayString="B+:jpsik",
    variables=variables,
    filename="ntuple_gbasf2.root",
    treename="Bp",
    path=main,
    storeEventType=False,
)

return main
```

Constructing a b2luigi script

- Execute the b2luigi task
 - This create a Scheduler that submits, monitors, and eventually downloads the jobs
- The first time you run the script, it will submit the project(s) and monitor them
- If you kill it and run it again later, it is smart enough to check that the projects already exist on the grid and skip straight to monitoring
- It will only download the output once all the jobs in a project are Successful

```
if __name__ == "__main__":  
    b2luigi.process(AnalysisTask(), batch=True)
```

Setting up

- It is recommended to use a basf2 virtual environment to run b2luigi
 1. Source your basf2 release
 2. Create your b2venv with whatever release you want
 3. Once it is created, for all future uses you can simply source from the existing venv directory

```
$ source /cvmfs/belle.cern.ch/tools/b2setup light-2505-deimos
Belle II software tools set up at: /cvmfs/belle.cern.ch/tools
Environment setup for release: light-2505-deimos
Central release directory      : /cvmfs/belle.cern.ch/el9/releases/light-2505-deimos
```

```
$ b2venv light-2505-deimos
Setting up virtual environment in directory venv for release light-2505-deimos
```

```
$ source venv/bin/activate
Warning: Changing existing PYTHONPATH from /cvmfs/belle.cern.ch/el9/externals/v02-03-00/
Linux_x86_64/common/root/lib:/cvmfs/belle.cern.ch/el9/releases/light-2505-deimos/lib/
Linux_x86_64/opt:/cvmfs/belle.cern.ch/tools to /cvmfs/belle.cern.ch/tools
Belle II software tools set up at: /cvmfs/belle.cern.ch/tools
Environment setup for release: light-2505-deimos
Central release directory      : /cvmfs/belle.cern.ch/el9/releases/light-2505-deimos
```

Installing b2luigi

- Once inside your virtual environment, use pip to install b2luigi

```
(venv) $ pip install b2luigi
```

- And run your b2luigi script in python

```
$ python b2luigi_test.py
INFO: Informed scheduler that task AnalysisTask_mumu_True_60de05985f has status PENDING
INFO: Done scheduling tasks
INFO: Running Worker with 1 processes
Certificate password:
INFO: Batch process Gbasf2Process running AnalysisTask(treeFit=True, event_type=mumu)

Submitting gbasf2 project: Gbasf2Task_B2SW25b4feed6a98

Using command:
gbasf2 /gpfs/home/belle2/qcampaign/public/B2SW2025/task_files/treeFit=True/event_type=mumu/AnalysisTask/steering_file_wrapper.py -p
Gbasf2Task_B2SW25b4feed6a98 -s light-2505-deimos -f serialized_basf2_path.pkl --input_dslist /gpfs/home/belle2/qcampaign/public/B2SW2025/charged.txt
--force

*****
***** Project summary *****
** Project name: Gbasf2Task_B2SW25b4feed6a98
** Dataset path: /belle/user/qcampaign/Gbasf2Task_B2SW25b4feed6a98
** Steering file: /gpfs/home/belle2/qcampaign/public/B2SW2025/task_files/treeFit=True/event_type=mumu/AnalysisTask/steering_file_wrapper.py
** Additional input sandbox files: ['serialized_basf2_path.pkl']
** Job owner: qcampaign @ belle (23:53:27)
** Preferred site / SE: None / None
** Input files for first job: LFN:/belle/MC/release-08-03-00/DB00003308/MC16rd_prompt/prod00049792/s00/e0030/4S/r00000/charged/mdst/sub00/
mdst_000291_prod00049792_task290000291.root
** Number of input files: 3
** Number of jobs: 3
** Processed data (MB): 6729
** Processed events: 654872 events
** Estimated CPU time per job: 3639 min
*****
Initialize metadata for the project:
No attribute. Initialize Dataset...
Dataset initialization: OK
Dataset metadata attributes already exist (30): OK
Successfully finished.
JobID = [490998403, 490998404, 490998405]
Jobs in gbasf2 project "Gbasf2Task_B2SW25b4feed6a98" at 2025-06-20 08:16:27: 'Waiting': 3
```

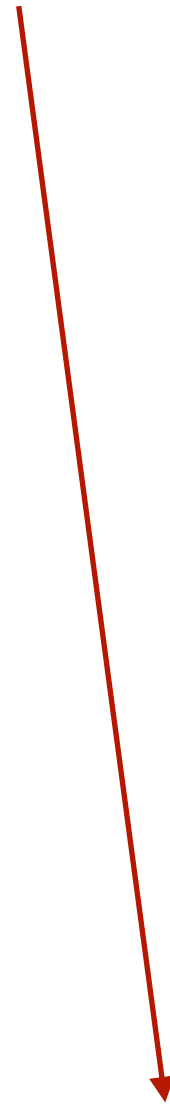
Other stuff about b2luigi

- In principle there is a way to generalize this to multiple projects in one script, although I haven't fully figured that out yet
 - I have it most of the way there, it submits and monitors all the projects but gets weird about the download
 - I've included the script in the B2SW2025 folder in case you want to play around with it
- There is also a way to automate rescheduling jobs, although I haven't included that in the example scripts
- I enthusiastically recommend using b2luigi for project submission and monitoring
- Still recommend although not as enthusiastically for downloads
 - b2luigi doesn't start downloading files until the project is fully complete, which can take a while
 - If you're in a hurry, maybe it is more useful to you to still download manually
- Overall, this is a great tool that will help you keep track of your grid jobs easily
 - Already working with the developers to make the experience even better



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/light-2505-deimos/sphinx/skim/doc/04-experts.html#skim-registry>

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 xWiki page](#)
 - [Computing glossary](#)

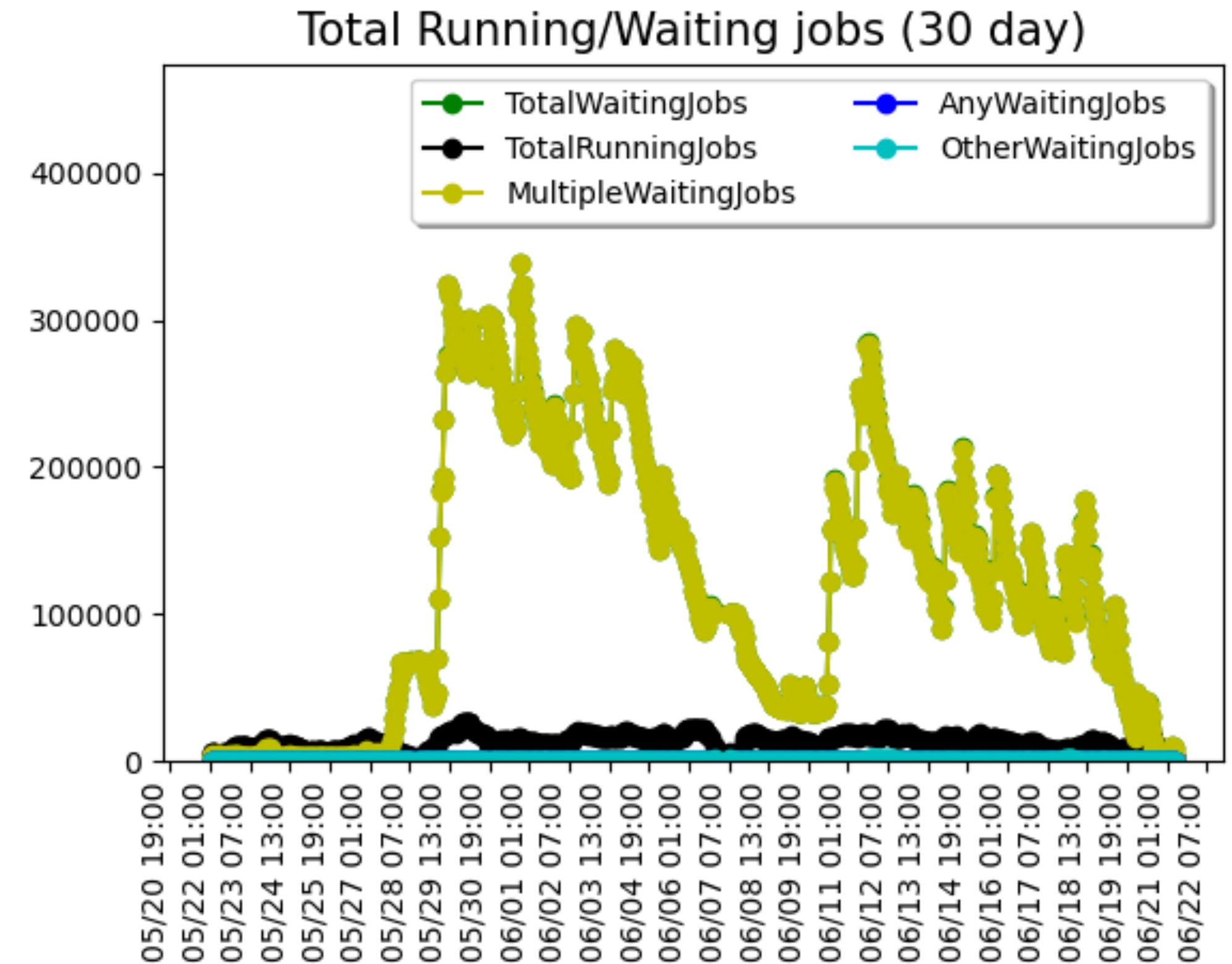


Final remarks

Total Running Job = 12741

Total Waiting Job = 28456

- 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!



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

If you find the easter egg hidden in here let me know

Backup Slides: Some Common Comp-Users-
Forum problems and solutions/advice

Jobs Running/Waiting for Too Long

- Has it been less than a week?
 - Wait
- Has it been assigned to a site?
 - Yes
 - Wait, we don't have control over how sites schedule jobs
 - No
 - Did you set the cputime when you submitted your job?
 - If not, consider resubmitting with the cputime included
- Has it been longer than 2 weeks?
 - Are all of the jobs in the project stuck in Waiting/not finishing?
 - Yes
 - Contact expert
 - No
 - Is there a steady stream of jobs completing?
 - If yes, wait
 - If not, contact expert

Failed Jobs

- Use `gb2_diagnostic --failed_job`
- Most of the time it will be a problem with the steering script
- Could be that your basf2 release can't be run on certain sites
 - See [here](#) to see what operating systems your release is guaranteed to run on, then use `--resource_tag` at job submission to restrict the possible sites
- Could be problem with Conditions Database
 - Notify the experts
- Often, there is just a problem at the site
 - Keep rescheduling until you get a new site or the problem resolves

Maximum Number of Reschedulings Reached

- Check that your jobs aren't failing for some important reason
 - See previous slide
- If not, ask experts to reset your jobs

Failed downloads

- Use `gb2_diagnostic --failed_download`
- Usually a problem with the site
 - Try using `gb2_ds_rep` to create a replica of your file on a different site
- If the diagnostic tool doesn't show you any problems, it was probably just a transient issue, try downloading again
- If this doesn't work, contact experts