

Methods for automating error handling in the Event Picking Service of the ATLAS experiment.

*Joint Institute for
Nuclear Research
(Dubna)*

***E. Alexandrov,
I. Alexandrov,
D. Barberis,
A. Yakovlev***



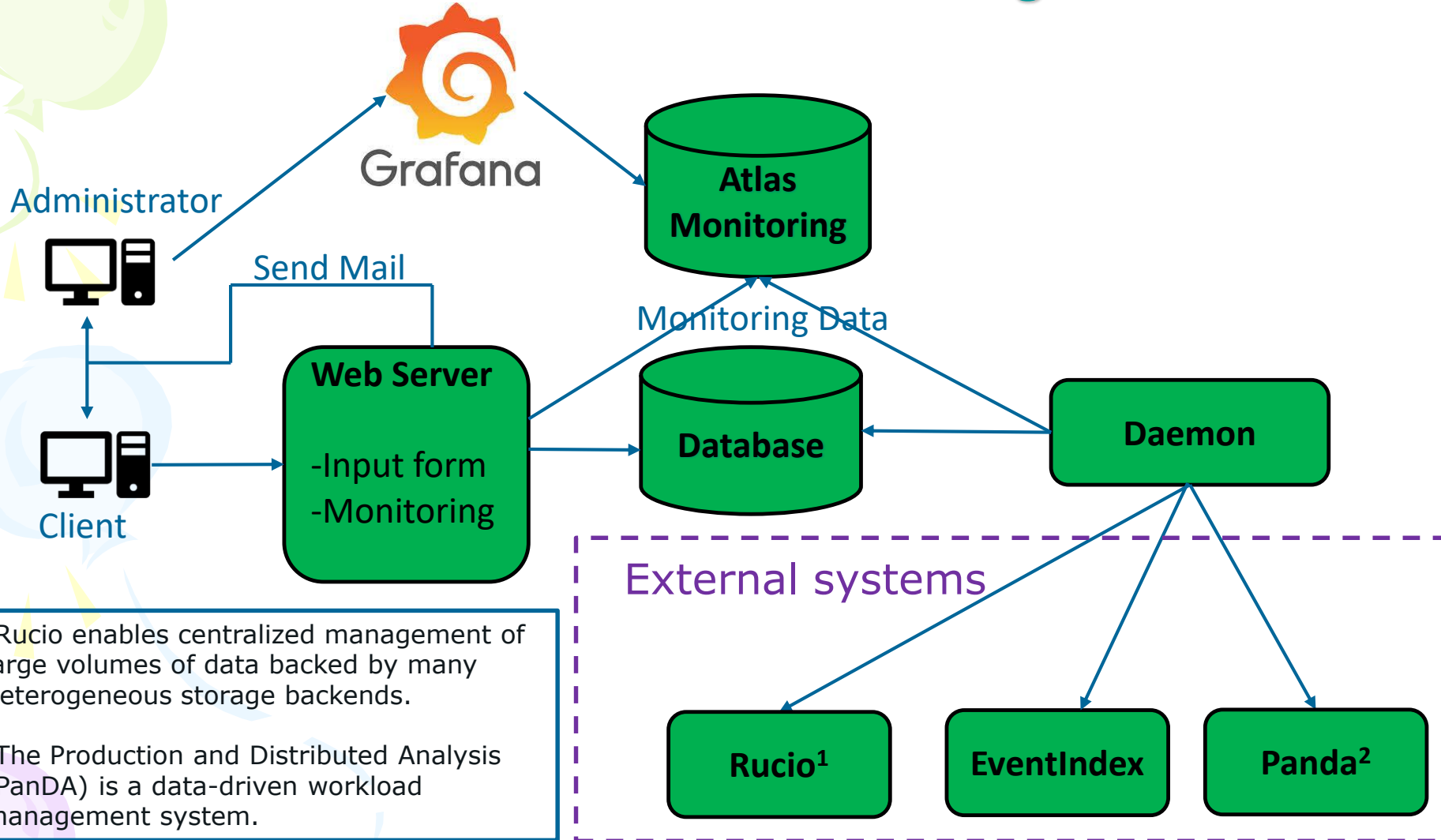
ATLAS EventIndex

- **The EventIndex is the global catalogue of all ATLAS events**
- **For each event, each data format and each processing version, it contains:**
 - **Event identifiers (run and event number);**
 - **Location (GUID of the file containing it) and provenance;**
 - **Trigger and other useful metadata;**
- **Main use case is event picking for detailed analysis and/or displays**

Event Picking Service

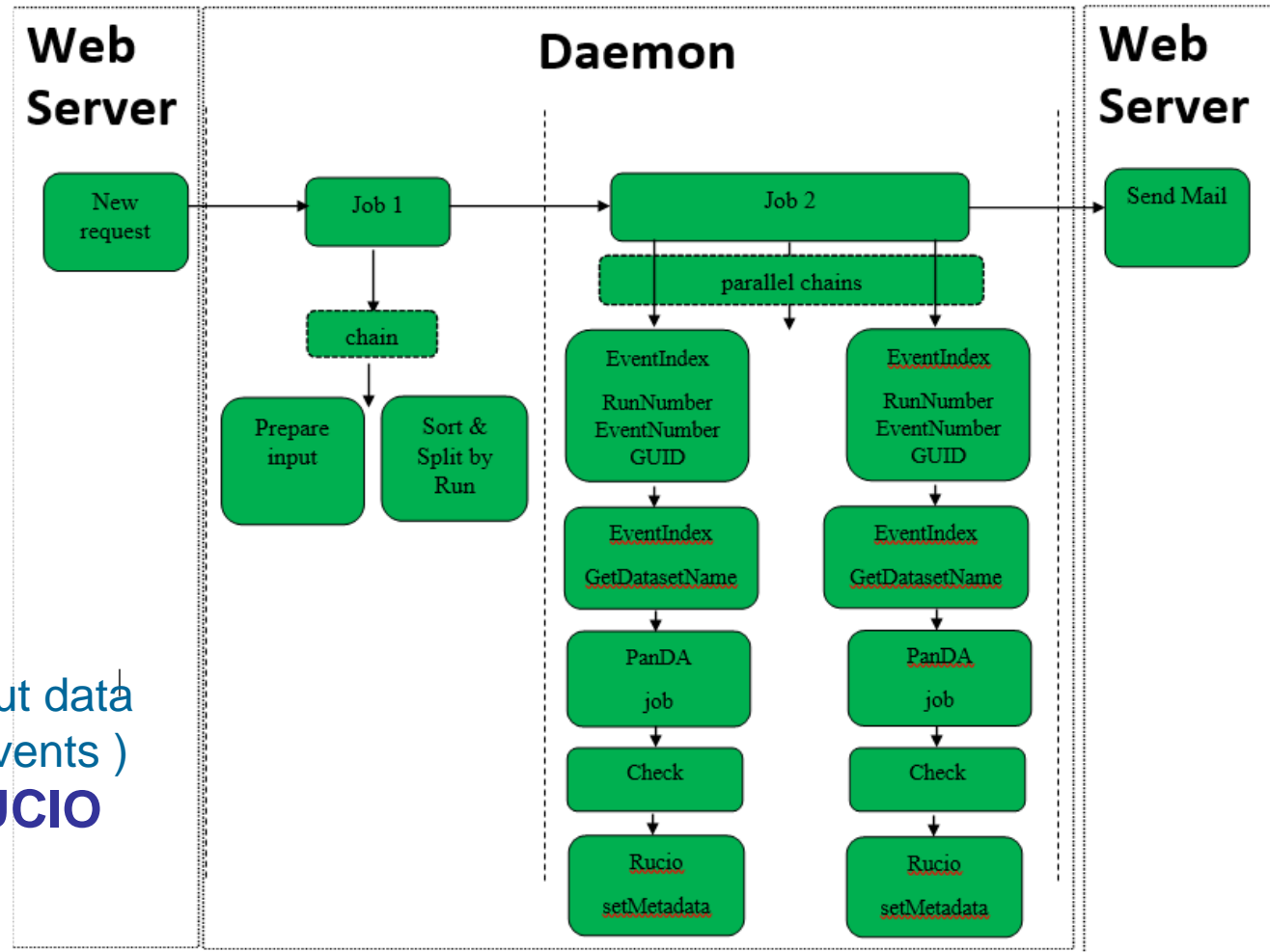
- **Some physics analyses need to extract many events in order to process them with enhanced algorithms**
 - **$\gamma\gamma \rightarrow WW$ analysis:**
 - The first round 50k events (2019)
 - The second round 136k events (2021)
 - **$B_c^* \rightarrow B_c$ gamma analysis:**
 - 16K events (2023)
 - **$Z \rightarrow \text{TauTau}$ selections:**
 - 11K events (2023)
 - **R2R4 Milestone test:**
 - 1M RAW events (2025)
 - **“Pixel dE/dx” analysis:**
 - ~3M RAW events (2025)

Architecture of the Event Picking Service



Event Picking Service : Main stages

- **Get request**
- **Split tasks by Run**
- **Get GUIDs**
(require for PanDa job)
- **Get Dataset Name**
(require for PanDa job)
- **Start PanDA Job**
(Real copy events)
- **Verification**
(should validate the output data possible duplicate/ skip events)
- **Set Metadata in RUCIO**
- **Send mail**
(user notification)

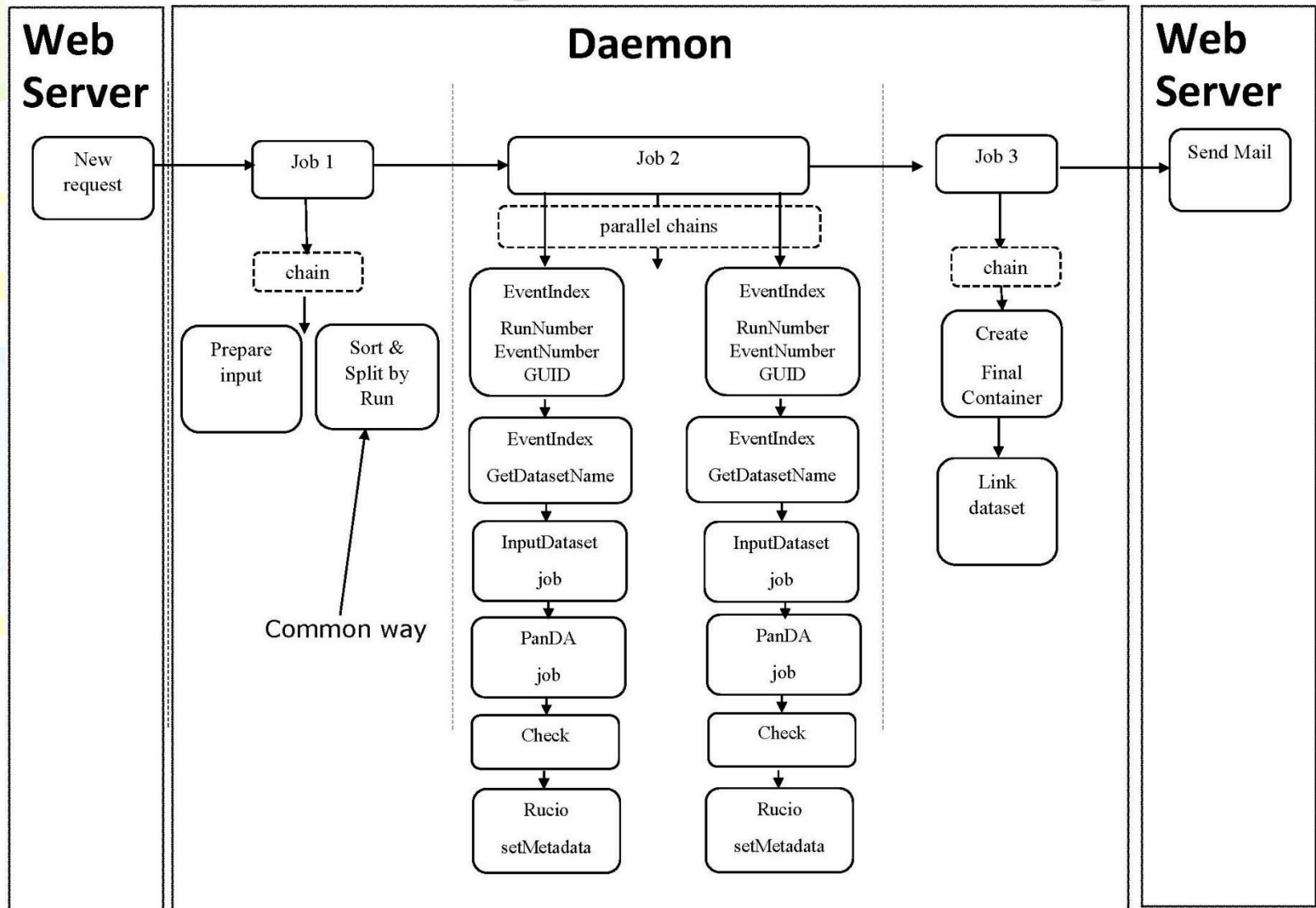


10.07.2025

Current scheme !!

In the new scheme the set of components will be changed.

Event Picking Service : Main stages



New scheme under development

Event Picking Service : error handling

Optimal case: (no errors)

- all chains (and tasks in the chain) are completed;
- all external services worked correctly;
- all necessary data received and collected;

L	124 350121	Chain finished (state: DONE_CHAIN)	02.06.2025 18:13:12 ⬇	03.06.2025 05:03:24 ⬇	Finalized (Done without error and warning)	
Tasks:	Task Name	Initial state	Current state	Created	Changed	Task status
L	getIndex	Panda job (state: PANDA_PART)	Get dataset name for first guid (state: GET_DATASET_NAME)	02.06.2025 18:13:13 ⬇	03.06.2025 01:36:27 ⬇	Finalized (Done without error and warning)
L	getDatasetName	Get dataset name for first guid (state: GET_DATASET_NAME)	Chain workflow. Get GUIDs from EventIndex (state: INDEXED_TASK)	03.06.2025 01:36:27 ⬇	03.06.2025 01:36:49 ⬇	Finalized (Done without error and warning)
L	panda	Chain workflow. Get GUIDs from EventIndex (state: INDEXED_TASK)	Chain workflow. Start panda task (state: START_PANDA_TASK) PanDA job	03.06.2025 01:36:49 ⬇	03.06.2025 03:43:40 ⬇	Finalized (Done without error and warning)
L	check	Chain workflow. Start panda task (state: START_PANDA_TASK)	Set metadata to rucio (state: Metadata)	03.06.2025 03:43:40 ⬇	03.06.2025 04:50:05 ⬇	Finalized (Done without error and warning)
L	metadata	Set metadata to rucio (state: Metadata)	Chain finished (state: DONE_CHAIN)	03.06.2025 04:50:06 ⬇	03.06.2025 05:03:24 ⬇	Finalized (Done without error and warning)

Event Picking Service : error handling

Real cases:

During the processing of a request, various errors may occur. These errors can be grouped by the "severity" of their impact on the final result and by the methods of error handling.

L	166 362661	Error (state: ERROR)	02.06.2025 18:13:56 ↓	18.06.2025 15:41:33	task finished with error (Finished with fatal error)	
Tasks:	Task Name	Initial state	Current state	Created	Changed	Task status
L	getIndex	Panda job (state: PANDA_PART)	Get dataset name for first guid (state: GET_DATASET_NAME)	02.06.2025 18:13:59 ↓	03.06.2025 08:57:47 ↓	Finalized (Done without error and warning)
L	getDatasetName	Get dataset name for first guid (state: GET_DATASET_NAME)	Chain workflow. Get GUIDs from EventIndex (state: INDEXED_TASK)	03.06.2025 08:57:47 ↓	03.06.2025 08:58:25 ↓	Finalized (Done without error and warning)
L	panda	Chain workflow. Get GUIDs from EventIndex (state: INDEXED_TASK)	Error (state: ERROR) <u>PanDA job</u>	03.06.2025 08:58:25 ↓	18.06.2025 15:41:33 ↓	java.lang.Exception: Panda time out error (Error state of task)

EPS error handling : Automatically restart task

- external services may not respond;
- connection to an external service may be lost during operation;
 - such errors have a fatal effect on the result;
 - however, such errors can be resolved automatically or manually;

L	35 356259	Chain finished (state: DONE_CHAIN)	18.05.2025 21:35:38 	19.05.2025 16:33:25 	Finalized (Done with restart task)		
	Tasks:	Task Name	Initial state	Current state	Created	Changed	Task status
	L	getIndex	Panda job (state: PANDA_PART)	Panda job (state: PANDA_PART)	18.05.2025 21:35:38 	19.05.2025 05:31:22	Wait restart (Error state of task)
	L	getIndex	Panda job (state: PANDA_PART)	Panda job (state: PANDA_PART)	19.05.2025 05:41:23 	19.05.2025 05:41:25	Wait restart (Error state of task)
	L	getIndex	Panda job (state: PANDA_PART)	Get dataset name for first guid (state: GET_DATASET_NAME)	19.05.2025 05:51:25 	19.05.2025 06:35:22 	Finalized (Done without error and warning)
	L	getDatasetName	Get dataset name for first guid (state: GET_DATASET_NAME)	Chain workflow. Get GUIDs from EventIndex (state: INDEXED_TASK)	19.05.2025 06:35:24 	19.05.2025 06:36:09 	Finalized (Done without error and warning)
	L	panda	Chain workflow. Get GUIDs from EventIndex (state: INDEXED_TASK)	Chain workflow. Start panda task (state: START_PANDA_TASK) PanDA job	19.05.2025 06:36:10 	19.05.2025 15:30:44 	Finalized (Done without error and warning)
	L	check	Chain workflow. Start panda task (state: START_PANDA_TASK)	Set metadata to rucio (state: Metadata)	19.05.2025 15:30:45 	19.05.2025 16:31:20 	Finalized (Done without error and warning)
	L	metadata	Set metadata to rucio (state: Metadata)	Chain finished (state: DONE_CHAIN)	19.05.2025 16:31:22 	19.05.2025 16:33:24 	Finalized (Done without error and warning)

EPS error handling : Automatically restart task

In this case, the automatic restart failed, since the permissible number of repeated restarts has been exceeded.

L	32 360348	Chain finished (state: DONE_CHAIN)	18.05.2025 21:35:38 ↓	19.05.2025 05:31:18 ↓	Finalized (Done with error in chain)
---	--------------	---------------------------------------	--------------------------	--------------------------	---

Tasks:	Task Name	Initial state	Current state	Created	Changed	Task status
L	getIndex	Panda job (state: PANDA_PART)	Panda job (state: PANDA_PART)	18.05.2025 21:35:38 ↓	19.05.2025 04:41:11	Wait restart (Error state of task)
L	getIndex	Panda job (state: PANDA_PART)	Panda job (state: PANDA_PART)	19.05.2025 04:51:12 ↓	19.05.2025 04:51:13	Wait restart (Error state of task)
L	getIndex	Panda job (state: PANDA_PART)	Panda job (state: PANDA_PART)	19.05.2025 05:01:14 ↓	19.05.2025 05:01:14	Wait restart (Error state of task)
L	getIndex	Panda job (state: PANDA_PART)	Panda job (state: PANDA_PART)	19.05.2025 05:11:14 ↓	19.05.2025 05:11:15	Wait restart (Error state of task)
L	getIndex	Panda job (state: PANDA_PART)	Panda job (state: PANDA_PART)	19.05.2025 05:21:15 ↓	19.05.2025 05:21:16	Wait restart (Error state of task)
L	getIndex	Panda job (state: PANDA_PART)	Panda job (state: PANDA_PART)	19.05.2025 05:31:17 ↓	19.05.2025 05:31:18	Max number of restart (Error state of task)

If chains cannot be restarted automatically, they will terminate with an error status. Data from such chains were not received, the received dataset of events is incomplete

EPS error handling : Manually restart request

Restart request - In terms of the Event Picking Service, it means creating a new request using the input from the parent request.

Request view :

Data format: RAW	Request ID: 468	Category: Normal
Project name: data18_13TeV	Client name: Dario Barberis	Created: 18.05.2025 21:35 WebInterface
Stream: physics_Main	Client e-mail: dario.barberis@cern.ch	Last change: 02.06.2025 15:18 ReqFinJob with reqId=482
AMI tag:		Current status: Error
Input file (run and event numbers):		Detail status: Error (Done with error in chain)

Restart request

Restart request : Request ID: 468

"Restart request" - In terms of the Event Picking Service, it means **creating a new request** using the input from the parent request. A new request will be launched with **your credentials**.

Attention !! There are several options for restarting a request with different options.

Parent request	* User (client) name: Aleksandr Iakovlev
Request ID: 468	* User e-mail (identifier): alexander.iakovlev@cern.ch
Data format: RAW	
Stream: physics_Main	
AMI tag:	
Created: 18.05.2025 21:35 WebInterface	
Last change: 02.06.2025 15:18 ReqFinJob with reqId=482	
Current status: Error	<input checked="" type="radio"/> Full Restart
Detail status: Error (Done with error in chain)	<input type="radio"/> Restart Warning & Error branches
Client name: Dario Barberis	<input type="radio"/> Restart only check tasks
Client e-mail: dario.barberis@cern.ch	

Restart request Clear form

Close

Three types of restart

EPS error handling : Manually restart request

Full restart: The request starts from the beginning, the results from the parent request are not used in the new request.

Restart Warning & Error branches: Only chains with the "error" status are restarted. The new request will contain the processing results only for these chains.

Jobs:	Job Name	Initial state	Current state	Created	Changed	Job Status
L	SplitJob	Sort and split (state: SORT_SPLIT)	Panda job (state: PANDA_PART)	18.05.2025 21:35 ↓	18.05.2025 21:35 ↓	Finalized. Total=1, errors=0, warnings=0 (Done without error and warning)
L	PandaJob	Panda job (state: PANDA_PART)	Error (state: ERROR)	18.05.2025 21:35 ↓	02.06.2025 15:22 ↓	Running job:137 from 741 (Done with error in chain) <div>Ok chains: 562 Error chains: 176 Warning chains: 0 Working chains: 0</div>

Chains:	Chain number / Run number	Current state	Created	Changed	Chain status
L	46 358541	Chain finished (state: DONE_CHAIN)	18.05.2025 21:35:39 ↓	19.05.2025 08:03:42 ↓	Finalized (Done without error and warning)
L	47 357713	Panda job (state: PANDA_PART)	18.05.2025 21:35:40 ↓	18.05.2025 21:35:40	Finalized by ext req (Done with error in chain)
L	48 357713	Chain finished (state: DONE_CHAIN)	18.05.2025 21:35:40 ↓	20.05.2025 15:15:41 ↓	Finalized (Done without error and warning)
L	49 357713	Panda job (state: PANDA_PART)	18.05.2025 21:35:40 ↓	18.05.2025 21:35:40	Finalized by ext req (Done with error in chain)
L	50 357713	Panda job (state: PANDA_PART)	18.05.2025 21:35:40 ↓	18.05.2025 21:35:40	Finalized by ext req (Done with error in chain)
L	51 359010	Chain finished (state: DONE_CHAIN)	18.05.2025 21:35:40 ↓	19.05.2025 13:08:49 ↓	Finalized (Done without error and warning)

EPS error handling : Check errors

Data received from PanDA should be checked for missing or duplicate events.

- there may be cases where the files returned by PanDA are not currently available. In this case, an error occurs and the user is returned a list of missing events.
 - the error is not fatal, the data has already been received;
 - just need to repeat the verification step again;



















1081625	Proccess for start check with name: root://eosatlas.cern.ch:1094//eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/31/5b/group.proj-evind.44762906._000119.data.root has not output:/tmp/ei_servise/78794/out.txt	20.05.2025 04:47:48
1081630	Proccess for start check with shot name has error.	20.05.2025 04:47:50
1081634	Proccess for start check with shot name: /eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/e9/df/group.proj-evind.44762906._000120.data.root has not output	20.05.2025 04:47:51
1081639	Proccess for start check has error. Required restart.	20.05.2025 04:47:52
1081641	Proccess for start check with name: root://eosatlas.cern.ch:1094//eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/e9/df/group.proj-evind.44762906._000120.data.root has not output:/tmp/ei_servise/78794/out.txt	20.05.2025 04:47:53
1081644	Total number:1000 Skipped:577 Duplicated:0	20.05.2025 04:47:54
1081647	Skipped:["364292, 2412205137","364292, 2733270339","364292, 2439529669","364292, 2482182713","364292, 2501483795","364292, 2762200468","364292, 2483966366","364292, 2469104494","364292, 2917824897","364292, 2479482621","364292, 2489277455","364292, 2614637835","364292, 2695903805","364292, 2593771044","364292, 2858145627","364292, 2573212787","364292, 2418328216","364292, 2476902281","364292, 2643868833","364292, 2917659097","364292, 2512342344","364292, 2446959016","364292, 2513040762","364292, 2931736569","364292, 2430718709","364292, 2547707197","364292, 2421418139","364292, 2702973976","364292, 2473502220","364292, 2789042959","364292, 2610514783","364292, 2666764858","364292, 2776620198","364292, 2931893122","364292, 2524353688","364292, 2938656327","364292, 2866040081","364292, 2435956106","364292, 2736489149","364292, 2543350838","364292, 2414636480","364292, 2755056497","364292, 2793448378","364292, 2509531706","364292, 2737968870","364292, 2658956278","364292, 2740271357","364292, 2557991642","364292, 2649450011","364292, 2902433360","364292, 2734835915","364292, 2646482844","364292, 2477087919","364292, 2695455482","364292, 2696689619","364292, 2731772043","364292, 2817397413","364292, 2711362677","364292, 2742045791","364292,	20.05.2025 04:47:54

EPS error handling : Check errors

Special type of restart for this case (Restart only check tasks)

Only those chains with check errors are started, starting only with the "check" task.

In this case, PanDA data is taken from the parent request.

Chains:	Chain number / Run number	Current state	Created	Changed	Chain status																		
L	1 360402	Chain finished (state: DONE_CHAIN)	24.06.2025 09:50:52 	24.06.2025 10:10:30 	Finalized (Done without error and warning)																		
L	2 363910	Chain finished (state: DONE_CHAIN)	24.06.2025 09:50:52 	24.06.2025 10:15:31 	Finalized (Done without error and warning)																		
<div><div>Tasks:</div><table><tr><th>Task Name</th><th>Initial state</th><th>Current state</th><th>Created</th><th>Changed</th><th>Task status</th></tr><tr><td>L check</td><td>Check job (state: CHECK_OUTPUT)</td><td>Set metadata to rucio (state: Metadata)</td><td>24.06.2025 09:50:52 </td><td>24.06.2025 10:11:34 </td><td>Finalized (Done without error and warning)</td></tr><tr><td>L metadata</td><td>Set metadata to rucio (state: Metadata)</td><td>Chain finished (state: DONE_CHAIN)</td><td>24.06.2025 10:11:34 </td><td>24.06.2025 10:15:31 </td><td>Finalized (Done without error and warning)</td></tr></table></div>						Task Name	Initial state	Current state	Created	Changed	Task status	L check	Check job (state: CHECK_OUTPUT)	Set metadata to rucio (state: Metadata)	24.06.2025 09:50:52 	24.06.2025 10:11:34 	Finalized (Done without error and warning)	L metadata	Set metadata to rucio (state: Metadata)	Chain finished (state: DONE_CHAIN)	24.06.2025 10:11:34 	24.06.2025 10:15:31 	Finalized (Done without error and warning)
Task Name	Initial state	Current state	Created	Changed	Task status																		
L check	Check job (state: CHECK_OUTPUT)	Set metadata to rucio (state: Metadata)	24.06.2025 09:50:52 	24.06.2025 10:11:34 	Finalized (Done without error and warning)																		
L metadata	Set metadata to rucio (state: Metadata)	Chain finished (state: DONE_CHAIN)	24.06.2025 10:11:34 	24.06.2025 10:15:31 	Finalized (Done without error and warning)																		
L	3 360209	Chain finished (state: DONE_CHAIN)	24.06.2025 09:50:52 	24.06.2025 10:07:05 	Finalized (Done without error and warning)																		

EPS error handling : Algorithm refinement

- **external service returns outdated data;**
 - **the next service in the workflow will not be able to use this information;**
 - **the problem is solved by further refinement of the algorithm;**

EventIndex indexed the file and believed it to exist. However, the file was deleted.

When calling the getIndex service, it may return a path to a file that no longer exists.

To resolve this conflict, the algorithm had to add a call to the AMI service to check whether the file actually exists.

EPS error handling : Sending error messages

(New feature, currently being tested)

For some types of errors in external services, EPS will send an error email either to the official address of such service (if any) or to the actual experts of these services.

The list of experts is created manually by Event Picking Service administrators.

EPS error handling : Sending error messages

When an error occurs, the engine records information about it in a special table in the database.

The EPS periodically monitors this table. When an error occurs, the web interface sends a message to the list of experts according to the type of error.

1	Common errors
6	File_not_created
7	Error_in_file
8	Output_not_created
2	getIndex
9	AMI_check_error
10	Output_lengt_is_null
3	getDatasetName
11	Dataset_not_found
4	panda
12	Not_started
13	Check_status
14	Check_error
5	check
15	Check_duplicated

ID	Method name	Expert name	Expert email	Error type
2	getDatasetName	test2	test2@google.com	Dataset_not_found
1	getIndex	test	test@cern.ch	AMI_check_error
3	panda	test	test3@cern.ch	Check_error

List of message recipients for the corresponding error type

Types of errors specific to a particular task

Three balloons (green, blue, and purple) are positioned vertically on the left side of the slide. Each balloon has a string and several yellow triangular streamers attached to it.

Thanks for attention !