



Development of the ATLAS Event Picking Server

Evgeny Alexandrov¹, Igor Alexandrov¹, Dario Barberis², Fedor Prokoshin¹, Alexander Yakovlev¹ on behalf of the ATLAS collaboration

¹Joint Institute for Nuclear Research, Dubna, Russia.

²University and INFN Genova (Italy).

**GRID Conference at JINR
05 Jule 2021**



Introduction

- The main use case for the EventIndex is event picking, i.e. finding one or a few events in the many billions of ATLAS events, stored in several million files, and extracting them.
- Some physics analyses need to extract many events in order to process them with enhanced algorithms
 - One example in 2019 was the “ $\gamma\gamma 2WW$ ” analysis, which required the extraction of 50k events
 - A second round of that analysis is now requiring 136k events
- An automatic system to extract the requested events and deliver them to the requestors is therefore needed: the Event Picking Service

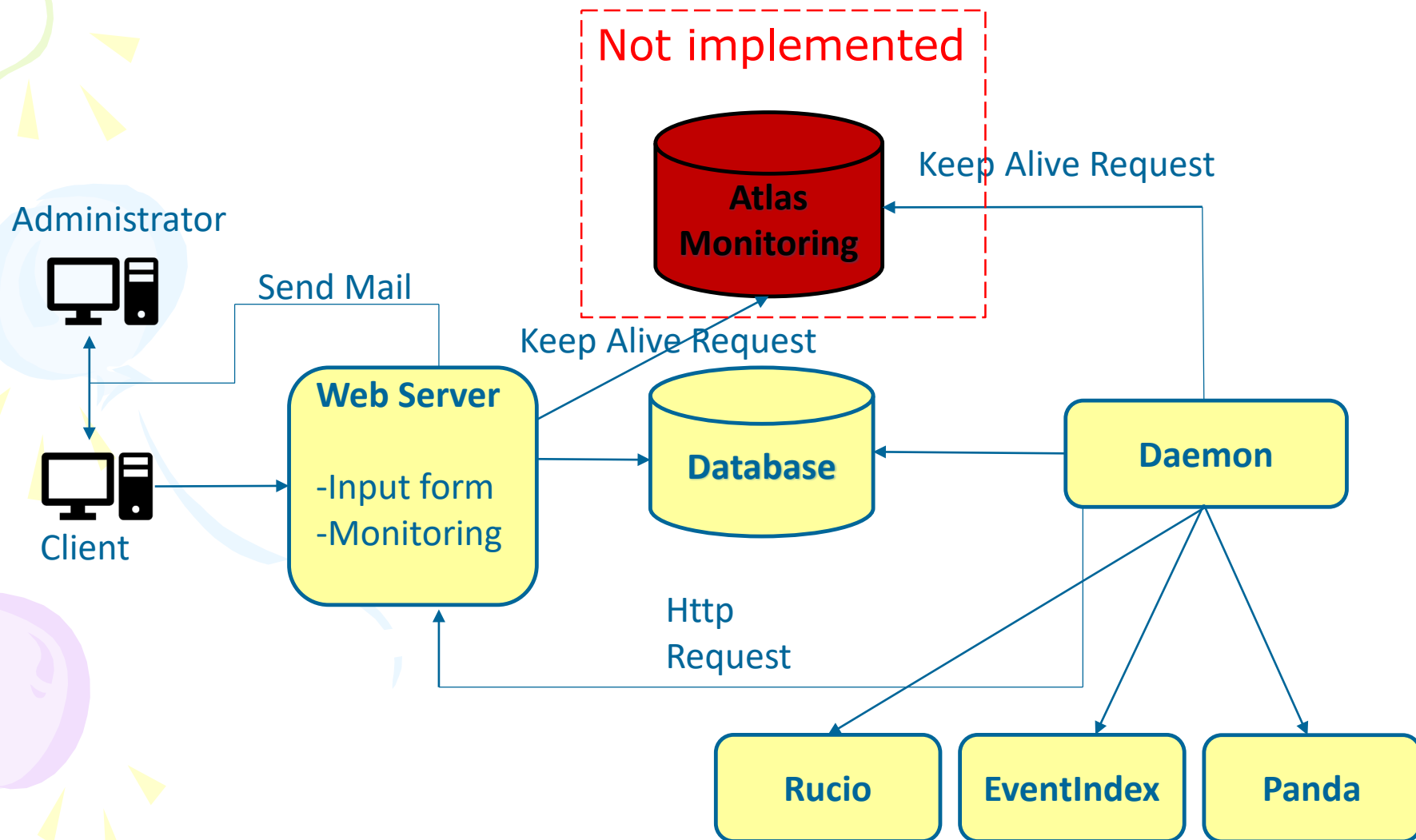


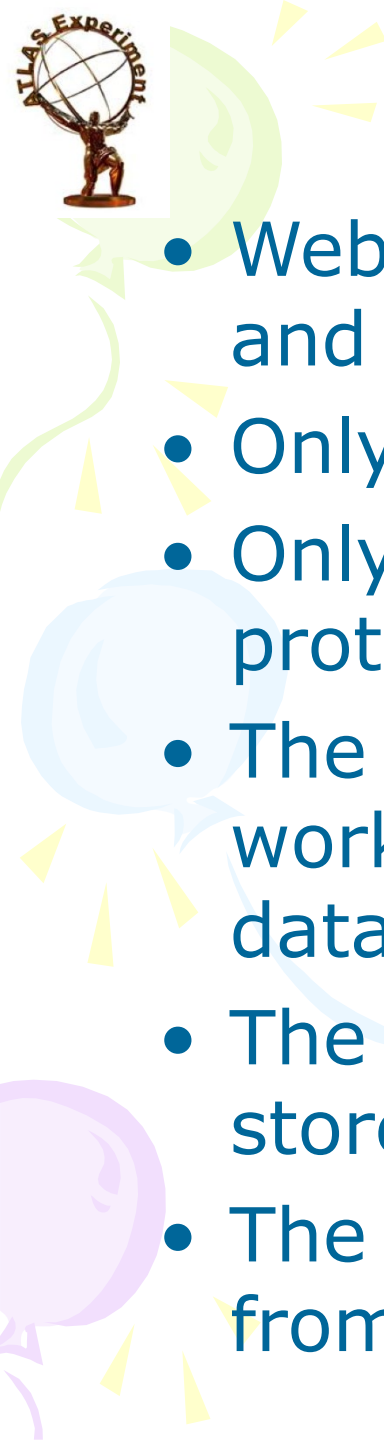
Main Stages

- Split task
- Panda task
 - Copy events
- Verification
 - Check that all run-events exists in output file
- Send mail



Architecture of the Event Picking service





Guidelines



- Web server and Daemon are independent and communicate using database.
- Only web server communicates with Client
- Only the Daemon communicates with protected systems (like Panda, Rucio).
- The Daemon has flexible workflow. The workflow can be modified only using database.
- The results of all steps of workflow are stored in database.
- The input data to any step is the output from the previous one.



Basic terms

Request

Initial data of user

Job

A step of request workflow. Each job contains at least one chain. The number of chains can be different for different jobs. All chains of the job can be executed in parallel mode.

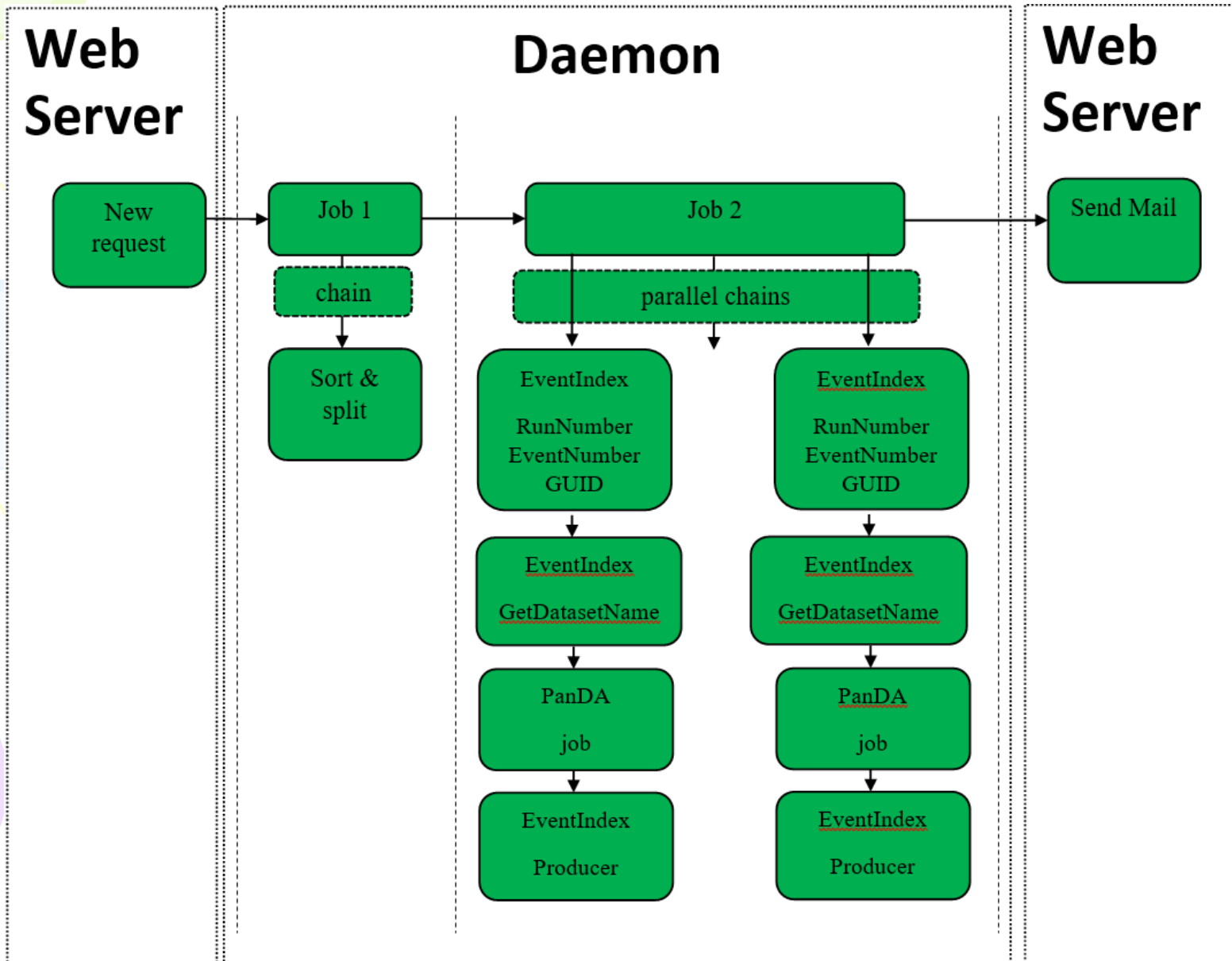
Chain

A combination of tasks with special workflow.

Task

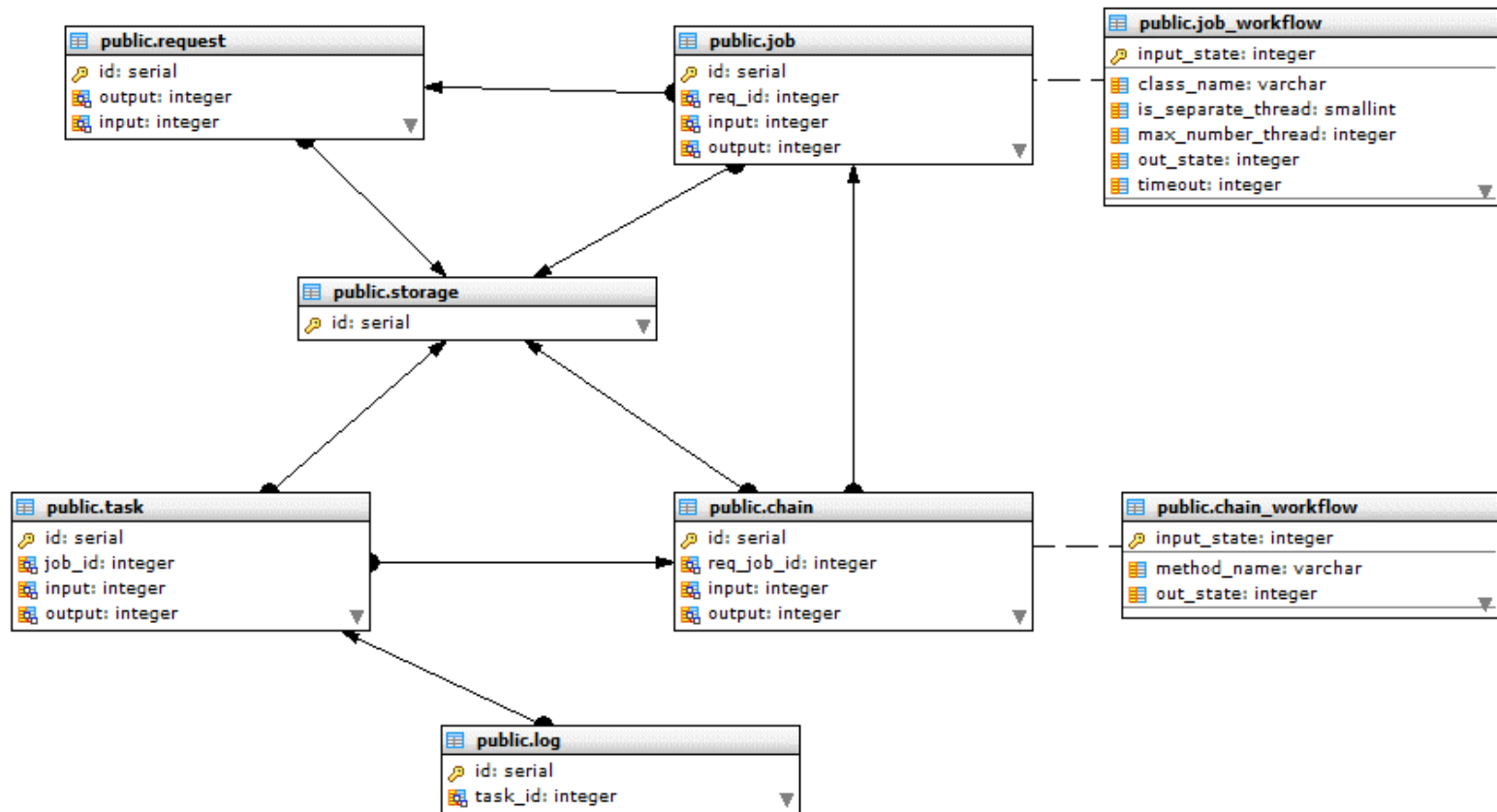
A workflow chain step that performs a real task.

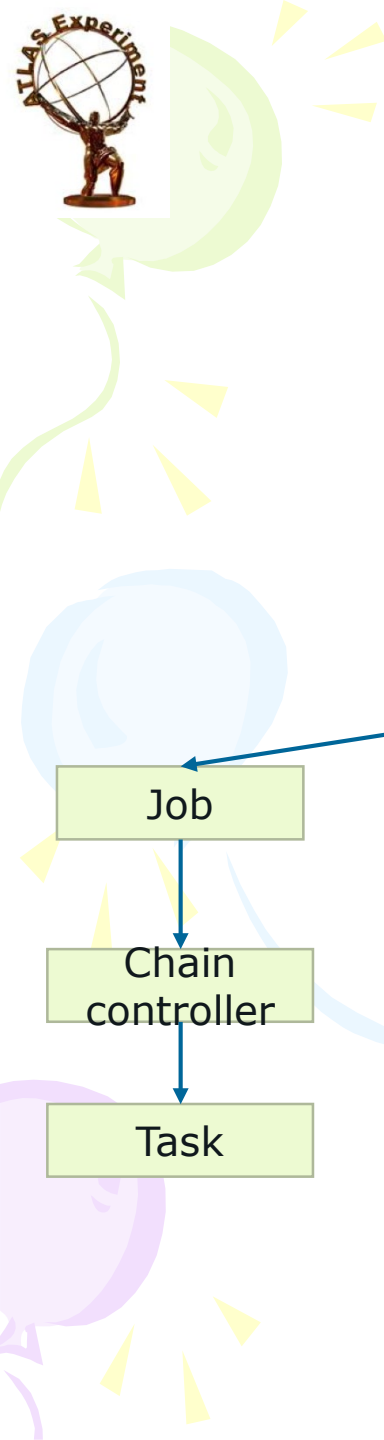
Request Workflow



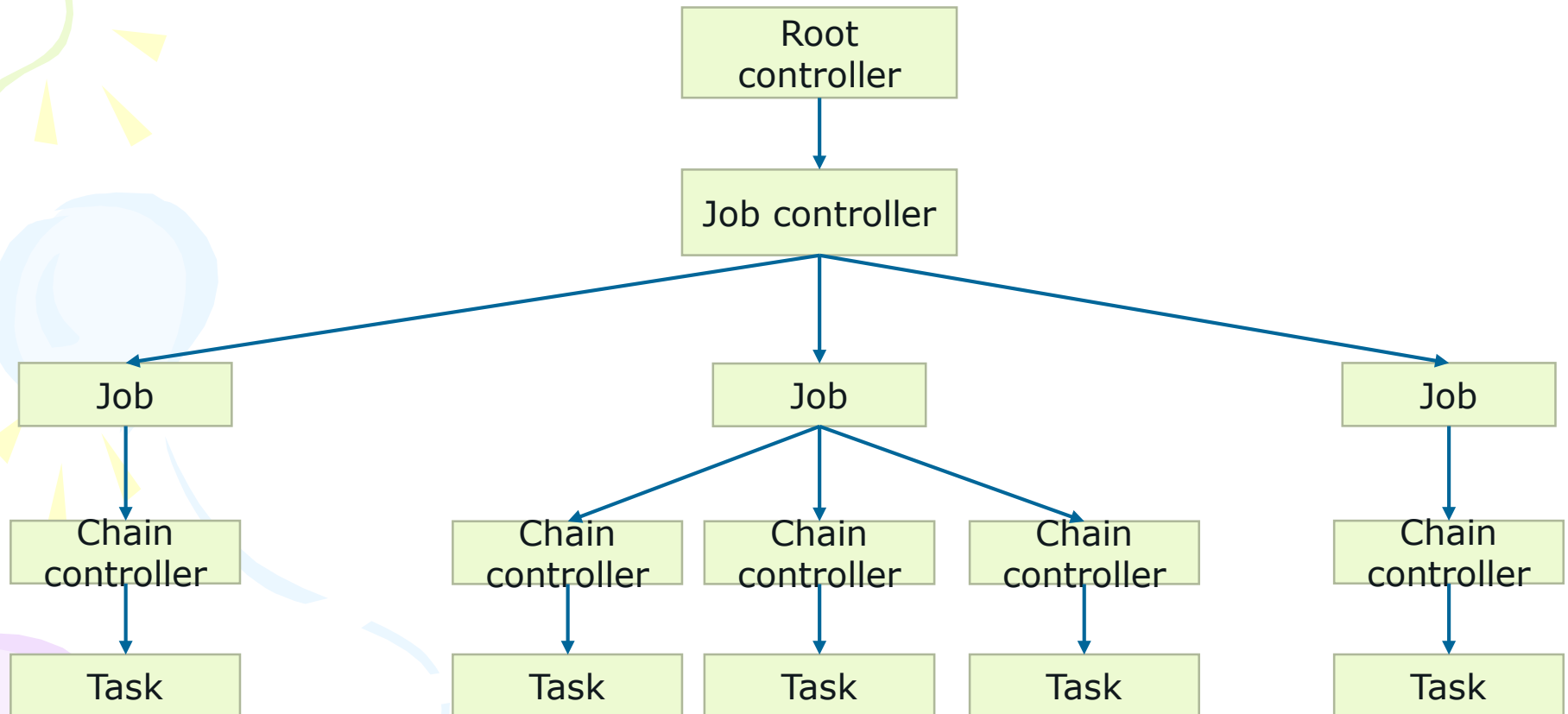


Object model (briefly)





Daemon structure





Web server

Basic technologies:

- PostgreSQL;
- ApacheTomcat;
- JAVA;
- WALT*.

Version 0.3.05 available

*Web Application Lego Toolkit (Developer Sergey Kunyaev from JINR)

Web interface: Create new request

Event Picking Service

+

← → ↻ 🏠

🔒 aiatlas027.cern.ch/ep_test_pg/service

⋮ 🛡️ ☆ 🔍 Search

📶

Event Picking Service v. 0.3.05

Common Info

New Request

Requests Monitoring

Data Tables

* Data format:

Select Data format ▼

Project: ☒ real data ☐ Monte Carlo

Trigger stream:

AMI tag:

File containing run and event numbers :

Browse...

 No file selected.

* User (client) name:

* User e-mail (identifier):

WEB_REQ_ID = 118
web_req_id = "

Submit request

Clear form

Data format - required field - choose one of several formats.

Project - enables (if real data) or disables (if Monte Carlo) the field "Trigger stream".

Trigger stream - the field is active if real data.

AMI tag - the field is active if the data format is not "RAW".

File containing run and event numbers - required field - a text file containing strings like "XXX YYYYY", first column - **run number**; second column - **event number**.

Caution !! If the file format does not meet this requirement, the server will return an error.

User name - required field

User e-mail - required field - e-mail, to which the results of the request processing will be sent.

Web interface: Monitoring requests

Event Picking Service

+

← → ↺ 🏠

aiatlas027.cern.ch/ep_test_pg/service

... 🛡️ ☆

🔍 Search

Event Picking Service v. 0.3.05

Common Info

New Request

Requests Monitoring

Data Tables

Search request

All requests

search filters

Find

Request ID	Client Name	Client E-mail	Data Format	Stream	AMI Tag	CREATED	CHANGED	STATE
1	MailUser	user@server.com	RAW			28.04.2021 15:06, WebInterface	30.05.2021 17:16, Job.start	Request finished
2	test_usewr	test_addr@test_server.com	RAW	trigger_stream		24.05.2021 19:51, WebInterface	24.05.2021 19:52, JobController.run	Critical Error
3	user_to027	mail_to027	RAW			24.05.2021 20:45, WebInterface	24.05.2021 20:45, JobController.run	Critical Error
4	aleksand	aleksand@jinr.ru	RAW			30.05.2021 23:03, WebInterface	31.05.2021 07:04, Job.start	Request finished
5	aleksand	aleksand@jinr.ru	RAW			31.05.2021 07:38, WebInterface	31.05.2021 10:39, Job.start	Request finished
6	aleksand	yakovleva@jinr.ru	RAW			31.05.2021 10:55, WebInterface	31.05.2021 14:10, Job.start	Request finished
7	aleksand	aleksand@jinr.ru	RAW			10.06.2021 16:08, WebInterface	10.06.2021 16:46, Job.run	Critical Error
8	Evgeny	aleksand@jinr.ru	RAW			11.06.2021 09:35, WebInterface	11.06.2021 14:53, Job.run	Request finished

Debug tools:
[LOG log](#) | [LOG zaprosob](#) | [CFG DOC](#) | [CSS](#) | [JS](#) | USER_ID: 4790; FIO:



Web interface: Monitoring requests

Event Picking Service

aiatlas027.cern.ch/ep_test_pg/service

Search

Request ID	Client Name	Client E-mail	Data Format	Stream	AMI Tag	CREATED	CHANGED	STATE
1	MailUser	user@server.com	RAW			28.04.2021 15:06, WebInterface	30.05.2021 17:16, Job.start	Request finished
2	test_usewr	test_addr@test_server.com	RAW	trigger_stream		24.05.2021 19:51, WebInterface	24.05.2021 19:52, JobController.run	Critical Error
3	user_to027	mail_to027	RAW			24.05.2021 20:45, WebInterface	24.05.2021 20:45, JobController.run	Critical Error
4	aleksand	aleksand@jinr.ru	RAW			30.05.2021 23:03, WebInterface	31.05.2021 07:04, Job.start	Request finished
5	aleksand	aleksand@jinr.ru	RAW			31.05.2021 07:38, WebInterface	31.05.2021 10:39, Job.start	Request finished
6	aleksand	yakovleva@jinr.ru	RAW			31.05.2021 10:55, WebInterface	31.05.2021 14:10, Job.start	Request finished
7	aleksand	aleksand@jinr.ru	RAW			10.06.2021 16:08, WebInterface	10.06.2021 16:46, Job.run	Critical Error
8	Evgeny	aleksand@jinr.ru	RAW			11.06.2021 09:35, WebInterface	11.06.2021 14:53, Job.run	Request finished

Data format: RAW

Stream:

AMI tag:

Request ID: 8

Client name: Evgeny

Client e-mail: aleksand@jinr.ru

Created: 11.06.2021 09:35

Last change: 11.06.2021 14:53

Current status: Request finished

WebInterface

Job.run

DatasetName

group.proj-evind.data15_13TeV.00284006.physics_Main.evtpick.DRAW_EVT PICK_n8d674

group.proj-evind.data15_13TeV.00281411.physics_Main.evtpick.DRAW_EVT PICK_n8d675

file_path

/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/5c/f6/group.proj-evind.25845295_000001.event.dat

/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/1d/9b/group.proj-evind.25845294_000004.event.dat /eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/22/0d/group.proj-evind.25845294_000005.event.dat /eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/6e/e9/group.proj-evind.25845294_000002.event.dat /eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/a3/af/group.proj-evind.25845294_000003.event.dat /eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/cf/1c/group.proj-evind.25845294_000001.event.dat /eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/da/e9/group.proj-evind.25845294_000006.event.dat

Request progress

Jobs:	Job Name	Initial state	Current state	Created	Changed	Job Status
L	SplitJob	Sort and split (state: SORT_SPLIT)	Panda job (state: PANDA_PART)	11.06.2021 09:36	11.06.2021 09:37	Finalized
L	PandaJob	Panda job (state: PANDA_PART)	Request finished (state: DONE_REQ)	11.06.2021 09:37	11.06.2021 14:53	Finalized

Web interface: Monitoring requests

Request progress

Jobs:	Job Name	Initial state	Current state	Created	Changed	Job Status
L	SplitJob	Sort and split (state: SORT_SPLIT)	Panda job (state: PANDA_PART)	11.06.2021 09:36 ↓	11.06.2021 09:37 ↓	Finalized
L	PandaJob	Panda job (state: PANDA_PART)	Request finished (state: DONE_REQ)	11.06.2021 09:37 ↓	11.06.2021 14:53 ↓	Finalized

Chains:	Chain number	Current state	Created	Changed	comment
L	1	Chain finished (state: DONE_CHAIN)	11.06.2021 09:37:10 ↓	11.06.2021 14:53:32 ↓	Finalized
L	2	Chain finished (state: DONE_CHAIN)	11.06.2021 09:37:10 ↓	11.06.2021 14:53:59 ↓	Finalized

Tasks:	Task Name	Initial state	Current state	Created	Changed	comment
L	getIndex	Panda job (state: PANDA_PART)	Get dataset name for first guid (state: GET_DATASET_NAME)	11.06.2021 09:37:10 ↓	11.06.2021 09:37:30 ↓	Finalized
L	getDatasetName	Get dataset name for first guid (state: GET_DATASET_NAME)	Chain workflow. Get GUIDs from EventIndex (state: INDEXED_TASK)	11.06.2021 09:37:30 ↓	11.06.2021 09:37:49 ↓	Finalized
L	panda	Chain workflow. Get GUIDs from EventIndex (state: INDEXED_TASK)	Chain workflow. Start panda task (state: START_PANDA_TASK)	11.06.2021 09:37:49 ↓	11.06.2021 14:52:45 ↓	Finalized
L	check	Chain workflow. Start panda task (state: START_PANDA_TASK)	Chain finished (state: DONE_CHAIN)	11.06.2021 14:52:46 ↓	11.06.2021 14:53:58 ↓	Finalized



Result Mail

Dear **Evgeny**,

Your request **8**
completed **successfully**.

Start time at **11.06.2021 09:35**; End time at **11.06.2021 14:53**.

Results :

DatasetName	file_path
group.proj-evind.data15_13TeV.00284006.physics_Main.evtpick.DRAW_EVTPICK._n8d674	/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/5c/f6/group.proj-evind.25845295._000001.event.dat
	/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/1d/9b/group.proj-evind.25845294._000004.event.dat
	/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/22/0d/group.proj-evind.25845294._000005.event.dat
	/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/6e/e9/group.proj-evind.25845294._000002.event.dat
group.proj-evind.data15_13TeV.00281411.physics_Main.evtpick.DRAW_EVTPICK._n8d675	/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/a3/af/group.proj-evind.25845294._000003.event.dat
	/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/cf/1c/group.proj-evind.25845294._000001.event.dat
	/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/da/e9/group.proj-evind.25845294._000006.event.dat

More details can be found at
http://aiatlas027.cern.ch/ep_test_pg/service

Second Round of Event Picking

for yy2WW analysis

New request:

- 593 runs
- 136k events

Results:

- Small request (1 and 4 runs in request) was done without big problems. The small bug fix was done.
- Big request (589 runs):
 - Some runs finished ok
 - Some runs finished with problems, being addressed now
 - not enough quota
 - not found dataset

Conclusion

- After half year of development, prototype of the ATLAS Event Picking Server implemented.
- The development will continue, to adapt for possible errors on different tasks.
- The real task “Second Round of Event Picking for yy2WW analysis” was started.
- The algorithms of service will improve use experience of problem data.