Development, creation and further evolution of the Event Picking Service within the ATLAS Event Index project

A.V. 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

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

Introduction

- Some physics analyses need to extract many events in order to process them with enhanced algorithms
 - γγ -> WW analysis:
 - The first round 50k events (2019)
 - The second round 136k events (2021)
 - B_c* -> B_c gamma analysis:
 - 16K events (2023)
 - Z -> TauTau selections:
 - 11K events (2023)

Event Picking serviceIntroduction

Example is the "γγ→WW" analysis (50k events in RAW format out of the 18 billion events in Run 2 (about 10 million files)) described in https://indico.cern.ch/event/939372/contributions/4017178/attachments/2107421/3544354/spagan_exclWW.pdf

and discussed in the Jira ticket https://its.cern.ch/jira/browse/DATREP-169.

An automatic system to extract the requested events and deliver them to the requestors is therefore needed: the Event Picking Service

User requirement

https://docs.google.com/document/d/1WxB_5OjyAuiTujxblu_GrDm6SpE0BU4B4_lx8zYxdjc/edit#heading=h.2mbya2n6yh2h

Event Picking service: Main stages

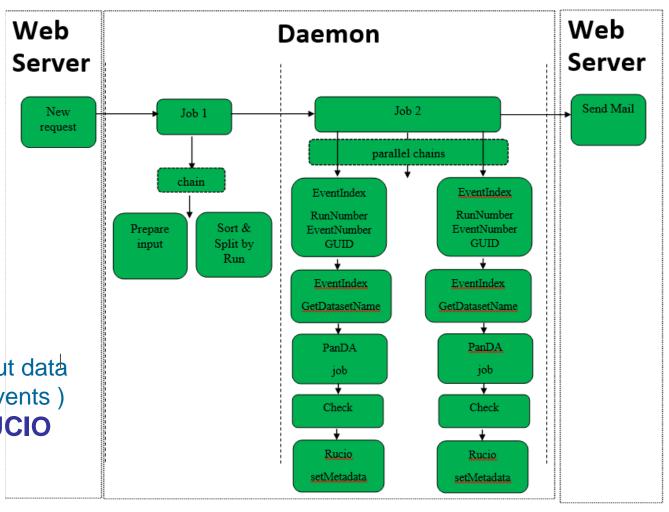
- Get request
- Split tasks by Run
- Get GUIDs
 (require for PanDa job)
- Get Dataset Name (require for PanDa job)
- Start PanDA Job

(Real copy events)

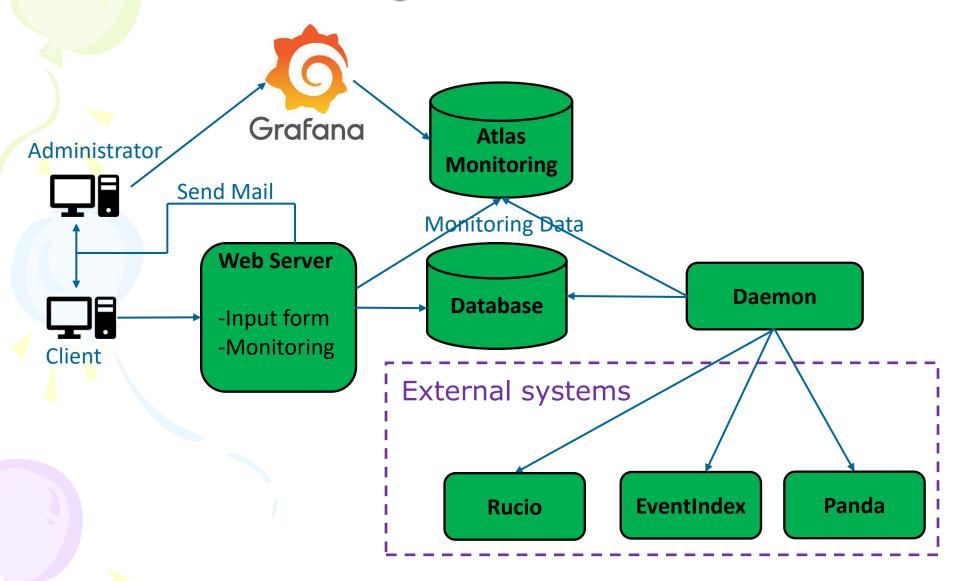
Verification

(should validate the output data possible duplicate/ skip events)

- Set Metadata in RUCIO
- Send mail (user notification)



Event Picking service: Architecture



16.10.2024

6

Event Picking service: Web Server

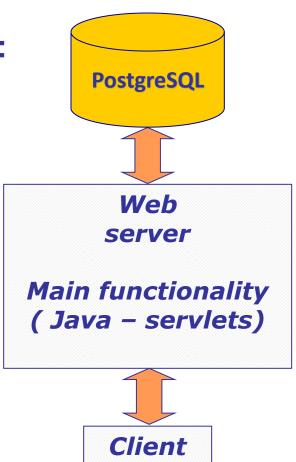
Basic technologies:

- PostgreSQL;
- Apache Tomcat;
- JAVA;
- JQuery;
- HTML / CSS / JAVA-script;
- WALT

Start page:

https://atlas-event-picking.cern.ch/eventpicking/

Web service is available outside of CERN (need CERN SSO authorization)

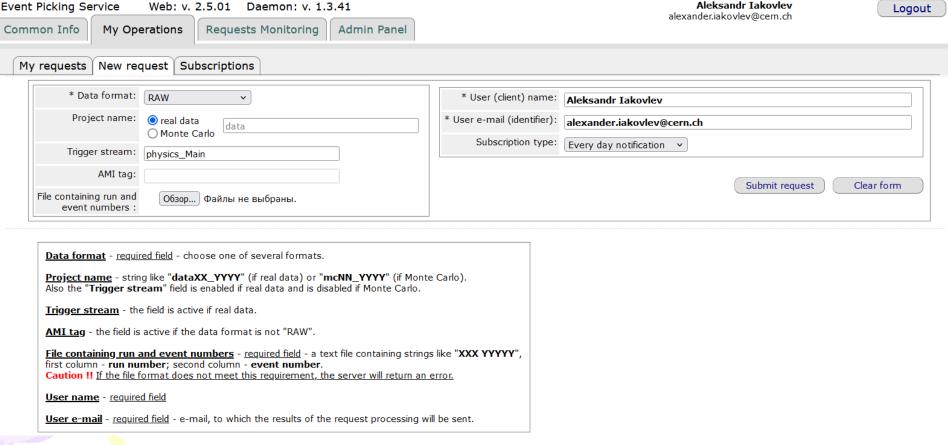


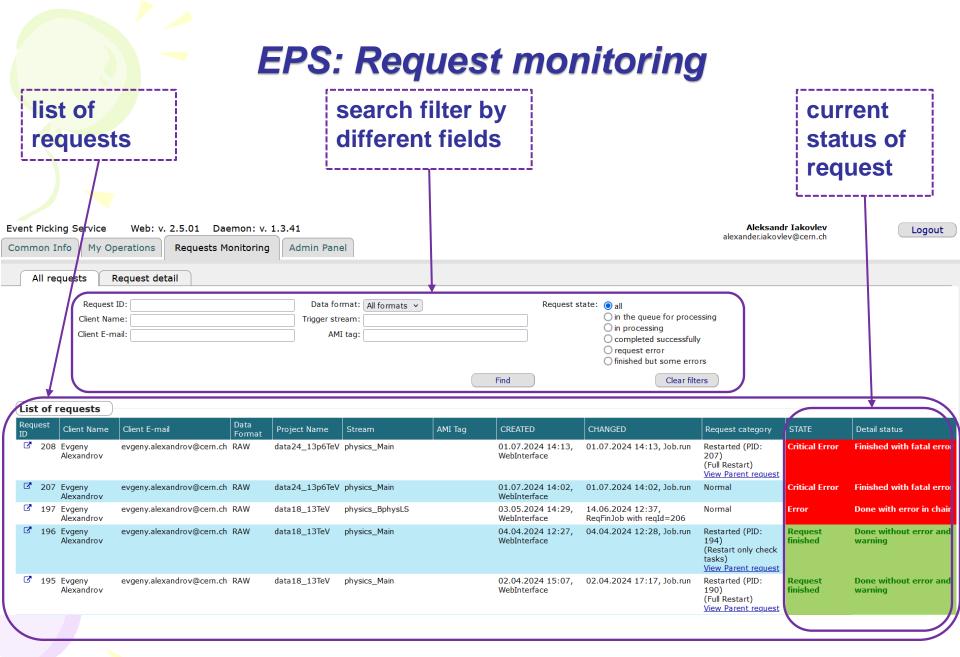
(Web-

browser)

EPS: Create new request

- all required fields must be filled
- text file must contain only strings like XXX YYYYY
 (first column run number, second column event number)



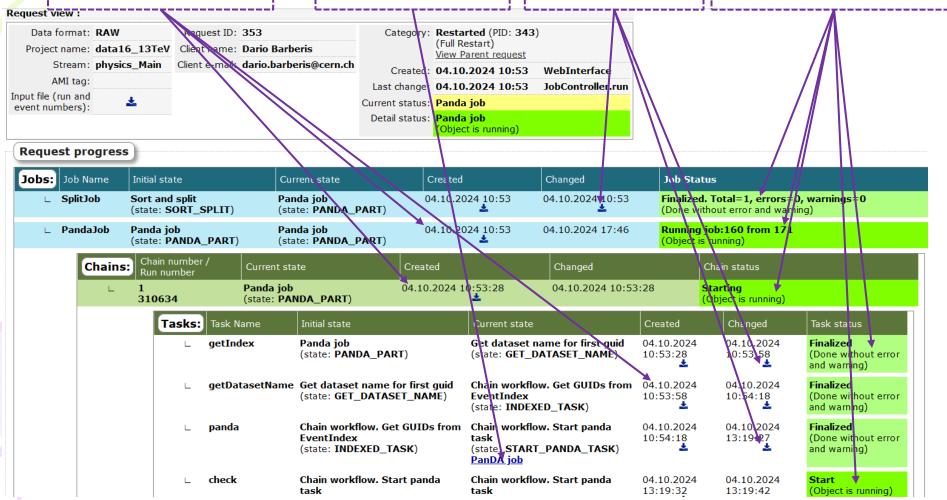


EPS: Request monitoring

Request detail information:

start and stop time for jobs, chains and tasks link to Panda page for Panda task output result in JSON format

current statuses of jobs, chains and tasks



EPS: Request monitoring Request results panel:

Data format:	RAW	Request ID:	95	Created:	23.04.2022 23:21	WebInterface
Stream:	physics_Main	Client name:	aleksand	Last change:	24.04.2022 15:50	Job.run
AMI tag:		Client e-mail:	aleksand@jinr.ru	Current status:	Request finished	*
				Detail status:	Request finished (Done with restart task)	
Request	result :					

Download Summary Report file (+)

Event Processing Progress Log (+)

Result files (+)

typical request result (for completed successfully)

Request result

DatasetName

group.projevind.data15_13TeV.00276954.physics_Main.evtpick.DRAW_EVTPICK.r60t22992

evind.data15_13TeV.00276689.physics_Main.evtpick.DRAW_EVTPICK.r60t23003

evind.data15_13TeV.00276952.physics_Main.evtpick.DRAW_EVTPICK.r60t23002

evind.data15_13TeV.00279345.physics_Main.evtpick.DRAW_EVTPICK.r60t23013

evind.data15_13TeV.00276790.physics_Main.evtpick.DRAW_EVTPICK.r60t23001

evind.data15_13TeV.00278880.physics_Main.evtpick.DRAW_EVTPICK.r60t23016

group.proj-

evind.data15_13TeV.00279764.physics_Main.evtpick.DRAW_EVTPICK.r60t22996

evind.data15_13TeV.00279685.physics_Main.evtpick.DRAW_EVTPICK.r60t23017

evind.data15_13TeV.00279259.physics_Main.evtpick.DRAW_EVTPICK.r60t22994

evind.data15_13TeV.00279284.physics_Main.evtpick.DRAW_EVTPICK.r60t23010

file path

/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/67/c2/group.proj-evind.27840942._000001.event.dat

/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/5a/69/group.proj-evind.27840962._000001.event.dat

/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/a1/f7/group.proj-evind.27840961._000001.event.dat

/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/

/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/36/d1/group.proj-evind.27840958._000002.event.dat

/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/1c/78/group.proj-evind.27840979._000002.event.dat /eos/group/proj-evind/aa/a4/group.proj-evind.27840979._000001.event.dat /eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/ed/54/group.proj-evind/rucio/group/proj-evind/ed/54/group.proj-evind/ed/54/grou

/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/19/99/group.proj-evind.27840952._000002.event.dat /ec/group/proj-evind/6c/a3/group.proj-evind.27840952. 000001.event.dat

/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/

/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/43/05/group.proj-evind.27840946._000001.event.dat /ec/group/proj-evind/ba/aa/group.proj-evind.27840946._000002.event.dat

/eos/atlas/atlasgroupdisk/proj-evind/rucio/group/proj-evind/a7/80/group.proj-evind.27840969._000002.event.dat /eo/group/proj-evind/d1/0e/group.proj-evind.27840969._000001.event.dat

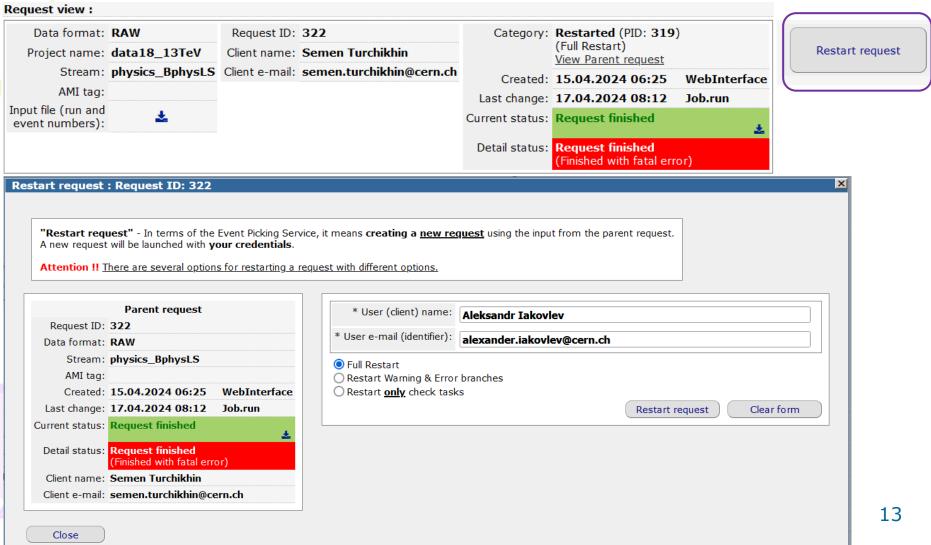
EPS: Request monitoring

Request detail information: (Finished but some errors exists)

Finalized. Total=165, errors=13 (Done with error in chain)

L	2 303892	Chain fi (state:	inished DONE_CHAIN)	24.08.2024 10:00:48	31.08.20	024 07:38:54		Finalized (Done with check errors	5)
	Tasks:	Task Name	Initial state	Current	state	Cre	eated	Changed	Task status
	L	getIndex	Panda job (state: PANDA_PART)		aset name for first guid GET_DATASET_NAME)		08.2024 :00:49	24.08.2024 10:01:35	Finalized (Done without error and warning)
	L	getDatasetName	Get dataset name for fi (state: GET_DATASET_I	NAME) EventIn	orkflow. Get GUIDs fro dex NDEXED_TASK)		08.2024 :01:35	24.08.2024 10:01:55	Finalized (Done without error and warning)
	L	panda	Chain workflow. Get GU EventIndex (state: INDEXED_TASK)	(state: S	orkflow. Start panda ta START_PANDA_TASK) j <u>ob</u>		08.2024 :01:55	30.08.2024 03:39:57	Finalized (Done without error and warning)
	L	check	Chain workflow. Start p (state: START_PANDA_		orkflow. Start panda ta START_PANDA_TASK)		08.2024 39:57	31.08.2024 07:33:54	Wait restart (Error state of task)
	_	check	Chain workflow. Start p (state: START_PANDA_		nished DONE_CHAIN)		08.2024 :38:54	31.08.2024 07:38:54	ru.jinr.atlas.eventindex.RestartExcepti (Error state of task)
L	3 303892	Chain fi (state:	inished DONE_CHAIN)	24.08.2024 10:00:48	30.08.20)24 22:51:56		Finalized (Done with warning: sk	ip or duplicate events)
	Tasks:	Task Name	Initial state	Currer	nt state		Created	Changed	Task status
	_	getIndex	Panda job (state: PANDA_PART)		ataset name for first gu : GET_DATASET_NAME		24.08.2024 10:00:49	24.08.2024 10:19:53	Finalized (Done without error and warning)
	L	getDatasetName	Get dataset name for fi (state: GET_DATASET_I	NAME) Event	workflow. Get GUIDs fi Index : INDEXED_TASK)	rom	24.08.2024 10:19:54	24.08.2024 10:40:55	Finalized (Done without error and warning)
	L	panda	Chain workflow. Get GU EventIndex (state: INDEXED_TASK)	(state:	workflow. Start panda : START_PANDA_TASK <u>A job</u>		24.08.2024 10:40:56	30.08.2024 04:12:25	Finalized (Done without error and warning)
	L	check	Chain workflow. Start p (state: START_PANDA_		etadata to rucio : Metadata)		30.08.2024 04:12:25	30.08.2024 22:35:20	Finalized (Done with warning: skip or duplicate events)
	L	metadata	Set metadata to rucio (state: Metadata)		finished : DONE_CHAIN)		30.08.2024 22:35:20	30.08.2024 22:51:56	Finalized (Done without error and warning)
	4	Chain fi	inished	24.08.2024 10:00:49	27.08.20	024 08:52:38		Finalized	
L	303638		DONE_CHAIN)	±		*		(Done with error in chai	n)

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



EPS: Time result

4	Request Number of events		Version	Time	
	\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\	FOI:	1.0.0	2 weeks	
	γγ -> WW	50k	manual	3 months	
	γγ -> WW	136k	Beta version	3 months	
	$B_c^* -> B_c$	16K	1.2.37	84h	
	Z -> TauTau	11K	1.2.37	40h	

EPS: Available workflows

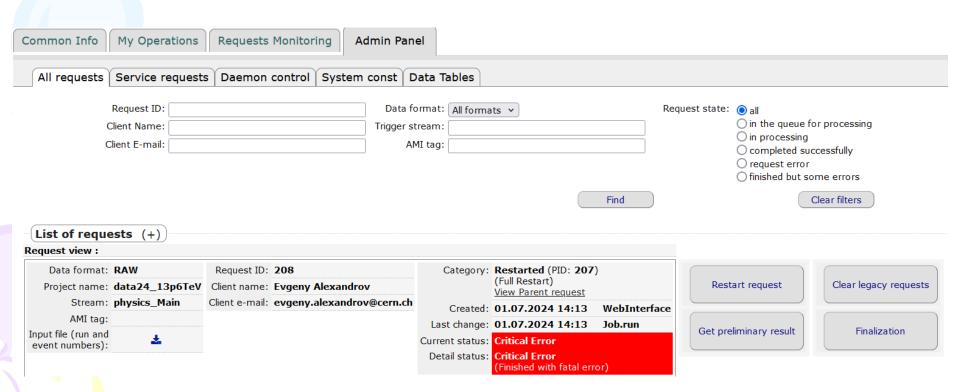
- Main
 - Common workflow for get selected events.
- Full restart
 - Restart request with the same parameters.
- Restart errors and warnings job
 - Restart only error and warning job.
- Restart check job
 - Restart job with check error. Start job since check job.
- Finalize request (special service request type)
 - Use in case of critical error of request. Set error for all not finished job.

EPS: Admin Panel

Special tools for Event Picking Service administrators.

Service requests

- Clear information from legacy requests;
- Getting preliminary results;
- Finalization of "asleep" requests;



EPS: Conclusion

- The Event Picking Service is running on the production server.
- Error handling and automatic fixes have been improved, which is why the speed of the new version has been increased.
- The number of users of the service is growing.
- The Event Picking Service is flexible and can be used in other experiments once the required workflow is implemented.

MLIT License management system

E. Alexandrov, I. Alexandrov, D. Belyakov, N. Davydova, P. Zrelov, L. Kalmykova, M. Lyubimova, T. Sapozhnikova, T. Syresina, A. Yakovlev

Main goals:

- managing the distribution of licenses
- audit and analysis of the effectiveness of the use of licenses.
- making decisions on the purchase of licensed software

MLIT License management system

Start page:

http://dev-web.jinr.ru:8080/license/service

The JINR single sign-on (SSO) is used to authenticate the user and his role when entering the personal account of the LMS.

Detailed report (26.04.2024):

https://indico.jinr.ru/event/4635/

Thanks for attention!

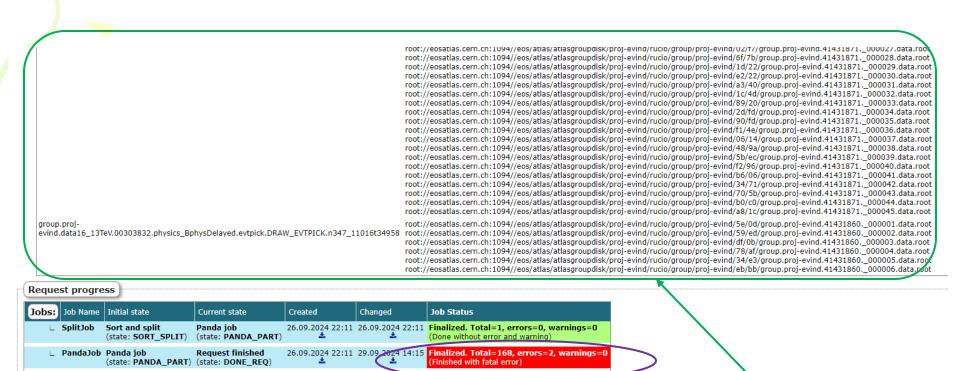
Backup slides

Critical error for most of chains (Full restart)

Event Picking Service	Web: v. 2.2.2	24 Daemon: v. 1	.3.39					
Common Info My C	perations Requ	uests Monitoring	Admin Panel					
All requests Serv	vice requests Da	emon control Sys	stem const Da	ta Tables				
Reques Client N Client E-	ame:		Data format: Trigger stream: AMI tag:			O in proce	ted successfully	
					Find		but some errors Clear filters	
List of requests	(+)							
Request view :								
Data format: RAW		Request ID: 337		Category:				
Project name: data1		Client name: Dario B			06.08.2024 16:34	WebInterface	Restart request	Finalization
	ics_BphysDelayed	Client e-mail: dario.b	arberis@cern.ch	-	24.08.2024 10:00	ReqFinJob with reqId=344		
AMI tag: Input file (run and event numbers):	<u>.</u>			Current status: Detail status:		nain)		
Request progress	3				Conc war ciror in ci			
Jobs: Job Name II	nitial state	Current state	Created	Changed	Job Status		l	
	ort and split tate: SORT_SPLIT)	Panda job (state: PANDA_PAR		:34 06.08.2024		tal=1, errors=0, warnings=0 error and warning)		
∟ PandaJob P a (s	anda job tate: PANDA_PART)	Error) (state: ERROR)	06.08.2024 16	:34 24.08.2024	10:00 Finalized. To (Done with ex			
(Hidden requests	(+)							

168 Errors! Require restart all request!

Critical error for a few of chains (Restart warning & error branches)



Only 2 errors. 166 done.

Correct results

Error for a check tasks (Restart only check tasks)

343 warning (error but result file exists)

