

The pilot sends detailed information about file transfers to Rucio. Here is a list of the different fields contained in the trace report.

Field name	Explanation
appid	PanDA job id
catStart	time stamp when rucio has been queried for replica information
clientState	state at the time the trace is sent; e.g. INIT_REPORT, STAGEIN_FAILED, NO_REPLICA, STAGEIN_NOTALLOWED, STAGEIN_ATTEMPT_FAILED, STAGEOUT_ATTEMPT_FAILED, DONE
dataset	dataset name (prodDBlock from job definition - or destinationDblock)
duid	(currently not set by the pilot)
eventType	type of trace; currently the following event types are sent by the pilot (ordered by frequency; in the case of user jobs, an _a is added to the type): <ol style="list-style-type: none"> 1. get_sm_a: stagein inputs for analysis jobs 2. get_sm: stagein inputs for non-analysis jobs 3. put_sm: stageout outputs for non-analysis jobs 4. get_es: stagein inputs for eventservice jobs 5. download: EXPLANATION 6. put_sm_logs: stageout logs for non-analysis jobs 7. put_sm_a: stageout outputs for analysis jobs 8. put_sm_logs_a: stageout logs for analysis jobs 9. upload: EXPLANATION 10. put_es: stageout outputs for eventservice jobs 11. put_sm_logs_os: stageout logs to objectstore (special transfer defined in agis per panda queue)
eventVersion	pilot version
filename	the local file name (LFN)
filesize	file size
guid	GUID with removed '-' signs, <code>guid.replace('-', '')</code>
hostname	host name, <code>socket.gethostbyaddr(socket.gethostname())[0]</code>
ip	IP number of local host: <code>socket.gethostbyname(socket.gethostname())</code>
localSite	PanDA site name (if set by pilot's Mover.py) or DDM endpoint (if set by pilot's movers/mover.py) - should only be DDM endpoints?
protocol	name of copy tool used by the pilot; e.g. xrdep
relativeStart	time stamp when stage-in or stage-out begins
remoteSite	DDM endpoint
scope	replica scope (note: missing in pilot's TraceReport class init function)
stateReason	error message or explanation; e.g. BAD_COPYTOOL, OK, 'skip stagein file'
suspicious	(currently not being set by the pilot - a corrupt file is currently reported to rucio via <code>declare_suspicious_file_replicas()</code> in the <code>rucio.client</code>)
taskid	PanDA task ID
timeEnd	time stamp when transfer or replica lookup, etc has been failed by the pilot
timeStart	start time of trace report; <code>time.time()</code>

PilotRucioTraces < PanDA < TWiki

transferStart	currently the same time stamp as relativeStart, see above
url	TURL in case of direct access for given replica. Note: in the case of direct access, the copy tool is not involved by rucio is queried for the replica info. The corresponding trace will contain the TURL, as well as clientState='FOUND_ROOT' and stateReason='direct_access'. In event service mode and when Prefetcher is used, the TURL, clientState='FOUND_ROOT' and stateReason='prefetch' are also added to the trace report
usr	hash of the user DN (usrdn field): <code>hashlib.md5(job.prodUserID).hexdigest()</code>
usrdn	user distinguished name (DN)
uuid	hash of the pilot id (if it exists): <code>hashlib.md5('ppilot_%s' % job.jobDefinitionID).hexdigest()</code> otherwise <code>commands.getoutput('uuidgen -t 2> /dev/null').replace('-', '')</code>
validateStart	time stamp when pilot performs checksum verification during stage-in and stage-out
version	(currently not used by the pilot)

-- PaulNilsson - 2018-02-16

This topic: PanDA > PilotRucioTraces

Topic revision: r4 - 2018-02-28 - IlijaVukotic



Copyright &© 2008-2021 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

or Ideas, requests, problems regarding TWiki? use Discourse or Send feedback