Service Incident Report for CASTOR DB Upgrade + Hurricane
on 29 July 2012

Incident Start: 2012-07-29, 21:02 [UTC]
Incident End: 2012-08-07, 08:30 [UTC]
Author: Jhen-Wei Huang

Description

On July 29 an Oracle RAC for CASTOR Stager suffered local SAS error due to hard drive overheating issue. Some transfer failures were observed after service recovery, the root cause was some DB indexes were broken by incident.

Unfortunately we were not able to do works for DB fixing imminently because a typhoon was approaching Taiwan. We discussed with experiments and decided to raise a scheduled downtime for CASTOR maintenance combined with CASTOR upgrade on August 6 after typhoon. After one week service degraded period, CASTOR was full back on August 7.

Impact

CASTOR efficiency degraded for one week from July 30 to August 7 due to DB index broken. ATLAS and CMS transfers failed from time to time if bad index files were referred by data access requests. ATLAS excluded TAIWAN-LCG2_DATATAPE from T0 export and CMS stopped transfer to Taiwan during that week before we fix index problem.

Analysis

CASTOR efficiency degraded after DB hardware failure.

Timeline

- Sun Jul 29 21:02 UTC 2012: Warning message sent by alarm system, CASTOR Stager DB crashed due to hardware failure.
- Sun Jul 29 23:25 UTC 2012: DB was recovered. Service was set online.
- Mon Jul 30 00:07 UTC 2012: Received ticket from experiment, transfer error appeared from time to time because of broken DB index issue. Site was looking for solution.
- Wed Aug 01 13:00 UTC 2012: After tracing DB problem recovery procedures had been made. Also we decided to upgrade CASTOR version to 2.1.11-9 together with this intervention. But there was a strong typhoon approaching Taiwan, for safety reason we postpone downtime to next Monday (Aug 06).
- Mon Aug 06 01:00 UTC: Downtime started for DB recovery and CASTOR upgrade.

- Mon Aug 06 16:00 UTC: Downtime extended because DB recovery took longer time than we expected.
- Tue Aug 07 08:30 UTC: DB recovery finished, brought CASTOR online. Service was fully operational after DB recovery. Experiments resumed transfer at the same day.

Future plan

We tuned some to CASTOR DB and also we finished upgrade to CASTOR system. Now it's providing more stable service to experiments.

Degraded service was delay to resume due to unexpected typhoon came to Taiwan. We will try to shorten affecting period as we can after we confirm operation people and equipment are safe.