Report sent on July 18th to: wlcg-scod@cern.ch

Type of Incident: dCache outage and performance issues
Location: IN2P3-CC
Duration: 30.3 hours
Date: July 1st 2012, 3:40AM to July 2nd 2012, 10:00AM CEST
Author: Rolf Rumler

Description

The servers for LCG dCache became inoperative. To re-establish the service a batch downtime was needed in order to avoid more unsuccessful job completions.

Timeline

July 1st (Sunday)

- 03:40 NAGIOS signalled the dCache servers for LCG being down.
- 08:00 The engineer on duty didn't succeed to restart the service.
- 08:25 dCache experts alerted, downtime declared.
- 09:00 Analysis showed one of the core servers being out of memory.
- 09:30 Core server rebooted, but dCache didn't restart correctly.
- 12:20 Downtime ended but erratic performance of dCache mixed with SRM command failures.
- 19:00 All jobs requesting dCache hold in queue.
- 22:00 Restart of core servers for Chimera and SRM, followed by a successful restart of dCache, new downtime declared, decision to keep jobs in queue.

July 2nd (Monday)

- 09:15 Freeing hold jobs after verification of service running correctly.
- 10:00 End of downtime.

Analysis

The origin of the incident was a bug in the operating system, SL6, due to the leap second introduced in the night from Saturday to Sunday. Initially it was not recognized that the problem came from the operating system and not from dCache itself.

Impact

The dCache servers for LCG run SL6 and so were subject to the bug mentioned, making the service unusable. As most of the incoming grid jobs require it, the batch was stopped for grid usage. Running jobs which tried to use dCache failed.

Corrective actions

Servers running under SL6 were rebooted.

SL6 has been updated with the correction for the bug encountered.