

Report sent on December 1st to: wlcg-scod@cern.ch

Type of Incident: partial network outage

Location: IN2P3-CC

Duration: 1.6 hours

Date: November 26th, 8:15AM to 9:50AM CET

Author: Rolf Rumler

Description

A switch connecting a whole rack to the rest of the site broke. Various services located on the machines of this rack became unusable, in particular all VO-boxes.

Timeline

November 26th (Wednesday)

- 08:15 NAGIOS signals the unavailability of a bunch of machines, all connected via the same switch. Network experts start analysis of the problem.
- 08:34 Batch closed i. e. no new jobs starting any longer, the running ones are let alone.
Network analysis continues and is complicated by the fact that the status indicators of the switch itself shows that all ports are ok. However, no traffic goes through. Decision is made to replace the switch.
- 09:50 Replacement done, connections re-established, all services back.
- 10:15 Batch re-opened.

Analysis

The alerts given by NAGIOS were sufficiently clear to start immediate action. The search for the failure origin was complicated by the fact that the switch itself gave wrong information about its state. After replacement investigation is ongoing for the reasons of this misbehaviour.

Impact

Various internal services and all VO-boxes were cut off the rest of the site's LAN. All jobs running at the time of the incident were unaffected. However, ALICE's monitoring tools did not see those jobs any longer and declared them "lost". To re-establish the situation the local ALICE support had to kill all of them.

Corrective actions

The defective switch was replaced.

Further corrective actions might be necessary depending on the outcome of the continuing investigation about the malfunction of the switch's status indicators.