www.egi.eu
@EGI_eInfra

Major incidents report

Alessandro Paolini

EGI Foundation

Major incident definition

- An incident is classified as Major when:
  - high number of users impacted **AND/OR** a large part of the infrastructure is impacted
  - breaking seriously agreed service levels or affecting EGI reputation or causing serious productivity loss for users

- For the Core Services:
  - Any incident producing a service loss is (must be) dealt within the (required) AvCo plan where the several risks have been assessed and the correspondent countermeasures planned/implemented.
  - In case of major loss of service, we create an entry in the Major Incidents DB to keep history of incidents
    - Useful information for the future

Major incidents list

- Jan 9th: Confluence is down for Java memory problems
- Feb 10th: MediaWiki on server wiki.egi.eu down after a change in PHP configuration
- Feb 11th: RT server not responding
- Feb 16th: DocumentDB went down - Internal Server Error
- Feb 20th: Jira not responding
- Mar 6th: GOCDB went down
- Apr 20th: Marketplace went down

Confluence is down for Java memory problems

Start time	Jan 9th 2020, 05:14 am
End time	Jan 9th 2020, 11:26 am

- GGUS ticket: https://ggus.eu/index.php?mode=ticket_info&ticket_id=144758

- Error message in the logs:

  [Full GC (Allocation Failure) 1266M->1120M(1280M), 3.5772970 secs] [GC concurrent-mark-abort]

  …

  java.util.concurrent.ExecutionException: java.lang.OutOfMemoryError: Java heap space

  java.lang.OutOfMemoryError: Java heap space

  …

- Solution: **increased maximum memory of 512 MB, from 1280 MB to 1792 MB, and restarted the server**

wiki.egi.eu down after a change in PHP configuration

Start time	Feb 10th 2020, 9:30 am
End time	Feb 10th 2020, 11:03 am

- The server went down, and we got a failure notification

- IT Support sent an email to us saying that the service was unexpectedly down after a change of the PHP configuration they just did

- The initial state of Apache modules have been restored
  - The same change on CSIRT Wiki was fine
  - On the EGI wiki there was a different apache version, with different modules, so the change would have required different settings

- Improvement suggestions:
  - The change was done without notifying us in advance, and without scheduling a downtime for the intervention. Even if this problem could not have been foreseen, we asked IT Support to always notify us about any changes affecting the Collaboration Tools, by using the usual CHM channels.

RT server not responding

Start time	Feb 11th 2020, 11:32 am
End time	Feb 11th 2020, 11:43 am

- We got the failure alert notification

- Opened a ticket: https://ggus.eu/index.php?mode=ticket_info&ticket_id=145493

- Problem:
  - The old host certificate was deleted, but the Apache conf was still pointing to it.

- Solution:
  - The settings were properly modified to point to the new certificate, and the server started regularly.

DocumentDB went down

Start time	Feb 16th 2020, 4:48 pm
End time	Feb 17th 2020, 10:40 am

- We got the failure alert notification (it was Sunday)
- Opened ticket on Monday morning:
  https://ggus.eu/index.php?mode=ticket_info&ticket_id=145573

- Problem: the server ran out of memory
- Solution:
  - the server was rebooted to restore its conditions
  - it wasn't possible determining what caused the high memory consumption
- Improvement:
  - installed a tool to monitor the CPU and the memory usage

- ticket closed after one week: no anomalies detected after the reboot

Jira not responding

Start time	Feb 20th 2020, 5:03 am
End time	Feb 20th 2020, 9:52 am

- We got the failure alert notification
- We got notified by email from IT Support about an unplanned outage of the servers jira.eosc-hub.eu and jira.egi.eu.
  - Misconfiguration of Apache modules
- What happened:
  - it was scheduled a system update in the early morning (outside working hours) which broke the apache configuration after the (automatic) restart of the machine.

- Recommendation:
  - It would have been preferable doing the upgrade during working hours with people following it, and informing in advance EGI Operations of the planned upgrade using the CHM process.

GOCDB went down

Start time	Mar 6th 2020, 9:33 am
End time	Mar 6th 2020, 10:27 am

- incident noticed first by EGI Operations https://ggus.eu/index.php?mode=ticket_info&ticket_id=145958

- first hypothesis: the web-server hit its maximum number of concurrent sessions.
    - Maybe similar to a high load issue occurred in December
    - the mix of API queries being served by GOCDB changes in the run up to these outages, whether this is malicious or an unintended side effect of legitimate use is currently unknown. This change then causes a high number of concurrent sessions and a high load on the underlying server - causing the web server to become unresponsive.
    - continued to investigate the working theory by comparing traffic in that morning with the day before one.
    - updated the frequency of the polling of internal monitoring

- Mar 9th: this outage was caused by one database nodes losing connectivity to its storage, resulting in a decrease in performance of the database and a backlog concurrent web server sessions.

Marketplace went down

Start time	Apr 20th 2020, 8:53 pm
End time	Apr 21st 2020, 10:28 am

- We got the failure alert notification
- The morning after it was opened the ticket:
  - https://ggus.eu/index.php?mode=ticket_info&ticket_id=146586

> *Proxy Error*
> *The proxy server could not handle the request GET /.*
> *Reason: Error during SSL Handshake with remote server*

- It was fixed the proxy configuration in the apache service

EGI: Advanced Computing for Research

**Thank you
for your attention.**

Questions?