Incident report for WR#346257

Timeline – Moodle LDAP incident – 27th October 2020

Time	Event
19::00	AWS Aurora DB works commence to reduce the instance sizes of Aurora nodes
19:30	AWS Aurora Writer node is failed over to new writer instance
19:48	AWS Aurora scale down actions completed
19:50	Catalyst engineer notes that they are able to log into the site (manual account) and able to observe the change as a small blip in new relic. Task is marked as complete
~20:00	Alistair Spark notes that LDAP has been lost
20:15	Catalyst Engineer observes the note and returns to their console to commence analysis/fix works
20:29	Catalyst engineer notes that investigation is underway
20:29	Alistair Spark notes that the issue is resolved by updating the config – as outlined below.

Executive summary

During the catalyst works to reduce the size of the database instances in use, a database value that enabled LDAP module was altered to the off/disabled state and as a result all DLAP logins were failing with an error to end users.

The incident lasted exactly 60 minutes and the fix applied was to re-enable the LDAP module in Moodle's authentication plugins interface.

Incident detailed overview

A major incident related to LDAP Single-Sign-On was incurred on 27th of October between the hours of 7:29pm and 8:30pm. During this period of time the system was having infrastructure maintenance carried out – with actions to reduce the system resources on – specifically the scale of database service in use from AWS.

Catalyst completed the managed scale down of AWS Aurora RDS service, moving from 3 DB nodes at scale 4xlarge to the same 3 nodes running at 2xLarge. The reason for this action is to deliver cost savings to UCL and align resource sizes to consumption.

This action was deemed suitable to be delivered during a low traffic standard operation window (evening times) as it presented as a low risk change, with no outage to end users and a graceful pathway to do so – by failing over each DB node to a smaller instance, one-by-one.

Upon completion of the action, the Catalyst engineer logged into the Moodle system (using the manual account utilised by Catalyst) and performed basic testing to prove the site was still operational. No anomalies were caught, but clearly no test for LDAP was carried out.

At 20:15 the Catalyst engineer who performed the work was notified by Alistair Spark that LDAP had become unavailable, and as a result users were unable to access the site. All users attempting to access the site via the LDAP authentication method received a consistent error in response.

At 20:29 The catalyst engineer reached out Alistair to pick up the issue but was notified that the issue had been rectified by Alistair Spark by re-enabling the LDAP service inside the Moodle application.

All investigations show that the database value in Moodle which configures Moodle LDAP authentication module to 'on' was updated to 'off'. The remedial action applied by Alistair was simply to revert this Moodle config setting to the 'on' state once again.

Subsequent detailed analysis has shed no real light on the cause of this – Catalyst have never seen a similar occurrence in our time working with Moodle on AWS RDS services or Aurora services specifically. It is very unclear how a DB value can be updated as a result of the work completed and we suggest further analysis on the staging instances in UCL to try to recreate the issue.

The cause of such an update in the DB is not fully known, Catalyst teams have spent today hypothesising on the cause and have the suggestions below on how something may trigger this issue. But it is a never before seen incident which also does not present a clear and reasonable cause.

The strongest suggestion we have is as follows, but this is extremely unlikely, rare and unfortunate if proven to be the case.

During the DB switchover, it is possible we could see a very specific edge case where that might occur.

$CFG->auth- if not expressly set in config.php will be set during the early part of setuplib by backfilling from the config cache and/or the config table.

We could see for very specific cases where you could get the wrong value come out of the config table for this but the confluence of circumstances is extremely rare for it happening during the DB rollover on a read only client returning an empty row instead of the real value for a config that doesn't have a cacheable value.

That's one confluence where you could get $CFG->auth be an empty string.

Beyond this – Catalyst request more time is afforded to try and recreate the issue, whilst implementing the mitigations below.

Results and impact

The result of this outage was severe/critical. All users who utilise LDAP would have been unable to log into the site for ~60 minutes. Full incident response was slower than usual to commence as no critical ticket status was raised in the WR or no alarm sounded for the Catalyst engineers / on call team.

Incident response actions taken

As noted above, a full and complete incident response was not commenced due to no clear flag of critical incident being in play. Once the engineer was underway with resolution works, the issue was already resolved by UCL.

Ongoing risks

There is a clear and present risk this may occur again in any of the following events;

- A database failover action due to loss of availability zone
- A database scaling action

- A database remedial action in the event of an issue or incident.

Mitigation for future

A number of actions are proposed to mitigate this risk and ensure no future occurrences arise.

1. Complete the implementation of clear LDAP monitoring to ensure that LDAP availability and function are verified frequently by probes – allocated as immediate action required.

2. Additional tests applied to Catalyst release and AWS update steps, to ensure LDAP is checked after every change.

3. Additional analysis of the scenario which led to this incident, trialling on the staging site to attempt to recreate the issue so as to better understand the cause and avoidance measures – To be discussed with UCL.

4. Additional discussions and training to ensure that remedial works are clearly notified and also that escalation to critical incident (according to the SLA definitions) is actioned ASAP in the event of incidents.

5. Agreement that formal maintenance windows are used for ***all*** actions. Whilst this is a suggestion, Catalyst would welcome discussion with UCL as this mitigation does reduce the pace and flexibility that the AWS cloud affords to UCL Moodle service (s).