

Root Cause Analysis: System Outage

The following is a detailed accounting of the service outage that Rally users experienced on September 5th, 2019.

Root Cause Analysis Summary

Event Date	09/05/2019
Event Start	12:00 AM MDT
Time Detected	8:04 AM MDT
Time Resolved	8:59 AM MDT
Event End Time	11:14 PM MDT
Root Cause	<p>Rally has been moving from Amazon S3 buckets for storage, to Google Storage buckets for storage. As part of this, we were running a sync job that copied all files from S3 to Google Storage (GS), but did not realize that if a file was present in GS, but not in S3, it would delete the GS file. On 9/4, we disabled storing files to S3, and at midnight on 9/4-9/5, the sync job kicked off, removing the files from GS.</p> <p>When a file isn't in GS, Rally falls back to the database attachment content. Rally then tries to decompress the data from the database, but the content returned by the database isn't compressed content. Rally's error handling was poor around this situation, and it would go into an infinite loop of trying to decompress uncompressed data. While it did this, a database connection was in use, causing exhaustion of our connection pool, and eventually an outage.</p>
Customer Impact	55 minutes of downtime / 12 hour time window of missing files / 19 support cases

Future Preventative Measures

Actions that should be taken to prevent this Event in the future.

Actions	Description
ALM needs to handle attachments better	Remove fallbacks and respect its environment (Google Cloud/On-prem/etc). Should simply fail if it cannot get its "source of truth" attachment or if it cannot upload its attachment.
ALM not handling exceptions well	Resolving the infinite loop which was holding database connections open. (Done)
Write multi-request restore script	Decrease the restore time in case a larger incident of this type occurs by writing a multi-request restore script