

DSB Production Root Cause Analysis

For Production Outage on 25th March 2018

Prepared by: Technical.Support@anna-dsb.com
Date: 4th April 2018

Table of Contents

IMPACT ASSESSMENT & CATEGORIZATION	3
INTRODUCTION	3
EXECUTIVE SUMMARY - FINDINGS AND ROOT CAUSE	4
CORRECTIVE ACTIONS TAKEN & PLANNED	4
EVENT DESCRIPTION	5
CHRONOLOGY OF EVENTS / TIMELINE	6

Revision History

Version	Date	Reason
1.0	4 th April 2018	Published RCA



IMPACT ASSESSMENT & CATEGORIZATION

Critical/System Down (Severity One – S1)

Start: 13:07

Resolved: 13:28

Total: 21 minutes

Production application down or major malfunction resulting in a product inoperative condition. Users unable to reasonably perform their normal functions.

The specific functionality is mission critical to the business and the situation is considered an emergency.

Condition 1- When a critical system, network component or key application is under outage (or imminent outage) with critical impact to all clients.

Condition 2 - Total loss of service to entire user base which includes total unavailability of critical applications for entire end users and in all locations.

Significant Impact (Severity Two - S2)

Start: 12:02

Resolved: 14:00

Total: 1 hour 58 minutes

Critical loss of application functionality or performance resulting in high number of users unable to perform their normal functions. Major feature/product failure; inconvenient workaround or no workaround exists. The program is usable but limited.

Condition 1: A key component of the solution, an application across all users, a set of users or intermittent network degradation or instability leading to performance and degradation of service.

Condition 2: An incident which is not yet S1, but might lead to a potential S1 incident.

Condition 3: Partial users at a particular location are affected but not all the users in all locations

INTRODUCTION

The purpose of this Root Cause Analysis (RCA) is to determine the cause that contributed to the recent loss of service and “Something went wrong” error message and response code “HTTP 500” encountered by clients in the DSB production environment on 25th March 2018 between the hours of 12:02 UTC and 14:00 UTC. This RCA determines what happened during the event, how it happened, and why it happened. To accomplish this, an investigation took place internally between the DSB support, Development teams and senior analysts to ascertain the primary root cause or a list of root causes that contributed to this issue.



EXECUTIVE SUMMARY - FINDINGS AND ROOT CAUSE

Sunday 25th March 2018

This was due to performance issues on SOLR (<http://lucene.apache.org/solr/>) resulting in a lack of responsiveness as the DSB observed an increase in CPU load and memory consumption on the SOLR servers. After examination of log files and alerts received both pre- and post-event, the DSB technology team stopped all API end points and restarted all SOLR servers in sequence to stabilize the service as initial health checks showed 2 SOLR servers were not in active state.

Validation of the optimized ISIN Engine codebase continues in UAT (see date for Industry Wide Load Testing below). Production release of said code will follow within Q2 2018.

CORRECTIVE ACTIONS TAKEN & PLANNED

- All SOLR servers restarted in sequence to recover service
- Q2 2018 - Target production date for Authentication caching configuration change



EVENT DESCRIPTION

On 25th March 2018 at 12:02 pm UTC, the production environment experienced an issue with SOLR services, causing some established FIX and ReST API connections to drop without successfully reconnecting. This took place between the hours of 12:02 and 14:00. Clients also had difficulty to login via the web GUI and therefore the web portal was put into maintenance mode at 13:07.

“Something went wrong” error messages were experienced by clients when searching or creating ISIN’s due to the Cordra (<https://cordra.org/>) service timing out on their SOLR services connections. This was due to the SOLR service being unresponsive during this period because of extended Garbage Collection within the Java Virtual Machine.

“HTTP 500: error messages were encountered by clients utilizing the Production GUI during this time due to the SOLR service being unresponsive.

Between the hours of 12:02 and 14:00, users experienced reconnects and disconnections via FIX. ISIN creation and search services for API and GUI were unavailable between the S1 start and finish times specified on page 2.

As a result of the issues experienced, the DSB technology team initiated a restart of all SOLR servers in sequence after all end points were shut down in order to stabilize and restore the service.

After the SOLR restart, all SOLR service health checks reported in active state and therefore GUI functionality was restored as well as all FIX and ReST endpoints were opened. The service then returned to a normal state.

. The latest optimized codebase was deployed in UAT on March 27th with specific enhancements to address capacity bottlenecks. Deployment in production is expected in Q2 2018 after successful industry wide load testing in UAT scheduled this April 6th to 7th.

CHRONOLOGY OF EVENTS / TIMELINE

Sunday 25th March 2018

12:02 PM UTC – Sunday 25th March 2018

Alerts were triggered on SOLR and Cordra services due to high CPU and memory consumption. Technical support started investigation

12:11 PM UTC – Sunday 25th March 2018

Health checks on SOLR services confirm only 1 SOLR cluster member is active

12:19 PM UTC – Sunday 25th March 2018

As part of troubleshooting efforts, one of the three SOLR servers restarted but services did not stabilize

12:25 PM UTC - Sunday 25th March 2018

Health checks on SOLR services show 2 SOLR cluster members are active

12:34 PM UTC - Sunday 25th March 2018

Notification email was sent to the clients

12:42 PM UTC - Sunday 25th March 2018

All FIX endpoints were stopped.

12:49 PM UTC- Sunday 25th March 2018

SOLR services restarted – in sequence

12:53 PM UTC - Sunday 25th March 2018

Health checks on SOLR services show 1 SOLR cluster member is active

13:07 PM UTC - Sunday 25th March 2018

Placed the WEB GUI and all ReST endpoints in maintenance

13:12 PM UTC - Sunday 25th March 2018

Health checks on SOLR services show 2 SOLR cluster members are active

13:17 PM UTC - Sunday 25th March 2018

Updated notification was sent to the clients advising ongoing status



13:19 PM UTC - Sunday 25th March 2018

All SOLR cluster members back to reporting active status

13:36 PM UTC - Sunday 25th March 2018

All FIX endpoints were started

14:00 PM UTC - Sunday 25th March 2018

Maintenance lifted on all ReST endpoints and Web GUI

14:07 PM UTC - Sunday 25th March 2018

GUI, ReST and FIX validation checks are passed

14:13 PM UTC - Sunday 25th March 2018

Resolution Notification email sent to the clients