Stellar Streaming Solution POC – draft v.01

Contents

[1. GOAL 1](#_Toc68012622)

[DATA SOURCE IN SCOPE 1](#_Toc68012623)

[REPORTING [Replicate shintec edl reports] 1](#_Toc68012624)

[2. Architecture 2](#_Toc68012625)

[omni360 data 2](#_Toc68012626)

[Stellar STORE TEAM 3](#_Toc68012627)

[Omni360 platform TEAM 3](#_Toc68012628)

[3. Cybersecurity 3](#_Toc68012629)

[4. PRIVACY 4](#_Toc68012630)

[5. RACI 4](#_Toc68012631)

[6. Timeline 4](#_Toc68012632)

[7. COST 4](#_Toc68012633)

[8. Definition of Done 4](#_Toc68012634)

[9. APPENDIX 4](#_Toc68012635)

# GOAL

Design a production-ready solution to extract data from the India Magento database and build the reports listed in the scope to support the Store BI.

## DATA SOURCE IN SCOPE

Only the Magento data needed to recreate the India Store BI reports are in scope. We are going to list the necessary tables after analysing each one of the reports.

## REPORTING [Replicate shintec edl reports]

Only the listed reports with structure defined by March 2021 are in scope. Any change on structure will require to conduct an evaluation exercise to estimate the amount of effort and plan accordingly.

Table containing the list of reports in scope:

|  |  |
| --- | --- |
| 2 categories, Todo Arti |  |
|  |  |
|  |  |

# Architecture

The proposed architecture leverages the design of the work presented by Emmanuel Espina from AWS in the blog title "Streaming Changes in a Database with Amazon Kinesis" 1. The author suggested enabling the "Binary Log" 2. on the Mysql database and build a fetcher process to capture the logs and push them to a kinesis stream.

The stellar team will enable the binary log parameter in the read replica from the India Magento Database in the proposed solution. Then the Stellar team will build a process running on an EC2 to capture the changes on the database and push those changes to the C360 kinesis stream.

C360 will take the Kinesis stream's information and transform and load it to an AWS Aurora Postgres RDS. After loading the data to the RDS, the Stellar BI team will recreate the reports needed.

Figure 1 shows a simplified version of the solution proposed for this POC.

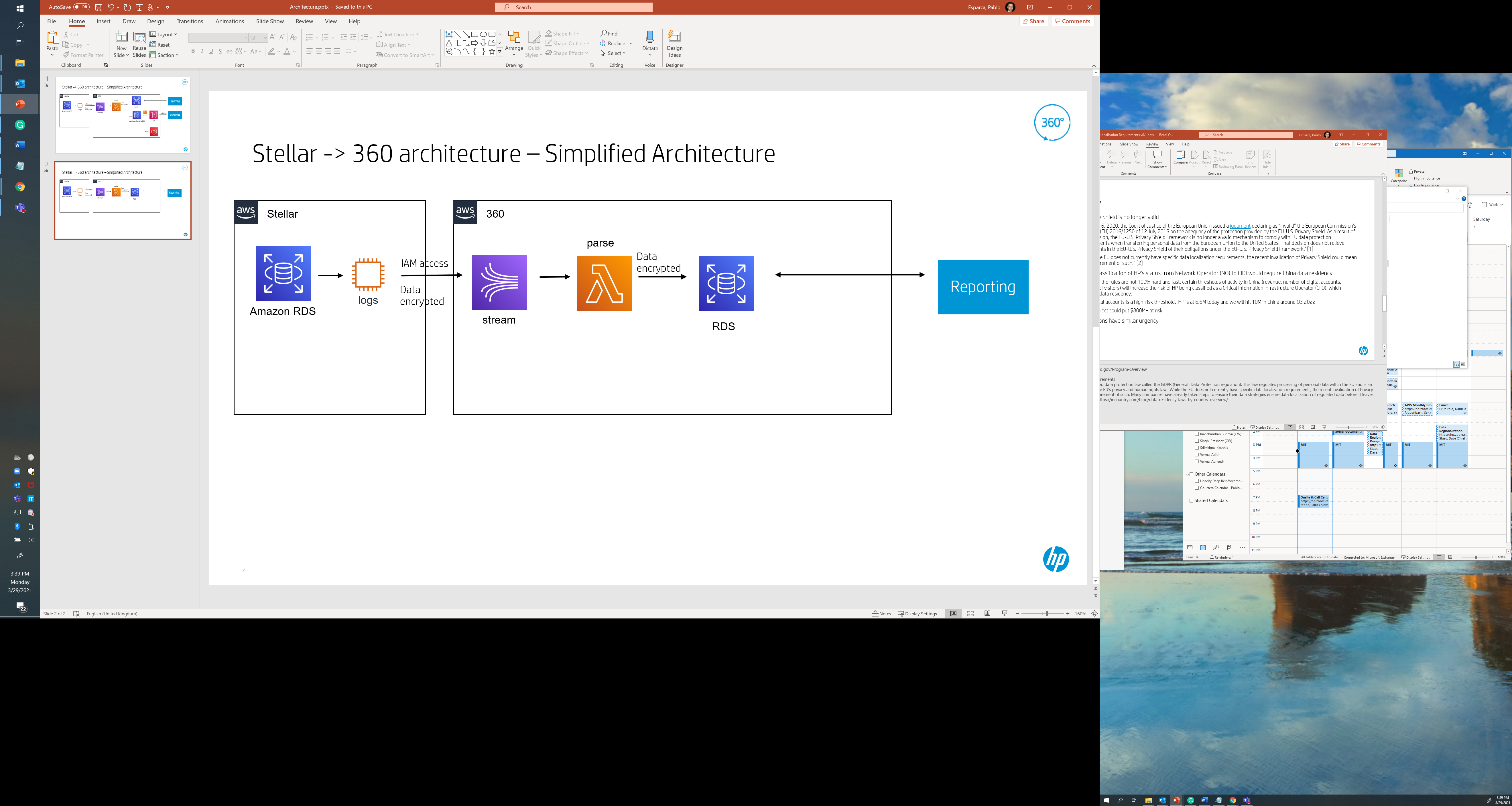


Figure 1

## omni360 data TEAM

The Omni360 data team will get the India database catalog from the Stellar Store Team and reverse engineering to identify the tables and columns needed to generate the Tableau reports.

The Omni360 data team will provide a list of tables required to the Stellar Store Team to filter only the logs needed. They also will send a list of columns that requires encryption to the Omni360 platform team.

After the log is loaded to the AWS Aurora Postgres, the Omni360 data team will create the necessary views and ETLS to support the tableau dashboards.

## Stellar STORE TEAM

The Stellar Store Team will use the India database read replica to test the impact performance on the server with the “Binary Log” process enabled.

The performance impact will be using CPU % utilization.

In the same AWS account that the databases are running, the stellar store team will provision an EC2 instance and develop a process to continue monitor for changes on the database. Per the original blog, a python library call ”mysql-replication==0.23” can be used to reduce the development effort.

In the same python process the logs will be filter by just the table list provided by the Omni360 data Team and the format of the logs will be json and it will include the database name as one of the attributes.

As a security, the data is going to be encrypted in transit using the python library “rfernet==0.1.3” using a common set of predefined encryption keys.

Finally, the logs will be push to a kinesis stream own by the Omni360 platform team assuming an AWS IAM role attached to the EC2 machine as authentication.

## Omni360 platform TEAM

The Omni360 platform team will enable a kinesis streams and grant access to the IAM role name provided by the Stellar Store Team.

A lambda function will be developed to received the encrypted data from the kinesis stream. The process will decrypt the data, identify any PII data according to Cybersecurity guidelines and tokenize the data using the Omni360 platform team tokenizer service.

Once the data is tokenize it will be load to the AWS Aurora Postgres data. The log contains the information of the schema, and operation required. (update, delete and insert)

# Cybersecurity

A cybersecurity review is required to validate cross account access. The Omni360 platform team will create an IAM role and grant access to the Stellar Store Team. Per Cybersecurity any cross-account policies require approval.

The enablement of the IAM cross-account access has been tested and is working, but it requires that cybersecurity whitelist that role name ARN.

A cybersecurity review request was submitted on March 25 with number SRS2021-455.

# PRIVACY

No privacy assessment is required since personal data is following HP Cybersecurity standards on encryption at rest, in transit and at a field level.

# RACI

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Responsible | Accountable | Consulted | Informed |
| Changes on database tables | Stellar Store Team |  |  | Omni360 Data Team |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# Timeline

3-6 Months plus 1-2 months

# COST

|  |  |
| --- | --- |
| **Stellar Store Team** | **Omni360 platform team** |
| Total Cost $8.8K | Covering 15 countries  Total $25,000 |
| **Omni360 resources** |  |
| ~$30K |  |

# Definition of Done

|  |  |  |
| --- | --- | --- |
| Done | Responsible | Description |
|  | Stellar Store Team | Share data catalog of the India database with the Omni360 data team |
|  | Stellar Store Team | Enable logs in the India Magento database |
|  | Stellar Store Team | Conduct an Impact analysis on the performance on the database after enabling the logs. Compare current CPU % AVG Utilization before and after |
|  | Stellar Store Team | Provision an EC2 to process the database logs |
|  | Stellar Team | Provide the role-name to the Omni360 Platform Team |
|  | Stellar Store Team | Share encryption credentials with the Omni360 Platform Team |
|  | Stellar Store Team | Build the process to send the logs to the Kinesis Stream |
|  | Stellar Store Team | Filter only the log information required base on the list of tables provided by the Omni360 Data Team |
|  | Omni360 Data Team | Send the list of tables needed to build the reports to the Stellar Store Team |
|  | Omni360 Data Team | Identify PII data on the tables required. |
|  | Omni360 Data Team | Send PII data reference to the Omni360 Platform Team. |
|  | Omni360 Data Team | Build the necessary views and ETLs to support the Tableau Dashboards |
|  | Omni360 Platform Team | Enable the a dev and pro Kinesis Stream |
|  | Omni360 Platform Team | Enable the Stellar Store Team IAM role access to the Kinesis Team |
|  | Omni360 Platform Team | Lead the Cybersecurity engament to enable IAM cross account access |
|  | Omni360 Platform Team | Transform the logs received from the Kinesis stream and load it into the AWS Aurora Postgres. |
|  |  |  |
|  |  |  |
|  |  |  |

# APPENDIX

1. Streaming Changes in a Database with Amazon Kinesis

<https://aws.amazon.com/blogs/database/streaming-changes-in-a-database-with-amazon-kinesis/>

1. Binary Log

<https://dev.mysql.com/doc/refman/5.7/en/binary-log.html>