AWS DWH Design Document for Kotak Prime

Logo, company name

Description automatically generated

Table of Contents

[1 Introduction / Overview 3](#_Toc84942335)

[1.1 Project Overview 3](#_Toc84942336)

[2 Environment Details for Extracts 4](#_Toc84942337)

[2.1 AWS Architecture Overview 4](#_Toc84942338)

[2.2 AWS-Storage(S3) 5](#_Toc84942339)

[2.3 Source Information 5](#_Toc84942340)

[2.4 Destination Information 5](#_Toc84942341)

[3 Source Details 6](#_Toc84942342)

[3.1 Oracle db 6](#_Toc84942343)

[4 High Level Approach 6](#_Toc84942344)

[5 Detailed Technical Process 7](#_Toc84942345)

[6 Data Model Diagram 10](#_Toc84942346)

[7 Distribution Style and Sort Key 11](#_Toc84942347)

[8 Metadata information For INGESTION, PREP & SERV lAYERS 13](#_Toc84942348)

[9 Mapping Document 13](#_Toc84942349)

[10 Appendix 13](#_Toc84942350)

Introduction / Overview

## Project Overview

This project is intended to replicate the Oracle on premise databases (Asset core and OGL) on AWS Redshift data warehouse and to create the necessary tables and views/MVs to support the reporting requirements on Amazon QuickSight for Kotak Prime.

In Scope:

1. Report Mapping and Data Modelling
2. DMS Full/CDC data capture
3. S3 to Redshift data load and automation
4. Report creation on Amazon QuickSight

Out of Scope:

# Environment Details for Extracts

## AWS Architecture Overview

Diagram

Description automatically generated

Current Scope

## AWS-Storage(S3)

2.2.1 Buckets

## Source Information

* KMP CORE PROD
* OGL PROD

## Destination Information

DMS End Points

|  |  |
| --- | --- |
| Source Endpoint | Target Endpoint |
|  |  |

Source Details

## Oracle Source DB

Details for Asset core DB:

|  |  |
| --- | --- |
| Database | Asset core DB |
| Host | 10.240.55.28 |
| Port | 1530 |
| SID | KCF |
| DB Version | 19C |

Details for OGL DB:

|  |  |
| --- | --- |
| Database |  |
| Host |  |
| Port |  |
| SID |  |
| DB Version |  |

# High Level Approach

1. DMS to pull Full/CDC data to S3 bucket
2. Redshift COPY command to load pipe (|) delimited files from s3 to ingestion layer of Redshift
3. Redshift script to load data from ingestion to prep layer
4. Redshift Materialized views/Views to be used in Analytics layer
5. Amazon QuickSight to connect to Analytics layer

# Detailed Technical Process

Data transfer from Oracle to S3:  
  
1. AWS DMS will connect to the Oracle instance and pull the records for the tables defined in the task definition.  
  
2. Records pulled from Oracle will be placed into the following S3 bucket folders

* fullload - full load dump of selected tables in task
* cdc - data changes of selected tables in task

Format:

<S3 bucket name>/fullload/schema\_name/table\_name/year/month/date  
<S3 bucket name>/cdc/schema\_name/table\_name/year/month/date

3. 2 sets of DMS tasks will be created for the tables to be replicated from Oracle to Redshift

* FULL Load
* Ongoing Replication

"FULL Load" - the task will be run when we want to SYNC all data from Oracle to Redshift  
"Ongoing Replication" - the task will be running all the time to track and replicate CDC changes  
  
4. Only INSERTs and UPDATEs will be replicated to Redshift. DELETEs will not be replicated.  
  
5. Target file format in S3 will be CSV with pipe "|" as the delimiter  
  
Sample:  
  
FULL LOAD:  
  
1|abc|BLR  
2|efg|CHN  
3|xyz|BLR  
4|qrl|MUM  
  
CDC:  
  
I|2021-10-05 06:00:37.671205|1|abc|BLR  
I|2021-10-05 06:00:37.671205|2|efg|CHN  
I|2021-10-05 06:00:37.671205|10|MLN|BLR  
U|2021-10-05 06:01:06.248178|4|qrl|DEL  
U|2021-10-05 06:01:06.248178|4|qrl|DEL  
  
  
FULL Load data transfer from S3 to Redshift:  
  
1. COPY command will be used to transfer the data from S3 to Redshift ingest layer table directly.   
  
2. Before full load, ingest layer table will be truncated to avoid duplicate values.

# Data Model Diagram

# Distribution Style and Sort Key

Distribution style:

All - whenever table size is small (< 100,000 rows) and if it involves in JOIN, then we will be making it ALL distribution style

Key - Table is large and getting involved in JOIN, we will use JOIN KEY as DIST KEY to co-locate the data and avoid data movement across compute node during runtime.

Note: While selecting JOIN KEY make sure it will do EVEN/Near to even distribution of data, else we may see RUN TIME skew impacting performance.

Even - If table will not involve in any JOINs or does not have proper DIST KEY which will make EVEN/Near to even distribution of data and table size is large, then we opt for EVEN. This will make sure even distribution of data

SORT KEY:

Sort key help to speed up data retrieval by eliminating the blocks not containing data - this is called block pruning.

2 types of SORT KEYs - compound and interleaved. In this project we will use compound SORT KEY only as interleaved comes with extra baggage of costly maintenance.

We can define multiple columns in SORT KEY but defining more than 3 columns will not add any performance improvements.

Columns to choose for SORT KEY are based on FILTER predicates, JOIN clause, GROUP BY clause

Note: To maintain sort of table frequent VACUUM is required (there is automated process as well which takes care from backend)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Schema Name | Table Name | PK | FK | Joining column | Master data | # of rows | Distribution Style | Sort Key (ingestion) | Joining table | Joining column |
| RCT | ALLOC\_SET |  |  |  |  |  |  |  |  |  |
| AP | APAC |  |  |  |  |  |  |  |  |  |
| AP | APAC\_AMRT\_DETAILS |  |  |  |  |  |  |  |  |  |
| APAC | APAC\_AMRT\_DETAILS |  |  |  |  |  |  |  |  |  |
| AP | APAC\_APAC\_RELN |  |  |  |  |  |  |  |  |  |
| AP | APAC\_ASSET\_RELN |  |  |  |  |  |  |  |  |  |
| DTS | APAC\_DOC |  |  |  |  |  |  |  |  |  |
| AP | APAC\_EVENTS |  |  |  |  |  |  |  |  |  |
| AP | APAC\_FIN\_PARAM |  |  |  |  |  |  |  |  |  |
| CT | APAC\_FIN\_PARAM\_LOOKUP\_TABLE1 |  |  |  |  |  |  |  |  |  |
| AP | APAC\_FLEXIBLE\_INFO |  |  |  |  |  |  |  |  |  |
| AP | APAC\_INTEREST\_PARAM |  |  |  |  |  |  |  |  |  |
| AP | APAC\_PARTY\_RELN |  |  |  |  |  |  |  |  |  |
| AST | APAC\_REFINANCE |  |  |  |  |  |  |  |  |  |
| RPT | APAC\_REVISED\_IRR\_DATA |  |  |  |  |  |  |  |  |  |
| AST | ASSET |  |  |  |  |  |  |  |  |  |
| AST | ASSET\_FLEXIBLE\_INFO |  |  |  |  |  |  |  |  |  |
| OLB | BALANCES |  |  |  |  |  |  |  |  |  |
| AP | CASH\_FLOW |  |  |  |  |  |  |  |  |  |
| CT | CITY |  |  |  |  |  |  |  |  |  |
| RCT | INSTRUMENT\_BANK |  |  |  |  |  |  |  |  |  |
| RCT | INSTRUMENT\_BOUNCE |  |  |  |  |  |  |  |  |  |
| IFC | KYC\_TEMPLATE\_DATA\_TAB\_B1\_B6 |  |  |  |  |  |  |  |  |  |
| CT | LOCATION |  |  |  |  |  |  |  |  |  |
| CT | LOCKER |  |  |  |  |  |  |  |  |  |
| CT | LOOKUP\_TABLE1 |  |  |  |  |  |  |  |  |  |
| CT | LOOKUP\_TABLE2 |  |  |  |  |  |  |  |  |  |
| CT | LOOKUP\_TABLE3 |  |  |  |  |  |  |  |  |  |
| AST | MODEL |  |  |  |  |  |  |  |  |  |
| RPT | MONTHLY\_ACCRUAL@RPT |  |  |  |  |  |  |  |  |  |
| RPT | MONTHLY\_ACCRUAL\_FORECAST\_V |  |  |  |  |  |  |  |  |  |
| RPT | MONTHLY\_ACCRUAL\_H |  |  |  |  |  |  |  |  |  |
| DTS | PAC\_DOC |  |  |  |  |  |  |  |  |  |
| PRM | PARTY |  |  |  |  |  |  |  |  |  |
| PRM | PARTY\_ADDRESS |  |  |  |  |  |  |  |  |  |
| PRM | PARTY\_CORRESPONDANCE\_ADDR |  |  |  |  |  |  |  |  |  |
| PRM | PARTY\_FIRM |  |  |  |  |  |  |  |  |  |
| PRM | PARTY\_FLEX\_INFM |  |  |  |  |  |  |  |  |  |
| PRM | PARTY\_INDIVIDUAL |  |  |  |  |  |  |  |  |  |
| PRM | PARTY\_IT |  |  |  |  |  |  |  |  |  |
| PRM | PARTY\_MIS\_CODE\_DTL |  |  |  |  |  |  |  |  |  |
| RCT | PAYIN\_SLIPS |  |  |  |  |  |  |  |  |  |
| RCT | PDC\_INSTR |  |  |  |  |  |  |  |  |  |
| RCT | PDC\_LEDGER |  |  |  |  |  |  |  |  |  |
| AP | PREDEFINED\_OL\_TRANS |  |  |  |  |  |  |  |  |  |
| PRM | PARTY\_FLEX\_INFM |  |  |  |  |  |  |  |  |  |
| APPL | PRODUCT |  |  |  |  |  |  |  |  |  |
| AP | PROPOSAL |  |  |  |  |  |  |  |  |  |
| CT | RBI\_MICR |  |  |  |  |  |  |  |  |  |
| RCT | RECEIPT |  |  |  |  |  |  |  |  |  |
| RCT | RECEIPT\_INSTRUMENT |  |  |  |  |  |  |  |  |  |
| CREDIT | SANCTION\_CONFIG |  |  |  |  |  |  |  |  |  |
| CREDIT | SANCTIONS |  |  |  |  |  |  |  |  |  |
| APPL | SCHEME |  |  |  |  |  |  |  |  |  |
| RCT | SI\_ALLOCATION |  |  |  |  |  |  |  |  |  |
| RPT | SOH |  |  |  |  |  |  |  |  |  |
| RPT | SOH\_DAYWISE\_V |  |  |  |  |  |  |  |  |  |
| RPT | SOH\_H |  |  |  |  |  |  |  |  |  |
| RCT | STANDING\_INSTR |  |  |  |  |  |  |  |  |  |
| TAXES | TAX\_DETAILS |  |  |  |  |  |  |  |  |  |
| IFC | TEMP\_GST\_EXPENSE\_DATA |  |  |  |  |  |  |  |  |  |
| IFC | TEMP\_GST\_INCOME\_DATA |  |  |  |  |  |  |  |  |  |
| OLB | TRANS |  |  |  |  |  |  |  |  |  |
| OLB | TRANS\_DETAILS |  |  |  |  |  |  |  |  |  |
| AP | V\_APAC\_ASSET\_INFO |  |  |  |  |  |  |  |  |  |
| IFC | VW\_APAC\_PAYMENT\_INFO |  |  |  |  |  |  |  |  |  |
| RPT | VW\_OGL\_ACCOUNT\_SPLIT |  |  |  |  |  |  |  |  |  |
| PRM | VW\_PARTY\_CITY |  |  |  |  |  |  |  |  |  |
| RPT | SOH\_V |  |  |  |  |  |  |  |  |  |
| RPT | IDR\_V |  |  |  |  |  |  |  |  |  |
| PYT | PAYMENT\_TRANS |  |  |  |  |  |  |  |  |  |
| PYT | PAYINSTR |  |  |  |  |  |  |  |  |  |
| AFP | APPL |  |  |  |  |  |  |  |  |  |
| APAC | LOCATION |  |  |  |  |  |  |  |  |  |
| APAC | SUB\_LOCATION |  |  |  |  |  |  |  |  |  |
| LOC | STATE\_CODE |  |  |  |  |  |  |  |  |  |
| AFP | APAC |  |  |  |  |  |  |  |  |  |
| AFP | CONTRACT\_NO |  |  |  |  |  |  |  |  |  |
| AFP | AGR\_DATE |  |  |  |  |  |  |  |  |  |
| AFP | PHYSICAL\_AGR\_DATE |  |  |  |  |  |  |  |  |  |
| AFP | APAC\_EFFECTIVE\_DATE |  |  |  |  |  |  |  |  |  |
| AFP | AGR\_PERIOD\_SIZE |  |  |  |  |  |  |  |  |  |
| AFP | MATURITY\_DATE |  |  |  |  |  |  |  |  |  |
| AFP | FINANCE\_TYPE\_CD |  |  |  |  |  |  |  |  |  |
| AFP | INTEREST\_TYPE\_LU\_CODE1 |  |  |  |  |  |  |  |  |  |
| AFP | AGR\_VALUE |  |  |  |  |  |  |  |  |  |
| AFP | PRODUCT\_ID |  |  |  |  |  |  |  |  |  |
| APAC | MAIN\_PARTY\_ID |  |  |  |  |  |  |  |  |  |
| PRM | F\_GET\_PARTY\_NAME |  |  |  |  |  |  |  |  |  |
| AFP | DEAL\_IRR |  |  |  |  |  |  |  |  |  |
| AFP | CUST\_FLOW\_IRR |  |  |  |  |  |  |  |  |  |
| AFP | AUTHZ\_BY |  |  |  |  |  |  |  |  |  |
| AFP | AUTHZ\_DT |  |  |  |  |  |  |  |  |  |
| AFP | SCHEME\_ID |  |  |  |  |  |  |  |  |  |

# Metadata information For INGESTION, PREP & SERV lAYERS

# Mapping Document

# Appendix

Sample DDL scripts