Golden Gate Implementation with ODI as the ETL

One of India's largest private sector banks
Mumbai

Industry
Banking

Employees
5001-10,000

Services
- Golden Gate Replication
- ODI ETL

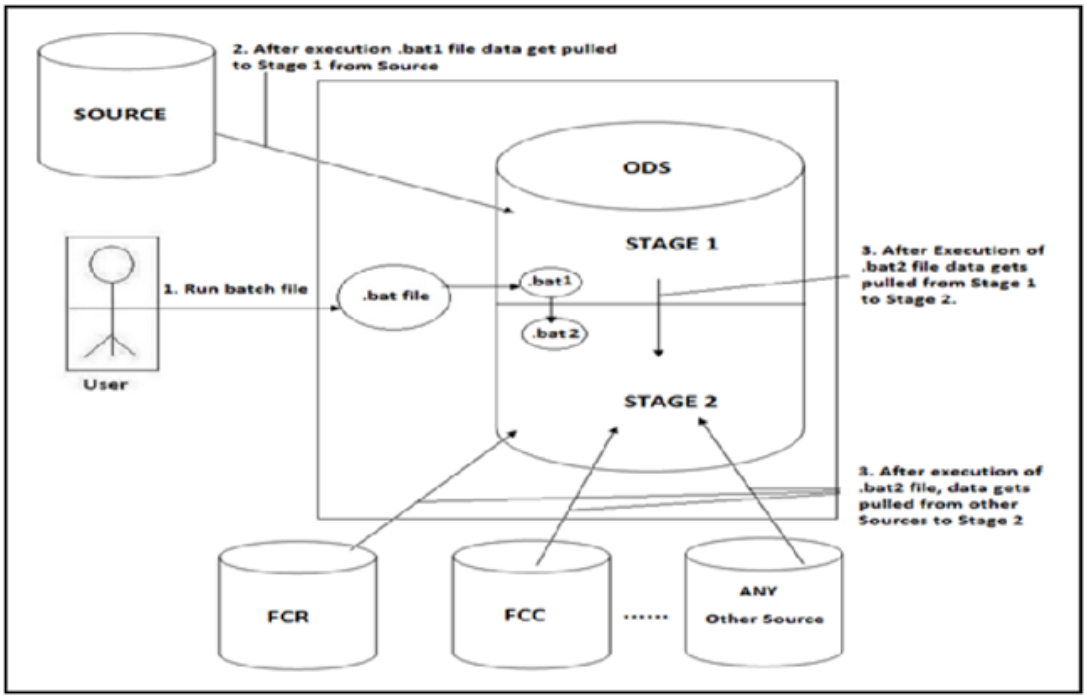
Overview

The client is one of India's largest private sector banks. It has a widespread branch network of over 572 branches across 375 cities, with 1170+ ATMs and 2 National Operating Centers in Mumbai and Gurgaon.

Problem Statement:

The Bank was extracting data from various sources after completing the systems EOD activities. The process was a manual run of the batch files of each system after their EOD. Since each system has its own processes and dependent systems, the process for each system ran at a different time. The process was prone to manual errors (missing out a run, multiple runs, wrong batch run, person dependent), different run-times, sequence dependent, missing of transactions during the run and many other issues. The result was that the quality and data integrity on the BI reports generated gave an erroneous view to the business decision makers.

Current Approach is elicited in the diagram below:



- A batch file (.bat file) is run to extract data from the main system to populate the Operational Data Store (ODS) – Stage 1. This database acts as the designed database to integrate data from multiple sources for additional operations on the data. Data is then passed to the reporting system. This is a sequential run performed manually.

- The data is then subjected to operational transformations as required by the business, post which, another batch file (.bat1 file) is run to load the Stage 2 of ODS.
- Allied data from other systems (such as FCC, FCR etc.) is brought into Stage 2 depending on EOD closures in these systems and the ability to fetch interlinked data from the source data.
- After complete loading in stage 2, data is transformed and sent to BI systems.

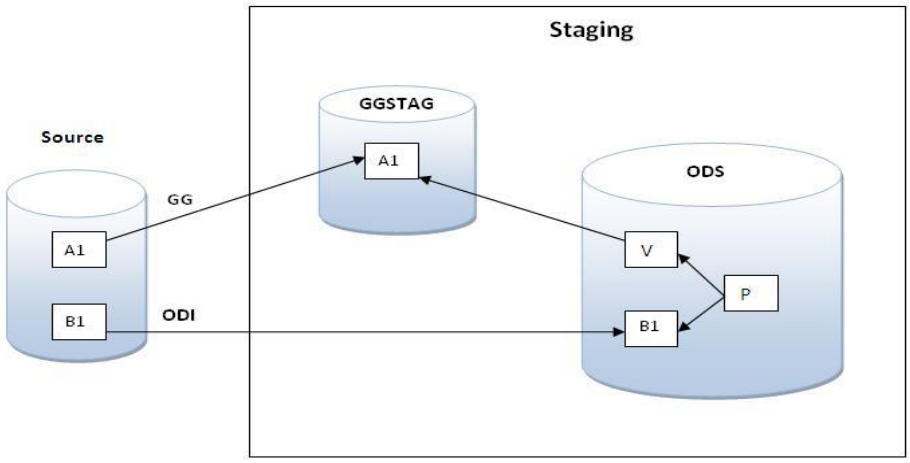
Challenges in the existing approach:

- Jobs need to be run manually as each system had dependencies on EOD run of the respective system.
- While the EOD process execution is on, if any transaction gets recorded into the system, then it gets captured into the TEMP table/file. These transactions may be different in different systems.
- The events cannot be scheduled, because their run-times are not fixed.
- To run the batch files, user need to login to the server which is a security concern.
- User may forget to run the batch files which could result in critical issues, missing data and failed processes.
- User may run a wrong batch file if there is dependency.
- User may run batch file twice hence the system is prone to human errors

Solution:

Clover in consultation with the client decided to automate and remove the manual dependencies. It was suggested that the processes and report generation be near Real-time with checkpoint mechanism for report runs. This ensures data integrity in the BI reports. **It was decided to use Oracle Golden Gate for real-time replication of the data and use Oracle Data Integrator (ODI) for ETL run to transform the data as per business use-cases.**

- Oracle Golden Gate replicated the data in real-time from various systems into the ODS. Golden Gate was configured on the Source System and Target System to pull data from the source systems to the target system. Requisite ports were opened for each system for the data transfer.
- Golden gate parameters were configured to stop the replication process post the EOD event for each of the source system. Any data logged into the source system, post EOD, was stored in trail files but not replicated immediately.
- Post completion of all the EOD's of all the system (Stage 1), the ETL transformation was triggered using ODI. Triggers were created in ODI to run the ETL processes of all the connected systems.
- Post completion of the transformation phase (Stage 2), Golden gate replication process was restarted. The replication started from the trail files ensuring no data loss and complete data integrity in the reports generated by the BI systems.
- As there were a large number of mutually independent reports,, parallel processes were configured in ODI to generate the reports independently resulting in faster BI reporting.



Value Propositions:

- Automation of the process and removal of all manual dependencies, runs, privileges and associated issues.
- Data integrity and authenticity to give exact view of the business by ensuring the processes of all connected systems are run exactly after their EOD process
- Real-time reporting of the business is available and the data can be fetched from the BI systems, as and when required.
- Majority of the replication, ETL and reporting processes are configured to run simultaneously thereby reducing the TAT by 60%.