**Executive Summary**

Following a review of the RFPs received, the decision was made in favor of setting up a Data Warehouse. We were supplied with a series of Data Definition Language text files to guide us through building the database as per the client’s requirements. The Data Warehouse has been created using Oracle Cloud due to its consistent placement in the Gartner Magic Quadrant – which is a testament to how robust and top-of-the-shelf this product is. To prevent incurring costs for a new Data Warehouse, we reused a Data Warehouse we had created previously. All the older tables were deleted and new tables were created.

Furthermore, we added data to the Data Warehouse using various techniques. We added the dates to the Date dimension using a script that auto-generates all the fields within the Date table when given a starting range. The client had asked us to update the date range after the first draft of the Date table was shared with them. We then added data to the Customers table & the Products table using SQL’s INSERT statements. Finally, we also loaded data to the tables using Talend Open Studio – another great ETL (Extract Transform Load) tool that is leading the Gartner Magic Quadrant. This was a mission-critical part of the project since it involved ironing out tasks such as connecting Talend to our Oracle Cloud environment while ensuring that all security protocols for authentication were followed so it could lead to a successful connection.

After updating the tables with the data contained in the .CSV files using Talend, we identified that there was a discrepancy in the number of products in the .CSV file provided by the client, and the Products table that we had previously loaded using SQL INSERT statements.

To ensure that our database has the ability to look back into time, we added Slowly Changing Dimensions to our database using Talend. Slowly Changing Dimensions allow us to preserve history. These were enabled on the Customer, Product, and Sales tables. For the Product table to define SCD type, the field ‘productid’ is identified as Source Key, and the other relevant fields that are expected to be modified over the course of business and are expected to be retained history for were marked as SCD Type 2 fields. Similarly, for the Customer table, the ‘custid’ is identified as the Source key, ‘birthday’ as SCD type0 (as it does not change), ‘name’ as SCD type 1(as it can be changed), and the remaining fields are marked as type 2.

Eventually, once all the other tables and conditions had been set, we loaded the Sales Fact Table with data. The Sales table ties together the entire Database as we follow the Kimball methodology. It is the proverbial ‘center of our star schema’.

After all the Dimensions and Fact tables have been successfully loaded, we run a ‘SELECT \*’ query to view the finished dataset and confirm that there are no anomalies or unintended data points that were not a part of our ETL process but might have crept in as an outcome of an incorrect action taken during the process. This marks the end of the ETL Pipeline process and the Data Warehouse is now ready for the final stage of the Data Warehouse lifecycle – adding a Business Intelligence layer to derive actionable insights to enable impactful business decisions.

**Appendix:**

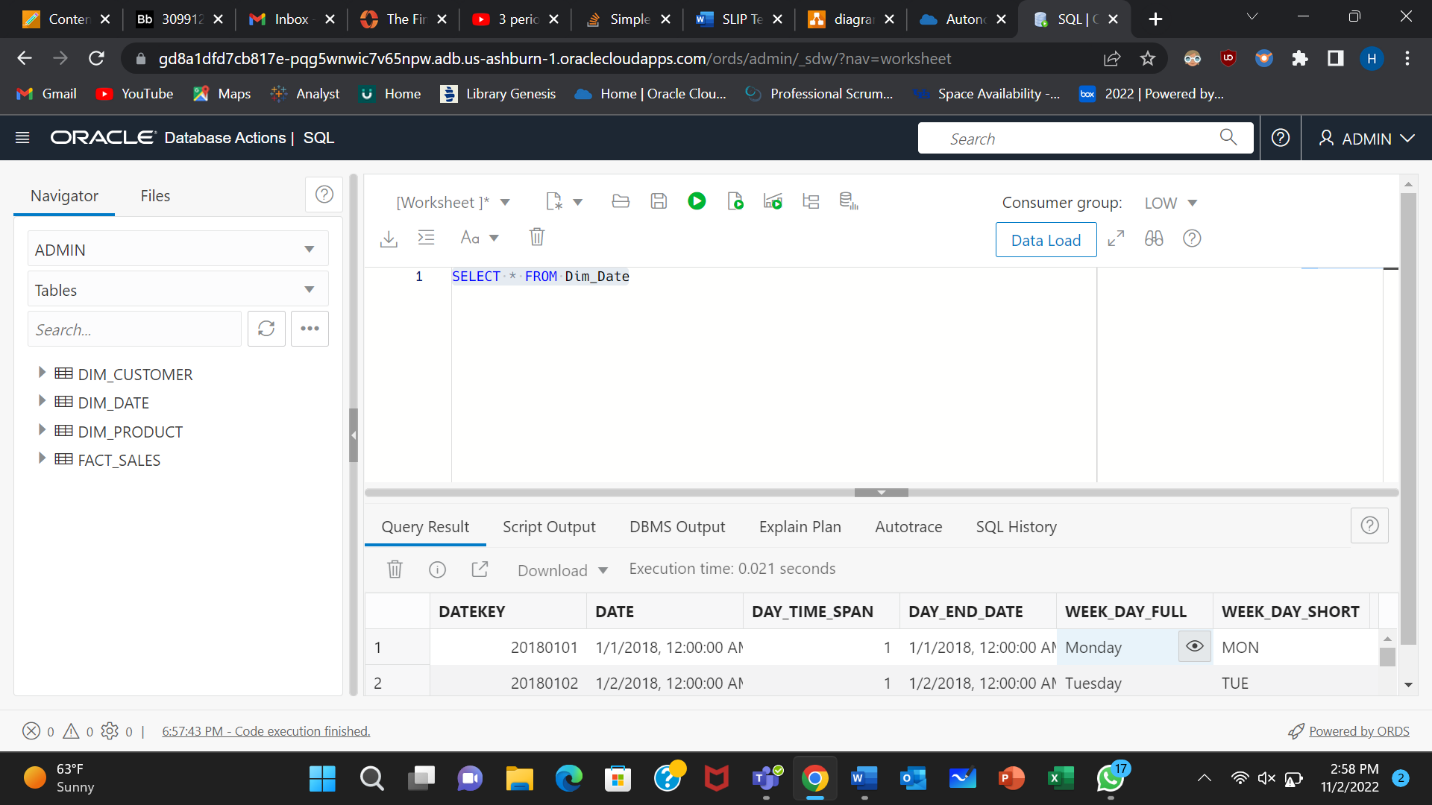
A picture containing timeline

Description automatically generated

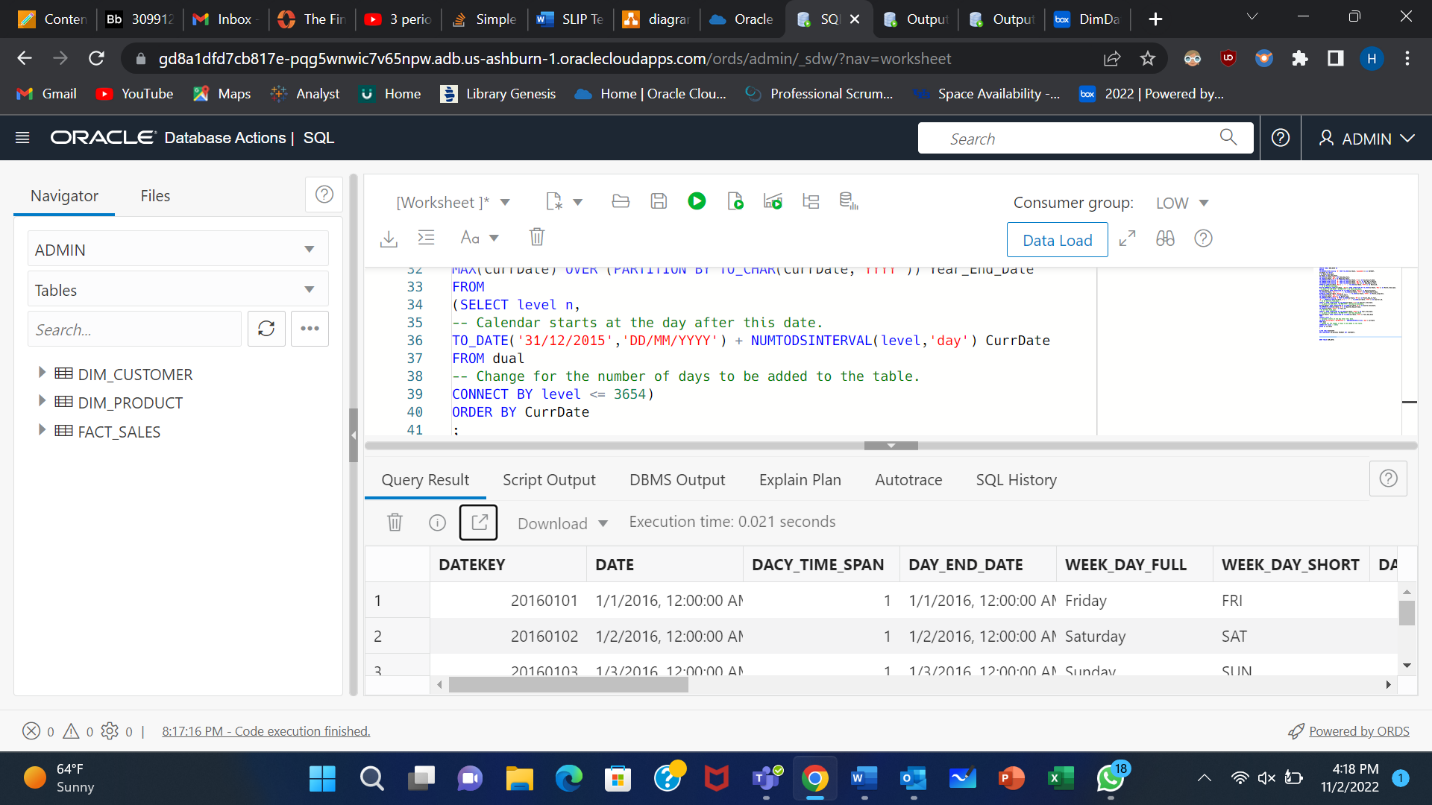
A pivotal detail missing in the diagram above is the foreign key relationships between tables in the Oracle Cloud environment. This is due to the ETL process we are following on this project. These relationships will be established later using the tMAP component in Talend Open Studio. tMAP transforms and directs data from single or multiple sources to single or multiple destinations ([Talend](https://help.talend.com/r/en-US/7.3/tmap?tocId=tPOEpGz3hZf4MiowThLTlA)). There are considerable advantages with respect to Referential Integrity by doing this in the ETL pipeline rather than the Database ([Information Management – Point 5](https://web.archive.org/web/20201125230646/https:/www.information-management.com/news/best-practices-for-data-warehouse-database-developers)).

**DATE TABLE:**

“Select \* from Dim\_Date” output before updating the script:



“Select \* from Dim\_Date” output with updated script:



**PRODUCT TABLE:**

“Select \* from Dim\_Product” query result:

A screenshot of a computer

Description automatically generated

“SELECT \* FROM DIM\_PRODUCT” query results after updating the records through Talend ETL

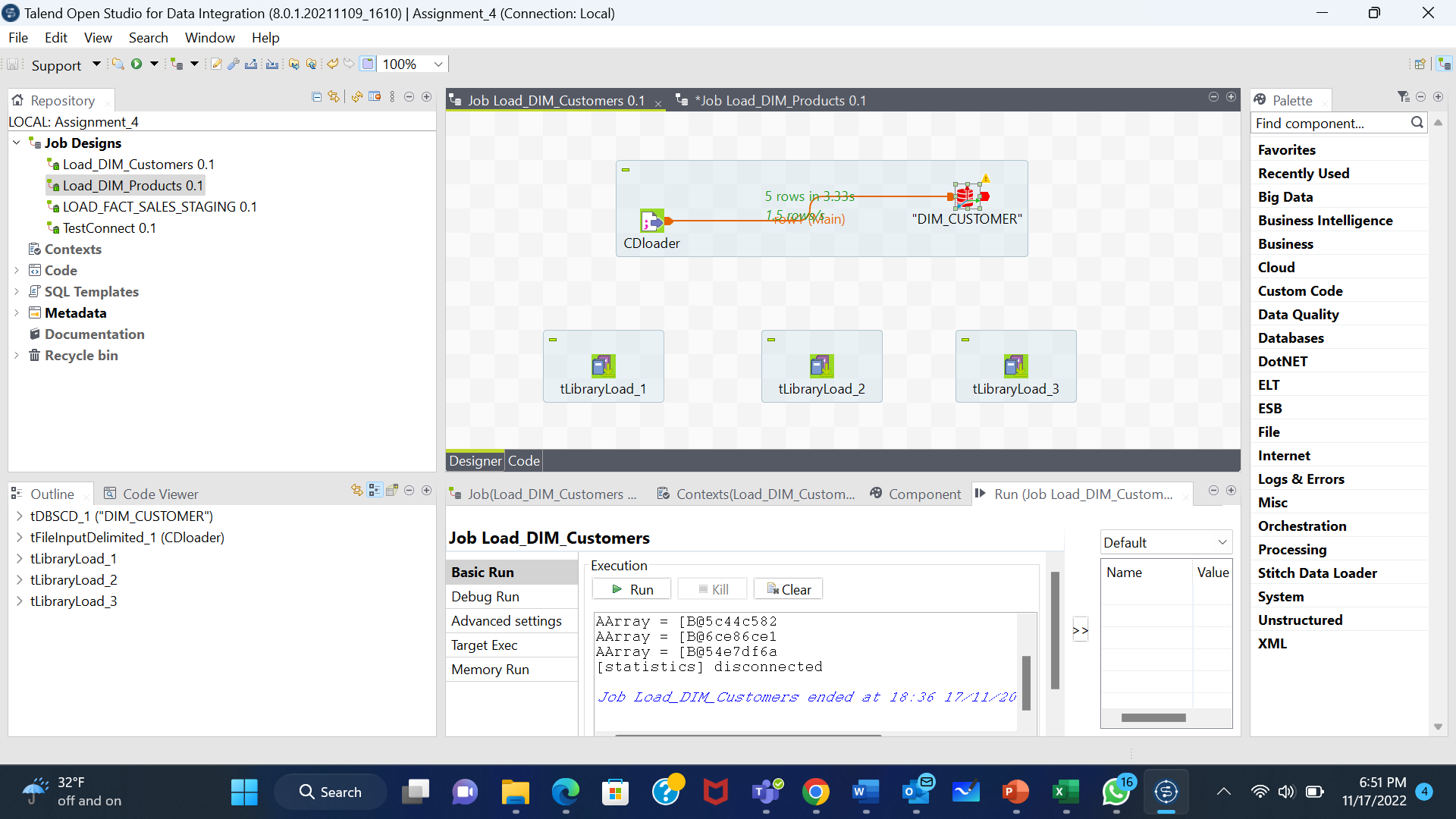
tool:

A screenshot of a computer

Description automatically generated

**CUSTOMER TABLE:**

Talend Job of dim\_customer:



“SELECT \* FROM Dim\_Customer” query result:

A screenshot of a computer

Description automatically generated

**FACT TABLE:**

“SELECT \* FROM Fact\_sales” query results:

A screenshot of a computer

Description automatically generated

Talend Job of fact\_sales:

Graphical user interface, application

Description automatically generated

Screenshot of tMap screen:

A screenshot of a computer

Description automatically generated with medium confidence