

ETL and Reporting for Employee Projects

Contents

Introduction	1
High Level Process.....	1
Source Database	2
Transformations.....	2
Destination Data Warehouse.....	3
Reporting	3
Files	4
Troubleshooting.....	4
Conclusions	5

Introduction

This document specifies the ETL process for ingesting data about Employees, projects and the number of hours worked by an employee on a project from a source database to a data warehouse. The ETL process extracts information from a given EmployeeProjects relational database, performs transformations on the field names and data, and finally loads the data into a data warehouse with relevant dimensions and fact tables for consistent and usable data.

The create, update, delete actions performed in the ETL process are logged in a meta data table and a basic report is created showing the log actions and time using MS report builder.

High Level Process

The ETL process (Extract Transform and Load) from the source EmployeeProjects database into the destination DWEmployeeProjects database, as shown in Figure 1 below.

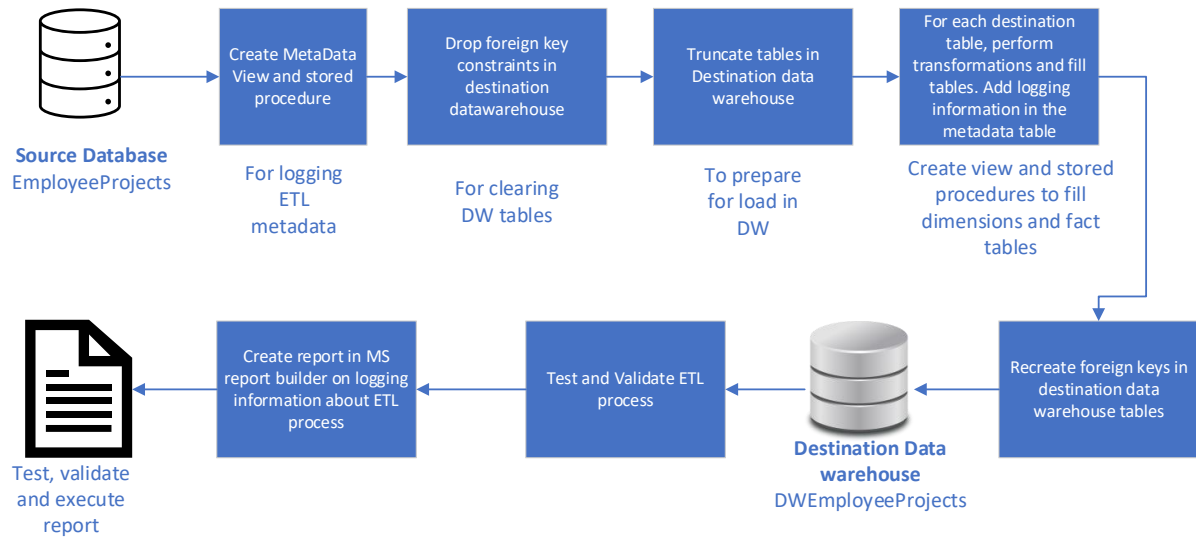


Figure 1 ETL Process from EmployeeProjects into DWEmployeeProjects

Source Database

The data model for the source database EmployeeProjects is shown in Figure 2 below. There are three tables: Employees, Projects and EmployeeProjectHours.

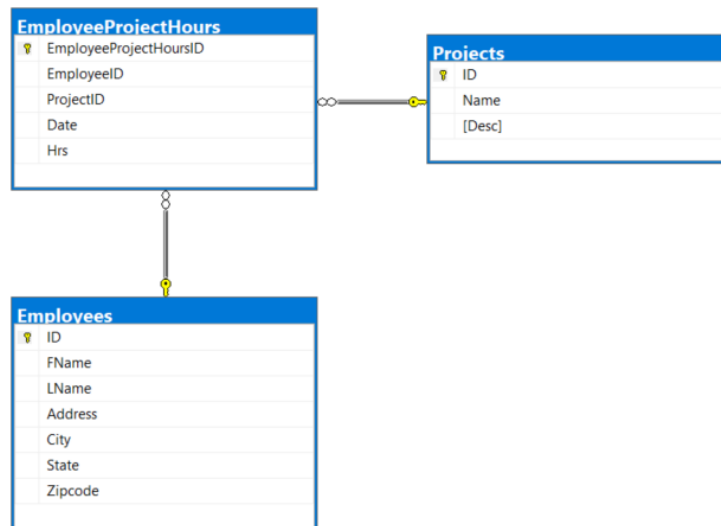


Figure 2 EmployeeProjects Data Model

Transformations

The transformations in the ETL process to clean the data, create dimension and fact tables for Data warehouse are listed in the 3_Assignment03MetadataWorksheet.xls metadata Excel file. Surrogate keys are added to all tables. Non essential fields are dropped. The arbitrary textual datatypes are changed to varchar(100). A dimension table for dates (called DimDates) is created. In total there are 31

transformations. Please refer to 3_Assignment03MetadataWorksheet.xls metadata Excel file for more details.

Destination Data Warehouse

The data model for the destination database DWEmployeeProjects is shown in Figure 3 below. The dimension tables are DimProjects, DimEmployees, DimDates. The fact table is FactEmployeeProjectHours.

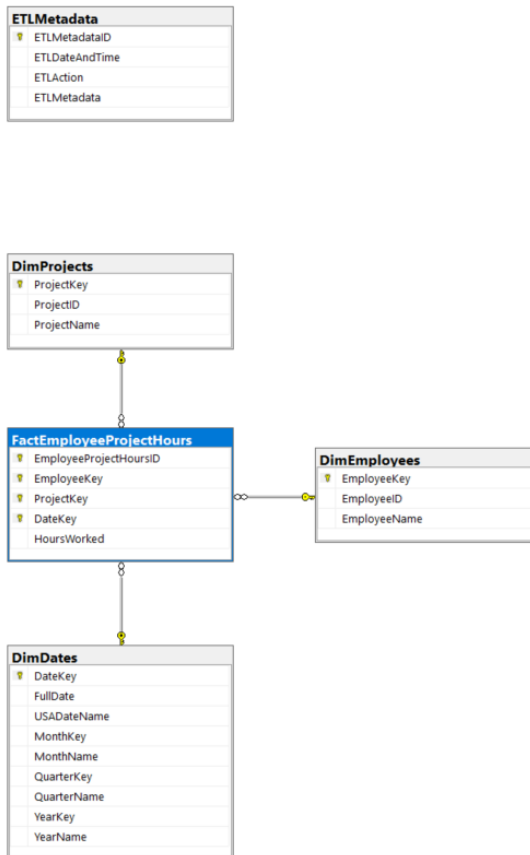


Figure 3 DWEmployeeProjects DW

Reporting

The ETL log for DWEmployeeProjects is created in MS Report Builder. To obtain the report, launch the ETLogReport.rdl and click Run to generate the report. See Figure 4 and Figure 5. ID, Date and Time columns can be sorted in the output.

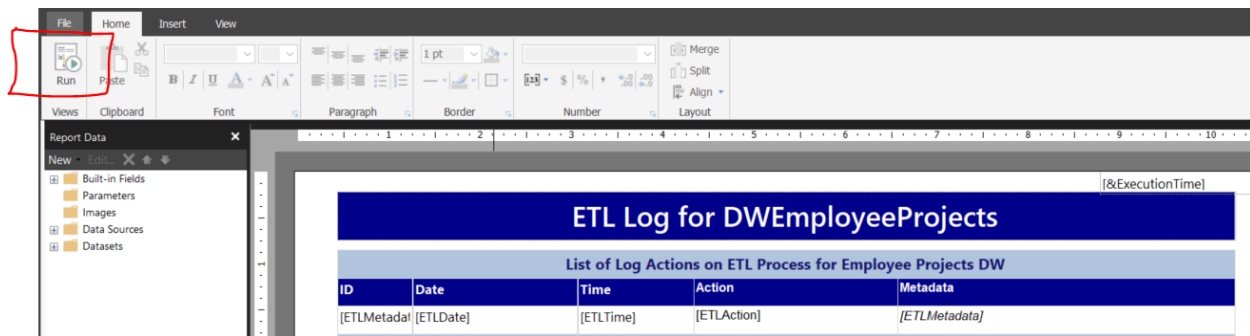


Figure 4 Run ETL Log Report – Design Mode

ETL Log for DWEmployeeProjects				
List of Log Actions on ETL Process for Employee Projects DW				
ID	Date	Time	Action	Metadata
1	Tuesday, July 26, 2022	00:12	pETLDropFks	Dropped Foreign Keys
2	Tuesday, July 26, 2022	00:12	pETLTruncateTables	Truncated Tables
3	Tuesday, July 26, 2022	00:12	pETLDimEmployees	Filled DimEmployees (30 rows)
4	Tuesday, July 26, 2022	00:12	pETLDimProjects	Filled DimProjects (10 rows)
5	Tuesday, July 26, 2022	00:12	pETLDimDates	Filled DimDates (368 rows)
6	Tuesday, July 26, 2022	00:12	pETLReplaceFks	Replaced Foreign Keys
7	Tuesday, July 26, 2022	00:31	pETLDropFks	Dropped Foreign Keys
8	Tuesday, July 26, 2022	00:31	pETLTruncateTables	Truncated Tables
9	Tuesday, July 26, 2022	00:31	pETLDimEmployees	Filled DimEmployees (30 rows)
10	Tuesday, July 26, 2022	00:31	pETLDimProjects	Filled DimProjects (10 rows)
11	Tuesday, July 26, 2022	00:31	pETLDimDates	Filled DimDates (368 rows)

Figure 5 ETL Log Report Output

Files

- 2_Assignment03SourceDatabase.sql: SQL script to create and fill source tables, EmployeeProjects database
- 4_Assignment03DWDDatabase.sql: SQL script to create DW empty tables
- 3_Assignment03MetadataWorksheet.xlsx: Transformations metadata sheet/
- 5_Assignment03ETL.sql : ETL SQL script
- ETLLogReport.rdl – ETL logging report in MS report builder

Troubleshooting

- Connect SS Management Studio to local host. If trouble connecting to local host, ensure that SQL Server is running. Check SQL Server Configuration Manager. See [SQL Server Configuration Manager - SQL Server | Microsoft Docs](#)
- Report Builder is useful for simple tables, but hard to show filters and drill through. Would suggest not to edit the rdl file to increase functionality.
- In case the report does not run, ensure that the datasource is LogData and the DataSets include LogData

Conclusions

This project performs an ETL process to clean and transform data about Employees, projects and hours worked into a data warehouse with dimensions and a fact table. The ETL actions are logged in a metadata table and a report is generated about these actions. The source and data warehouse are simple and intuitive enough to enable easy maintenance and support.