Jacob Dineen, Denise Dunham, Anuj Vinod Dubey, QunfangWu

IST 722- Data Warehouse

Group #1

Final Project

Contents

[Project Charter 1](#_Toc501306432)

[Introduction 1](#_Toc501306433)

[Objectives 2](#_Toc501306434)

[Risks and Constraints 2](#_Toc501306435)

[Project Scope 2](#_Toc501306436)

[Strategy 2](#_Toc501306437)

[Timeline 3](#_Toc501306438)

[Project Plan 3](#_Toc501306439)

[Functional Requirements 4](#_Toc501306440)

[Business Processes/ Business Value 5](#_Toc501306441)

[Role Assignment 5](#_Toc501306442)

[High-Level Dimensional Modeling 5](#_Toc501306443)

[Detail-Level Dimensional Modeling 5](#_Toc501306444)

[Generate SQL Code and Create Target Tables 6](#_Toc501306445)

[Source to Target Map 6](#_Toc501306446)

[Initial ETL 6](#_Toc501306447)

[ETL Documentation of Employee Timesheets Business Process (Business Process 8) 7](#_Toc501306448)

[ETL Documentation for Customer Revenue Business Process (Business Process 1) 8](#_Toc501306449)

[Business Intelligence 13](#_Toc501306450)

[SQL Server Analysis Services: 13](#_Toc501306451)

[Excel Pivot Tables 14](#_Toc501306452)

[PowerBI 14](#_Toc501306453)

[Conclusion: 16](#_Toc501306454)

# Project Charter

## Introduction

Fudgemart, Inc. has hired Group 1 to build a Data Warehouse/Business Intelligence solution from scratch. Fudgemart, Inc. is a fictitious conglomerate with two subsidiary companies:

1. Fudgemart: Fudgemart is a fictitious online retailer, similar to Amazon.com or Walmart.com. The database consists of customers, products, and vendors, and has familiar business processes you would find in any online retailer. The database for Fudgemart is called Fudgemart\_v3.

2. Fudgeflix: Fudgeflix is a fictitious online DVD-by-mail and video-on-demand service, similar to Amazon Instant Video or Netflix. The database for Fudgeflix is called Fudgeflix\_v3 and contains concepts such as accounts, subscriptions, and video titles, as well as other things associated with an online video-streaming service.

A third database, ExternalSources, has some useful datasets for data warehousing in general as well as some Fudgemart sets with fm in their object names.

## Objectives

Group 1 will build a Data Warehouse and Business Intelligence (DW/BI) solution for Fudgemart, Inc., a fictitious conglomerate with two subsidiary companies. The DW/BI solution will address the functional requirements outlined in this charter by modeling the underlying business processes in the functional requirements. By modeling the underlying business processes the DW/BI solution can be used for related analyses and reporting needs not yet required.

## Risks and Constraints

* It will be challenging to aggregate customer attributes, billing history, sales, and other important business processes at a hierarchical level, because Fudgemart, Inc. consists of two separate subsidiaries, each having a separate data store.
* There is a time constraint with the solution required by 18 December.
* The team has limited domain knowledge on these specific companies and their respective source data.
* There is a lack of access to actual business users for functional requirement research.

## Project Scope

Group 1 will deliver a data warehouse that meets the needs of the client - Inclusive of KPI and reporting requirements as noted by the functional requirements and business processes seen below (1d).

|  |  |
| --- | --- |
| In Scope | Out of Scope |
| * Define functional requirements * High level dimensional model * Detail-level dimensional model * Data Warehouse on SQL Server * Initial ETL * Business Intelligence * Presentation and Demonstration | * Use of data sources not listed in Background * Additional business processes not listed in Charter * On-going ETL updates |

## Strategy

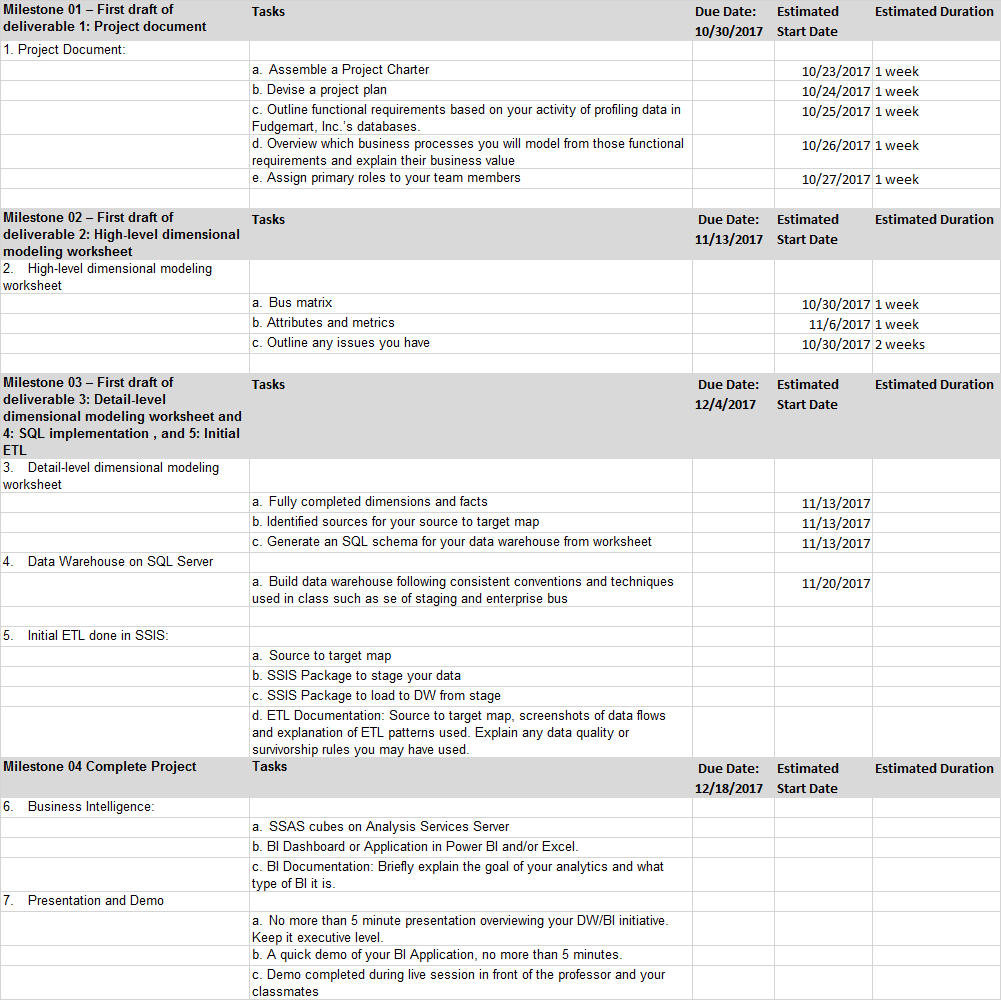
We will follow the Enterprise Bus Architecture, otherwise known as the Kimball Technical Architecture, with our main goal being to create a unified representation of master data and business processes within the specified landscape. With this approach, the data moves from the source systems through the staging area and into the Dimensional Data Stores, with conformed dimensions.

Our communication plan will be comprised of the following guidelines: We will meet once a week, every week. Initial discussion about milestones will take place during class, and will be followed by a subsequent meeting in the 2U platform which will feature more in-depth conversation and role delegation. All group members must be present, at least during the kick off call. When all roles are delegated, and a clear understanding is established, meetings can be attended only by personnel essential to the relevant task. All members are expected to complete their assigned projects, or reach out for assistance from another member of the team.

## Timeline

The project should be completed before Week 11 i.e 18 December. Updates will be provided in the form of milestones (mentioned below). Feedback from these updates may help to determine further direction.

## Project Plan

****

## Functional Requirements

1. The business users need to be able to analyze sales (defined as a customer completing a purchase) over a series of time, by geographic area, by customer demographic, by sales territory (if exists), by product, by product category, by vendor, and by shipped via.
   * Business Process: Sales Reporting
   * Business Value: Geographic analysis. Product analysis. Order fulfillment.

1.1 Combine the Fudgeflix rental sales with Fudgemart sales.

* + Business process: Sales Reporting and Comparison
  + Business Value: Sales Comparison for the two subsidiary companies.

1. The business users need to be able to analyze subscription sales (defined as a customer subscribing to a package, rather than purchasing a product, eg. Fudgeflix) over a series of time, by geographic area, by customer demographic, by title, by genre, by ‘Plan’,
   * Business Process: Subscription Sales
   * Business Value: Pricing analysis
2. The business users need to be able to track order delivery status, order queuing, order time elapsed between purchase and delivery, or delivery and return over a series of time, by product/title, by geographic area, by category/genre.
   * Business Process: Order fulfillment and Delivery
   * Business Value: shipment optimization, cost saving, customer satisfaction improvement
3. The business users need to be able to track inventory for specific products at any point in time.
   * Business Process: Product Inventory
   * Business Value: Inventory levels, Inventory projection, Reorder Triggers
4. The business users need to be able to track inventory for specific titles at any point in time.
   * Business Process: Rental History
   * Business Value: Inventory levels, Inventory projection
5. The business users need to be able to look at customer reviews/Review Count by product/title, over a series of time, by geographic area, by customer demographic, by title, by genre.
   * Business Process: Customer Reviews
   * Business Value: Customer Sentiment analysis
6. The business users need to be able to look at rental history by geographic location over a series of time.
   * Business Process: Rental History
   * Business Value: Inventory levels, Inventory projection
7. The business users need to be able to report monthly payroll expenses by Department and by Job title.
   * Business Process: Payroll reporting
   * Business Value: Payroll accounting, budget projections
8. The business users need to be able to determine which employees generate the most sales, the least sales, and analyze their compensation.
   * Business Process: Compensation Analysis
   * Business Value: Retain top sales employees by keeping salaries competitive, offering bonuses. Incentivize low sales employees

## Business Processes/ Business Value

The processes above will help us to better understand internal financials, order fulfillment and our customers’ purchasing habits/patterns. The structure of these dimensional models will lay the groundwork for us to analyze and predict future behavior within our organization, and refine existing processes.

## Role Assignment

Project Manager: All

Data Modeling: All

ETL: All

BI: All

Business Lead: All

Business Analyst: All

Lead Tester: All

\*Role assignment was initially delegated, but we ended up following more of an all-hands approach, file-sharing between milestones. Each member of the group wanted more hands-on experience with the different data warehousing processes, and this allowed for us to be knowledgeable about the specifics of the project without one person holding all of the key knowledge.

# High-Level Dimensional Modeling

For Milestone 2, we created a high level dimensional model worksheet for four of our business processes. The HLDM helped us to define the dimensions and facts for our business processes, and acted as a place for us to flag potential issues that we might not understand as well as the business user.

We decided to use our first, third, eighth and ninth business processes to model for milestone 2.

*High level dimension model worksheet can be found here:*

<https://docs.google.com/spreadsheets/d/1t2fv-FjrFwX_WEkOeDsXq0Hy5jvgz-g8KA5CRhajG8o/edit?usp=sharing>

# Detail-Level Dimensional Modeling

Our group chose business process #1 and #8. With a better understanding of dimensional models, we added clarity on the relationships possible between our specific facts and dimensions. As a result, the scope of our business processes was changed.

Business Process 1: Show revenue generated by Source, from both sources (Fact). We will have a conformed customer dimension table that draws from both Fudgeflix and Fudgemart. (We will utilize the pre-queried view to source our stage/target tables).

Business Process 8: Show Total Hours and Total Pay by Employee. Only Fudgemart has employee timesheets, so this business process will be specific to Fudgemart.

***Reference source to target map below for a high level view of our business processes.***

***Detailed Dimension Model worksheet can be found here:***

*https://drive.google.com/drive/folders/1xQadeS0uQL7t-rbfY8Nr2L\_\_ZEcCPpkA*

# Generate SQL Code and Create Target Tables

SQL code in the drive folder

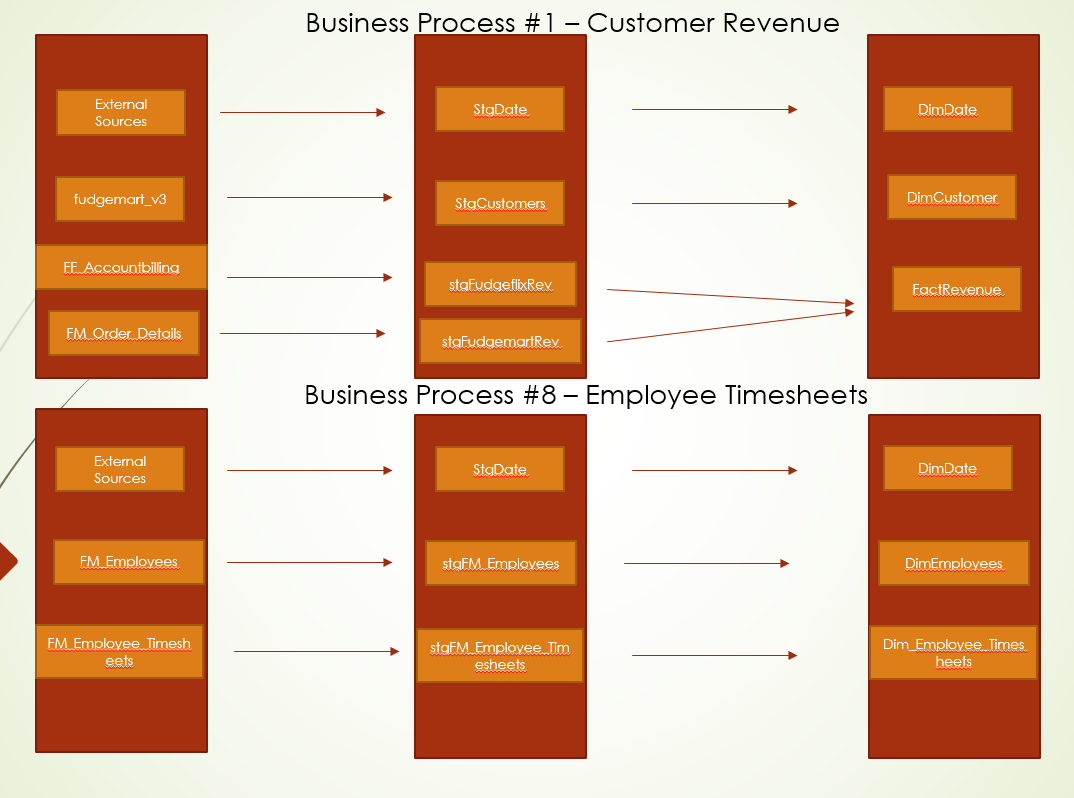
Using DB: [ist722\_mafudge\_1\_dw]

**SQL text file location:**

https://drive.google.com/drive/folders/1urTisrwJRjXf9fzSlGXFtOLSAeKCtnUj

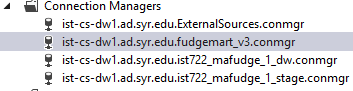
# Source to Target Map

**(In the Drive Folder)**



# Initial ETL

We established four separate connects to our server. ExternalSources contains the source data to populate our date dimension.

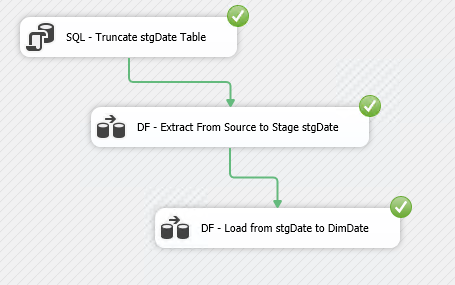


## ETL Documentation of Employee Timesheets Business Process (Business Process 8)

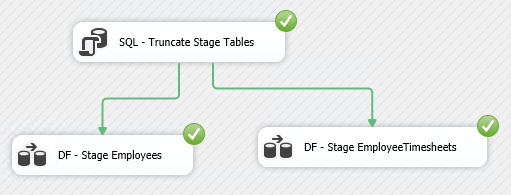
Our 8th business process focuses on our client’s need to view Fudgemart Payroll information by a series of dimensional attributes, such as job title and department. In order to get our data from the Source to the Target, we need to perform ETL - In this case we will be using SSIS (Microsoft Visual Studio 2015).

For this business process, we need two dimensions, those being DimDate and DimEmployee, and one FactTable, being FactEmployeeTimesheets. The below screenshots shows the flow of our ETL process with some idiosyncrasies noted in each data dataflow.

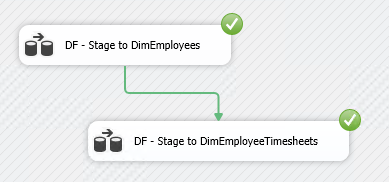
**DimDate**



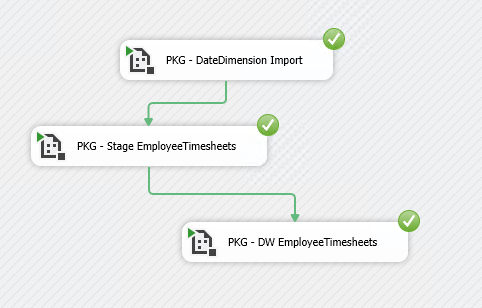
**DimEmployees and DimEmployeeTimesheets**



In order to populate Supervisor columns, we had to perform a self-join on our created stage table.



**All Packages**

****

**SQL Target Tables:**

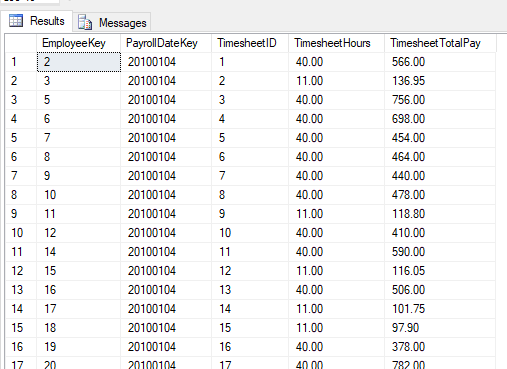
use ist722\_mafudge\_1\_dw

select \* from fudge.DimDate

select \* from fudge.DimEmployees

select \* from fudge.FactEmployeeTimesheets

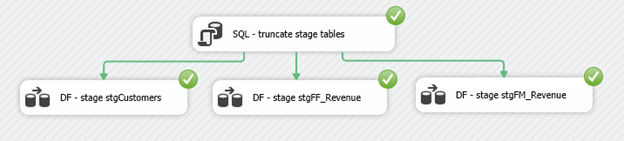
**SQL Fact Table Check**

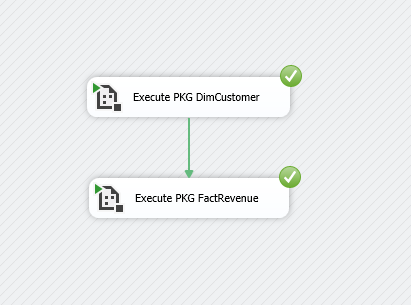
****

## ETL Documentation for Customer Revenue Business Process (Business Process 1)

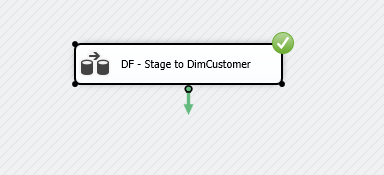
Our 1st business process focuses on our client’s need to see revenue totals by date, and at the customer level for both subsidiary companies side by side. Not only does this allow for comparative analysis of the sectors of the conglomerate that is Fudgemart, but it will also allow us to look deeper into customer segmentation- IE we will be able to see which people are customers of both companies. This business process will also allow for simple time series analysis of generated revenue across both companies. We will only have two dimensions here, Customer and Date, with our conformed fact table comprised of a single fact, which is revenue.

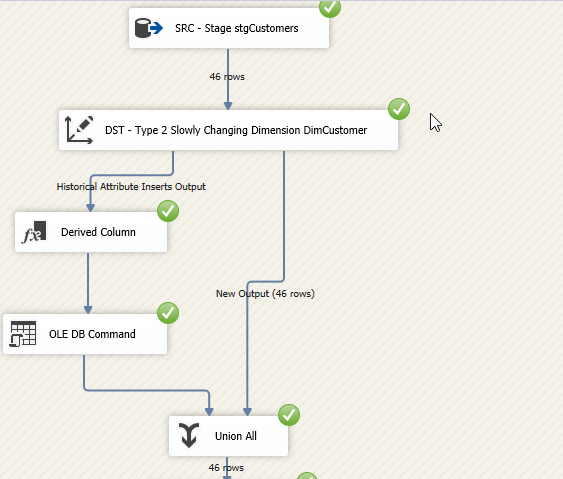
**Execute Packages**

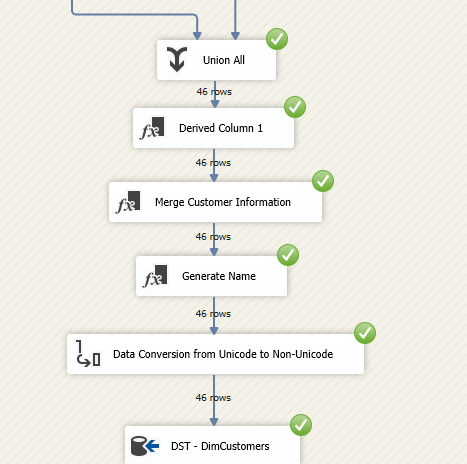
******

****

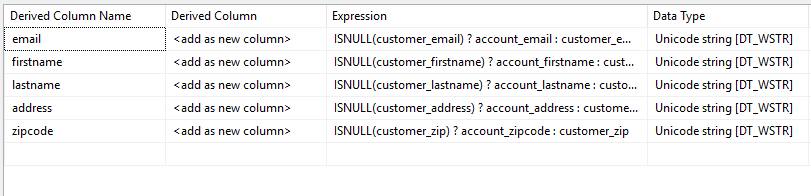
**Package DimCustomer**

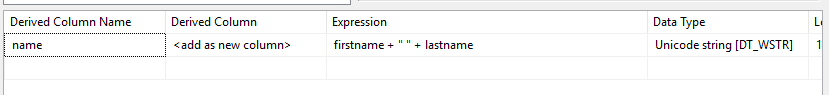
****

****

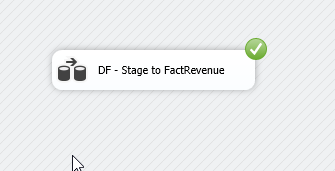
****

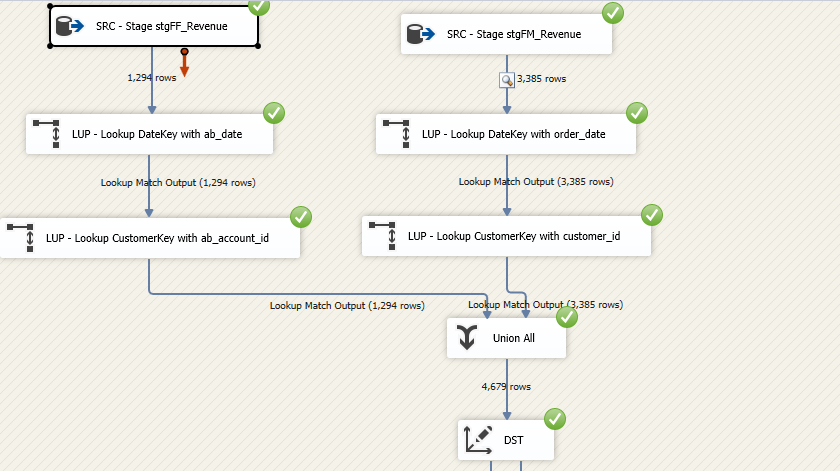
**Derive Columns for Customer Information**

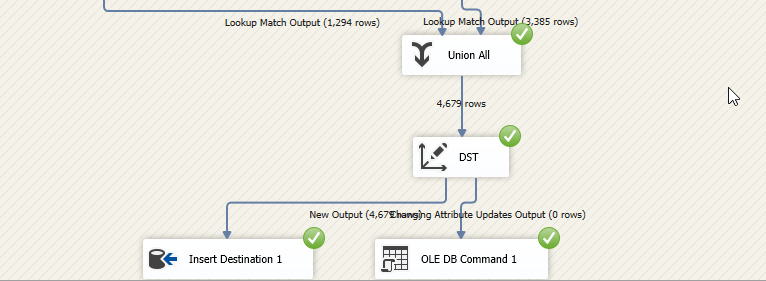
****

****

**Package FactRevenue**

******

****

****

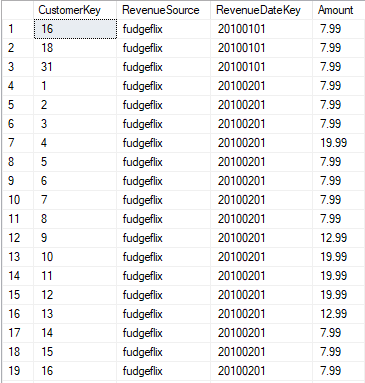
**SQL Target Tables:**

use ist722\_mafudge\_1\_dw

select \* from fudge.DimCustomers

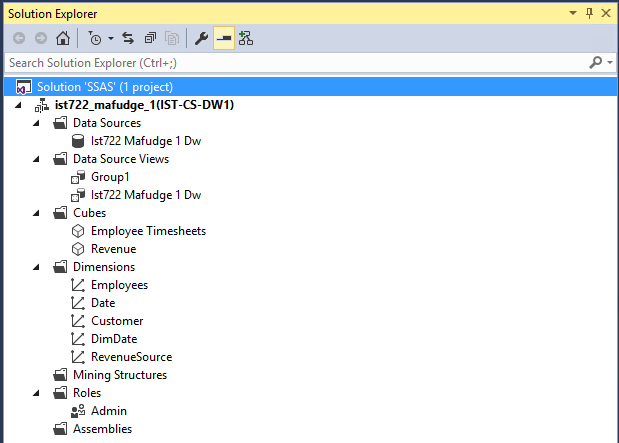
select \* from fudge.FactRevenue

**SQL Fact Table Check**

******

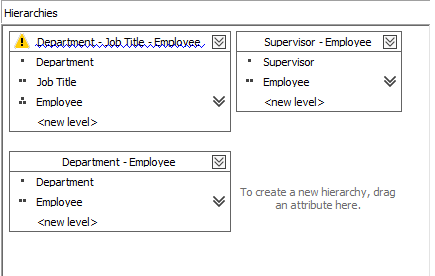
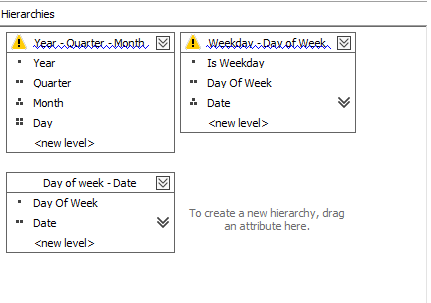
# Business Intelligence

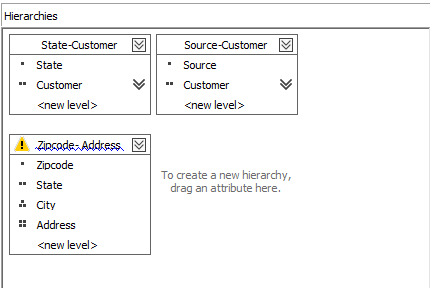
## SQL Server Analysis Services:



\*Note that Revenue Source is a degenerate dimension from the Revenue Fact table (Revenue Business Process). We wanted to be able to show revenue by the clustered transaction, and not just by the customer source attribute. See example below under ‘Excel Pivots’.

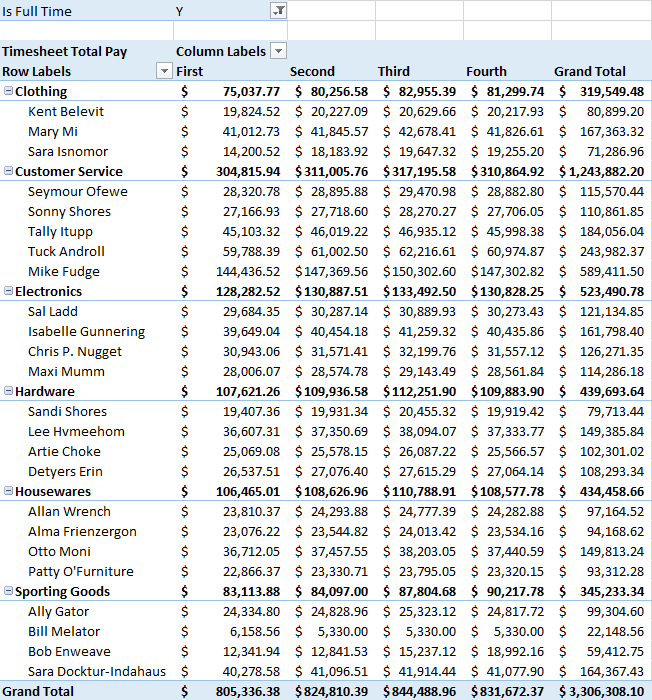
**Established Hierarchies**



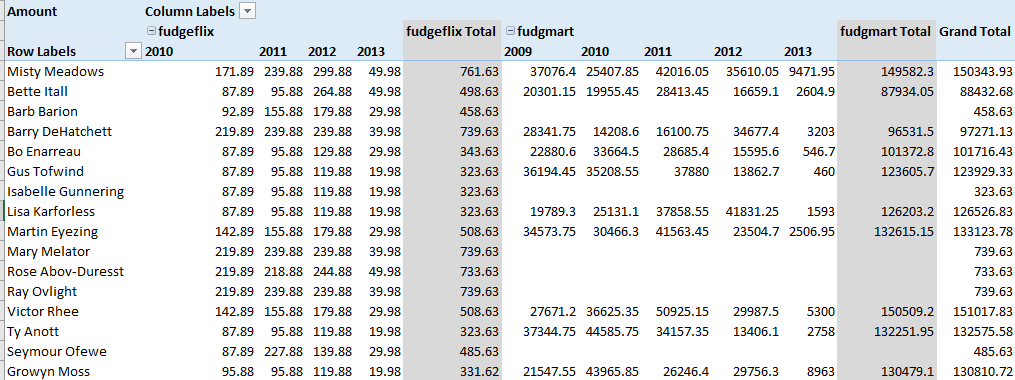


## Excel Pivot Tables

**Employee Timesheets business process**



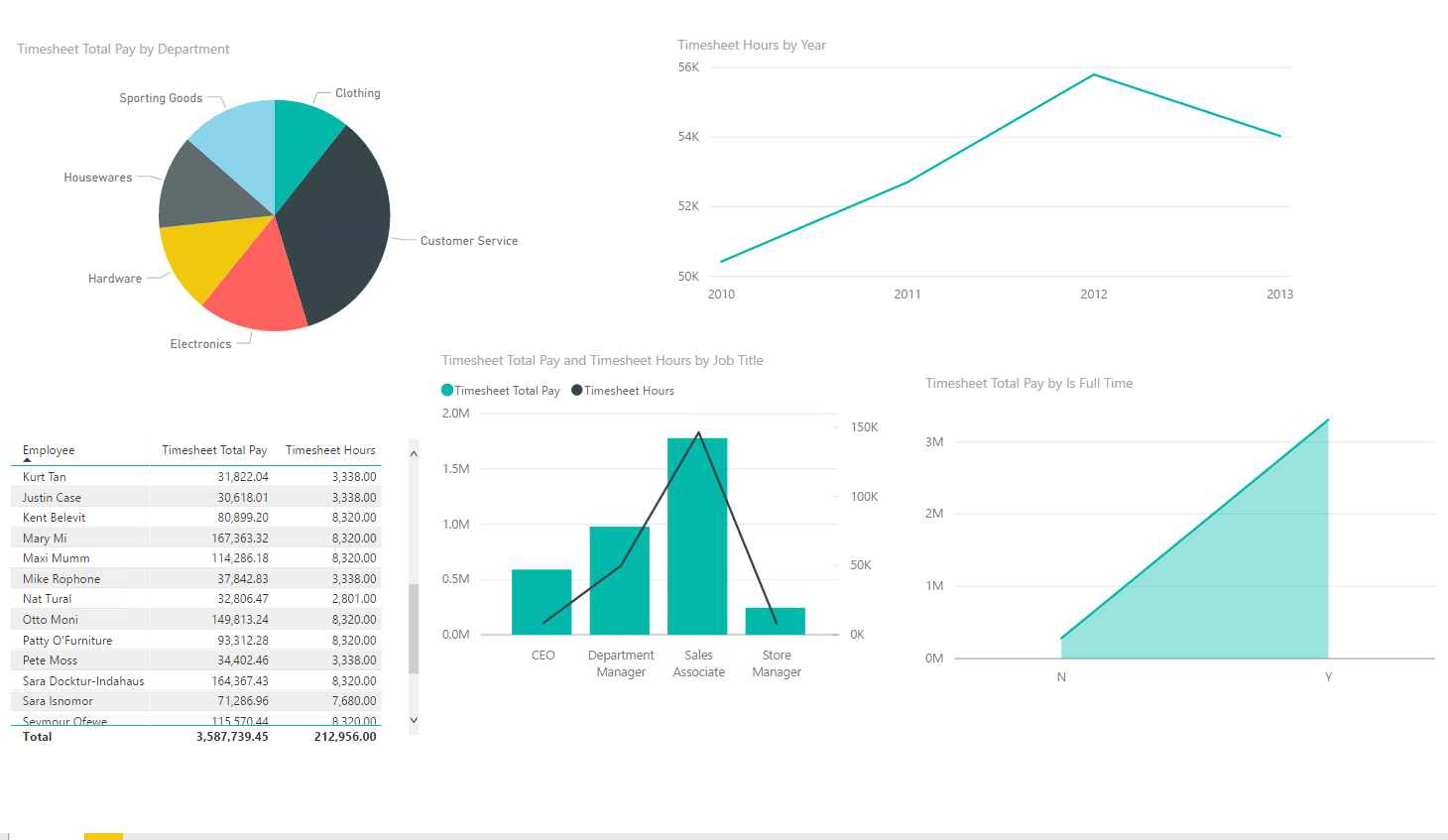
**Customer Revenue business process**

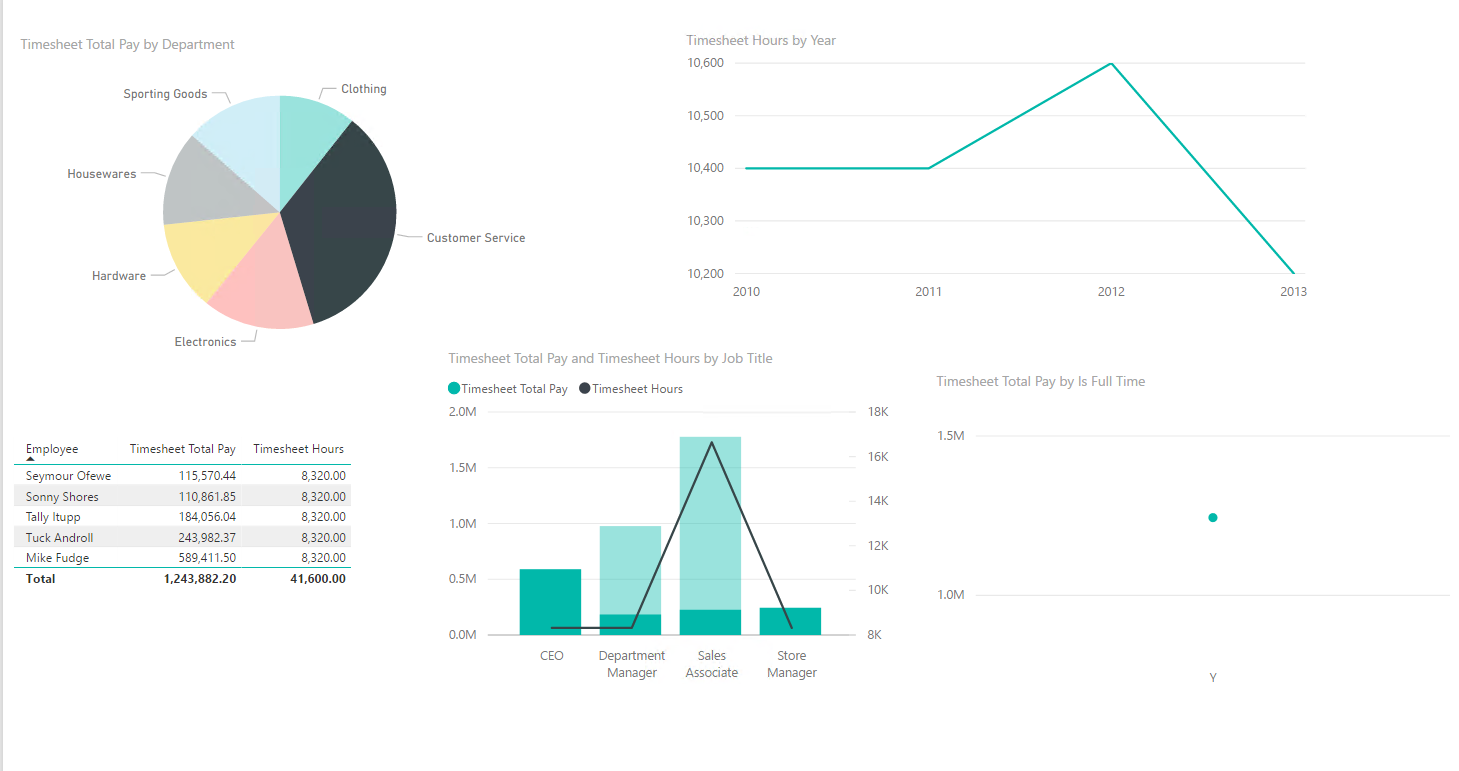


## PowerBI

The dashboard below allows for interactive querying of employee timesheet data, slice and dicable by the applicable dimensions. Each of the interactive charts is linked to the others, so we can change the filter of a single chart and see the results of the other charts per that filter. Example: Say we want to only look at timesheet facts (Hours/Total Pay) for a single department, like Customer Service, we can simply click on the Customer Service portion of the Pie chart and see the following results, which note the employees within that department and their pay, along with pay by job title, and timesheet hours submitted by year. A table was also added to this dashboard to allow for filtering on a certain employee. This is a digestible way of providing granular insights at the employee level for supervisors to monitor.

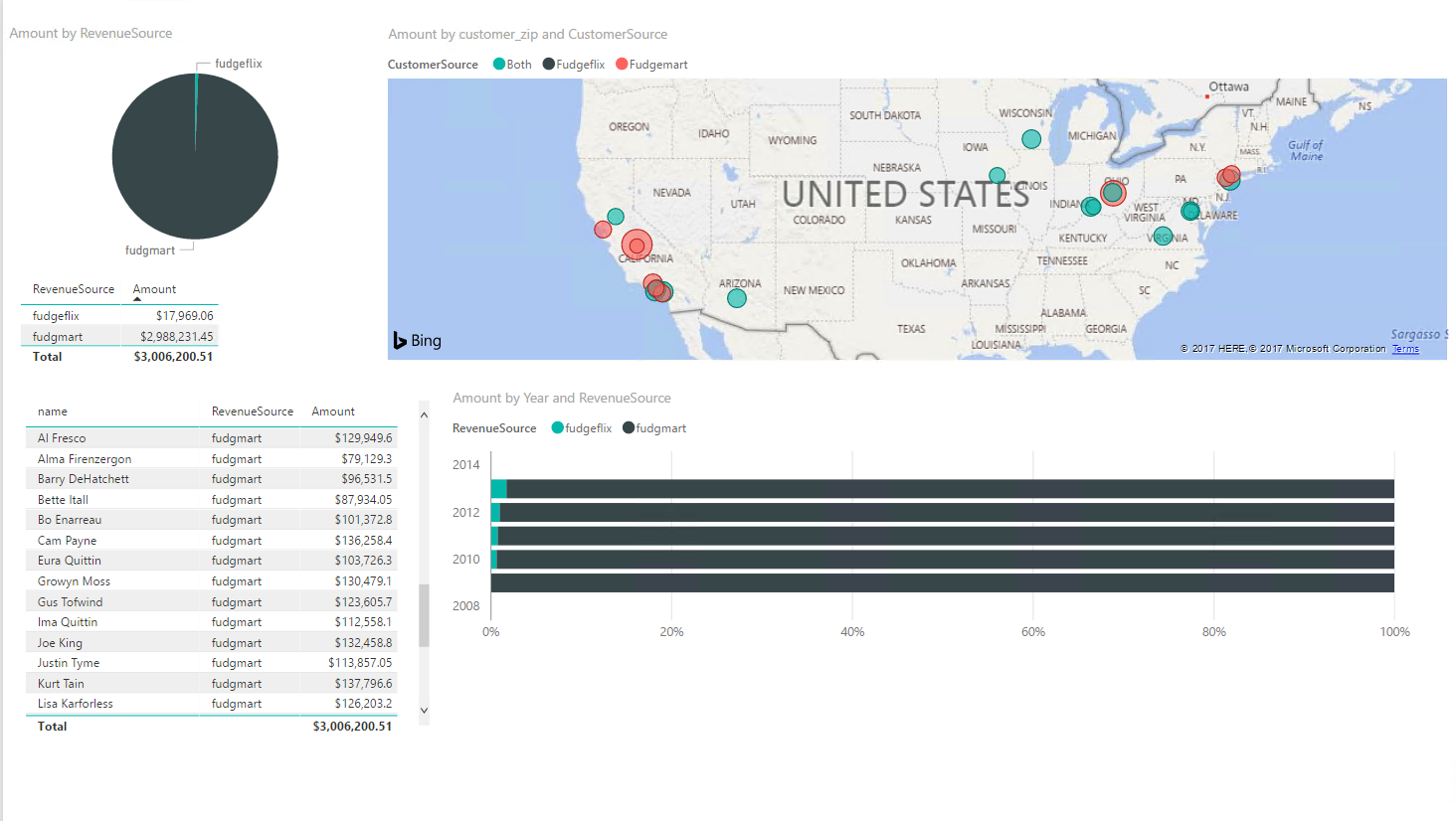
**Interactive PowerBI chart for Employee Timesheet business process:**

******



The dashboard below, created using PowerBI, is a reflection of our Revenue Cube created in SSAS. We originally wanted to be able to show revenue generated as it relates to time across the two subsidiaries of Fudgemart for comparative analysis, but we were able to show more concise audience segments. We can see, through this dashboard, that the subsidiary ‘Fudgemart’ takes in 99.4% of the total revenue generated by the conglomerate. Part of our business process was to be able to show whether customers are Fudgemart customers only, Fudgeflix customers only, or customers of both subsidiaries. This was nice, but we wanted to go to a deeper level, so we utilized a degenerate dimension in our fact table in order to show the source of a particular transaction, we called this dimension ‘Revenue Source’. We were also able to create a map based on the zip codes of our conformed customer dimension table. This step was trickier, as we had to create new columns in our target table with implemented logic to source them (Reference Derived columns in our ETL process). This map provides a visualization for unpenetrated markets, as well as density of our top performing regions for each subsidiary.

**Interactive PowerBI chart for Customer Revenue business process***:*



***PowerBI files (PBIX) are available at this link:***

https://drive.google.com/drive/folders/1vzPowyTBq3rSSPI75WXm3BdVy-K1Jbnx

# Conclusion:

The documentation of our project process noted throughout this paper allows for reproducible results and necessary explanation of process quirks evident throughout our work. The culmination of our work satisfies the original needs of our business processes and allows for even more granularity than we may have initially expected, which should leave the end users satisfied and occupied with analysis and potential findings. To eliminate congestion, we have displayed links within this document that point towards individualized tasks, such as the high level dimensional model, detailed dimension model, ETL packages, SSAS sln file and PowerBI dashboards.