Infibeam:
Data Warehouse Implementation

Group 9: Deeksha Paliwal, Jake Fabrizio, Pratyush Kulwal, Farees Patel, Sanchit Mehra

Spring 2017

School of Information Studies
Syracuse University

We create Magic!

Table of Contents

Infibeam

"Infibeam e-Commerce" is an online retail e-commerce website that sells digital and consumer electronics. Infibeam is a leading provider of technology products, services and solutions. The company offers expert service at an unbeatable price more than 1.5 billion times a year to the consumers, small business owners and educators who visit our e-commerce websites.

Infibeam provide the fundamental technology infrastructure and marketing reach to help merchants, brands and other businesses that provide products, services and digital content to leverage the power of the Internet to engage with their users and customers. Our businesses are comprised of core e-commerce where we also participate in the logistics and local services sectors.

Through innovation, we're creating a seamless experience to let customers shop anytime and anywhere online, through mobile devices and in stores. We are creating opportunities and bringing value to customers and communities around the globe.

Everyday Low Price (EDLP) is the cornerstone of our strategy, and our price focus has never been stronger. Today's customer seeks the convenience of one-stop shopping that we offer. From electronics and entertainments to sporting goods and crafts, we provide the deep assortment that our customers appreciate when they're shopping online at Infibeam.com

Since there is a huge competition in the online - retail sales market, the organization wants to maximize its profit by optimizing its Supply Chain, Sales Channels and Inventory Management. Infibeam wants to strategize their marketing campaign by targeting the right people that can yield in lucrative margins and help attain more user acquisition. Therefore, as a part of our strategic initiative, the organization has decided to implement a data warehouse project in order to generate insights, predict, and make recommendations out of the historical transactional data.

Project Scope

Need for a Data Warehouse

Currently, Infibeam has a user-interface from where the customer makes a purchase and a backend relational database system, where the data being stored in the form of a relational format in the traditional Online transactional processing system (OLTP) . Thereby, using this transactional data, a data warehouse can be built that could not only answer critical business questions based on historic data, but also enable the organization with Business Intelligence capabilities.

Business Questions

Our company is focussed on strengthening our marketing campaigns to drive value for the customers and the company. Hence, the initiative is segmented into two focus areas:

- Strategic Level Analysis
- Operational Level Analysis

More specifically, each focus area can be branched out for further detail analysis.

By implementing a data warehouse and BI, Infibeam hopes to be able to several business questions pertinent to marketing campaigns and customer habits.

1. **Strategic Level Analysis**

At the the strategic level, Infibeam wants to answer the following questions:

- Who will be the target age demographic for adoption of IoT?
- What should be the marketing strategy for targeting these customers?
- What is current cell-phone adoption and how much a consumer is willing to spend for latest smartphones?
- How much consumer will be willing to spend in future?

2. **Operational Level Analysis**

At the operational level, Infibeam wants to answer the following questions:

- Which products are sold most at Infibeam?
- Are customer satisfied with their purchase?
- Sales analysis based on age and gender demographics
- Which regions are most prominent for sales?

BI Team at Infibeam

The BI Team at Infibeam is one of the critical business units of the organization, which operates at an intersection of e-commerce, marketing and business strategy. The team supports all operational and strategic reporting for the organization.

The BI Team at Infibeam consists of highly-skilled individuals with numerous years of field experience in all facets of BI. Together, this team executed Infibeam's data warehouse project from the planning stages and business requirements definitions through all design phases and deployment.

Team Members and Roles

Below are the details of personnel and the specific role that each was assigned for the purpose of Infibeam's data warehouse implementation:

Team Member	**Role**
Deeksha Paliwal	Data Analyst
Pratyush Kulwal	Project Manager/ Data Architect
Sanchit Mehra	ETL Specialist
Jake Fabrizio	BI Engineer
Farees Patel	BI Engineer

Key Stakeholders

While this project is being carried out by the BI Team at Infibeam, they are not the only personnel with a vested interest or involvement in the project. The following is an additional list of stakeholders who were involved in various components:

- Database Administrators
- Data Architect
- End Users (registered and guest users)
- ETL Specialist
- Project Manager
- Data Analyst
- BI Engineer

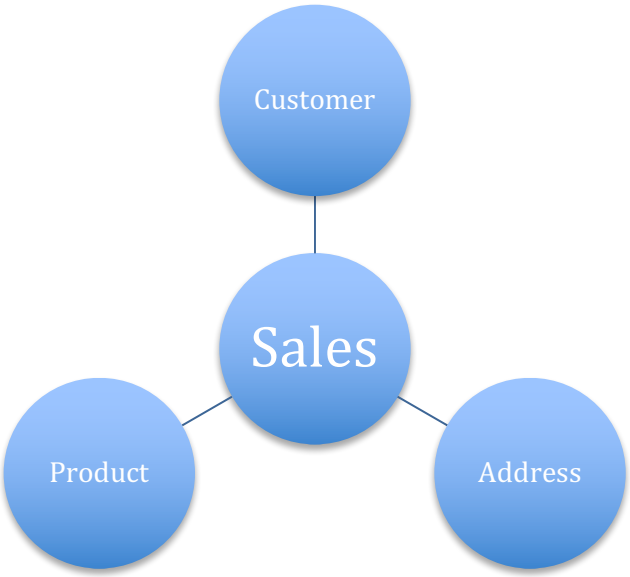
Data Warehouse Implementation

Bubble Chart

Included are two bubble charts which graphically summarizes the design of Infibeam's intended data warehouse business processes. These charts model data at the entity level and show the fact-dimension relationships for a given business process. The reason for including this tool is to facilitate discussion between technical and non-technical members of an organization and to show a streamlined view of the business processes.

The business processes that we are detailing are:

1. Sales Fact- Grain = 1 row per order detail



2. Product Rating Fact- Grain – 1 row per product rating

Bus Matrix

As part of the data warehouse logical design phase, we created a high level enterprise data warehouse bus diagram. The bus matrix provides detail about the business processes and respective fact tables. Fact grain type, granularity, facts, and dimensions are all included in the bus matrix. Bus matrices are key for planning and communicating across an organization. (Please see attached Excel file for complete Detailed Bus Matrix.)

Business Process Name	Fact Table	Fact Grain Type	Granulairty	Facts	Product	Address	Date	Customer	Product Review
Sales	SalesFact	Transaction	one row per order detail	ProductQuantity, SoldAmount, OrderDiscount	x	x	x		
Product Rating	ProductRatingFact	Transaction	one row per reviw	VerifiedCustomer	x		x	x	x

Attribute List

Please see "Appendix B: Attribute List" for a screenshot of our tables and their attributes.

Issue List

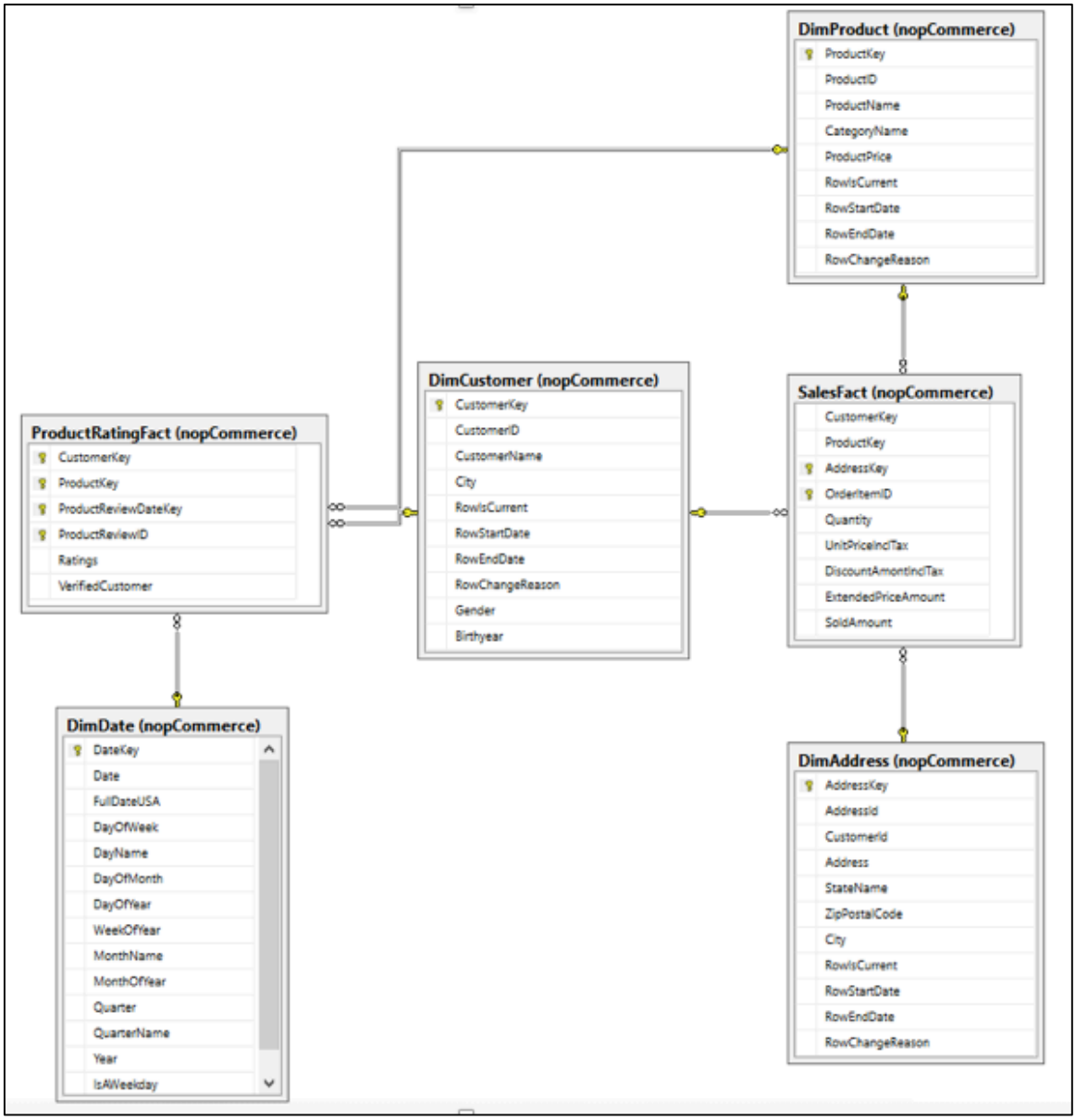
Below is a list of issues we encountered and resolved over the course of the project implementation:

Issue #	Task / Topic	Issue	Identified Date	Reported By	Respon-sible	Status	Priority	Date Closed
1	Dim	No attribute of 'gender' in OLTP system	2/26/2017	2/26/2017	OLTP system	Resolved	H	3/2/2017
2	Dim	No attribute of 'Age' or 'DOB' in OLTP system	3/7/2017	3/7/2017	OLTP system	Resolved	H	3/10/2017
3	SSIS	In SSIS, it was difficult to look up in one go	3/25/2017	3/25/2017	Team members	Resolved	L	3/28/2017
4	Dim	Customer demographics had to be imported using flat files	4/5/2017	4/5/2017	Team members	Resolved	L	4/10/2017
5	Dim	Database creates multiple record for the same customer in the customer table	4/5/2017	4/5/2017	OLTP system	Resolved	H	4/10/2017
6	Fact	Type mismatch error was resolved using CAST statements	4/15/2017	4/15/2017	Team members	Resolved	L	4/20/2017
7	SSAS	SSAS was not connecting to the database	4/20/2017	4/20/2017	SSAS	Resolved	H	4/23/2017

Data Warehouse Integration

Star Schema

The Star Schema, generated by SQL Server, is a dimensional model that shows the relationships between the business process fact and dimension tables. This visual representation shows each table within the data warehouse and its respective attributes, offering more detail than a high-level design.



ETL Specifications

Provided is a high-level source to target map of our ETL processes (facts and dimensions) to show the basic design flow:

- Customer Dimension:

		Target							Source				
Column Name	Display Name	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
CustomerKey	CustomerKey	int			PK ID		N		Derived				
CustomerID	CustomerID	int					N		group9-nopCommerce	dbo	Customers	CustomerID	int
CustomerName	CustomerName	nvarchar	201				N		group9-nopCommerce	dbo	Address	FirstName, LastName	nvarchar
City	City	nvarchar	100				Y		group9-nopCommerce	dbo	Address	City	nvarchar
Gender	Gender	nvarchar	200				Y		External Source	csv	External flat file	External flat file	nvarchar
Birthyear	Birthyear	int					Y		External Source	csv	External flat file	External flat file	int
RowIsCurrent	Row Is Current	bit					N	1	Derived			RowIsCurrent	
RowStartDate	Row Start Date	datetime					N	1/1/1900	Derived			RowStartDate	
RowEndDate	Row End Date	datetime					N	12/31/9999	Derived			RowEndDate	
RowChangeReason	Row Change Reason	nvarchar	200				Y		Derived			RowChangeReason	

- Product Dimension:

		Target							Source				
Column Name	Display Name	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
ProductKey	ProductKey	int			PK ID		N		Derived				
ProductID	ProductID	int					N		group9-nopCommerce	dbo	Product	ProductID	int
ProductName	ProductName	nvarchar	400				N		group9-nopCommerce	dbo	Product	Name	nvarchar(400)
CategoryName	CategoryName	nvarchar	400				N		group9-nopCommerce	dbo	Categories	Name	nvarchar(400)
ProductPrice	ProductPrice	decimal	18,4				N		group9-nopCommerce	dbo	Product	UnitPrice	decimal(18,4)
RowIsCurrent	Row Is Current	bit					N	1	Derived				
RowStartDate	Row Start Date	datetime					N	1/1/1900	Derived				
RowEndDate	Row End Date	datetime					N	12/31/9999	Derived				
RowChangeReason	Row Change Reason	nvarchar	200				Y		Derived				

- Address Dimension:

		Target							Source				
Column Name	Display Name	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
AddressKey	AddressKey	int			PK ID		N		Derived				
AddressId	AddressId	int					N		group9-nopCommerce	dbo	Address	Id	int
CustomerId	CustomerId	int					N		group9-nopCommerce	dbo	CustomerAddresses	Customer_Id	int
Address	Address	nvarchar	50				N		group9-nopCommerce	dbo	Address	Address1	nvarchar(max)
StateName	StateName	nvarchar	100				N		group9-nopCommerce	dbo	StateProvince	Name	nvarchar(100)
ZipPostalCode	ZipPostalCode	nvarchar	50				N		group9-nopCommerce	dbo	Address	ZipPostalCode	nvarchar(max)
City	City	nvarchar	50				Y		group9-nopCommerce	dbo	Address	City	nvarchar(max)
RowIsCurrent	Row Is Current	bit					N	1	Derived				
RowStartDate	Row Start Date	datetime					N	1/1/2000	Derived				
RowEndDate	Row End Date	datetime					N	12/31/1999	Derived				
RowChangeReason	Row Change Reason	nvarchar	200				Y		Derived				

- Date Dimension:

		Target							Source				
Column Name	Display Name	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
DateKey	DateKey	int			PK		N		External Sources			External Sources	int
Date	Date	datetime					Y		External Sources	dbo		External Sources	datetime
FullDateUSA	FullDateUSA	nchar	11				N		External Sources	dbo		External Sources	nchar
DayOfWeek	DayOfWeek	tinyint					N		External Sources	dbo		External Sources	tinyint
DayName	DayName	nchar	10				N		External Sources	dbo		External Sources	nchar
DayOfMonth	DayOfMonth	tinyint					N		External Sources	dbo		External Sources	tinyint
DayOfYear	DayOfYear	int					N		External Sources	dbo		External Sources	int
WeekOfYear	WeekOfYear	tinyint					N		External Sources	dbo		External Sources	tinyint
MonthName	MonthName	nchar	10				Y		External Sources	dbo		External Sources	nchar
MonthOfYear	MonthOfYear	tinyint					N		External Sources	dbo		External Sources	tinyint
Quarter	Quarter	tinyint					N		External Sources	dbo		External Sources	tinyint
QuarterName	QuarterName	nchar	10				Y		External Sources	dbo		External Sources	nchar
Year	Year	int					N		External Sources	dbo		External Sources	int
IsWeekday	IsWeekday	varchar	1				N	0	External Sources	dbo		External Sources	varchar

- Sales Fact:

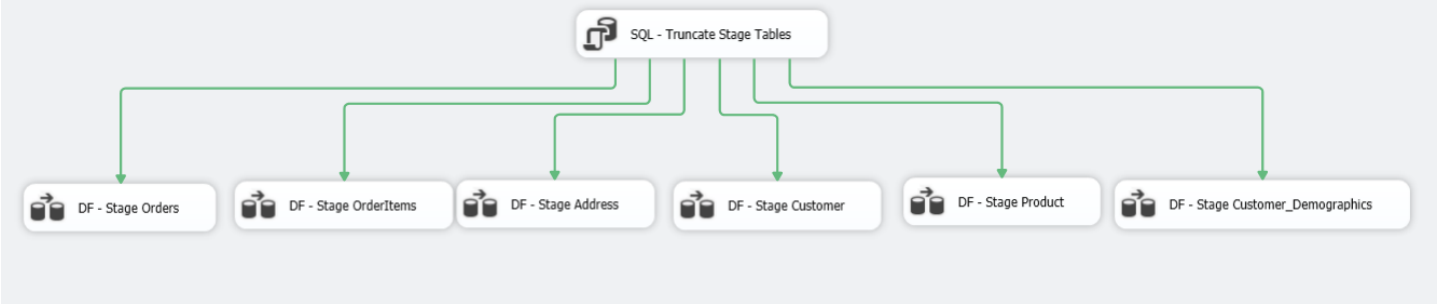
		Target							Source				
Column Name	Display Name	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
CustomerKey	CustomerKey	int			FK	DimCustomer.CustomerKey	N		ist722_group9_dw	dbo	DimCustomer	CustomerKey	int
ProductKey	ProductKey	int			FK	DimProduct.ProductKey	N		ist722_group9_dw	dbo	DimProduct	ProductKey	int
AddressKey	BillingAddress	int			PK, FK	DimAddress.AddressID	N		ist722_group9_dw	dbo	DimAddress	AddressID	int
OrderItemID	OrderItemID	int			PK	OrderItem.Id	N		group9-nopCommerce	dbo	OrderItem	Id	int
Quantity	Quantity	int					N		group9-nopCommerce	dbo	OrderItem	Quantity	int
UnitPriceInclTax	UnitPriceInclTax	decimal	(18,4)				N		group9-nopCommerce	dbo	OrderItem	UnitPriceInclTax	decimal (18,4)
DiscountAmountInclTax	DiscountAmountInclTax	decimal	(18,4)				N	0	group9-nopCommerce	dbo	OrderItem	DiscountAmountInclTax	decimal (18,4)
ExtendedPriceAmount	ExtendedPriceAmount	decimal	(18,4)				N		Derived	dbo			
SoldAmount	Sold Amount	decimal	(18,4)				N		Derived	dbo			

- Product Rating Fact

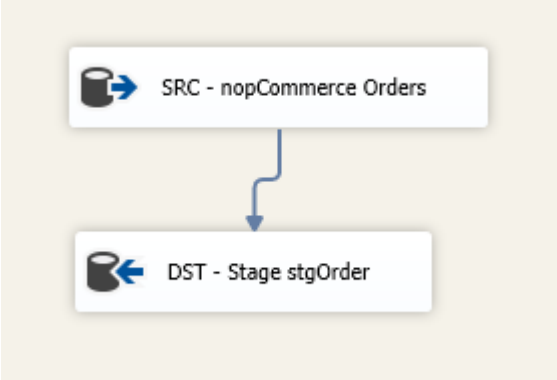
		Target							Source				
Column Name	Display Name	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
CustomerKey	CustomerKey	int			FK, PK	CustomerDimension.Custom	N		ist722_group9_dw	dbo	DimCustomer	CustomerKey	int
ProductKey	ProductKey	int			FK, PK	ProductDimension.ProductKe	N		ist722_group9_dw	dbo	DimProduct	ProductKey	int
ProductReviewDateKey	ProductReviewDateKey	int			FK, PK	DimDate.DateKey	N		ist722_group9_dw	dbo	DateKey	DateKey	int
ProductReviewID	Product Review ID	int			PK		N		[group9-nopCommerce]	dbo	ProductReview	Id	int
Ratings	Ratings	int					N		[group9-nopCommerce]	dbo	ProductReview	Rating	int
VerifiedCustomer	Verified Customer	int					N		Derived				int

Detailed ETL Flow for Business Processes

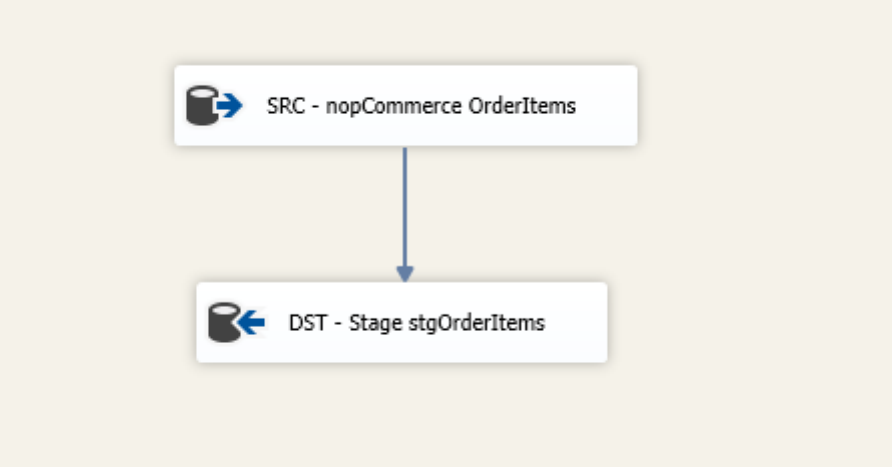
1. Control flow for extracting dimensions from OLTP to Staging databsase



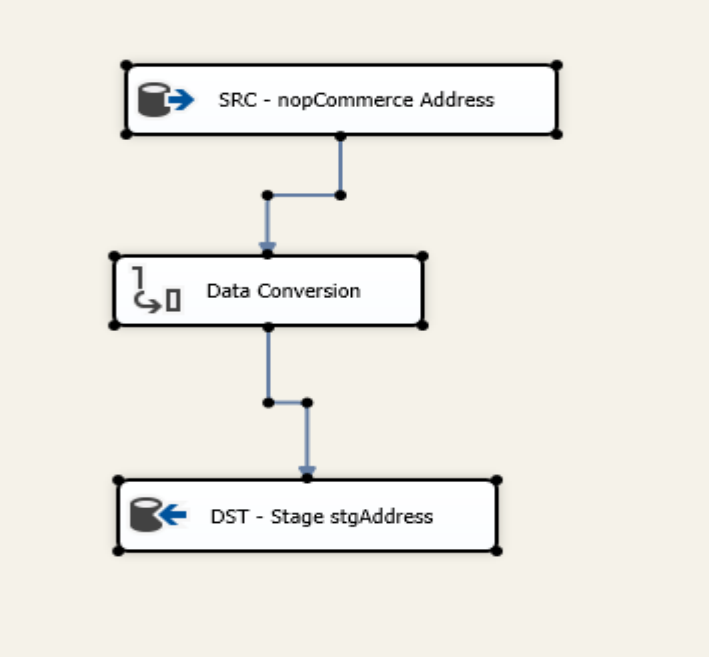
1.1 Data flow for extracting Order table from OLTP to Staging environment



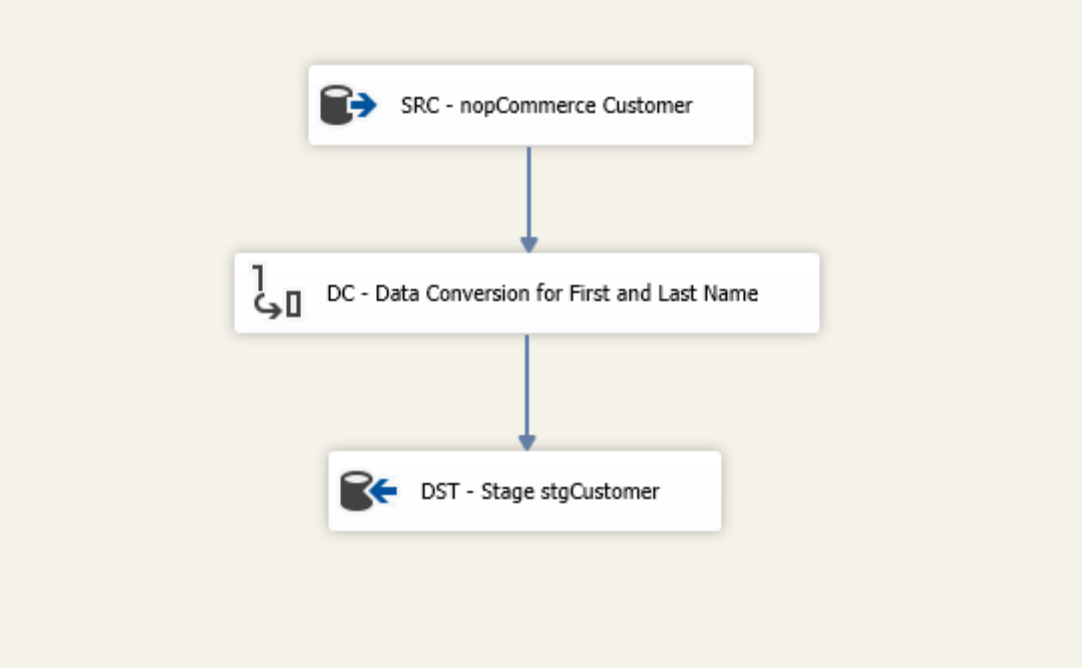
1.2 Data flow for extracting OrderItems table from OLTP to Staging environment



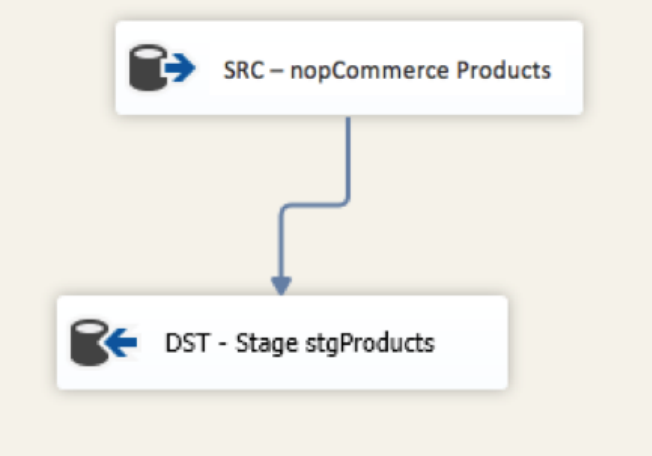
1.3 Data flow for extracting Address table from OLTP to Staging environment



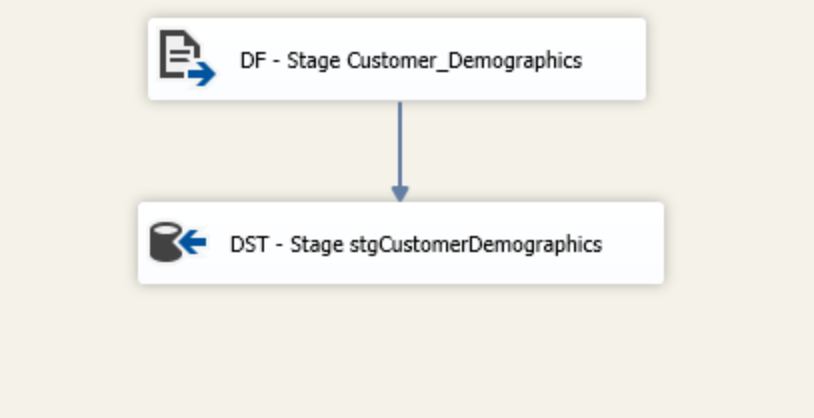
1.4 Data flow for extracting Customer table from OLTP to Staging environment and combining the First and last name of the customers



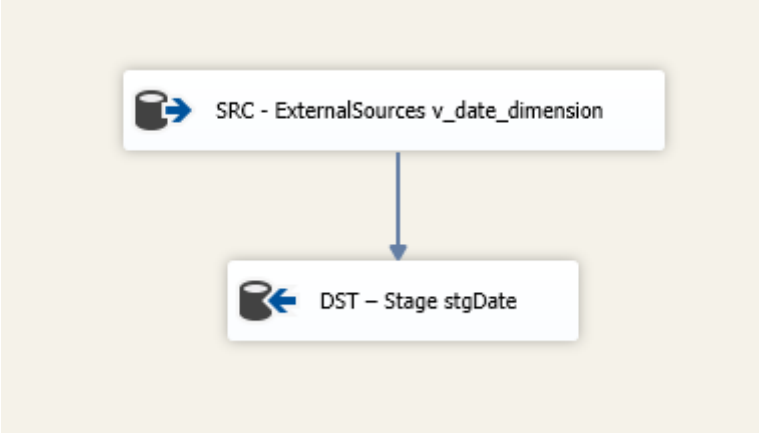
1.5 Data flow for extracting Product table from OLTP to Staging environment



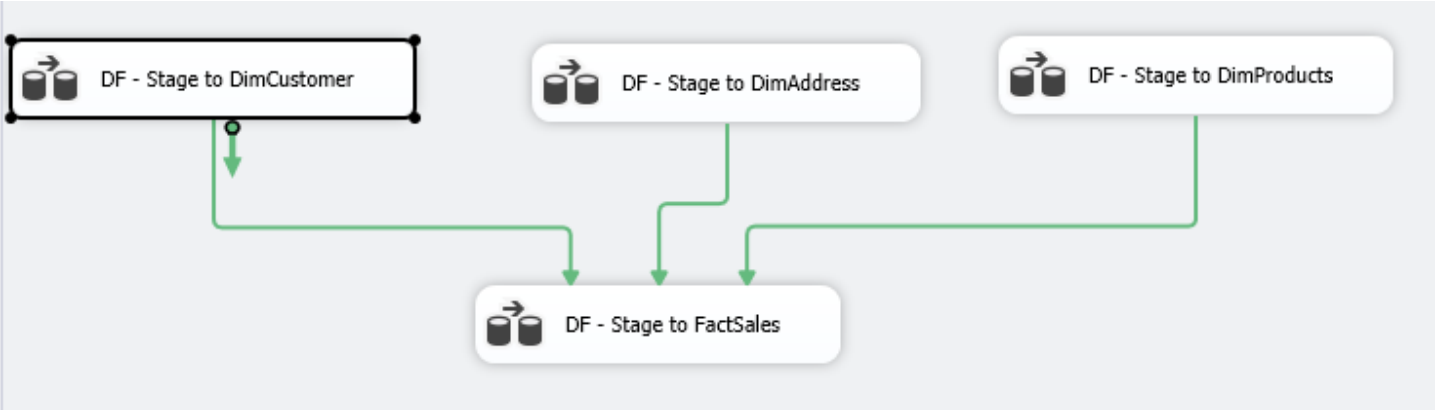
1.6 Data flow for extracting Customer demographics information from flat files to the Staging environment



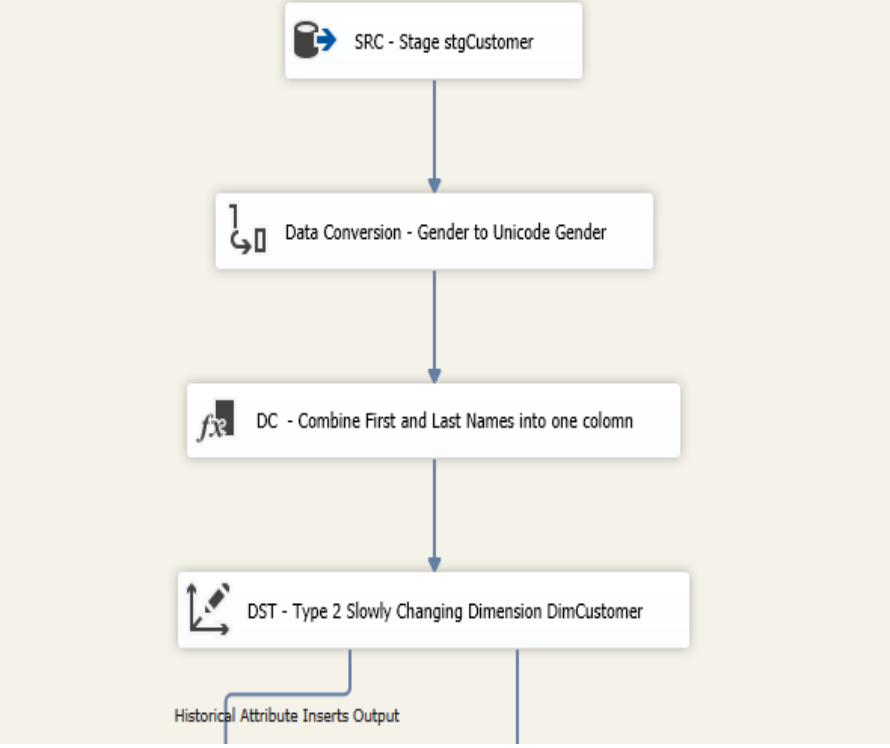
1.7 Data flow for extracting date table from External Sources to the Staging environment

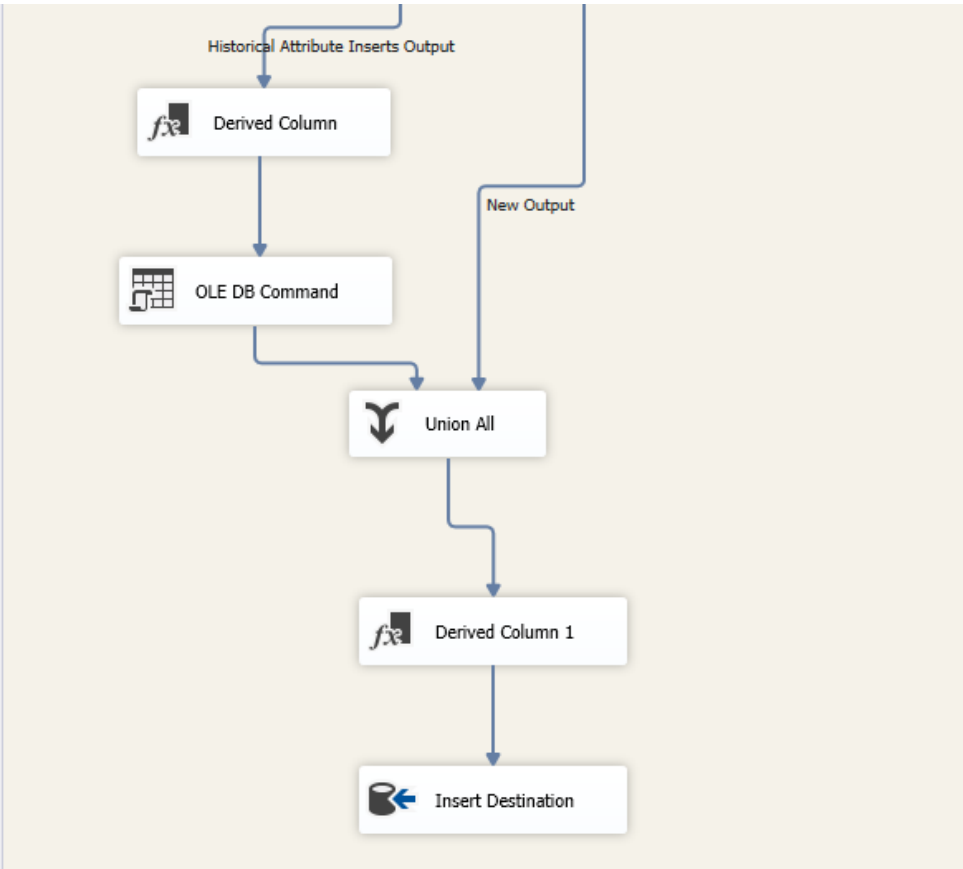


2. Control flow for extracting dimensions from Staging database to the data warehouse for Sales Fact table

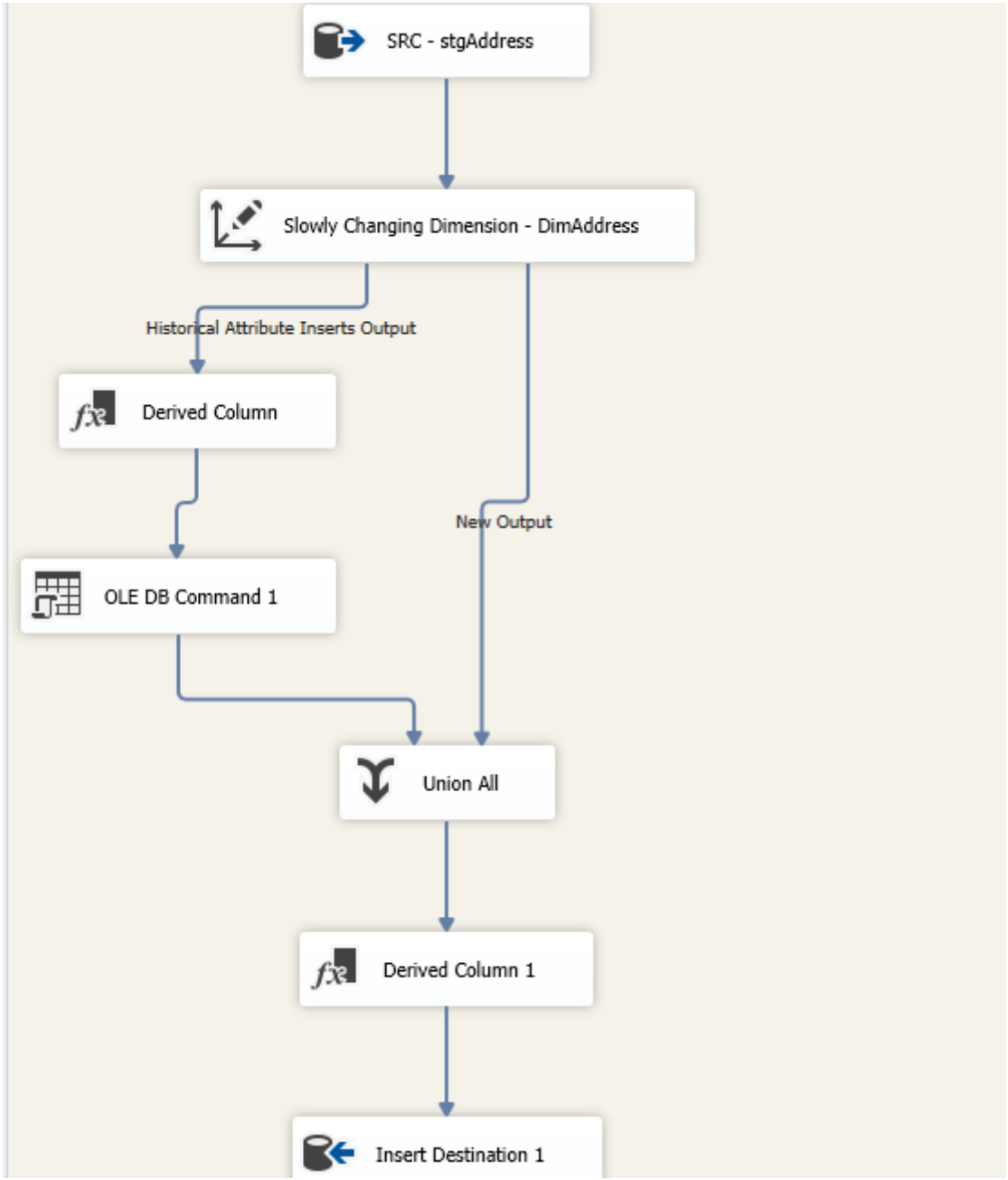


2.1 Data flow for extracting Customer table from Staging environment to the DW

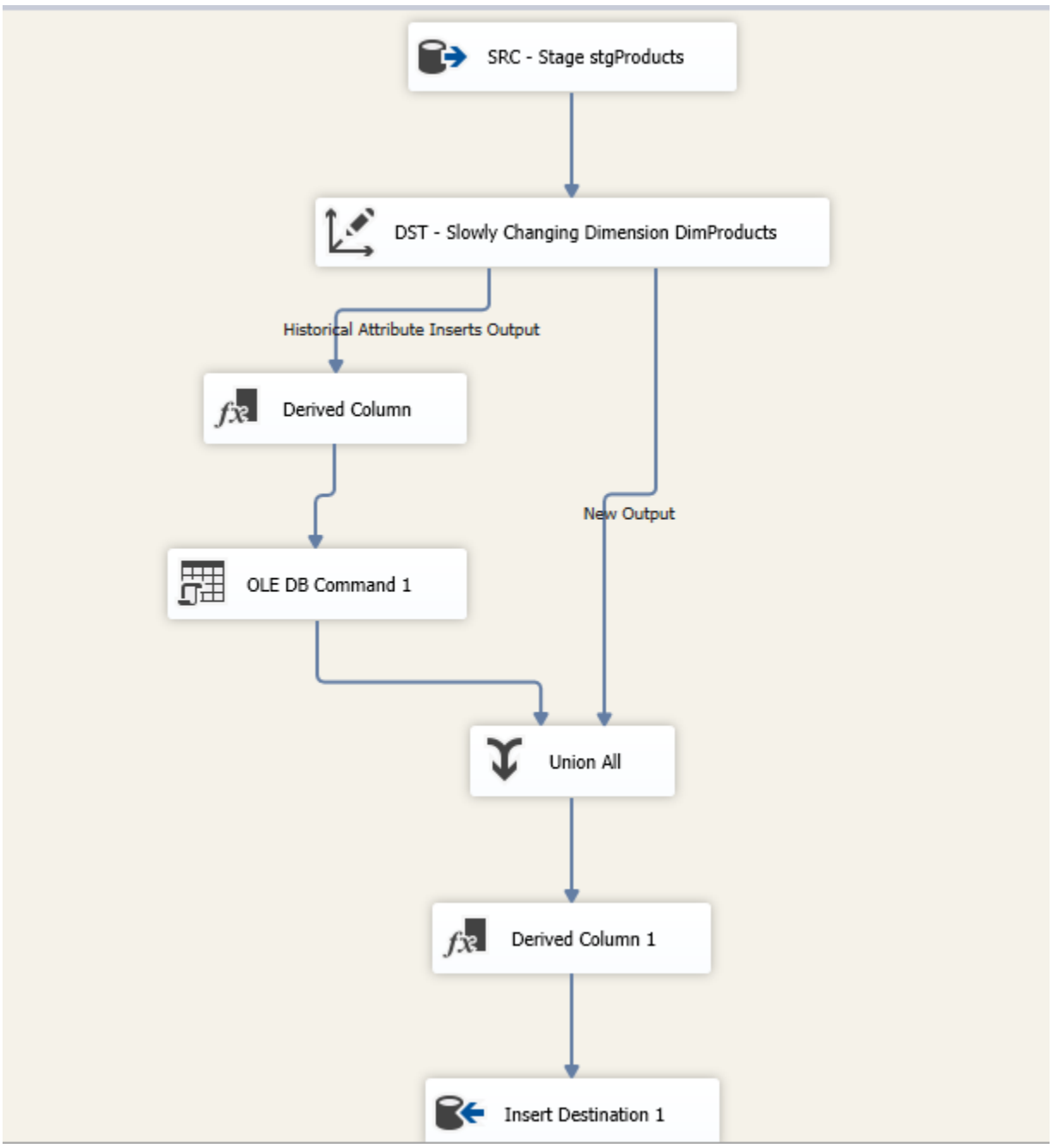




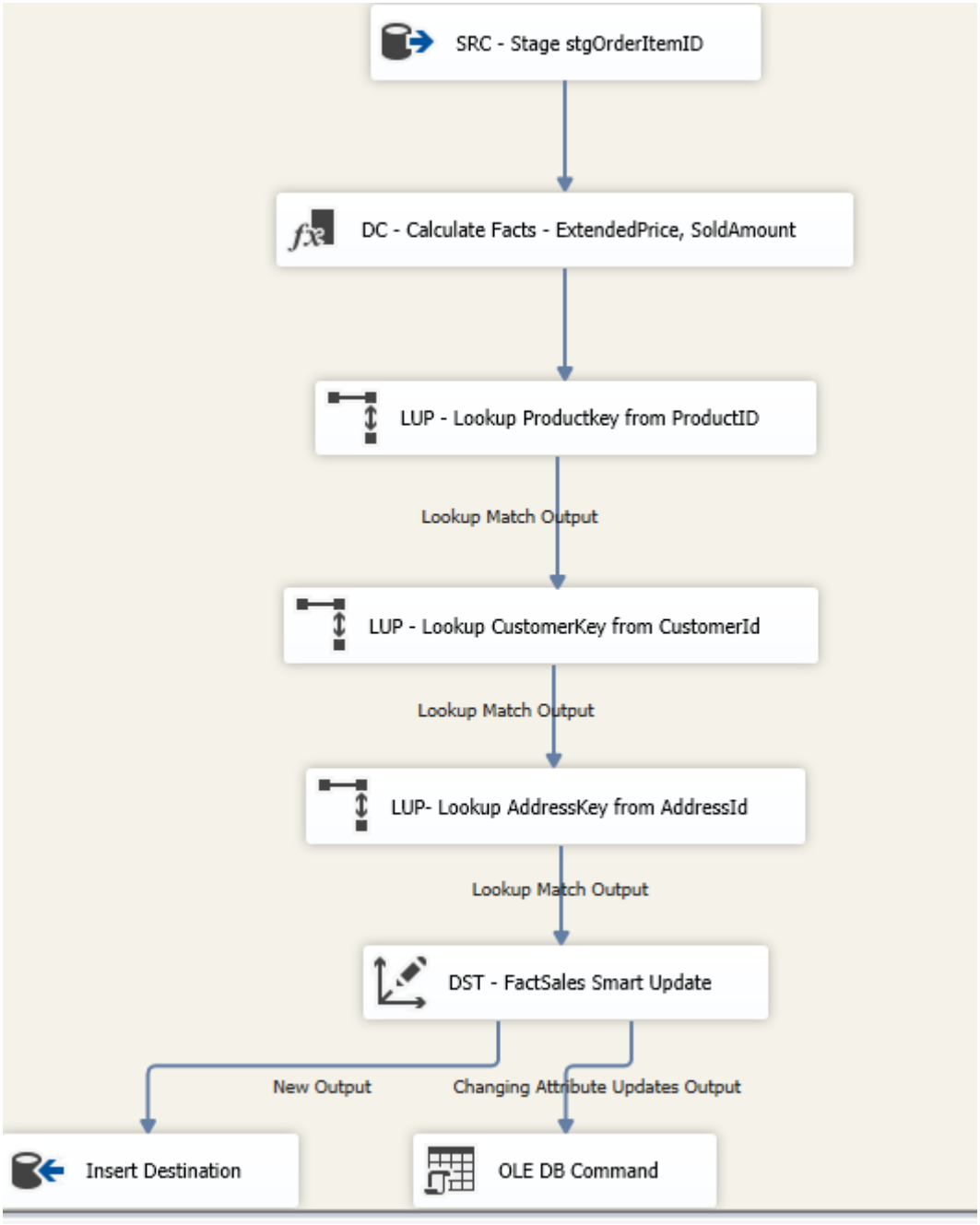
2.2 Data flow for extracting Address table from Staging environment to the DW



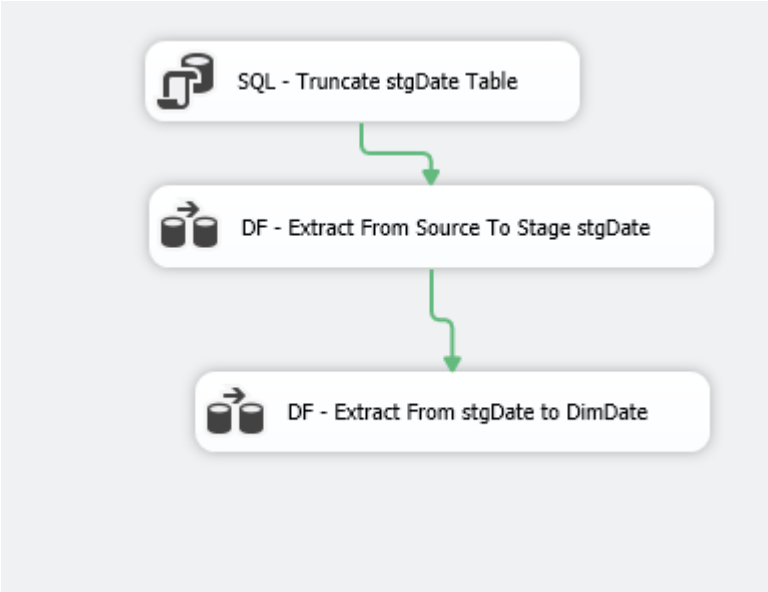
2.3 Data flow for extracting Product table from Staging environment to the DW



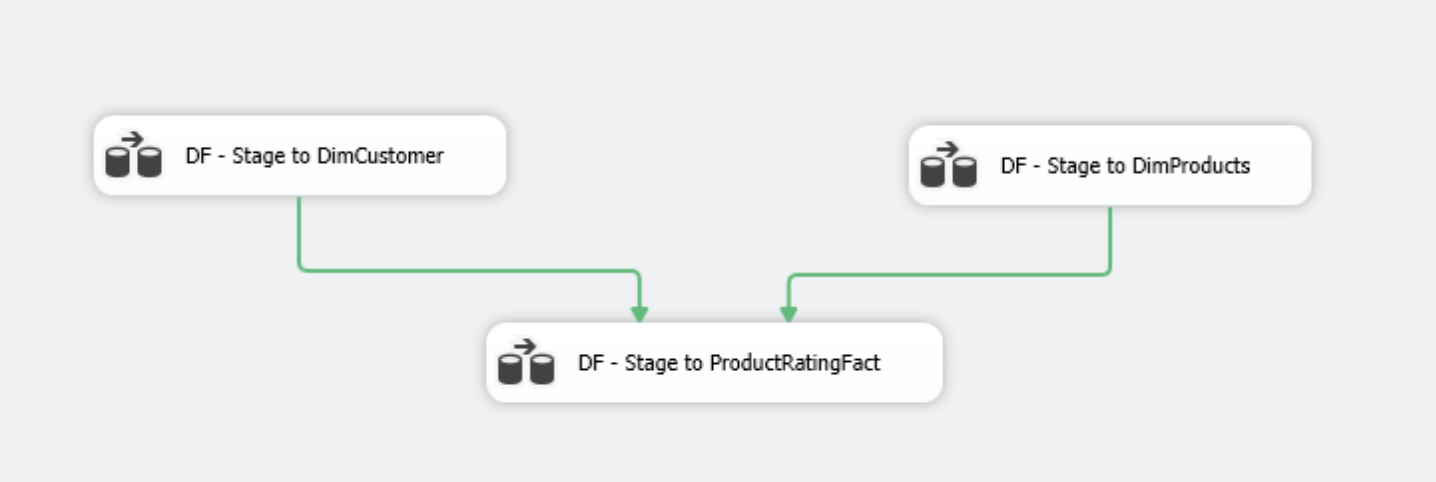
2.4 Data flow for extracting OrderItemID table from Staging environment and preparing the Fact table in DW



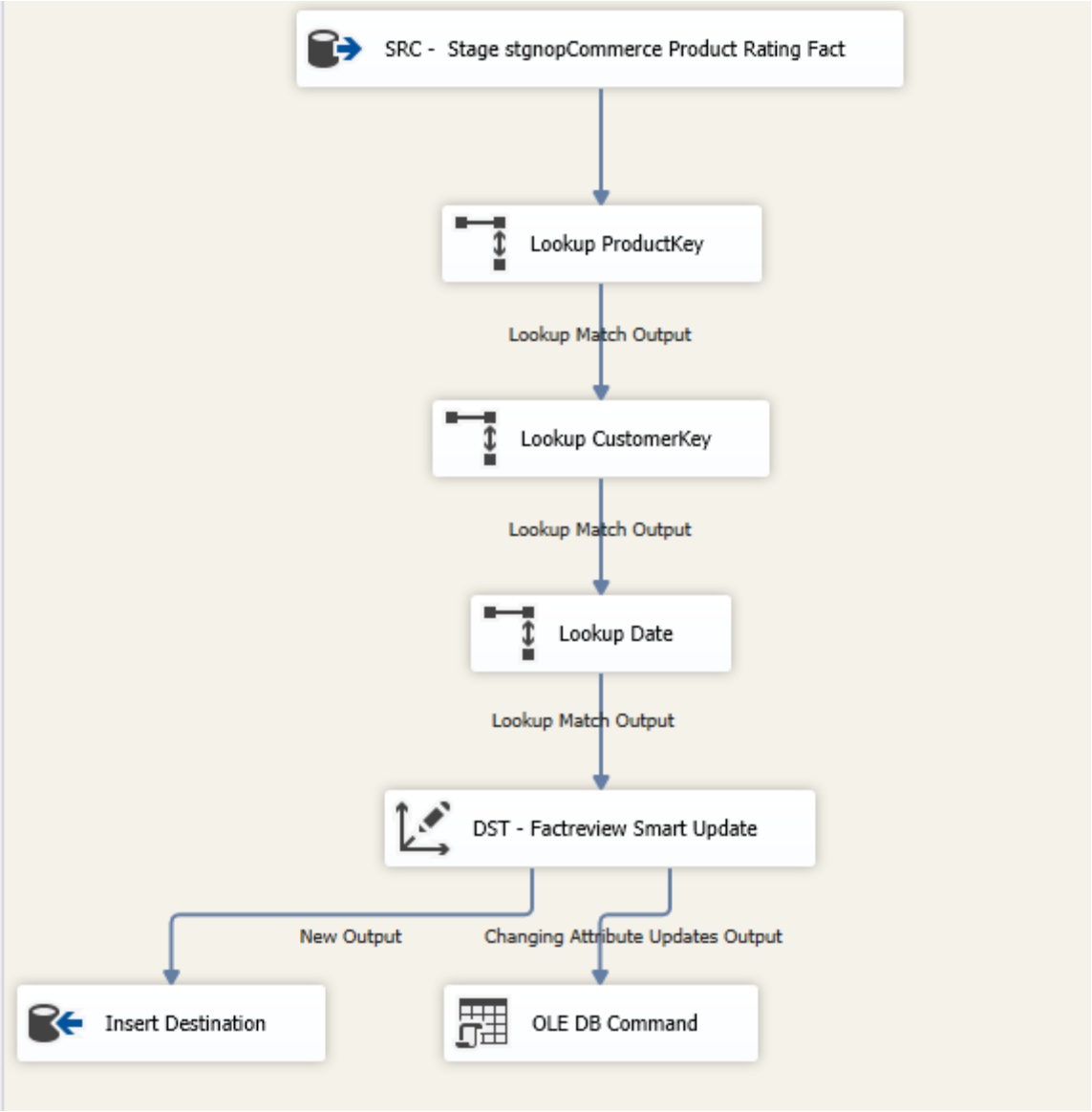
2.5 Data flow for extracting Date table from Staging environment to the DW



3. Control flow for extracting dimensions from Staging database to the data warehouse for Product Rating Fact table



3.1 Data flow for creating Product Rating fact table from in the DW using conformed dimension



Dimensional Hierarchies

Provided below is a screenshot of our Excel pivot table which shows dimensional hierarchy of products:

3	Product Categories	Sold Amount
4	⊞ **Books**	**129**
5	⊞ **Camera & photo**	**12540**
6	⊟ **Cell phones**	**79403.43**
7	Apple - Pre-Owned iPhone 5s 4G LTE with 16GB Memory Cell Phone (Unlocked) - Silver	10059.43
8	AT&T Classic Phone	775
9	G - V20 with 64GB Memory Cell Phone - Titan Gray	285
10	HTC One M8 Android L 5.0 Lollipop	37240
11	HTC One Mini Blue	7100
12	Huawei Mate 9	11700
13	Nokia Lumia 1020	8376
14	Refurbished Apple iPhone 4 16GB	1068
15	Samsung Galaxy 8	2800
16	⊞ **Clothing**	**73.5**
17	⊞ **Desktops**	**16745**
18	⊞ **Digital downloads**	**85**
19	⊞ **Gift Cards**	**575**
20	⊞ **Jewelry**	**2460**
21	⊟ **Notebooks**	**57470**
22	Apple MacBook Pro 13-inch	50400
23	Asus N551JK-XO076H Laptop	3000
24	HP Spectre XT Pro UltraBook	1350
25	Lenovo Thinkpad X1 Carbon Laptop	2720
26	⊞ **Others**	**5277.38**
27	⊞ **Shoes**	**278.04**
28	⊞ **Software**	**1844.99**
29	**Grand Total**	**176881.34**

Appendix A: Team Member Contribution Report

The team worked effectively together on most of the project assignment to completion. However, some team members' strengths were positioned in other areas than others. While there was equal contribution, various members specialized in what they were most proficient. Deeksha and Sanchit were an invaluable technical members spearheading most of the SSIS (ETL) and SSAS portions of the project. Jake oversaw data management and the development of business questions, including making transactions on the Infibeam website, and managing the backend such as adding products and inventory. He also contributed a great deal to the compilation of the final report. Farees also was a key member in the compilation of the report, as well as preparing, editing and finalizing the high level and detail level dimensional models essential for the planning and preparation of the basis of our project. He assisted with data management and saw that the scope of the project was reasonable given time constraints and the data and technology that we had. Sanchit and Pratyush put the most work into the physical design and construction of the presentation for class. They led the way in the implementation and analysis of the BI information generated. Additionally, Pratyush served essentially as the project manager, coordinating that all of us were able to both work together and complete necessary individual contributions when we were unable to meet all together. He also was key in SSIS and in utilizing Tableau software.

Overall, the team acted cohesively as a unit, and all members were generally present for most steps of the project, once we finally gained more concise direction following the midterm report. We worked off of each other's strength, and we believe that our blend of technical expertise and soft skills were the perfect ingredient to a successful team project. Each of us had the opportunity to enrich our skills and learn about the different components of implementing a data warehouse and business intelligence. We were especially grateful for the opportunity to work with team members that were able to fill in our knowledge gaps and offer us a better holistic understanding of course topics, skills, and technology.

Appendix B: Attribute List

Instructions!

Dimension / Fact Table	Attribute / Fact Name	Description	Alternate Names	Sample Values
SalesFact	CustomerKey	Key to Customer table		1,2,3..
SalesFact	ProductKey	Key to the Product table		1,2,3..
SalesFact	AddressKey	Key to Address table		1,2,3..
SalesFact	OrderItemID	Degenerate key		1,2,3..
SalesFact	Quantity	Quantity of the product		1,2,3..
SalesFact	UnitPriceInclTax	Price per product		100
SalesFact	DiscountAmountInclTax	Discount for a specific product		20
SalesFact	ExtendedPriceAmount	UnitPriceInclTax * Quantity		80
SalesFact	SoldAmount	ExtendedPriceAmount - (Quantity * DiscountAmountInclTax)		100
ProductRatingFact	CustomerKey	Key to Customer table		1, 2, 3
ProductRatingFact	ProductKey	Key to the Product table		1, 2, 3
ProductRatingFact	ProductReviewDateKey	Key to Date (for Product)		20120108
ProductRatingFact	ProductReviewID	ID to Product Review		34.35
ProductRatingFact	Ratings	Rating of the Product		1,2,3,4,5
ProductRatingFact	VerifiedCustomer	Indicates whether the review has been made a person who brought the product		0,1
DimDate	DateKey	Surrogate primary key		20041123
DimDate	Date	Full date as a SQL date		38314
DimDate	FullDateUSA	String expression of the full date, eg MM/DD/YYYY		23-Nov-2004
DimDate	DayOfWeek	Number of the day of week; Sunday = 1		1..7
DimDate	DayName	Day name of week, eg Monday		Sunday
DimDate	DayOfMonth	Number of the day in the month		1..31
DimDate	DayOfYear	Number of the day in the year		1..365
DimDate	WeekOfYear	Week of year, 1..53		1..52 or 53
DimDate	MonthName	Month name, eg January		November
DimDate	MonthOfYear	Month of year, 1..12		1, 2, ..., 12
DimDate	Quarter	Calendar quarter, 1..4		1, 2, 3, 4
DimDate	QuarterName	Quarter name eg. First		November
DimDate	Year	Calendar year, eg 2010		2004
DimDate	IsWeekday	Is today a weekday		1, 0
DimAddress	AddressKey	Surrogate primary key		1, 2, 3...
DimAddress	AddressId	Business key from source system (aka natural key)		1,25,26
DimAddress	CustomerId	Uniquely identifies Customer from source system		
DimAddress	Address	Address description		838 Westcott Street
DimAddress	StateName	Name of the State		New York
DimAddress	ZipPostalCode	Postal Code details		13210
DimAddress	City	City Name		Syracuse
DimAddress	RowIsCurrent	Is this the current row for this member (Y/N)?		TRUE, FALSE
DimAddress	RowStartDate	When did this row become valid for this member?		40567
DimAddress	RowEndDate	When did this row become invalid? (12/31/9999 if current row)		1/14/1998, 12/31/9999
DimAddress	RowChangeReason	Why did the row change last?		

DimProduct	ProductKey	Surrogate primary key		1, 2, 3...
DimProduct	ProductID	Business key from source system (aka natural key)		1,2,3,...
DimProduct	ProductName	Name of product		Laptop
DimProduct	CategoryName	Product type/category		Desktops
DimProduct	ProductPrice	The Price of the product		20
DimProduct	RowIsCurrent	Is this the current row for this member (Y/N)?		TRUE, FALSE
DimProduct	RowStartDate	When did this row become valid for this member?		40567
DimProduct	RowEndDate	When did this row become invalid? (12/31/9999 if current row)		1/14/1998, 12/31/9999
DimProduct	RowChangeReason	Why did the row change last?		
DimCustomer	CustomerKey	Surrogate primary key		1, 2, 3...
DimCustomer	CustomerID	Business key from source system (aka natural key)		1, 2, 3...
DimCustomer	CustomerName	Name of the Customer		John Smith
DimCustomer	City	Customer's City		Boston
DimCustomer	Gender	Gender of the custmer		Male
DimCustomer	Birthyear	Birth year of the customer		1992
DimCustomer	RowIsCurrent	Is this the current row for this member (Y/N)?		TRUE, FALSE
DimCustomer	RowStartDate	When did this row become valid for this member?		40567
DimCustomer	RowEndDate	When did this row become invalid? (12/31/9999 if current row)		1/14/1998, 12/31/9999
DimCustomer	RowChangeReason	Why did the row change last?		