INFO 7290
Data Warehousing & Business Intelligence

Group 03

Yixuan Feng
Yuwei Hou
Linduo Li

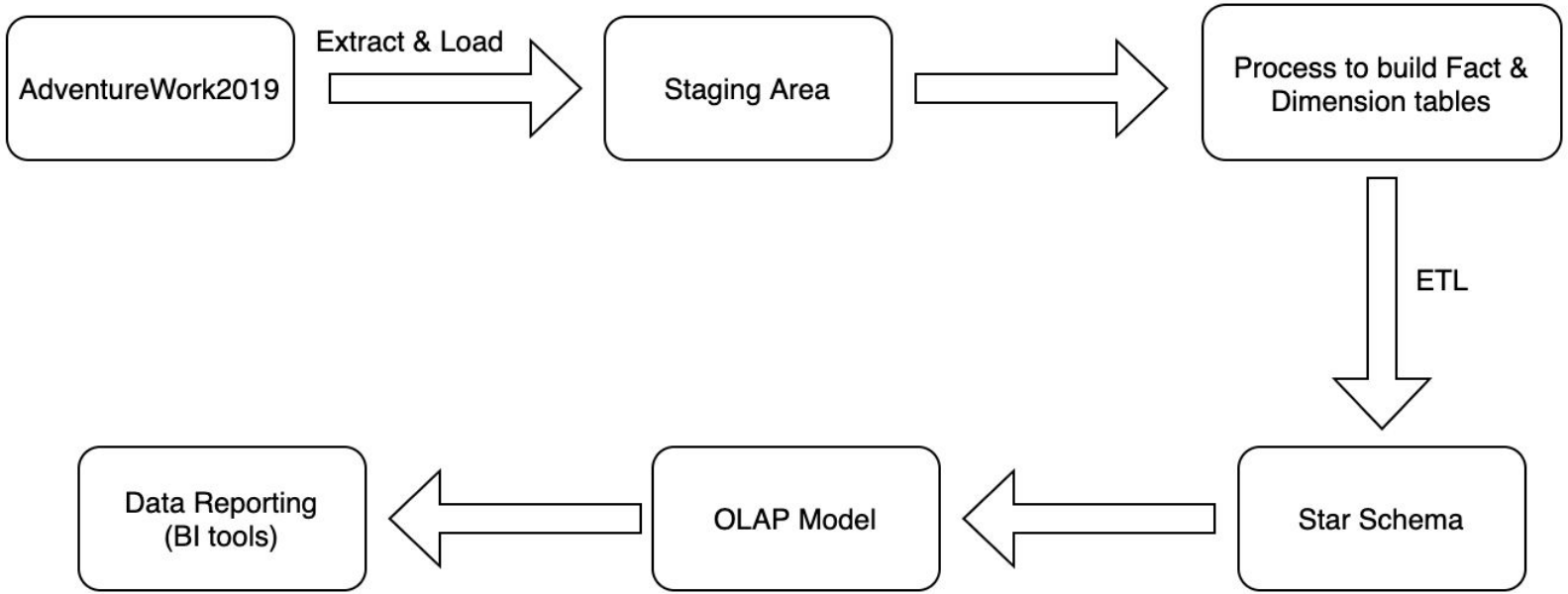
Introduction

- Create dimensional model for DW that will support:
  Sales analysis
  Inventory analysis

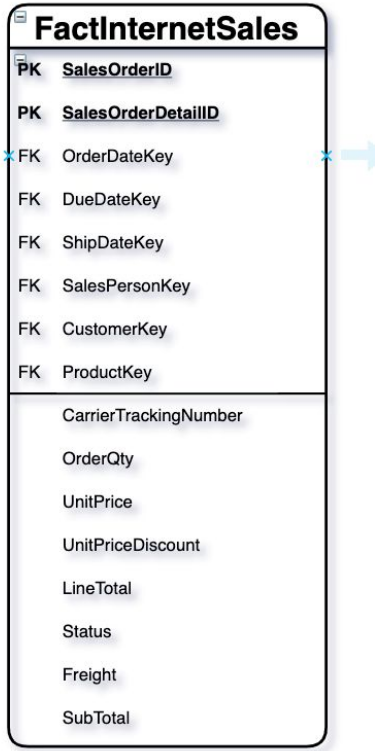
- DI: Load data sources into DW
  SOR: AdventureWorks 2019
  Tool: MS SQL Server (SSIS)

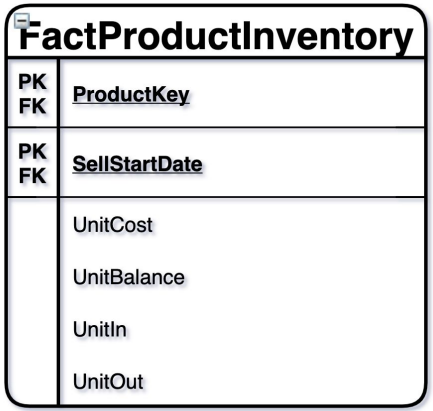
- BI: Develop visualization dashboards & reports
  Tool: R

Data Flow Process

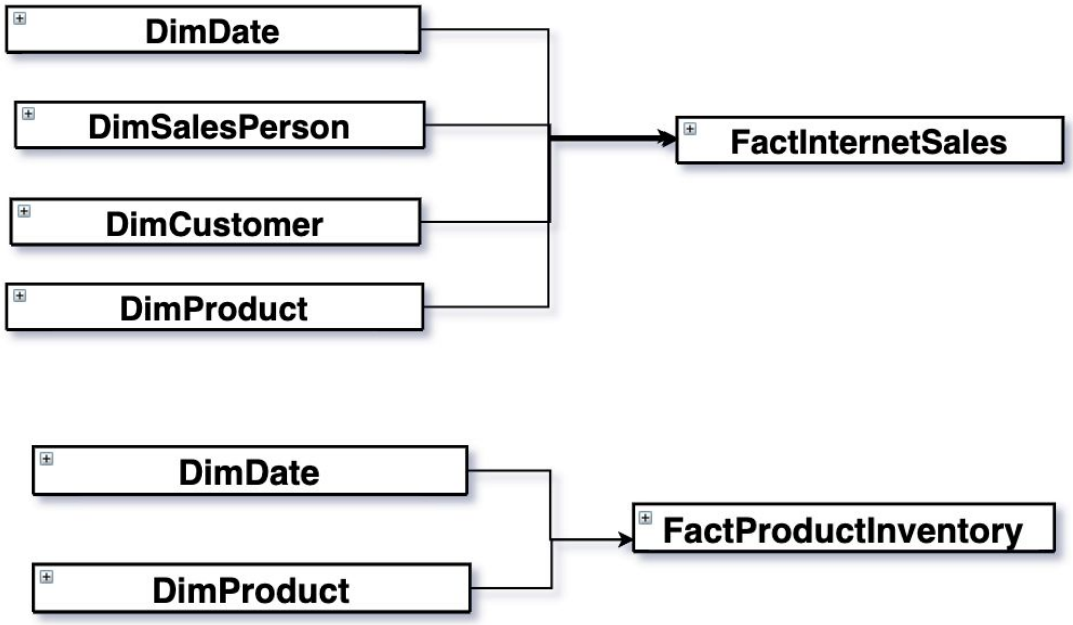


Data Warehouse Model ---- Facts and Dims

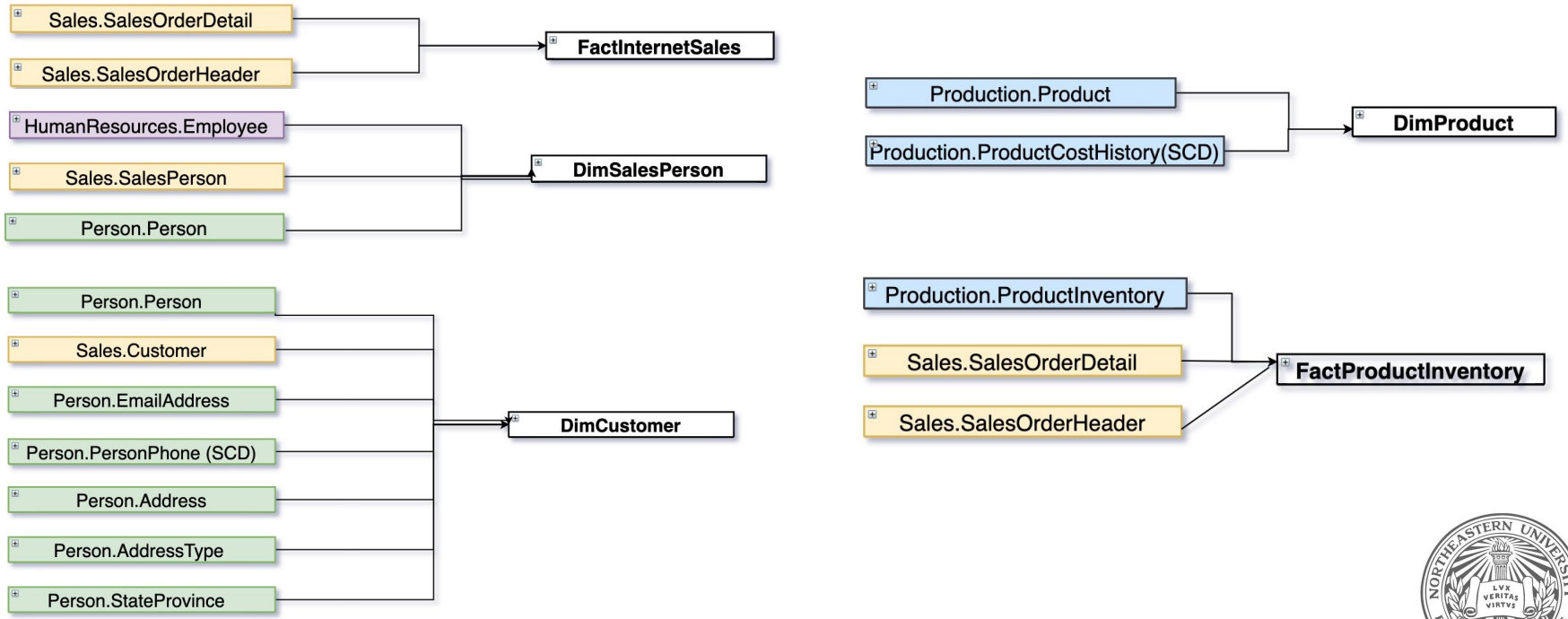




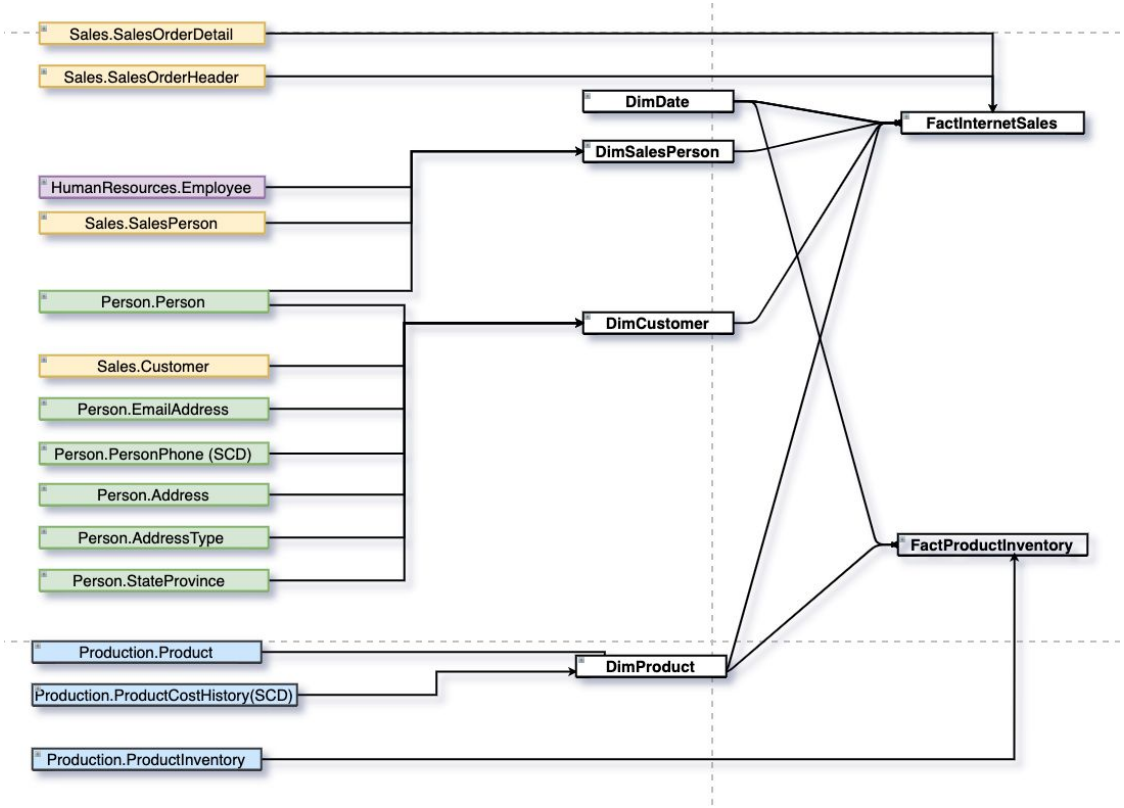
Data Warehouse Model ---- Facts and Dims



Data Warehouse Model ---- SOR



Data Warehouse Model



Data Integration ---- SCD

- DimCustomer: Customer Phone Number (Overwrite)

CustomerKey	BusinessEntityID	FirstName	LastName	Phone	UpdateDate
11000	11000	Jon	Yang	1(11) 500 555-0162	2011-08-30 11:34:29:403

Replace the orignal phone number

CustomerKey	BusinessEntityID	FirstName	LastName	Phone	UpdateDate
11000	11000	Jon	Yang	1(11) 500 555-0932	2011-11-04 10:01:24:522

- DimProduct: Product Unit Cost (Full History)

ProductKey	HistoryCostID	ProductUnitCost	CostStartDate	CostEndDate
301	1	33.25	2015-09-15	2016-03-20
301	2	34.75	2016-03-20	2016-05-25
301	3	40.25	2016-05-25	NULL

Data Integration

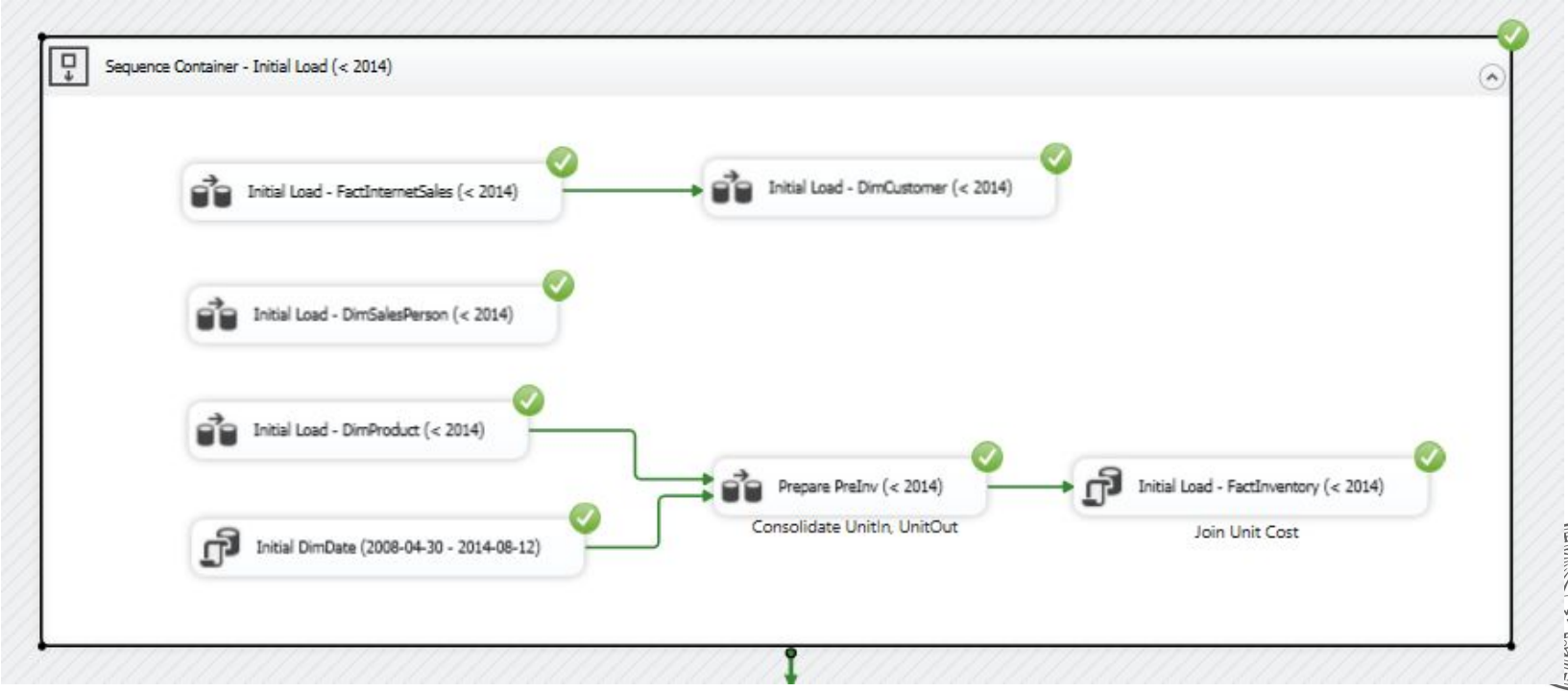
SQL Server Integration Services (SSIS)

SSIS

- **Initial Load:** load data before 2014
- **Incremental Loads:** 8 Batches (2014-01-01 to 2014-08-31)
- **SCD:** 2 test cases for customer Phone and product UnitCost respectively

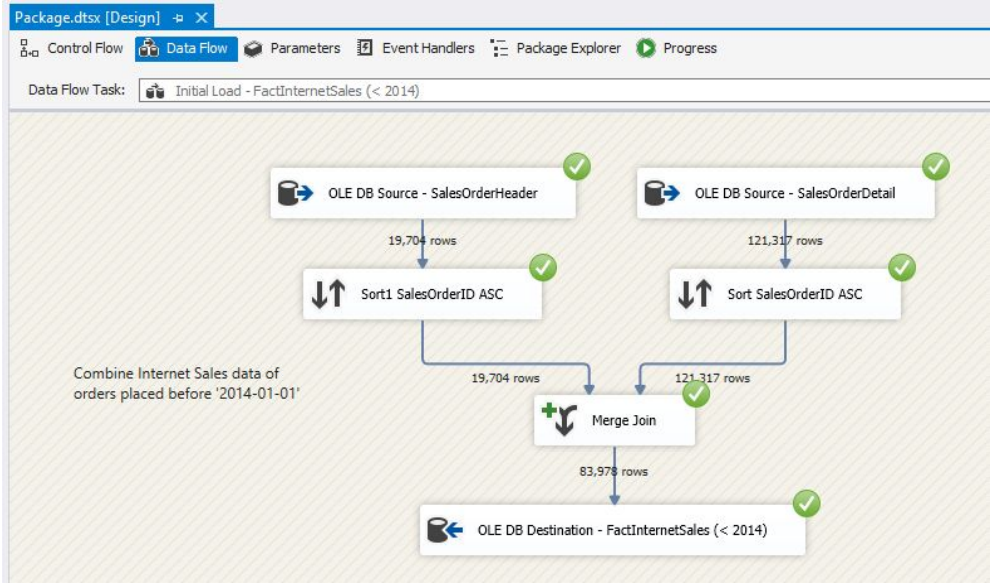
Initial Load

- Load data prior to 2014

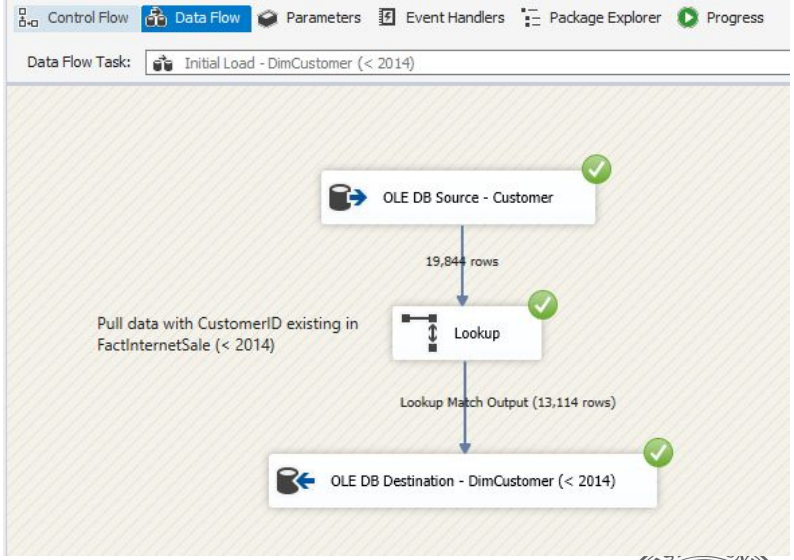


Initial Load

- Initial Load - FactInternetSales (< 2014)

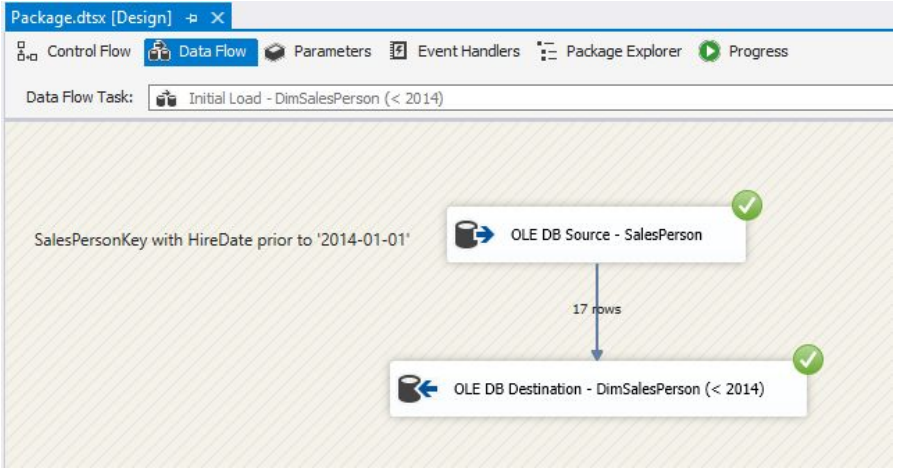


- Initial load - DimCustomer (< 2014)

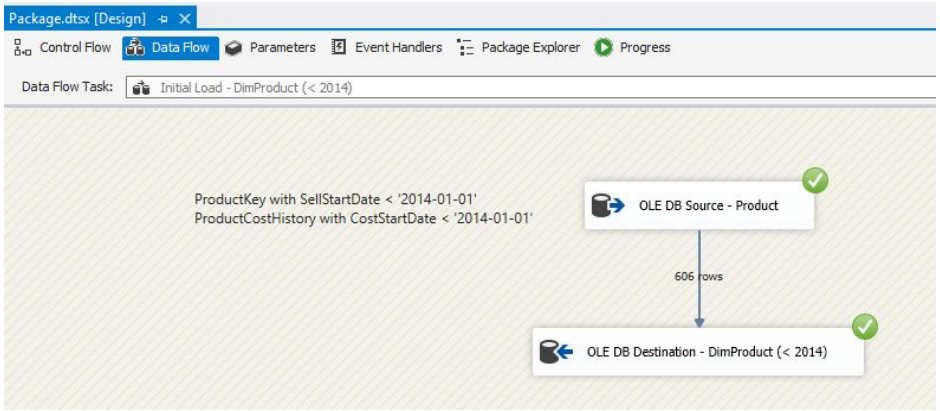


Initial Load

- Initial Load - DimSalesPerson (< 2014)

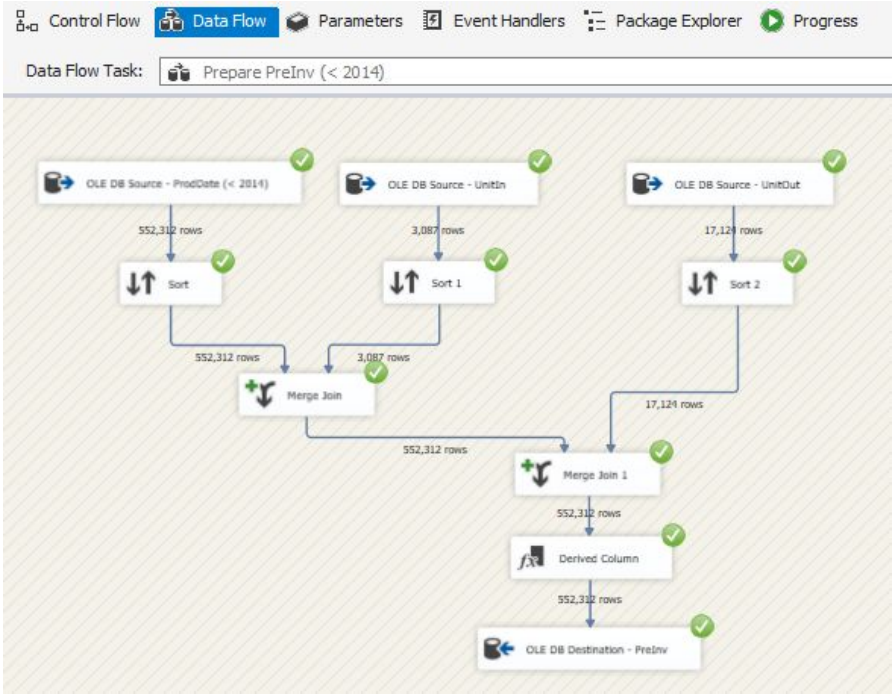


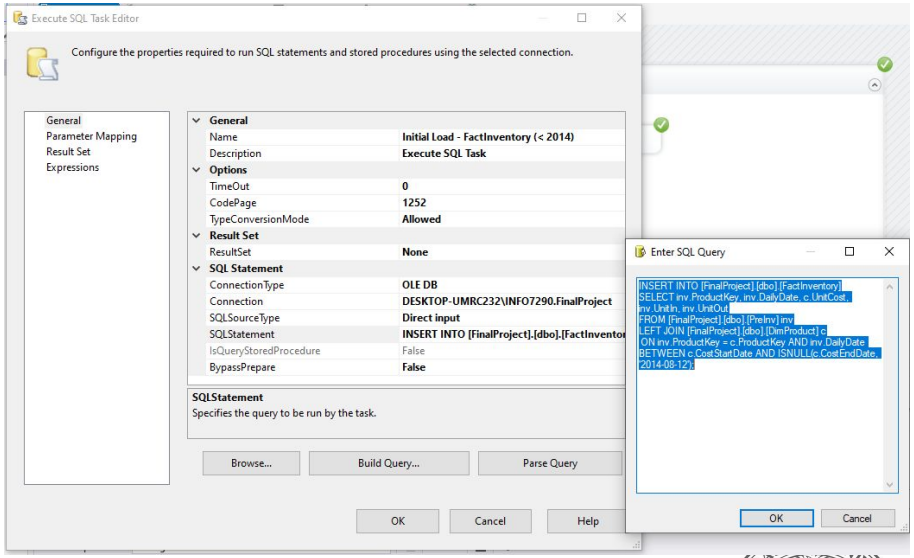
- Initial load - DimProduct (< 2014)



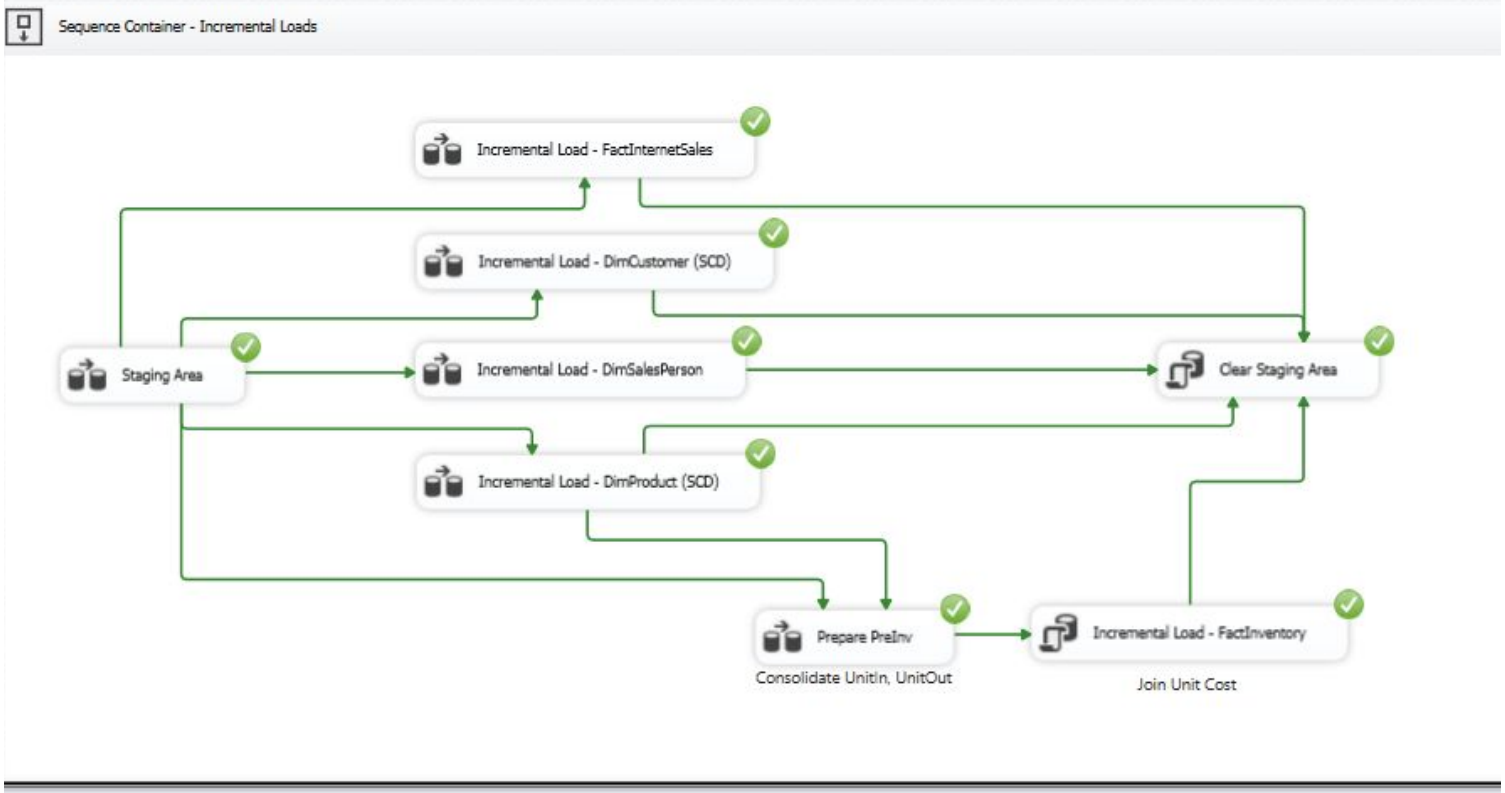
Initial Load

- Initial Load - FactInventory (< 2014)





Incremental Loads

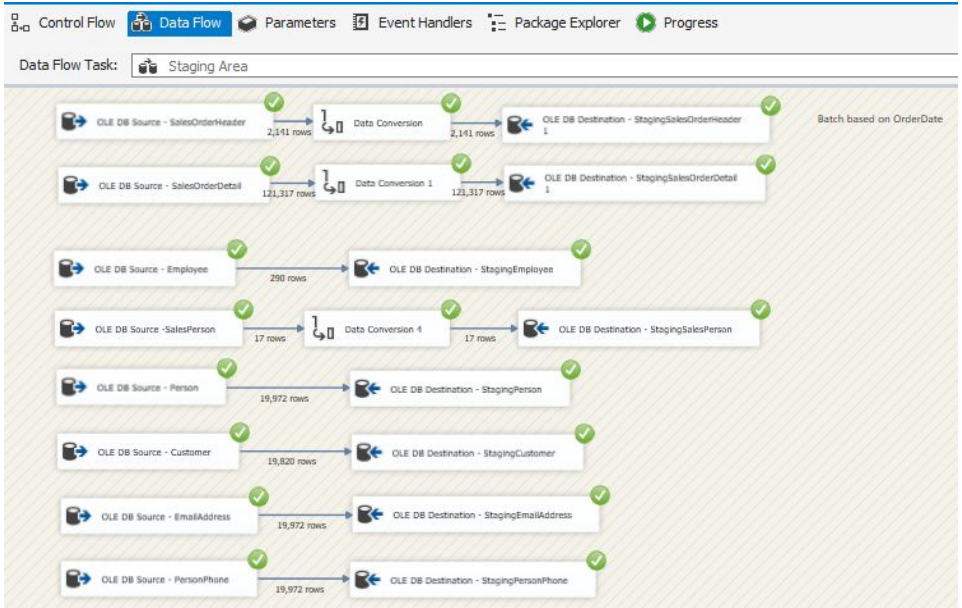


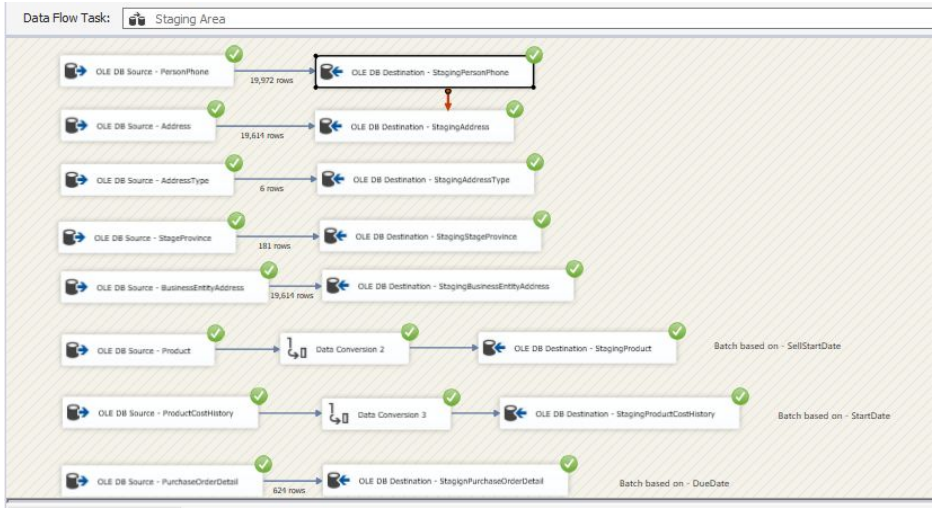
Incremental Loads

- Based on **month**
- Variables @BatchStartDate and @BatchEndDate to separate batches

Table \ Batch	Initial load (Prior to 2014)	Batch 1 (20140101-20140131)	Batch 2 (20140201-20140228)	Batch 3 (20140301-20140331)	Batch 4 (20140401-20140430)	Batch 5 (20140501-20140531)	Batch 6 (20140601-20140630)	Batch 7 (20140701-20140731)	Batch 8 (20140801-20140831)
FactInternetSales	83,978	91,023	95,304	105,259	110,561	119,187	121,317	121,317	121,317
FactInventory	552,312	564,898	576,266	588,852	601,032	613,618	625,798	638,384	638,384
DimCustomer	13,114	19,844	19,844	19,844	19,844	19,844	19,844	19,844	19,844
DimSalesPerson	17	17	17	17	17	17	17	17	17
DimProduct	606	606	606	606	606	606	606	606	606
DimDate	2,296	2,296	2,296	2,296	2,296	2,296	2,296	2,296	2,296

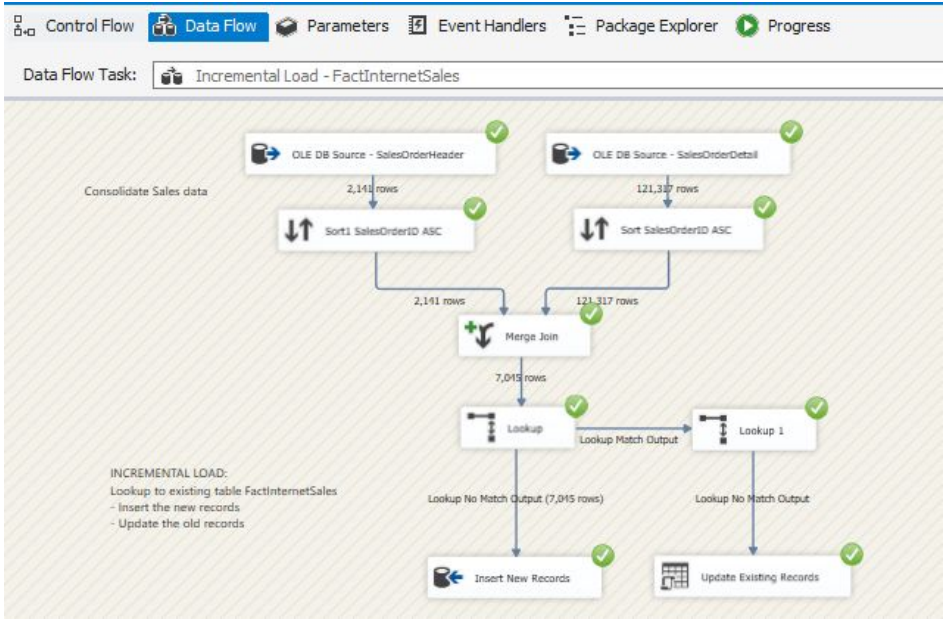
Incremental Loads - Staging Area



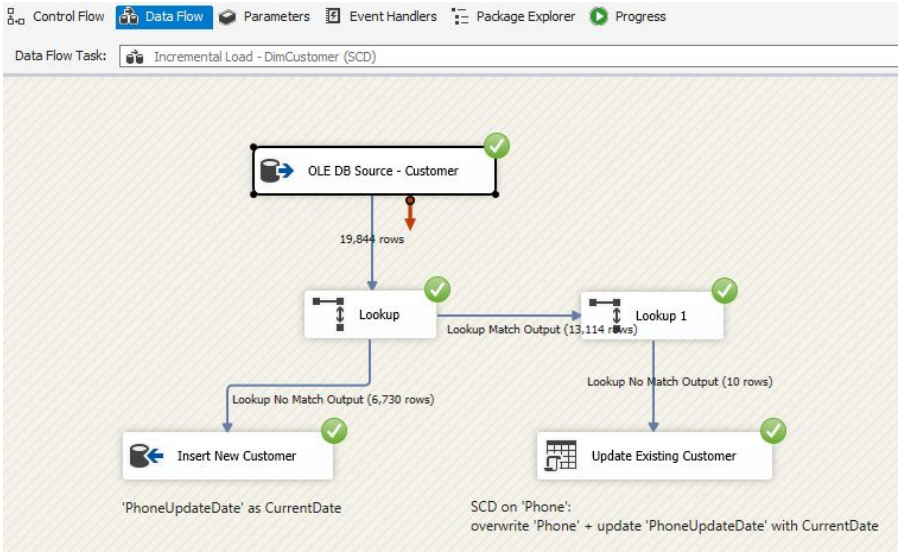


Incremental Loads - ETL Processing

- FactInternetSales

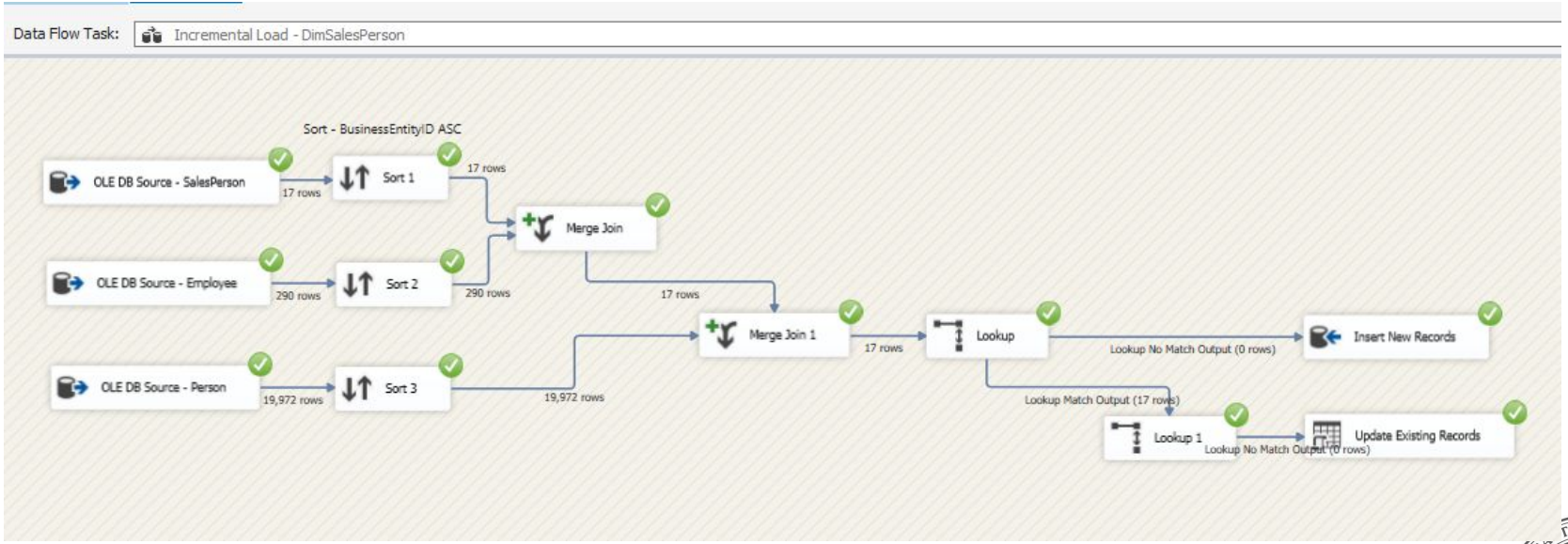


- DimCustomer



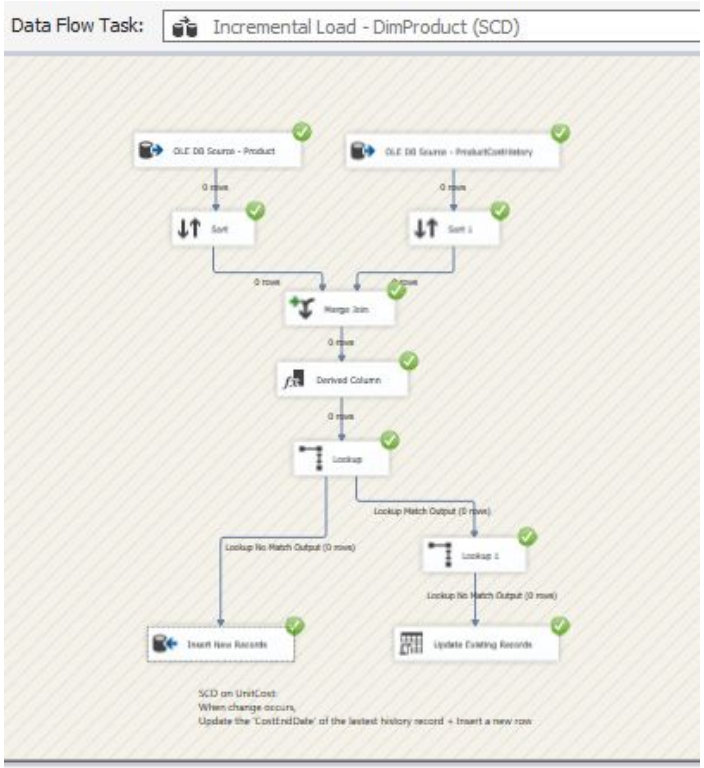
Incremental Loads - ETL Processing

- DimSalesPerson



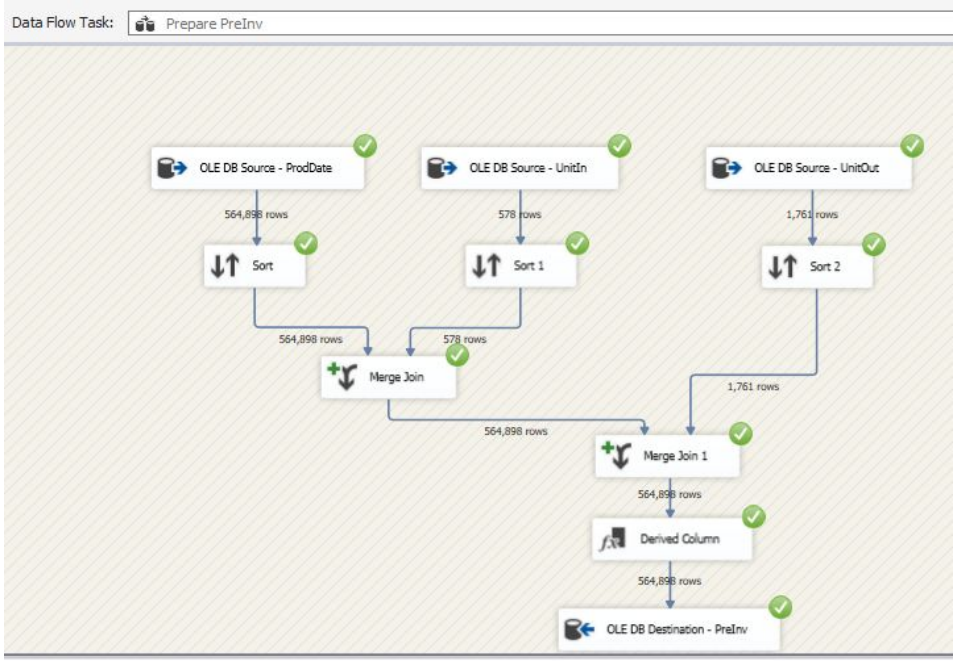
Incremental Loads - ETL Processing

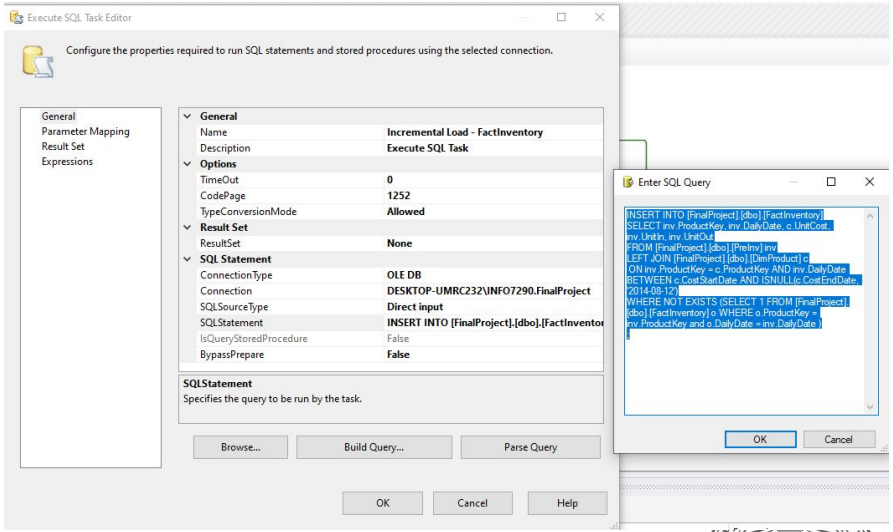
- DimProduct



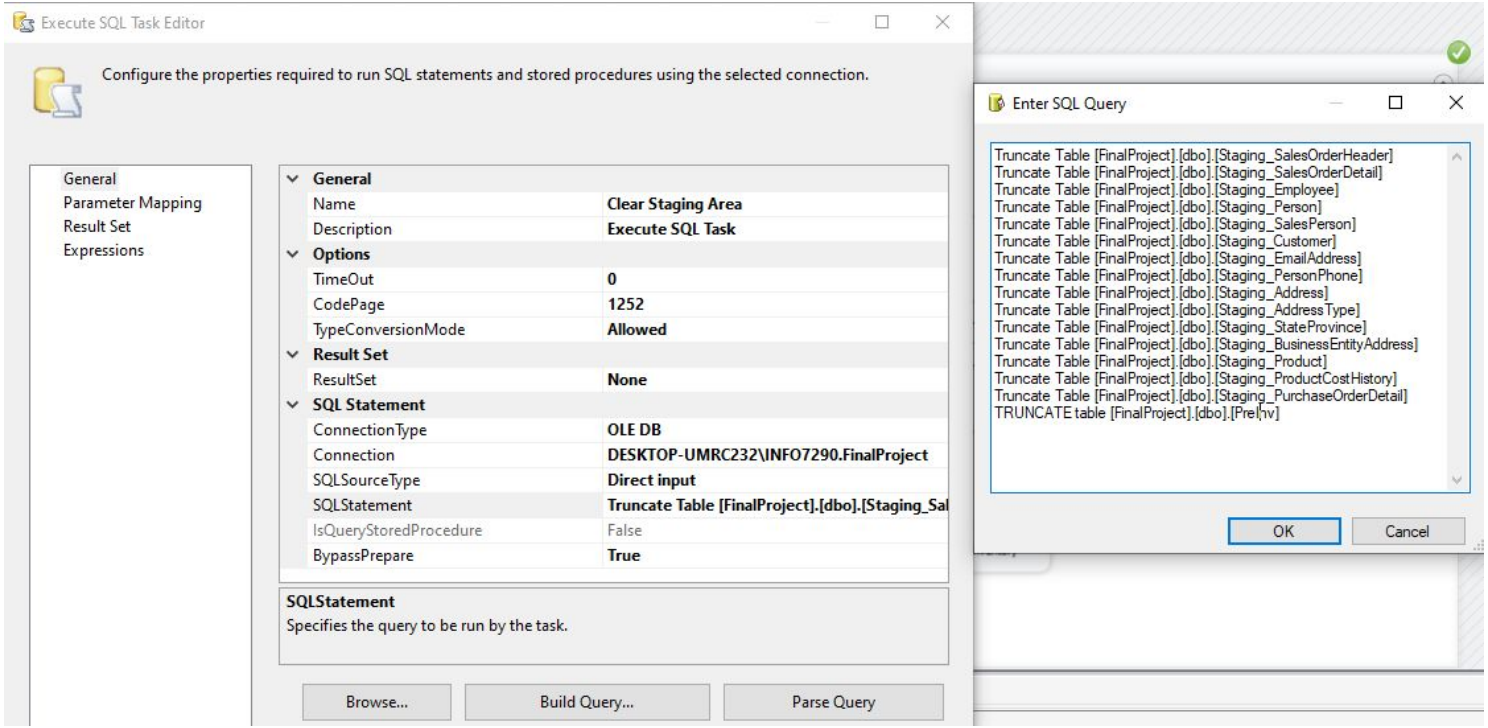
Incremental Loads - ETL Processing

- FactProductInventory





Incremental Loads - Clear Staging Tables



SCD1 - DimCustomer: Customer Phone Number (Overwrite)

Staging_PersonPhone

DimCustomer

Before

	BusinessEntityID	PhoneNumber	ModifiedDate
1	16722	855-555-0159	2013-11-01 00:00:00.000
2	16723	1 (11) 500 555-0161	2014-01-25 00:00:00.000

	CustomerKey	BusinessEntityID	Phone	PhoneUpdateDate
1	28360	16722	855-555-0159	2013-11-01 00:00:00.000
2	17835	16723	1 (11) 500 555-0161	2014-01-25 00:00:00.000

```
UPDATE [FinalProject].[dbo].[Staging_PersonPhone]
SET PhoneNumber = '99999', ModifiedDate = GETDATE()
WHERE BusinessEntityID = 16722 OR BusinessEntityID = 16723
```

After

	BusinessEntityID	PhoneNumber	ModifiedDate
1	16722	99999	2021-04-28 12:23:29.137
2	16723	99999	2021-04-28 12:23:29.137

	CustomerKey	BusinessEntityID	Phone	PhoneUpdateDate
1	28360	16722	99999	2021-04-28 12:23:29.137
2	17835	16723	99999	2021-04-28 12:23:29.137

SCD1 - DimProduct: Product Unit Cost (Full History)

Staging_ProductCostHistory

Before

	ProductID	StartDate	EndDate	StandardCost
1	707	2011-05-31	2012-05-29	12.0278
2	707	2012-05-30	2013-05-29	13.8782
3	707	2013-05-30	NULL	13.0863

```
UPDATE [FinalProject].[dbo].[Staging_ProductCostHistory]
SET EndDate = '9999-01-01'
WHERE ProductID = 707 AND StartDate = '2013-05-30';

INSERT INTO [FinalProject].[dbo].Staging_ProductCostHistory
VALUES ('707', '9999-01-01', Null, 99999)
```

After

	ProductID	StartDate	EndDate	StandardCost
1	707	2011-05-31	2012-05-29	12.0278
2	707	2012-05-30	2013-05-29	13.8782
3	707	2013-05-30	9999-01-01	13.0863
4	707	9999-01-01	NULL	99999.0000

DimProduct

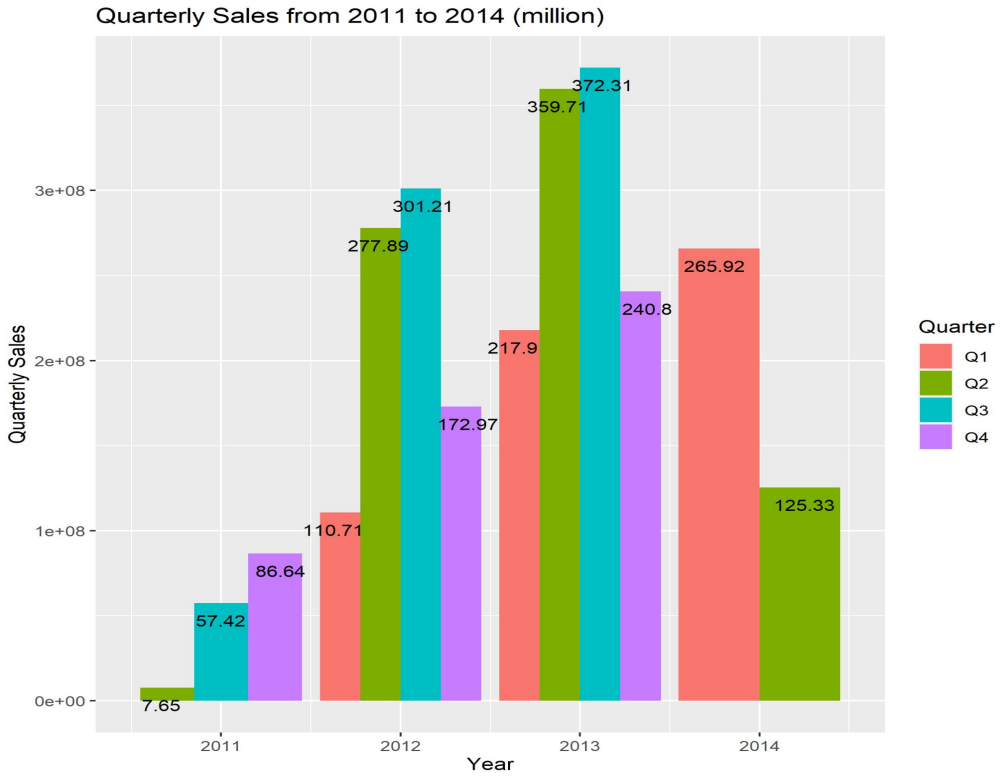
	ProductKey	HistoryCostID	ProductName	UnitCost	CostStartDate	CostEndDate
1	707	1	Sport-100 Helmet, Red	12.0278	2011-05-31	2012-05-29
2	707	2	Sport-100 Helmet, Red	13.8782	2012-05-30	2013-05-29
3	707	3	Sport-100 Helmet, Red	13.0863	2013-05-30	NULL

	ProductKey	HistoryCostID	ProductName	UnitCost	CostStartDate	CostEndDate
1	707	1	Sport-100 Helmet, Red	12.0278	2011-05-31	2012-05-29
2	707	2	Sport-100 Helmet, Red	13.8782	2012-05-30	2013-05-29
3	707	3	Sport-100 Helmet, Red	13.0863	2013-05-30	9999-01-01
4	707	4	Sport-100 Helmet, Red	99999.0000	9999-01-01	NULL

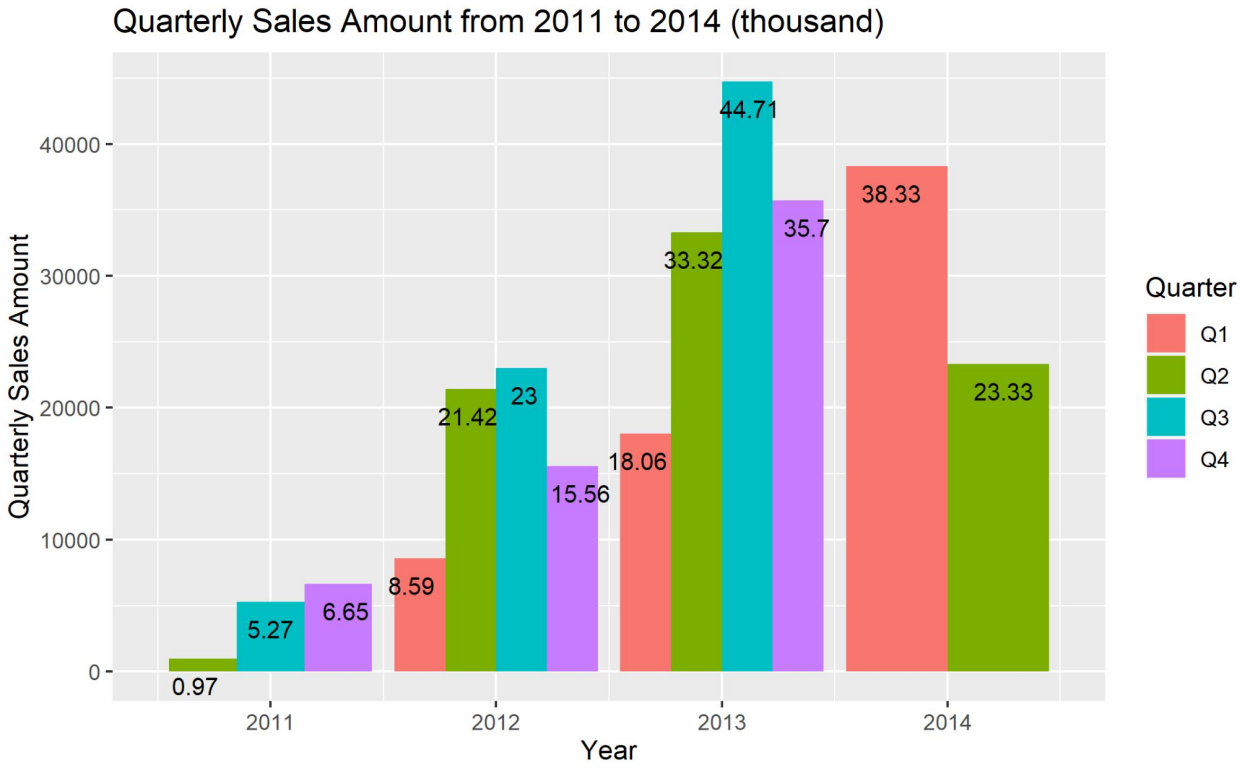
Data Visualization

Online Sales Analysis

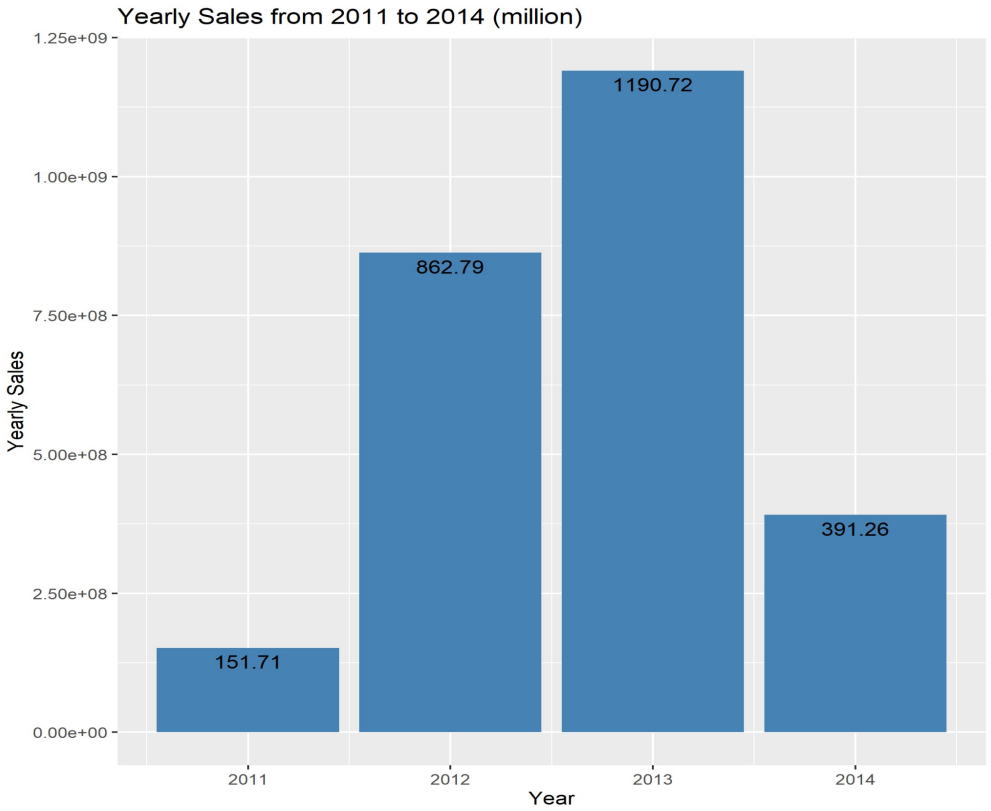
Data Visualization --- Quarterly Sales



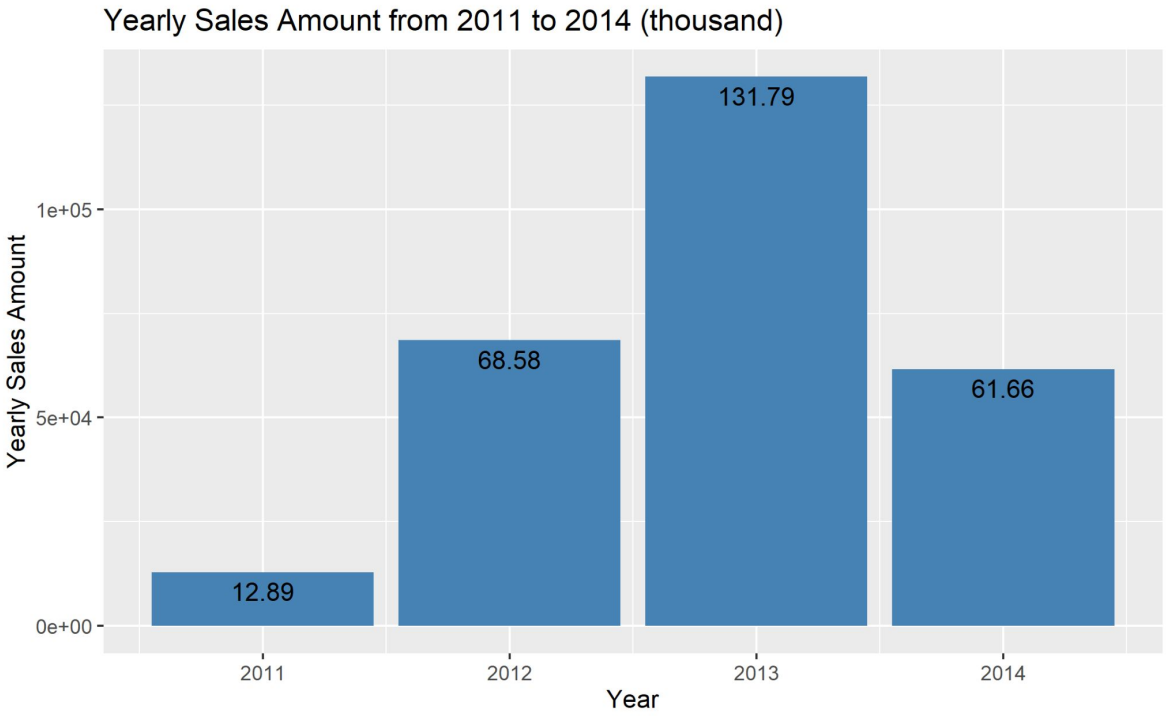
Data Visualization --- Quarterly Sales Amount



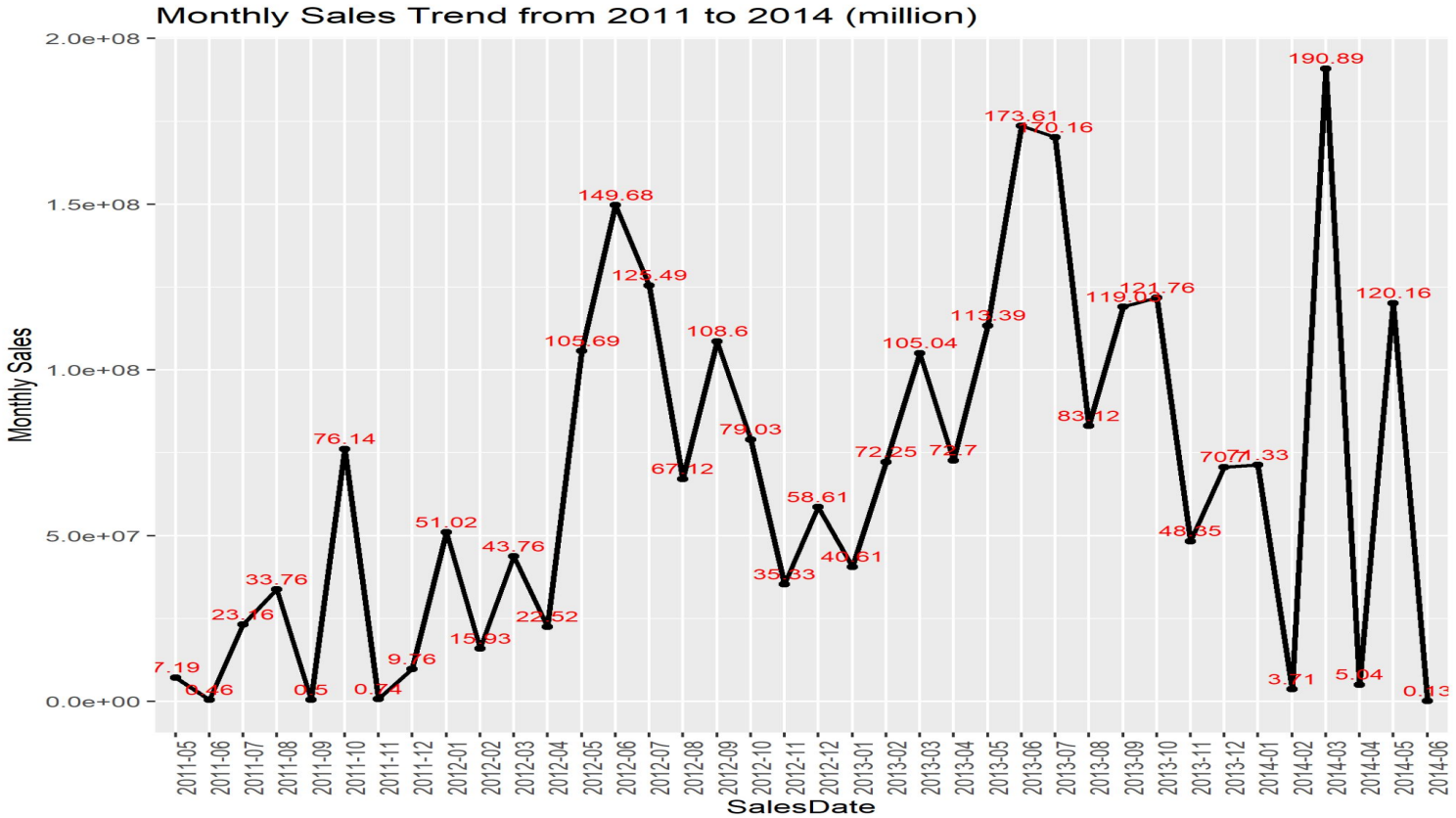
Data Visualization --- Yearly Sales



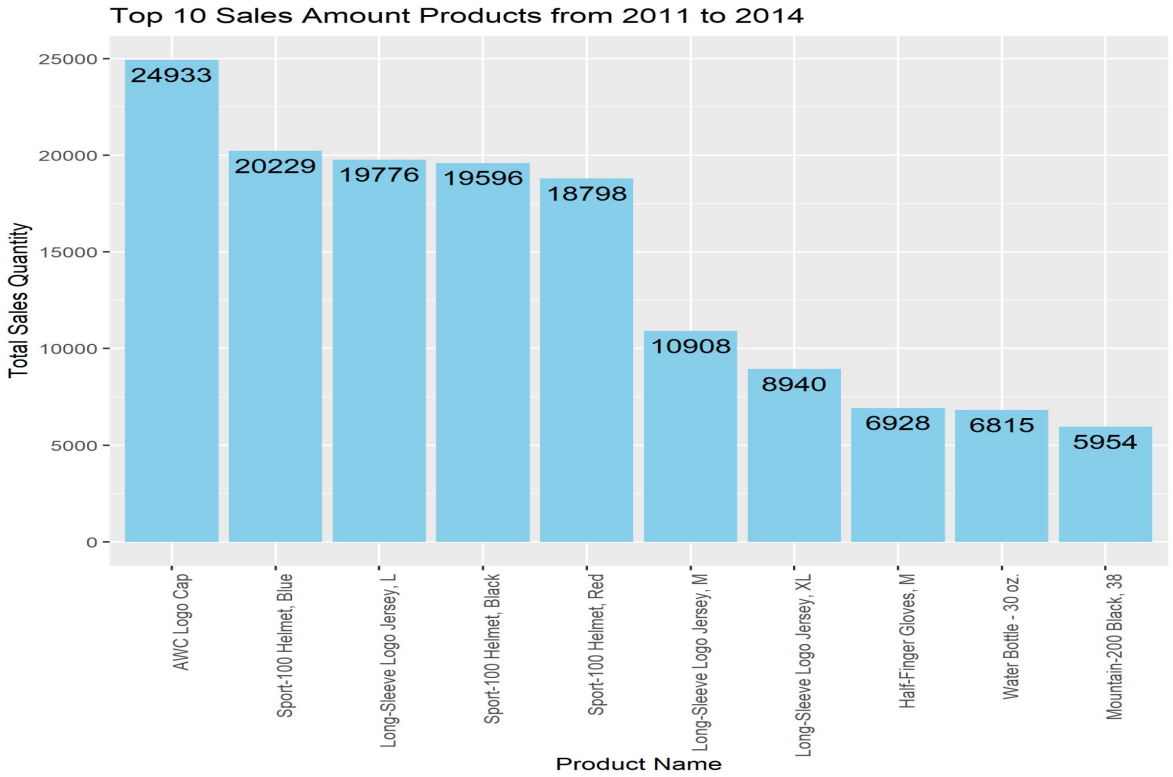
Data Visualization --- Yearly Sales Amount



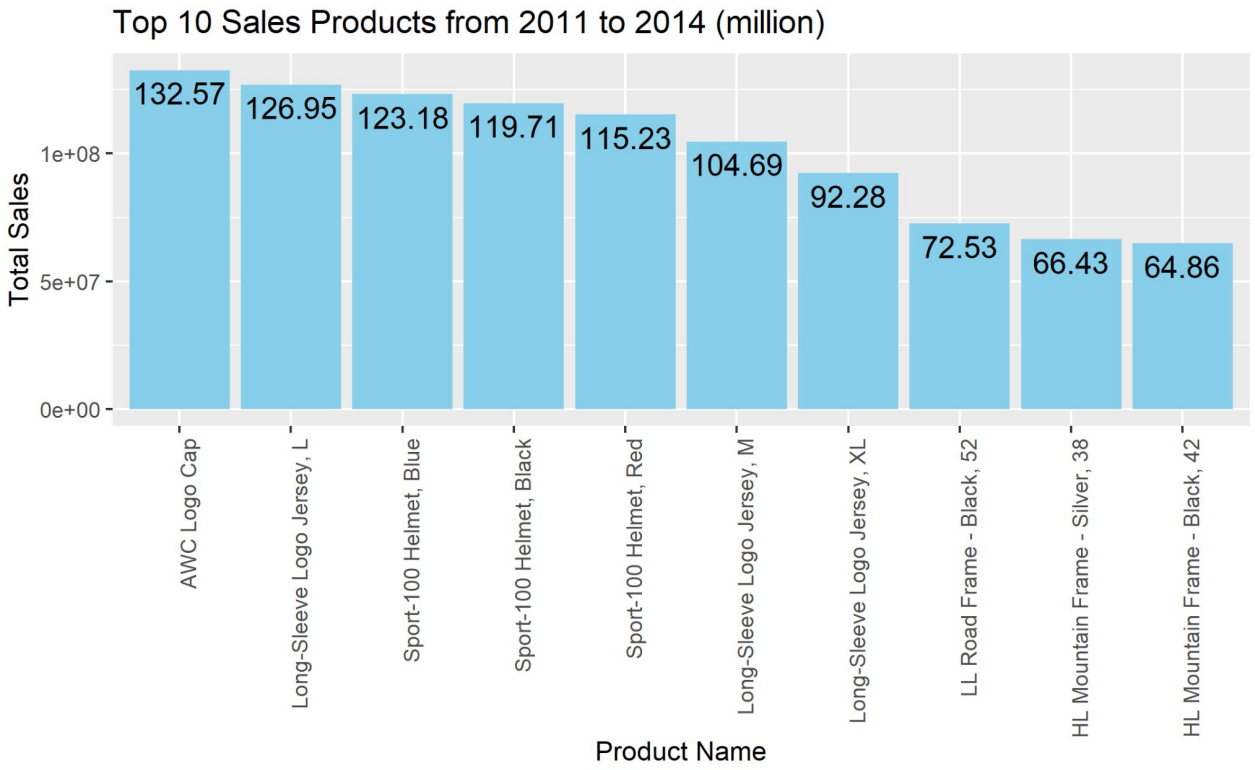
Data Visualization --- Sales Trends



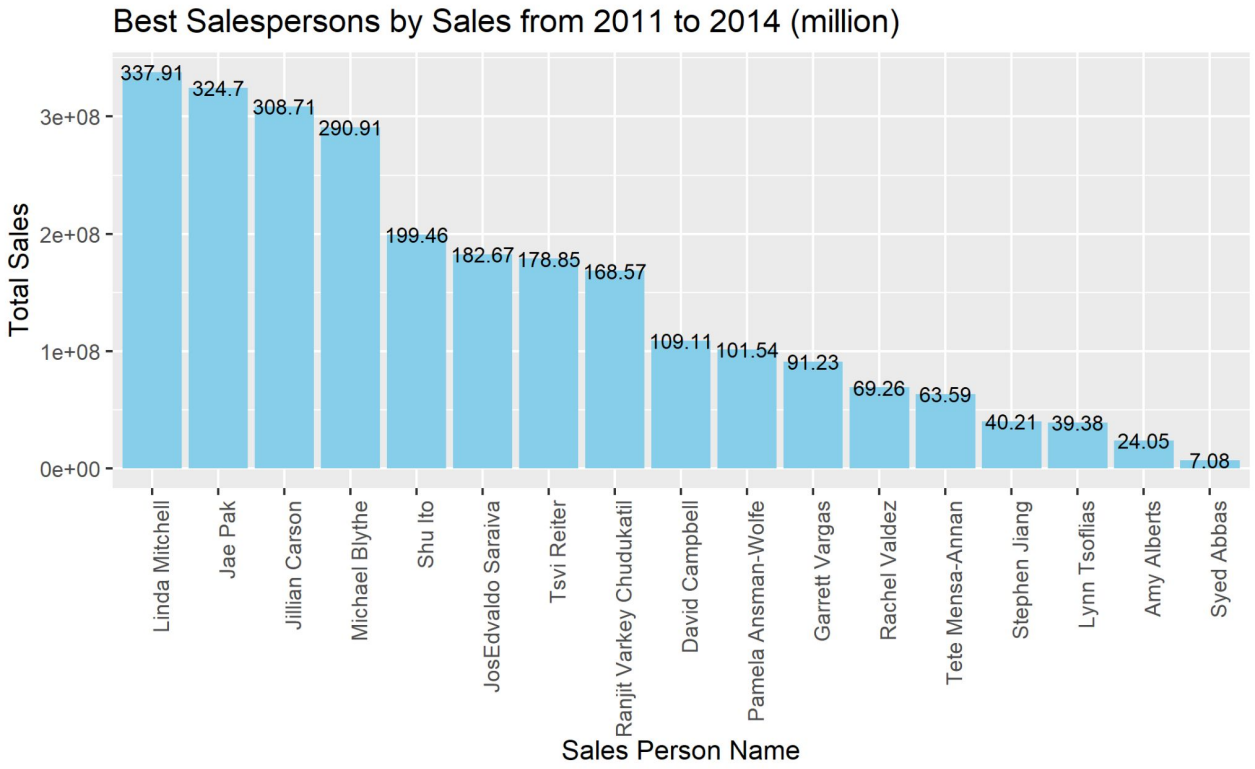
Data Visualization --- Most Popular Products



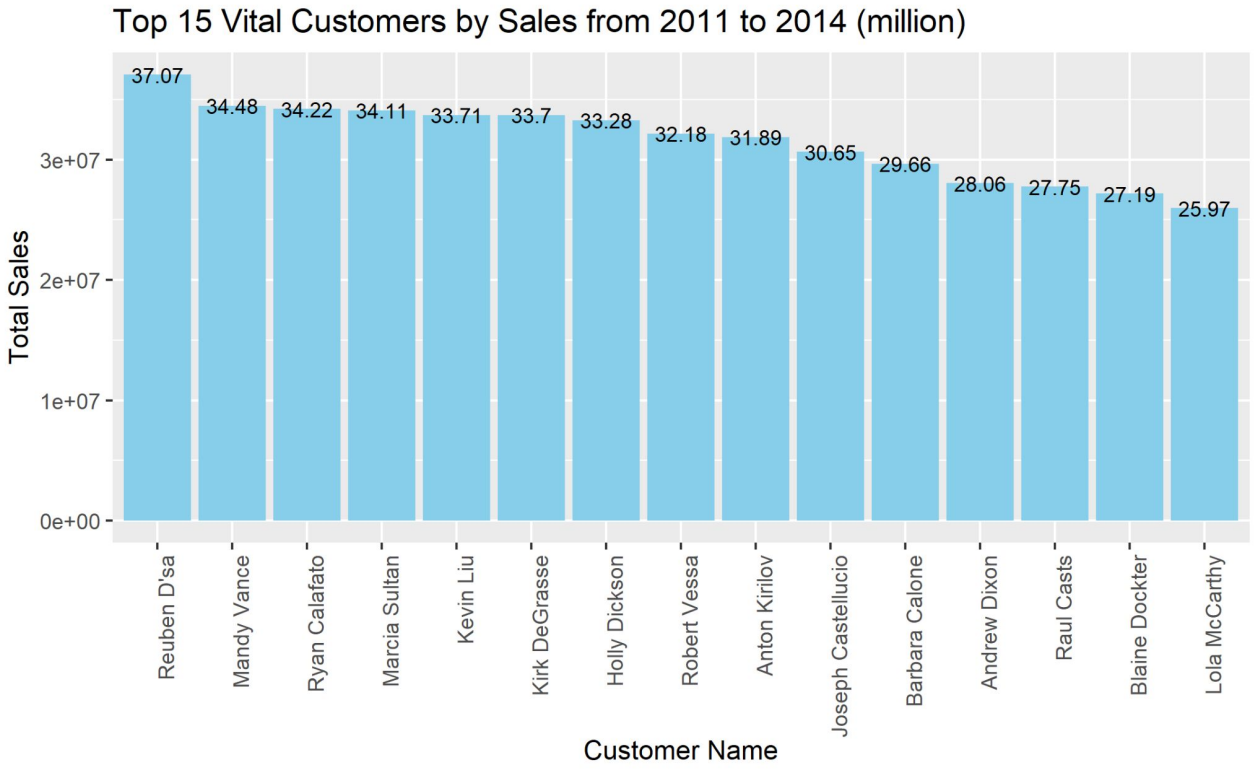
Data Visualization --- Most Popular Products



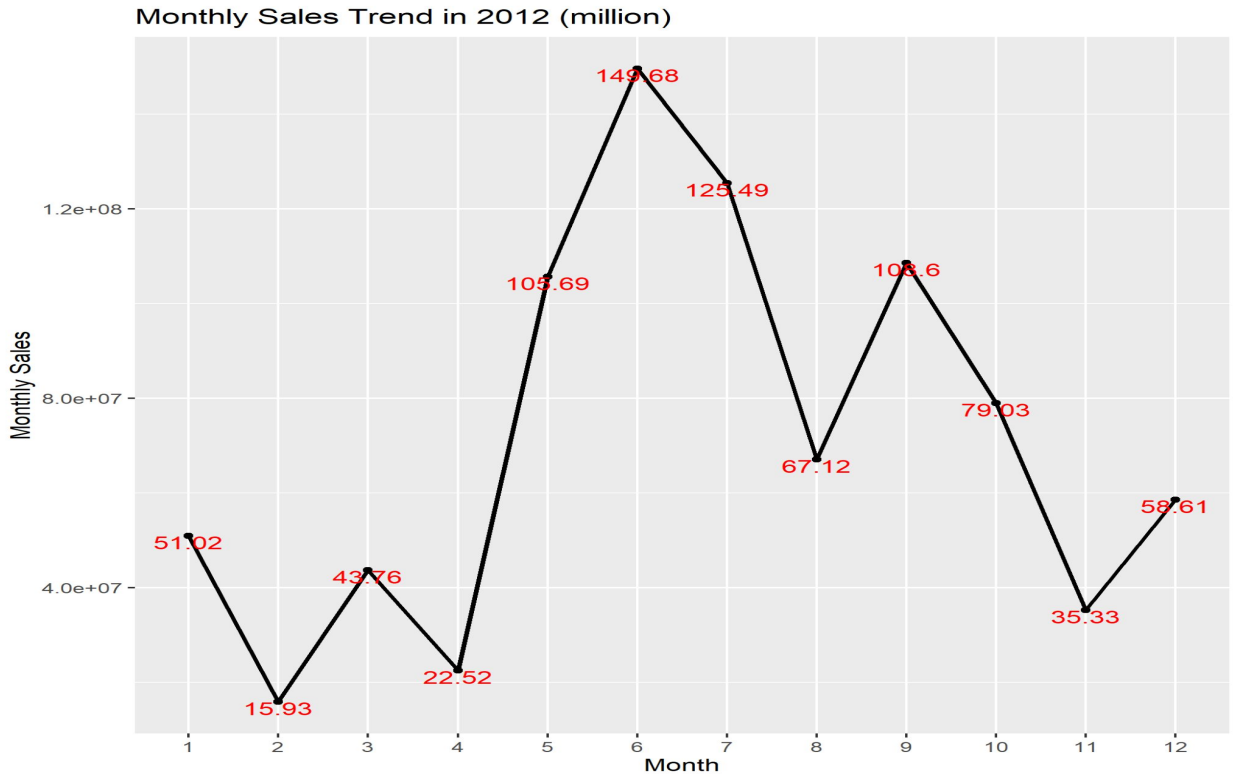
Data Visualization --- Best Salespersons



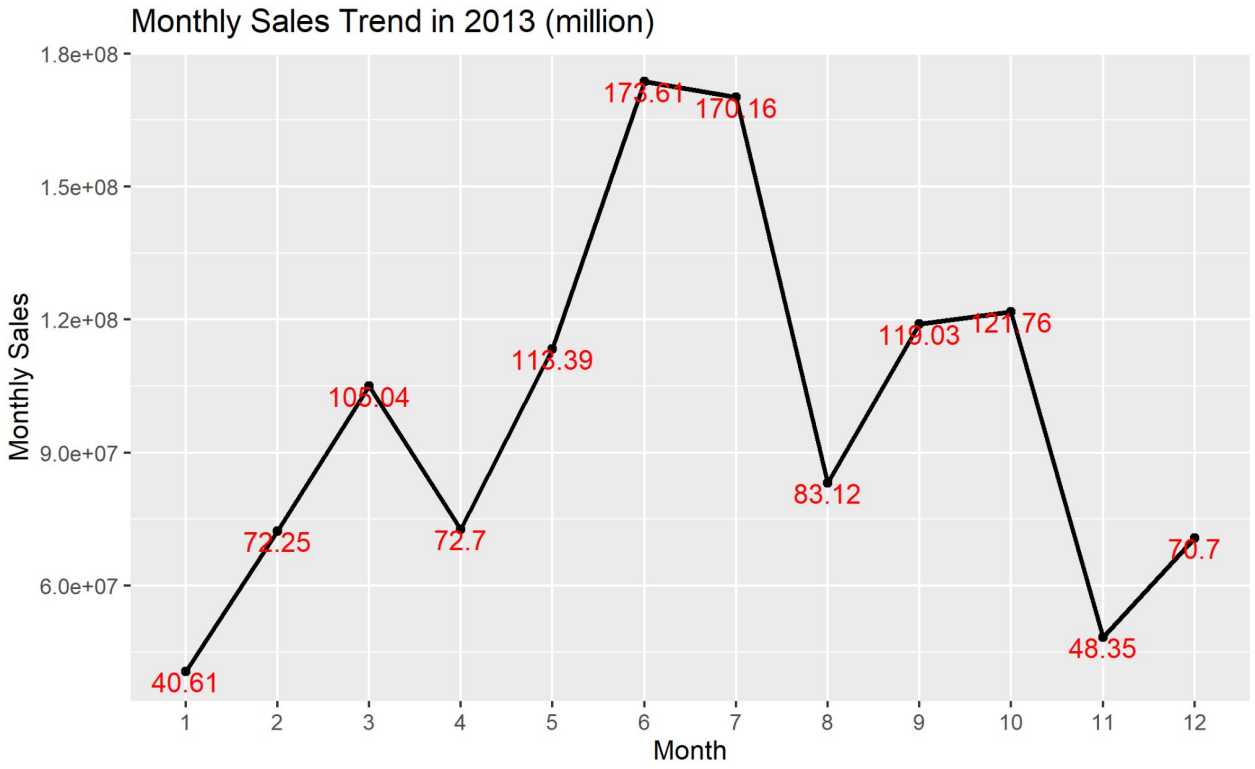
Data Visualization --- Vital Customers



Data Visualization --- Sales Trend in 2012



Data Visualization --- Sales Trend in 2013



Thanks!


NORTHEASTERN UNIVERSITY
LVX
VERITAS
VIRTVS
FOUNDED
A.D.
1898
BOSTON, MASSACHUSETTS