# INFO 7290:
# Data Warehousing
# & Business Intelligence

## Fall 2015

## BI & Data Integration

## Final Team Project

Northeastern University

Rick Sherman
**Athena IT Solutions**
ri.sherman@northeastern.edu

## Deliverables:

1. DI -- Load data sources into DW

   • SOR: flat files, SQL Server & MySQL

   • DW: Microsoft Contoso BI dataset (customized version)

2. BI -- Develop BI Dashboards, Reports & Visualizations

   • Store analysis

   • Online sales analysis

   • Inventory analysis

ATHENA
IT SOLUTIONS

## Data Integration Tools:

- Talend Enterprise Data Integrator

- SSIS


- Deliverables:
  - ✓ Load DW from data sources
    - ▪ Document all jobs
    - ▪ Provide load statistics
  - ✓ Handle data quality & error messages
    - ▪ Document error handling results
  - ✓ Comparison of each tool – key differences – strengths & weaknesses

## BI Tools:

- **Tableau**

- **Qlik**
  - ✓ Qlik Sense
  - ✓ Optional: QlikView

- **Microsoft BI**
  - ✓ PowerPivot, PowerView, PowerMap, Excel 2103
  - ✓ Optional: PowerBI, Office365

- **Deliverables:**
  - ✓ Dashboards with visualization for each analysis & for each BI tool
  - ✓ Comparison of each tool – key differences – strengths & weaknesses

- ## Contoso – Fictional Retail Company

| TableName | row_count |
|---|---|
| DimAccount | 24 |
| DimChannel | 4 |
| DimCurrency | 28 |
| DimCustomer | 18,869 |
| DimDate | 3,652 |
| DimEmployee | 293 |
| DimEntity | 421 |
| DimGeography | 674 |
| DimMachine | 7,816 |
| DimOutage | 303 |
| DimProduct | 2,517 |
| DimProductCost | 2,517 |
| DimProductPrice | 2,517 |
| DimPromotion | 28 |
| DimSalesTerritory | 265 |
| DimScenario | 3 |
| DimStore | 306 |

| TableName | row_count |
|---|---|
| FactExchangeRate | 773 |
| FactInventory | 8,013,099 |
| FactITMachine | 23,283 |
| FactITSLA | 4,925 |
| FactOnlineSales | 12,627,608 |
| FactSales | 3,334,098 |
| FactSalesQuota | 7,465,911 |
| FactStrategyPlan | 2,750,628 |

<u>Note:</u> The final dataset may have slight schema changes and row counts may vary

ATHENA
IT SOLUTIONS

# Dashboards - Tabular Reports, Visualizations or Maps
## Deliverables – Sample Questions

- **Online Sales Analysis**
  - ✓ Sales amount, profit & profit margin by Category/SubCategory and Time (Year/Qtr/Month)
  - o Sales & profit with Period over Period analysis
  - o Sales & profit by customer demographics such as education, income, etc.

- **Store Sales Analysis**
  - o Sales & profit by channel and Time (Year/Qtr/Month)
  - o Rank sales
    - ▪ By Year
    - ▪ Product Category
    - ▪ Country & State

- **Sales Analysis – Geo analysis**
  - o Display on map
  - o Provide contribution analysis

ATHENA
IT SOLUTIONS

- Use existing schema from Contoso_Retail_BI as your target DW schema. Truncate existing data on your data integration jobs.

- Data is being sourced from 3 geographic area (continents) databases:
  - ✓ Contoso_Retail_SOR_NorthAmerica  - Microsoft SQL Server
  - ✓ Contoso_Retail_SOR_Europe - MySQL
  - ✓ Contoso_Retail_SOR_Asia – PostgreSQL

- In addition many tables are sourced from various files in Excel, csv or text delimited file format

ATHENA
IT SOLUTIONS

# BI Team Project: 12/9/2015
## Systems of Record (SOR)

| Contoso_Retail_SOR_NorthAmerica |
|---|
| DimCustomer_Company |
| DimCustomer_Person |
| DimProduct |
| DimProductCategory |
| DimProductSubcategory |
| DimPromotion |
| DimStore |
| FactCatalogSales_NA |
| FactOnlineSalesOrder_NA |
| FactOnlineSalesOrderDetail_NA |
| FactResellerSales_NA |
| FactStoreSales_NA |
| DimGeography |
| FactStrategyPlan |

| Contoso_Retail_SOR_Europe |
|---|
| DimCustomer_Company |
| DimCustomer_Person |
| DimProduct |
| DimProductCategory |
| DimProductSubcategory |
| DimPromotion |
| DimStore |
| FactInventory |
| FactOnlineSalesOrder_EU |
| FactOnlineSalesOrderDetail_EU |
| FactResellerSales_EU |
| FactStoreSales_EU |

| Contoso_Retail_SOR_Asia |
|---|
| DimCustomer_Company |
| DimCustomer_Person |
| DimProduct |
| DimProductCategory |
| DimProductSubcategory |
| DimPromotion |
| DimStore |
| FactInventory |
| FactOnlineSalesOrder_AS |
| FactOnlineSalesOrderDetail_AS |
| FactResellerSales_AS |
| FactStoreSales_AS |

| Files |
|---|
| DimAccount.txt |
| DimChannel.csv |
| DimCurrency.csv |
| DimEntity.csv |
| DimProduct_Cost_CNY_Step_1.csv |
| DimProduct_Cost_CNY_Step_2.csv |
| DimProduct_Cost_CNY_Step_3.csv |
| DimProduct_Cost_CNY_Step_4.csv |
| DimProduct_Cost_EUR_Step_1.txt |
| DimProduct_Cost_EUR_Step_2.txt |
| DimProduct_Cost_EUR_Step_3.txt |
| DimProduct_Cost_EUR_Step_4.txt |
| DimProduct_Price_USD_Step_1.csv |
| DimProduct_Price_USD_Step_2.csv |
| DimProduct_Price_USD_Step_3.csv |
| DimProduct_Price_USD_Step_4.csv |
| DimSalesTerritory.csv |
| DimScenario.txt |
| FactExchangeRate.xlsx |

ATHENA
IT SOLUTIONS

# Systems of Record (SOR)

- There are 4 sales channels for this company:
  - ✓ Catalog
  - ✓ Retail
  - ✓ Stores
  - ✓ Online Sales

- In DW Sales are broken into:
  - ✓ FactSales – includes all 4 channels
  - ✓ FactOnlineSales – only includes Online Sales

- IN SOR sales are broken into 4 sales channels & 3 continents (North America, Europe & Asia):
  - ✓ Catalog – note: US-based only
  - ✓ Retail
  - ✓ Stores
  - ✓ Online Sales – further broken into Header & Detail (line) tables

ATHENA
IT SOLUTIONS

- DW has all data in US dollars (USD)
- SORs have prices, costs & sales in "continent" currency
  - ✓ North America – USD
  - ✓ Europe - Euro
  - ✓ Asia - China Yuan

- Sales, Returns & Costs are in "constant" currency, i.e. recorded using published unit prices & costs
  - ✓ Daily currency exchange rate should be used in converting Euro & Yuan to USD

- Unit Price & Unit Cost should NOT stored in Fact Sales related tables nor in the DimProduct dimension

- Unit Prices & United Costs were independently changed 3 times during 2012-2014. You need to create SCD dimension for both Unit Price & Unit Cost Dimensions.

  ✓ Step 1 – initial unit prices or costs
  ✓ Step 2 – prices or costs revised
  ✓ Step 3 – prices or costs revised
  ✓ Step 4 – prices or costs revised

| Cost_Step | Effective_Date |
|-----------|----------------|
| 1 | 1/1/2012 |
| 2 | 10/1/2012 |
| 3 | 10/1/2013 |
| 4 | 10/1/2014 |

| Pricing_Step | Effective_Date |
|--------------|----------------|
| 1 | 1/1/2012 |
| 2 | 7/1/2012 |
| 3 | 7/1/2013 |
| 4 | 7/1/2014 |

ATHENA
IT SOLUTIONS

- Error Handling Standard will be to reject any rows that have incorrect FKs such as:
  - ✓ Product
  - ✓ Customer
  - ✓ Geography
  - ✓ Promotion
  - ✓ Store

ATHENA
IT SOLUTIONS