INFO 7290:
Data Warehousing
& Business Intelligence

BI & Data Integration

Final Team Project

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

Deliverables:

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

2. DI -- Load data sources into DW
   - SOR: flat files, SQL Server, Oracle, PostgeSQL & MySQL
   - DW: Retail BI – SQL Server (SSIS), MySQL (Talend)

3. BI -- Develop BI Dashboards, Reports & Visualizations Sales (Channel) analysis
   - Sales analysis
   - Inventory analysis
   - Sales quota analysis

ATHENA
IT SOLUTIONS

BI Tools:

- Tableau

- Qlik
  - Qlik Sense

- Microsoft BI
  - PowerBI
  - PowerPivot, PowerView, PowerMap, Excel 2103

Deliverables:

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

ATHENA
IT SOLUTIONS

Data Integration Tools:

- Talend Enterprise Data Integrator
- Microsoft SSIS

Deliverables:

- Load DW from data sources
  - Document all jobs
  - Provide load statistics
  - Provide analysis of load jobs using one of BI tools used in this projecy
- Handle data quality & error messages
  - Document error handling results
- Load rejection records
  - Track reasons for rejections
  - Provide analysis of rejections & reasons using one of BI tools used in this project

- ## Note:
  - Follow project standards

Team Project
Data Sourceing Requirements

ATHENA
IT SOLUTIONS

BI Team Project:
Systems of Record (SOR)

retail_sor_catalog
DimGeography
DimProduct
DimProductCategory
DimProductSubcategory
DimPromotion_NA
DimStore_NA
FactCatalogSales_NA

Retail_SOR_NorthAmerica
DimCustomer_Company
DimCustomer_Person
DimGeography
DimProduct
DimProductCategory
DimProductSubcategory
DimPromotion_NA
DimStore_NA
FactInventory
FactOnlineSalesOrderDetail_NA
FactOnlineSalesOrderHeader_NA
FactResellerSales_NA
FactStoreSales_NA

Retail_SOR_Europe
DimCustomer_Company
DimCustomer_Person
DimGeography
DimProduct
DimProductCategory
DimProductSubcategory
DimPromotion_EU
DimStore_EU
FactInventory
FactOnlineSalesOrderDetail_EU
FactOnlineSalesOrderHeader_EU
FactResellerSales_EU
FactSalesQuota_EU
FactStoreSales_EU

Retail_SOR_Asia
DimCustomer_Company
DimCustomer_Person
DimGeography
DimProduct
DimProductCategory
DimProductSubcategory
DimPromotion_AS
DimStore_AS
FactInventory
FactOnlineSalesOrderDetail_AS
FactOnlineSalesOrderHeader_AS
FactResellerSales_AS
FactSalesQuota_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

Note: These may be revised

ATHENA
IT SOLUTIONS

BI Team Project:
Systems of Record (SOR)

- Data is being sourced from 3 geographic area (continents) databases:

  o Retail_SOR_NorthAmerica - Microsoft SQL Server

  o Retail_SOR_Europe - MySQL

  o Retail_SOR_Asia – PostgreSQL

  o Retail_SOR_Catalog - Oracle

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

ATHENA
IT SOLUTIONS

- 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

ATHENA
IT SOLUTIONS

- 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

Note: These will be revised

ATHENA
IT SOLUTIONS

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

- Fact tables should have a "rejects" table that contains the rows with errors and a error reason column

Team Project
BI Requirements

Deliverables – Analysis

- **Data Subjects:**
    - Online Sales Analysis (FactOnlineSales)
    - Sales Analysis (FactSales)
    - Inventory Analysis (FactInventory)

- **Types of analysis:**
    - Trending
    - Ranking
    - Comparison
    - Period over Period
    - Geo Map
    - Contribution

- **Measures:**
    - Sales $, Profit, Profit Margin, Avg Order Size,…

- **Dimensions:**
    - Customers: Company & Person, demographics
    - Product: Product Hierarchy (Category, Subcategory, Product), Brand, other attributes
    - Store: Type, other attributes
    - Dates
    - Geography

ATHENA
IT SOLUTIONS

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

ATHENA
IT SOLUTIONS

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

- Store Sales Analysis
  - Sales & profit by channel and Time (Year/Qtr/Month)
  - Rank sales
    - By Year
    - Product Category
    - Country & State
  - Sales Analysis – Geo analysis
  - Provide contribution analysis

ATHENA
IT SOLUTIONS