INFO 7290:
Data Warehousing
& Business Intelligence

BI & Data Integration

Final Team Project

Version: 1.0

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

- This project involves a retail company that sells a variety of products to people and to businesses across a variety of sales channels.

- The Retail Company would like to:
  - Create a data warehouse (DW) from its various systems of record (SORs)
  - Create BI applications that enable analysis of business performance

- The DW will store (facts) in the following subject areas:
  - Sales
  - Inventory
  - Sales Quotas
  - Strategy Plans

- The categories of products sold:
    - Audio
    - TV and Video
    - Computers
    - Cameras and camcorders
    - Cell phones
    - Music, Movies and Audio Books
    - Games and Toys
    - Home Appliances

- The Retail Company would like to:
    - Create a data warehouse (DW) from its various systems of record (SORs)
    - Create BI applications that enable analysis of business performance

ATHENA
IT SOLUTIONS

BI Team Project:
Retail Company

- The company has 4 sales channels:
  - Retail
  - Stores
  - Catalog
  - Online

- The company has three top-level geographic regions:
  - North America
  - Europe
  - Asia

- Customers who purchased products are tracked in:
  - Catalog
  - Online

ATHENA
IT SOLUTIONS

- There are 4 SORs:

Database	Business Area	DBMS
o Retail_SOR_NA	North America	Microsoft SQL Server
o Retail_SOR_EU	Europe	MySQL
o Retail_SOR_AS	Asia	PostgreSQL
o Retail_SOR_CAT	Catalog	Oracle

- Data is also sourced from files in Excel, csv or text delimited file format

- The Continent oriented SORs contain:
  - o Sales – Store, Reseller & Online Channels data

- The Catalog SOR contain:
  - o Sales – Catalog Channel data from all three continents

- All SORs contain:
  - o Inventory
  - o Sales Quotas
  - o Strategy Plans
  - o Supporting dimensions

ATHENA
IT SOLUTIONS

- SORs have prices, costs & sales in "continent" currency
  o North America & Catalog Operations – US Dollars (USD)
  o Europe - Euro (EUR)
  o Asia - China Yuan (CNY)

- DW has prices, costs & sales:
  o In USD
  o In "local" currency used in SOR

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

ATHENA
IT SOLUTIONS

- Unit Prices & Costs were 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

- Unit Prices & Unit Costs should NOT stores in Sales related tables nor in the Product dimension

ATHENA
IT SOLUTIONS

- DW sales related tables
  - o FactSales – contains data for all 4 sales channels
  - o FactOnlineSales – contains Online sales channel only and has customer data

- Sales Quotas & Strategy Plan:
  - o Have data for three scenarios
    - Actuals
    - Budget
    - Forecast
  - o Data granularity is monthly
  - o For Sales Quota need to aggregate actual sales data to populate
  - o For Strategy Plan need to aggregate data for actuals in 2 accounts:

Account Key	Account Name
4	Sale Revenue
5	Cost of Goods Sold

ATHENA
IT SOLUTIONS

Database Table Names & Row Counts

TableName	retail_sor_na	retail_sor_eu	Retail_SOR_AS	retail_sor_cat	Total
DimAccount	24	24	24	15	
DimCustomer_Company	276	43	67		
DimCustomer_Person	9,395	5,505	3,593	10,000	
DimDate	45,656	45,656	45,656	45,656	
DimEntity	209	56	41	4	
DimGeography	517	517	517	249	
DimProduct	2,517	2,517	2,516	2,517	
DimProductCategory	8	8	8		
DimProductSubcategory	44	44	44		
DimPromotion	10	19	10	28	
DimSalesTerritory	183	53	29		
DimStore	209	56	41	1	
FactInventory	5,951,174	1,714,784	1,386,841	557,768	9,610,567
FactOnlineSalesOrderDetail	5,528,466	4,463,166	4,796,456		14,788,088
FactOnlineSalesOrderHeader	748,865	756,090	405,349		1,910,304
FactCatalogSales				915,111	915,111
FactResellerSales	169,985	165,476	164,514		499,975
FactStoreSales	1,697,107	470,989	459,278		2,627,374
FactSalesQuota	3,618,470	1,137,386	1,100,597	646,166	6,502,619
FactStrategyPlan	2,017,741	581,659	398,690	35,479	3,033,569

ATHENA
IT SOLUTIONS

Database Table Names & Row Counts

Dimension or Fact	Currency	Continent	File
Product Cost	CNY	AS	cost_cny_step_1_of_4.txt
Product Cost	CNY	AS	cost_cny_step_2_of_4.txt
Product Cost	CNY	AS	cost_cny_step_3_of_4.txt
Product Cost	CNY	AS	cost_cny_step_4_of_4.txt
Product Cost	EUR	EU	cost_eur_step_1_of_4.csv
Product Cost	EUR	EU	cost_eur_step_2_of_4.csv
Product Cost	EUR	EU	cost_eur_step_3_of_4.csv
Product Cost	EUR	EU	cost_eur_step_4_of_4.csv
Product Cost	USD	NA	cost_usd_steps_all.xlsx
Product Price	CNY	AS	price_cny_step_1_of_4.txt
Product Price	CNY	AS	price_cny_step_2_of_4.txt
Product Price	CNY	AS	price_cny_step_3_of_4.txt
Product Price	CNY	AS	price_cny_step_4_of_4.txt
Product Price	EUR	EU	price_eur_step_1_of_4.csv
Product Price	EUR	EU	price_eur_step_2_of_4.csv
Product Price	EUR	EU	price_eur_step_3_of_4.csv
Product Price	EUR	EU	price_eur_step_4_of_4.csv
Product Price	USD	NA	price_usd_steps_all.xlsx
Currencies		All	DimCurrency.csv
Exchange Rates		All	FactExchangeRate.xlsx

ATHENA
IT SOLUTIONS

Team Project
Data Sourcing Requirements

ATHENA
IT SOLUTIONS

Deliverables:

1. DW has dimensional model to support:
   - Sales analysis
   - Inventory analysis
   - Sales quota & planning analysis

2. DI -- Load data sources into DW
   - SOR: flat files, SQL Server, PostgreSQL & MySQL, Oracle
   - DW: SQL Server
   - ETL: Talend

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

BI Tools:

- Tableau

- Qlik
  - Qlik Sense

- Microsoft BI
  - PowerBI

Deliverables:

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

ATHENA
IT SOLUTIONS

BI Team Project:
Deliverables - DI

<u>Data Integration Tools:</u>

- Talend Enterprise Data Integrator to DW (MySQL)

- Microsoft SSIS to DW (SQL Server)

<u>Deliverables:</u>

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

- <u>Note:</u>
  - o Follow project standards

ATHENA
IT SOLUTIONS

Team Project

BI Requirements

ATHENA
IT SOLUTIONS

- **Data Subjects:**
  - Online Sales (FactOnlineSales)
  - Sales (FactStoreSales)
  - Inventory (FactInventory)
  - Sales Quota
  - Strategy Plan

- **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

- Fact: FactOnline Sales

- Dimensions: Customer, Store, Product, Promotion, Sales Territory, Geo – Sales (or Store) & Customer

- Measures: Sales, Cost, Profit, Profit Margin, etc.

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

Online Sales Analysis

- Product sales analysis
  - Contribution, Ranked (Top x)
    - Product Hierarchy, Brand
- Sales by time
  - Period over Period analysis
  - Trends
- Customer sales analysis
  - B2C - demographics such as education, income, etc. & geo
  - B2B – by annual sales, geo
- Sales Analysis by Geography
- Promotion Analysis
  - Impact of sales & margin with promotions

ATHENA
IT SOLUTIONS

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

Overall Sales Analysis

- Fact: FactStoreSales
- Dimensions: Channel, Store, Product, Promotion, Sales Territory, Geo – Sales (or Store)
- Measures: Sales, Cost, Profit, Profit Margin, etc.

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

Sales Analysis

- Channel sales analysis
  - Contribution
  - Trends
- Product sales analysis
  - Contribution, Ranked (Top x)
    - Product Hierarchy, Brand
- Sales by time
  - Period over Period analysis
  - Trends
- Sales Analysis by Geography
- Promotion Analysis
  - Impact of sales & margin with promotions

ATHENA
IT SOLUTIONS

Overall Sales Analysis

- Fact: FactStoreSales

- Dimensions: Channel, Store, Product, Promotion, Sales Territory, Geo – Sales (or Store)

- Measures: Sales, Cost, Profit, Profit Margin, etc.

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

Reject Analysis

- Fact: FactSales, FactOnlineSales, FactInventory

- Analysis by Region SOR and/or by Source Table
  - Row counts & type of rejects

ATHENA
IT SOLUTIONS

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

Sales Quota & Strategy Plan Analysis

- Fact: FactSalesQuota, FactStrategyPlan, FactSales
- Scenario: Budget, Forecast, Actual (from FactSales)

- Analysis
  - Compare Budget, Forecast & Plan
    - By product or Product Category
    - By store & geography

ATHENA
IT SOLUTIONS

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

ETL Load Performance - Talend

- Table Row Counts

- Time to load entire DW