Final Team Project Fall 2019

- Retail DW & BI Project

BUSINESS INTELLIGENCE GUIDEBOOK

From Data Integration to Analytics

RICK SHERMAN

FOREWORD BY **CLAUDIA IMHOFF**
PRESIDENT OF INTELLIGENT SOLUTIONS, INC.

MORGAN KAUFMANN

ATHENA IT SOLUTIONS

Final Team Project

- Sourcing Requirements

ATHENA
IT SOLUTIONS

Retail Company SOR

Business Unit	SORs	Database	Database Filename
North America	Retail_SOR_NA	Oracle	retail_sor_na_sqlserver.zip retail_sor_na_orcl_20191129.sql migrate_na_mssql_orcl_all_v2.zip
Europe	Retail_SOR_EU	SQL Server	retail_sor_eu_sqlserver.zip
Asia	Retail_SOR_AS	MySQL	retail_sor_as_mysql.zip
Catalog	Retail_SOR_CAT	PostgreSQL	retail_sor_cat_post.backup
Quota & Plan	Retail_SOR_PLAN	PostgreSQL	retail_sor_plan_post.backup
Cross-unit	Flat Files	OneDrive \Final Project\SOR_Files	

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

- Continent oriented SORs contain:
  - Sales – Store, Reseller & Online Channels data
- The Catalog SOR contain:
  - Sales – Catalog Channel data from all three continents
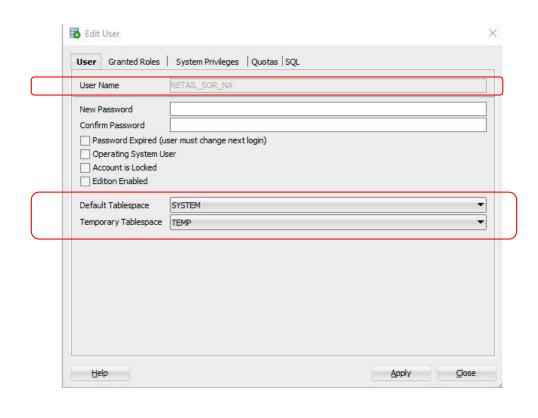
ATHENA
IT SOLUTIONS

Oracle Data Source

- Restore SQL Server backup: retail_sor_na_sqlserver.zip
- Create retail_sor_na schema in Oracle
- Create tables in above: retail_sor_na_orcl_20191129.sql
- Import archive file into Talend: migrate_na_mssql_orcl_all_v2.zip
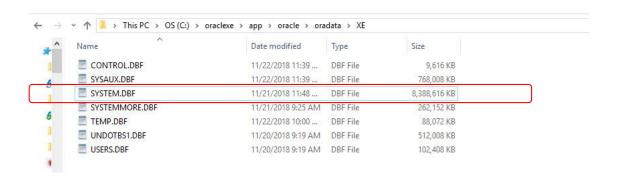- Run Table jobs above to import retail_sor_na from SQL Server into Oracle

ATHENA
IT SOLUTIONS

Retail_SOR_NA – SQL Server to Oracle Migration
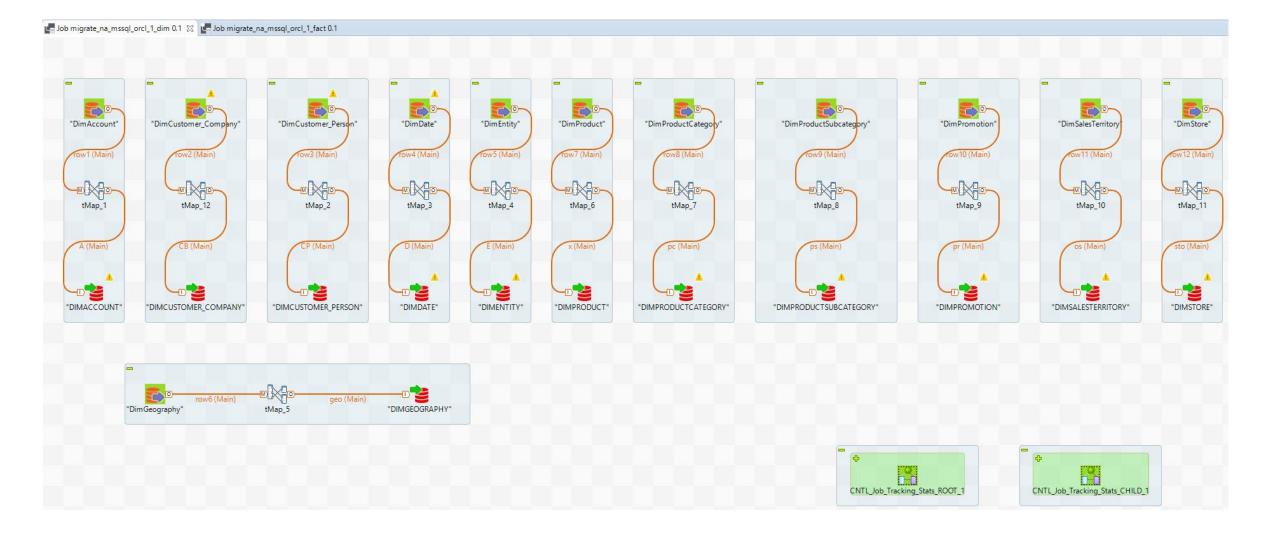
Need to increase size of SYSTEM tablespace:

ALTER DATABASE DATAFILE
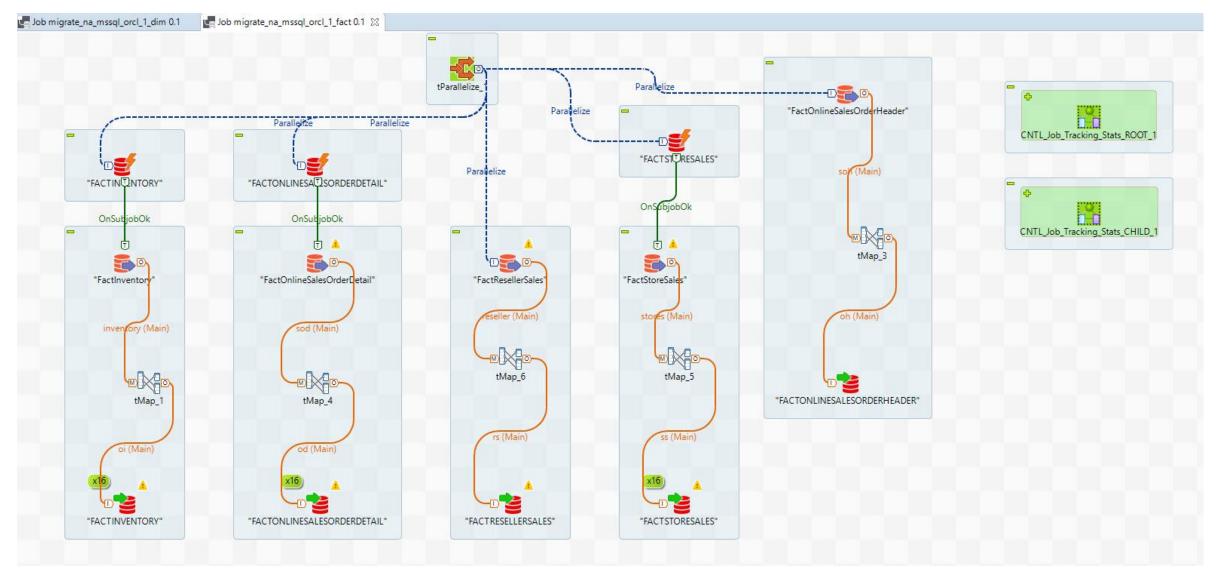'C:\oraclexe\app\oracle\oradata\XE\system.dbf'
RESIZE 8g;

Retail_SOR_NA – SQL Server to Oracle Migration

ATHENA
IT SOLUTIONS

Retail_SOR_NA – SQL Server to Oracle Migration

Final Team Project

- Overview

ATHENA
IT SOLUTIONS

Overview

- 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

ATHENA
IT SOLUTIONS

Retail Company

- 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

Retail Company

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

- Three top-level geographic regions:
    - North America
    - Europe
    - Asia

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

Retail Company – SOR Table Row Counts

TableName	retail_sor_eu	retail_sor_na	retail_sor_as	retail_sor_cat	retail_sor_plan	Totals	Distinct Count
dimchannel					4	4	4
dimcurrency					28	28	28
dimscenario					3	3	3
factsalesquota					5,856,453	5,856,453	5,856,453
factstrategyplan					1,998,355	1,998,355	1,998,355
dimaccount	24	24	24	24	24	120	24
dimcustomer_company	43	276	67			386	386
dimcustomer_person	5,505	9,395	3,593	9,395		27,888	18,493
dimdate	45,656	45,656	45,656	45,656	45,656	228,280	45,656
dimentity	421	421	421	421	421	2,105	421
dimgeography	147	303	67	303	517	1,337	517
dimproduct	2,517	2,517	2,517	2,517	2,517	12,585	2,517
dimproductcategory	8	8	8		8	32	8
dimproductsubcategory	44	44	44			132	44
dimpromotion	19	10	10	10		49	39
dimsalesterritory	53	265	29	183		530	265
dimstore	56	209	41	1	306	613	306
factcatalogsales				632,781		632,781	632,781
factinventory	1,745,094	5,917,637	1,411,894	138,560		9,213,185	9,213,185
factonlinesalesorderdetail	12,702,836	15,454,238	12,851,807			41,008,881	41,008,881
factonlinesalesorderheader	2,352,402	2,750,852	1,302,057			6,405,311	6,405,311
factresellersales	399,333	441,054	394,077			1,234,464	1,234,464
factstoresales	1,139,890	4,456,654	1,092,629			6,689,173	6,689,173
Totals	18,394,048	29,079,563	17,104,941	829,851	7,904,292	73,312,695	73,107,282
Database	sql server	oracle	mysql	postgresql	postgresql		

Note: Fact tables may have reject rows based on referential integrity violations.

ATHENA
IT SOLUTIONS

Systems of Record (SORs)

- SORs have prices, costs & sales in "continent" currency
  - North America & Catalog Operations – US Dollars (USD)
  - Europe - Euro (EUR)
  - Asia - China Yuan (CNY)
  - Quota & Plan has data in local or "continent" currency
- DW has prices, costs & sales:
  - In USD
  - In "local" currency used in SOR
- Prices, costs & sales 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

Systems of Record (SORs)

- Unit Prices & Costs were changed during 2012-2016

- You need to create SCD dimension for both Unit Price & Unit Cost Dimensions.
    - Step 1 – initial unit prices or costs
    - Step 2-n – prices or costs revised

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

ATHENA
IT SOLUTIONS

Systems of Record (SORs)

- DW sales related tables
  - FactSales – contains data for all 4 sales channels
  - FactOnlineSales – contains Online sales channel only and has customer data
- Sales Quotas & Strategy Plan:
  - Have data for three scenarios
    - Actuals
    - Budget – from SORs directly
    - Forecast – from SORs directly
  - Data granularity is monthly
  - For Sales Quota need to aggregate actual sales data to populate Actuals scenario
  - For Strategy Plan need to aggregate data for actuals in 2 accounts:
    - AccountKey = 4 'Sale Revenue'   - sales
    - AccountKey = 5 'Cost of Goods Sold' – total product cost

ATHENA
IT SOLUTIONS

Final Team Project

- Deliverables

- Grading Criteria

Deliverables

- DW has dimensional model to support:
  - Sales analysis
  - Inventory analysis
  - Sales quota & planning analysis
- DI -- Load data sources into DW
  - SOR: see data sources
  - DW:
    - Even # teams: Target is SQL Server - retail_dw - Create Schema - 2019 – sql server.sql
    - Odd # teams: Target is MySQL - retail_dw - Create Schema - 2019 - my sql.sql
  - ETL: Talend

ATHENA
IT SOLUTIONS

Deliverables

- BI -- Develop BI Dashboards, Reports & Visualizations analysis
  - Sales analysis
  - Inventory analysis
  - Sales quota & planning analysis

ATHENA
IT SOLUTIONS

Deliverables - BI

BI Tools:

- Tableau

- Microsoft BI
  - PowerBI

Deliverables:

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

Deliverables - BI

Note:

Sample Retail_DW databases have been provided to start working on BI while developing DI code

- retail_dw_test_sqlserver.zip
- retail_dw_Test_2019_03_mysql.sql

ATHENA
IT SOLUTIONS

Deliverables - DI

- Use Talend Enterprise Data Integrator to DW (TBD)

- Load DW from data sources

  o Document all jobs

  o Provide load statistics

  o Provide analysis of load jobs using one of BI tools used in this project

- Handle data quality & error messages

  o Document error handling results

- Load rejection records

  o Track reasons for rejections

  o Provide analysis of rejections & reasons using one of BI tools used in this project

- Note: Follow project standards

ATHENA
IT SOLUTIONS

Deliverables – Data Integration

- Table Row Counts
  - Total Rows - Dimensions, Facts & Reject Tables
  - Row counts by region (SOR source) - Dimensions, Facts & Reject Tables
  - Row counts by reject reason – Fact Tables

- Sales Amounts (USD)
  - By Region
  - By Channel (Internet versus Store)

- Time to load DW
  - Note: Need one root job that loads all sources – databases and files

ATHENA
IT SOLUTIONS

Deliverables – Analysis

- Data Subjects:
    - Customer Sales - Online & Catalog both have customers
    - Sales - All four sales channels
    - Inventory
    - Sales Quota – Budget & Forecast provided, need to generate Actuals
    - Strategy Plan – Budget & Forecast provided, need to generate Actuals
    - Reject Analysis

ATHENA
IT SOLUTIONS

Deliverables – Analysis

Data Subjects	Even # Teams	Odd # Teams
Customer Sales	Power BI	Tableau
Sales Inventory	Tableau	Power BI
Reject Analysis	Power BI	Tableau
Sales Quota	Tableau	Power BI
Strategy Plan	Tableau	Power BI

ATHENA
IT SOLUTIONS

Deliverables – Analysis

- 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.

ATHENA
IT SOLUTIONS

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

Overall Sales Analysis

- Fact: FactStoreSales

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

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

ATHENA
IT SOLUTIONS

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

Overall Sales Analysis

- Fact: FactStoreSales

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

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

ATHENA
IT SOLUTIONS

Reject Analysis

- Fact: FactSales, FactOnlineSales, FactInventory, FactSalesQuota, FactStrategyPlan

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

ATHENA
IT SOLUTIONS

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

Grading Criteria

- Grading Weight:
  - Data Integration: 65%
  - BI: 35%

- Data Loaded – all tables, row counts correct
  - 1 year
  - 3 years
  - 5 years

- Data Integration
  - Loaded in one orchestrator or master job
  - Correctly understand loading of fact tables
  - Correctly process rejects
  - Correctly handle currency
  - Correctly process SCDs

- BI Data Visualizations
  - Data History analyzed
    - 1 year
    - 3 years
    - 5 years
  - Data Visualizations answer business questions
  - Data Visualizations that follow best practices discussed in course

- Bonuses:
  - Faster 3 jobs
  - Most effective data visualizations for business analysis

ATHENA
IT SOLUTIONS