

# Final Team Project

## Fall 2018

- Retail DW & BI Project



# BUSINESS INTELLIGENCE GUIDEBOOK

From Data Integration to Analytics

RICK SHERMAN



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

# SOR Retail Company SOR - FALL 2018

Database	SOR Target	Backup Files
Retail_SOR_NA	Oracle	retail_sor_na_sqlserver.zip (use SQL Server backup & run Talend) Talend_Migrate_RETAIL_SOR_NA_MSSQL_to_ORCL.zip
Retail_SOR_EU	SQL Server	retail_sor_eu_sqlserver.zip (SQL Server)
Retail_SOR_AS	MySQL	retail_sor_as_mysql.zip
Retail_SOR_CAT	Azure SQL	retail_sor_cat_azure.bacpac
Retail_SOR_PLAN	PostgreSQL	retail_sor_plan_postgresql.sql
Flat Files		

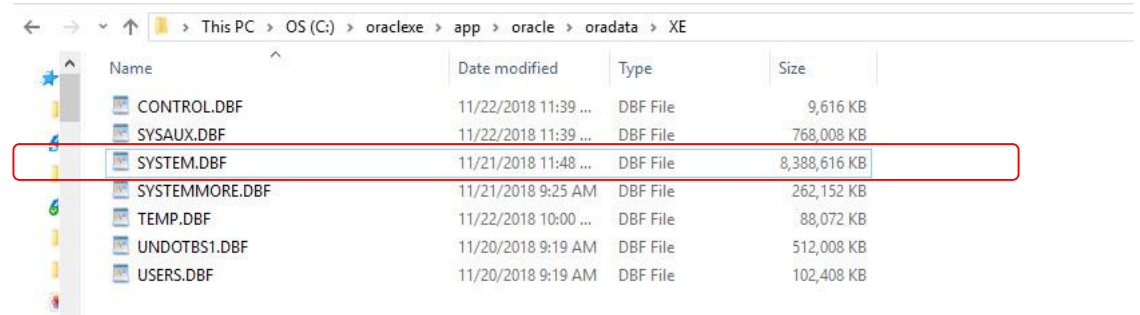
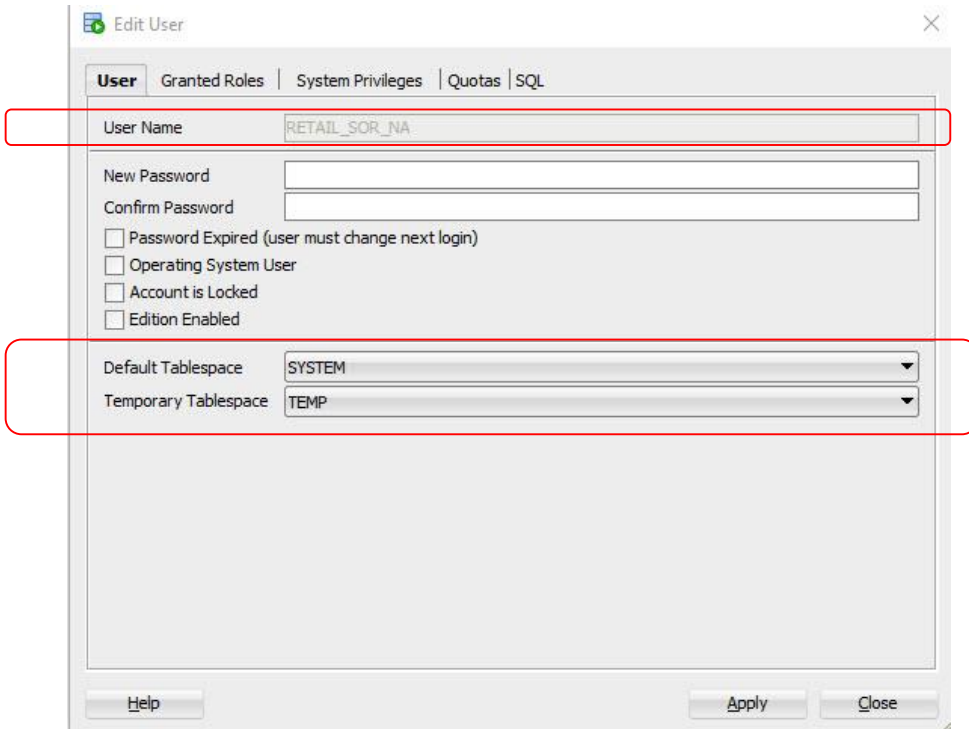
# Retail\_SOR\_NA – SQL Server to Oracle Migration

Need to increase size of SYSTEM tablespace:

ALTER DATABASE DATAFILE

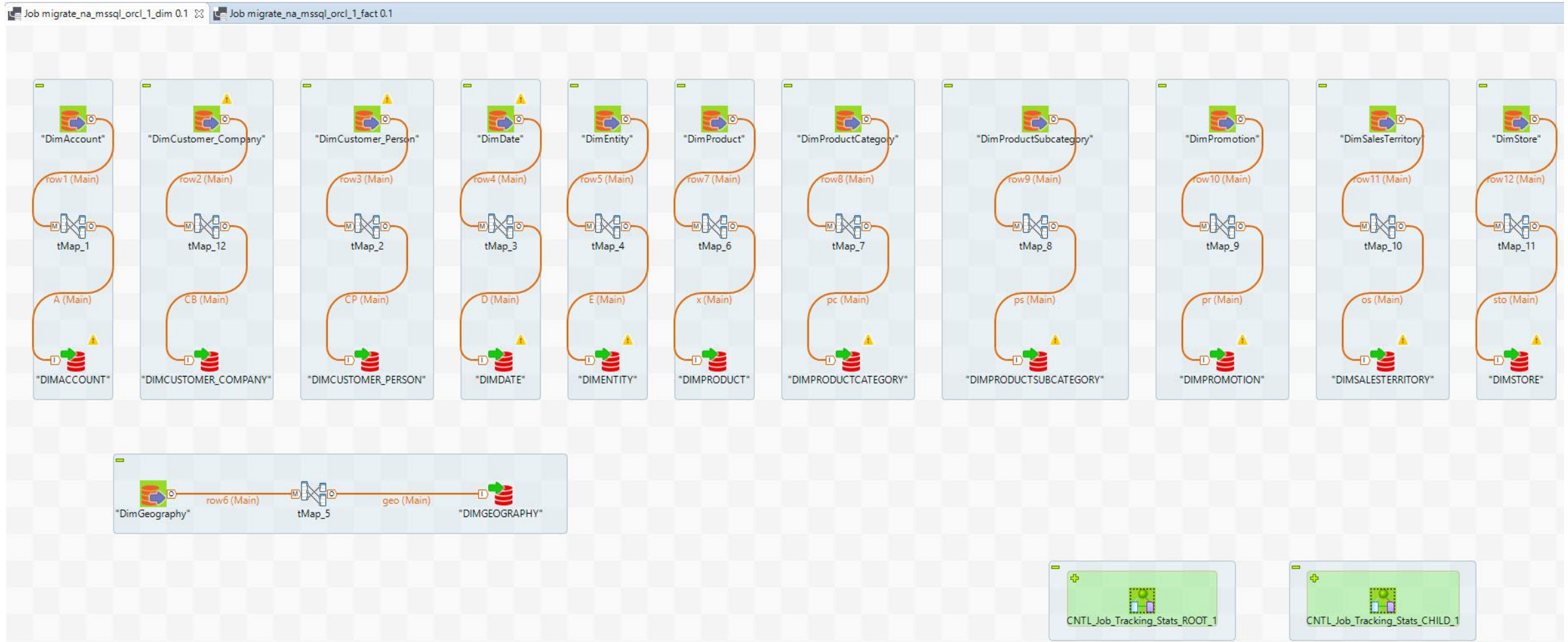
'C:\oraclexe\app\oracle\oradata\XE\system.dbf'

RESIZE 8g;



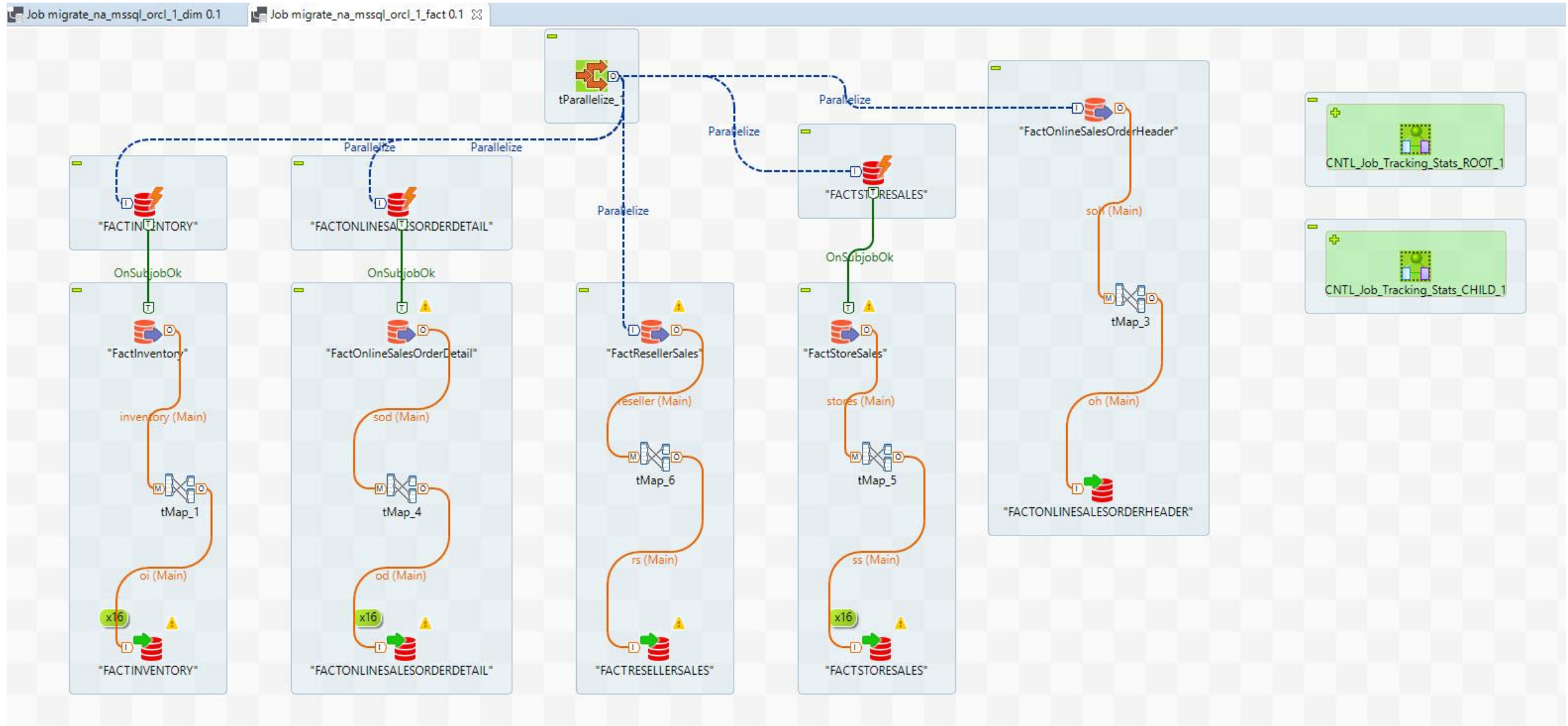
Name	Date modified	Type	Size
CONTROL.DBF	11/22/2018 11:39 ...	DBF File	9,616 KB
SYSAUX.DBF	11/22/2018 11:39 ...	DBF File	768,008 KB
SYSTEM.DBF	11/21/2018 11:48 ...	DBF File	8,388,616 KB
SYSTEMMORE.DBF	11/21/2018 9:25 AM	DBF File	262,152 KB
TEMP.DBF	11/22/2018 10:00 ...	DBF File	88,072 KB
UNDOTBS1.DBF	11/20/2018 9:19 AM	DBF File	512,008 KB
USERS.DBF	11/20/2018 9:19 AM	DBF File	102,408 KB

# Retail\_SOR\_NA – SQL Server to Oracle Migration



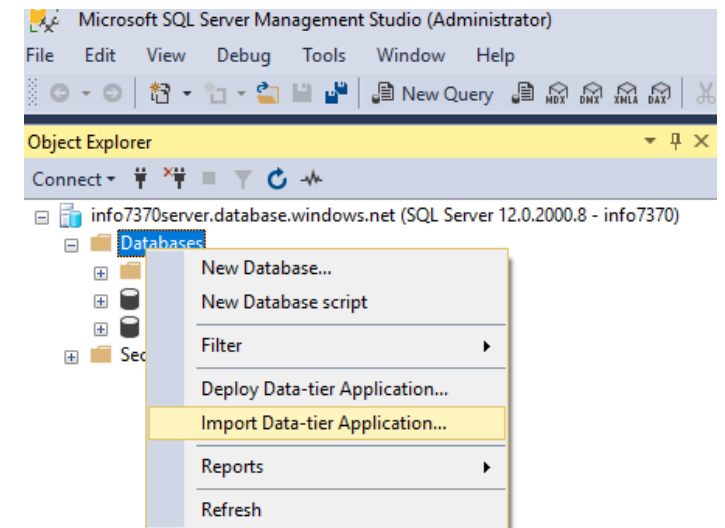
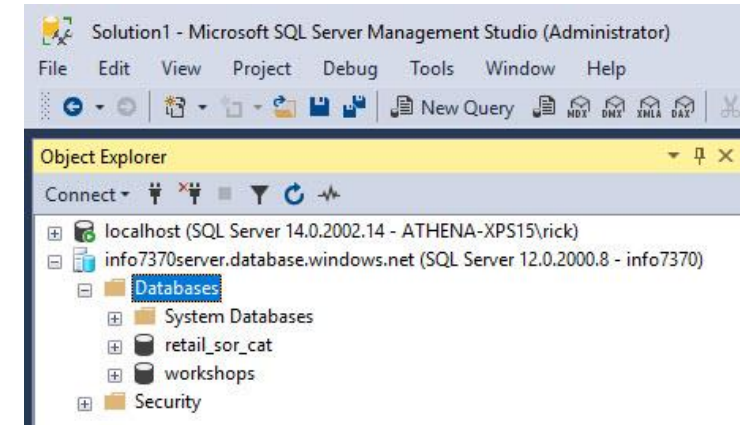


# Retail\_SOR\_NA – SQL Server to Oracle Migration

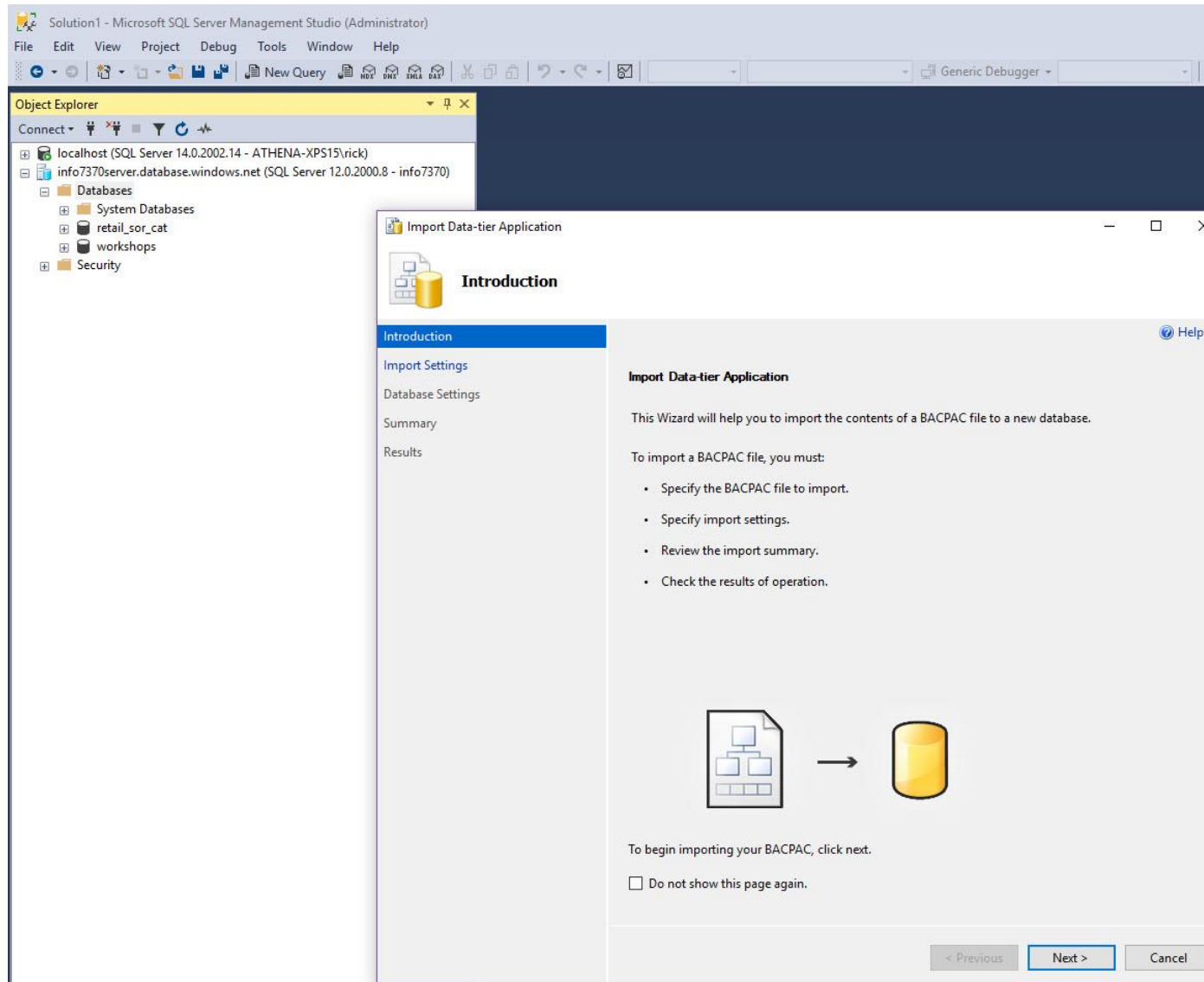


# Retail\_SOR\_CAT – Azure SQL

- Azure SQL Backup - retail\_sor\_cat\_azure.bacpac
- Login to your Azure SQL server
- Right click on Databases & select Import Data-Tier Application

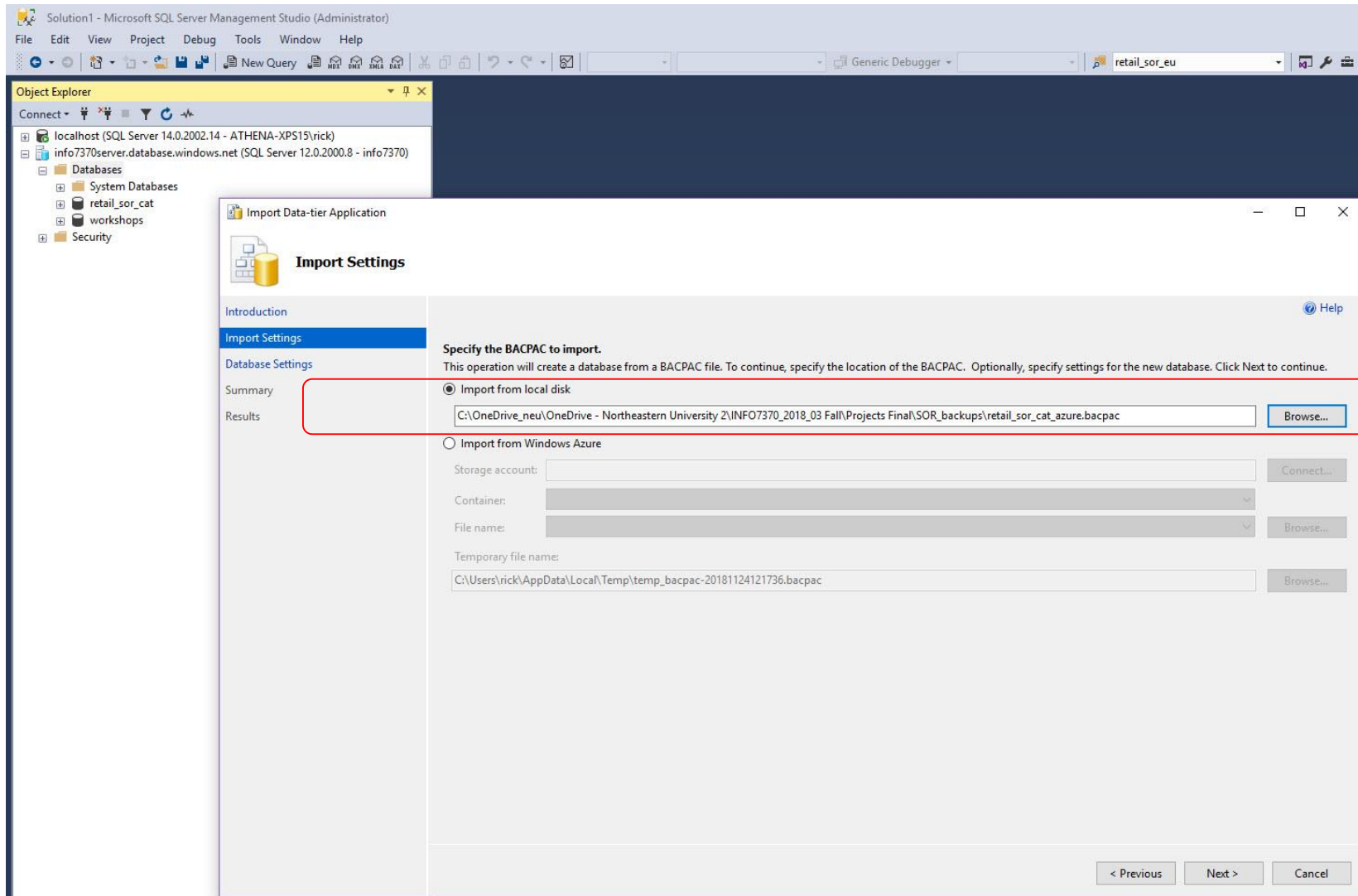


# Retail\_SOR\_CAT – Azure SQL



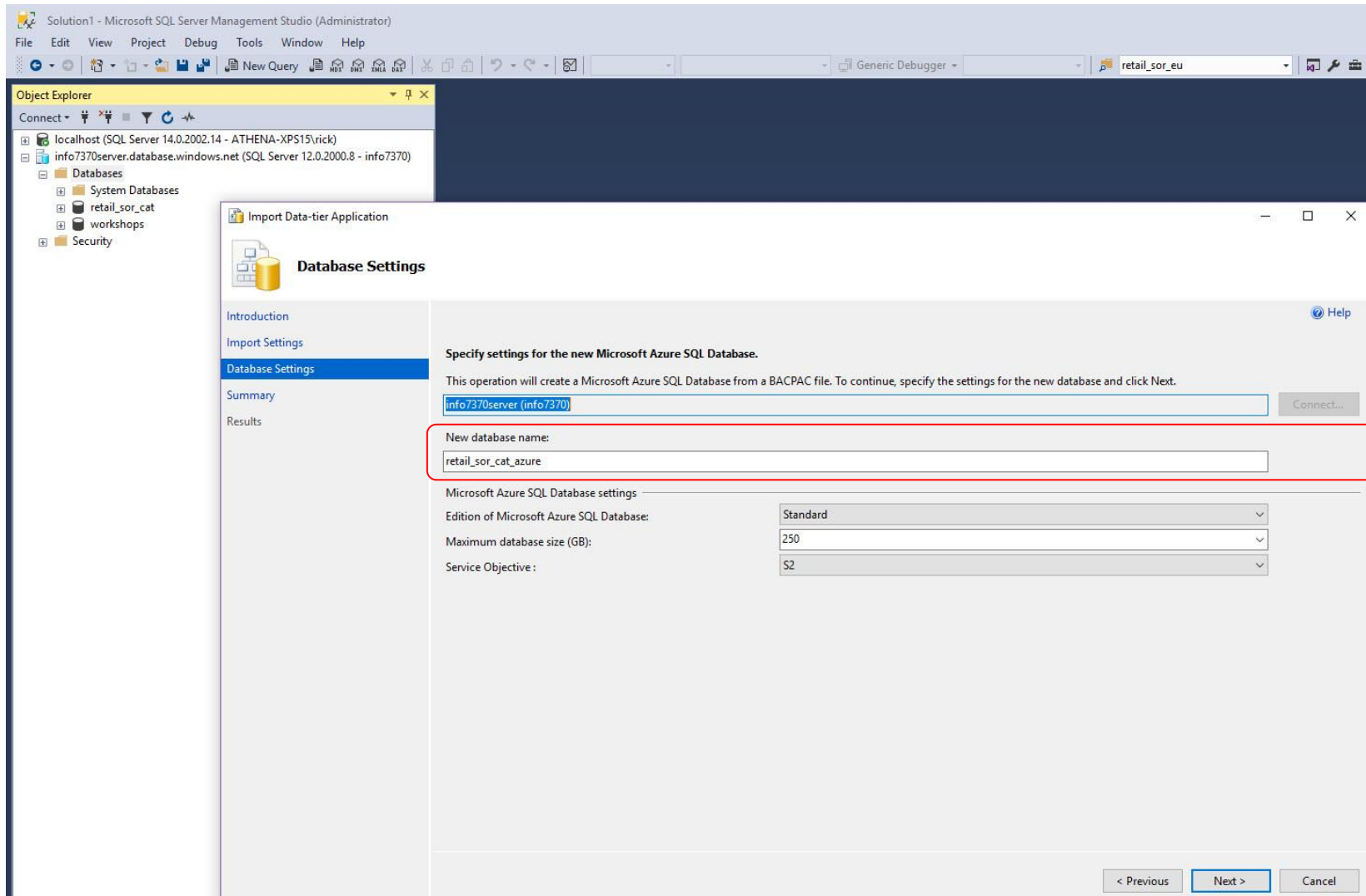
# Retail\_SOR\_CAT – Azure SQL

- The Azure backup file is on OneDrive. Use, of, course, whatever your file location is.



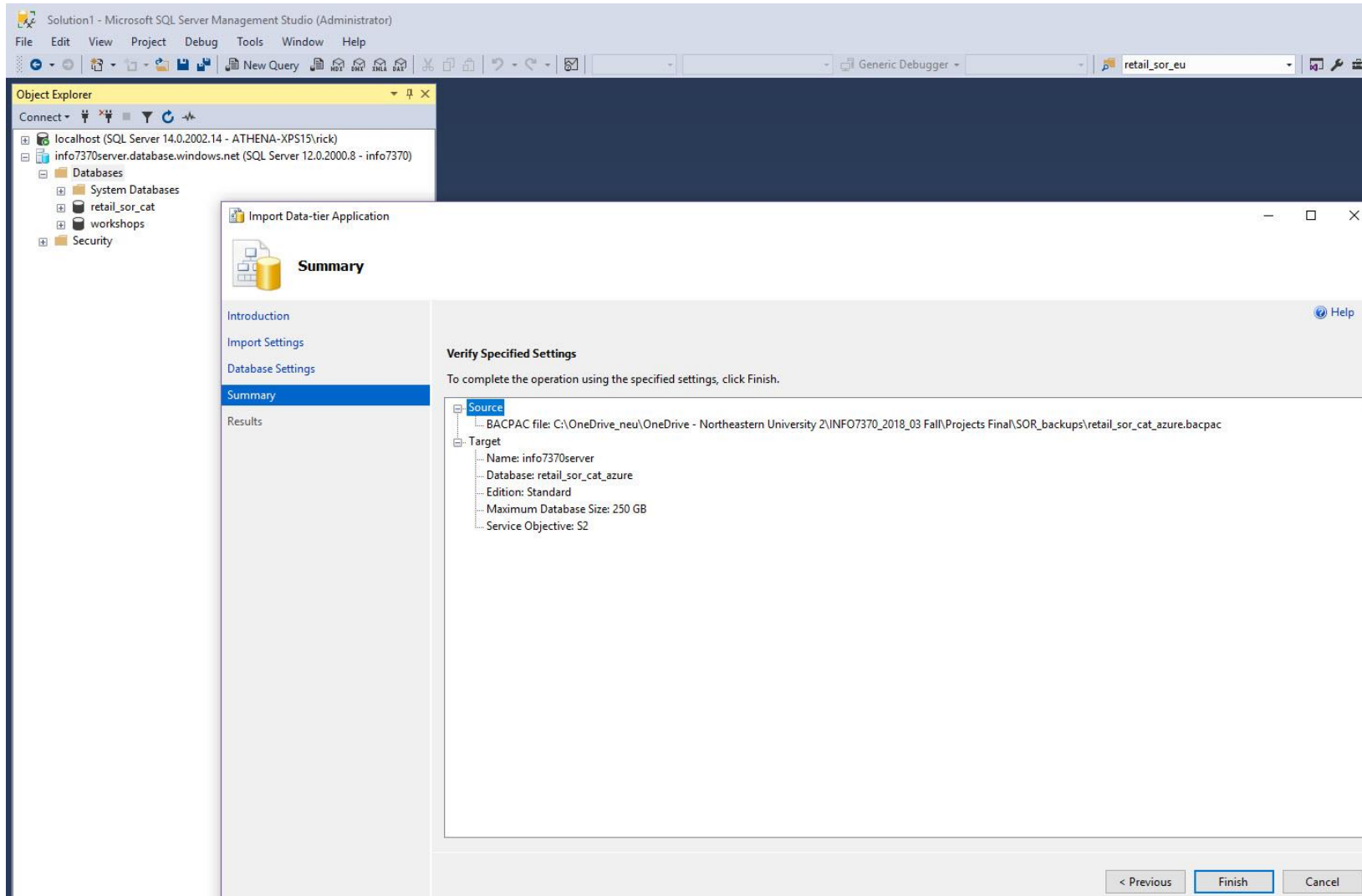


# Retail\_SOR\_CAT – Azure SQL



- Input a new database name.
- Use Retail\_SOR\_CAT (note: I added suffix \_Azure because I already had this db created & loaded via Talend)

# Retail\_SOR\_CAT – Azure SQL



- Select Finish
- You should see you db deployed on Azure Portal

# Retail\_SOR\_CAT – Azure SQL

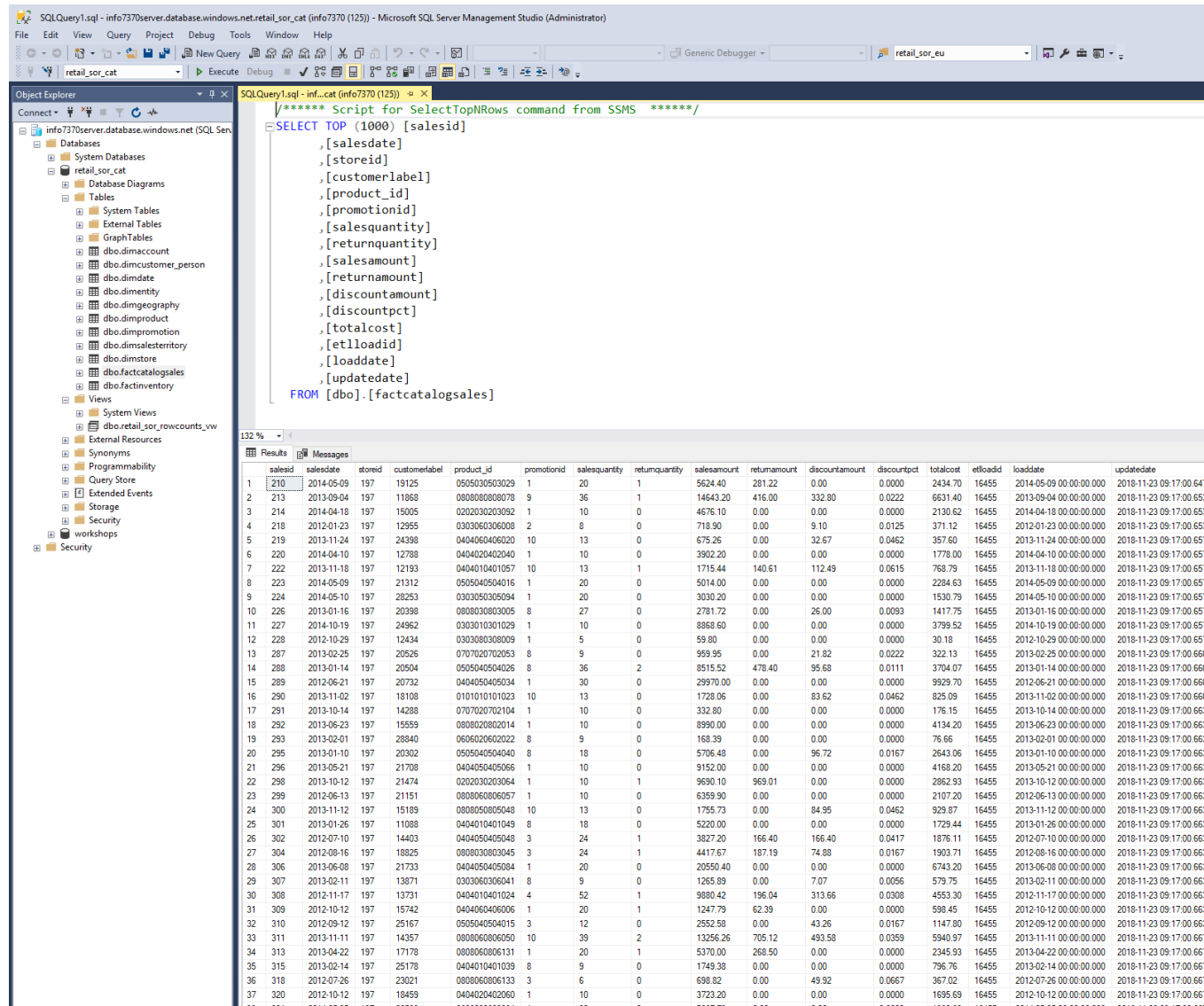
The screenshot displays the Microsoft Azure portal interface for the 'retail\_sor\_cat' SQL database. The left sidebar shows navigation options like 'Create a resource', 'All services', and 'FAVORITES'. The central pane shows the 'retail\_sor\_cat' database with an object explorer listing tables, views, and stored procedures. The right pane shows the 'Query editor (preview)' with a query: `1 select top 100 * from factcatalogsales`. Below the query editor, the 'Results' tab displays a table with columns: SALESID, SALESDATE, STOREID, CUSTOMERLABEL, PRODUCT\_ID, PROMOTIONID, and SALESQUANTITY. The table contains 100 rows of data.

SALESID	SALESDATE	STOREID	CUSTOMERLABEL	PRODUCT_ID	PROMOTIONID	SALESQUANTITY
210	2014-05-09T00:00:00.0000000	197	19125	0505030503029	1	20
213	2013-09-04T00:00:00.0000000	197	11868	0808080808078	9	36
214	2014-04-18T00:00:00.0000000	197	15005	0202030203092	1	10
218	2012-01-23T00:00:00.0000000	197	12955	0303060306008	2	8
219	2013-11-24T00:00:00.0000000	197	24398	0404060406020	10	13
220	2014-04-10T00:00:00.0000000	197	12788	0404020402040	1	10
222	2013-11-18T00:00:00.0000000	197	12193	0404010401057	10	13
223	2014-05-09T00:00:00.0000000	197	21312	0505040504016	1	20
224	2014-05-10T00:00:00.0000000	197	28253	0303050305094	1	20
226	2013-01-16T00:00:00.0000000	197	20398	0808030803005	8	27
227	2014-10-19T00:00:00.0000000	197	24962	0303010301029	1	10
228	2012-10-29T00:00:00.0000000	197	12434	0303080308009	1	5
287	2013-02-25T00:00:00.0000000	197	20526	0707020702053	8	9

- Azure Portal – work with SQL db

# Retail\_SOR\_CAT – Azure SQL

- SSMS connected to Azure Server & databases is alternative to Azure Portal



The screenshot displays the Microsoft SQL Server Enterprise Manager (SSMS) interface. The left pane shows the 'Object Explorer' with the 'retail\_sor\_cat' database selected. The central pane shows a query window with the following SQL script:

```
***** Script for SelectTopNRrows command from SSMS *****  
SELECT TOP (1000) [salesid]  
,[salesdate]  
,[storeid]  
,[customerlabel]  
,[product_id]  
,[promotionid]  
,[salesquantity]  
,[returnquantity]  
,[salesamount]  
,[returnamount]  
,[discountamount]  
,[discountpct]  
,[totalcost]  
,[etlloadid]  
,[loaddate]  
,[updatedate]  
FROM [dbo].[factcatalogsales]
```

The right pane shows the 'Results' grid with 15 columns: salesid, salesdate, storeid, customerlabel, product\_id, promotionid, salesquantity, returnquantity, salesamount, returnamount, discountamount, discountpct, totalcost, etlloadid, loaddate, and updatedate. The grid displays 37 rows of data, with the first row having salesid 210 and salesdate 2014-05-09.

# Final Team Project

## Retail DW



# 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

# 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

These  
numbers  
will change

Table	retail_sor_eu	retail_sor_as	retail_sor_na	retail_sor_cat	retail_sor_plan	Worlwide
DimAccount	24	24	24	24		24
DimChannel						-
DimCustomer_Company	43	67	276			386
DimCustomer_Person	5,505	3,593	9,395	9,395		18,493
DimDate	45,656	45,656	45,656	45,656		45,656
DimEntity	421	421	421	421		421
DimGeography	147	67	303	303		517
DimProduct	2,517	2,517	2,517	2,517		2,517
DimProductCategory	8	8	8			8
DimProductSubcategory	44	44	44			44
DimPromotion	19	10	10	10		39
DimSalesTerritory	53	29	183	1		265
DimScenario						-
DimStore	56	41	209			306
FactInventory	1,745,148	1,410,895	5,912,445	143,512		9,212,000
FactOnlineSalesOrderDetail	4,463,166	4,796,456	5,528,466			14,788,088
FactOnlineSalesOrderHeader	756,126	405,523	748,650			1,910,299
FactSales						
FactCatalogSales				118,537		118,537
factresellersales	175,047	173,574	175,047			523,668
factstoresales	497,523	484,058	1,940,185			2,921,766
FactSalesQuota_All					5,856,453	5,856,453
FactStrategyPlan_All					2,998,090	2,998,090
Total	7,691,503	7,322,983	14,363,839	320,376	8,854,543	38,397,577

# 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



# 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 store in Sales related tables nor in the Product dimension

# 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

# Final Team Project

- Sourcing Requirements



# Deliverables

- DW has dimensional model to support:
  - Sales analysis
  - Inventory analysis
  - Sales quota & planning analysis
- DI -- Load data sources into DW
  - SOR: TBD
  - DW: TBD with ETL: Talend
  - DW: TBD with ETL: SSIS

# Deliverables

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



# 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

# Final Team Project

- Deliverables



# Deliverables - DI

This will be updated

- Use Talend Enterprise Data Integrator to DW (TBD)
- Use SSIS to DW (TBD)
- 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
- 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

# 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

# Deliverables – Analysis

***This will be updated***

- Data Subjects:

- Online Sales (FactOnlineSales)
  - Online & Catalog both have customers
- Sales (FactStoreSales)
  - All four sales channels
- 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



# Online Sales Analysis

***This will be updated***

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

# Online Sales Analysis

***This will be updated***

- 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

# Overall Sales Analysis

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

# 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

***This will be updated***

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

# Reject Analysis

***This will be updated***

- Fact: FactSales, FactOnlineSales, FactInventory
- Analysis by Region SOR and/or by Source Table
  - Row counts & type of rejects

# Sales Quota & Strategy Plan Analysis

***This will be updated***

- Fact: FactSalesQuota, FactStrategyPlan, FactSales
- Scenario: Budget, Forecast, Actual (from FactSales)
- Analysis
  - Compare Budget, Forecast & Plan
    - By product or Product Category
    - By store & geography