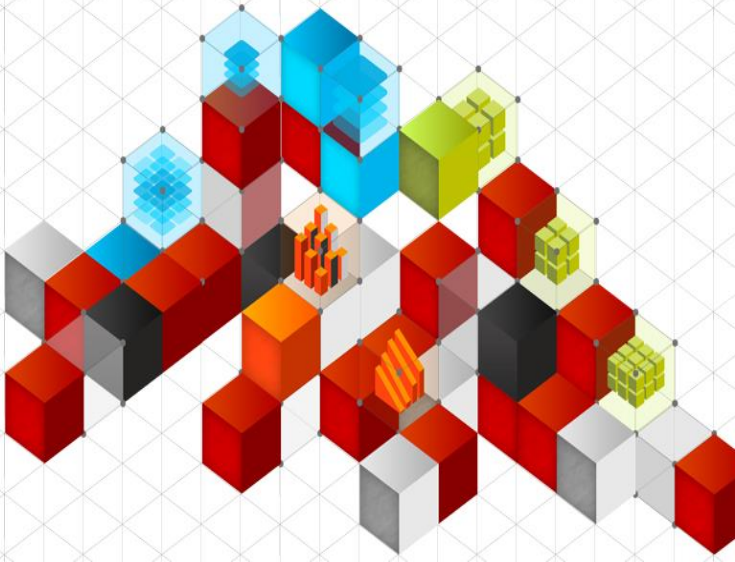
Microsoft Business Intelligence and Data Mining

Nigel Peter Sammy



TTSSUG PASS

http://www.nigelpsammy.com

Questions

- ❑ Familiar with Business Intelligence
- ❑ Familiar with Data Mining
- ❑ Familiar with SQL Server

Agenda

- Definitions
- Microsoft Tools
- Microsoft Data Mining

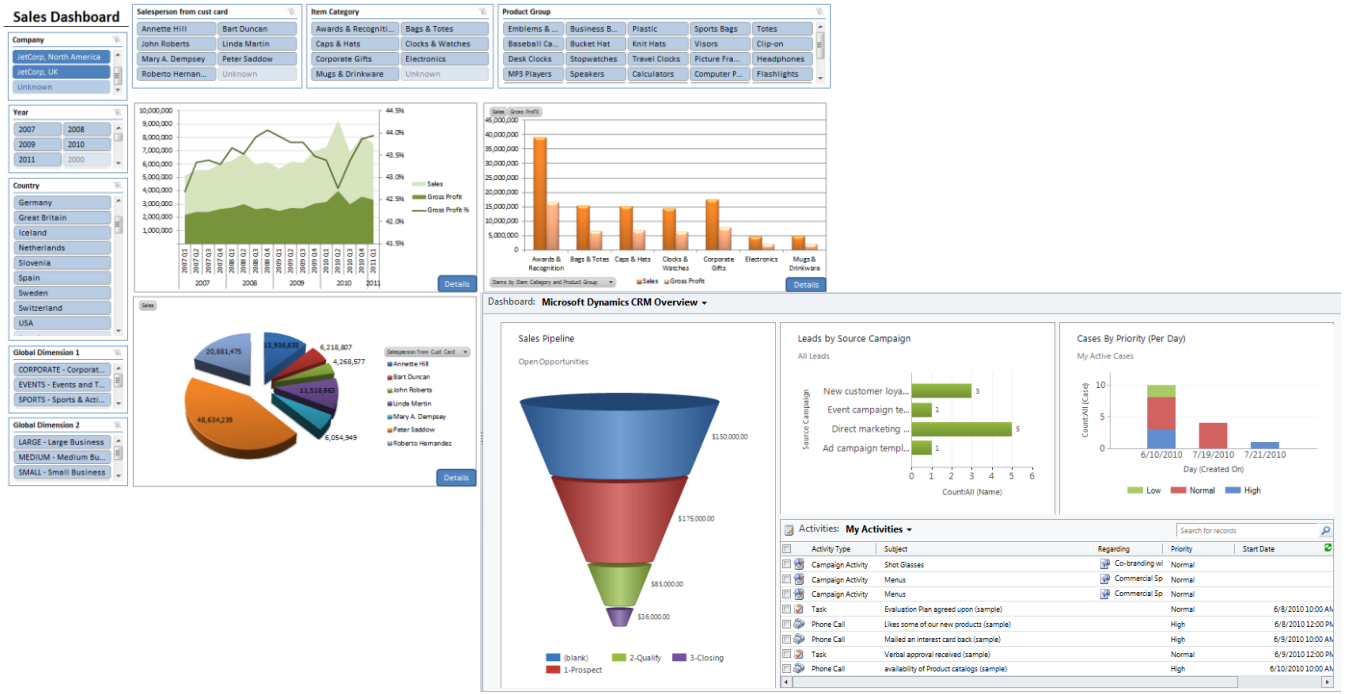
Definitions

OLTP

Online Transactional Processing

OLAP

Online Analytical Processing



OLTP vs OLAP

	OLTP System	OLAP System
Source of data	Operational data; OLTPs are the original source of the data.	Consolidation data; OLAP data comes from the various OLTP Databases
Purpose of data	To control and run fundamental business tasks	To help with planning, problem solving, and decision support
Inserts and Updates	Short and fast inserts and updates initiated by end users	Periodic long-running batch jobs refresh the data
Queries	Relatively standardized and simple queries Returning relatively few records	Often complex queries involving aggregations
Processing Speed	Typically very fast	Depends on the amount of data involved; batch data refreshes and complex queries may take many hours; query speed can be improved by creating indexes
Space Requirements	Can be relatively small if historical data is archived	Larger due to the existence of aggregation structures and history data; requires more indexes than OLTP
Database Design	Highly normalized with many tables	Typically de-normalized with fewer tables; use of star and/or snowflake schemas

Business Intelligence

Business intelligence, or BI, is an umbrella term that refers to a variety of software applications used to analyze an organization's raw data.

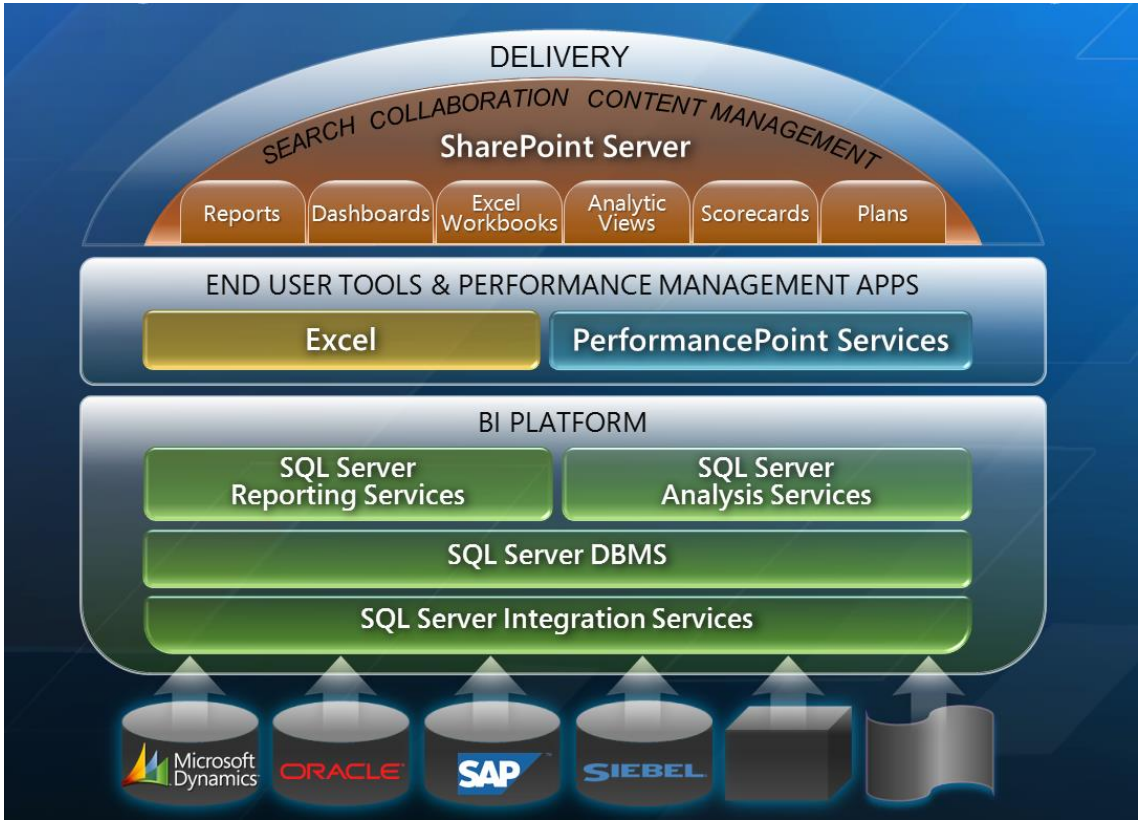
Data Mining

The process of extracting valid, previously unknown, comprehensible and actionable information from large databases and using it to make crucial business decisions

Microsoft Tools

TTSSUG PASS

Microsoft Business Intelligence Stack



SQL Server Components

Data Mining

- Exploring your data
- Finding patterns and rules
- Making predictions

Features

- Multiple data sources:
- Integrated data cleansing, data management, and ETL
- Multiple customizable algorithms
- Model testing infrastructure
- Querying and drill through
- Client tools
- Scripting language support and managed API
- Security and deployment

Data Mining Concepts

- **Forecasting**: Estimating sales, predicting server loads or server downtime

- **Risk and probability**: Choosing the best customers for targeted mailings, determining the probable break-even point for risk scenarios, assigning probabilities to diagnoses or other outcomes

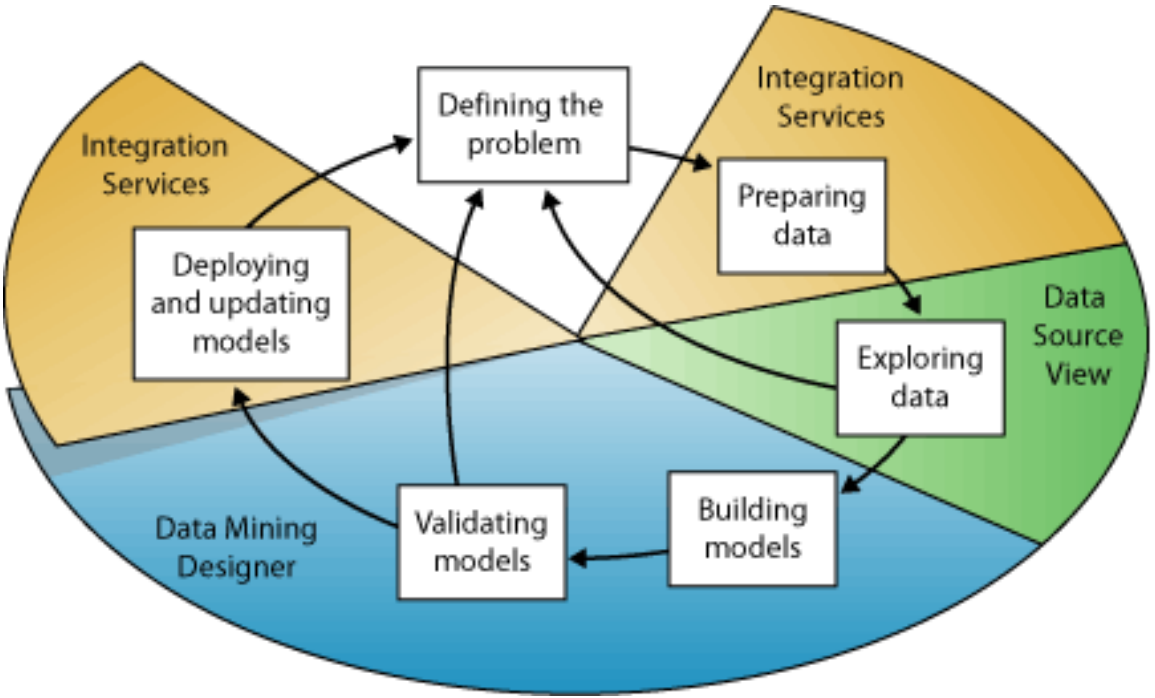
- **Recommendations**: Determining which products are likely to be sold together, generating recommendations

Data Mining Concepts

- **Finding sequences**: Analyzing customer selections in a shopping cart, predicting next likely events

- **Grouping**: Separating customers or events into cluster of related items, analyzing and predicting affinities

Data Mining Model



Algorithms

- Microsoft Association Algorithm
- Microsoft Clustering Algorithm
- Microsoft Decision Trees Algorithm
- Microsoft Linear Regression Algorithm
- Microsoft Logistic Regression Algorithm
- Microsoft Naive Bayes Algorithm
- Microsoft Neural Network Algorithm
- Microsoft Sequence Clustering Algorithm
- Microsoft Time Series Algorithm

DEMO

Contact Information

Follow us 

Contact us: info@ttssug.org

Website: http://ttssug.org/

Personal

Email: nigelpsammy@live.com

Blog: www.nigelpsammy.com

TTSSUG PASS

Microsoft SQL Server 2012