Part 23

Data Mining
OLAP (On-Line Analytical Processing)

Oracle OLAP Products
- Oracle Express
- Oracle Express Web Agent
- Oracle Express Objects
- Oracle Express Analyzer
- Oracle Financial Analyzer
- Oracle Financial Controller
- Oracle Sales Analyzer

OLAP Uses
- Lets users play with information stored in the data warehouse
- Can view the data from a variety of angles
- Can analyze sales over time, by region, by product type
Categories of OLAP Tools

DOLAP - Desktop OLAP
  PC tools with limited analysis of small warehouses

ROLAP - Relational OLAP
  Server-based tools for savvy relational database users

MOLAP - Multidimensional OLAP
  Server-based tools for read-only pre-computed cubes of data
Star Schemas
Data Mining

Discover hidden patterns in data
   Also known as Knowledge Discovery in Databases (KDD)
   Pattern has not yet been identified by the end user

Discover anomalies in the data
   One pattern is inconsistent with other patterns
   Identify bad data, bad reporting, bad modeling, vs. real differences that need to be dealt with or exploited

Pattern:
   A pattern is an inherent relationship and/or distributions in the data under consideration

Data:
   Data is the collection stored in the Data Warehouse environment
Data Mining Initial Process

Problem identification
  Determine the scope of the problem
  Determine if the data is available to address the problem

Data preparation
  Data collection
  Data cleaning
  Data reduction, including removal of useless variables and sampling of large datasets
  Data transformation to a limited set of numerical values

Model formulation
  Select a model and technique for mining the data

Data analysis/Pattern exploration/Data classification
  Exact approach will depend on model
Data Mining Continuing Process

Validation and iteration
   Test model against unknown cases that were not used to generate the model
   Verify validity of model
   Modify as necessary
   Must be able to predict unusual cases

Knowledge acquisition
   Run model against all cases
   Generate final statistics

Interpretation
   Translate numbers into words
   Build a theory of behavior based on subject matter knowledge that is consistent with the facts

Implementation
   Develop a plan of action that is theoretically sound under the theory developed
Fields used by Data Mining

Statistics
Means, variances, normal distribution, regression (single and multiple, linear and non-linear), ANOVA

Artificial Intelligence
Pattern recognition, deductive reasoning, vision algorithms, expert systems, genetic algorithms

Databases
Clustering, pattern bit maps

Operations Research
Linear programming, Lagrangian multipliers, branch and bound, integer programming
Business Applications of Data Mining

Market segmentation
Market basket analysis
Fraud detection
Customer retention
Forecasting
Investment management
Manufacturing and production
Data Mining Tasks

Classification
Clustering
Association
Prediction
Summarization
Visualization
Data Mining Methods

Statistical regression
Neural networks
Decision trees
Multivariate statistical methods
Mathematical programming
Case-based reasoning
Genetic algorithms