COURSE SYLLABUS

Data Warehouse & ETL
(Overview)

50 Cragwood Rd, Suite 350
South Plainfield, NJ 07080

Victoria Commons, 613 Hope Rd Building #5,
Eatontown, NJ 07724

130 Clinton Rd,
Fairfield, NJ 07004
Avtech Institute of Technology Course

Instructor: 
Course Duration: 10 
Date/Time: 
Training Location:

Course: Data Warehouse Overview

Text / Lab Books

http://www.datawarehousing.com

Course Description

The data warehouse is best known for taking disparate and disordered transaction data and presenting it in a cohesive, orderly way for the business to make intelligent, fact-based decisions.

The data warehouse ETL designer is charged with the task of applying a set of consistent techniques for delivering conformed dimensional data. Precisely designing and building reusable processes to extract, clean, conform and deliver dimensional data is the foundation for a successful, reduced cost, data warehouse implementation.

Dimensional modeling is the proven technique for developing understandable, high-performance data warehouses and data marts. Dimensional analysis and design closes the gap between business requirements and traditional dimensional modeling. The rigorous and practical use of dimensional analysis and design improves productivity and communication between business users and IT by supporting incremental development and more fully capturing analytical requirements.

This course is appropriate for anyone involved or interested in learning the latest techniques for planning, designing and managing dimensional data warehouses and ETL processes. Beginner, intermediate and experienced data warehouse practitioners, data architects, DBA’s and ETL designers & developers will benefit from this course.

This course will give a general concept and view of Data Warehouse and ETL process, which including but not limit to: W/H, PLTP application, OLAP, ROLAP, MOLAP, HOLAP, O/S, EDI, designing tools to build a Data Warehouse, Database types, Modeling types, schemas, queries, Visio, Erwin, Business Process Management (BPM), and Transformation Manager (TM).
Learning Objectives

1.0 Business Process & Business Modeling
   1.1. Modeling Types
       1.1.1. Process Modeling
           1.1.1.1. FTL: Affordable, Fast ROI
           1.1.1.2. Master Data Management: Centralize master data across all BI, ERP, warehousing applications
       1.1.3. Data Modeling (Conceptual Data Modeling, Enterprise Data Modeling, Logical Data Modeling, Physical Data Modeling, Relational Data Modeling, Dimensional Data Modeling)
   1.2. Modeling Methods
       1.2.1. Business Modeling Methods: IDEF (acronym), IDEFO (methods)
       1.2.2. Process Modeling: Software, solutions and consulting Download Whitepapers
       1.2.3. Process Methodology: World-class process metrology
       1.2.4. 2007 Workflow Handbook: IDEF3, IDEF4
       1.2.5. Data Modeling Methods: IDEFIX, Dimensional Modeling Notation
   1.3. Business Process
       1.3.2. Different kinds of Business Processes: Product manufacturing, payment of insurance and taxes, HR and Recruitment, services calls in call centers, information stored by bank tellers, the transformation in data warehouse)
   1.4. Business Process Tools
       1.4.1. Smart BPM, Process modeling, Business Management Class and Business Process Modeling
   1.5. Business Process Management (BPM)
       1.5.1. Business Requirements
       1.5.2. B2B Process Integration
       1.5.3. Business Process Report
       1.5.4. Process Management
       1.5.5. Need for Business Process Management
   1.6. Advantages of BPM
       1.6.1. Process Modeling, BPM & Rules Whitepaper, Business Requirements, Reliable BPM Solution
   1.7. Business Process Re-engineering
   1.8. Business Process Modeling
       1.8.1. Business Requirements, Process modeling, and Workflow
   1.9. Business Process Tools

1.10. Business Process Modeling Example
1.10.2. Business Requirements, Smart Rules-Driven BPM, Business Process Mgmt

1.11. Process Flow Modeling
1.11.1. Business Process Report and Business Requirements

1.12. Data Flow Modeling
1.12.2. Activities or Processes, Data Flows or Arrows, Date Store, External References, Physical Resources

1.13. Work Flows
1.13.2. Advantages of Workflow

1.14. Business Activity Monitoring
1.14.2. Advantages
1.14.3. Workflow software

1.15. Sources of Modeling Data
1.16. Pre-Processing the Data
1.17. Alternative Modeling Strategies

2.0 Data Modeling (DM)
2.1. Overview
2.1.1. Easy ER Diagram Software (SmartDraw)
2.1.2. Database Modeling Tool (DeZign for Databases V4-datanamic)
2.1.3. ModelRight
2.1.4. Data Warehousing Video
2.2. Data Modeling Tools: Erwin, Embarcadero, Rational Rose, Power Designer, Oracle Designer, Xcase

2.3. Tools
2.3.1. Logical Data Model
2.3.2. Physical Data Model

2.4. DM Tools-Erwin
2.5. ED Tools-Xcase

2.6. Development Cycle
2.6.1. Gathering Business Requirements-First Phase
2.6.2. Conceptual Data Modeling (CDM)-Second Phase
2.6.3. Logical Data Modeling (LDM)-Third Phase
2.6.4. Physical Modeling (PDM)-fourth Phase
2.6.5. Database -Fifth

2.7. DM Standards
2.7.1. Standardization Needs/Modeling data
2.7.2. Table Names Standardization
2.7.3. Column Names Standardization
2.7.4. Database Parameters Standardization

2.8. Steps to Create a Data Model

2.9. Data Modeler Role
2.9.1. Business Requirement Analysis
2.9.2. Development of data model
2.9.3. Reports
2.9.4. Review
2.9.5. Creation of database
2.9.6. Support & Maintenance

2.10. Modeling Reports
2.10.1. Logical Data Model Report
2.10.2. Physical Data Model Report

2.11. Conceptual Data Modeling (CDM)
2.12. Logical Data Modeling (LDM): entities, attributes, key groups, and relationships

2.13. Physical DM

2.14. Logical vs. Physical
2.14.1. Logical DM
2.14.1.1. Represents business information and defines business rules
2.14.1.2. Entity, attribute, primary Key, Inversion Key Entry, Rule, Relationship, and Definition
2.14.2. Physical DM
2.14.2.1. Represents the physical implementation of the model in a database
2.14.2.2. Table, column, Primary Key Constraint, Unique Constraint or Unique Index, Non Unique Index, Check Constraint, Default Value, Foreign Key, Comment

2.15. Relational (OLTP) DM
2.15.1. Data is stored in RDBMS
2.15.2. Tables are units of storage
2.15.3. Data is normalized and used for OLTP. Optimized for OLTP processing
2.15.4. Several tables and chains of relationships among them
2.15.5. Volatile(several updates) and time variant
2.15.6. SQL is used to manipulate data
2.15.7. Detailed level of transactional data
2.15.8. Detailed level of transactional data

2.16. Relational vs. Dimensional
2.16.1. Dimensional DM:
2.16.2. Data is stored in RDBMS or Multidimensional databases
2.16.3. Cubes are units of storage
2.16.4. Data is denormalized and used in data warehouse and data mart. Optimized for OLAP
2.16.5. Few tables and fact tables are connected to dimensional tables
2.16.6. Non volatile and time invariant
2.16.7. MDX is used to manipulate data
2.16.8. Summary of bulky transactional data (Aggregates and Measures) used in business decisions
2.16.9. User friendly, interactive, drag and drop multidimensional OLAP Reports

2.17. Dimensions
2.17.1. Dimension Table
2.17.2. Location Dimension
2.18. Slowly Changing Dimensions
2.19. Supertype & Subtype
2.20. Creating a Data Model in Visio

3.0 Database & Data Modeling
3.1. Database Overview
3.1.1. Database Types: Database Management Systems (DBMS), Relational Database Management Systems (RDBMS), Object Oriented Databases, Multidimensional Databases
3.1.2. Often Used Databases (RDBMS) In Most Of The Practical Applications: Oracle, Sql Server, Informix, Terradata, DB2
3.1.3. Data Modeling Tool: For example, Erwin to generate DDL scripts from the Data Modeling tool
3.1.4. Oracle Database Objects: Instance, Schema, Table, Column, Datatype, Primary, Key Constraint, Unique Constraint, Check Constraint, Null, Not Null, Index, Sequence, View, Materialized View, Synonym, Procedure, Function, Package, Trigger
3.2. Database Objects
3.3. Database Sample Data and Sample Data Analysis
3.4. Create Object Commands, Alter Object Commands, and Drop Object Commands
3.5. DML Statements
3.6. Other Important Commands
3.7. Data Dictionary Commands

4.0 Data Warehouse (DW) & ETL
4.1. DW Concepts
4.2. DW & Data Mart
4.2.1. Enterprise Data Warehouse
4.2.2. ODS (Operational Data Store)
4.2.3. Data Mart
4.3. Star Schema
4.3.1. What is Star Schema?
4.3.2. Steps in designing Star Schema
4.3.2.1. Identify a business process for analysis (like sales)
4.3.2.2. Identify measures or facts (sales dollar)
4.3.2.3. Identify dimensions for facts (product dimension, location dimension, time dimension, organization dimension).
4.3.2.4. List the columns that describe each dimension. (Region name, Branch name).
4.3.2.5. Determine the lowest level of summary in a fact table (sales dollar).
4.3.3. Snowflake Schema
   4.3.3.1. Important aspects of Star Schema & Snow Flake Schema
4.4. Fact Table
   4.4.1. Measure Types
      4.4.1.1. Additive - Measures that can be added across all dimensions.
      4.4.1.2. Non Additive - Measures that cannot be added across all dimensions.
      4.4.1.3. Semi Additive - Measures that can be added across few dimensions and not with others
   4.4.2. Steps in designing Fact Table
4.5. ETL Tools
   4.5.1. What is ETL
   4.5.2. Popular ETL Tools: Informatics, DGT/Studio. Data Stage. Ab initio, Data Junction, Oracle Warehouse Builder, Microsoft SQL Integration, TransformOnDemand, Transformation Manager
4.6. ETL Concepts
   4.6.2. Source System
   4.6.3. Mapping
   4.6.4. Metadata Data
   4.6.5. Staging Area
   4.6.6. Cleansing
   4.6.7. Transformation
   4.6.8. Transportation
   4.6.9. Target System
4.7. Learn Informatica
4.8. Informatica-Transformations
   4.8.1. Active Transformation
   4.8.2. Passive Transformation
   4.8.3. Connected Transformation
   4.8.4. Un-Connected Transformation
   4.8.5. List of Transformations available in Informatica:
4.9. ETL Tools-Transformation Manager

5.0 Transformation Manager(TM)
   5.1. Overview and Executive Summary
   5.2. Key Features of TM
   5.3. Available Solutions
   5.4. Hand Coding and Code Generators
   5.5. Engine Based ETL Data Integration Solution, Database Embedded ETL
   5.6. META-DATA Text Repository
   5.7. Data Quality and Data Access to Data of any Format
5.8. Model Management and Development Environment
5.9. Transformation Features
5.10. Test and Debugger Tools
5.11. Scalability and Performance
5.12. Higher Productivity
5.13. Handle the Most Complex Requirements
5.14. Deployment, Deployment Flexibility to Suit Project Requirements
5.15. Improved Management and Reporting
5.16. Reduced Project Risk and Future Maintainability
5.17. Example Cases
5.18. Database-RDMS
5.19. Popular RDBMS Databases: Oracle, IBM DB2 UDB, IBM Informix, Microsoft SQL Server, Sybase, Teradata

6.0 ERP (Enterprise Resources Planning)
6.1. Enterprise Resource Planning (ERP)
6.2. What an ERP System should be
   6.2.1. Flexible to the changing needs of an organization
   6.2.2. Protecting their existing investment
   6.2.3. Increasing the customer service by satisfying the needs of the customer
   6.2.4. Able to talk with other business-to-business transactions
   6.2.5. Providing a unified data model to single, accurate view of the business application
   6.2.6. Able to provide quick implementation, optimize performance, streamline support
   6.2.7. Able to support latest technologies like Electronic Fund Transfer, Electronic Data
   Interchange (EDI), Internet, Intranet, Video Conferencing, E-Commerce
   6.2.8. Providing Business Intelligence Tools for Decision Support Systems
   6.2.9. Helping Managers to do best project management
   6.2.10. Integrating all the departments of a company and across the companies under the
   same management.

6.3. ERP Tools
6.4. Oracle Applications
6.5. Oracle Apps Implementation

7.0 Metadata & BI
7.1. Metadata Tools
7.2. Business Metadata
7.3. Technical Metadata
7.4. Metadata & ETL
7.5. Metadata Reports
7.6. BI Overview
7.7. BI Tools
7.8. OLAP & Hybrids
7.9. OLAP Analysis
7.10. OLAP Database-Multidimensional
7.11. Key Performance Indicators  
7.12. BI Dashboards  
7.13. BI Scorecars  
7.14. What is Data Mining?

8.0 Information Technology Overview  
8.1. Information Technology  
8.2. Operating System  
8.3. Server, Mainframe  
8.4. Computer Networking and Computer Testing  
8.5. Visio, C Language, XML  
8.6. Enterprise Application Integration (EAI)

Prerequisite

Knowledge in Computer Programming and Database

Contact Hours

_______ Contact Hours (Lecture ___ Hours / Lab ____ Hours)

Semester Credit Hours

__________ semester credit hours

Teaching Strategies

A variety of teaching strategies may be utilized in this course, including but not limited to, lecture, discussion, written classroom exercises, written lab exercises, performance based lab exercises, demonstrations, quizzes and examinations. Some quizzes may be entirely or contain lab based components. A mid-course and end course examination will be given.

Method of Evaluating Students

Grade Distribution

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Attendance</td>
<td>10</td>
</tr>
<tr>
<td>Mid Term</td>
<td>30</td>
</tr>
<tr>
<td>Finals</td>
<td>50</td>
</tr>
<tr>
<td>Special Projects</td>
<td>10</td>
</tr>
<tr>
<td>Makeup projects</td>
<td>10</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

AVTECH Institute
Grading Policy

At the end of each course, each student is assigned a final grade as follows:

<table>
<thead>
<tr>
<th>Point Range</th>
<th>Interpretation</th>
<th>Grade</th>
<th>Quality Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>90 – 100</td>
<td>Excellent</td>
<td>A</td>
<td>4.0</td>
</tr>
<tr>
<td>80 – 89</td>
<td>Very Good</td>
<td>B</td>
<td>3.0 – 3.9</td>
</tr>
<tr>
<td>70 – 79</td>
<td>Average</td>
<td>C</td>
<td>2.0 – 2.9</td>
</tr>
<tr>
<td>60 – 69</td>
<td>Poor</td>
<td>D</td>
<td>1.0 – 1.9</td>
</tr>
<tr>
<td>Below 60</td>
<td>Failure</td>
<td>F</td>
<td>0</td>
</tr>
<tr>
<td>N/A</td>
<td>Withdrawal</td>
<td>W</td>
<td>0</td>
</tr>
<tr>
<td>N/A</td>
<td>Pass</td>
<td>P</td>
<td>0</td>
</tr>
<tr>
<td>N/A</td>
<td>Incomplete</td>
<td>I</td>
<td>0</td>
</tr>
</tbody>
</table>

A student earning a grade of D or above is considered to have passed the course and is eligible to pursue further studies. A student receiving a grade of F has failed the course. A failed course must be repeated and passed to meet Avtech Institute’s graduation requirements, in addition to an overall program GPA of 2.0.

Requirements for Successful Completion of the Course

At a minimum, students must achieve the following:

- A passing grade of D or above
- Completion of all required examinations
- Submission of all required lab exercises and projects and;
- Adherence to the school attendance policy.

Equipment Needed

Industry standard desktop computer for lab exercises.

Equipment Breakdown

Lab room
Videos and Projector

Library Assignments

To be determined by the instructor.

Portfolio Assignment

Student program outcome portfolios are required to demonstrate student competencies. In conjunction with your course structure, please select a project/paper that best demonstrates what you have learned in this course and add it to your program portfolio.
Course Policies

Disruptive Behavior

Disruptive behavior is an activity that interferes with learning and teaching. Inappropriate talking during class, surfing inappropriate website, tardiness, cheating, alcohol or drug use, use of cell phone, playing loud music during class, etc. all disrupt the learning process.

Copyright Infringement

Specific exemptions to copyright infringement are made for student use in the context of learning activities. Graphic design students often download images from the Internet, or scan images from publications. As long as this work is for educational purpose, and subject to faculty permission, this is not a problem.

Plagiarism

Faculty cannot tolerate the misrepresentation of work as the student’s own. This often involves the use by one student or another student’s design, whether voluntarily or involuntarily. In the event that plagiarism is evident and documented, all students involved in the conscious decision to misrepresent work must receive an F as the grade for the project. A second occurrence may result in suspension for the rest of the quarter, and return to the school only after a review by the Academic Standards Committee.

Attendance

Attendance and Lateness

In education and the workplace, regular attendance is necessary if individuals are to excel. There is a direct correlation between attendance and academic success. Attendance is mandatory. All students must arrive on time and prepared to learn at each class session. At the faculty member’s discretion, students may be marked absent if they arrive more than 15 minutes late to any class. More than five absences in a class that meets twice per week or more that two absences in a class that meets once per week may result in a failure.

Make-Up Work

Late Projects and Homework

All projects and homework must be handed in on time. Homework should be emailed to your instructor if you are going to miss a class. Work that is submitted one week late will result in the loss of one full grade; and work that is submitted two weeks late will result in the loss of two full grades; more than two weeks late you will receive a failing grade on the project.