



# COURSE SYLLABUS

## **Data Warehouse & ETL** (Overview)



## 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

#### Modeling Types

Process Modeling

FTL: Affordable, Fast ROI

Master Data Management: Centralize master data across all BI, ERP, warehousing applications

Business Process Modeling (Business Process Modeling, Process Flow Modeling, Data Flow Modeling)

Data Modeling (Conceptual Data Modeling, Enterprise Data Modeling, Logical Data Modeling, Physical Data Modeling, Relational Data Modeling, Dimensional Data Modeling)

#### Modeling Methods

Business Modeling Methods: IDEF (acronym), IDEF0 (methods)

Process Modeling: Software, solutions and consulting Download Whitepapers

Process Methodology: World-class process metrology

2007 Workflow Handbook: IDEF3, IDEF4

Data Modeling Methods: IDEFIX, , Dimensional Modeling Notation

#### Business Process

Process Modeling, Business Process Modeling, Mapping Process, Business Process Modeling

Different kinds of Business Processes: Product manufacturing, payment of insurance and taxes, HR and Recruitment, services calls in call centers, inforion stored by bank tellers, the transformation in data warehouse)

#### Business Process Tools

Smart BPM, Process modeling, Business Management Class and Business Requirements

(Popular) Business Process Tools: BPM Suite, Process Suite, Business Manager, Paga Rules Process Commander, E Work Vision, Team Works, Intalio, Bizflow, EugeoBPM, Business Process Manager

#### Business Process Management (BPM)

Business Requirements

B2B Process Integration

Business Process Report

Process Management

Need for Business Process Management

#### Advantages of BPM

Process Modeling, BPM & Rules Whitepaper, Business Requirements, Reliable BPM Solution

#### Business Process Re-engineering

Business Requirements, Process modeling, and Business Process Report

#### Business Process Modeling

Business Requirements, Process modeling, and Workflow

#### Business Process Tools

Popular BPM tool Name: All fusion Process Data Modeler, Visio, Corporate Modeler, Procarta, Aris Toolset, Live Model, Workflow Modeler, Aion, Holosofx, System Architect BPR, Designer/2000 BPR software, Provision, Smart Draw

#### Business Process Modeling Example

Business Process Modeling: BPM Suite from EMC, Migrate to WS-BPEL, Business Process, Business Process Discover  
Business Requirements, Smart Rules-Driven BPM, Business Process Mgmt

#### Process Flow Modeling

Business Process Report and Business Requirements  
Business Process: IDEF3, Fan Out Junctions: Fan In Junctions: Synchronous AND, Synchronous AND, Asynchronous OR, OR

#### Data Flow Modeling

Business Process and Data Flow Diagram: Data Warehouse Institute. Free DB Modeling Trial, Talend: 100% free ETL  
Activities or Processes, Data Flows or Arrows, Data Store, External References, Physical Resources

#### Work Flows

BPM & Rules Whitepaper, Workflow and Business Process Mgmt, 2007  
Workflow Handbook  
Advantages of Workflow

#### Business Activity Monitoring

BPM & Rules Whitepaper, Process modeling, (Business Process Reduce BP cycle Time with Sonic ESB. Free Webinar)  
Advantages  
Workflow software

#### Sources of Modeling Data

#### Pre-Processing the Data

#### Alternative Modeling Strategies

## 2.0 Data Modeling (DM)

### Overview

Easy ER Diagram Software (SmartDraw)  
Database Modeling Tool (DeZign for Databases V4-datanamic)  
ModelRight  
Data Warehousing Video

Data Modeling Tools: Erwin, Embarcadero, Rational Rose, Power Designer, Oracle Designer, Xcase

### Tools

Logical Data Model  
Physical Data Model

### DM Tools-Erwin

### ED Tools-Xcase

### Development Cycle

Gathering Business Requirements-First Phase  
Conceptual Data Modeling (CDM)-Second Phase

- Logical Data Modeling (LDM)-Third Phase
- Physical Modeling (PDM)-fourth Phase
- Database -Fifth
- DM Standards
  - Standardization Needs/Modeling data
  - Table Names Standardization
  - Column Names Standardization
  - Database Parameters Standardization
- Steps to Create a Data Model
- Data Modeler Role
  - Business Requirement Analysis
  - Development of data model
  - Reports
  - Review
  - Creation of database
  - Support & Maintenance
- Modeling Reports
  - Logical Data Model Report
  - Physical Data Model Report
- Conceptual Data Modeling (CDM)
- Logical Data Modeling (LDM): entities, attributes, key groups, and relationships
- Physical DM
- Logical vs. Physical
  - Logical DM
    - Represents business information and defines business rules
    - Entity, attribute, primary Key, Inversion Key Entry, Rule, Relationship, and Definition
  - Physical DM
    - Represents the physical implementation of the model in a database
    - Table, column, Primary Key Constraint, Unique Constraint or Unique Index, Non Unique Index, Check Constraint, Default Value, Foreign Key, Comment
- Relational (OLTP) DM
  - Data is stored in RDBMS
  - Tables are units of storage
  - Data is normalized and used for OLTP. Optimized for OLTP processing
  - Several tables and chains of relationships among them
  - Volatile(several updates) and time variant
  - SQL is used to manipulate data
  - Detailed level of transactional data
  - Detailed level of transactional data
- Relational vs. Dimensional
  - Dimensional DM:
    - Data is stored in RDBMS or Multidimensional databases
    - Cubes are units of storage

Data is denormalized and used in data warehouse and data mart. Optimized for OLAP

Few tables and fact tables are connected to dimensional tables

Non volatile and time invariant

MDX is used to manipulate data

Summary of bulky transactional data(Aggregates and Measures) used in business decisions

User friendly, interactive, drag and drop multidimensional OLAP Reports

Dimensions

Dimension Table

Location Dimension

Slowly Changing Dimensions

Supertype & Subtype

Creating a Data Model in Visio

### **3.0 Database & Data Modeling**

Database Overview

Database Types: Database Management Systems (DBMS), Relational Database Management Systems (RDBMS), Object Oriented Databases, Multidimensional Databases

Often Used Databases (RDBMS) In Most Of The Practical Applications: Oracle, Sql Server, Informix, Terradata, DB2

Data Modeling Tool: For example, Erwin to generate DDL scripts from the Data Modeling tool

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

Database Objects

Database Sample Data and Sample Data Analysis

Create Object Commands, Alter Object Commands, and Drop Object Commands

DML Statements

Other Important Commands

Data Dictionary Commands

### **4.0 Data Warehouse (DW) & ETL**

DW Concepts

DW & Data Mart

Enterprise Data Warehouse

ODS (Operational Data Store)

Data Mart

Star Schema

What is Star Schema?

Steps in designing Star Schema

Identify a business process for analysis (like sales)

Identify measures or facts (sales dollar)

Identify dimensions for facts (product dimension, location dimension, time dimension, organization dimension).

List the columns that describe each dimension. (Region name, Branch name).

Determine the lowest level of summary in a fact table (sales dollar).

Snowflake Schema

Important aspects of Star Schema & Snow Flake Schema

Fact Table

Measure Types

Additive - Measures that can be added across all dimensions.

Non Additive - Measures that cannot be added across all dimensions.

Semi Additive - Measures that can be added across few dimensions and not with others

Steps in designing Fact Table

ETL Tools

What is ETL

Popular ETL Tools: Informatica, DGT/Studio. Data Stage. Ab initio, Data

Junction, Oracle Warehouse Builder, Microsoft SQL Integration,

TransformOnDemand, Transformation Manager

ETL Concepts

Glossary of ETL (Reference:[www.Oracle.com](http://www.Oracle.com))

Source System

Mapping

Metadata Data

Staging Area

Cleansing

Transformation

Transportation

Target System

Learn Informatica

Informatica-Transformations

Active Transformation

Passive Transformation

Connected Transformation

Un-Connected Transformation

List of Transformations available in Informatica:

ETL Tools-Transformation Manager

## **5.0 Transformation Manager(TM)**

Overview and Executive Summary

Key Features of TM

Available Solutions

Hand Coding and Code Generators

Engine Based ETL Data Integration Solution, Database Embedded ETL

META-DATA Text Repository

Data Quality and Data Access to Data of any Format

Model Management and Development Environment  
Transformation Features  
Test and Debugger Tools  
Scalability and Performance  
Higher Productivity  
Handle the Most Complex Requirements  
Deployment, Deployment Flexibility to Suit Project Requirements  
Improved Management and Reporting  
Reduced Project Risk and Future Maintainability  
Example Cases  
Database-RDMS  
Popular RDBMS Databases : Oracle, IBM DB2 UDB, IBM Informix, Microsoft SQL Server, Sybase, Terradata

## **6.0 ERP (Enterprise Resources Planning)**

Enterprise Resource Planning  
(ERP)

What an ERP System should be

- Flexible to the changing needs of an organization
- Protecting their existing investment
- Increasing the customer service by satisfying the needs of the customer
- Able to talk with other business-to-business transactions
- Providing a unified data model to single, accurate view of the business application and an enterprise data model to view the complete functions of the enterprise
- Able to provide quick implementation, optimize performance, streamline support and maximize the return on your investment
- Able to support latest technologies like Electronic Fund Transfer, Electronic Data Interchange (EDI), Internet, Intranet, Video Conferencing, E-Commerce
- Providing Business Intelligence Tools for Decision Support Systems
- Helping Managers to do best project management
- Integrating all the departments of a company and across the companies under the same management.

ERP Tools

Oracle Applications

Oracle Apps Implementation

## **7.0 Metadata & BI**

Metadata Tools

Business Metadata

Technical Metadata

Metadata & ETL

Metadata Reports

BI Overview

BI Tools

OLAP & Hybrids

OLAP Analysis

OLAP Database-Multidimensional



Key Performance Indicators  
 BI Dashboards  
 BI Scorecars  
 What is Data Mining?

### 8.0 Information Technology Overview

Information Technology  
 Operating System  
 Server, Mainframe  
 Computer Networking and Computer Testing  
 Visio, C Language, XML  
 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

|                                  |             |
|----------------------------------|-------------|
| Class Attendance                 | 10          |
| Mid Term                         | 30          |
| Finals                           | 50          |
| Special Projects Makeup projects | 10          |
| <b>Total</b>                     | <b>100%</b> |

## Grading Policy

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

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

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 than 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.