

Habib University
CSE 457 – Data Science
Fall' 2020
Assignment 1 - Maximum Miniatures, Inc.
35 Points

Objective:

The objective of this assignment is to make students practice dimension modeling on a real-world business scenario and gain hands-on experience with Extract-Transform-Load(ETL) to populate a data warehouse.

You are given a business scenario of Maximum Miniatures Inc. that you need to understand and then attempt the following questions:

- a) [12 points] Design a star schema for the requirements mentioned in the scenario for Manufacturing and Sales data marts.
- b) [03 points] Generate a SQL script for your star schema and run that script to setup your data warehouse MaxMinDW in SQL server.
- c) [20 points] Build an ETL workflow in SSIS to load the data from given source databases to MaxMinDW data warehouse.

Submissions: You will do this assignment in pairs. Only one member needs to submit the assignment. The assignment will be submitted on the LMS as a zip containing a folder (Ass01_Student1Name_Student2Name). The folder will in turn contain:

- the diagram of your star schema
 - sql script to create the schema of your Data warehouse
 - complete SSIS project that runs the ETL process
-

Case Study – Maximum Miniatures Inc.

Maximum Miniatures, or Max Min, Inc., manufactures and sells small, hand-painted figurines. It has several product lines, including the Woodland Creatures collection of *North American animals*; the *Mythic World* collection, which includes dragons, trolls, and elves; the *Warriors of Yore* collection containing various soldiers from Roman times up through World War II; and the *Guiding Lights* collection featuring replica lighthouses from the United States. The miniatures are made from clay, pewter, or aluminum.

Max Min markets these miniatures through three different channels. It operates five of its own "Maximum Miniature World" stores dedicated to selling the Max Min product line. Max Min also operates MaxMin.com to sell its products online. In addition, Max Min sells wholesale to other retailers.

Business Needs

Max Min, Inc. has experienced rapid growth in the past three years, with orders increasing by over 300%. This growth has put a strain on Max Min's only current Source of business intelligence, the printed report. Reports that worked well to support decision making just a few years ago now take an hour or more to print and even longer to digest. These reports work at the detail level with little summarization. Max Min's current systems provide few, if any, alternatives to the printed reports for viewing business intelligence. In addition, Max Min, Inc. is facing tough competition in a number of its product areas. This competition requires Max Min to practice effective decision making to keep its competitive edge. Unfortunately, Max Min's current business intelligence infrastructure, or lack thereof, is making this extremely difficult. Because of these issues, Max Min has launched a new project to create a true business intelligence environment to support its decision making. This project includes the design of a data warehouse structure, the population of that data warehouse from its current systems, and the creation of analysis applications to serve decision makers at all levels of the organization.

The new business intelligence platform is based on SQL Server 2005. After an extensive evaluation, it was decided that the SQL Server 2005 platform would provide the highest level of business intelligence capability for the money spent. SQL Server 2005 was also chosen because it features the tools necessary to implement the data warehouse in a relatively short amount of time. Let's take a quick look at Max Min's current systems.

Current Systems

Max Min has five data processing systems that are expected to serve as sources for DW.

Manufacturing Automation

The manufacturing automation system tracks the materials used to make each product. It also stores which products are manufactured on which production lines. Finally, this system tracks the number of items manufactured during each shift. The manufacturing automation system uses a proprietary data storage format. Data can be exported from the manufacturing automation system to a comma-delimited text file. This text file serves as the source for loading the manufacturing data into the business intelligence systems.

Order Processing

The order processing system manages the inventory amounts for all products. It tracks wholesale orders placed by non-Max Min retailers. The system also records product amounts sold through the Max Min retail stores and the Max Min online store to maintain inventory amounts.

The order processing system tracks order fulfillment including product shipping. It also generates invoices and handles the payment of those invoices. In addition, this system records any products returned from the retailer. The order processing system uses a Microsoft SQL Server database as its backend.

Point of Sale

The point of sale (POS) system manages the cash registers at each of the five Max Min-owned retail stores. This system also tracks the inventory at each retail store using UPC bar code stickers placed on each item. The POS system handles both cash and credit card transactions. It also tracks information on any products returned by the customer.

Information from each of the five POS systems is exported to an XML file. This XML file is transferred, using FTP, to a central location nightly. These XML files serve as the source for loading the POS data into the business intelligence systems.

MaxMin.com

The MaxMin.com online store is an ASP.NET application. It uses SQL Server as its backend database. All sales through the online store are paid with a credit card. All customers of the online store must provide name, address, phone number, and e-mail address with each purchase.

The online store tracks the shipping of orders. It also handles any products returned by customers. Finally, the online store saves information on product promotions and discounts that are run on the store site.

Accounting

The accounting system tracks all the financial transactions for Max Min, Inc. This includes the purchase of raw materials for manufacturing. The accounting system uses a Visual FoxPro database for its backend.

Designing the Maximum Miniatures Manufacturing Data Mart

Business Need

The vice president (VP) of production for Max Min, Inc. wants to analyze the statistics available from the manufacturing automation system. He would like an interactive analysis tool, rather than printed reports, for this analysis. In keeping with Max Min's new business intelligence strategy, Microsoft SQL Server 2005 Analysis Services is the platform for this analysis tool. Because the manufacturing automation system does not use a database, logging everything to comma-delimited text files instead, a data mart must be designed and built as a repository for this information. The manufacturing automation system controls all the machines used by Max Min to create its figurines.

Each machine handles all the steps in the manufacturing process of a figurine. This includes the following:

- Filling a mold with the raw material (clay, pewter, or aluminum)
- Aiding the hardening of this material
- Removal from the mold when hardening is complete

- Computerized painting of the figurine, if necessary (pewter figurines are not painted)
- Curing the paint, if necessary

Multiple painting and curing cycles may be necessary, depending on the intricacy of the paint job required by a product. A quality assurance check is done by the machine operator as the figurine is completed. Operators log onto a machine. As part of this logon process, the operator tells the manufacturing automation system what product is being produced, along with the batch number of the raw material being used by that machine. The operator also makes an entry in the system when a figurine is rejected.

An interview with the VP of production yielded the following data requirements for effective decision making:

- Number of accepted products by batch by product by machine by day
- Number of rejected products by batch by product by machine by day
- Elapsed time for molding and hardening by product by machine by day
- Elapsed time for painting and curing by paint type by product by machine by day
- Product rolls up into product subtype, which rolls up into product type
- Machine rolls up into machine type, which rolls up into material (clay, pewter, or aluminum)
- Machine also rolls up into plant, which rolls up into country
- Day rolls up into month, which rolls up into quarter, which rolls up into year

The information should be able to be filtered by machine manufacturer and purchase date of the machine. The export file from the manufacturing automation system contains one row for each product produced.

Each row includes the following information:

- Product
- Batch number of the raw material
- Machine number
- Operator employee number
- Start of manufacture date and time (when the batch run begins)
- End of manufacture date and time (when the batch is complete)
- Number of items produced
- Number of items rejected
- The manufacturing automation system does not include information on the hierarchies within Max Min. To include these hierarchies in the data mart, we need to pull data from another system. It turns out that the Accounting System has the data we need.

Designing the Maximum Miniatures Sales Data Mart

Business Need The VP of sales for Max Min, Inc. would like to analyze sales information. This information is collected by three OLTP systems: the Order Processing System, the Point of Sale (POS) System, and the MaxMin.com Online System. The Order Processing System and the MaxMin.com Online System both use Microsoft SQL Server as their back-end databases. The POS System uses XML files to transfer data from each of the five Max Min-owned retail stores.

The VP of sales would like to be able to analyze the following numbers:

- Dollar value of products sold
- Number of products sold
- Sales tax charged on products sold
- Shipping charged on products sold

These numbers should be viewable by:

- Store
- Sales Promotion
- Product
- Day, Month, Quarter, and Year
- Customer
- Sales Person

An analysis of the three data sources shows that all of this information is available from at least one data source. The three data sources have been attached with the assignment.