# A Realistic Data Warehouse Project:
# An Integration of Microsoft Access® and
# Microsoft Excel® Advanced Features and Skills

## Michael A. King
## Virginia Polytechnic Institute and State University
## Blacksburg, VA, USA

### Michael.king@vt.edu

## Executive Summary

Business intelligence derived from data warehousing and data mining has become one of the most strategic management tools today, providing organizations with long-term competitive advantages. Business school curriculums and popular database textbooks cover data warehousing, but the examples and problem sets typically are small and unrealistic. The purpose of this paper is to provide an overview of how to construct a realistic data warehouse using numerous advanced features available in Microsoft Access and Microsoft Excel.

Large fact table creation is demonstrated, which subsequently allows for the development of meaningful queries and cross tab analysis utilizing pivot tables. Fact table sizes of one million records can be iteratively developed and quickly imported into databases such as Microsoft Access or MySQL. A short discussion on the benefits of using Microsoft Access Query by Example and completely bypassing the complexities of advanced SQL is included.

With the resulting fact table, students can experiment with several indexing techniques, usually only conceptually discussed in textbooks, and measure a series of index effectiveness. This paper includes a brief discussion of enterprise-level data requirements, the differences between dimensional and relational modeling, data warehouse schemas, and enterprise data flow concepts, along with a demonstration of business modeling concepts, such as random variable generation and probability distributions.

As a case example, this data warehouse project utilizes a public retail corporation with an excellent online presence to provide the student with a real data extract, transform and load hands on experience. General financial data and colorful background information about the corporation is provided.

**Keywords:** Data warehouse, star schema, dimensional modeling, pivot tables, Microsoft Access, Microsoft Excel, random variable generation.

## Introduction

Due to the current global business environment and the resulting competitive pressures, an increasing number of corporations have implemented data warehouses to assist senior management with

strategic policies and decisions.  Edgar F. Codd, the father of the relational database model once said:

*"Today's markets are much more competitive and dynamic than those in the past. Business enterprises prosper or fail according to the sophistication and speed of their information systems, and their ability to analyze and synthesize information using those systems. The numbers of individuals within an enterprise who have a need to perform more sophisticated analysis is growing."*
(Codd, Codd, & Salley, 1998)

There are numerous cases where companies such as Continental Airlines, Wal-Mart, Toyota, Hewlett-Packard and Snyder's of Hanover have developed decision support systems along with the underlying data warehouse infrastructure and have sustained competitive advantages within their respective industries.  While it is apparent that data warehouses are increasingly common within the for-profit and even in the non-profit sectors, full courses offered by business schools are rare and the popular database textbooks written by Robb and Coronel, and Hoffer, Prescott, and McFadden offer limited coverage of data warehouses.  Data warehouse treatment from these authors is usually confined to one chapter and the database examples are small and unrealistic. The Price College of Business at the University of Oklahoma (http://price.ou.edu/) offers an excellent semester data warehouse course; however, the course examples and exercises, again, are quite small and unrealistic.  Teradata Student Network (http://tunweb.teradata.ws/tunstudent/) offers valuable business intelligence and analytic exercises but no exercises on basic data warehouse design and data population.

The purpose of this paper is to describe and demonstrate how to create a realistic data warehouse schema with a one year time frame, develop and import real data, and simulate online analytical processing.  As an additional pedagogical extension, advanced skills and features from both MS Access and MS Excel are outlined and demonstrated.  The MS Excel features web query, string editing techniques, and random number generation are covered, along with MS Access concepts such as crosstab queries, pivot tables, pivot charts and primary key management and manipulation are detailed.  Lowe's Corporation (www.lowes.com) was chosen as a corporate model to add the much needed realism to the project, due to the availability of product information, store locations, and financial data.  However, any corporation with a strong online presence, such as Target, Best Buy or Barnes and Noble, would be sufficient.

# Lowe's Corporation

What started as a small hardware store in Wilkes County North Carolina, U.S., grew into 48th on the Future 500 list of top U.S. public corporations.  Lowe's is number two in the home improvement industry, after Home Deport.  Lowe's customer base includes the do-it-yourself market segment to professional contractors.  Key product groups for Lowe's are lumber, millwork, appliances, tools, hardware, and lawn care.  Lowe's is second only to Sears in U.S. appliance sales. The company is located in all 50 U.S. states, with 1534 retail locations, and booked $48,283 (mil.) in fiscal 2007 revenues.  Like most public corporations, Lowe's has seen its market capitalization fall from $32.9 (Bil.) in early 2008 to approximately $20.5 (Bil.) in November 2008. While Lowe's market capitalization has fallen, it has maintained an impressive 5 year average return on equity of 19.97%.  During fiscal 2007, Lowe's recorded over 720 million customer transactions with an average of $67.05 per ticket.  The retailer's stated strategy is to focus on domestic sales, specifically targeting baby boomers and women with upscale stores, additional service offerings, along with a full online shopping presence.

As an aside, and a testament to Lowe's long legacy of entrepreneurship, during the 1940s Lowe's general store was the only merchant in western North Carolina that was able to stock a sizable inventory of copper tubing.  The World War II support effort had essentially depleted most sup-

plies of raw and formed metals.  Lowe's could sell all the copper tubing to the "local community" it could locate and for roughly a decade sustained a profitable business by meeting the "customized needs" of the rural region.  The boot legging business continued to be profitable until the early 1950s when the price of sugar tripled and then the "shine drivers" began driving around mud tracks to pass the time.  The drivers realized that they could actually charge spectators an admission fee to watch the races, thus the birth of NASCAR.  Lowe's Corporation continues to support its' legacy as sponsor of the Lowe's Motor Speedway

# The Data Warehouse

Before a discussion of dimensional modeling and data warehouse schema design, a short review of the differences between operational data and decision support data is warranted.  Table 1 outlines the differences between three types of enterprise data: transactional, decision support, and online analytical processing.  The functionality of enterprise data spans a continuum, for example, the typical use of transactional data is for daily updates, while OLAP data is intended for strategic analysis.   Another example of this enterprise data attribute continuum relates to the functionality of User Views.  Transactional views are usually static and application oriented, while decision support data lends itself to more flexible user views of the data, and finally OLAP data allows management complete ad hoc report ability.

| Attribute | Transactional | Decision Support | OLAP |
|-----------|---------------|------------------|------|
| Typical Use | Update | EIS Reports | Data analysis |
| Analytical Need | Low | Medium | High |
| User Views | Static | User Defined | Ad hoc |
| Data Flow Size | Small | Medium | Large |
| Data Aggregation | Low | Summary | Business Analytics |
| Time Context | Current | Near Time | Archived |

**Table 1: Enterprise Data Comparison**

A discussion relating to the difference between relational modeling and dimensional modeling is also useful because, as with most database courses, extensive coverage is given to relational modeling and usually only cursory attention given to dimensional modeling.  While dimensional modeling employs many of the same relational database constructs, such as tables, multiplicity, and integrity constraints, the dimensional modeling process is not necessarily intuitive and in some respects is a mirror image to relational modeling.  Table 2 describes four major differences between the two data modeling perspectives and it should be noted that probably the most important difference relates to schema design.  The relation or table normalization process maximizes data integrity and minimizes data redundancy; however, the process actually has a negative effect on a dimensional data model.  Dimensional models are "denormalized" which, in effect, optimizes the database for query retrieval.  Data modelers organize data warehouses with the goals of ease of reporting and semantic understanding, in contrast to the relational modeling goals of operational efficiency and transactional accuracy.

A hybrid star-snowflake database schema, which is based on a denormalized relational model, was chosen for this project because of the additional concepts and complexities that could be modeled and discussed.  Figure 1 illustrates the entity relationship diagram for a typical retail data warehouse and provides detailed information on the three data dimensions: time, product, and location.  These three dimensions are standard dimensional modeling attributes frequently used in retail environments, which makes the schema quite adaptable.  Each data dimension is organized

| Attribute | Relational Modeling | Dimensional Modeling |
|---|---|---|
| Schema Focus | Optimized for update | Optimized for retrieval |
| Data Requirements | Minimize data redundancy | Maximize data meaning |
| Entities | One table per entity | One fact table per data gain |
| Process Function | Transactional | Analytical |

**Table 2: Relational and Dimensional Data Modeling**

into a hierarchy that supports the capability of OLAP tools to "drill down" into the data for finer details or to "roll up" data for aggregation analysis. Using Lowe's as an example, the Location Dimension Table has 540 records and is organized top down beginning with the Region, then state, moving all the way down to a specific Store name and store number. The Time Dimension Table has 365 records, each representing one day of the fiscal year, the week number, month number and quarter number. One strategy that improves data navigation and organizational efficiency within a dimension is to normalize the dimension table. In this specific example, the Product Dimension is normalized into three tables, each with a one-to-many relationship, i.e., the appliance product group has many product types such as washers, dryers, and refrigerators. The refrigerator product type has many unique products, such as a GE side-by-side 25.5 cu. ft. stainless steel refrigerator. The Product Dimension has 13 main products groups, 148 products types and 1720 individual products.
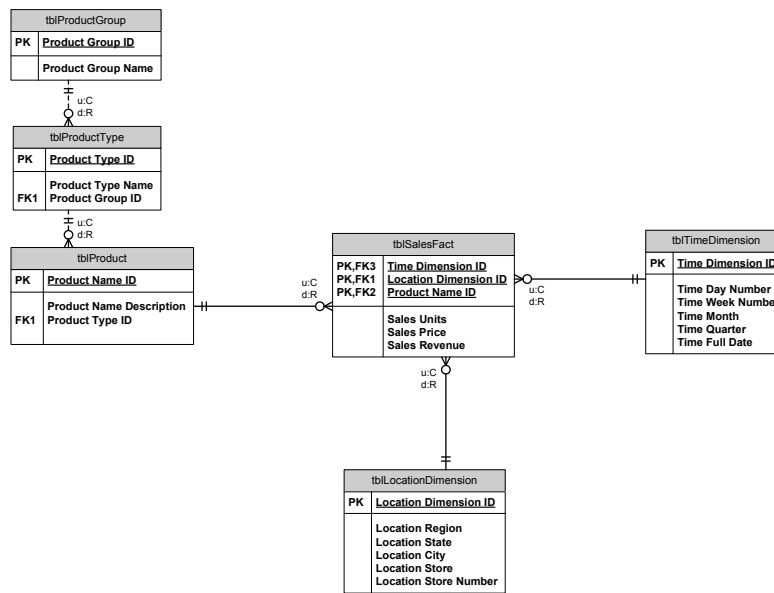


**Figure 1: Hybrid Snowflake Star Schema**

The junction table in the data warehouse is called a fact table and it contains very detailed metrics, measures or quantitative data from business activities. A data modeler has a choice of what degree of grain to use, and for this project, the modeler choose a fine grain fact table to represent individual transactions. For example, Figure 2 illustrates that a transaction occurred, on day 116 from store 240, with product 294 that included 26 units priced at $341.00 for a total ticket of $8,866.00.

| Time Dimension ID | Location Dimension ID | Product Name ID | Sales Units | Sales Price | Sales Revenue |
|---|---|---|---|---|---|
| 116 | 240 | 294 | 26 | $341.00 | $8,866.00 |
| 93 | 178 | 294 | 29 | $437.53 | $12,688.23 |
| 126 | 14 | 294 | 26 | $278.48 | $7,240.36 |

**Figure 2: Partial Fact Table**

The fact table in this project contains 20,000 records, which is an arbitrary number; however, this large fact table supports realistic data analysis with pivot tables, pivot charts, and crosstab queries.  Typically, course or textbook supplied data sets are small and provide limited data analysis opportunities.  In contrast, this project provides a complex and extensive fact table, which theoretically, could hold a minimum of 365 x 1720 x 540 records. Certainly, this size is minuscule compared to the multi-terabyte data warehouses in operation at the previously mentioned companies.

Although data warehouses store historical data, the analytical and transactional processes are not isolated data flows.  In reality, the data warehouse process is intricately networked within the enterprise data infrastructure.  Figure 3 illustrates the data flow beginning in the transactional data systems, then flowing to the informational system and subsequently flowing through the decision support system.  The operational managers, that began the original data flow, use the resulting data analysis to make better decisions.
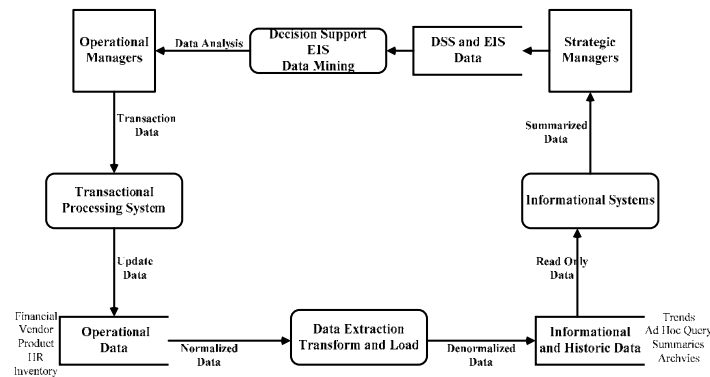


**Figure 3: Enterprise Data Context Diagram**

# Modeling the Fact Table

As mentioned, most data warehouse examples used in database courses are problematic because of their small fact table, and thus the crosstab queries and pivot table exercises are far from realistic.  One goal of this project was to utilize advanced features available in MS Excel to construct a realistic fact table with 20,000 records.  Most business students are very familiar with MS Excel, which makes the mechanics of modeling straight forward.  Once created, the student imports the spreadsheet directly into the MS Access database populating the entire fact table.

The first step in creating the fact table was developing a method to model the Time Dimension, which is essentially a proxy for sales demand volume.  As you may recall the Time Dimension has 365 records that represent the actual days of the year.  From a discussion with a Lowe's store manager, a general estimation of store transaction volume was developed.  As you would aspect for a home improvement retailer, transaction volume is lower in the winter months and highest in late spring and early summer.  Lowe's employee's affectionately call April, May and June the "100 days of hell."  As indicated, any retail company could be used a model for this project, as long as monthly sales revenue can be obtained. Lowe's strong online presences and unique his-

tory helps to keep the case interesting. One method to model over a specific interval, and in this case, 1 to 365, is to use the Beta distribution. This flexible distribution has four parameters: two shape parameters and two range parameters. The next step included using Excel Solver to determine the shape parameters for the distribution, assuming a 150 day mean and range parameters 1 and 365. Figure 4 illustrates a realistic distribution of sales transactions for a typical Lowe's store. The use of the Beta distribution is easily generalized to any cyclical demand modeling problem.
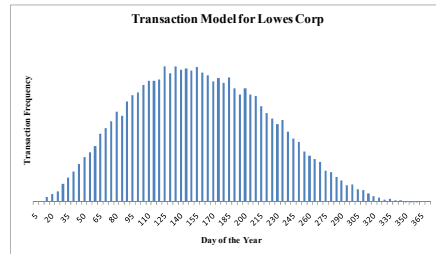


**Figure 4: Beta Distribution Transaction Model**

The final step in modeling sales demand was to randomly generate 20,000 integers based on the hypothesized Beta distribution using the MS Excel function,
**=int(betainv(rand(),3.233,4.6651,1,365))**. A series from 1 to 20000 was automatically created in column A beginning at A2 using the Fill Series tool as shown in Figure 5.
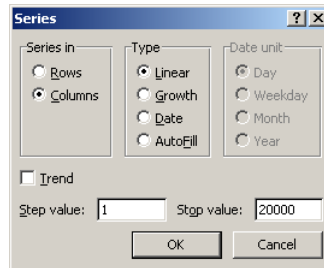


**Figure 5: Excel Fill Series Options**

Cell B2 contains the random number generating formula as shown in Figure 6. Double clicking the Fill Handle in the lower right hand corner of the Cell Indicator copies the formula down. Since there is an adjacent column with entries, Excel is smart enough to auto copy the formula down to the last adjacent cell with an entry, thus alleviating the arduous job of painting down 20,000 rows.
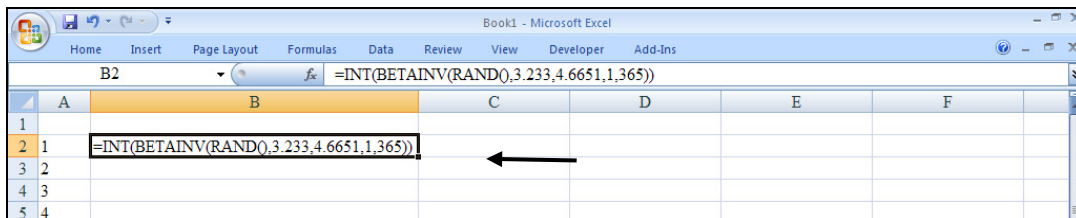


**Figure 6: Fact Table Modeling**

Modeling the remaining dimensions and fields, in the fact table, follows the same process. The Location Dimension entry was modeled by uniformly generating integers using the Excel function **=randbetween(1,540)** because there are 540 stores in the Location Dimension table. The Product Dimension entry was modeled by the same method, because there are 1720 unique products in the Product Name table. The sales unit volume was modeled based on a triangular distri-

bution by utilizing the XLSim (http://www.analycorp.com/) function add-in **=gen_triangular(1,5,50)**. The sales price was modeled by using the Excel function **=norminv(rand(),67,500)**. All of the distribution parameters can certainly be changed to model any retail environment. While only the Time Dimension is realistic, the remaining dimensions and fields offer more realism than manually creating 20,000 records to populate the fact table. By using these advanced modeling methods, a student can create the fact table in approximately thirty to sixty minutes! Figure 7 shows the completed formulas and that the Fill Handle tab is ready for double clicking to auto copy the row down 20,000 times. The spreadsheet now has 120,000 calculations that update automatically with any change to the worksheet. This powerful Excel feature actually can be an irritation. One of two methods can be used to prevent the large worksheet from automatically recalculating, by either setting the calculation method to manual or copying the range and only pasting the values using the Paste Special options.

The student then simply imports the spreadsheet, using the Get External Data – Excel Spreadsheet wizard, into the empty tblSalesFact table in MS Access, which has all the data types predefined before the import.



**Figure 7: Fact Table Modeling**

# Extract, Transform, and Load

Although the primary goal of this project is to demonstrate how to create a realistic data warehouse with a sufficiently large fact table that makes the resulting data analysis meaningful, importing accurate data is just as important. For added realism, the actual product groups, product types, and ten to thirteen individual products from each project type, along with store information from twelve states were extracted from the Lowe's website. The product information is contained in the normalized product dimension tables. The store information is contained in the unnormalized tblLocationDimension table. MS Excel Web Query along with several user defined Excel macros automated the download process and minimized manual data input, although some manual data input and manipulation remained. For example, a state was selected from Lowe's website and the Web Query tool imported the data into an empty worksheet as shown in Figure 8.



**Figure 8: Extracted Store Data**

A data "scrubber" macro deleted all unneeded information and blank lines, set the data type to text and inserted several column titles. The macro contains steps that includes the Find and Replace tool to locate duplicated text and to replace it with a blank. The macro integrates the Substitute function to remove unneeded data from strings within cells, by replacing it with a blank. The macro includes the Right function that was useful in copying the store number information into an adjacent cell. The macro process sorted the worksheet several times separating store names from unwanted information. The resulting import worksheet is shown in Figure 9.



**Figure 9: Transformed Store Data**

As previously indicated, MS Access will automatically assign a surrogate primary key to each record when imported. One interesting item to note is that MS Access will at times skip primary key values when importing external data sets, which normally does not matter because primary keys should carry no meaning. However, in this specific project, while the actual value does not matter, a continuous range of primary key values does. For example, when modeling the number of daily transactions, 20,000 random numbers in the range between 1 and 365 were generated because the Time Dimension table has 365 records. To reset the primary key for a specific table, a backup of the entire database was completed and a new Time Dimension table created using the Make Table query. Then the original primary key field was deleted, the database closed and compacted and then reopened. When the new primary key was recreated using the AutoNumber data type, a new sequential range of values was generated. It must be reiterated that primary key values should not contain any business meaning, but the ability to manipulate primary key values was essential for modeling realism in this project.

While not nearly as sophisticated as commercially available ETL applications, the download and text manipulation process does give the student an appreciation of the need to automate the extract, transform, and load process. Creating the data dimensions does provide hands-on experience with data type conversion, parsing, string manipulation, deleting duplicated or unnecessary data, and loading the resulting worksheet into a database table.

# Data Analysis

Numerous realistic total and crosstab queries were developed in MS Access using Query by Example for the Lowe's data warehouse application, as shown in the appendix. The screen shots give an idea of representative categories of reports. QBE was used because of the ease the graphical user interface affords the query builder. Table 3 illustrates the intimidating SQL code for a Product Group by Quarter Sales crosstab query. It would require four separate SELECT queries to accomplish what the specialized TRANSFORM and PIVOT SQL extensions complete illustrated in the example in Table 3. These queries take the fine grained transaction data from the tlbSalesFact table, which is too detailed to be meaningful to managers, and aggregates the raw data into more meaningful information such as views, reports or charts. Instead of presenting a report indicating that "on day 152 store 324 sold 10 units of product 528 for $1.52 each," busi-

ness intelligence tools such as OLAP, aggregate and summarize data into higher levels of more meaningful units, i.e., "the top product group in terms of sales for the Southeast region for 2007 was building supplies."

```
TRANSFORM Sum(qryFullDataWarehouseView.[Sales Revenue]) AS [SumOfSales Revenue]

SELECT qryFullDataWarehouseView.[Product Group Name],
Sum(qryFullDataWarehouseView.[Sales Revenue]) AS [Total Of Sales Revenue]

FROM qryFullDataWarehouseView

GROUP BY qryFullDataWarehouseView.[Product Group Name]

PIVOT qryFullDataWarehouseView.[Time Quarter];
```

**Table 3: SQL Example**

Pivot table and pivot chart views are also included in the application. Figure 10 shows the appliance product group aggregated for two states. These high-level analytic tools allow users to "slice and dice" the data into any form they see fit. Instead of requiring a manager to view static queries from a traditional relational database, OLAP tools permit complete analysis flexibility allowing the manager to respond to his or her changing business environment. The user can select from a list of variables and develop new summaries of the data. You may recall that the hierarchal design of the data dimensions actually provides the data structures that support the option to "roll up or drill down the data." Actual meaningful analysis and realism is achieved because of the extensive amount of data contained the tblSalesFact table.

| Pivot Table For Sales Fact Table | | | |
|---|---|---|---|
| **Product Group Name ▾** Appliances | **Location Region ▾** All | | |
| | | **Location State ▾** | |
| | NC | VA | Grand Total |
| **Product Type Name ▾** | Sum of Sales Revenue | Sum of Sales Revenue | Sum of Sales Revenue |
| Air Conditioners & Fans | $156,004.07 | $9,724.72 | $165,728.79 |
| Air Purifiers & Accessories | $98,916.12 | $84,095.46 | $183,011.58 |
| Beverage Chillers & Centers | $90,477.01 | $18,783.96 | $109,260.97 |
| Compactors & Disposers | $122,559.67 | $31,676.99 | $154,236.66 |
| Cooking | $133,511.75 | $15,315.92 | $148,827.67 |
| Dishwashers | $80,631.21 | $10,699.72 | $91,330.93 |
| ENERGY STAR ® Appliances | $76,731.76 | $32,100.82 | $108,832.58 |
| Floor Care | $193,227.10 | $43,175.85 | $236,402.95 |
| Freezers & Ice Makers | $103,218.96 | $102,431.04 | $205,650.00 |
| Humidifiers & Dehumidifiers | $116,578.19 | $114,751.80 | $231,329.99 |
| Parts & Accessories | $90,342.32 | $72,323.99 | $162,666.32 |
| Refrigerators | $108,583.05 | $55,164.70 | $163,747.76 |
| Washers & Dryers | $62,471.14 | $42,468.42 | $104,939.55 |
| Water Heaters | $156,870.50 | $50,874.90 | $207,745.40 |
| Grand Total | $1,590,122.85 | $683,588.30 | $2,273,711.15 |

**Figure 10: Pivot Table Example**

# Classroom Efficacy

The data warehouse fact table modeling demonstration was designed with the idea of student hands-on class participation, supported by the ubiquitous nature of Microsoft Office and laptop computers. In general, a short lecture explaining the basics of the data warehouse schema is used as a lead in and functions as a review for students with database backgrounds or a high-level overview of data warehouse design for students that have not completed a general database course. It is the author's general observation that hands-on activities reinforce the important course concepts and technical skills much more effectively than standard lectures and assigned homework. The instructor or graduate assistant is present at content delivery and is available for immediate assistance, which is typically very important when teaching technical skills such as

Microsoft Excel or Microsoft Access. Of course, classroom activities involve more preparation, require more class time, and represent the classic tradeoff between content delivered versus achieved learning objectives. The lecture and hands-on demonstration was presented by the author in two different undergraduate courses with the expectation of comparing student reactions and their value assessments.

The first course is titled Business Information Technology 2405 and has the following course description: *Data collection, descriptive statistics, probability theory, and statistical inferential procedures, utilizing Excel.* The course is required for undergraduate business students at the Pamplin College of Business. It is typically taken in summer school following the freshman year or during the sophomore term. For the majority of the students, the course is their first in depth statistics course. Quantitative modeling with Microsoft Excel is most often a new experience for these students. Basic business statistics, such as descriptive statistics, point estimates, and hypothesis testing, are covered. The ability to model random variables utilizing standard Microsoft Excel functions, such as RAND(), RANDBETWEEN(x,y) and NORMINV(p,m,s), represent the applied learning aspect of the course and are by far the most important learning objective for the class. The section sizes for BIT 2405 are, unfortunately, large and make in class activities difficult at best.

The second course is titled Business Information Technology 4514 and is described as follows: *Study of the design of databases and data structures for supporting business. Topics include basic database structure and design, structured query language, database management systems, integration of backend database servers, data warehousing and mining, on-line analytical processing, and database application, security, and management.* This course is an upper level elective and is normally taken during the junior or senior terms. The course covers a set of standard database concepts such as database types, data modeling, entity relationship diagrams, the relational model, and structure query language. There is one required semester project where the students construct a relational database using Microsoft Access and populate the database. The students are required to document their project with an entity relationship diagram, describe all constraints such as primary keys and document several SQL statements and display the results. The section size for higher-level electives such as BIT 4514 are small and lend themselves to more participatory lectures as well as structured class activities.

While students from both courses, appreciated the active and contextual learning supported by a real business case, the learning results were surprisingly divergent between the two courses. The students in the quantitative methods course, BIT 2405, focused on a totally different aspect of the modeling exercise than the author expected. The author found that the students in the quantitative methods class really did not connect with the database concepts or with the value of being able to model a fact table with 500,000 records in minutes. What these students appreciated was learning how to model random variables in respect to a business context. Modeling the number of transactions per day over one year, using Lowe's Corporation as a supporting case, seemed to be the most interesting data warehouse dimension to model for the students. Undergraduate students are a "tough audience," but the students were appreciative of the opportunity to reinforce the concept of a random variable in an actual business context as opposed to a sterile textbook example.

The students in the database class quickly gravitated to the benefits of modeling a large data warehouse fact table. Their memory of random variables was essentially limited, so more discussion and explanation of the Microsoft Excel functions was required before the actual modeling of the data warehouse dimensions could continue. Student comments such as "now I see the value of auto copy" proved to the author that these students did understand how simple it is to construct a fact table of sufficient size that the existence and type of index design does matter.

# Implication for Future Practice

While the resulting cross tab queries and pivot tables are much more realistic than standard text-book examples, the project could benefit from further development by adding more realism to the data dimensions. The product and location data contained in the tblSalesFact table were created using the Excel function **=randbetween(x,y)** which produces integers over a given range and is based on a uniform probability distribution. Of course, not all Lowe's stores have the same trans-action volume, and not all products have the same demand. Public sales data is available at the product group level that could be used to create a custom discrete probability distribution. The addition of real product prices to the Product Dimension would create more realistic data analy-sis. Obtaining data from commercial retail marketing databases could yield more realistic sales data modeling. Additional refinement of the beta distribution parameters would also be benefi-cial.

Improving the Excel Web Query process and the associated "ETL" macros could provide for a more straightforward data download that includes price capture from the Lowe's website. Actu-ally, by creating the entire "ETL" process in Visual Basic for Applications, the process could contain more flexibility, editing power, and user control than provided by the macro environment. An additional idea for a separate project is incorporating SAS® into the ETL process. SAS® is an extremely power statistical analysis application freely available to students at most universities. The application contains an impressive list of data manipulation functions as well as numerous text manipulating functions.

Be increasing the fact table, tblSalesFact, by an order of magnitude, the student could experiment with index creation, partitioning, and memory management and how they affect overall database performance. The larger data set would allow the student to move beyond DSS, which is a reac-tion to a problem or opportunity, to data mining, which uncovers trends and opportunities.

Lastly, while a MS Access exercise lends itself easily to an undergraduate database class, a MySQL implementation would offer a graduate level database course additional advanced DBMS concept for investigation, such as index type selection and creation, query analysis, and query performance measures.

# Acknowledgement

The MS Access database and a MySQL script for creating an empty database are available by email.

# References

Codd, E. F., Codd, S. B., & Salley, C. T. (1998). Providing OLAP (On-line Analytical Processing) to user-analysts: An IT mandate. E. F. Codd & Associates.

# Supplemental Reading

Cornell, P. (2005). A complete guide to pivot tables: A visual approach. Berkeley: Apress.

Coronel, C., & Rob, P. (2007). Database systems: Design, implementation, and management. Boston: Thompson Course Technology.

Davis, R. (2008). Teaching project simulation in Excel using PERT-Beta distributions. INFORMS Trans-actions on Education, 8(3), 10.

Dhar, V., & Sten, R. (1997). Intelligent decision support methods: The science of knowledge work. Upper Saddle River: Prentice Hall.

Hoffer, J. A., McFadden, F. R., & Prescott, M. B. (2002). Modern database management. Upper Saddle River: Pearson Education.

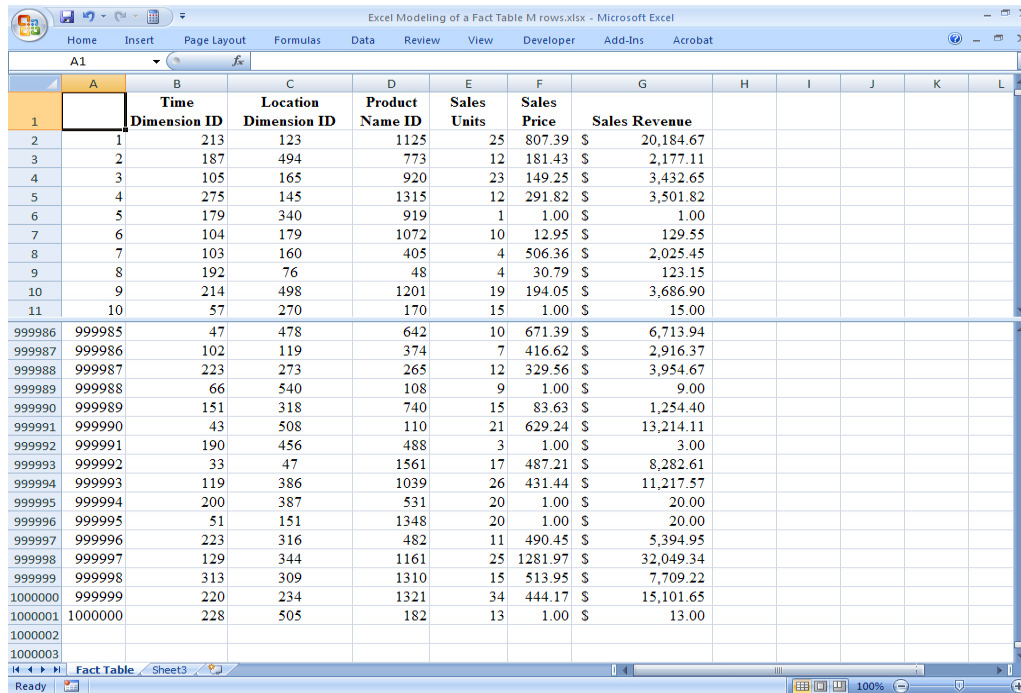Hoover's, Inc., Lowe's Corporation research page. University of Virginia Camp Library, Charlottesville, VA 25 November 2008. http://premium.hoovers.com/subscribe/co/factsheet.xhtml?ID=rfstfffjsskrct

McDonald, M. (2007). Access 2007: The missing manual. Sebastopol: O'Reilly Media.

Olson, D. L., & Shi, Y. (2007). Introduction to business data mining. Boston: McGraw-Hill Irwin.

Winston, W. L. (2004). Microsoft Excel data analysis and business modeling. Redmond: Microsoft Press.

# Appendix

## Microsoft Excel Screen Shot of Model Fact Table

| | A | B Time Dimension ID | C Location Dimension ID | D Product Name ID | E Sales Units | F Sales Price | G Sales Revenue |
|---|---|---|---|---|---|---|---|
| 1 | | | | | | | |
| 2 | 1 | 213 | 123 | 1125 | 25 | 807.39 | $ 20,184.67 |
| 3 | 2 | 187 | 494 | 773 | 12 | 181.43 | $ 2,177.11 |
| 4 | 3 | 105 | 165 | 920 | 23 | 149.25 | $ 3,432.65 |
| 5 | 4 | 275 | 145 | 1315 | 12 | 291.82 | $ 3,501.82 |
| 6 | 5 | 179 | 340 | 919 | 1 | 1.00 | $ 1.00 |
| 7 | 6 | 104 | 179 | 1072 | 10 | 12.95 | $ 129.55 |
| 8 | 7 | 103 | 160 | 405 | 4 | 506.36 | $ 2,025.45 |
| 9 | 8 | 192 | 76 | 48 | 4 | 30.79 | $ 123.15 |
| 10 | 9 | 214 | 498 | 1201 | 19 | 194.05 | $ 3,686.90 |
| 11 | 10 | 57 | 270 | 170 | 15 | 1.00 | $ 15.00 |
| 999986 | 999985 | 47 | 478 | 642 | 10 | 671.39 | $ 6,713.94 |
| 999987 | 999986 | 102 | 119 | 374 | 7 | 416.62 | $ 2,916.37 |
| 999988 | 999987 | 223 | 273 | 265 | 12 | 329.56 | $ 3,954.67 |
| 999989 | 999988 | 66 | 540 | 108 | 9 | 1.00 | $ 9.00 |
| 999990 | 999989 | 151 | 318 | 740 | 15 | 83.63 | $ 1,254.40 |
| 999991 | 999990 | 43 | 508 | 110 | 21 | 629.24 | $ 13,214.11 |
| 999992 | 999991 | 190 | 456 | 488 | 3 | 1.00 | $ 3.00 |
| 999993 | 999992 | 33 | 47 | 1561 | 17 | 487.21 | $ 8,282.61 |
| 999994 | 999993 | 119 | 386 | 1039 | 26 | 431.44 | $ 11,217.57 |
| 999995 | 999994 | 200 | 387 | 531 | 20 | 1.00 | $ 20.00 |
| 999996 | 999995 | 51 | 151 | 1348 | 20 | 1.00 | $ 20.00 |
| 999997 | 999996 | 223 | 316 | 482 | 11 | 490.45 | $ 5,394.95 |
| 999998 | 999997 | 129 | 344 | 1161 | 25 | 1281.97 | $ 32,049.34 |
| 999999 | 999998 | 313 | 309 | 1310 | 15 | 513.95 | $ 7,709.22 |
| 1000000 | 999999 | 220 | 234 | 1321 | 34 | 444.17 | $ 15,101.65 |
| 1000001 | 1000000 | 228 | 505 | 182 | 13 | 1.00 | $ 13.00 |
| 1000002 | | | | | | | |
| 1000003 | | | | | | | |

**Microsoft Access Screen Shot Examples**

# Biography

**Michael A. King** is currently a Ph.D. student and graduate assistant in the Business Information Technology department of the Pamplin College of Business at Virginia Polytechnic Institute and State University. He holds a master's of science in management of information technology from the University of Virginia, a master's in business administration from the University of North Carolina at Greensboro and a bachelor's of science in business administration from the University of North Carolina at Chapel Hill. He has over 20 years of professional experience as a systems and network engineer. His most recent position in industry was with Northrop Grumman where he held the title of senior network systems engineer.