



This is “Business Intelligence: Analysis of App Sales Data”, chapter 10 from the book [Designing Business Information Systems: Apps, Websites, and More \(index.html\)](#) (v. 1.0).

This book is licensed under a [Creative Commons by-nc-sa 3.0](http://creativecommons.org/licenses/by-nc-sa/3.0/) license. See the license for more details, but that basically means you can share this book as long as you credit the author (but see below), don't make money from it, and do make it available to everyone else under the same terms.

This content was accessible as of December 29, 2012, and it was downloaded then by [Andy Schmitz](#) (<http://lardbucket.org>) in an effort to preserve the availability of this book.

Normally, the author and publisher would be credited here. However, the publisher has asked for the customary Creative Commons attribution to the original publisher, authors, title, and book URI to be removed. Additionally, per the publisher's request, their name has been removed in some passages. More information is available on this project's [attribution page](http://2012books.lardbucket.org/attribution.html?utm_source=header).

For more information on the source of this book, or why it is available for free, please see [the project's home page](#) (<http://2012books.lardbucket.org/>). You can browse or download additional books there.

Chapter 10

Business Intelligence: Analysis of App Sales Data

10.1 Business Intelligence

LEARNING OBJECTIVES

1. Query sales data to spot meaningful trends
2. Distinguish between static reports, dynamic reports, and data mining
3. For a given situation, determine what type of business intelligence report is required to solve the problem

Introduction

In order to make strategic decisions about which products to feature in our store, we need to carefully analyze the sales and clickstream data. This type of data analysis is one form of business intelligence.

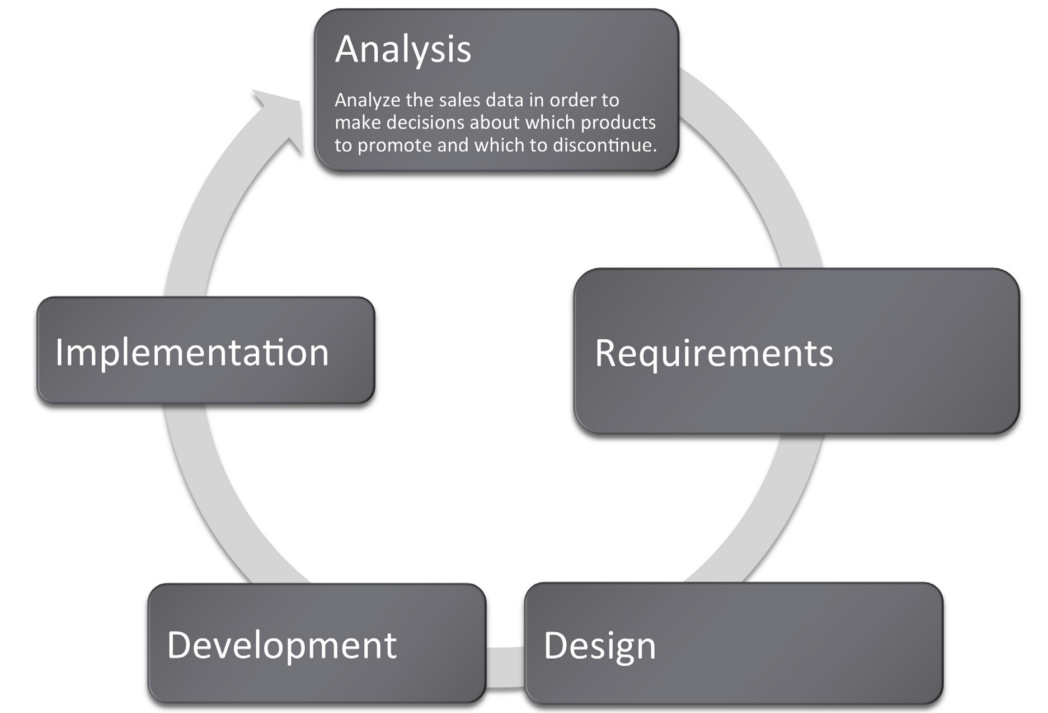
If there is one thing plentiful in the world today, it is data. At the heart of every information system is a database that captures transactional data. For example, who bought what, when, for how much, and so forth. It is useful to know about the architecture of the transactional systems so that it is not a complete mystery how the data is captured.

However, it is critical to know how to distill and analyze the captured data in order to make managerial decisions. For example, after summarizing thousands of records we might find a product selling particularly well with women in a particular age range living in a particular area. That meaningful information could be actionable in terms of the supply chain and marketing initiatives.

If anything in the world today there is perhaps too much data. Distilling that data into meaningful information is a key skill. There are a number of tools available to perform data analysis. These include spreadsheet programs such as Excel and database systems such as Access. Learning to use these tools will enhance your marketability.

Where Are We in the Life Cycle?

Many information systems projects are conceived of in a life cycle that progresses in stages from analysis to implementation. The diagram below shows the stages that we touch in the current chapter:



Kiva: Summarize Data to Produce Information

To illustrate the power of summary data, we will first show how it can be used as a marketing vehicle for a website. Impressive statistics can help encourage repeat business. The same marketing principles operate even for nonprofit organizations.

Kiva is a website that lets you make small loans (typically under \$500) to entrepreneurs in developing countries. The field of small loans is called **microfinance**¹. Microfinance institutions are an incredibly important resource to help third world citizens rise out of poverty. Surprisingly, the repayment rate of the world's poor ranges from 95 to 98%, far higher than the loan repayment rate in the United States. Over 80% of Kiva's loans are made to female entrepreneurs. They invest profits back into the businesses and improve the lives of their families.

Kiva works by pooling resources so that for example 50 people could lend \$10 each to total \$500. As part of its marketing effort Kiva maintains fast facts about their activities to date. For example, they report that they have nearly half a million lenders who together have lent \$161 million dollars over the last three years. These fast facts are gathered from the website's database after scanning millions of records and represent business intelligence. Not only does the information serve a marketing purpose, but it is also an internal scorecard to track the progress of Kiva's mission and influence decisions.

1. Making very small loans (typically under \$500) to entrepreneurs in developing countries. Most loans are repaid in six months to a year.

loans that change lives

LEND ABOUT COMMUNITY JOURNALS MY PORTFOLIO

Gifts | Help | Login | Register | My Basket 0 items

Facts & History

[Facts & Statistics](#) | [History](#)

Latest Statistics

The following statistics are updated nightly (between 1 - 3am US Pacific time).

Total value of all loans made through Kiva:	\$161,899,525
Number of Kiva Users:	759,281
Number of Kiva Users who have funded a loan:	482,346
Number of countries represented by Kiva Lenders:	206
Number of entrepreneurs that have received a loan through Kiva:	417,338
Number of loans that have been funded through Kiva:	222,389
Percentage of Kiva loans which have been made to women entrepreneurs:	82.06%
Number of Kiva Field Partners (microfinance institutions Kiva partners with):	120
Number of countries Kiva Field Partners are located in:	53
Current repayment rate (all partners):	98.90%
Average loan size (This is the average amount loaned to an individual Kiva Entrepreneur. Some loans - group loans - are divided between a group of borrowers.):	\$382.89
Average total amount loaned per Kiva Lender (includes relented funds):	\$212.87
Average number of loans per Kiva Lender:	6.28

[Field Partner Statistics](#)

[How Kiva Works](#) | [Press](#) | [Risk and Due Diligence](#) | [Terms of Use](#) | [Privacy Policy](#) | [Blog](#) | [Developers](#) | [Kiva Fellows](#) | [Jobs](#) | [Volunteer](#) | [Help](#) | [Donate](#)

© 2005-2010 Kiva. All rights reserved. Kiva is a U.S. 501(c)3 non-profit organization.

Lending to the working poor through Kiva involves risk of principal loss. Kiva does not guarantee repayment nor do we offer a financial return on your loan.

Kiva's Facts and History page is a business intelligence report. Note the sentence that appears under "Latest Statistics," which announces that the statistics are updated nightly (between 1 - 3 am). This is typical of business intelligence systems. Searching millions of records puts such a drain on the system that these activities are usually run during off peak hours.

What Is Business Intelligence?

The Kiva example is a form of **business intelligence**². Business intelligence (BI) is the delivery of accurate, useful information to the appropriate decision makers within the necessary time frame to support effective decision making.

By this definition all the work we have done with Excel would qualify as business intelligence since our deliverables contained accurate and useful information to support effective decision making. However, business intelligence is commonly understood to include distilling and analyzing large data sets such as those found in corporate databases. Extracting and analyzing information stored in databases is the subject of this chapter. It is very likely that at multiple points in your work career you will be asked to engage in just this type of analysis.

2. The delivery of accurate, useful information to the appropriate decision makers within the necessary time frame to support effective decision making.

Business intelligence is part of the big picture information systems architecture. Most systems in existence can be classified either as enterprise systems, collaboration systems, or business intelligence systems. The enterprise systems—taking orders for example—feed their data to the data warehouse, which in turn is queried to support business intelligence.

From a managerial standpoint, there are three factors necessary to make an effective decision:

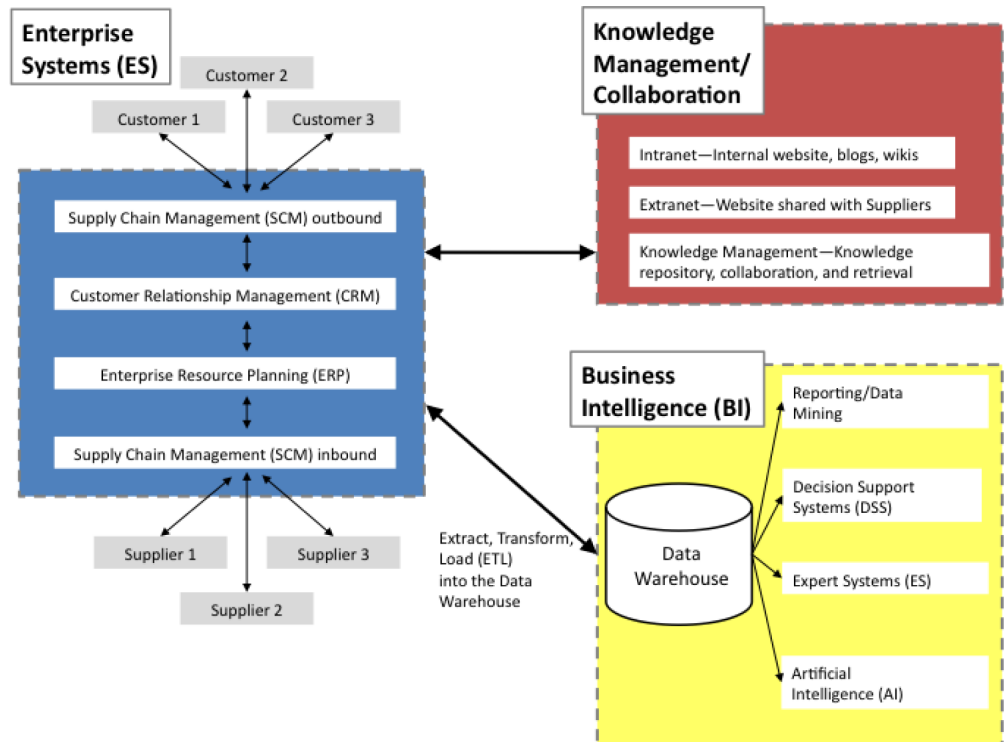
1. Construct a set of goals to work toward.
2. Determine a way to measure whether a chosen path is moving closer or farther from those goals.
3. Present information on those measures to decision makers in a timely fashion

For example, let's say our goals are to develop a clothing business that produces high quality products while lowering costs. We further determine that we will measure product quality by the percentage of products rejected by inspectors at each station. (Think about those inspector 99 tags that you find in pockets of your new clothes. The clothes you are wearing are the ones accepted by the inspector.) A relatively high rejection rate is a red flag to management requiring further analysis. Is this an overzealous inspector? Is there any pattern to the rejected products? Does one station in the factory tend to produce more rejects than the others?

We also need to see performance over time. Is product quality improving or getting progressively worse?

Let's say that our analysis determines that the high rejection rate comes from just one factory in Southeast Asia. We report the problem to management. They dispatch a team to review the plant. The review discovers child labor, abusive conditions, and very low morale at the plant. The horrible conditions are quickly reversed and the rejection rate returns to average.

Business Intelligence: Analysis of App Sales Data



3. A form of BI reporting with which we are most familiar—summary reports distributed at regular intervals.
4. Look like standard reports but with a major difference. They are interactive and allow the user to drill down to discover the source for the summary numbers.
5. A high level management overview of the data—sometimes depicted using dials and needles similar to an automobile dashboard. In a car you might not need to know your exact RPMs but you do need to know if you are red lining. Similarly, upper management might not need to know exact sales figures but they do need to know if sales are out of the normal range.
6. The process of uncovering the numbers that contribute to creating a summary number. Drill down is like being shocked at your ATM balance and then calling the to get a list of the withdrawals and payments made against your account.

The business intelligence portion of the information systems architecture. Note that business intelligence systems typically operate off of a data warehouse—a repository of data for the corporation. Each enterprise system contains one or more databases. The contents of those databases is routinely copied into the data warehouse to enable the BI analysis. The process of copying is called extract, transform, and load (ETL).

Business Intelligence Process

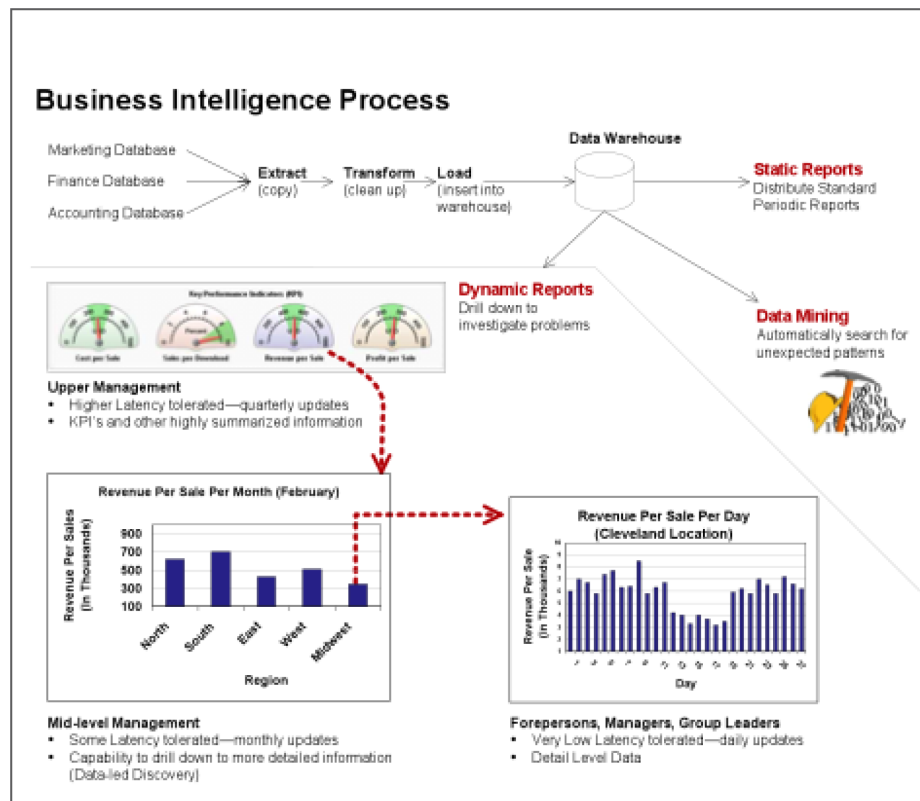
We will look at three types of business intelligence—static reports, dynamic reports, and data mining.

Static reports³ are by far the most common form of business intelligence. Most businesses have summarized standard reports already laid out and printed to assist in managerial decision making. For example, universities use enrollment reports to gauge which departments might need to hire more faculty. Credit card companies will request reports of persons with high credit scores to target credit card promotions. Similarly, the companies might target college students with good future earning potential. Marketers might look at sales figures for different stores and regions to determine where there are opportunities to run a sales promotion.

Dynamic reports⁴ look similar to static reports but online and interactive. A manager curious as to where a certain summary number on his **dashboard**⁵ comes from can **drill down**⁶ to expose the detail that contributed to that number. In

essence it is a fact finding tour where information discovered in each step gives clues on where to search next for information. For example, if sales in North America are down, then drill down to discover a problem in the Midwest region. Then drill down farther to discover a problem in the Cleveland, Ohio plant.

Data mining⁷ uses computer programs and statistical analyses to search for unexpected patterns, correlations, trends, and clustering in the data. In essence, it is fishing through the data to see if there are patterns of interest. One often cited example of data mining was the discovery that beer and diapers are frequently purchased on the same trip to the grocery store. Upon further inquiry marketers discovered that Dad picks up some beer on his trip to the grocery store to buy diapers. Marketers can use this information to place the two items in close proximity in the store.



The business intelligence process for dynamic reports is depicted here. The top half of the diagram shows how data finds its way into the data warehouse through the extract, transform, and load process. The dynamic report begins with an executive dashboard providing a high level view of the business. The dashed red arrows represent drilling down to find a reason for a pattern in the data. In this example, a downturn in North American sales is traced all the way back to a Cleveland, Ohio plant.

7. The process of fishing for patterns in the data using computing power because you really do not know what to look for.

KEY TAKEAWAYS

- Business intelligence is a way of uncovering trends and patterns in corporate data that might have strategic or operational significance.
- Most corporations already have the data that they need for business intelligence. However analyzing the data, presenting the results, and then following through on where the data leads, separates the winners from the losers in a competitive environment.
- Static reports, dynamic reports, and data mining are three different forms of business intelligence.

QUESTIONS AND EXERCISES

1. Managers are often most interested in exceptions—data that does not fit pre-established expectations. Describe how business intelligence can aid in this process.
2. Why do lower level managers require higher level detail in their information?
3. In what ways does fantasy football rely on business intelligence?

10.2 Databases

LEARNING OBJECTIVES

1. Determine which tables and fields in a database are needed to complete a query
2. Explain how data is captured in our Class App store
3. Explain how the Class App store data can be used for business intelligence

Introduction

In all of the forms of BI described above, you must actually store data to analyze. Organizations store their data in databases connected to their production systems. Here are some examples:

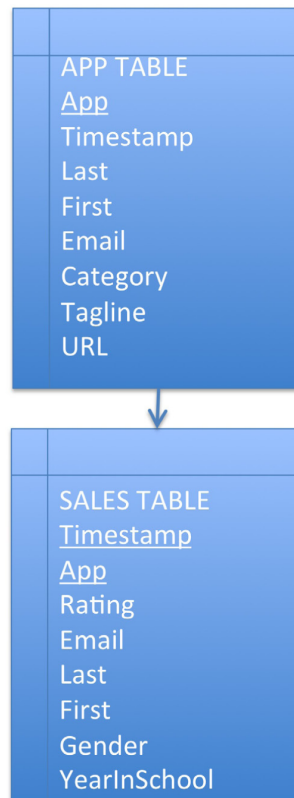
8. A database that consists of related tables and nothing but tables. The relationships among tables are established by repeating key fields. For example, the customer id field would be repeated in the account table for a bank.
9. One or more fields that together uniquely identify each record in a table.
10. To avoid duplicate data, store information common to each child in the parent table. Parent tables point to child tables in a diagram. However, the real link between them is made using a foreign key.
11. Repeat the primary key from the parent table in each corresponding record of the child table as a foreign key to link tables together.

- Banking transaction systems store data in databases containing information about customers, accounts, and transactions against those accounts.
- University enrollment systems store data in databases containing information about students, faculty, courses, and enrollment in those courses.
- Cell phone billing systems store data in databases containing information about customers, rate plans, and calls made.
- Credit card billing systems store data in databases containing information about customers, credit plans, and items charged.
- Supermarket checkout systems store data in databases containing information about customers, products, and buying habits of their customers. The loyalty card that you have swiped at the checkout ties all your purchases back to your name.

What do these databases actually look like? They consist of tables of data that are related to each other. This is called a **relational database**⁸. Each table must have a unique identifier that is called a **primary key**⁹. The database is organized into **parent and child tables**¹⁰ to avoid duplicating data. Data common to each child is stored in the parent table. Diagrammatically a parent table points to its child tables. Each parent record can have zero or more child records. To logically link the tables together simply repeat the primary key as a **foreign key**¹¹ in each corresponding record of the child table. To get information in and out of a relational database

requires a relational database management system (RDBMS) such as Microsoft Access. The goal of the system is to facilitate transactions while safe guarding the integrity of the data.

The theory behind database design is one of the most elegant areas in all of information systems. If you continue in information systems, you will see it in detail. However, for our purposes all we need to know is that data is typically stored in multiple files even if the report that we get is contained in a single file. Why? The simple answer is that we want to avoid **duplicate data**¹² by storing information common to each child in the parent table. Why do we care? Because duplicate data opens up the possibility that one of the duplicates will be different in an important way. For example you would not want your bank balance to be sometimes one number, sometimes another depending on which record happens to be called up by the database.



The data from the Class App store is stored in a relational database consisting of two tables—an APP table and a SALES table. The primary key of the APP table is App name. The primary key of the SALES table is the combination of Timestamp and App name. App name in the SALES table is also a foreign key linking each sale with its corresponding App.

12. Storing some fact about the world more than once. It is the number one sin in database design because duplicate data might become inconsistent.

Architecture of Class App Store

The Class App store created for this course has at its heart a simple database. Nonetheless, that database supports some fairly sophisticated functionality. The beauty of the Class App store is that it was created almost entirely without writing code, by using Google Sites and Google Docs.

The database consists of two tables—an App table and a Sales table. The App table captures registration information about each app. The Sales table captures sales information—who bought what and when.

Conceptually the tables are linked by what is called a **one to many relationship**¹³. One app has many sales. Every database has one to many links of this sort. The relationships are formed by the primary key to foreign key correspondence.

Once the architecture is established the next step is to get data in and out of the database. Data is entered into a database using **forms**¹⁴. For the App table, use the Register App form. For the Sales table, use the Purchase App form.

Data is extracted from the database using **reports**¹⁵. The listing of apps on the Class App store home page is a report.

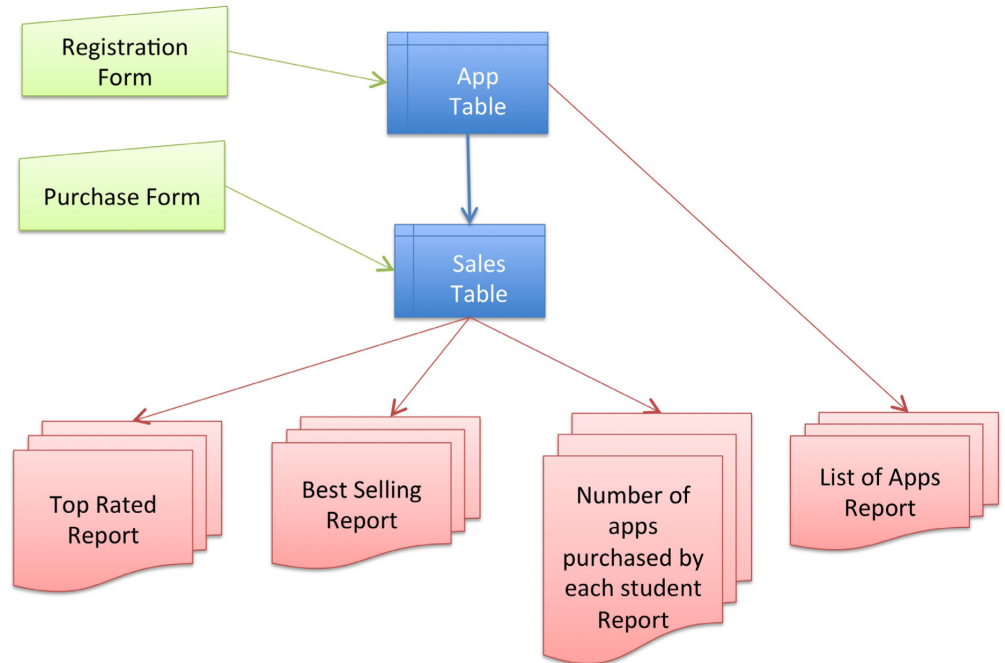
When the reports involve summary data, we would characterize that as meaningful information. For example, listing the best selling apps and the top rated apps qualifies as information. The number of apps purchased by each student is also information—it reveals how many students have completed the assignment.

And there are a variety of reports that can come out of even a simple database such as this. For example, a report might list the best selling apps for men who are freshmen. One can be quite specific as to the information extracted for analysis.

13. The type of relationship formed between parent and child tables. One parent record has zero or more child records.

14. A way to get information into a database. Each field in the form corresponds to a field in the database table.

15. A way to get information out of a database. Often a report will gather information together from multiple tables and present it as a single table.



Architecture of the Class App store. Even this simple database requires two forms and four reports.

Group and Summarize Data

We will analyze the sales data for our own app store to find trends in buying patterns for the class. Distilling that data and finding meaningful patterns is a form of business intelligence.

The important concepts here are to group and summarize data, and then to order and compare groups. For example, showing a list of the best selling apps. Creating this list requires counting total sales for each app and then listing those totals in descending order.

To do this in real time requires sending a query to the store typically written in a language called **Structured Query Language (SQL)**¹⁶. This is how we were able to get the store to display tables of best selling and top rated apps. The query looks similar to this:

16. Structured Query Language (SQL) is the language used by most relational database management systems. It requires very little code to accomplish powerful operations.

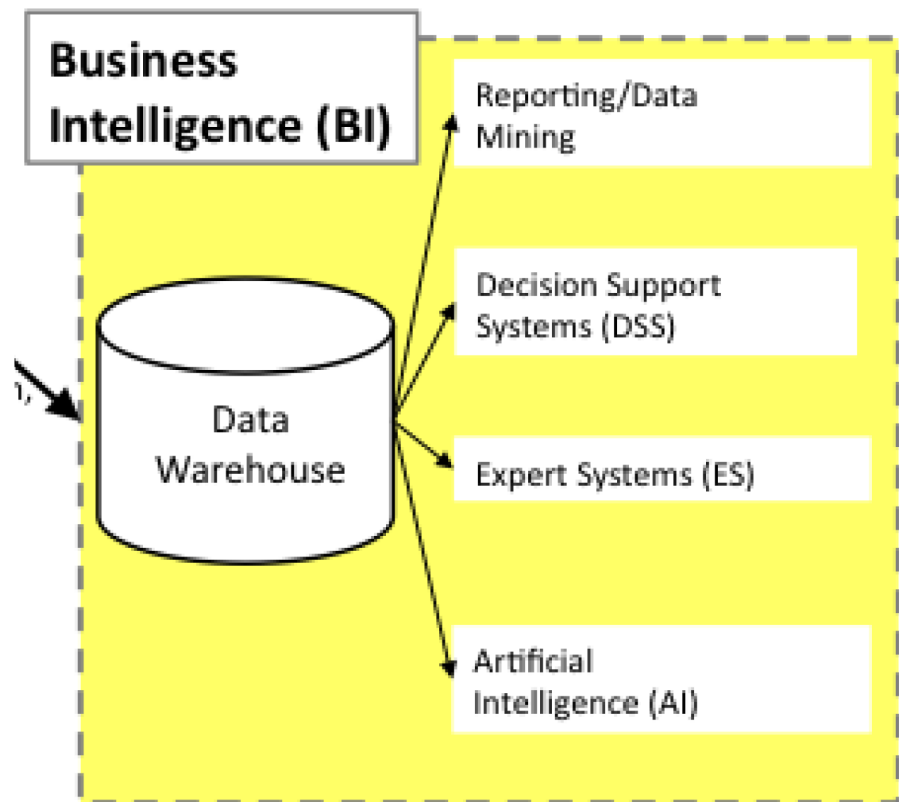
```
select App, count(Timestamp)
group by App
```


Periodically, perhaps once a day, data is copied from the company's many databases to a very large database called the **data warehouse**¹⁸. The process of copying the data is called **extract, transform, and load (ETL)**¹⁹.

- Extract — Copies data from one or more databases systems.
- Transform — Cleans the data so that related records in different databases appear in a consistent format.
- Load — Inserts the cleansed data into the data warehouse.

Why go to all this trouble? One of the main reasons is that analyzing the data on the production system would slow it down considerably leading to poor customer service. Another reason to copy the data is so that multiple databases can be merged into a single data warehouse.

It is the data warehouse that is analyzed to produce management reports.



18. A giant database that contains periodic dumps from many databases throughout the company. BI systems query the data warehouse to spot patterns and trends.

19. The process of copying data from many databases throughout the enterprise into the data warehouse.

Note the role of the data warehouse as the central repository for all the business intelligence data.

Latency²⁰ is the amount of time between the occurrence of a transaction and the loading of that transaction's information into the business intelligence system. In other words it is the amount of time that passes before a manager has a distilled report in hand analyzing the operation. Some managers are content to get a monthly update, others need daily or even hourly updates. It depends on the nature of the job. Ironically, lower level managers tend to need more up to the minute data. This is because they control the systems in real time. Upper level managers, by contrast, tend to focus on the big picture over a larger time horizon.

KEY TAKEAWAYS

- Multiple corporate databases feed into a large data warehouse that is used for querying the data.
- The greatest sin in database design is allowing duplicate data. Duplicate data has the potential to become inconsistent—sometimes one value, sometimes another.
- The higher up a manager is in the organization, the less detail he or she needs to see in the data. In fact, detail only becomes important to an upper manager when it is needed to explain an unexpected trend.

QUESTIONS AND EXERCISES

1. The transform step in the ETL process can be quite involved. Research and find an example of data that needs to be cleaned.
2. Explain why databases beyond one table require relationships among the tables.

Techniques

The following techniques, found in the Excel section of the software reference, may be useful in completing the assignments for this chapter: **Pivot Table**

20. The amount of time between the occurrence of a transaction and the loading of that transaction's information into the business intelligence system.

L3 ASSIGNMENT: SALES DATA ANALYSIS

How do you increase sales of your app in the store? In order to answer that question you need to examine your competitive position in the store. Your competitive position is defined by comparisons with other apps selling in the same category. So if you designed a music app, then you should compare with other music apps. There are a number of dimensions along which you can examine your competitive position: market share, unique visitors, conversion rate, personal sales, or cross selling.

Setup

To complete this assignment, you will need two files from your professor. The first is the sales file from the class store. The second is the content drilldown report from Google Analytics. Then create a new blank Excel spreadsheet with the column headings shown in the example. You need to include a row for every app that sold in your category. So if your category is music and there are ten music apps in the store, then you need to have ten rows including your own. Your row should be boldfaced.

Content and Style

Number and answer all of the following questions in the space below your spreadsheet. (Use merge cells and text wrap to make sure that your answers do not exceed the width of your spreadsheet.)

- **Market share:** Of all the sales in your category, what percentage does your app account for? How does that compare with the competition?
- **Unique visitors:** How many unique visitors came to your page in the store? How does that compare to the competition? What could you do to encourage more visits?
- **Conversion rate:** Of all the visitors to your page, what percentage actually bought your app? This is called the conversion rate. How does your conversion rate compare with the competition? What could you do to improve your conversion rate?
- **Personal sales:** The sales records reveal who bought your app. Some of those sales may be the result of you personally promoting the app to others in the class. What percentage of your sales are the result of personal selling? How many people did you try to sell that did not buy your app? What is your closing rate?
- **Cross Selling:** Of the people that bought your app, what other apps did they buy? What apps cross sell well with your app? Perhaps you could

promote your app on those pages and vice versa. To find this answer you need to import the sales table into Microsoft Access and then run both of the queries listed below. It is so worth it; the output is really interesting.

Deliverables

Electronic submission: Submit the Excel file electronically

Paper submission: Please print out the Excel file in landscape view using fit to page.

Competitive Data Analysis
22-Apr-11

Music Category (boldface the row corresponding to your app)

App	Category	Sales	Market Share	Unique PageViews	Conversion Rate (Sales/PageViews)	Bounce Rate (Visitors Exited Immediately)	Average Time on Page	Top 5 cross sells
1 WeTheDJ	Music	98		197	50%	71%	0:42	
2								
.								
.								
.								
n								
Total				Average				

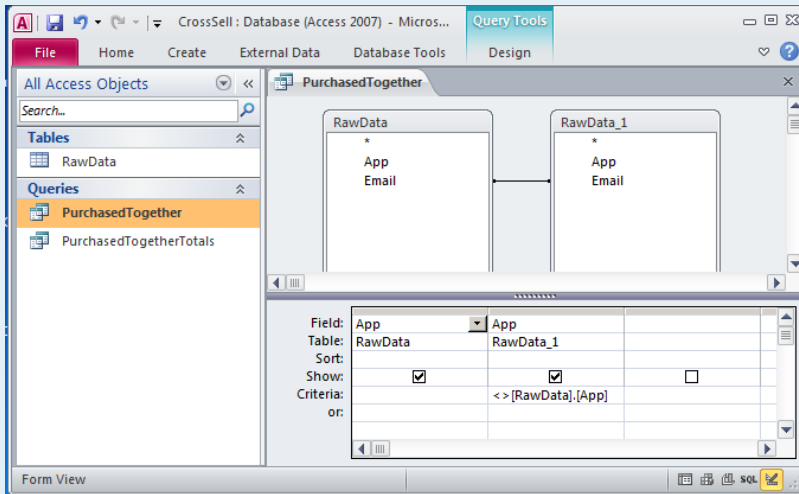
How to Find the Cross Selling Information

Begin by importing the sales data into a blank database in MS Access. Under the External Data tab select Excel and follow the screen prompts. The worksheet should come in as the RawData table or whatever name your professor calls it.

Now you will create two queries to run against the table. The first query, PurchasedTogether, creates a new row for each combination bought by a customer. For example, (WickedCrazyApp, CoolMusicApp), (BogusFlowerApp, IntenseAwesomeApp) and so forth. The logic of this query is to find all records from both tables where the emails match but the apps purchased do not. To create this query go to Create > Query Design and add the RawData table *twice* in the query design process. The second version of the table is called RawData_1. Drag a connector from one email field to the other to join the Email fields from both tables. (The example shown is simplified, showing only two fields.) Fill out the grid at the bottom to match the example. Run the query by clicking the red exclamation point.

The second query, PurchasedTogetherTotals, counts how many times each combination appears. The logic of this query is to count combinations no

matter who bought them. We have further limited the results to those counts greater than 4, but you can change this number as need be. To create this query you must add the PurchasedTogether query in the query design process. In other words you are doing a query of a query! Add the Total row to the grid by clicking the Σ , then fill out the rest of the grid as shown. Run the query and you have your cross selling data!



PurchasedTogether creates a new row for each combination bought by a customer. PurchasedTogetherTotals counts how many times each combination appears no matter who bought it.

