

# Understanding Business Intelligence:

## An Introduction to Business Intelligence Concepts

Looking for a Business Intelligence (BI) solution for your company can be intimidating. BI uses its own special terminology and the database design concepts can be difficult to grasp.

So where do you start?

ArcherPoint has put together this document in the hopes that it will help anyone looking to implement a Business Intelligence solution understand the underlying concepts and be in a position to ask the right questions of any BI vendor.

### Why Business Intelligence is important

Most companies track and record thousands of transactions daily. Not just customer purchases – which might include information such as the customer, the products/items sold, the store from which the purchase was made, and the date and time of the purchase – but also transactions such as warehouse activity, inventory purchases, employee hours and time off, and daily operating costs.

In fact, most companies are virtually swimming in data. If only there was a way to collect this data in one place and make sense of it all with a simple report (or set of reports). Companies who can successfully extract pertinent data from their mountain of information are gaining unique perspectives of their business, enabling them to become leaner and more competitive.

The process of collecting, organizing, and analyzing business data and turning it into useful and actionable information is commonly referred to as Business Intelligence. With business intelligence, companies have greater insight into their organization, yielding new opportunities, corrections to existing procedures or processes, competitive advantages, and more, including the ability to:

- Identify top-selling products by region, store, or sales person
- Identify trends, both good and bad, early on
- Generate ad-hoc financial reports
- Track competitors in their area
- Compare information about customers, products, prices, and costs over time

But before you run and write an SQL query to start extracting your business intelligence information from your company's database, there are some things you might want to consider first. And to do that, we'll have to define a few...no, many...terms.

### What's different about ArcherPoint?

Our clients see their business systems as unique and important to their competitive position. They prefer software and technology to support their business processes, not the other way around. Our clients are looking for a strategic partnership with the company that implements their software, not just a vendor. They demand the most talented, experienced professionals who understand how to quickly, efficiently solve their business issues. If this describes you, we invite you to contact us to learn more about our people, company and solutions.

## Not all databases are the same

Well, it can be argued that they kind of are, but I am referring here to how the underlying database is set up and configured – or what is known as the *database schema*.

To simplify things, let's think of a database as a group of tables. Each table has rows and columns, and where each row and column intersect is a cell...and each cell contains a single piece of data. Think of a table as a single Excel spreadsheet and a database as a set of Excel spreadsheets.

Most databases today are *relational databases*, and, typically, the goal is to minimize repetition of data across the tables. So, for example, there might be a single table for customer data (name, address, phone, email, and, of course, the primary key of the customer ID) and another table for products (manufacturer, product name, SKU, price, and, of course, the product ID). When a sale is made, one customer may purchase multiple products – but each product is not listed exhaustively for each transaction, nor is the customer information. Instead, in the table that records this sales transaction, all that would be required is to record the product IDs and the customer ID along with the time and date information and the store location (probably a store ID) and maybe the ID of the sales person. In this way, the data has been minimized – with just the store ID, the customer ID, and the product IDs, the company can find out everything about the purchase and the person making the purchase, from the type of person who buys a particular product to the time of day when the most sales are made at this store.

---

**The process of collecting, organizing, and analyzing business data and turning it into useful and actionable information is commonly referred to as Business Intelligence.**

---

## Online Transaction Processing (OLTP)

Relational databases are best when the primary goal is to record many transactions – which is the case for most companies. In fact, in retail especially, the most important consideration is being able to record multiple transactions simultaneously across the enterprise, with safeguards to prevent two or more people changing the same information at the same time as well as the ability to keep backups of the transaction database in real time.

This type of system is referred to as an *Online Transaction Processing* or OLTP database, and these systems are great for recording transactions, but not so great at extracting and analyzing strategic data quickly. A different type of database is used for analysis.

## Online Analytical Processing (OLAP)

There are several reasons why running analytics against the production transaction database is a bad idea. The numerous tables create a complexity that makes analytics difficult to perform quickly, and this will impact database performance, which, in turn, will cause delays at the check-out line.

But there are other reasons.

During the time it takes to run the analytics, new transactions are being added – which may alter the accuracy of the analytics you are trying to perform.

Another problem is that the production database requires multiple access from multiple departments, including the sales associates, the warehouse personnel, and the accounting staff.

Typically, that is not the case for analytical data. Instead, the analysis is only performed on historical data by a few people – so simultaneous access by many users and the ability to record simultaneous transactions is not necessary. What IS necessary, however, is the ability to crunch massive amounts of data quickly. This necessitates a different way of looking at how the database is put together.

A database specifically designed for business analytics is called an *Online Analytical Processing* (OLAP) database. It might also be referred to as a Data Warehouse or a Data Mart.

## Characteristics of an OLAP database

One of the most distinguishing characteristic of an OLAP database is that is generally isolated from the transactional database. This guarantees that operations performed for analysis will not impact the production database. It also means that the OLAP database contains only historical data, with little or no updates to the data itself.

Another distinguishing characteristic is that there are far fewer tables in an OLAP database than is usually found in an OLTP database, and the idea of repeating data in the tables is no longer a consideration. Instead, the focus is on how quickly the database can access the information needed to perform the analyses required.

A third characteristic is that the OLAP database is primarily concerned with numerical data, also called measures or facts.

## Facts, Dimensions, and OLAP Cubes

The main table in an OLAP database is the *Fact table*, where the information of interest is stored. The data stored in the Fact table are taken from various *Dimension tables*, which represent some aspect needed to perform the data analysis. For instance, if we wanted to analyze the total sales figures of each of our stores for each quarter, we would have a Time dimension table (representing the time period from which data is taken) and a Store dimension (representing each of the stores being analyzed). The Fact table would then contain the sales figures for each store for each time period. This would be a Fact table joined to two Dimension tables. We could add another (third) dimension, for example, Product. Now we can show the total sales by product for each store for each quarter.

In mathematics, two dimensions define an area or an X-Y rectangular grid (length and width, like the surface of a tabletop), while three dimensions define a space (length, width, and depth) or X-Y-Z coordinate system. A rectangle with depth defines a box, or cube (like a six-sided die or an alphabet block). So, an OLAP Fact table with three dimensions is called an *OLAP Cube*. Now, technically, a cube is equal along all its edges, and an OLAP Cube is not necessarily symmetrical, so it should probably be called an "OLAP Box". But the term "OLAP Cube" is commonly used in this sense.

Mathematically, we can define more than just three dimensions; we just can't draw or visualize more than three dimensions very easily. For instance, in addition to store, quarter, and product, we could also look at the sales data by customer, by payment type, by sales associate, and on and on. This would give us a Multi-dimensional OLAP Cube, sometimes called a *Hypercube*. However, many analysts use the term OLAP Cube to include Hypercubes as well.

Each dimension table contains facts, called *Members*. When possible, members are grouped into hierarchies, which are represented as parent/child relationships. For instance, Time can be represented as years, quarters, months, days, hours, or minutes. Day would be a "child" of Month, but a "parent" of Hour. This allows the analyst to look at data by year, but also by months, days, and hours.

As much as possible, Fact tables usually only contain the key fields needed to connect the Fact table to the various Dimension tables and Measure fields which contain the *Measures*, or facts, being analyzed.

The Dimension tables, by contrast, not only contain facts, but can also provide additional information associated with them, called *Attributes*. For instance, a Product dimension table may contain attributes such as the product name, manufacturer, and SKU.

## OLAP Schema

An OLAP database schema will typically take the form of either a *star schema* or a *snowflake schema*. Both seek to minimize the number of tables needed to get to the data of interest.

In the simplest schema, the star schema, the Fact tables can reference one or more Dimension tables, forming a pattern resembling a star.

The snowflake schema has at least one Dimension table that is itself associated with another Dimension table. The snowflake schema is used to improve performance for specific database queries.

There are basically three types of OLAP databases: *Multidimensional OLAP* (MOLAP), *Relational OLAP* (ROLAP), and *Hybrid OLAP* (HOLAP).

- **MOLAP** – Data is stored in multidimensional cubes and is not relational, which helps speed up query performance, but limits the amount of data it can process.
- **ROLAP** – Data is stored in a relational database, which increases the amount of data it can handle, but causes performance to suffer.
- **HOLAP** – Combines both types to take advantage of MOLAP for faster performance when working with summary data, but uses ROLAP when detailed information is needed.

Common operations performed by OLAP databases are *Slice and Dice*, *Drill Down/Up*, *Roll Up*, and *Pivot*.

- **Slice and Dice** – A "slice" of a cube is selecting data from the cube by fixing one dimension to a single value. For instance, the analyst might look at sales for all products for all months in the year for a single store, or sales for all stores and all products for a specific month. "Dicing" refers to selecting a range of possible values along one or more dimensions, for example, sales for the first quarter for all products in only the southeastern stores.
- **Drill Down/Up** – Allows the analyst to look at summary data ("drill up") or more detailed data ("drill down").
- **Roll Up** – Summarizes or aggregates the data along one or more dimensions.
- **Pivot** – Rotates the cube so the analyst can see data along a different axis. For example, instead of looking at total sales of each product at each store year-by-year, the analyst might pivot the table to show how each product sold every year store-by-store, representing the same data from a different point of view.

## Moving from OLTP to OLAP

At some point, data from the OLTP database needs to be moved to the OLAP database, a process called *Extract, Transform, and Load* (ETL).

- **Extract** – The data that will be placed into the data warehouse (OLAP database) does not necessarily all come from the OLTP database. In fact, the data may come from different data sources and appear in different formats. Part of the extraction process is to convert the data to a format for use in the transformation stage.
- **Transform** – All extracted data is manipulated into the same data type for storing into the data warehouse. For instance, one system might use '1' for "yes" and '0' for "no". Another might use the characters 'Y' and 'N'. The transform process would format all of these into a format the target database will use.
- **Load** – Here, the transformed data is loaded into the OLAP database.

ETL is a time when data can be "scrubbed", for example, invalid entries are discarded or corrected (null values for required fields) or conflicting data is reconciled (addresses may have changed).

ETL is also when some critical decisions are made. Part of the reasons for creating the data warehouse is to quickly analyze the numerical data being generated across the enterprise. Unless all the data from every transaction is going to be preserved (requiring a large and complex database), some data will have to be *aggregated*, or summarized. For instance, rather than store all individual transaction records, the company may decide to simply store the data as hourly, daily, or monthly sales figures.

This is actually an important concept. If the OLAP database is initially configured to aggregate the data by month, the analyst will not be able to later see the data at a finer granularity, like days or hours, without going back to the original data source and extracting it again. Therefore, it is important to understand how the data will be utilized before setting up the dimensions, or OLAP Cubes.

## Conclusion

Obviously, there is a lot more that can be said on the subject of Business Intelligence and Data Warehouse design that we did not cover here, however, the goal of this paper is to assist the reader in understanding the basic concepts that play a part when considering any Data Warehouse or Business Intelligence solution, as well as the differences between an OLTP database and an OLAP database and some of the principles used in setting up a data warehouse. This information should also help you evaluate Business Intelligence solutions providers.

If you are looking for a BI solution, your next steps will be to research the tools your prospective vendors have available for:

- Establishing your OLAP schema,
- Migrating your production data to the data warehouse, and
- Analyzing your data.

---

### About Microsoft Dynamics NAV

Microsoft Dynamics NAV, formerly Navision, is the best-selling mid-market ERP product in the world. With over 60,000 companies and over 1,000,000 users relying on Microsoft Dynamics NAV solutions every day, it is the clear favorite among growing companies with unique requirements.

---

**ArcherPoint**

(866) 343-4517

[www.archerpoint.com](http://www.archerpoint.com)

1735 North Brown Rd, Suite 425  
Lawrenceville, Georgia 30043

 **Microsoft** Partner  
Gold Enterprise Resource Planning

 Microsoft Dynamics NAV