

The Blueberry Muffin and the Data Warehouse

by Lyn Ryan

At the supermarket recently I spied punnets of blueberries for the keen price of \$1.99, so I bought a box. When I returned home, I realised that, apart from just eating the fresh blueberries, I didn't really know how else to prepare them. I first scanned my recipe books for ideas, but found none that were appealing. Like most people, I have eaten blueberry muffins and thought I would like to make some of those, but my recipe books appeared to contain no recipes for muffins. The next stop was the internet, where I scanned millions of ideas for blueberry pancakes, blueberry pies, blueberry sauces, blueberry cheesecakes, blueberry desserts of all descriptions, and of course, countless recipes for blueberry muffins.

There were moist blueberry muffins, blueberry muffins with a crusty top, blueberry and chocolate chip muffins, and many more. Having selected a recipe, the thought occurred to me...the quest for a blueberry muffin recipe is not unlike the process for using a data warehouse.

Instead of looking for a blueberry muffin recipe, imagine I was the CEO of a company, looking for a new product or perhaps a way of increasing profits. In the marketplace, I see the punnet of blueberries and something makes me very interested. I am not interested in the plain blueberries, I am interested in value-adding to the basic product, that is, I am interested in making the blueberries into something more enticing. Perhaps a flash of imagination, a stroke of inspiration, and, in my mind's eye, I picture a delicious morning tea...or a fabulous new product that will shake up the market place, or make the bottom line look extraordinary.

Just as I returned to my kitchen to check my resources, the CEO may return to the office wanting to check her own resources. Firstly we each have to formulate some questions. What shall I make with my secret ingredient? Do I know how to make that product? Do I need help with this task? What help do I have available to me? Is this help or information adequate for the task at hand?

I determine I need to research these questions further. I turn to my recipe shelf. One recipe book may organise information well, and have an index that enables me to look up the words 'blueberry' or 'muffin' so I can quickly determine whether there something in this resource to help me or not. Another recipe book may not have an index, or may classify recipes according to cuisine types (Italian, French, Thai) or meal courses (entrees, main meals, desserts). These structures do not allow me to easily search for blueberry muffin recipes.

Even if each book had a detailed index, what becomes obvious is that I don't have an overarching index that enables me to locate information across all recipe books or enable me to determine if that information is lacking from my recipe book collection. While I may have much information on my recipe shelf, it is difficult to use and poorly organised. The only way to answer my question is to manually study many books and collate my results in a time-consuming way.

Note that I may have purchased these recipe books separately, thinking each one was a valuable addition to my kitchen skill base, however, this motley collection of recipe books may make the current blueberry muffin task harder, not easier.



The Blueberry Muffin and the Data Warehouse, Continued

Imagine yourself back in the CEO's shoes. You have lots of systems generating lots of reports. Some systems have loads of data, and perhaps generate information-rich reports; others are perhaps not so good or so well organised. To make matters worse, the information contained in a single computer system is probably more like an encyclopaedia than a single book, and, over time, that encyclopaedia has been added to and rewritten by several authors all with different ideas about how to organise, classify and process information. Most importantly, there is no way I can get one report collating data from all systems to answer my relatively simple question: 'Do I have a recipe for blueberry muffins?' Here's where a data warehouse can come in handy.

A data warehouse takes all the information from other computer systems and reclassifies the information into a single structure, or a data model, so that the database can locate and report on all information collated there. Through the magic of technology, a data warehouse data model resolves problems with duplicated data or similar but different information. For example a data model is able to ensure that the record of a customer called Mr Bill Jones in one system is associated with the record of the same customer who happens to be called Mr William Jones in another system.

The internet search for blueberry muffin recipes is a closer analogy to a data warehouse search. Multiple resources of various kinds are available to me at the touch of a few keystrokes. A search engine like Google enables me to tailor queries, search across various sources and scan the results which have been placed in an order of relevance. However, unlike the contents of the Internet, the contents of a data warehouse is carefully designed, controlled and maintained.

A 'blueberry muffin' search on Google yields just over half a million results, far too many for a single person to read through and comprehend. In order to make sense of the query results, I may need to reformulate my questions and refine my search again. For example, I may only want to show blueberry recipes that contain ingredients that I know are in my pantry, or to only show recipes that contain ingredients I can obtain easily. Needless to say that requires me to inventory the contents of my pantry and be sure I have sufficient quantities of the relevant ingredients.

But, you might ask, how did the blueberry muffins turn out? Ah, actually making the muffins is another project altogether!

