

Spreadsheet vs. Database: Hidden cost of a spreadsheet solution

Andrew Powell
Paisley Software Solutions, Inc.

Reading this could save you 25% of your salary overhead!

Today more than ever, companies employ electronic record keeping. Doing so has been a proven cost effective solution. When it comes to storing data in an electronic medium, companies have many choices. Companies can save data in text files, word processing files, spreadsheets, database files, and many other formats.

Although company data can be maintained in many formats, not all formats are as cost effective as they initially seem. While some formats have small initial cash investments for purchasing the software, over time the cost of running a business with that solution can be much, much greater. That is, some formats that are easy to implement often have hidden costs spent maintaining the data and deriving information.

These costs are hidden in the expense of your employees working and re-working the same information to maintain data and produce reports.

A simple look at the definitions below will provide a basis for the following paper.

Data: Atomic pieces of information composed of simple elements

Information: Processed Data

There is a cost to process and store information. Depending on the type of business you run, your needs and sophistications may vary. All businesses, especially small businesses, can benefit from a database solution when they have more than one person working with a set of data.

Below this section there is a description of the limitations of collaborative use of spreadsheets for data maintenance as compared to a database. But here is a simple calculation that can show how a database solution can help you.

Company A has 5 employees that maintain information in varying formats (Word, Excel, Outlook, Access, ACT...).

The Average employee salary is \$40,000.

Annual Salary costs for Company A are \$200,000.

If a database solution could improve efficiency and help reduce data errors and lost data by a conservative 25%, Company A could have \$50,000 more productivity each year.

There is also an intangible benefit: EMPLOYEE SATISFACTION. When it is easier for your employees to be effective and perform well, they will be happier. And so will you!

Data Storage, Maintenance, and Information Processing

When data is stored it is usually stored in a grouping of related elements: grids, tables, and spreadsheets.

Grids consisting of rows and columns can be implemented in: word processors such as Microsoft Word, spreadsheets such as Microsoft Excel, or databases such as Microsoft SQL Server. The choice depends on what you need for your application. If you are putting together a document that needs formatting flexibility, neither a spreadsheet nor a database can provide what is needed. They are simply too awkward and lack the features of a word processor. On the other hand, trying to do grid calculations in a word processing application can be difficult.

Microsoft Excel and similar spreadsheets are good choices when you need to organize simple data in a fast and cost effective manner. It is easy to use Excel to save customer and/or prospect data when it is a small set to be used by a single person or a small group.

When your data grows and more people need access to it, then a database is a better choice. Spreadsheets will record and collect information in an organized manner. As a spreadsheet grows in number of rows and complexity, it becomes more difficult to search, change, save and use.

There are easier ways to find answers than to scroll through rows of data. These easier ways also provide an intangible benefit, employee satisfaction. When it is easier for your employees to be effective and perform well, they will be happier.

Spreadsheets have the following 'growth' problems:

Repeating data.

At a hypothetical business, someone may make several contacts. Each contact would have his/her own phone, but they may share a common address. In order to record this information in a spreadsheet, the person would have to repeat the address information several times.

With a database, the person would create two tables: One with company information and another with contact information. The contact table would contain one column (CompanyID) for linking the two tables. No repeating information.

Updating multiple table links can require a lot of tweaking with a spreadsheet application.

Databases were designed to handle multiple table linking. One of the main database design goals is to create as many sub-tables as possible, in order to eliminate duplicated data. For information on database design look up 'Relational database modeling':

<http://www.utexas.edu/cc/database/datamodeling/rm/>

Sparse data.

If you are collecting sales data from customers, with sales data to be collected after each sale, one possible spreadsheet design is to use one row per customer, one column per sale. With this design, there might be many empty cells for sales on which a given customer did not buy items. This creates a lot of wasted space.

In a database, you would define two tables: one for customers and a second table for sales (and possibly a third or fourth depending on the complexity of information you wish to maintain).

The first table could have many fields that define the attributes of a customer, for this example we will use 3 fields (CustomerID, CustomerName, and CustomerWebSite).

The sales table will have the attributes that define a sale, for this example it will again be a simple set of attributes for a database design service sale (CustomerID, SaleDate, NumberOfHours, & HourlyRate).

The sales table would only have entries for actual sales there would not be any placeholder rows for non-existent sales. This solves the problem of wasted space.

Another benefit from this solution may be obvious, it's easy to select data for either customers or date ranges. You will be selecting from a known source and format to find information vs. piecing data to provide a one-time, makeshift report.

Data Entry & Validation.

With a spreadsheet, by default any cell can contain any kind of data. With a database, the data fields are typed (that is, information in that field must be specified as a number, as a character, or as a date) – only a certain type of data is allowed in a cell. Validations are optional in Excel; validations must be applied to a database. Most spreadsheet users do not validate their data. Experienced database developers always validate data.

Data Sharing.

Typically, just one person updates a spreadsheet. Databases are designed for multiple users, and each user can have different data permissions. Only a few users can access spreadsheets at a time (two with Office XP Excel). This creates a time and resource issue. If a spreadsheet is on a network and an employee is using it, an error that the file is in use and can't be opened will display (or the application may allow someone to open it as read only). This means that the employee desiring the spreadsheet data will have to wait, or that they can open up a copy that may or may not be up to date. This situation is further complicated when resourceful employees make copies of spreadsheets and work on them, hoping to eventually to combine the data back together.

Security.

A spreadsheet can be password protected, which assigns the specific password to the sheet or workbook. This is mild security at best. Many individuals often know the password. There is a specific user password that is assigned to only one user, such as a user's network login. Databases, on the other hand, have tremendous security features. The security can be applied to every object in the database. A single table can be viewable by one group, updateable and viewable another, and a third separate group can have permission to delete rows from the table.

Database technologies were built with multiple users in mind. Security was a primary concern, so the security is both flexible and strong.

Backup and fault tolerance.

Again with Databases designed for multiple users implementations, backup and data consistency are strong. Databases can be backed up to the second; data can live past a drive failure; or power outage. Databases also process data within a *transaction*. That is, data is only saved after all of the changes in the transaction group are verified and committed. With a spreadsheet data is only saved after the users remembers to click save (or if you installed the AutoSave add-in for Excel, when the time period elapses).

Search and Retrieval.

In a spreadsheet, sorting rearranges how the data is presented and stored. If the spreadsheet contains calculations that depend on particular data being in particular cells, then sorting may destroy the results of the calculations. If you insert or delete rows or columns, the spreadsheet is usually smart enough to change its formulas to follow the data. But sorting often involves data movement that is too complex for the spreadsheet to follow. Sorting presents a major opportunity to corrupt your data.

In a database, calculation formulas are not stored in the cells. Sorting the information only changes data presentation; it has no impact on data storage or calculations. Spreadsheets are cell-oriented; databases are column and table oriented. This implies different construction and debugging strategies for each.

Orphan records.

If your data is organized into header information/ detail information, it is possible with a spreadsheet to have 'orphan' detail records (e.g., no associated header record). With a database, it is possible to turn on a feature called 'referential integrity' which will guarantee that all detail records have header records.

Spreadsheet Errors

Errors are a natural part of the activity of building a spreadsheet, spell check helps fix many word processing documents, but there is no equivalent logic check functionality for either a database or a spreadsheet.

Spreadsheets are often temporary pieces of work. Generally, they are used to meet a particular need and then discarded, or used very little. This means that the full rigors of software engineering are often absent from their development. The ease with which an inexperienced user can produce plausible output also tends to work against the adoption of a more rigorous approach

How different Products rate in terms of processing Grid information

Spreadsheets

| Advantages | Disadvantages |
|--------------------------------|-------------------------------------|
| • Simple to use | • Repeated Data |
| • Graphics easier to setup | • Data entry |
| • Easy data duplication | • Data validation & Checking |
| • Cell formulas & Calculations | • Data sharing / Collaboration |
| | • Search and Retrieval |
| | • Error Debugging |
| | • Upper row limit (65000 for Excel) |
| | • Easy data duplication |
| | • Formula errors |
| | |

Databases, such as MS SQL Server

| Advantages | Disadvantages |
|--|---|
| • Multi- user sharing | • Harder to setup |
| • User level Security | • Expense of Database application greater |
| • Data validation & checking | • Knowledge of Structured Query Language (SQL) takes time to learn. |
| • Strong Search and Retrieval | |
| • Can eliminate repeated data | |
| • Data entry can be done with strong validation | |
| • Rules can be applied to data relationships (i.e. no sale can exist without a customer) | |
| • No upper row limit | |
| • Multiple Presentation formats | |
| | |

Contact Paisley Software Solutions for a free consultation on improving your data storage and maintenance.

Paisley Software Solutions, Inc.
4436 Utica Street
Denver, CO 80212

720-252-9240