2) A listing of the tools or features that you would like to see in a data management tool. If the desired tools already exist, then please reference their Web site addresses and clearly explain what the tools do. If the tools and/or features do not already exist, then you should clearly describe the functionality that you are requesting.

Tools/Features of Data Management Tools

The tool/ features of data management tools have been broken down into two groups:

* features that belong in nearly every data management tool or features that are very popular. For example, the following features can all be found in SQLite or phpMyAdmin, which can be downloaded at <http://www.sqlite.org/download.html> and <http://www.phpmyadmin.net/home_page/downloads.php>, respectively.

1. Store/retrieve consistent data – One of the main purposes of a data management tool is the ability to store and retrieve data by using a specific search or query. Without this feature data storage would be useless.

1. Modify/delete - More main features of a data management tool is the ability to modify and even delete data once you store it into the database. This simplifies the process of retrieving data, modifying it, and then storing it.

1. Analyze data(patterns) - After storing and eventually modifying data, a data management tool needs to be able to allow the user to analyze this data which can be done by showing patterns.

1. Performance/efficiency - Obviously the performance of a data management tool is very important to the user and data management systems can employ different techniques in order to improve efficiency. For example, a technique known as clustering can be used in order to “cluster” or group data that is related closer in memory in order to improve the time in sorting.
2. Atomicity - An “all or nothing” rule that states that if any part of a transfer fails then the entire transfer fails. This is important because if this is not the cause data could be lost if a transfer failed. For example, if an individual was transferring money from their savings account to their checking account and the transfer failed halfway, without atomicity, that money could just become corrupted and be gone forever.

1. User friendly/simplicity - User friendliness is important because if a database is difficult to use then its performance is irrelevant as more time will be wasted in trying to figure out the program. A user friendly system will also be chosen over one that is not.
2. Sorting - Data can be nearly useless if it is not sorted. In order to see how data compares to other data sorting is a necessary in a data management tool. For instance, if managing a group of people’s birth years, it would be useful to sort these into either descending ascending order.
3. Data abstraction - In order to simplify use for different types of users, abstraction creates a layer of security by preventing some users from accessing certain areas of data. A regular user should not have access to the same files as a programmer or a database manager.

1. Safety/security (encryption) - Data management tools should provide safety and security to the user. When using a data management tool the user should never feel that their data is at risk.
2. Sharing - When working with a database, users often work in groups and therefore it is necessary to be able to share the work with multiple users easily and quickly.

1. Reliability - Reliability is important because a data management system cannot be used if the data is not correct. So even if it can give an answer to a query immediately, it is useless to the user if it is not the correct answer.

* tools that already exist and can be seen in a present data management tool. For example, cloud storage is a feature that exists in numerous data management tools, however, there are many databases that do not possess this feature.

1. Scalability - This is important because a database is needed to be able to scale itself so that if given a larger task it is able to complete it or will be able to complete it if the system is enlarged itself. For example, if a database system cannot handle a growing amount of work, but it is upgraded and given a faster CPU and more RAM, it should be able to make use of this and become more efficient and handle the growing work. This feature can be found in Twitter (found at <https://twitter.com/>) which uses its own framework ontop of MySQL in order to scale its databases to handle the vast amounts of incoming tweets daily.

1. Portability/compatibility - If the database can be ported to a different environment (such as another operating system) or it is compatible to another environment, then it is much more useful as the user is able to manage their data from different computing environments. This feature can be found in Firefox (which can be found at <http://www.mozilla.org/en-US/>) which is availbale on nearly every operating system available.

1. Crash tolerance - As with any computer tool or software it is desirable to not have to deal with crashes. However, having a crash free system is nearly impossible and the next best thing is having a good crash tolerance. Or in other words, the ability to recover from an unexpected crash without any data loss or other problems. This feature can be found in Thunderbird (which can be found at [*https://www.mozilla.org/en-US/thunderbird/*](https://www.mozilla.org/en-US/thunderbird/)) as it autosaves your emails so that if it were to crash you will not lose the current state of your email when restarting the service.

1. Visualization - This roughly means representing data in visual way that allows it to be communicated easily and effectively. For example, placing data into a chart can dramatically improve the way that data is presented. This feature can be found in LibreOffice Calc (which can be found at <http://www.libreoffice.org/>) as it allows the user to select data and convey it into a table or graph very easily.

1. Redundancy - A major issue with data management is data redundancy. Having multiple copies of files is redundant and can take up storage for no reason. Thus, a data management tool should have the ability to take care of redundant files and determine whether it necessary to have two copies of the same file. This feature can be found in Duplicate Cleaner (which can be found at <http://www.digitalvolcano.co.uk/content/duplicate-cleaner>) as it scans a system and checks for duplicate files and deletes them.

1. Compression - Databases can be very large and storage consuming and compressing them can be very effective in order to save space. Compression is also useful for other features such as sharing. This feature can be found in SQLite (which can be found at <http://www.sqlite.org/download.html>). As the name suggests, this is a lightweight piece of software compared to a tool that is large and take up more storage space.

1. Isolation - Isolation, or multiple users accessing same file simultaneously is important for group purposes like sharing. For example, this lab is being done using Google docs so that our entire group can access and edit each part of the lab. As just mentioned, this feature can be found in Google Docs (which can be found at [docs.google.com/](http://docs.google.com/)).

1. Computation - Computation is important as it can save vast amounts of time as data is often needed to be computed to be useful to the user. For instance, data such as a company’s revenue needs to be computed with math operations in order for it to be useful to the company. This feature can be found in Microsoft Excel (which can be found at <http://office.microsoft.com/en-us/excel/>) as users are able to select amounts of data in various “cells” and perform mathematical operations.
2. Change/access logging - In order to keep track of changes made to a database a log is needed to be kept. This is important in order to know who is making changes or accessing the database. For example, if the data system crashes after changes are made the log will show what was being accessed at the time of the crash. This feature can be found in FileZilla (which can be found at <http://filezilla-project.org/>) as it provides the user with logging to file.
3. Cloud storage - Being able to store data into the cloud will eliminate the problems of limited storage. Cloud storage is available from Dropbox (which can be found at <https://www.dropbox.com/>) as users have the ability to store their files into the cloud and access them from any location.
4. Open source - This feature is not necessarily necessary, however, it is a feature that we would want in a data management system because it allows the user to have full control of the data management software. An example of an open source data management system is Drupal (which can be found at <http://drupal.org/>) as it is maintained and developed by a community of 630,000+ users and developers.