Database Security Settings:

Database security concerns the use of a broad range of security controls to protect databases including the data, the database systems, and the database servers against compromises of their secrecy, integrity and availability. It involves various types or methods of controls, such as technical, administrative and physical. A security mechanism allows us to enforce a chosen security policy. Two main mechanisms we have learned at the DBMS level are Discretionary access control, Mandatory access control.

**Security Risks that we have addressed in our project are:**

* Unintended activity or misuse by authorized database users, or database administrators.
* Malware infections causing incidents such as unauthorized access, leakage or disclosure of personal or proprietary data, deletion of or damage to the data or programs.
* Data corruption and/or loss caused by the physical damage of system or entry of invalid data, mistakes in database etc.

In order to overcome the problem of unintended activities of authorized users of our database, we are giving only read permission to all users and admin can only modify and delete the data. So we have given only select permission to all users.

Sample query for granting read permission on data is:

**GRANT SELECT ON gene\_fpkm TO mdc6d WITH GRANT OPTION;**

In order to over the problem of unauthorized user access, we have designed our web site in such way that user can only access the system by entering valid username and password. If user failed to provide valid credentials won’t access the database. And we are storing all passwords of users with MD5 encryption techniques, so that valid user can’t see other user password credentials.To overcome the problem of data loss and corruption, we are storing our database on two systems as a backup. In case of any loss we can restore it from backups.

**Other topics:**

**Hive:** Hive is a data warehousing infrastructure based on [Hadoop](http://hadoop.apache.org/). Hive is designed to enable easy data summarization, ad-hoc querying and analysis of large volumes of data by providing Hive QL

***Uses of Hive:***

1) Hive QL queries are implicitly converted into map reduce jobs.

2) Hive QL allows traditional map/reduce programmers to be able to plug in their custom mappers and reducers to do more sophisticated analysis that may not be supported by other languages.

3) Hive provides compaction and [Bitmap index](http://en.wikipedia.org/wiki/Bitmap_index)ing to provide acceleration and have different storage types such as plain text, [RCFile](http://en.wikipedia.org/wiki/RCFile), [HBase](http://en.wikipedia.org/wiki/HBase" \o "HBase), and others.

4) Hive has some built-in user defined functions (UDFs) to manipulate dates, strings, and other data-mining tools.

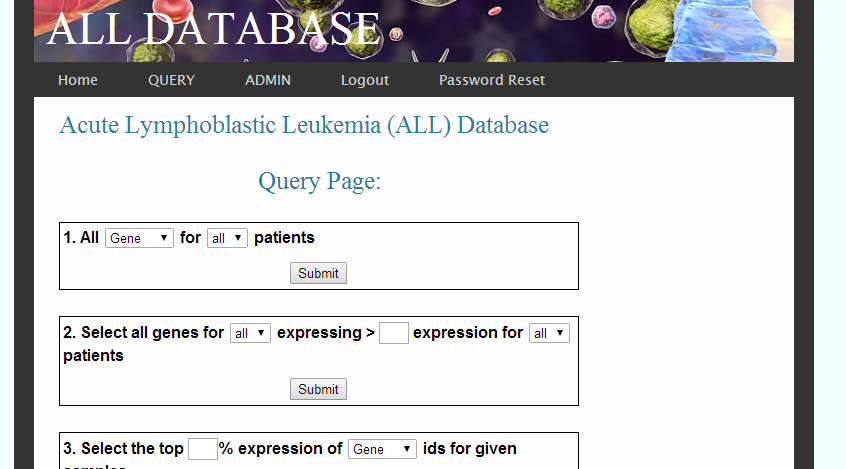
5) Hive operates on compressed data stored into Hadoop ecosystem, algorithm including [gzip](http://en.wikipedia.org/wiki/Gzip" \o "Gzip), [bzip2](http://en.wikipedia.org/wiki/Bzip2), [snappy](http://en.wikipedia.org/wiki/Snappy_(software)), etc.

Hive has lot of potential uses, we thought of implementing our project on Hive, but because of time constraints and other issues we did not that. But as our future ambition of our project, we are going to develop on Hive.

**Website design:**

We have chosen php to develop user interface for our database. In our databases, there are two kinds of users like admin, normal user. Admin can register new users, delete users. And we have query page which can be used to query the results from our database.

Screen shot of query page:

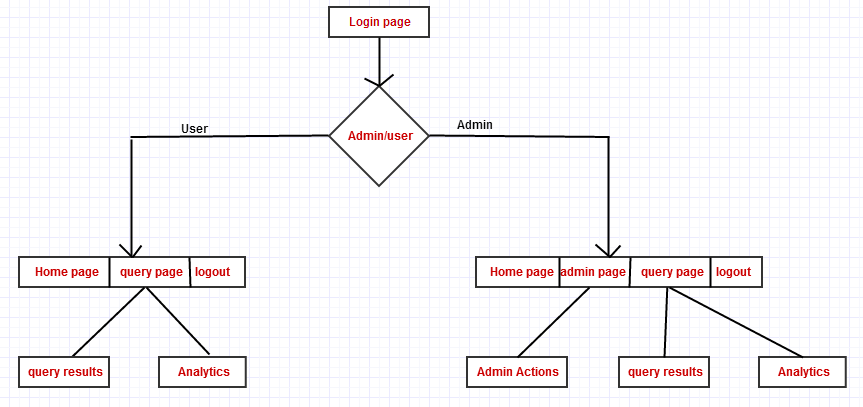


And we will show the results in table manner in corresponding pages.

Web site link:

<http://babbage.cs.missouri.edu/~cs4380sp14grp2/FinalProject/index.html>

**Flow chart of our website:**



**Analytics functionalities:**

While displaying results for queries, we are only displaying only top 10 rows. And providing a link to download entire results in excel files, which can be used for further analytics.

We are using pchart library which is built in PHP to make pie charts and bar charts for our query results. pChart is a PHP framework that will help you to create anti-aliased charts or pictures directly from your web server. You can then display the result in the client browser, sent it by mail or insert it into PDFs. We are using this pChart library and connecting to our database and displays the graphs.

**Simple bar chart for percentage of reads trimmed for sample\_ids of gene**

