Initial Plan: *Recommend Songs Cloud Project*

Group 18
Group member:
No. 15211173 Name: Sean O'Sullivan;
No. 15210749 Name: Mingyao Chen;
No. 15211374 Name: Siyi Kong;
No. 15210540 Name: Ye Qiu.

1. Data

The million song dataset was chosen for this project which includes huge numbers of songs and each "song" contains many details. Such as, title of the song, artist of the song, published time of the song, energy of the song, duration of the song, etc.
Dataset copy from http://labrosa.ee.columbia.edu/millionsong/

Clean and merge the dataset by Pig , Hive or Bash shell.

2. Application Description

Majority the application could be used in two ways, query a song by a part of artist name or a part of title, once the user input a word through our web page, such as "love", the application will query that songs contain "love" in its title or artist. Then, return a list includes both of the artist name and title of songs.

The user could also type in a whole title of a song. Then, the application will automatically find out ten similar songs refer to the "type in" song. The similarity searching algorithm is refer to the feature of the song, such as energy of the song, similar artist of the song, but not limit on it.

3. Management

3.1 Code:

Group members upload their code or data to Github in regularly. To make sure all the resource of each member that related to this project could be shared.

3.2 Time table of Meeting:

Every evening of Monday and Thursday, additional meeting noticed by email.

3.3 Task assignment:

Split the whole project into small units which include data cleaning and merging, query part

implementation in mapreduce, processing of “find similar songs” implementation in mapreduce, front-end web development, servlet development, configuration of application.

Details:

Data cleaning and merging:	Sean O'Sullivan
Query part implementation:	Ye Qiu
Processing of “find similar songs” implementation:	Mingyao Chen
Servlet develop:	Siyi Kong
Configuration of application:	Ye Qiu, Siyi Kong
Web front-end develop:	Sean O'Sullivan, Mingyao Chen

3.4 schedule

1. Week 7 Download and merge the dataset, Discussion of application design.

2. Week 8 Coding of MapReduce functionality.

3. Week 9 Implement the server part, Implement the web front-end

4. Week 10 .application configuration

5. Week 11 Testing