Project Description

The project that we are working on will be using MLB data. As we are using a dataset involving sports statistics, we will be making a client for users to easily look up stats and information through a command-line client application. The command line interface will act as a menu for the user to navigate through to choose what specific type of information they want to view, whether it be the plays that took place during a specific game, or the stats on a specific player, etc.

Project Functionality

As mentioned in the project description, the application will focus on two main aspects: game data and player data.

If the user selects game data, they will have to select a game, or even possibly, they will be able to select multiple games if they wish to do so. If they select multiple games, then the stats that will come up will be an average as well as computed totals over the selected games for statistics like average game time, average home team score and away team score, etc. If they instead choose a single game, they will be offered much more information to work with as they will be to view. Initially, most of the data that will be displayed on the application will be computed statistics that are collected through aggregations on the data in the tables. These can include global stats like total hits from each team, total runs, as well as individual stats for each batter and pitcher that played in the specific game. There are many different types of plays that can take place in a game whether it be the event that took place from a batter, the number of strikeouts from the pitcher, the speed of the throw, etc. We can take all this data and compute

interesting information that will be presented to the user and allow them to see a player's performance for that game.

The functionality for the player data will follow a similar route to the game data where they will be able to either view an individual player's performance over the course of the several seasons of games that exist in the database, or multiple players total and average performances. If the player is a pitcher or a batter, the information displayed will be different, but ultimately, the information will be computed using the tables that exist in the database similar to the stats for the game data. The main difference between game data and player data is that game data will have stats for the performance of the players in the game(s), whereas player data will be the performance of the player(s) over all the games in the database that they have participated in.