# Project background

My mother used to capture the Cardinals games on a score card as she watched the game.

I never did this and I never learned how to do it.

I noticed that no one seems to do this at the ballgame anymore.

My first inclination was to develop a scorecard application but there are already plenty of these types of applications available.

I did notice that many of the scorecard applications are marketed to little league and school age teams to capture the statistics of the games.

Then I thought that somewhere out there on the world wide web there had to be all of this scorecard play by play data for major league baseball.

I was right!

# Project description

Retrosheet, Inc. has gathered these play-by-play statistics for major league baseball games from 1921 to 2016. There are a few years in there that are no available. Retrosheet has the data for the regular season, post-season, and all-star games.

My project presents this data to the user

* Play by play
* Within a game
* For a major league team
* Within a major league baseball season

I am focusing on the 2016 Regular Season although my design should accommodate data from any season – regular season, post season, and all-star game.

# What problem does it solve

The Retrosheet data is available in comma delimited text files grouped together by season and then by team within the season.

My project takes this cryptic data for a major league baseball season

* analyze it
* parse it and load to individual text files based on data type
* load it into a database to individual tables based on data type
* pulls the data by play within game and presents to the user.

NOTE: there are 400k+ records that make up the 2016 major league baseball regular season.

The secondary result of my project is that the data is loaded into a database for that can be used for data mining.

# Who will use it

Baseball fans who want to relive a particular major league baseball game play by play.

Baseball data researchers

# What will you create?

The results will be –

* Individual pipe delimited text files for the different types of game data
* A database that holds the different types of game data
* A UI that displays the game data play by play

Note: Retrosheet, Inc. does require the following statement to be prominently displayed:

The information used here was obtained free of charge from and is copyrighted by Retrosheet.

Interested parties may contact Retrosheet at "www.retrosheet.org".

# Screen mock ups

# Sample reports

No reports are within this project scope

# Do you need to interview experts?

The Retrosheet.org site has a lot of information about their data although it takes some time to interpret and decode.

So far I have been able to figure out the game data and the reference codes – but the website does provide contact information.

I have researched different baseball questions via Google. There is plenty of information about baseball on the internet.

# Figure out your questions first

# Tally the results

# How will you store your data?

* First I build individual pipe delimited text files – one file for each type of game data.
  + These files are created by data
  + Within team
  + Within season
* Second the data from the individual text files are loaded into SQL Server database tables
* The database tables have foreign keys to the other tables to link of the data together

# Mock up data table and their relations to each other

# Determine best method of storing data; SQL, Excel, flat file, etc...

It is necessary to split the one file holding all game data for the season for a team into individual files.

I do most of the parsing of the data while building the individual pipe delimited text files.

The loading of the text files to the database tables is fairly simple.

# Break up the work into steps

* Research the structure and organization of the downloaded files
* Research the reference data that defines some of the codes
* Research the structure and organization of the data in the individual Team files
* Define each of the different types of data in the files
* Determine the parsing to be done on each of the different types of data
* Determine the file structure of the output files.
* Code the input of the team game data, parsing of the data, and output of the individual data files.
* Code the input, parsing, and output of the reference data files.
* Code the load of the data to the database
* design of Set Up UI – how does the user get the data into the application database
* design of season and game selection interface
* design of game play-by-play presentation
* code the Set Up UI, season/game selection interface
* code the game play-by-play presentation