

# Phase 1 report

## System Description

In this project we are looking to extract, organize and display election's data that ranges from 1789 to 2016. The report will describe how we will scrape each source website and organize its content to be inserted into the database. The database we're setting up using apache will be using mysql instead of the class's oracle. Organizing the data will be based on the E-R model we learned in class; in BCNF. The server will be responsible for receiving user requests, that is in select query form, and will request from the database; thus, giving the server a temporary relation that is turned into an html page to be displayed to the user.

## Data Sources

- <https://www.loc.gov/rr/program/bib/elections/statistics.html>
- <http://www.globalelectionsdatabase.com/index.php/index>
- <http://www.270towin.com/historical-presidential-elections/>
- <https://fusiontables.google.com/DataSource?dsrclid=225439#rows:id=1>
- <http://www.infoplease.com/ipa/A0781450.html>
- [https://en.wikipedia.org/wiki/List\\_of\\_Presidents\\_of\\_the\\_United\\_States\\_by\\_home\\_state](https://en.wikipedia.org/wiki/List_of_Presidents_of_the_United_States_by_home_state)
- [https://en.wikipedia.org/wiki/List\\_of\\_political\\_parties\\_in\\_the\\_United\\_States](https://en.wikipedia.org/wiki/List_of_political_parties_in_the_United_States)
- [https://en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_presidential\\_candidates](https://en.wikipedia.org/wiki/List_of_United_States_presidential_candidates)
- <http://www.fec.gov/disclosure/pnational.do>
- [http://www.huffingtonpost.com/2011/02/21/the-net-worth-of-the-american\\_825939.html](http://www.huffingtonpost.com/2011/02/21/the-net-worth-of-the-american_825939.html)
- <http://www.usatoday.com/story/news/politics/elections/2015/08/26/24-7-wall-st-net-worth-presidential-candidates/32409491/>

## Assumptions and Limitation

- User speaks/reads English language.
- Understands US election process.
- Know basic commands for SQL queries.
- User does not need authentication to access.
- Basic web browsing skills.

The top-level information flow diagram

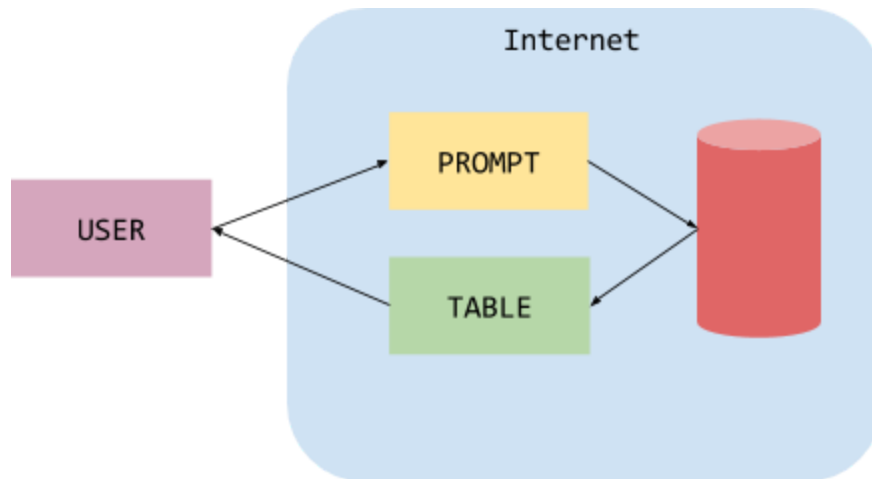


Figure: User Perspective

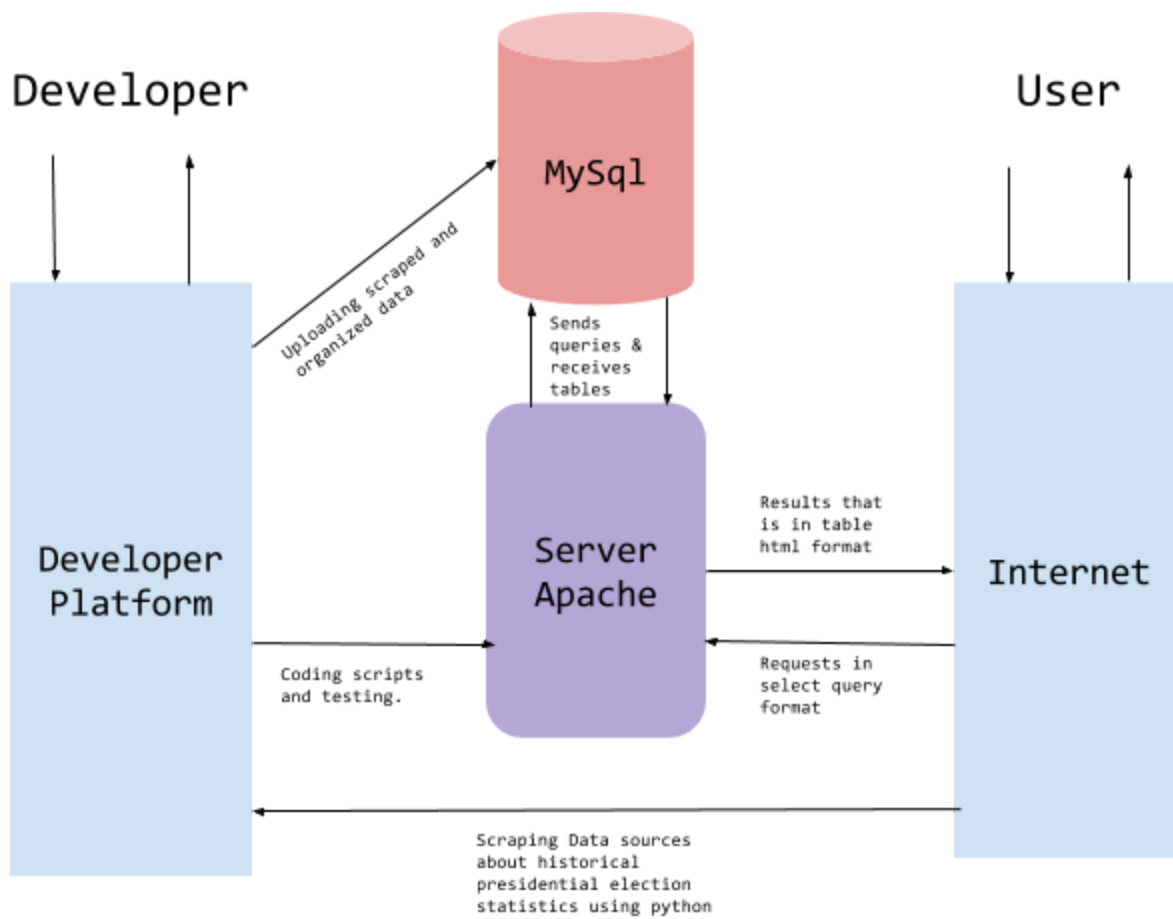


Figure: Developer Perspective

## The data documents that carry data between tasks

Data carried between the user and the server is transferred using Hypertext Transfer Protocol. However, data from web sources are going to be scraped using python tools such as scrapy and selenium. Uploading to the database will be done using python too using the following library: <https://pypi.python.org/pypi/MySQL-python/>

## Progress Report and Web Server

We are planning to use XAMPP since it comes equipped a control panel to interact with the Apache web server and MySQL database. Instead of using the php, we would like to link our own python script to scrape webpages as we build our database.

