Mattias Huber

Zoltan Batoczki   
Preston Peterson

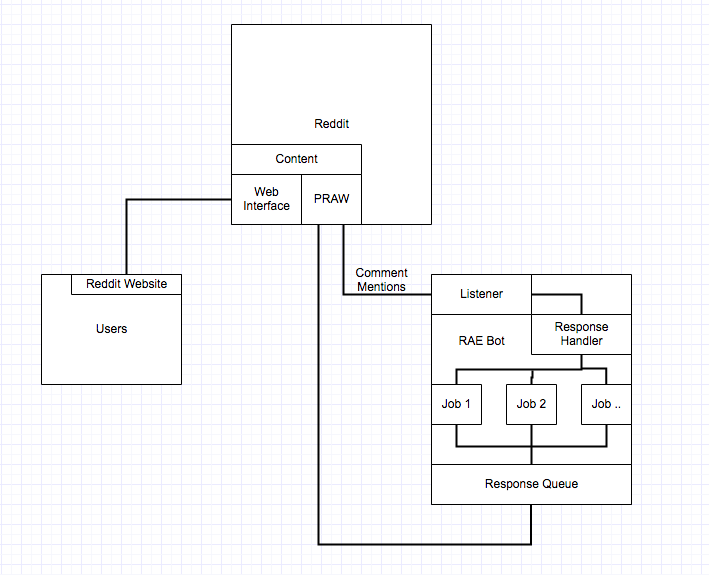
Phillip Porter

Heather Bradfield

Jesus Bamford

Reddits Analytics Engine Architectural Design

1. Include a conceptual diagram of the architectural design of your project (like the one shown at page 5 of the lecture slides):



2. Answer the architectural design decision questions shown at page 13 of the slides:

*Is there a generic application architecture that can act as a template for the system that is being designed?*

On the Python Reddit API Wrapper documentation website there is a ['call and response' bot template.](http://praw.readthedocs.io/en/stable/pages/call_and_response_bot.html)

*How will the system be distributed across hardware cores or processors?*

The bot client will need to run on a server and is not intended to be run independently by users. However, to be able to install the bot on new servers there will be scripts which install all the dependencies, such as python, PRAW, and matplotlib.

*What architectural patterns or styles might be used?*

The system is modeled after the pipe and filter architectural pattern, since transformations on the data are sequential:

1. The listener takes input in the form of a command and target from the user.
2. The command and target are piped to the request handler.
3. The request handler determines the nature of the request and pipes the data to the analytics modules.
4. The analytics modules generate the reports and send to the image upload module to upload the report.
5. The image upload module uploads the image to imgur.com and returns the URL of the uploaded image to the response handler.
6. The response handler replies to the reddit user with the link to newly generated and uploaded report.

*What strategy will be used to control the operation of the components in the system?*

The request handler’s job is to take the command and target found by the listener, determine the correct action, and respond by executing the correct analytics module.

*How should the architecture of the system be documented?*

The architecture of the system will be documented by using the 4+1 architectural views, namely the class diagram for the logical view, activity diagram for the process view, deployment diagram in the physical view, component diagram for the development view, and the use case diagram for the scenario view.

*What architectural organization is best for delivering the non-functional requirements of the system?*

A layered architectural organization is best for delivering the system NFRs. Layered architectures are great for NFRs because they support the ability to make future changes. Oftentimes a change in one component of a software may have adverse effects on NFRs related to other components. However, with a layered architecture, one layer can be completely revamped and it will only have adverse effects on its adjacent layers.

*How will the structural components in the system be decomposed into sub-components?*

The Reddit analytics engine is subdivided into the listener, request handler, image uploader, and response handler.

*What will be the fundamental approach used to structure the system?*

The bot must be able to be on all the time, autonomously responding and generating reports to the user.

3. Explain how the system characteristics (like those listed at page 15 of the slides) affected the chosen architectural design:

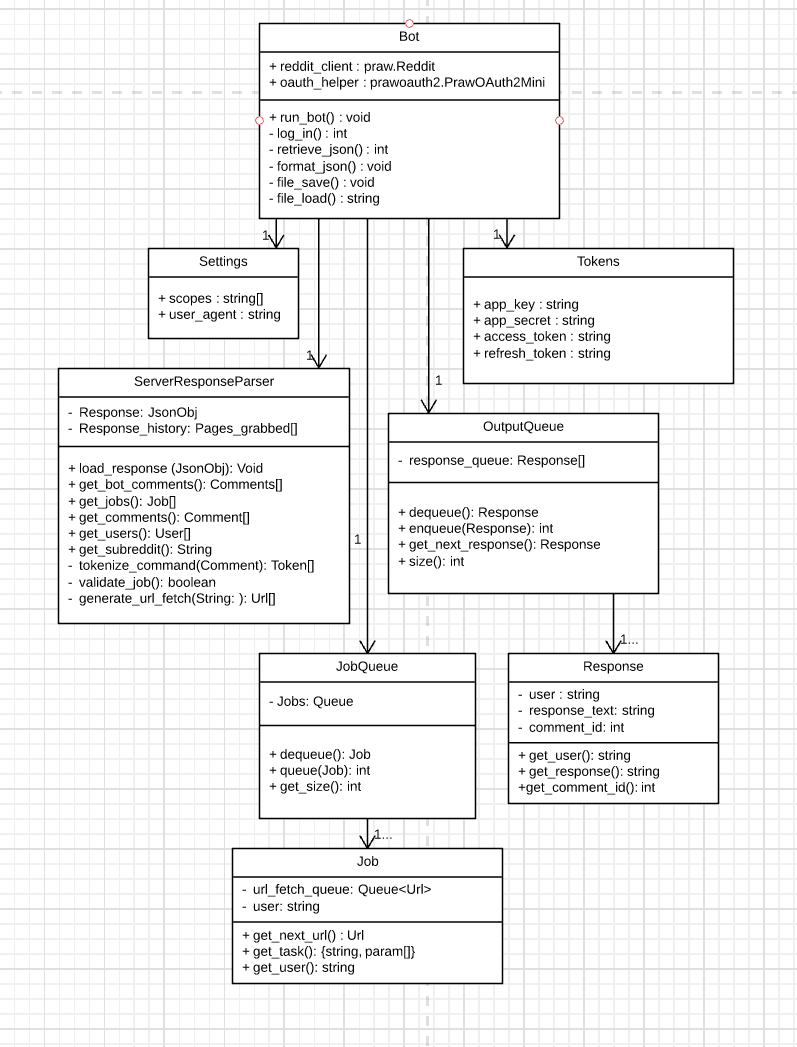
System Characteristics:

* Performance:
  + Uses “mentions” instead of scrapping Reddit every time our bot is requested. The use of “mentions” will allow us to not use a database which will minimise communications between components within RAE.
  + Utilizes Python’s multithreading capabilities as well as [PRAW’s multiprocessor capabilities](http://praw.readthedocs.io/en/stable/pages/multiprocess.html). The bot listener, request handler, and analytics jobs will all be in separate threads on separate cores, speeding up response time.
* Security:
  + Uses secret tokens and credentials for our Reddit bot and Imgur account. Only our immediate development group can modify the generation and/or posting of data. RAE is a public bot, but its features cannot be changed by users outside our development group.
  + Properly parse user input and detect injections of malicious code in order to prevent RAE from being compromised.
* Safety:
  + RAE will not hold any safety-critical data. All data from Reddit is public.
  + RAEis a system whose failure or malfunction will NOT result in death or serious injury to people, loss or severe damage to equipment/property, or environmental harm.
* Availability:
  + The RAE bot is infinitely running, waiting for a Reddit user to request it.
  + Includes server stability, bot stability, user error handling, API error and timeout handling in order to accommodate a large number of Reddit users using RAE.
* Maintainability:
  + Everything in the Analytics component is a method which requires a standard input and output. We chose this design in order to decouple the code.
  + Standard input and output increase the Analytics Component’s portability/extendability.
  + The Reddit bot was modularized so that more commands could be added later on.

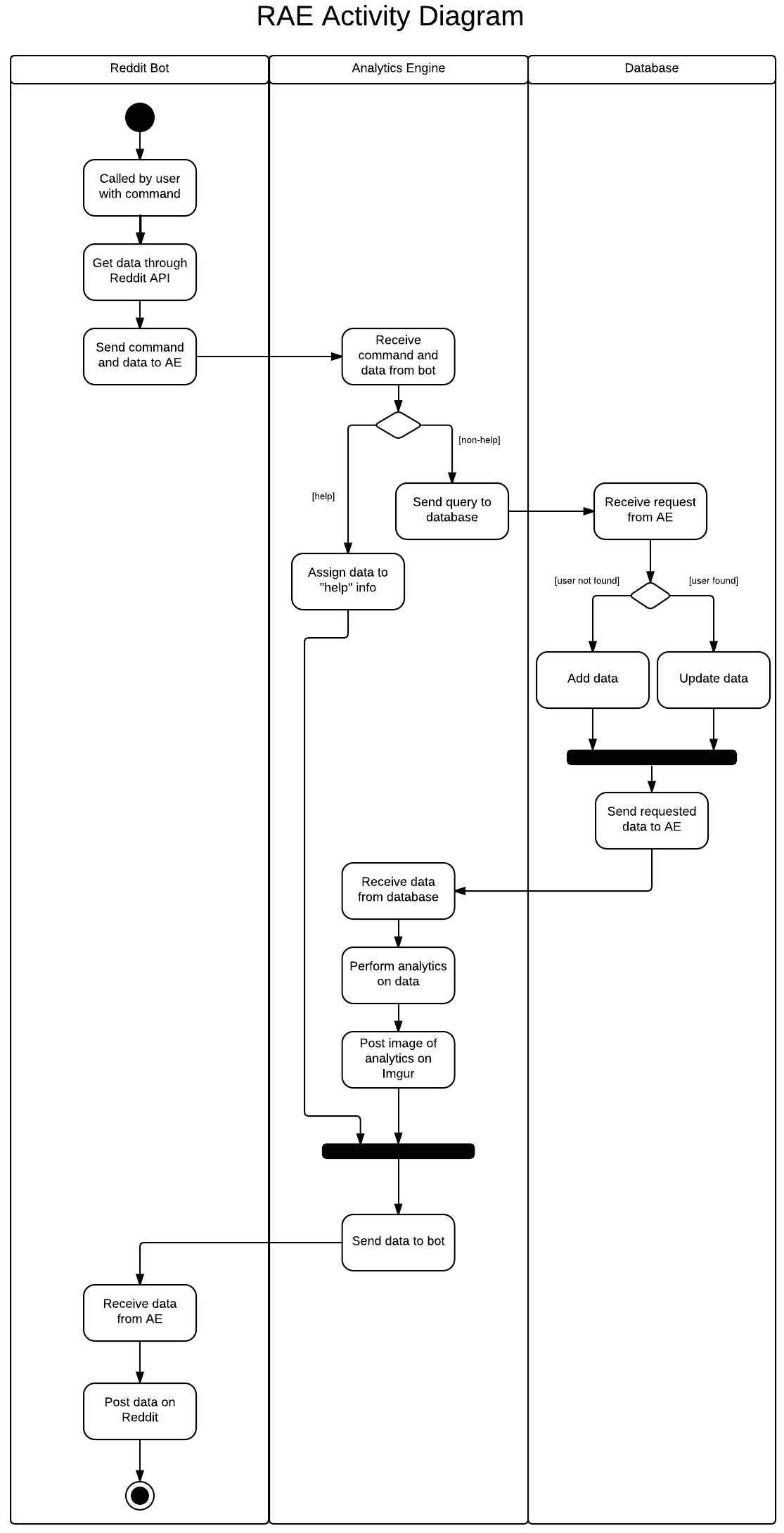
4. Prepare simplified 4+1 views of your system :

Diagrams were created using [Gliffy](https://www.gliffy.com/)

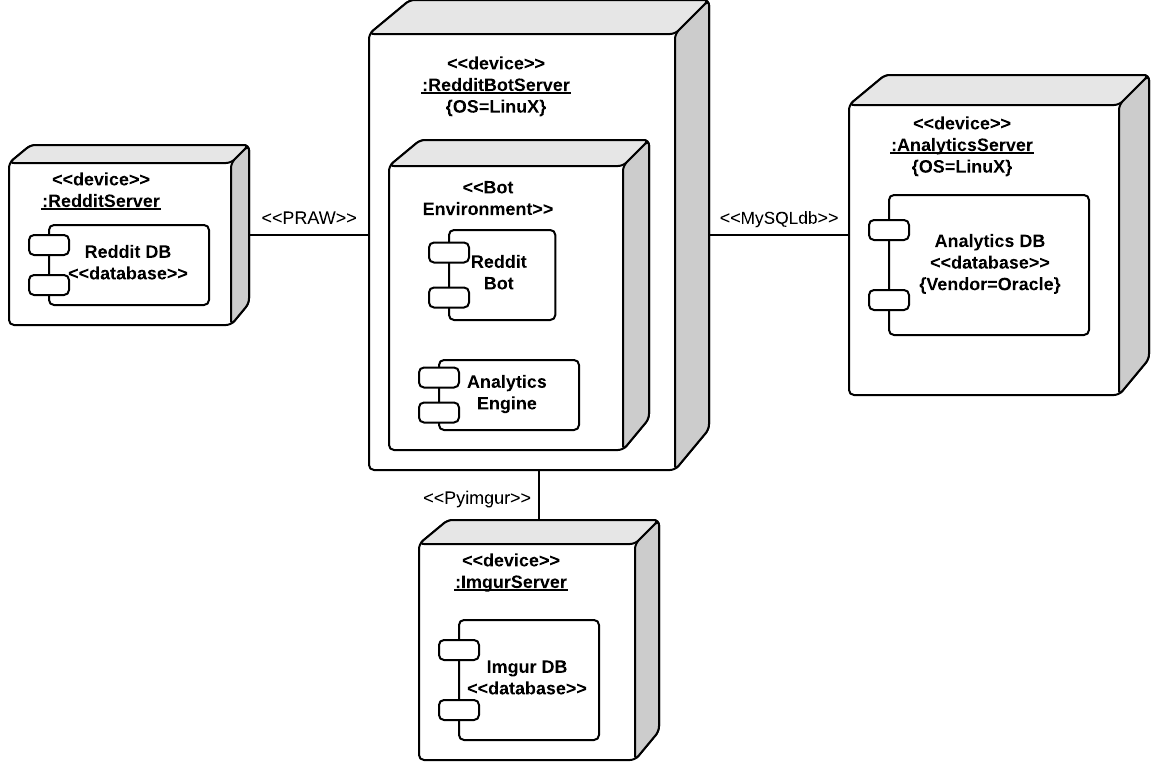
Logical View:



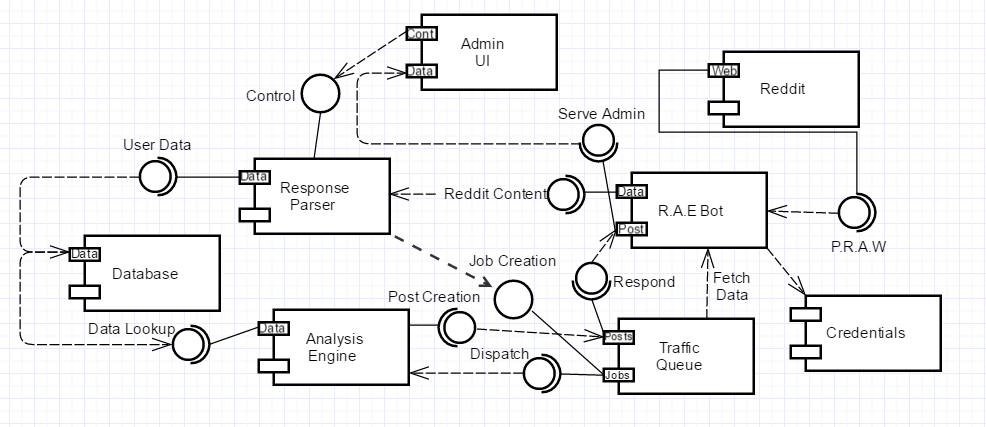
Process View:



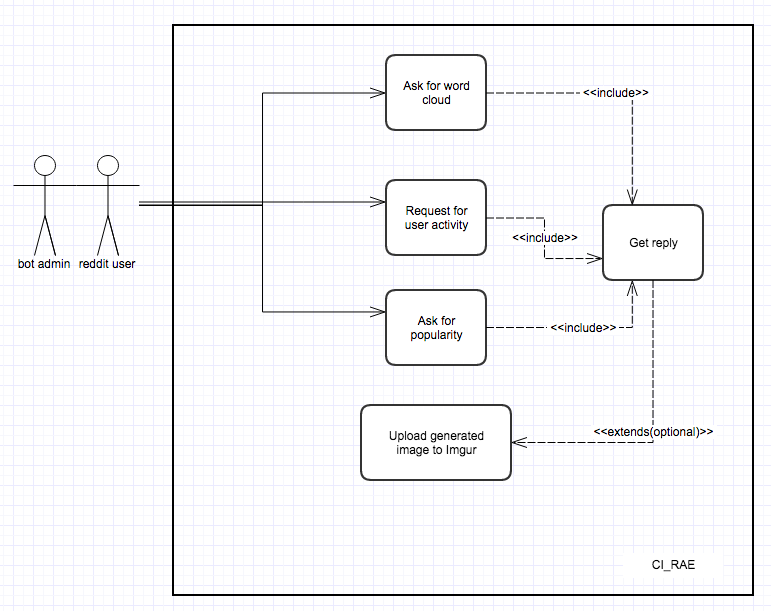
Physical View:



Development View:



Scenario View:

Based on a use-case diagram, a typical scenario would have a user submit a comment and depending on the command, the bot will respond with a comment. Depending on the command, it will upload a generated image to an Imgur account, and then post a comment back to the user (with a url link to the generated image). After looking into using PRAW to access the reddit api, a database is no longer necessary. 

5. Identify the architectural patterns (particularly those that we discussed in the class, like MVC, layers, repository, etc.) that you notice in your project; did you actually use any known patterns for the design of the architecture of your project?

We are using a client-server pattern in our architecture. This is because our software will be entirely run as a service on Reddit. Users will not use client software and will interact with our bot through the Reddit website. All the software we create will be running on our own machine, and since we are dependent on external library, the Reddit API Library (PRAW), we will not need to send code to Reddit to execute. For the core of our program, we are using the pipe and filter pattern. In our pattern, we lookup a reddit user object, which contains all the information about their account, we pipe the information to our analytics filter to transform the data into a graph. The product is a .png graph that we can then send to the IMGUR servers.