

Our first steps were to gather requirements, find possible web services, choose appropriate tools, and build prototypes for how we might solve this problem. Each prototype was designed for a different web service by a different member of the team. This was done to allow everyone to learn about the problem and the tools, and come up with unique solutions that we can look over. Our next step will be to analyze our different solutions to come up with the most simple and efficient method. After that we will start working with an actual database to store the scraping data, and then create an environment with a simple command line utility.

Initial Requirements

- Scrape pricing for various cloud services
 - Virtual machine and instance store services in particular
- Command line utility
- To help match customers with best cloud compute services
- Should be able to track history of pricing

Possible scraping tools

- **Scrapy (Python) - using this to start**
- BeautifulSoup (Python)
- Jaunt (Java)
- JSoup (Java)

Major web services to scrape data

- **Google**
- **Amazon**
- **Microsoft Azure**
- **Alibaba Cloud**
- CenturyLink
- IBM Softlayer

Create prototypes for the major cloud services using Scrapy

- Google - **Josh**
- Amazon - **Kevin**
- Microsoft Azure - **Chris**
- Alibaba - **Evan**