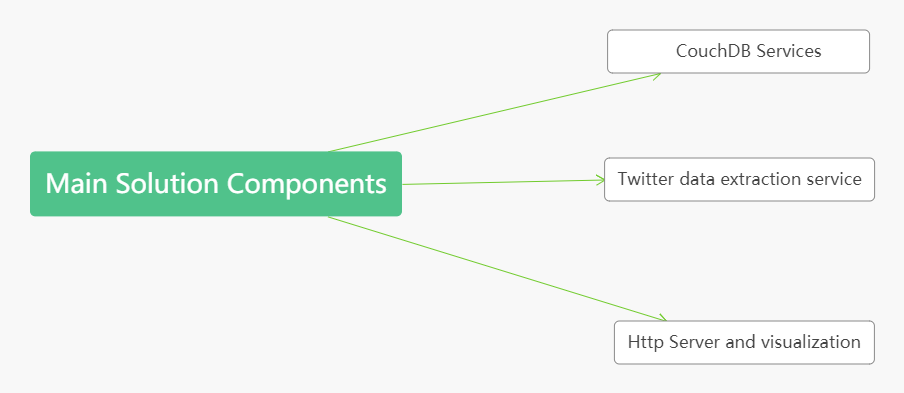
**System Design and Architecture**

In this part, we will mainly introduce some details of the design and architecture of the group, including the specific parts of the system, the various components used and the reasons why we adopted such a structure.

1. **System Architecture**

First of all, our project is based on UNIMELB research cloud, through the tweets obtained from twitter API and the analysis of Aurin data to draw a visual conclusion. Therefore, we divide the whole project task into five parts.



CouchDB services：

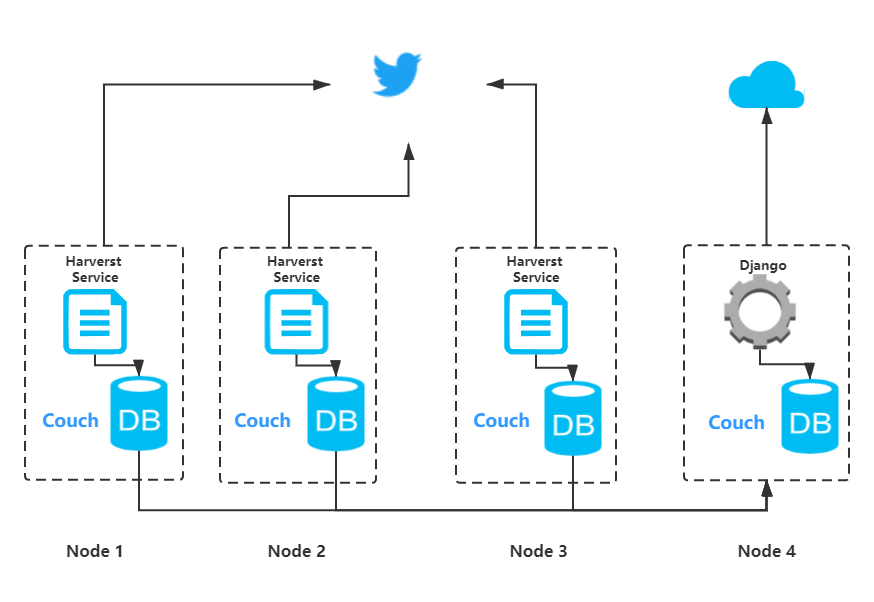
Data processing is done by using CouchDB, a NoSQL database that supports MapReduce. Because it provides an effective way to deal with big data. It can process data in parallel through map() and reduce() functions. And it stores data in JSON format, which means that we can store it directly, instead of splitting it into smaller parts and putting it in different tables.

Twitter data extraction service：

For the acquisition of twitter data, we obtain the data from the official tweetapi through multiple nodes at the same time and store the content containing the data in the local CouchDB, and the ID of the twitter will be set to the ID of the JSON document. Each node uses a different filtering function to ensure that the data repetition rate is reduced. For the duplicate data that may exist in different databases, we will integrate the data in the fourth node, delete the duplicate ID, so as to ensure the accuracy of the data.

Http Server and visualization

A Django framework is used to build a web server, and use some effective external tools such as Google map, apex chart to realize the visualization function, and show the data to the user through an intuitive graphical interface.



Challenges:

1. **System Components**

Ansible: Ansible is a new automatic operation and maintenance tool. Based on Python development, Ansible implements batch system configuration, batch program deployment, batch run command and other functions. In this case, it is used for Automated Deployment and configuration management. It is very important in the cluster configuration of this project. It is used to ensure that the configuration environment of each instance of CouchDB is the same and handle the connection between the instance and the cluster.

Ansible, as an automated unified configuration management tool, has the following advantages:

1. Remote execution and configuration management

Ansible can execute remote commands in batches, remotely operate multiple hosts and configure software services in batches. In addition, for the overall configuration, it can realize automation, unified service configuration management and start / stop

2. Manage public cloud and event driven

Through the Ansible module of API, different event driven services are implemented, such as:

1) Restart after modifying the configuration

2) Only modify the configuration file, not restart.

The second is to manage the public cloud through the API interface, and because the syntax is python, it is convenient for operation and maintenance to carry out secondary development.

In fact, in our use process, we don't feel too many shortcomings of Ansible. After looking up the relevant materials, we know that Ansible's large-scale concurrency capability may be relatively weak, especially when there are more servers, such as more than 200, because it is based on SSH service execution and adopts polling mode

Unimelb Research Cloud

It is a cloud facility supporting the research of the University of Melbourne. It provides cloud services to all students and scholars of the University of Melbourne to help them carry out scientific research and exploration. In our project, the system we developed is hosted in the Unimelb research cloud. And it provides us with powerful computing power support.

Django：

Django: is an open source web application framework written by python, which encourages rapid development and follows MVC design. It uses a method of separating business logic, data and interface display to organize code, and gathers business logic into a component. While improving and customizing interface and user interaction, it does not need to rewrite business logic.

The Django framework has a certain way of defining and executing tasks. It is a logical file structure. However, it is also mandatory that you cannot use your own file structure, because the framework has a way, commonly known as the "Django way" approach. If you cannot follow these rules, you may not be able to deploy anything using Django.

All of Django's features come with a lot of code. This requires server processing and time, which has some problems for low-end websites that can run on very small bandwidth.

Django also has many advantages, first of all, it is implemented by python, and its extensibility is quite good. At the same time, it provides one-stop solutions from template, ORM, session, authentication and so on. Even app division is well done. In a word, do as many things as possible for developers. Finally, the Django community is very large and various plug-ins are complete. In most cases, we can find what we want