1. This system is based on the foundation of Melbourne Research Cloud. Based on the MRC, there are four instances to consist of the system, all of them have 2 cores and 9GB RAM, which can satisfy the requirement of the project.
2. To manage instances of MRC in local PC, we use ansible playbook with openstack API to complete the creation of volumes, security groups, and instances. Next, let us turn to the architecture of the system.
3. There is the introduction chart of the system, each instance has its own tasks. At first, we look at the left of the chart, the demo-1 instance, which deploys the twitter harvest application, to collect data from twitters. And then, turn to the middle of the chart, the rest of the three instances consist of a CouchDB cluster to save data. Finally, we can focus on demo-4,, except as a CouchDB member, a web server is deployed at this instance for visualizing the analysis result
4. Basically, all applications run in instances are as docker containers, as we know, docker can make application separate from infrastructure, which can simplify the development process. Now, let us go into each instance and see the application operation situation. In the demo 1, we can see that, there are three twitter harvest applications. In these applications, we use tweepy as the tool to collect tweets and we have two approaches to do this. The first approach is using tweepy search API to find tweets in the past 7 days. We set a list of keywords as filters, because our topic is related to covid19 so we set out keywords to the terms that are relevant to covid19 such as coronavirus, covid19, epidemic, and such. The other approach is using twitter streamers to collect real-time tweets in Australia as supplementary.
5. Then turn to the demo2 and demo3, they both have a CouchDB container to save data, which we get from the demo-1 instance, which I will show you later. Finally, let us see the demo-4, except as a couchdb member, a nginx server with web pages also is deployed in this instance, for visualizing the data analysis result, and this part will be introduced in detailed in later section.
6. Now, let's talk about the CouchDB cluster, in the twitter harvest, we set the search region to Australia because we are more interested in the situation in Australia and the data could be used as comparisons to the data from AURIN, which is the database covid19\_au\_second. We also collect tweets outside Australia as some comparison dataset, which is covid19\_global. We collect both tweets in English and non-English so as to find connections or relationships that apart from the people speaking English, who else cares about the topic the most, covid19\_au\_with\_other\_language and covid19\_global\_with\_other\_language. Also, the realtime\_tweet is for the real-time tweets in Australis.

6.5 In addition, we also build some views for database, such as fulltext, hastag, location, and place, to accelerate the data reading speed and point out the key information.

1. Finally, let’s talk about our interactive dashboard. It is designed to help you get your insights at-a-glance. We use Apache ECharts, which is a powerful, interactive charting and visualization library for browser, to visualize data with report charts and dashboard components.
2. As you can see, there are five charts on the page. The first chart is to display the number of tweets and confirmed cases by state. You could obtain the exact data by selecting the exact state, estimate and compare the ratio of tweets to COVID-19 confirmed cases of each state, or just observe single class data. Besides, the original data could be shown as a table, the chart could be shown as a pure line chart or bar chart, and all charts could be downloaded by using the toolbox component.
3. The second chart is to display the top 10 hashtags contained in tweets. Cause we intend to know which hashtags have high numbers of occurrences when the topic comes to COVID-19, any hashtag containing “COVID” or ”coronavirus” has been excluded. You would notice several familiar hashtags in Australia, such as #auspol, #7news, and #9news, while you could also see some novel hashtags, such as #herdimmunity, #pandemic, and #kawasaki.
4. The third chart is to display the top 10 languages used in tweets. it can be asserted that English would take the first place in the chart, so we could just focus on other languages. The result shows Spanish tweets account for half all tweets. You could obtain the exact number and the percentage if you select an exact category.
5. The fourth chart is a tag cloud generated with the content of tweets. This format is useful for quickly perceiving the most prominent terms to determine its relative prominence. Bigger the word is, the higher frequency it has in the content of tweets.
6. The data from the fifth chart is from AURIN. It shows the population of each state of Australia in 2017. This chart could be connected with the first chart to show the ratio of confirmed cases to population and the ratio of tweets to population. Further research could be conducted to analyze the relationship between these several variables.
7. To make the web page render well on a variety of devices and window or screen sizes, we use media queries to implement responsive web design, that means you could access to this page via several different kinds of devices, such as your laptop, iPad, and iPhone. In that case, the agility of this dashboard could be improved and best practice could be accessed.
8. That’s all of our video, thank you for watching.
9. Hello, everyone, this is the project of the group 51, a cloud system to search scenario about COVID-19. Now, let’s begin.