# The Installation and Configuration of the Couchdb

## Opening port for external web server access

Couchdb is one kind of NoSQL database systems (DBS) for distributed system as well as abilities of this database are aligned with project’s requirements because of global twittering analysis and distribution-based architecture. Based on this conceptual idea, couchdb can be implemented in this project and installed in a virtual machine by the command line tool. Also, the couchdb can be accessed by HTTP so that the bind\_address is required to modify to 0.0.0.0 for client access. The diagram 1.1 illustrates the installation of the couchdb.

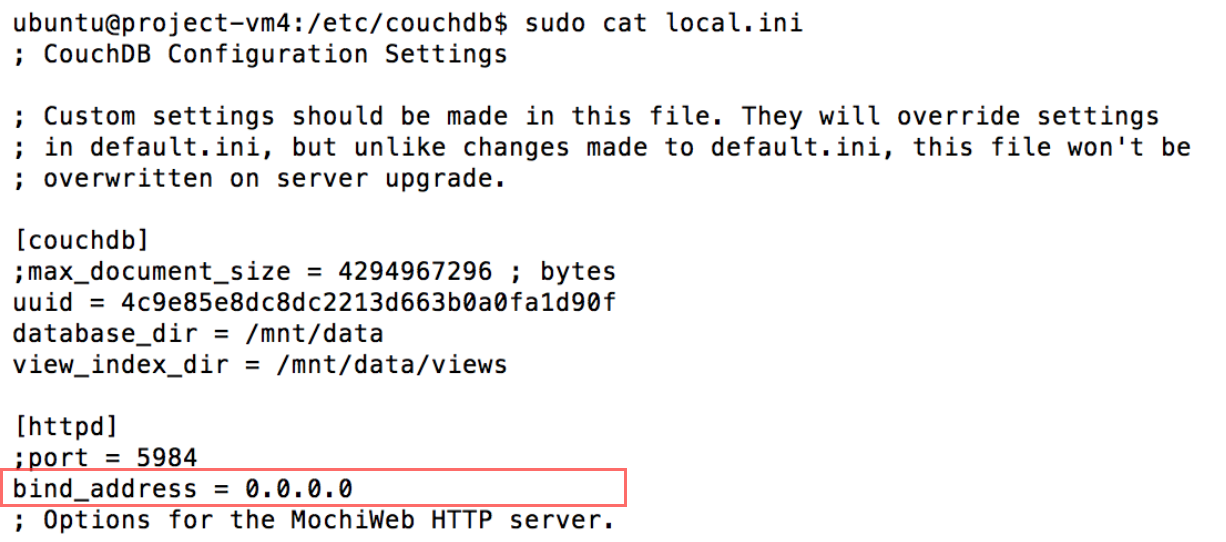


Diagram 1.1: Changing bind\_address for client access.

## Accessing to the Couchdb

Executing “curl –X Get <http://yourdatabaseip:5984/_all_dbs>” to the command line and <http://115.146.95.53:5984/_utils/> to the webpage, as a Futon webpage, are useful approaches, verifying whether the couchdb is working or not. The diagram 1.2 and 1.3 depict the process of installing and verifying couchdb.



Diagram 1.2: A command script to check all of DBS in the couchdb

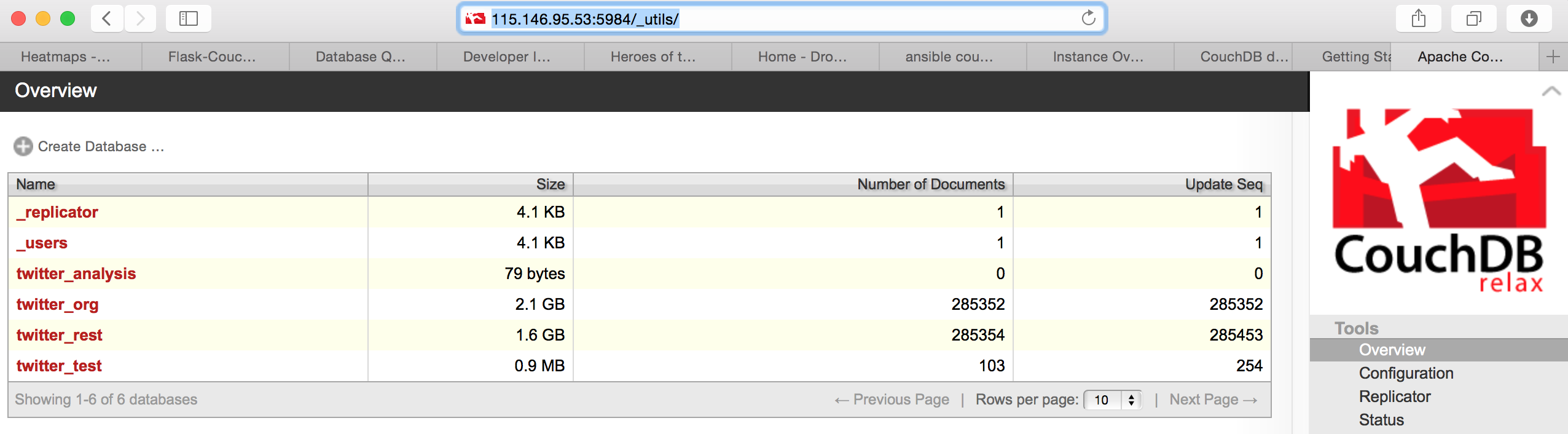


Diagram 1.3: Futon webpage for verifying all of DbS in the couchdb

## The attachment of volumes

The system also attaches a volume to the DBS because it may be increased incrementally due to inserting the tweet data on a daily basis for purposes of data and sentiment analysis. The diagram 1.4 shows a volume has been attached in couchdb partition.

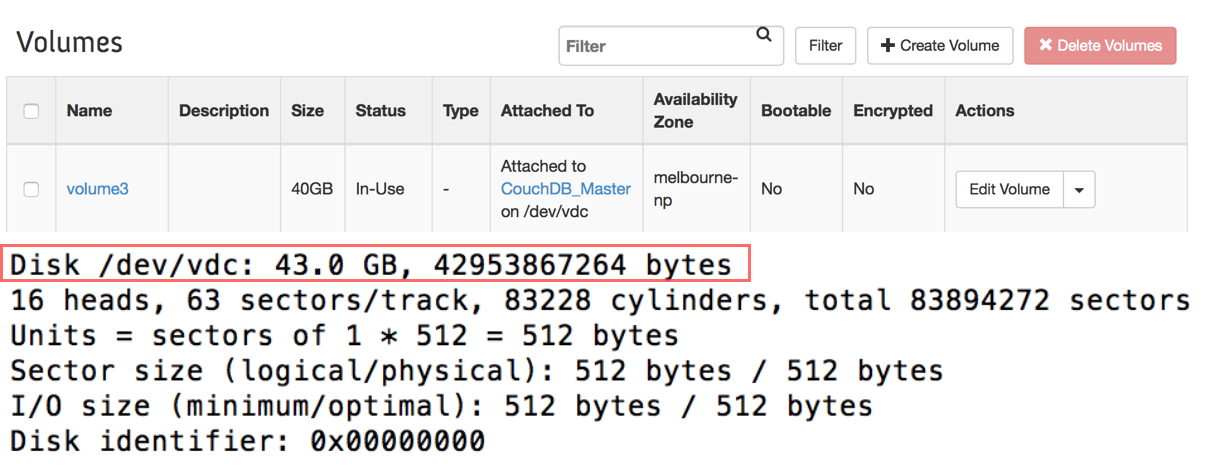


Diagram 1.4: The volume for the DBS partition

## Configurations of the replication

## Replication is essential for data backup as well as avoids accidental disaster. This disaster recovery in Couchdb can be implemented by web configuration as a continuous mode so that this recovery is executed constantly. The diagram 1.5 depicts the status of this recovery.





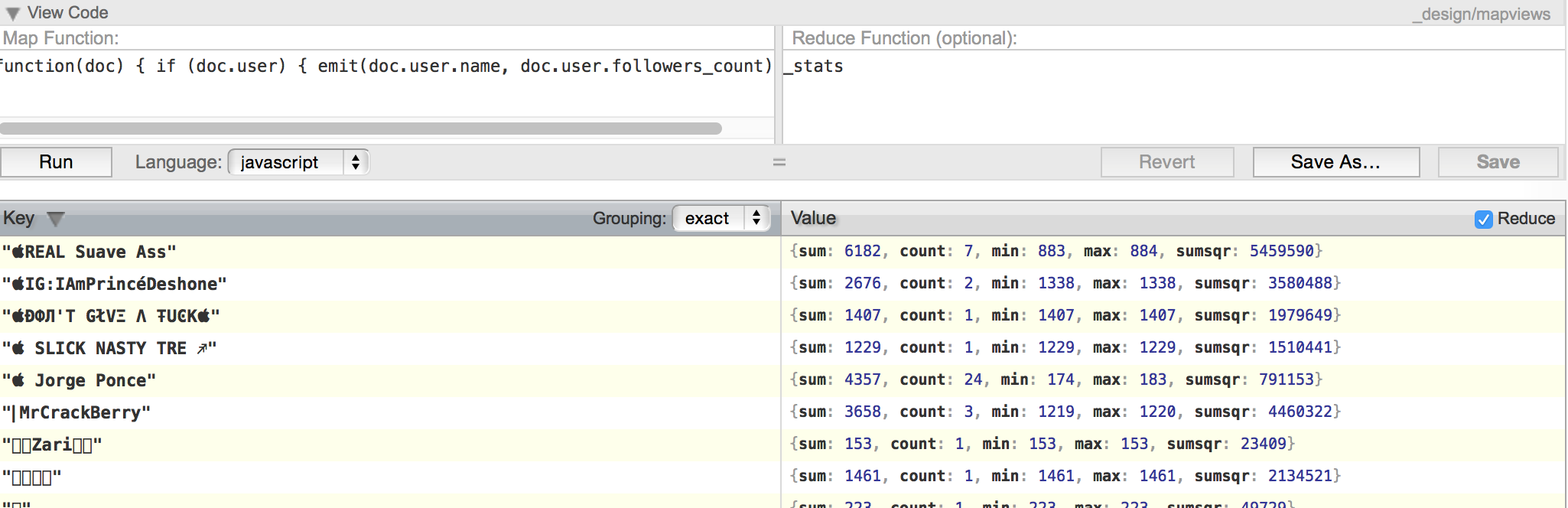
Diagram 1.5: the status of the replication

# Map-reduce view of the Couchdb for data analysis.

Based on the project’s requirements, the tweets are utilized for data and sentiment analysis so that Map-reduce is an approach to aggregate all of data and reduce the amount of data, summarizing the results. In the couchdb, the map-reduce function represents a view that is a high performance approach to acquire the results because of building the B-tree structure. There are several scenarios analyzed in this project: the top 10 most followers and 100 latest tweets.

## The top 10 most followers

By implementations of the map-reduce function, the tweeter user name and follower\_counts can be converged though map function and get the status of these data by reduce function; however, the couchdb only supports sorting by key so the list function is required for implementation as a filter for sorting by value and obtaining top 10 most followers. The diagram 2.1 depicts the results of most followers.

Diagram 2.1: the results of most followers.

## The latest 100 tweets

The vital point for this scenario is that the timeframe is sorted by key and geography and text are stored as an associative array, providing information for graphic results. The diagram 2.2 depicts the results of 100 latest tweets.

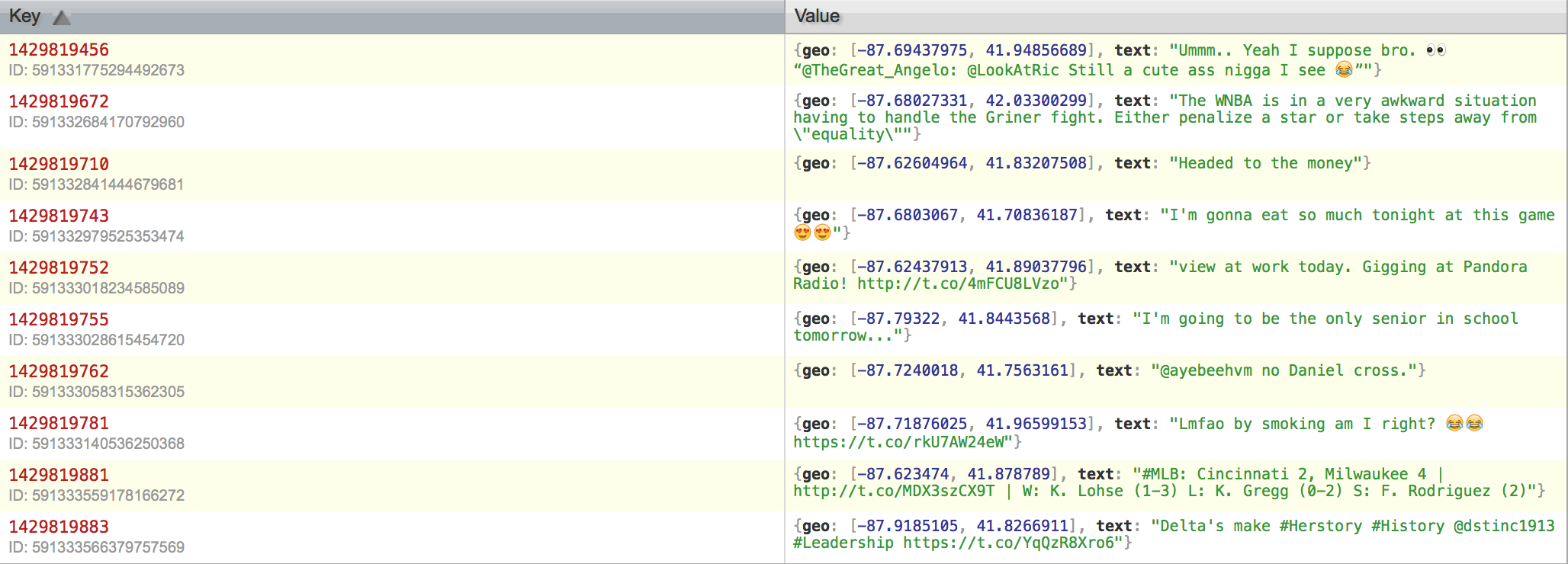


Diagram 2.2: the results of the latest 100 tweets.