

## TP2 Twitter Analytics

**Goal :** Try to find trends into tweets about GAFA and NATU. We add twitter, Microsoft, and IBM in the study.

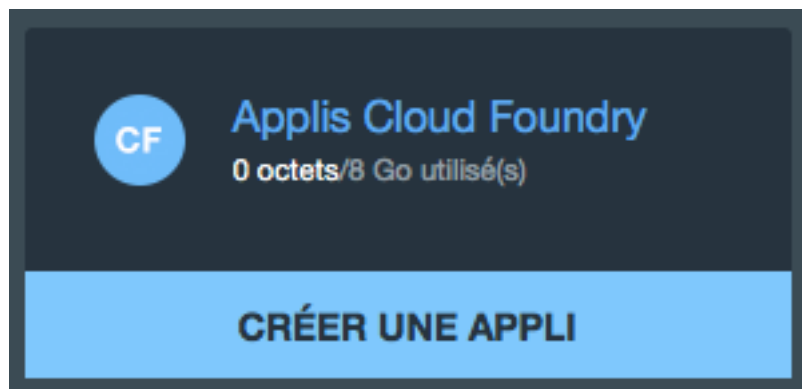
**Solution :** Make an application with node JS, nodeRED, cloudant, IBM BluDB, DashDB and R

### Requires :

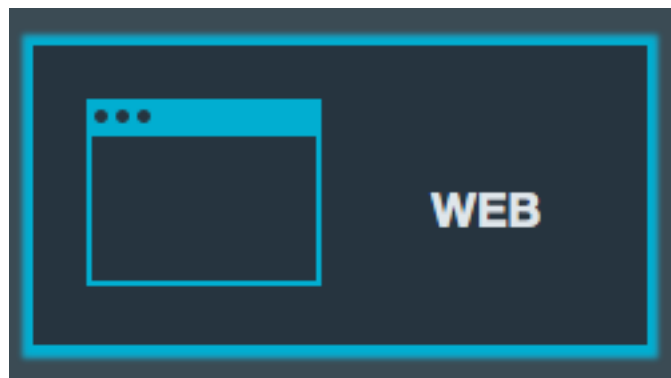
- 10 minutes to build an application to produce dataViz solution ! with R !
- Non stop twitter public stream stored into a NoSQL DB (couchDB).
- No infrastructure to deploy !
- No code to produce !
- That's bluemix solution !

### Let's start !

#### 1. Create a web application :



Web :



Select boilerplates :



Select NodeRED starter, boilerplates :

Be Careful : select an appropriate workspace. Think that workspaces hosted in US are more compliant with more possibilities

Créer une application :

---

**Espace :**

US ▼

**Nom :**

TwitterAnalyticsTP2

**Hôte :**

TwitterAnalyticsTP2

**Domaine :**

mybluemix.net

Plan sélectionné :

---

SDK for Node.js™

Par défaut ▼

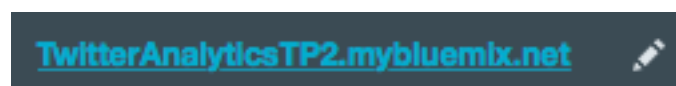
Cloudant NoSQL DB

Partagé ▼

**CRÉER**

## 2. Launch the app

Open the app, and go to nodeREd flow editor :



Go to your Node-RED flow editor

Select twitter social function.



Set credentials :

### Edit twitter in node

Twitter ID

Add new twitter-credentials...

Search

all public tweets

for

comma-separated words, @ids, #tags

Name

Name

### Add new twitter-credentials config node

Click here to authenticate with Twitter.

Add

Cancel

Set other values :

## Edit twitter in node

Twitter ID

@sgagneur

Search

all public tweets

for

cloud

Name

Twitter Public Stream

Tip: Use commas without spaces between multiple search terms. Comma = OR, Space = AND.  
The Twitter API WILL NOT deliver 100% of all tweets.  
Tweets of who you follow will include their retweets and favourites.

Add debug output :



3. Deploy service :



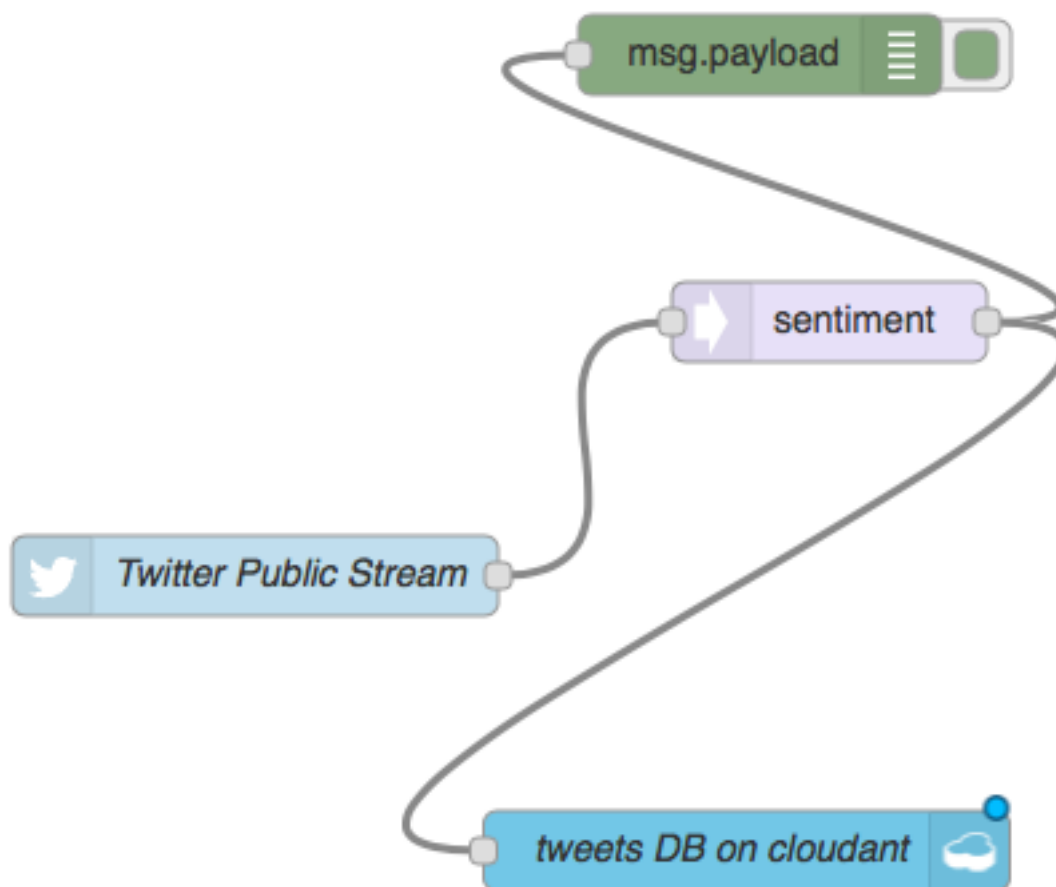
and you can see, in debug window, twitter public tweets :

2 novembre 2015 21:22:49 UTC+1 60edf6da.9f1208  
 weets/PdubPaula : msg.payload : string [132]  
 RT @Tricerat: Will your business benefit from the cloud? Absolutely. Here's how:  
<https://t.co/draD8zAilK> #datasecurity #disasterplan

2 novembre 2015 21:22:49 UTC+1 60edf6da.9f1208  
 weets/drinkwaterpaul : msg.payload : string [139]  
 Setting up Google Cloud Print is a Kafkaesque nightmare. @google documentation points to @brother\_UK, @brother\_UK points me back to @google

2 novembre 2015 21:22:50 UTC+1 60edf6da.9f1208  
 weets/HireaHire : msg.payload : string [83]  
 #Job for: Senior Cloud Engineer in Santa Clara, California. <https://t.co/vmpMucbRBQ>


Add cloudant (couchDB) service to store tweets in « tweets DB » :




**Edit cloudant out node**

Service

TwitterAnalyticsTP2-clou

 Database


tweets

 Operation

insert

☐

Only store msg.payload object?

 Name

tweets DB on cloudant

Ok

Cancel

#### 4. Add a noSQL DB

Create « tweets DB » in cloudant service :

Launch cloudant service from dashboard :



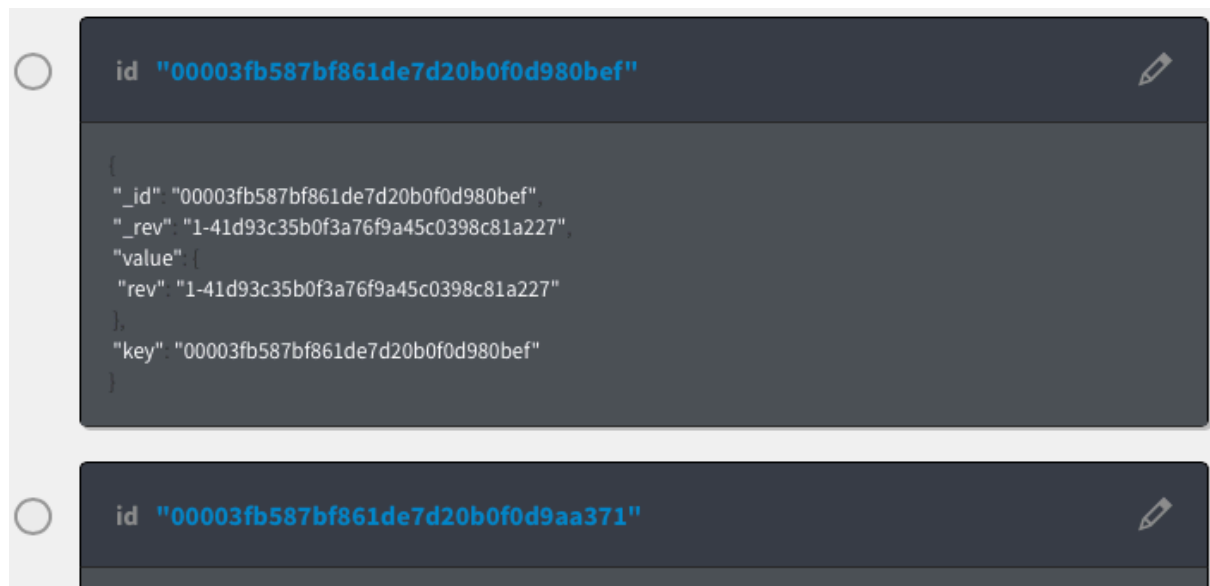
Add a new database :





Deploy nodeRED application every time you make a modification !

Now, you can see public tweets in cloudant DB :



Select one JSON record and you can see a tweet struct :

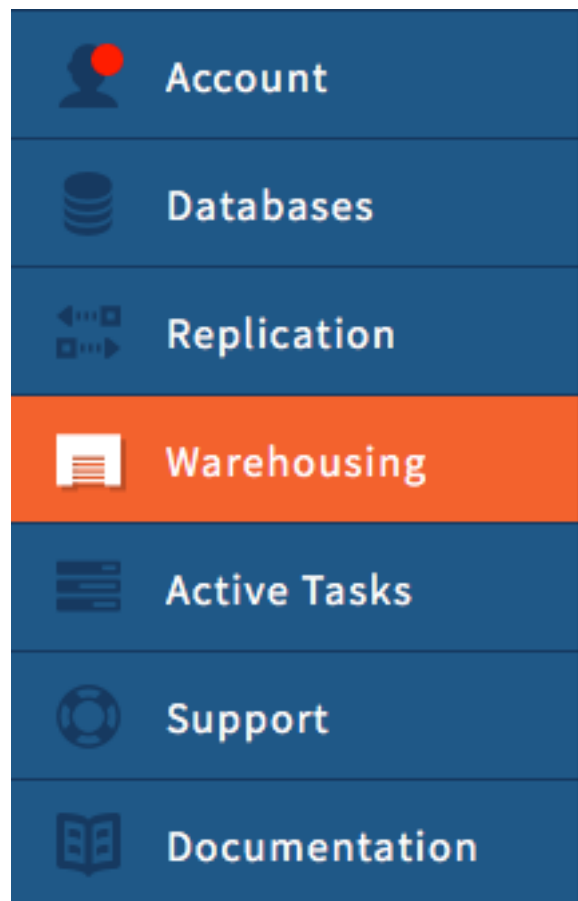
```
{
  "_id": "00003fb587bf861de7d20b0f0d980bef",
  "_rev": "1-41d93c35b0f3a76f9a45c0398c81a227",
  "topic": "tweets/funkd00bie",
  "payload": "@pishaSshisha да да,търпение нямам вече..и на 'The Little Girl Who Swallowed a Cloud as big as The Eiffel Tower'",
  "lang": "und",
  "tweet": {
    "created_at": "Mon Nov 02 20:28:58 +0000 2015",
    "id": 661278850438053900,
    "id_str": "661278850438053888",
    "text": "@pishaSshisha да да,търпение нямам вече..и на 'The Little Girl Who Swallowed a Cloud as big as The Eiffel Tower'",
    "source": "<a href=\"http://twitter.com/download/iphone\" rel=\"nofollow\">Twitter for iPhone</a>",
  }
}
```

## 5. Manage a data warehouse

Create now a data Warehouse

Select warehousing :





Create a Warehouse

Fill credentials, with your Bluemix account :

Create a warehouse :

Choose « tweets cloudant DB »

**Enter a name for your warehouse**

twitterWarehouse

**Add databases to your warehouse**

Type database name

tweets 9.1 MB X

**Optional: add the warehouse to an existing dashDB service instance or in a specific IBM Bluemix organization and space**

☒ Create new dashDB instance

☐ [no dashDB services available]

☐ Bluemix organization

**Create Warehouse**

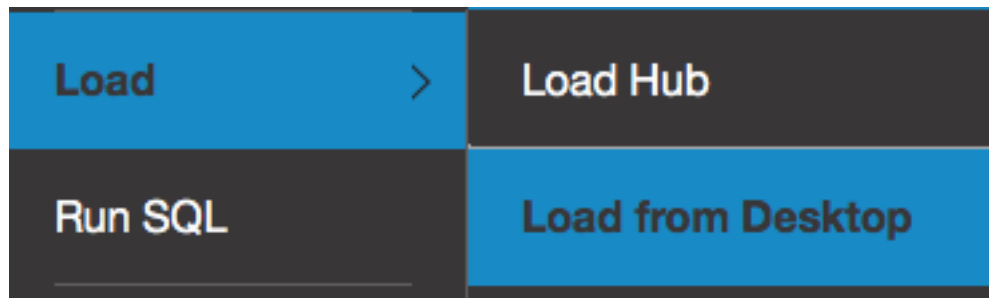
view builded warehouse :

WAREHOUSE NAME	SOURCES	VIEW IN DASHDB	ACTIONS
twitterWarehouse	1	<button>VIEW</button>	<button>+</button> <button>X</button>

Your are now in IBM dashDB :

We try to create a graphic visualization about top US companies correlate with public tweets !

Add your list of GAFA and NATU to warehouse



Supported file types: XLS, XLSX, CSV

File name:

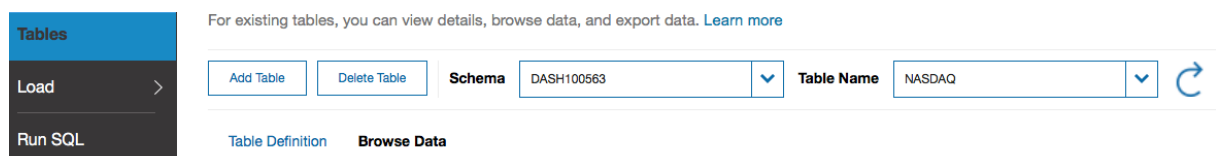
Does row one contain the column names? ☒ Yes ☐ No



- ☐ Load into an existing table  
☒ Create a new table and load

Type new table name and finish

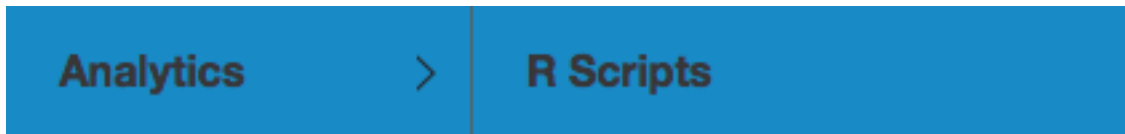
If an error occurs, go to tables and verify if tables are complete :



Errors don't affect file import, data are in tables, and you can use them !

## 6. Play with R

Goto Analytics and play with R :



Select tables and columns to play with :

Schema	Table Name	Columns
<i>Your schema</i>	<i>Quick Filter</i>	
DASH100563 >	Showing all items	
<i>Sample schemas</i>	NASDAQ	
DB2GSE >	NASDAQ2	
ERRORSCHEMA	TWEETS	
GOSALES	TWEETS_OVERFLOW	
GOSALESDW	TWEETS_SENTIMENT_N EGATIVE	
GOSALESHR	TWEETS_SENTIMENT_P OSITIVE	
GOSALESMR		

You need this code to finish your project. Copy and paste it, into code window.  
Sorry, but we won't learn to code with R language. But it seems important to know it, if you want to work as a data scientist.

#Code is here :  
#DASH100563 is your dashDB account, must be changed !  
#Tweets is cloudbant DB  
#Nasdaq is a dashDB tables build with (CSV or XLS) file import.  
#Columns are Payload (debug), Sentiment\_score and companies (CSV or XLS file)

```
library(ibmdbR)
mycon <- idaConnect("BLUDB", "", "")
idalnit(mycon)
```

```
df1437654379199t <- as.data.frame(ida.data.frame("DASH100563"."TWEETS"))[
,c('PAYLOAD')]
```

```
df1437654379199sc <- as.data.frame(ida.data.frame("DASH100563"."TWEETS")[
,c( 'SENTIMENT_SCORE')])
df1437658789453 <- as.data.frame(ida.data.frame("DASH100563"."NASDAQ")[
,c('COMPANIES')])
```

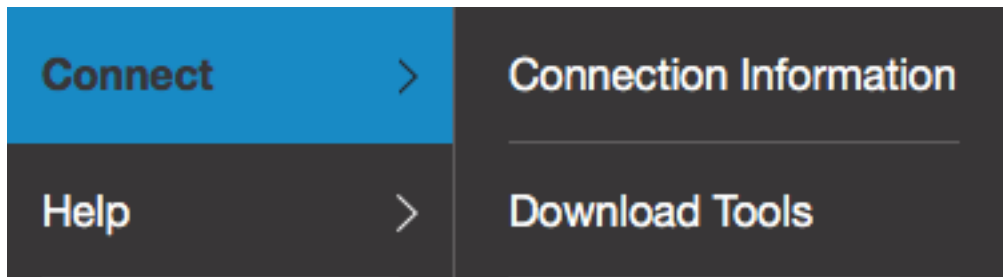
```
companiesPerTweet = apply(df1437654379199t,1,function(tweet) {
  tweet = tolower(tweet)
  mask = apply(df1437658789453,1,function(company) {
    company=tolower(company)
    if (grepl(company,tweet)) {
      company
    } else {
      NA
    }
  })
  #potentialCompanies = unique(mask)
  potentialCompanies = mask
  potentialCompaniesWithoutNone =
potentialCompanies[!is.na(potentialCompanies)]
  if (length(potentialCompaniesWithoutNone)==1) {
    potentialCompaniesWithoutNone
  } else {
    NA
  }
})
```

```
results = cbind(df1437654379199sc,companiesPerTweet)
```

```
myMean = function(valueList) {
  numericList=as.numeric(valueList)
  mean(numericList)
}
```

```
aggdata <-
aggregate(results$SENTIMENT_SCORE,by=list(results$companiesPerTweet),
FUN=myMean)
plot(aggdata)
plot(table(companiesPerTweet))
```

You can also use RStudio, an IDE to code with R.  
Go to connect informations panel :



Copy and paste user ID and password :

Host name:  
 Port number:  
 Database name:  
 User ID:  
 Password:  
 Version:

You can now work without install R Studio (on your computer) for your project :

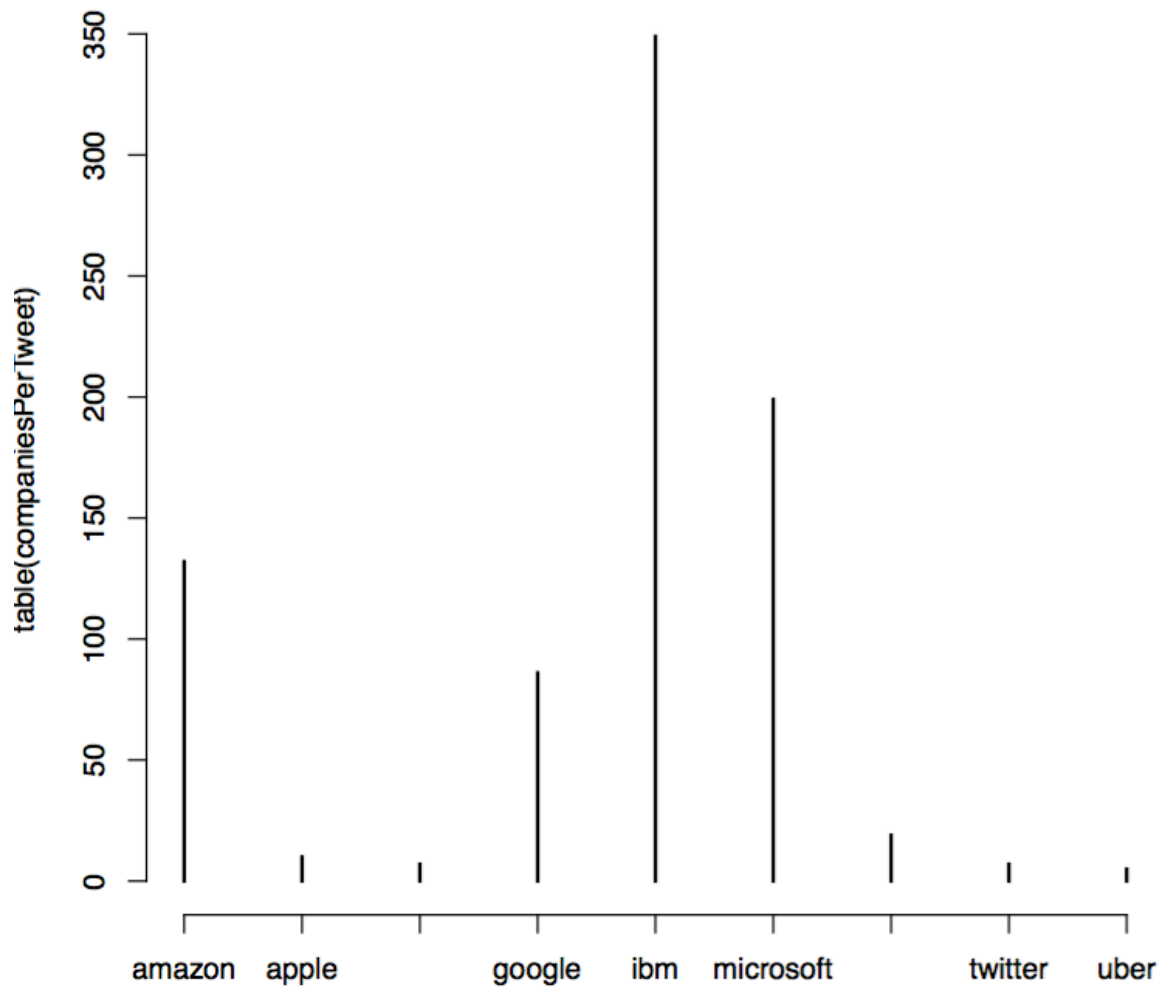
```

1 library(ibmdbR)
2 mycon <- idaConnect("BLUDB", "", "")
3 idaInit(mycon)
4
5
6 df1437654379199t <- as.data.frame(ida.data.frame("DASH100563"."TWEETS")[,c('PAYLOAD')]
7 df1437654379199sc <- as.data.frame(ida.data.frame("DASH100563"."TWEETS")[,c('SENTIM
8 df1437658789453 <- as.data.frame(ida.data.frame("DASH100563"."NASDAQ")[,c('COMPANIES
9
10
11 companiesPerTweet = apply(df1437654379199t,1,function(tweet) {
12   tweet = tolower(tweet)
13   mask = apply(df1437658789453,1,function(company) {
14     company=tolower(company)
15     if (grepl(company,tweet)) {
16       company
17     } else {
11:37 | (Top Level) | R Script
  
```

## 7. Run the project

Run your project, into your favorite workspace (IDE or Analytics R script)

You can enjoy your job in a wonderful PDF document :



Ok, we don't have colors, but you can add them !