Overview

Text Analysis in SAP HANA integrate with Twitter

In this lab, we are going to use the Twitter API to get the tweets, save the tweets into SAP HANA system using JDBC connection and run the Text Analysis on top of the tweets. After the lab, you will be able to learn:

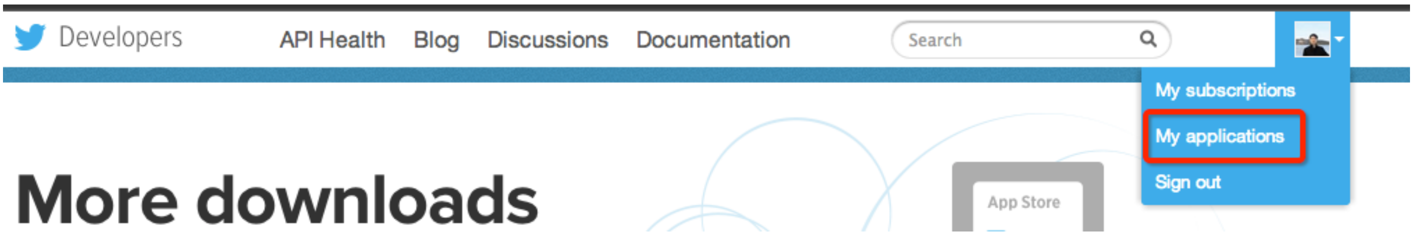
- SAP HANA integrates with Twitter
- Program with SAP HANA using JDBC in Java language
- SAP HANA Text Analysis

Prerequisites

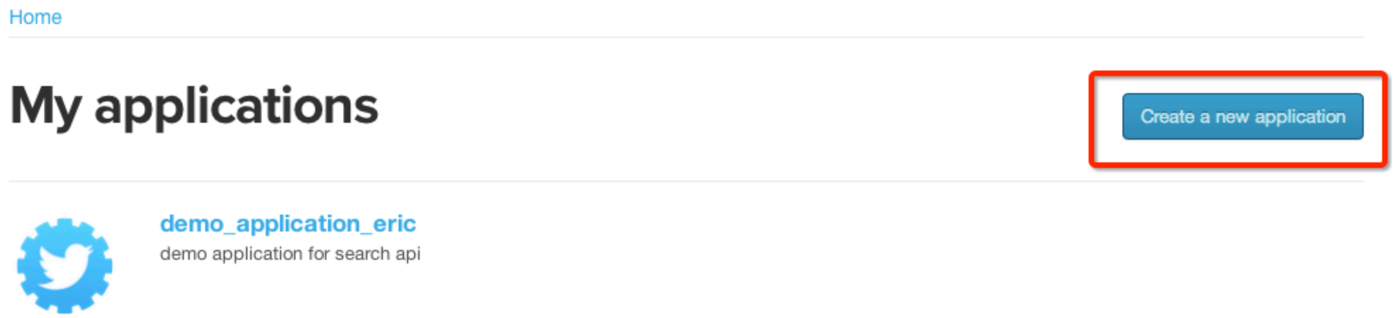
Register an Application at Twitter Developers

As we are going to use the Twitter API to extract the data from Twitter, it is required to create an application at Twitter Developer and we will need the authentication information of the application and use them to invoke the APIs later.

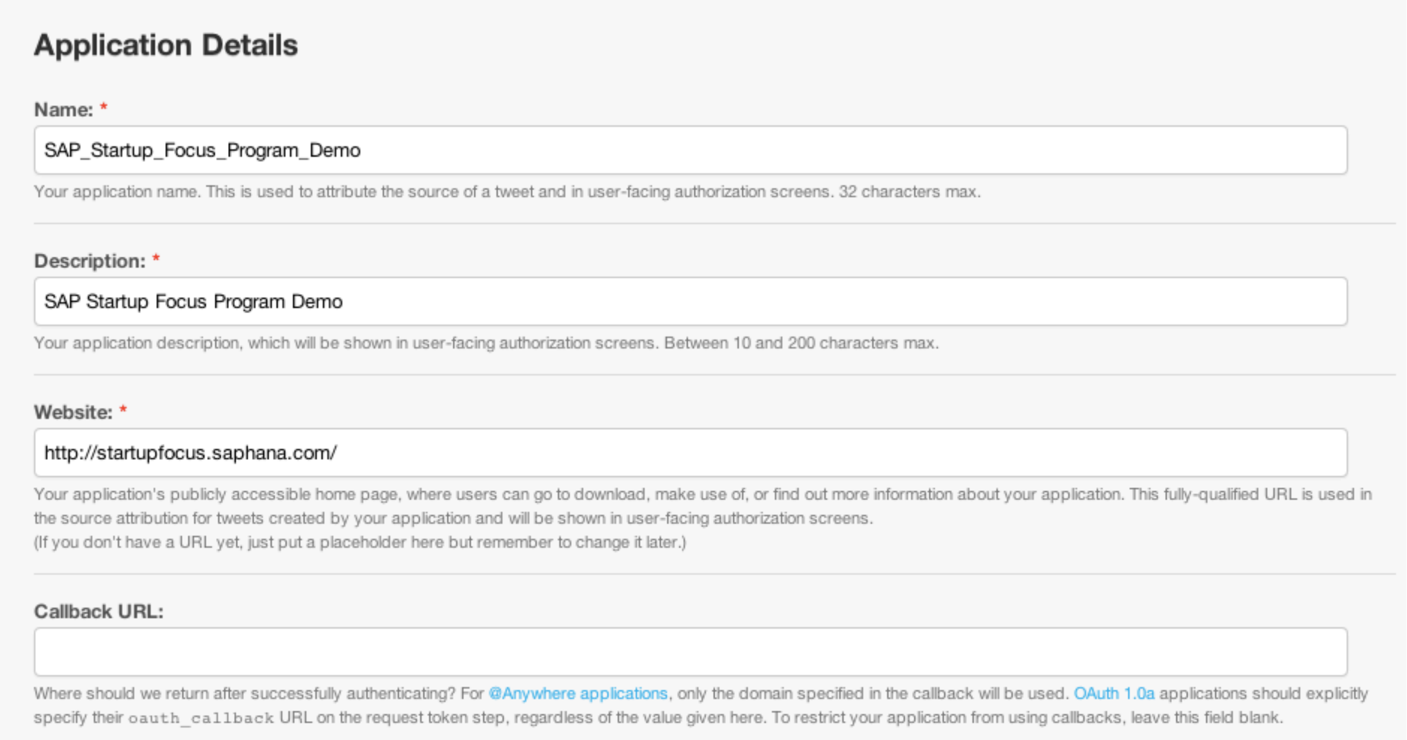
In case you haven't use Twitter before, you need to create your twitter account firstly. You can register an application and create your oAuth Tokens at https://dev.twitter.com/. Logon with your twitter account, click your profile picture and click the "My applications".



Click the button "Create a new application".



Follow the form instructions to complete the registration. You need to input the application name, description, your websites and leave the call back URL as blank. Accept the developer rules and click the button "Create your Twitter application".



Application Details

Name: *

SAP_Startup_Focus_Program_Demo

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description: *

SAP Startup Focus Program Demo

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website: *

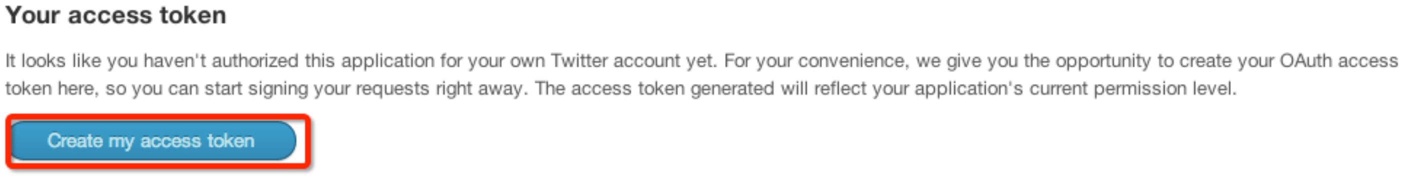
http://startupfocus.saphana.com/

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens.
(If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL:

Where should we return after successfully authenticating? For @Anywhere applications, only the domain specified in the callback will be used. OAuth 1.0a applications should explicitly specify their `oauth_callback` URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

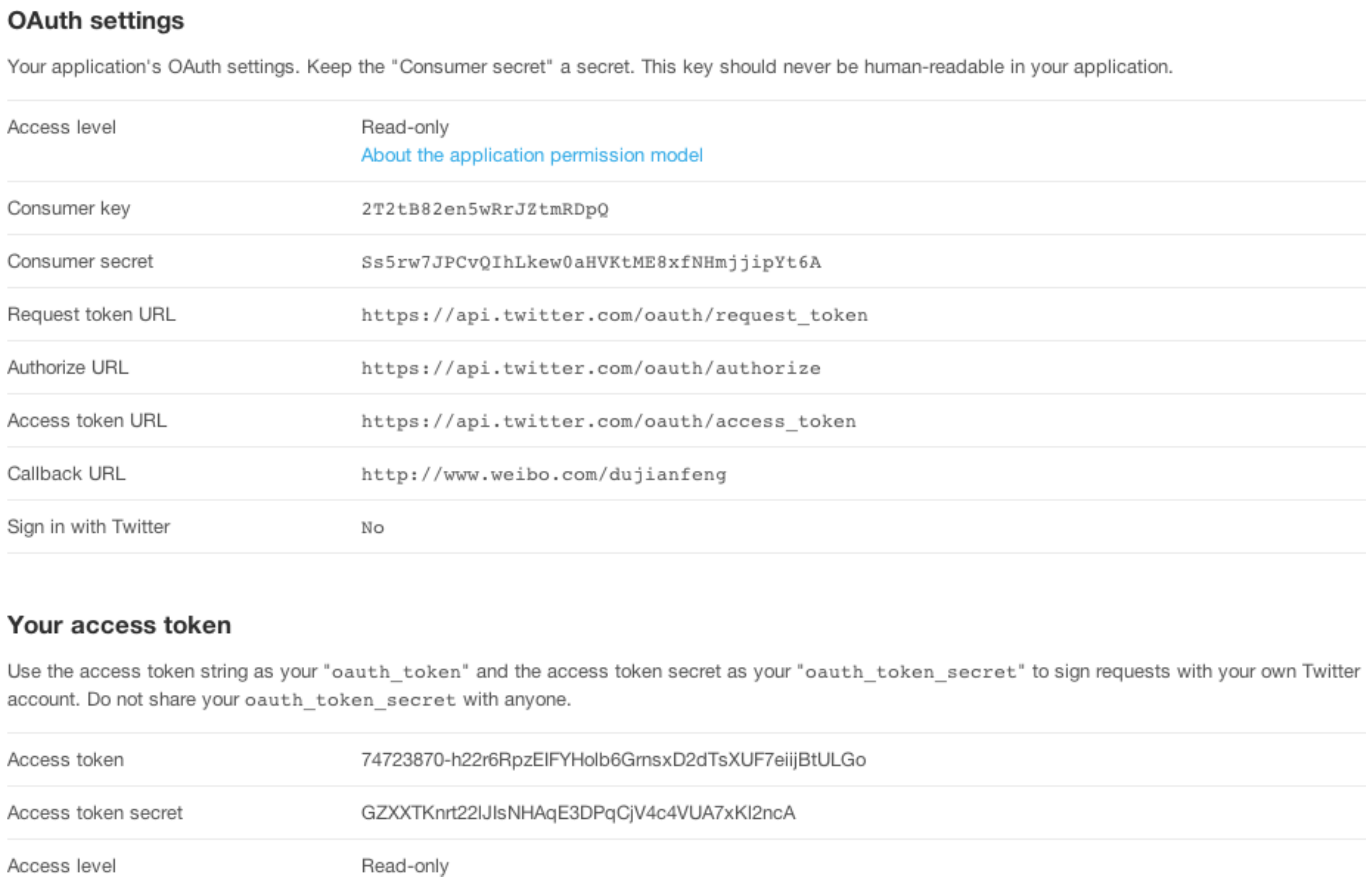
Scroll down the screen and you will see the button "Create my access token", click it to generate the token.



Your access token

It looks like you haven't authorized this application for your own Twitter account yet. For your convenience, we give you the opportunity to create your OAuth access token here, so you can start signing your requests right away. The access token generated will reflect your application's current permission level.

After that, you will be able to see the oAuth settings like below, save the values of Consumer Key, Consumer secret, Access token and Access token secret. We need to use them later in the APIs.



OAuth settings

Your application's OAuth settings. Keep the "Consumer secret" a secret. This key should never be human-readable in your application.

Access level	Read-only About the application permission model
Consumer key	`2T2tB82en5wRrJZtmRDpQ`
Consumer secret	`Ss5rw7JPCvQIhLkew0aHVKtME8xfNHmjjipYt6A`
Request token URL	`https://api.twitter.com/oauth/request_token`
Authorize URL	`https://api.twitter.com/oauth/authorize`
Access token URL	`https://api.twitter.com/oauth/access_token`
Callback URL	`http://www.weibo.com/dujianfeng`
Sign in with Twitter	`No`

Your access token

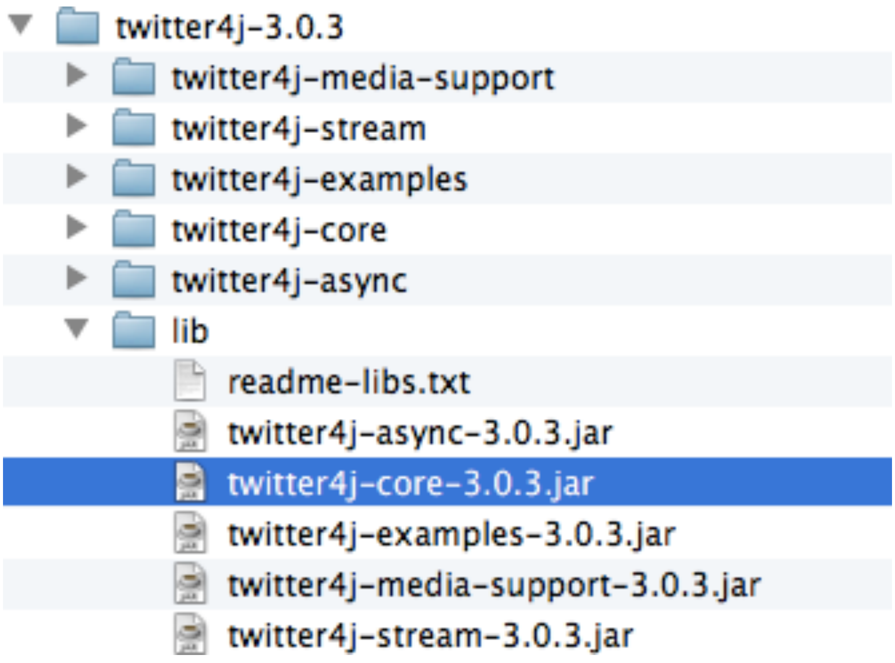
Use the access token string as your "`oauth_token`" and the access token secret as your "`oauth_token_secret`" to sign requests with your own Twitter account. Do not share your `oauth_token_secret` with anyone.

Access token	74723870-h22r6RpzEIFYHolb6GrnsxD2dTsXUF7eiijBtULGo
Access token secret	GZXXTKnrt22lJIsNHAqE3DPqCjV4c4VUA7xKI2ncA
Access level	Read-only

Download Twitter API Java library – Twitter4J

Twitter4J is an unofficial open source Java library for the Twitter API. With Twitter4J, you can easily integrate your Java application with the Twitter services. The link to download it is http://twitter4j.org/en/index.html.

Extracting the downloaded zip file, go the sub folder **lib** and you will see the file **twitter4j-core-3.0.3.jar**, which is the library we need in the Java project and it must be added as the library or class path in the java runtime.



- twitter4j-3.0.3
  - twitter4j-media-support
  - twitter4j-stream
  - twitter4j-examples
  - twitter4j-core
  - twitter4j-async
  - lib
    - readme-libs.txt
    - twitter4j-async-3.0.3.jar
    - twitter4j-core-3.0.3.jar
    - twitter4j-examples-3.0.3.jar
    - twitter4j-media-support-3.0.3.jar
    - twitter4j-stream-3.0.3.jar

There are some useful examples and you can simply check them to help yourselves getting familiar with the Twitter APIs.

Prepare the HANA jdbc library

In order to access SAP HANA from java, we will need the jdbc library, which you can find it at **C:\Program Files\SAP\hdbclient\ngdbc.jar** in windows and **/usr/sap/hdbclient/ngdbc.jar** in Linux by the default installation.

Note: In this exercise, the twitter4j and jdbc libraries are already included in the project you had received from your instructor thus you do not need to download them by yourselves.

Download Eclipse IDE for Java Developers

In this exercise, we will use Eclipse IDE for Java Developers to run the Java Project you already have from your instructor. You can add the Plugins in your HANA Studio or directly download the new IDE at http://www.eclipse.org/downloads/packages/eclipse-ide-java-developers/junosr1

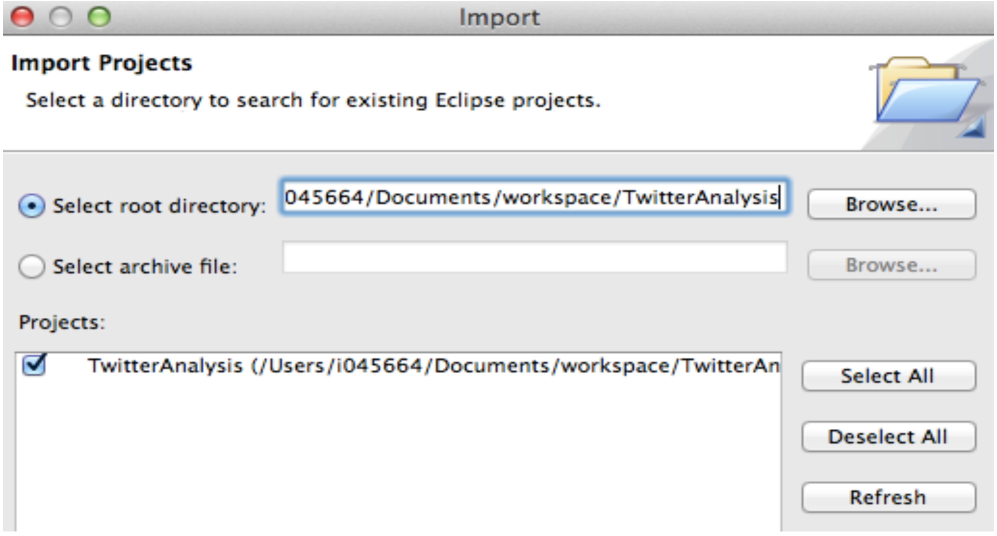
Exercise

Now it is ready to go, in the end of the exercise, we will understand the source code of the project and know how to connect HANA from java, how to use the twitter services in java and the most impressive thing is how simple it is to run the text analysis in HANA, which combines the unstructured data from various sources like twitter, documents with the structured data in RDBMS.

Import the Java Project in Eclipse

To save your time, we will ask you to import the existing java project instead of starting from scratch. Do not worry and we will explain all the components of the project in details below. Extracting the project file **TwitterAnalysis.zip** at your local folder, the file should have been sent to you before the session. Open your HANA Studio and follow the steps below:

1. In the **File** menu, choose **Import**....
2. Select the import source **General** > **Existing Projects into Workspace** and choose **Next**. You should have created the workspace in the XS exercise. Otherwise, you may need to have your workspace created first.
3. Select the root directory where your project files located, selects the project **TwitterAnalysis** and click **Finish** to complete the import.



Import

Import Projects

Select a directory to search for existing Eclipse projects.

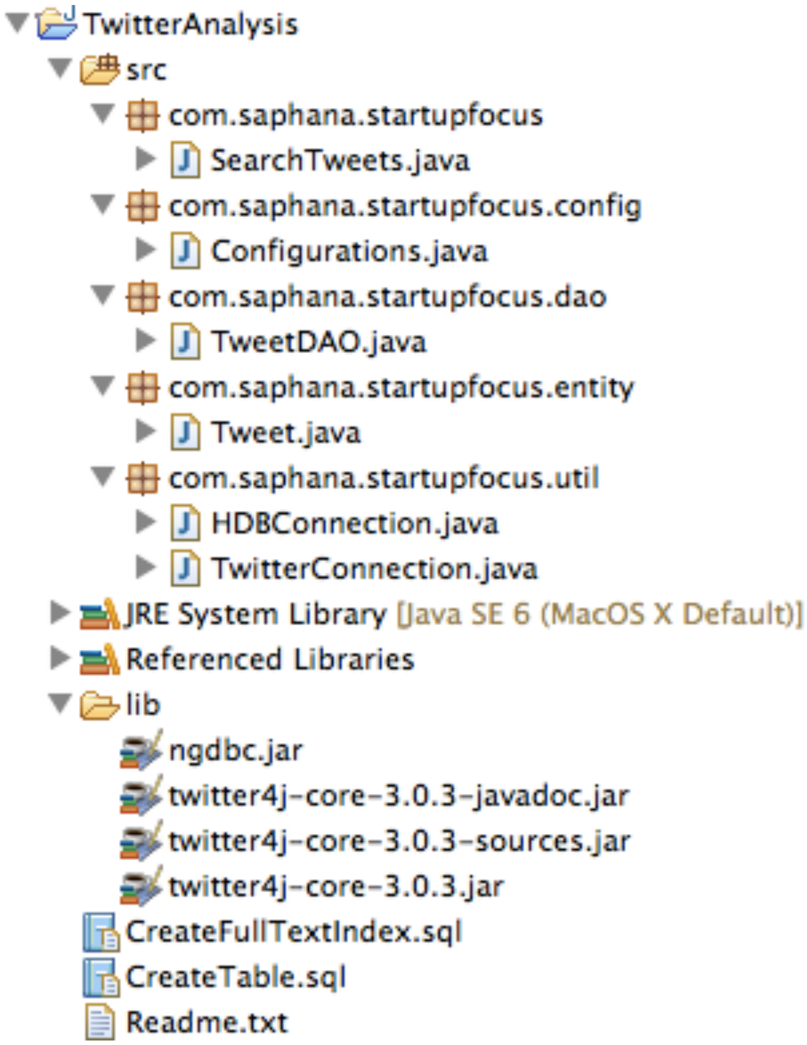
● Select root directory: 045664/Documents/workspace/TwitterAnalysis

○ Select archive file:

Projects:

☑ TwitterAnalysis (/Users/i045664/Documents/workspace/TwitterAn

4. Now you will be able to see the project with the structures like this:



```
▼ TwitterAnalysis
  ▼ src
    ▼ com.saphana.startupfocus
      ▶ SearchTweets.java
    ▼ com.saphana.startupfocus.config
      ▶ Configurations.java
    ▼ com.saphana.startupfocus.dao
      ▶ TweetDAO.java
    ▼ com.saphana.startupfocus.entity
      ▶ Tweet.java
    ▼ com.saphana.startupfocus.util
      ▶ HDBConnection.java
      ▶ TwitterConnection.java
  ▶ JRE System Library [Java SE 6 (MacOS X Default)]
  ▶ Referenced Libraries
  ▼ lib
      ngdbc.jar
      twitter4j-core-3.0.3-javadoc.jar
      twitter4j-core-3.0.3-sources.jar
      twitter4j-core-3.0.3.jar
    CreateFullTextIndex.sql
    CreateTable.sql
    Readme.txt
```

Understand the Java Project

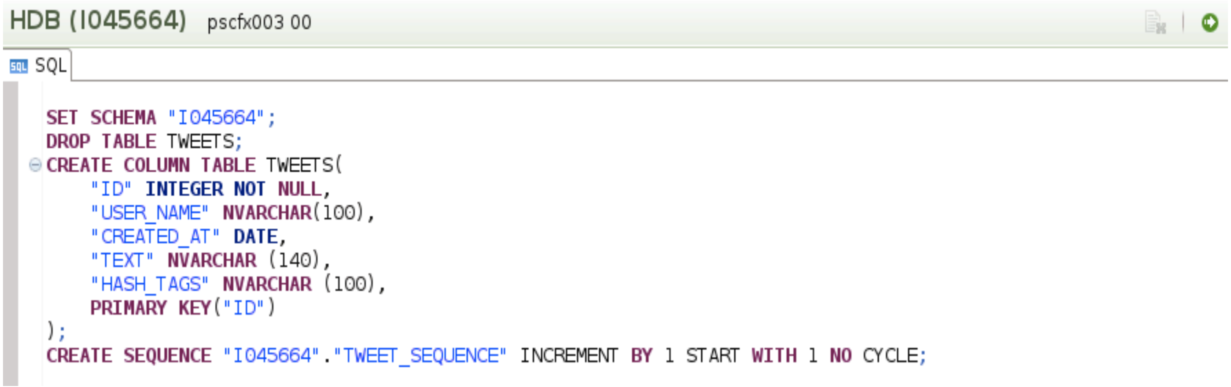
The following table lists the major files in the project and we will explain them in details later in the exercise.

Directory	File	Description
src	TwitterConnection.java	Build the connection to twitter services
	HDBConnection.java	Build the jdbc connection to HANA
	Configurations.java	The public interface for the network, twitter authentication configurations, override it by your own account or settings
	Tweet.java	The java bean class for the tweet objects
	TweetDAO.java	The data access object
lib	ngdbc.jar	SAP HANA jdbc library
	twitter4j-core-3.0.3.jar	Twitter4j library for twitter services in java
	CreateTable.sql	The SQL statement to create the column table in HANA
	CreateFullTextIndex.sql	The SQL statement to create the fulltext index for text analysis
	Readme.txt	The file describes the steps to execute the project

Create a column table in HANA

Firstly, we need to create a table in HANA, where we want to store the tweets we fetched from the twitter services.

1. Open HANA Studio, copy the SQL statement from the **CreateTable.sql** and execute it in the **SQL Console**. **You need to replace the current schema with your own schema**.



```
SET SCHEMA "I045664";
DROP TABLE TWEETS;
CREATE COLUMN TABLE TWEETS(
    "ID" INTEGER NOT NULL,
    "USER_NAME" NVARCHAR(100),
    "CREATED_AT" DATE,
    "TEXT" NVARCHAR (140),
    "HASH_TAGS" NVARCHAR (100),
    PRIMARY KEY("ID")
);
CREATE SEQUENCE "I045664"."TWEET_SEQUENCE" INCREMENT BY 1 START WITH 1 NO CYCLE;
```

2. Expand the **Catalog** folder in HANA studio, you should find the table **TWEETS** in your schema and the definition of the table is like:

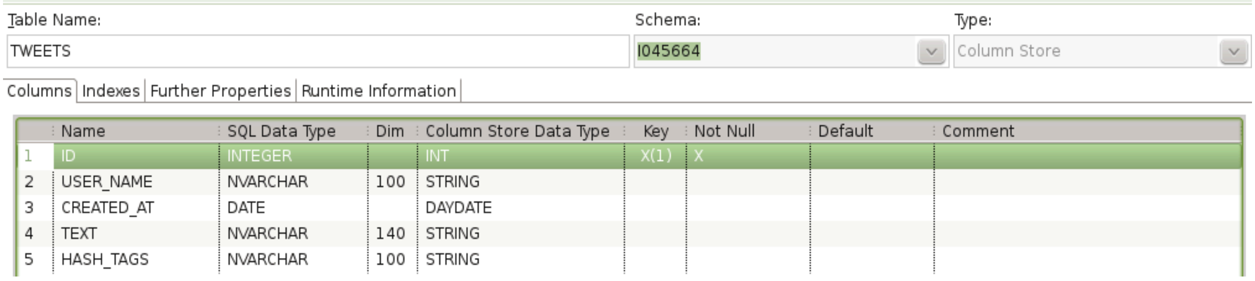


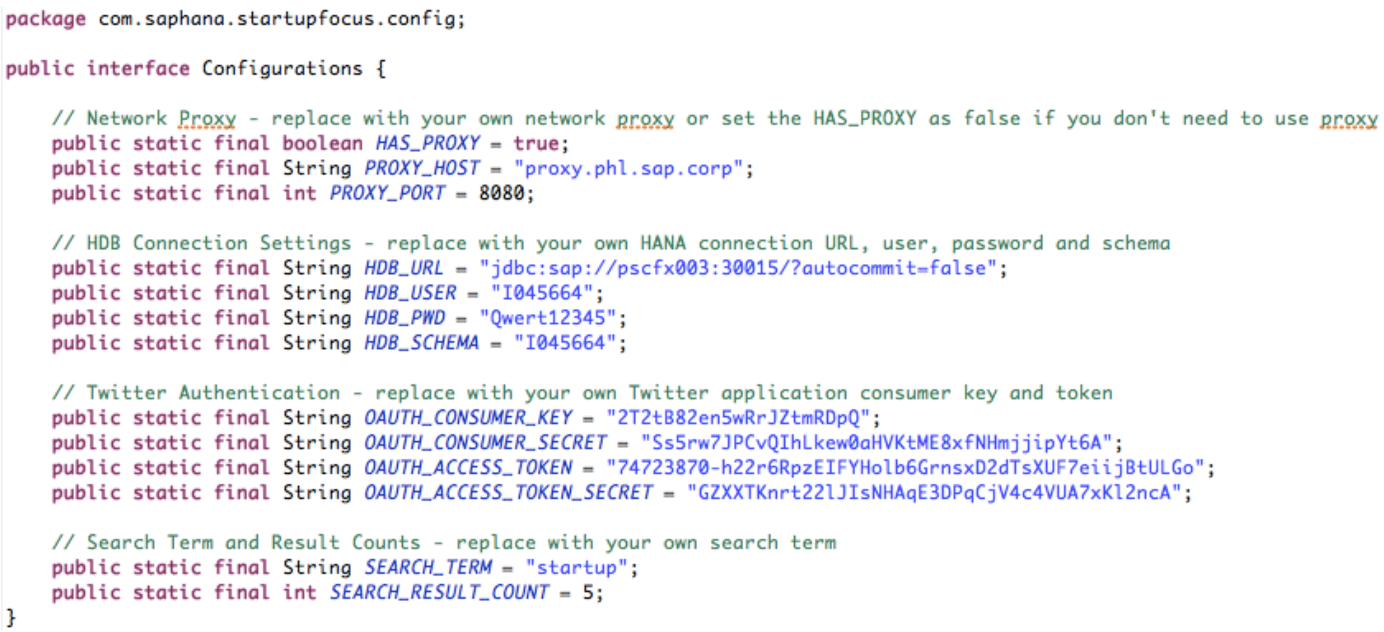
Table Name: TWEETS
Schema: I045664
Type: Column Store

	Name	SQL Data Type	Dim	Column Store Data Type	Key	Not Null	Default	Comment
1	ID	INTEGER		INT	X(1)	X		
2	USER_NAME	NVARCHAR	100	STRING				
3	CREATED_AT	DATE		DAYDATE				
4	TEXT	NVARCHAR	140	STRING				
5	HASH_TAGS	NVARCHAR	100	STRING				

Update the configurations

In the purpose to maintain the configurations easily, we put all the required information in a single interface and it is mandatory for you update it with your own account or settings before you can connect to either HANA or Twitter.

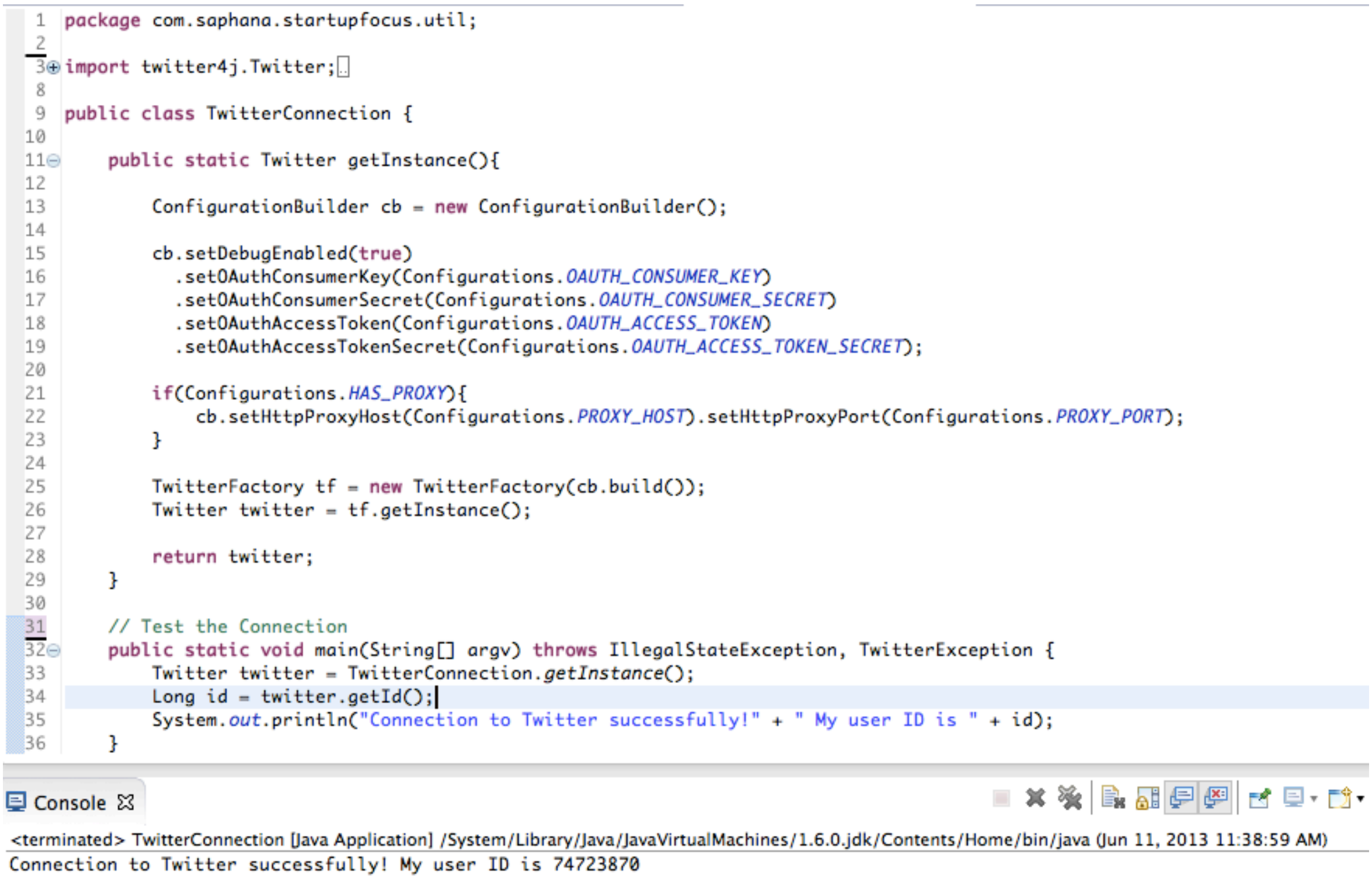
1. Open the file **Configurations.java** in your project. Basically, there are 4 category of setting you can override:
    - **Network Proxy Settings**: The proxy host and port, set the **HAS_PROXY** as false if you do not need to use proxy
    - **HANA Connection Settings**: Replace the HANA URL with your own HANA host and port, user, password and the schema where you created your table
    - **Twitter Authentication Settings:** Replace with your own authentication information from your twitter application as described in the prerequisites
    - **Search Term:** We will search the twitter based on the search term “startup” and we want to know what people were talking around the startups in twitter. You can always replace it with your own term if you are interested in other topics



```
package com.saphana.startupfocus.config;

public interface Configurations {

    // Network Proxy - replace with your own network proxy or set the HAS_PROXY as false if you don't need to use proxy
    public static final boolean HAS_PROXY = true;
    public static final String PROXY_HOST = "proxy.phl.sap.corp";
    public static final int PROXY_PORT = 8080;

    // HDB Connection Settings - replace with your own HANA connection URL, user, password and schema
    public static final String HDB_URL = "jdbc:sap://pscfx003:30015/?autocommit=false";
    public static final String HDB_USER = "I045664";
    public static final String HDB_PWD = "Qwert12345";
    public static final String HDB_SCHEMA = "I045664";

    // Twitter Authentication - replace with your own Twitter application consumer key and token
    public static final String OAUTH_CONSUMER_KEY = "2T2tB82en5wRrJZtmRDpQ";
    public static final String OAUTH_CONSUMER_SECRET = "Ss5rw7JPCvQIhLkew0aHVKtME8xfNHmjjipYt6A";
    public static final String OAUTH_ACCESS_TOKEN = "74723870-h22r6RpzEIFYHolb6GrnsxD2dTsXUF7eiijBtULGo";
    public static final String OAUTH_ACCESS_TOKEN_SECRET = "GZXXTKnrt22lJIsNHAqE3DPqCjV4c4VUA7xKl2ncA";

    // Search Term and Result Counts - replace with your own search term
    public static final String SEARCH_TERM = "startup";
    public static final int SEARCH_RESULT_COUNT = 5;
}
```

Test Connection to Twitter

Once have the twitter authentication maintained correctly in the previous step. You can open **TwitterConnection.java** and run it. You will see the message “Connection to Twitter Successfully!” following with your twitter user id in the console as the screenshot shows below.

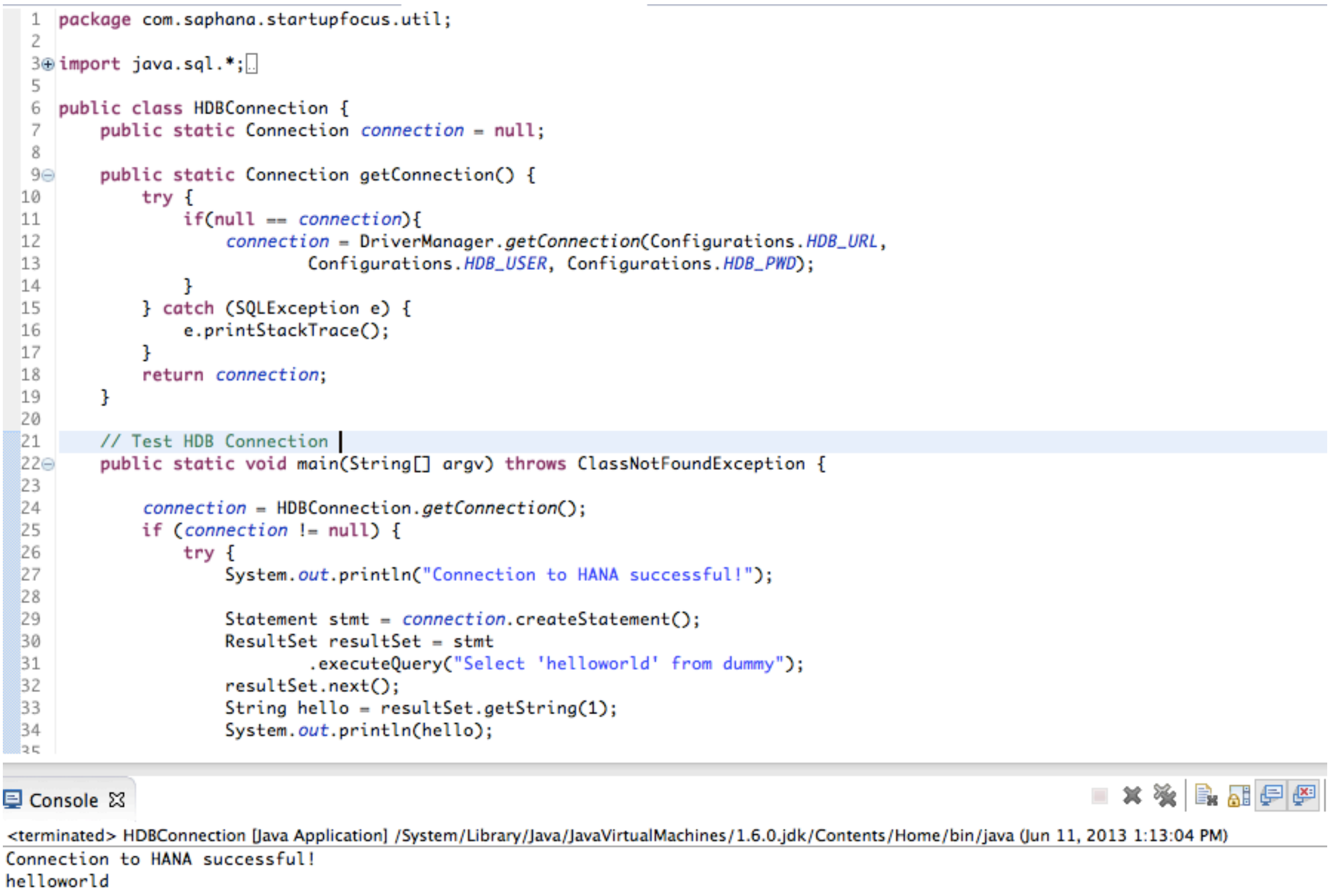


```
package com.saphana.startupfocus.util;

import twitter4j.Twitter;

public class TwitterConnection {

    public static Twitter getInstance(){

        ConfigurationBuilder cb = new ConfigurationBuilder();

        cb.setDebugEnabled(true)
          .setOAuthConsumerKey(Configurations.OAUTH_CONSUMER_KEY)
          .setOAuthConsumerSecret(Configurations.OAUTH_CONSUMER_SECRET)
          .setOAuthAccessToken(Configurations.OAUTH_ACCESS_TOKEN)
          .setOAuthAccessTokenSecret(Configurations.OAUTH_ACCESS_TOKEN_SECRET);

        if(Configurations.HAS_PROXY){
            cb.setHttpProxyHost(Configurations.PROXY_HOST).setHttpProxyPort(Configurations.PROXY_PORT);
        }

        TwitterFactory tf = new TwitterFactory(cb.build());
        Twitter twitter = tf.getInstance();

        return twitter;
    }

    // Test the Connection
    public static void main(String[] argv) throws IllegalStateException, TwitterException {
        Twitter twitter = TwitterConnection.getInstance();
        Long id = twitter.getId();
        System.out.println("Connection to Twitter successfully!" + " My user ID is " + id);
    }
```

```
<terminated> TwitterConnection [Java Application] /System/Library/Java/JavaVirtualMachines/1.6.0.jdk/Contents/Home/bin/java (Jun 11, 2013 11:38:59 AM)
Connection to Twitter successfully! My user ID is 74723870
```

Test Connection to SAP HANA

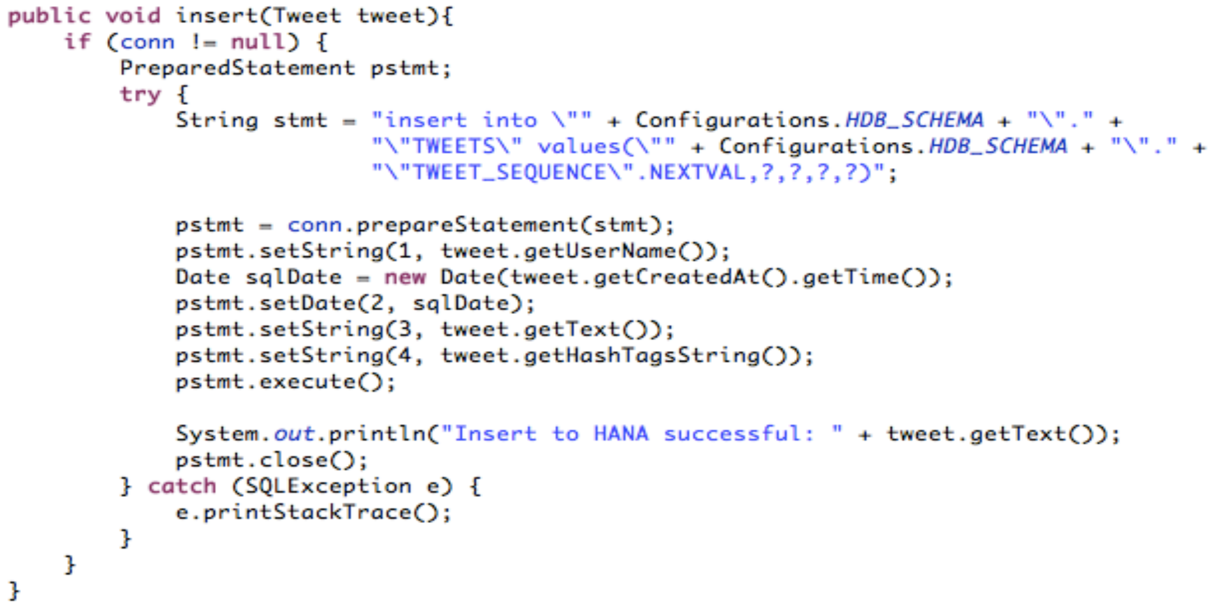
Now let us open the file **HDBConnection.java** and run it. You will see the message “Connection to HANA Successfully!” in the console as the screenshot shows below. Check the **Configurations.java** if you encountering any issue.



```java
package com.saphana.startupfocus.util;

import java.sql.*;

public class HDBConnection {
    public static Connection connection = null;

    public static Connection getConnection() {
        try {
            if(null == connection){
                connection = DriverManager.getConnection(Configurations.HDB_URL,
                        Configurations.HDB_USER, Configurations.HDB_PWD);
            }
        } catch (SQLException e) {
            e.printStackTrace();
        }
        return connection;
    }

    // Test HDB Connection
    public static void main(String[] argv) throws ClassNotFoundException {

        connection = HDBConnection.getConnection();
        if (connection != null) {
            try {
                System.out.println("Connection to HANA successful!");

                Statement stmt = connection.createStatement();
                ResultSet resultSet = stmt
                        .executeQuery("Select 'helloworld' from dummy");
                resultSet.next();
                String hello = resultSet.getString(1);
                System.out.println(hello);
```

```
<terminated> HDBConnection [Java Application] /System/Library/Java/JavaVirtualMachines/1.6.0.jdk/Contents/Home/bin/java (Jun 11, 2013 1:13:04 PM)
Connection to HANA successful!
helloworld
```

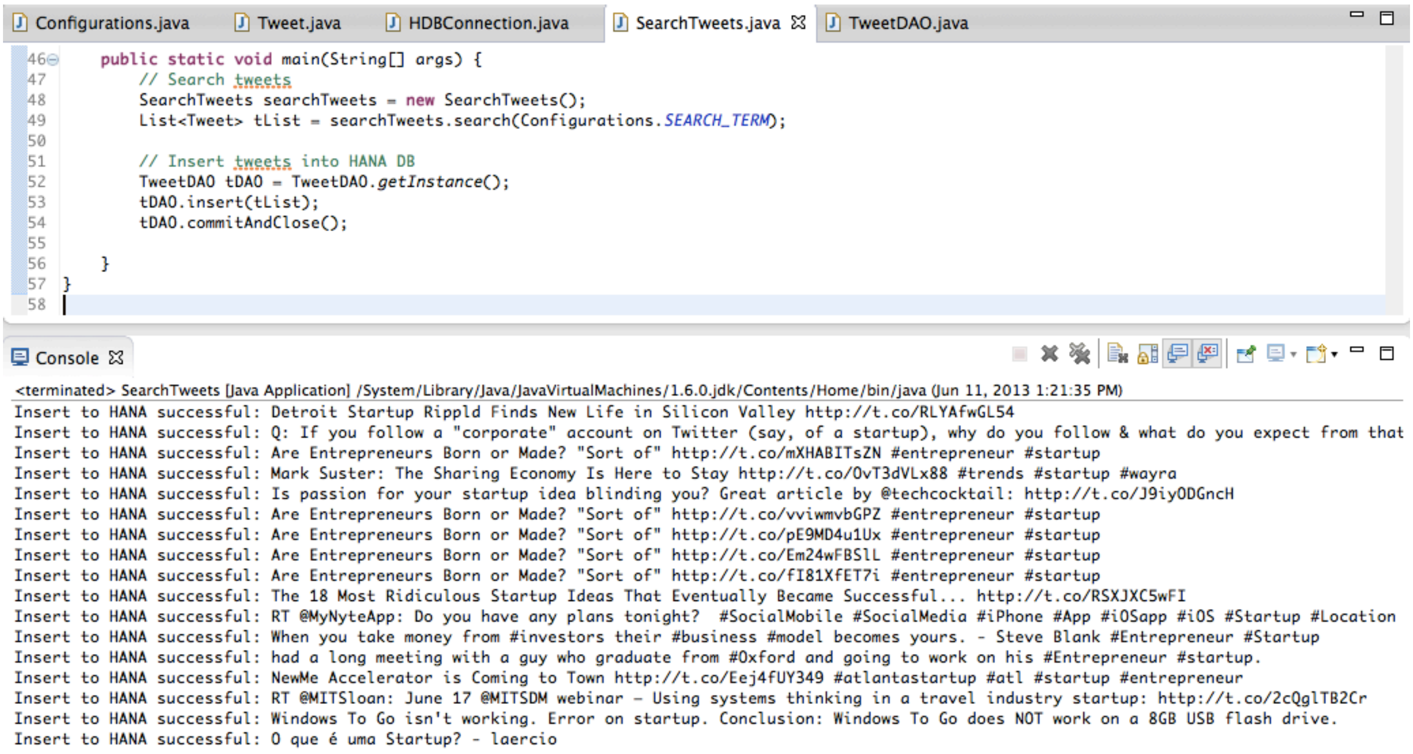
The data access object **TweetDAO** is the single point to communicate with HANA from java, take a look how the source code looks like and you will know the SQL statement and how to use the jdbc library.



```
public void insert(Tweet tweet){
    if (conn != null) {
        PreparedStatement pstmt;
        try {
            String stmt = "insert into \"" + Configurations.HDB_SCHEMA + "\"." +
                          "\"TWEETS\" values(\"" + Configurations.HDB_SCHEMA + "\"." +
                          "\"TWEET_SEQUENCE\".NEXTVAL,?,?,?,?)";

            pstmt = conn.prepareStatement(stmt);
            pstmt.setString(1, tweet.getUserName());
            Date sqlDate = new Date(tweet.getCreatedAt().getTime());
            pstmt.setDate(2, sqlDate);
            pstmt.setString(3, tweet.getText());
            pstmt.setString(4, tweet.getHashTagsString());
            pstmt.execute();

            System.out.println("Insert to HANA successful: " + tweet.getText());
            pstmt.close();
        } catch (SQLException e) {
            e.printStackTrace();
        }
    }
}
```

Invoke Twitter API and save the tweets into HANA

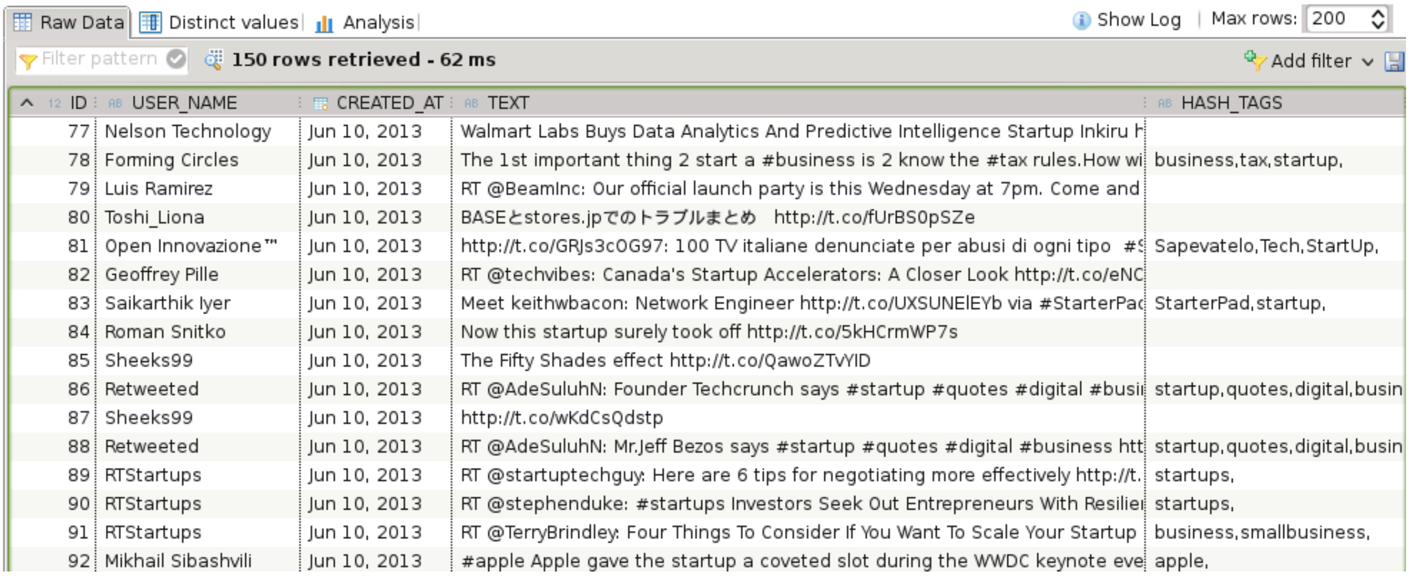
Now it's time to the do the real stuff. Open the file **SearchTweets.java** and run it, which will search the tweets based on the search term we specified in the **Configurations.java** and everything we got will saved to HANA table. You will see the messages in the console indicate the tweets have been inserted to HANA successfully like the screenshot shows:



```
    public static void main(String[] args) {
        // Search tweets
        SearchTweets searchTweets = new SearchTweets();
        List<Tweet> tList = searchTweets.search(Configurations.SEARCH_TERM);

        // Insert tweets into HANA DB
        TweetDAO tDAO = TweetDAO.getInstance();
        tDAO.insert(tList);
        tDAO.commitAndClose();

    }
}
```

```
<terminated> SearchTweets [Java Application] /System/Library/Java/JavaVirtualMachines/1.6.0.jdk/Contents/Home/bin/java (Jun 11, 2013 1:21:35 PM)
Insert to HANA successful: Detroit Startup Rippld Finds New Life in Silicon Valley http://t.co/RLYAfwGL54
Insert to HANA successful: Q: If you follow a "corporate" account on Twitter (say, of a startup), why do you follow & what do you expect from that
Insert to HANA successful: Are Entrepreneurs Born or Made? "Sort of" http://t.co/mXHABITsZN #entrepreneur #startup
Insert to HANA successful: Mark Suster: The Sharing Economy Is Here to Stay http://t.co/OvT3dVLx88 #trends #startup #wayra
Insert to HANA successful: Is passion for your startup idea blinding you? Great article by @techcocktail: http://t.co/J9iyODGncH
Insert to HANA successful: Are Entrepreneurs Born or Made? "Sort of" http://t.co/vviwmvbGPZ #entrepreneur #startup
Insert to HANA successful: Are Entrepreneurs Born or Made? "Sort of" http://t.co/pE9MD4u1Ux #entrepreneur #startup
Insert to HANA successful: Are Entrepreneurs Born or Made? "Sort of" http://t.co/Em24wFBSlL #entrepreneur #startup
Insert to HANA successful: Are Entrepreneurs Born or Made? "Sort of" http://t.co/fI81XfET7i #entrepreneur #startup
Insert to HANA successful: The 18 Most Ridiculous Startup Ideas That Eventually Became Successful... http://t.co/RSXJXC5wFI
Insert to HANA successful: RT @MyNyteApp: Do you have any plans tonight?  #SocialMobile #SocialMedia #iPhone #App #iOSapp #iOS #Startup #Location
Insert to HANA successful: When you take money from #investors their #business #model becomes yours. - Steve Blank #Entrepreneur #Startup
Insert to HANA successful: had a long meeting with a guy who graduate from #Oxford and going to work on his #Entrepreneur #startup.
Insert to HANA successful: NewMe Accelerator is Coming to Town http://t.co/Eej4fUY349 #atlantastartup #atl #startup #entrepreneur
Insert to HANA successful: RT @MITSloan: June 17 @MITSDM webinar – Using systems thinking in a travel industry startup: http://t.co/2cQglTB2Cr
Insert to HANA successful: Windows To Go isn't working. Error on startup. Conclusion: Windows To Go does NOT work on a 8GB USB flash drive.
Insert to HANA successful: O que é uma Startup? - laercio
```

After that, you can run the data preview in HANA studio and see the contents of the table **TWEETS** in your schema like this:



150 rows retrieved - 62 ms

ID	USER_NAME	CREATED_AT	TEXT	HASH_TAGS
77	Nelson Technology	Jun 10, 2013	Walmart Labs Buys Data Analytics And Predictive Intelligence Startup Inkiru h	
78	Forming Circles	Jun 10, 2013	The 1st important thing 2 start a #business is 2 know the #tax rules.How wi	business,tax,startup,
79	Luis Ramirez	Jun 10, 2013	RT @BeamInc: Our official launch party is this Wednesday at 7pm. Come and	
80	Toshi_Liona	Jun 10, 2013	BASEとstores.jpでのトラブルまとめ　http://t.co/fUrBS0pSZe	
81	Open Innovazione™	Jun 10, 2013	http://t.co/GRJs3cOG97: 100 TV italiane denunciate per abusi di ogni tipo #S	Sapevatelo,Tech,StartUp,
82	Geoffrey Pille	Jun 10, 2013	RT @techvibes: Canada's Startup Accelerators: A Closer Look http://t.co/eNC	
83	Saikarthik Iyer	Jun 10, 2013	Meet keithwbacon: Network Engineer http://t.co/UXSUNElEYb via #StarterPad	StarterPad,startup,
84	Roman Snitko	Jun 10, 2013	Now this startup surely took off http://t.co/5kHCrmWP7s	
85	Sheeks99	Jun 10, 2013	The Fifty Shades effect http://t.co/QawoZTvYID	
86	Retweeted	Jun 10, 2013	RT @AdeSuluhN: Founder Techcrunch says #startup #quotes #digital #busi	startup,quotes,digital,busin
87	Sheeks99	Jun 10, 2013	http://t.co/wKdCsQdstp	
88	Retweeted	Jun 10, 2013	RT @AdeSuluhN: Mr.Jeff Bezos says #startup #quotes #digital #business htt	startup,quotes,digital,busin
89	RTStartups	Jun 10, 2013	RT @startuptechguy: Here are 6 tips for negotiating more effectively http://t.	startups,
90	RTStartups	Jun 10, 2013	RT @stephenduke: #startups Investors Seek Out Entrepreneurs With Resilie	startups,
91	RTStartups	Jun 10, 2013	RT @TerryBrindley: Four Things To Consider If You Want To Scale Your Startup	business,smallbusiness,
92	Mikhail Sibashvili	Jun 10, 2013	#apple Apple gave the startup a coveted slot during the WWDC keynote eve	apple,

Run text analysis in HANA

Now we already have the tweets stored in the HANA table. The next step, we are going to run the text analysis to see what people are talking around the "startup" in twitter.

To run the text analysis, the only thing we need to do is creating a fulltext index for the column of the table we want to analysis and HANA will process the linguistic analysis, entity extraction, stemming for us and save the results in a generated table **$TA_YOUR_INDEX_NAME** at the same schema. After that, you can build views on top of the table and leverage all existing analysis tools around HANA to do the visualization even the predictive analysis.

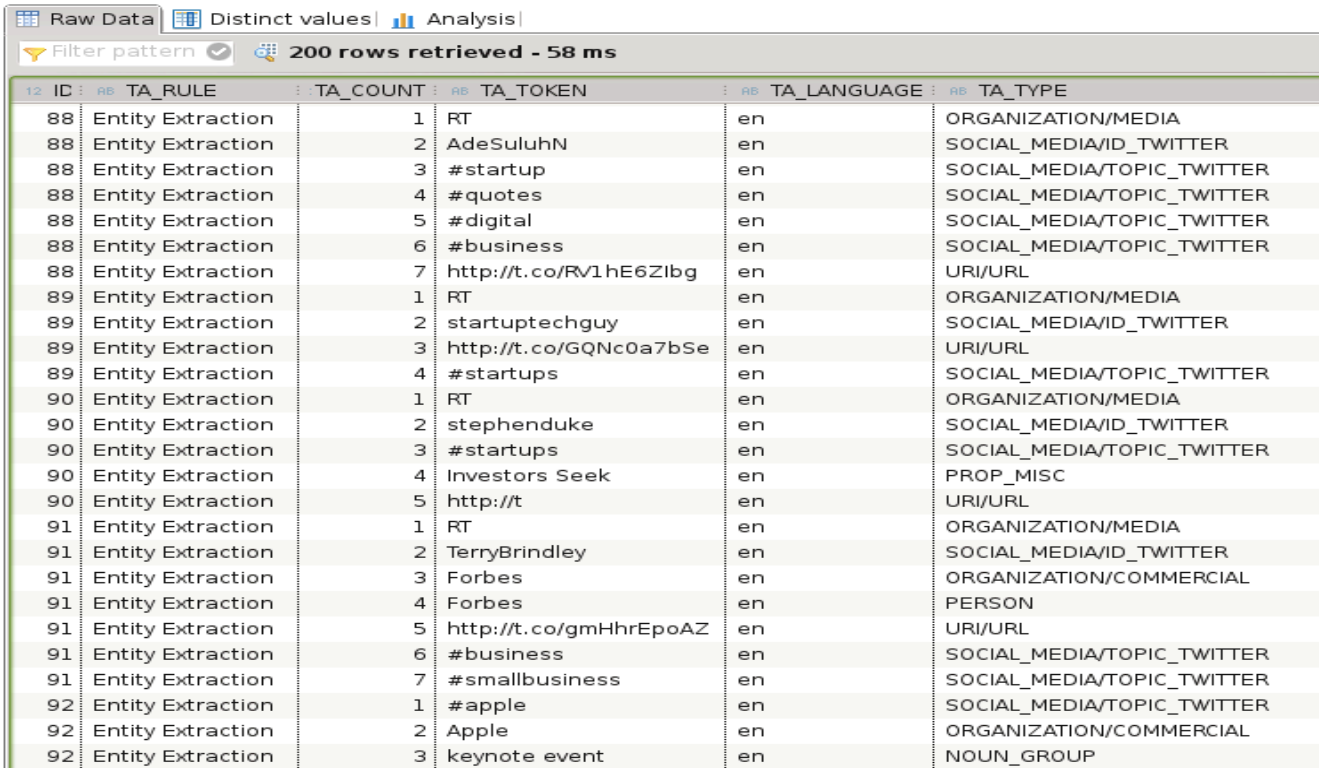
1. Copy the SQL statement from the **CreateFullTextIndex.sql** and execute it in SQL console:

```
-- Replace the Scheme with your own Schema! --
SET SCHEMA "I045664";
DROP FULLTEXT INDEX "TWEETS_FTI";
Create FullText Index "TWEETS_FTI" On "TWEETS"("TEXT")
TEXT ANALYSIS ON CONFIGURATION 'EXTRACTION_CORE';
```

2. Do you believe the text analysis is already done by HANA? Yes, it is. Now you know how simple it is! You will be able to find a generated table $TA_TWEETS_FTI in your schema. The structure of the table looks like this, which is the standardized format for the results of text analysis:

Column Name	Key	Description	Data Type
ID	Yes	This is the primary key of my table.	Same as in source table.
RULE	Yes	Stores the rule package that yielded the token. In my case: “Entity Extraction”	NVARCHAR(200)
COUNTER	Yes	Counts all tokens across the document or table column	BIGINT
TOKEN	No	Term or entity - depending on processing type. (The “who”, “what”, “where”, “when” and “how much”)	NVARCHAR(250)
LANGUAGE	No	You can either specify a language column when you create the fulltext index or it can be derived from the text.	NVARCHAR(2)
TYPE	No	The token type contains the linguistic or semantic type of the token; for instance "noun" (if option = LINGANALYSIS_*) or "company" (if option = EXTRACTION_*).	NVARCHAR(100)
NORMALIZED	No	Stores a normalized representation of the token. This becomes relevant e.g. for German with umlauts, or ß/ss. Normalization with regards to capitalization would not be as important as to justify this column.	NVARCHAR(250)
STEM	No	Stores the linguistic stemming information, e.g. the singular nominative for nouns, or the indicative for verbs. If text analysis yields several stems, only the first stem will be stored, assuming this to be the best match.	NVARCHAR(300)
PARAGRAPH	No	The paragraph number where my token is located in the document	INTEGER
SENTENCE	No	The sentence number where my token is located in the document	INTEGER
CREATED_AT	No	Creation timestamp	TIMESTAMP

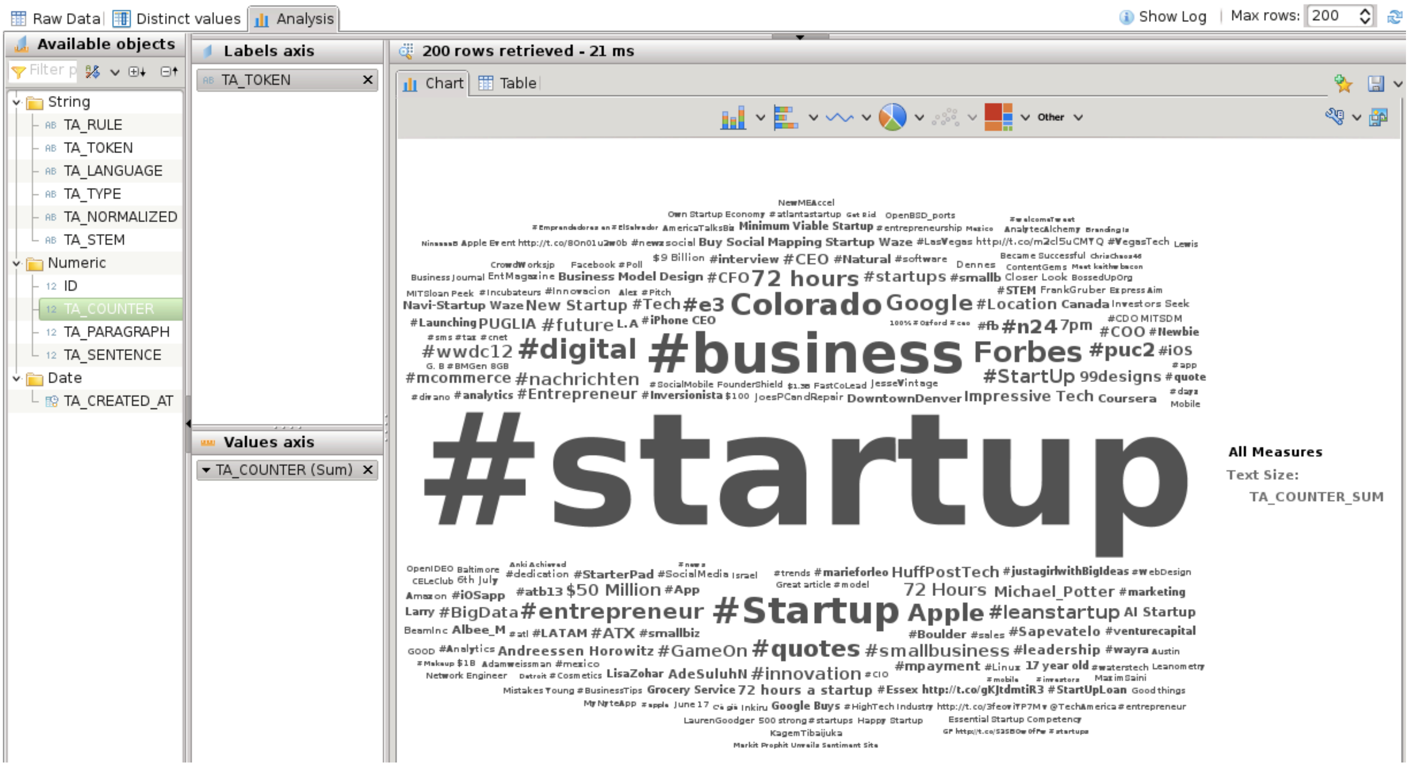
3. And here is the data preview of the $TA table, you will see the Tokens extracted from the tweets and the number of occurrence and entity type of each token.



200 rows retrieved - 58 ms

ID	TA_RULE	TA_COUNT	TA_TOKEN	TA_LANGUAGE	TA_TYPE
88	Entity Extraction	1	RT	en	ORGANIZATION/MEDIA
88	Entity Extraction	2	AdeSuluhN	en	SOCIAL_MEDIA/ID_TWITTER
88	Entity Extraction	3	#startup	en	SOCIAL_MEDIA/TOPIC_TWITTER
88	Entity Extraction	4	#quotes	en	SOCIAL_MEDIA/TOPIC_TWITTER
88	Entity Extraction	5	#digital	en	SOCIAL_MEDIA/TOPIC_TWITTER
88	Entity Extraction	6	#business	en	SOCIAL_MEDIA/TOPIC_TWITTER
88	Entity Extraction	7	http://t.co/RV1hE6ZIbg	en	URI/URL
89	Entity Extraction	1	RT	en	ORGANIZATION/MEDIA
89	Entity Extraction	2	startuptechguy	en	SOCIAL_MEDIA/ID_TWITTER
89	Entity Extraction	3	http://t.co/GQNc0a7bSe	en	URI/URL
89	Entity Extraction	4	#startups	en	SOCIAL_MEDIA/TOPIC_TWITTER
90	Entity Extraction	1	RT	en	ORGANIZATION/MEDIA
90	Entity Extraction	2	stephenduke	en	SOCIAL_MEDIA/ID_TWITTER
90	Entity Extraction	3	#startups	en	SOCIAL_MEDIA/TOPIC_TWITTER
90	Entity Extraction	4	Investors Seek	en	PROP_MISC
90	Entity Extraction	5	http://t	en	URI/URL
91	Entity Extraction	1	RT	en	ORGANIZATION/MEDIA
91	Entity Extraction	2	TerryBrindley	en	SOCIAL_MEDIA/ID_TWITTER
91	Entity Extraction	3	Forbes	en	ORGANIZATION/COMMERCIAL
91	Entity Extraction	4	Forbes	en	PERSON
91	Entity Extraction	5	http://t.co/gmHhrEpoAZ	en	URI/URL
91	Entity Extraction	6	#business	en	SOCIAL_MEDIA/TOPIC_TWITTER
91	Entity Extraction	7	#smallbusiness	en	SOCIAL_MEDIA/TOPIC_TWITTER
92	Entity Extraction	1	#apple	en	SOCIAL_MEDIA/TOPIC_TWITTER
92	Entity Extraction	2	Apple	en	ORGANIZATION/COMMERCIAL
92	Entity Extraction	3	keynote event	en	NOUN_GROUP

4. Based on this, you can use the knowledge you learned in the previous modelling exercises and use the table to build a view if you want. Here, we just go to the Analysis tab and build a tag cloud like this:



Raw Data
Distinct values
Analysis
Show Log
Max rows: 200
Available objects
Filter p
String
TA_RULE
TA_TOKEN
TA_LANGUAGE
TA_TYPE
TA_NORMALIZED
TA_STEM
Numeric
ID
TA_COUNTER
TA_PARAGRAPH
TA_SENTENCE
Date
TA_CREATED_AT
Labels axis
TA_TOKEN
Values axis
TA_COUNTER (Sum)
200 rows retrieved - 21 ms
Chart
Table
Other
#startup
#business
Colorado
Google
Forbes
#digital
#Startup
#entrepreneur
Apple
#quotes
72 hours
#startups
#Entrepreneur
#StartUp
#mcommerce
#nachrichten
#wwdc12
#n24
#puc2
New Startup
#Tech
#e3
#CFO
#CEO
Business Model Design
Minimum Viable Startup
Buy Social Mapping Startup Waze
Impressive Tech
DowntownDenver
HuffPostTech
72 Hours
Michael_Potter
#leanstartup
#smallbusiness
#innovation
Andreessen Horowitz
#GameOn
#BigData
$50 Million
#iOSapp
AdeSuluhN
LisaZohar
72 hours a startup
Grocery Service
Google Buys
#mpayment
#Sapevatelo
#venturecapital
#leadership
All Measures
Text Size:
TA_COUNTER_SUM