

# PFM

# HOTEL ADVISOR

## MANUAL DE INSTALACIÓN

AUTOR: FERNANDO MÁRQUEZ ARROYO



**KSCHOOL**

# ÍNDICE

1. INSTALACIÓN DE COTS (UBUNTU 16.04).....	3
1.1 INSTALACIÓN JAVA 8.....	3
1.2 INSTALACIÓN MAVEN 3.....	3
1.3 INSTALACIÓN SPARK 1.6.2 (PROCESAMIENTO IN-MEMORY).....	3
1.4 INSTALACIÓN PYCHARM (IDE PYTHON) .....	3
1.5 INSTALACIÓN KAFKA 2.11 (BRÓKER DE MENSAJERÍA) .....	4
1.6 INSTALACIÓN JSON DATA GENERATOR (SIMULADOR DE EVENTOS) .....	4
1.7 INSTALACIÓN INTELLIJ IDEA (IDE JAVA) .....	5
1.8 INSTALACIÓN MONGODB 3.4 (BASE DE DATOS ORIENTADA A DOCUMENTOS) .....	5
1.9 INSTALACIÓN ELASTIC SEARCH 5.2.2 (MOTOR DE BÚSQUEDA) .....	5
1.10 INSTALACIÓN MONGO-CONNECTOR Y ELASTIC-DOC-MANAGER.....	5
1.11 INSTALACIÓN KIBANA 5.2.2 (DASHBOARDS) .....	6
2. CREACIÓN BASE DE DATOS EN MONGODB.....	6
3. CREACIÓN MOTOR DE BÚSQUEDA EN ELASTICSEARCH.....	7
4. PROCESAMIENTO BATCH (BATCH PROCESSING).....	11
5. PROCESAMIENTO REAL TIME (REAL TIME PROCESSING).....	14
6. EXPLOTACIÓN DE LA INFORMACIÓN.....	16
6.1 EJECUCIÓN DE LA APLICACIÓN.....	16
6.2 DASHBOARDS / ANALYTICS.....	

## 1. INSTALACIÓN DE COTS (UBUNTU 16.04)

### 1.1 INSTALACIÓN JAVA 8

```
sudo apt-get update
```

```
sudo apt-get install default-jdk
```

### 1.2. INSTALACIÓN MAVEN 3

```
sudo apt-get install maven
```

Comprobar versión: `mvn -version`

### 1.3. INSTALACIÓN SPARK 1.6.2 (PROCESAMIENTO IN-MEMORY)

```
wget http://d3kbcqa49mib13.cloudfront.net/spark-1.6.2-bin-hadoop2.6.tgz
```

```
tar -xvf spark-1.6.2-bin-hadoop2.6.tgz
```

```
mv spark-1.6.2-bin-hadoop2.6 /opt
```

### 1.4. INSTALACIÓN PYCHARM (IDE PYTHON)

```
wget https://download.jetbrains.com/python/pycharm-community-2016.2.3.tar.gz
```

```
gzip -dc pycharm-community-2016.2.3.tar.gz | tar xf -
```

Configurar PySpark en PyCharm:

Editar `/home/<user>/bashrc`, y añadir las siguientes líneas:

```
export SPARK_HOME=/opt/spark-1.6.2-bin-hadoop2.6
export PYTHONPATH=$PYTHONPATH:$SPARK_HOME/python
export PYTHONPATH=$PYTHONPATH:$SPARK_HOME/python/lib/py4j-0.9-src.zip
export PYTHONPATH=$PYTHONPATH:$SPARK_HOME/python/pyspark
```

Ejecutar PyCharm: `/home/<user>/pycharm-community-2016.2.3/bin/pycharm.sh`

Instalar PyMongo (API de MongoDB para Python):

```
sudo apt-get install python-pip
```

```
python -m pip install pymongo==3.4.0
```

## 1.5. INSTALACIÓN KAFKA 2.11 (BRÓKER DE MENSAJERÍA)

```
wget http://apache.rediris.es/kafka/0.10.1.0/kafka_2.11-0.10.1.0.tgz
```

```
tar -xzf kafka_2.11-0.10.1.0.tgz
```

```
cd kafka_2.11-0.10.1.0
```

Arrancar Zookeeper: `bin/zookeeper-server-start.sh config/zookeeper.properties`

Arrancar Kafka server: `bin/kafka-server-start.sh config/server.properties`

Instalar python-kafka:

```
cd $SPARK_HOME
```

```
sudo apt-get install python-setuptools python-dev build-essential
```

```
sudo easy_install pip
```

```
sudo apt-get install git
```

```
sudo git clone https://github.com/mumrah/kafka-python
```

```
sudo pip install kafka-python
```

```
cd $SPARK_HOME/lib
```

```
wget http://central.maven.org/maven2/org/apache/spark/spark-streaming-kafka-assembly\_2.10/1.6.2/spark-streaming-kafka-assembly\_2.10-1.6.2.jar
```

Cómo ejecutar un programa de streaming python-spark:

```
cd $SPARK_HOME
```

```
bin/spark-submit --jars lib/spark-streaming-kafka-assembly_2.10-1.6.2.jar ../MainStreaming.py
```

## 1.6. INSTALACIÓN JSON DATA GENERATOR (SIMULADOR DE EVENTOS)

```
git clone git@github.com:acesinc/json-data-generator.git
```

```
cd json-data-generator
```

```
mvn clean package
```

```
cp target/json-data-generator-1.2.1-bin.tar /opt
```

```
cd /opt
```

```
tar xvf json-data-generator-1.2.1-bin.tar
```

```
cd json-data-generator
```

```
java -jar json-data-generator-1.2.1.jar <config_file_simulation>.json
```

## 1.7. INSTALACIÓN INTELLIJ IDEA (IDE JAVA)

Descargar el fichero <ideaIC or ideaIU>-\*.\*.tar.gz de la página de descargas: <https://www.jetbrains.com/idea/>

Descomprimir el fichero <ideaIC or ideaIU>-\*.\*.tar.gz a una carpeta diferente:

```
tar xfz <ideaIC or ideaIU>-*.*.tar.gz <new_archive_folder>
```

Cambiar al directorio bin: `cd <new archive folder>/<ideaIC or ideaIU>-*.*.bin`

Ejecutar el fichero idea.sh

## 1.8. INSTALACIÓN MONGODB 3.4 (BASE DE DATOS ORIENTADA A DOCUMENTOS)

Importar la clave pública usada por el sistema gestor de paquetes:

```
sudo apt-key adv --keyserver hkp://keyserver.ubuntu.com:80 --recv 0C49F3730359A14518585931BC711F9BA15703C6
```

Crear un fichero de lista para MongoDB:

```
echo "deb [ arch=amd64,arm64 ] http://repo.mongodb.org/apt/ubuntu xenial/mongodb-org/3.4 multiverse" | sudo tee /etc/apt/sources.list.d/mongodb-org-3.4.list
```

Recargar el paquete local de la base de datos:

```
sudo apt-get update
```

Instalar los paquetes de MongoDB:

```
sudo apt-get install -y mongodb-org
```

## 1.9 INSTALACIÓN ELASTIC SEARCH 5.2.2 (MOTOR DE BÚSQUEDA)

```
wget https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-5.2.2.tar.gz
```

```
tar -xzf elasticsearch-5.2.2.tar.gz
```

## 1.10 INSTALACIÓN MONGO-CONNECTOR Y ELASTIC-DOC-MANAGER

```
sudo pip install --upgrade pip
```

```
sudo pip install mongo-connector
```

```
sudo pip install elastic-doc-manager[elastic5]
```

```
sudo chmod 777 /var/log
```

```
sudo chmod 777 /opt/
```

```
sudo chmod 777 /usr/local/bin
```

## 1.11. INSTALACIÓN KIBANA 5.2.2 (DASHBOARDS)

```
wget https://artifacts.elastic.co/downloads/kibana/kibana-5.2.2-linux-x86_64.tar.gz
```

```
tar -xzf kibana-5.2.2-linux-x86_64.tar.gz
```

## 2. CREACIÓN BASE DE DATOS EN MONGODB

1. Creación del replicaSet de MongoDB:

```
sudo mongod --replSet "rs0" --dbpath "/var/lib/mongodb"
```

2. Arrancar MongoDB:

```
mongo
```

3. Iniciar el Replica Set:

```
rs.initiate()
```

```
rs.conf()
```

4. Crear base de datos de HotelAdvisor:

```
use HotelAdvisor
```

5. Crear índices para la colección “hotels”:

```
db.hotels.createIndex({"coord" : "2dsphere"})
```

```
db.hotels.createIndex({"continent_id" : 1})
```

```
db.hotels.createIndex({"cc1" : 1})
```

```
db.hotels.createIndex({"city_hotel":1})
```

```
db.hotels.createIndex({"min_rate_EUR" : 1})
```

```
db.hotels.createIndex({"review_score" : 1})
```

```
db.hotels.createIndex({"nr_rooms" : 1})
```

```
db.hotels.createIndex({"id":1})
```

```
db.hotels.createIndex({"name":1})
```

6. Crear índices para la colección “pois”:

```
db.pois.createIndex({"poi_coord" : "2dsphere"})
```

```
db.pois.createIndex({"poi_type":1})
```

```
db.pois.createIndex({"poi_name":1})
```

```
db.pois.createIndex({"poi_osm_id":1})
```

### 3. CREACIÓN MOTOR DE BÚSQUEDA EN ELASTICSEARCH

1. Arrancar Elastic Search:

```
~elasticsearch-5.2.2/bin$ ./elasticsearch
```

2. Crear índice de hoteladvisor para el tipo “hotels” (es importante mapear campo "coord" a Geo-Point (para los Tile Map de Kibana) y “date” a Date):

```
curl -X PUT localhost:9200/hoteladvisor -d '
```

```
{
  "mappings" : {
    "hotels" : {
      "properties" : {
        "address" : {
          "type" : "text",
          "fields" : {
            "keyword" : {
              "type" : "keyword",
              "ignore_above" : 256
            }
          }
        },
        "cc1" : {
          "type" : "text",
          "fields" : {
            "keyword" : {
              "type" : "keyword",
              "ignore_above" : 256
            }
          }
        },
        "city_hotel" : {
          "type" : "text",
          "fields" : {
            "keyword" : {
              "type" : "keyword",
              "ignore_above" : 256
            }
          }
        },
        "city_preferred" : {
          "type" : "text",
```



```
"fields" : {
  "keyword" : {
    "type" : "keyword",
    "ignore_above" : 256
  }
},
"city_unique" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"class" : {
  "type" : "float"
},
"continent_id" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"coord" : {
  "type" : "geo_point"
},
"currencycode" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"desc_en" : {
  "type" : "text",
  "fields" : {
```

```
"keyword" : {
  "type" : "keyword",
  "ignore_above" : 256
}
},
"desc_es" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"hotel_url" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"id" : {
  "type" : "long"
},
"max_rate_EUR" : {
  "type" : "float"
},
"maxrate" : {
  "type" : "float"
},
"min_rate_EUR" : {
  "type" : "float"
},
"minrate" : {
  "type" : "float"
},
"name" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
```

```
    "type" : "keyword",
    "ignore_above" : 256
  }
},
"nr_rooms" : {
  "type" : "long"
},
"photo_url" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"preferred" : {
  "type" : "long"
},
"public_ranking" : {
  "type" : "long"
},
"review_nr" : {
  "type" : "long"
},
"review_score" : {
  "type" : "float"
},
"reviews" : {
  "properties" : {
    "cleanliness" : {
      "type" : "float"
    }
  },
  "comment" : {
    "type" : "text",
    "fields" : {
      "keyword" : {
        "type" : "keyword",
        "ignore_above" : 256
      }
    }
  }
},
```

```
"confort" : {
  "type" : "float"
},
"date" : {
  "type" : "date"
},
"installations" : {
  "type" : "float"
},
"lang" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"location" : {
  "type" : "float"
},
"personal" : {
  "type" : "float"
},
"score" : {
  "type" : "float"
},
"username" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
},
"valueMoney" : {
  "type" : "float"
},
"wifi" : {
  "type" : "float"
}
}
```

```

},
"ufi" : {
  "type" : "long"
},
"zip" : {
  "type" : "text",
  "fields" : {
    "keyword" : {
      "type" : "keyword",
      "ignore_above" : 256
    }
  }
}
}
}
}
}
}'

```

3. Crear indice de hoteladvisor para el tipo “pois” (es importante mapear campo "poi\_coord" a Geo-Point (para los Tile Map de Kibana)):

```
curl -X PUT localhost:9200/hoteladvisor/pois/_mapping -d '
```

```

{
  "pois" : {
    "properties" : {
      "poi_name" : {
        "type" : "text",
        "fields" : {
          "keyword" : {
            "type" : "keyword",

```

```

        "ignore_above" : 256
      }
    }
  },
  "poi_type" : {
    "type" : "text",
    "fields" : {
      "keyword" : {
        "type" : "keyword",
        "ignore_above" : 256
      }
    }
  },
  "poi_coord" : {
    "type" : "geo_point"
  }
}
}
'

```

## 4. PROCESAMIENTO BATCH (BATCH PROCESSING)

1. Copiar la carpeta “HotelAdvisor – Processing Python-Spark\home\fer\PycharmProjects\HotelAdvisor” de la entrega a la carpeta “/home/<user>/PycharmProjects”

2. Arrancar el conector de MongoDB para conectar la información generada en MongoDB al motor de búsqueda de Elastic Search:

```
mongo-connector -m localhost:27017 -t localhost:9200 -d elastic_doc_manager -n HotelAdvisor.hotels
```

3. Lanzar el job de Spark para el parseo de los ficheros del dataset de hoteles:

```
~/spark-1.6.2-bin-hadoop2.6$ bin/spark-submit --name "HotelAdvisor" --master local[4]
/home/<user>/PycharmProjects/HotelAdvisor/HotelReader.py
```

Aparecen trazas por consola indicando el número de hoteles parseados con éxito y cuáles tenían un formato incorrecto:

```
Read 24110 hotels in Africa, 24104 succesfully parsed, 6 with wrong format
```

```
Hotels Africa with wrong format: 6
```

```
1450598 Ararat Hotel      ASMARA ROAD
      Addis Ababa      et      -603097 0.0      USD      45      94
38.8182109594345 9.02638589061076 1      http://www.booking.com/hotel/et/ararat.html
http://aff.bstatic.com/images/hotel/max500/512/51272719.jpg      Featuring free WiFi and a restaurant, Ararat Hotel offers
accommodation in Addis Ababa. Guests can enjoy the on-site bar. Free private parking is available on site. All rooms come with a flat-
screen TV with satellite channels.      Отель
Ararat расположен в Аддис-Абебе, в 6 км от собора Святой Троицы и от конференц-центра UNECA. На территории работает бар.
```

يقدم فندق أارات مكان إقامة في أديس أبابا ويوفر خدمة الواي فاي مجانًا ومطعم، ويمكن للضيوف الاستمتاع بالبار المتوفر في الموقع، كما تتوفر مواقف خاصة للسيارات في الموقع مجانًا.

1521806	RedWood Place	No.2 Tanba Street Off Parakou Crescent	addis-ababa	Addis Ababa	5	1	1
USD	49.8	3	7.46796470722575	9.07913343237809	1		
http://www.booking.com/hotel/ng/redwood-place.html http://aff.bstatic.com/images/hotel/max500/822/82212481.jpg							

RedWood Place offers accommodation in Wuse 2, Abuja. Guests can enjoy the on-site restaurant. Free private parking is available on site. Some rooms feature a seating area where you can relax. RedWood Place features free WiFi throughout the property.

Situé à Abuja, le RedWood Place dispose d'un restaurant. Le stationnement privé sur place est gratuit. La connexion Wi-Fi est gratuite dans l'ensemble des locaux du RedWood Place.

Das RedWood Place bietet Ihnen eine Unterkunft in Abuja. Freuen Sie sich in der Unterkunft auf ein Restaurant. Die Privatparkplätze an der Unterkunft stehen Ihnen kostenfrei zur Verfügung.

RedWood Place 旅馆位于阿布贾 (Abuja) 的 Wuse 2 地区，设有内部餐厅以及免费内部私人停车场。部分客房设有供客人放松身心的休息区。RedWood Place 旅馆提供覆盖各处的免费 WiFi。客房提供卫星电视。旅馆提供共用休息室、客房服务和理发店。旅馆提供汽车租赁服务，距离最近的机场——纳姆迪·阿齐基韦国际机场 (Nnamdi Azikiwe International Airport) 23 公里。

Гостевой дом Redwood Place расположен в столичном городе Абуджа. К услугам гостей ресторан и бесплатная частная парковка на прилегающей территории. Кроме того, на всей территории гостевого дома Redwood Place работает бесплатный Wi-Fi.

1851303	Best Western Plus Setif Hotel CITE EL ALIA LA GARE	19000	Sétif	dz	-484254	4.0	DZD	9600	83	5.42212307453156
http://www.booking.com/hotel/dz/setif.html http://aff.bstatic.com/images/hotel/max500/827/82755949.jpg										

Offering a terrace and views of the city, Best Western Plus Setif Hotel is situated in Sétif. Guests can enjoy the on-site restaurant. Each room at this hotel is air conditioned and is fitted with a flat-screen TV with satellite channels.

Doté d'une terrasse avec vue sur la ville, le Best Western Plus Setif Hotel est situé à Sétif. Il possède un restaurant. Toutes ses chambres sont équipées de la climatisation et d'une télévision par satellite à écran plat.

Mit einer Terrasse und Stadtblick erwartet Sie das Best Western Plus Setif Hotel in Sétif. Das hauseigene Restaurant erwartet gern Ihren Besuch. Alle Zimmer dieses Hotels sind klimatisiert und mit einem Flachbild-Sat-TV ausgestattet.

Отель Best Western Plus Setif с террасой с видом на город расположен в городе Сетиф. На территории отеля работает ресторан. Все номера оснащены кондиционером и телевизором с плоским экраном и спутниковыми каналами.

يوفر فندق ويست ويسترن بلاس - سطيف في مدينة سطيف، ويوفر تراس وإطلالة على المدينة، ويمكن للضيوف الاستمتاع بالمطعم الموجود في الموقع.

1569160	Sabaku Tour Camp merzouga	, DMS N 31° 4' 48.603"	setif	Sétif	5	1	1
-39390	0.0	EUR	20		4	-4.01198387145996	31.079714884242
http://www.booking.com/hotel/ma/sabaku-tour-camp.html http://aff.bstatic.com/images/hotel/max500/790/79026941.jpg							

Offering free WiFi and a restaurant, Sabaku Tour Camp is situated in Merzouga, 2.1 km from Camel Trekking Company. Flamingo Birdwatching is 4 km from the property. Lake Daya Tamezguidat/Flamingo is 4.4 km from Sabaku Tour Camp.

Le Sabaku Tour Camp est situé à Merzouga, à 45 km d'Erfoad. Il propose un restaurant et une connexion Wi-Fi gratuite. Les serviettes sont fournies.

merzouga	Merzouga	5	1	1
----------	----------	---	---	---

El sistema continúa con el parseo de los ficheros para el resto de continentes (se han omitido las trazas con los registros erróneos):

Read 50009 hotels in Asia, 49981 succesfully parsed, 28 with wrong format

Read 50046 hotels in Asia2, 49974 succesfully parsed, 72 with wrong format

Read 32116 hotels in Asia3, 32106 succesfully parsed, 10 with wrong format

Read 6087 hotels in Caribbean, 6083 succesfully parsed, 4 with wrong format

Read 6062 hotels in Middle America, 6060 succesfully parsed, 2 with wrong format

Read 6989 hotels in Middle East, 6986 succesfully parsed, 3 with wrong format

Read 50000 hotels in South America, 49979 succesfully parsed, 21 with wrong format

Read 8882 hotels in South America 2, 8881 succesfully parsed, 1 with wrong format

Read 50001 hotels in North America, 49998 succesfully parsed, 3 with wrong format

Read 50000 hotels in North America 2, 50000 succesfully parsed, 0 with wrong format

Read 5438 hotels in North America 3, 5438 succesfully parsed, 0 with wrong format

Read 50001 hotels in Europe, 49998 succesfully parsed, 3 with wrong format

Read 50000 hotels in Europe 2, 49998 succesfully parsed, 2 with wrong format  
Read 50000 hotels in Europe 3, 49988 succesfully parsed, 12 with wrong format  
Read 50000 hotels in Europe 4, 50000 succesfully parsed, 0 with wrong format  
Read 50000 hotels in Europe 5, 49999 succesfully parsed, 1 with wrong format  
Read 50000 hotels in Europe 6, 50000 succesfully parsed, 0 with wrong format  
Read 50000 hotels in Europe 7, 49999 succesfully parsed, 1 with wrong format  
Read 50000 hotels in Europe 8, 49998 succesfully parsed, 2 with wrong format  
Read 50000 hotels in Europe 9, 49999 succesfully parsed, 1 with wrong format  
Read 50000 hotels in Europe 10, 49999 succesfully parsed, 1 with wrong format  
Read 50000 hotels in Europe 11, 49999 succesfully parsed, 1 with wrong format  
Read 50000 hotels in Europe 12, 50000 succesfully parsed, 0 with wrong format  
Read 11654 hotels in Europe 13, 11654 succesfully parsed, 0 with wrong format  
Total Hotels: 971478

Como resultado de este procesamiento, se ha creado la colección “hotels” en la base de datos de MongoDB con la información obtenida del parseo de los ficheros. El tiempo aproximado de la ejecución de este procesamiento es de alrededor de 15 minutos.

3. Arrancar el conector de MongoDB para conectar la información generada en MongoDB al motor de búsqueda de Elastic Search:

```
mongo-connector -m localhost:27017 -t localhost:9200 -d elastic_doc_manager -n HotelAdvisor.pois
```

4. Lanzar el job de Spark, para el parseo de los ficheros del dataset de pois:

```
~/spark-1.6.2-bin-hadoop2.6$ bin/spark-submit --name "HotelAdvisor" --master local[4] --conf "spark.driver.memory=6g"  
--conf "spark.executor.heartbeatInterval=7200s" --conf "spark.network.timeout=7200s" --driver-memory 6g  
/home/<user>/PycharmProjects/HotelAdvisor/PoiReader.py
```

Aparecen trazas por consola indicando el número de pois parseados con éxito y cuáles tenían un formato incorrecto:

Read 223763 pois in Central America, 223762 succesfully parsed, 1 with wrong format

Pois Central America with wrong format: 1

Dirección: Edif. Águila y Dragones, 3er. piso.AF

El sistema continúa con el parseo de los ficheros para el resto de continentes (se han omitido las trazas con los registros erróneos):

Read 578302 pois in Africa, 578299 succesfully parsed, 3 with wrong format

Read 2695943 pois in Asia, 2695898 succesfully parsed, 45 with wrong format

Read 7284202 pois in Europe, 7284157 succesfully parsed, 45 with wrong format

Read 2221582 pois in North America, 2221572 succesfully parsed, 10 with wrong format

Read 233111 pois in Oceania, 233111 succesfully parsed, 0 with wrong format

Read 780616 pois in South America, 780615 succesfully parsed, 1 with wrong format



Total POIs: 14.017.414

Como resultado de este procesamiento, se ha creado la colección “pois” en la base de datos de MongoDB con la información obtenida del parseo de los ficheros. El tiempo aproximado de la ejecución de este procesamiento es de alrededor de 100 minutos (1h 40’).

## 5. PROCESAMIENTO REAL TIME (REAL TIME PROCESSING)

1. Configurar el simulador de reviews (críticas) de hoteles. Copiar el contenido de la carpeta “Hotel Advisor - Simulador Reviews” en la carpeta “conf” del directorio de instalación de la utilidad Json Data Generator: /opt/json-data-generator-1.2.1/conf

HotelReviewConfig.json:

```
{
  "workflows": [{
    "workflowName": "HotelReview",
    "workflowFilename": "HotelReviewWorkflow.json"
  }],
  "producers": [{
    "type": "kafka",
    "broker.server": "localhost",
    "broker.port": 9092,
    "topic": "hotel_reviews",
    "flatten": false,
    "sync": false
  }]
}
```

HotelReviewWorkflow.json, genera eventos con una determinada frecuencia, asignando notas y comentarios a hoteles escogidos de forma aleatoria:

```
{
  "eventFrequency": 100,
  "varyEventFrequency": true,
  "repeatWorkflow": true,
  "timeBetweenRepeat": 50,
  "varyRepeatFrequency": true,
  "steps": [{
    "config": [{
      "date": "nowTimestamp()",
      "hotel_id": "integer(1,971478)",
      "age": "integer(18,90)",
      "username": "admin",
      "location": "double(3.0,10.0)",
      "cleanliness": "double(3.0,10.0)",
      "comfort": "double(3.0,10.0)",
      "installations": "double(3.0,10.0)",
      "personal": "double(3.0,10.0)",
      "wifi": "double(3.0,10.0)",
      "valueMoney": "double(3.0,10.0)",
      "lang": "es",
      "comment": "random('Hotel encantador. Personal muy amable. Baño amplio y cama cómoda. Barrio Tranquilo','Buen hotel. Algo de ruido en las habitaciones. Baño bueno. Desayuno normal. Personal de recepción muy amable y resolvió todas nuestras dudas','Hotel normal. Buena cama pero baño pequeño. Recepcionista diligente.','Hotel mediocre. Cama sucia. Baño pequeño. El barrio da miedo. El tipo de la recepción era muy borde','No es centrico pero hay dos paradas muy cerca que hace que puedas moverte muy rapido a las zonas de interes','El Hotel está muy bien, un poco caro pero cumple con lo pagado. El desayuno buenísimo aunque también algo caro. La única pega es que está alejado del centro y hay que ir en tranvía siempre a la ciudad. En las inmediaciones del Hotel no hay ningún servicio. Tiene parking propio. Si tuviera que volver no elegiría este Hotel por la lejanía del centro. Por el resto el Hotel es fantástico aunque demasiado lujoso para mis intereses','Moderno. Jovial. Limpio. Excelente instalaciones a pasos de un tram que te lleva al centro en unos minutos! El mejor precio y lo mejor que se ofrece! volveria definitivamente')"
```

```

        },
        "duration": 0
    }]
}

```

2. Crear el tópic de Kafka "hotel\_reviews", dónde se van a recibir los eventos con las reviews de los hoteles

```

~kafka_2.11-0.10.1.0$ bin/kafka-topics.sh --zookeeper localhost:2181 --create --topic hotel_reviews --
partitions 1 --replication-factor 1

```

3. Arrancar Zookeeper:

```

~kafka_2.11-0.10.1.0$ bin/zookeeper-server-start.sh config/zookeeper.properties

```

4. Arrancar Kafka:

```

~kafka_2.11-0.10.1.0$ bin/kafka-server-start.sh config/server.properties

```

5. Arrancar el conector de MongoDB para conectar la información generada en MongoDB al motor de búsqueda de Elastic Search:

```

mongo-connector -m localhost:27017 -t localhost:9200 -d elastic_doc_manager -n HotelAdvisor.hotels

```

6. Arrancar job de Spark Streaming:

```

~/spark-1.6.2-bin-hadoop2.6$ bin/spark-submit --name "Hotel Reviews Streaming" --jars lib/spark-streaming-kafka-
assembly_2.10-1.6.2.jar /home/<user>/PycharmProjects/HotelAdvisor/KafkaHotelReviewStreaming.py

```

7. Arrancar simulador de reviews:

```

~/json-data-generator-1.2.1$ java -jar json-data-generator-1.2.1.jar HotelReviewConfig.json

```

El sistema irá generando críticas (puntuaciones y comentario) sobre los hoteles de forma aleatoria y se irán almacenando tanto en la base de datos de MongoDB como en el motor de búsqueda de Elastic Search.

## 6. EXPLOTACIÓN DE LA INFORMACIÓN

### 6.1 EJECUCIÓN DE LA APLICACIÓN

1. Copiar la carpeta “HotelAdvisor - Explotacion Java\home\fer\IdeaProjects\HotelAdvisor” de la entrega a la carpeta “/home/<user>/IdeaProjects”

2. Compilar el proyecto mediante Maven:

```
~/IdeaProjects/HotelAdvisor$ mvn clean package
```

3. Ejecutar el fichero:

```
~/IdeaProjects/HotelAdvisor$ HotelAdvisor.sh
```

Esto ejecutará la aplicación (.jar) generada en la carpeta /target

4. Usar la aplicación, siguiendo el Manual de Usuario, ubicado en: “Hotel Advisor – Documentacion\Hotel Advisor - Manual de Usuario.odt” de la entrega.

### 6.2 DASHBOARDS / ANALYTICS

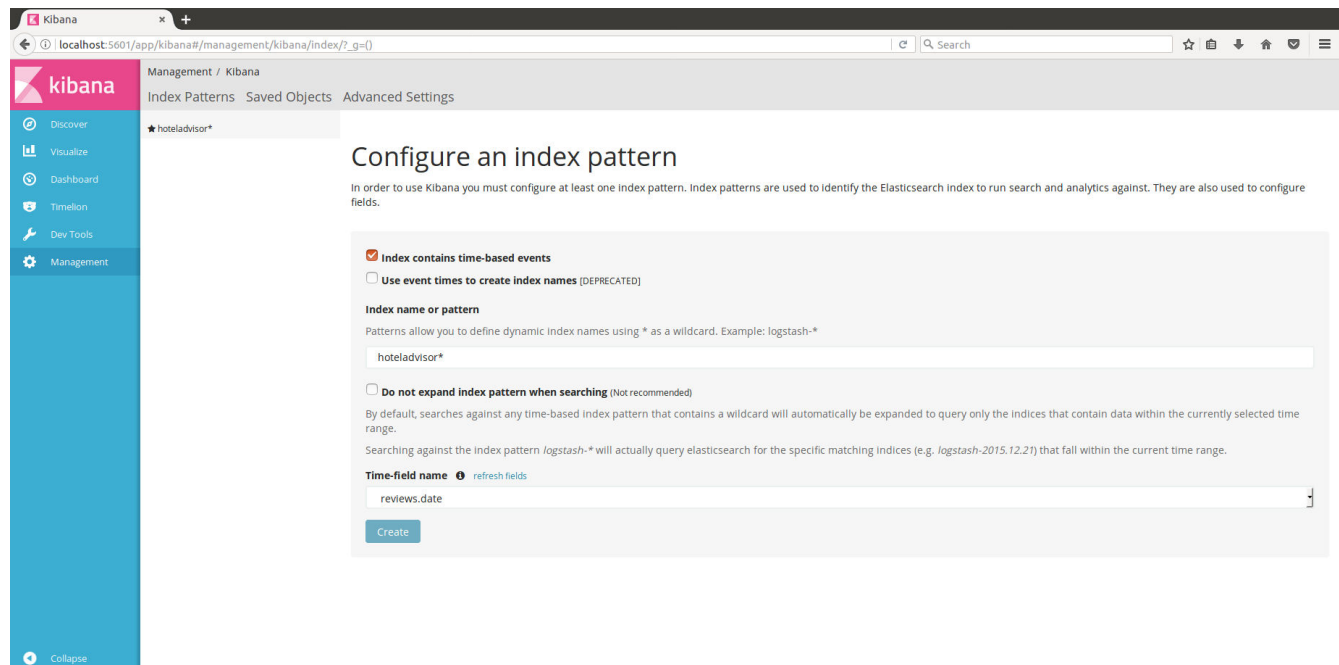
1. Arrancar Kibana (debe estar arrancado previamente Elastic Search):

```
~/kibana-5.2.2-linux-x86_64/bin$ ./kibana
```

2. Abrir la siguiente URL en un navegador, para tener acceso a la consola de Kibana:

<http://localhost:5601>

3. En Management>Kibana>Indexes, dar de alta el patrón: hoteladvisor\*:



4. Usar la consola de Kibana para crear los gráficos y dashboards deseados, siguiendo el Apartado 11 “Hotel Advisor – Dashboards (Kibana)” del Manual de Usuario, ubicado en: “Hotel Advisor – Documentacion\Hotel Advisor - Manual de Usuario.odt” de la entrega.