

## Лабораторная работа 6.

**Запуск HBase, создание таблицы, заполнение ее данными, разработка фильтров.**

### 1. Предварительные условия

На компьютере должен быть установлен и запущен hdfs hadoop версии 2.7.3.  
(См. лабораторную работу 1)

### 2. Настройка HBase.

Распаковываем hbase-1.2.3-bin.tar.gz в ~/hbase-1.2.3

Проверяем содержимое файла /etc/hosts  
и выставляем ip 127.0.0.1 для localhost и имени компьютера

Снимаем комментарий в файле conf/hbase-env.sh со строки  
export HBASE\_MANAGES\_ZK=true

Настраиваем файл конфигурации conf/hbase-site.xml следующим образом

```
<configuration>
  <property>
    <name>hbase.rootdir</name>
    <value>hdfs://localhost:9000/hbase</value>
  </property>
  <property>
    <name>hbase.cluster.distributed</name>
    <value>true</value>
  </property>
</configuration>
```

Добавляем в файл bin/hbase-config.sh первую строку

```
export JAVA_HOME=~/.jdk1.8.0_101
```

Добавляем в ~/.bashrc строку

```
export PATH=~/.hbase-1.2.3/bin:$PATH
```

### 3. Запуск HBase

Запуск HBase осуществляется скриптом /bin/start-hbase.sh

Остановка HBase осуществляется скриптом /bin/stop-hbase.sh

### 4. Консоль HBase

Консоль HBase вызывается скриптом /bin/hbase shell

Примеры команд консоли :

```
create 'testtable', 'colfam1','colfam2'
```

```
put 'testtable', 'myrow-1', 'colfam1:q1', 'value-1'
```

```
scan 'testtable'
```

### 5. Входные данные.

Аналогичны входным данным лабораторных работ 3,4,5

### 6. Задача

а. Требуется создать таблицу flights имеющую все поля файла и заполнить ее данными из файла

В качестве ключа требуется использовать составной ключ из даты перелета(FL\_DATE) + номер рейса (AIRLINE\_ID) + номер строки в исходном файле.

б. Требуется разработать фильтр, отбирающий записи перелетов с опозданием больше заданного или отмененные.

в. Требуется разработать отдельный main класс который использует фильтр а также отбирает записи перелетов в заданном интервале дат используя ограничения на ключ для сканера (методы setStartRow, setStopRow).

### 3. Подсказки.

а. Таблицу можно создать из консоли HBase

б. Пример файла проекта maven с настроенными зависимостями (pom.xml) приложен

в. Запуск standalone приложения с помощью maven осуществляется следующим образом

```
mvn compile exec:java -Dexec.mainClass="<Main class>"
```

г. Базовый класс фильтра - org.apache.hadoop.hbase.filter.FilterBase

д. Надо обратить внимание на корректность реализации методов сериализации

```
public byte[] toByteArray()
```

```
public static Filter parseFrom(byte[] pbBytes) throws DeserializationException
```

е. Для того чтобы подключить свой класс фильтра к HBase требуется добавить jar в HBASE\_CLASSPATH и прописать его в ~/hbase-1.2.3/conf/hbase-env.sh

Пример :

```
export HBASE_CLASSPATH="<путь>/hbase-examples-1.0-SNAPSHOT.jar"
```