

Отчёт. Лабораторная работа №2.

Студент: Турко С.А.

Группа: M16-522

Задание

A) Program which calculate linux syslog priority (is given by 7 – debug, 6 – info, 5 – notice, 4 – warning, warn, 3 – err, error, 2 – crit, 1 – alert, 0 – emerg, panic) count by hours.

B) Ingest technology – Flume from file agent

C) Storage technology – Ignite Native Persistence

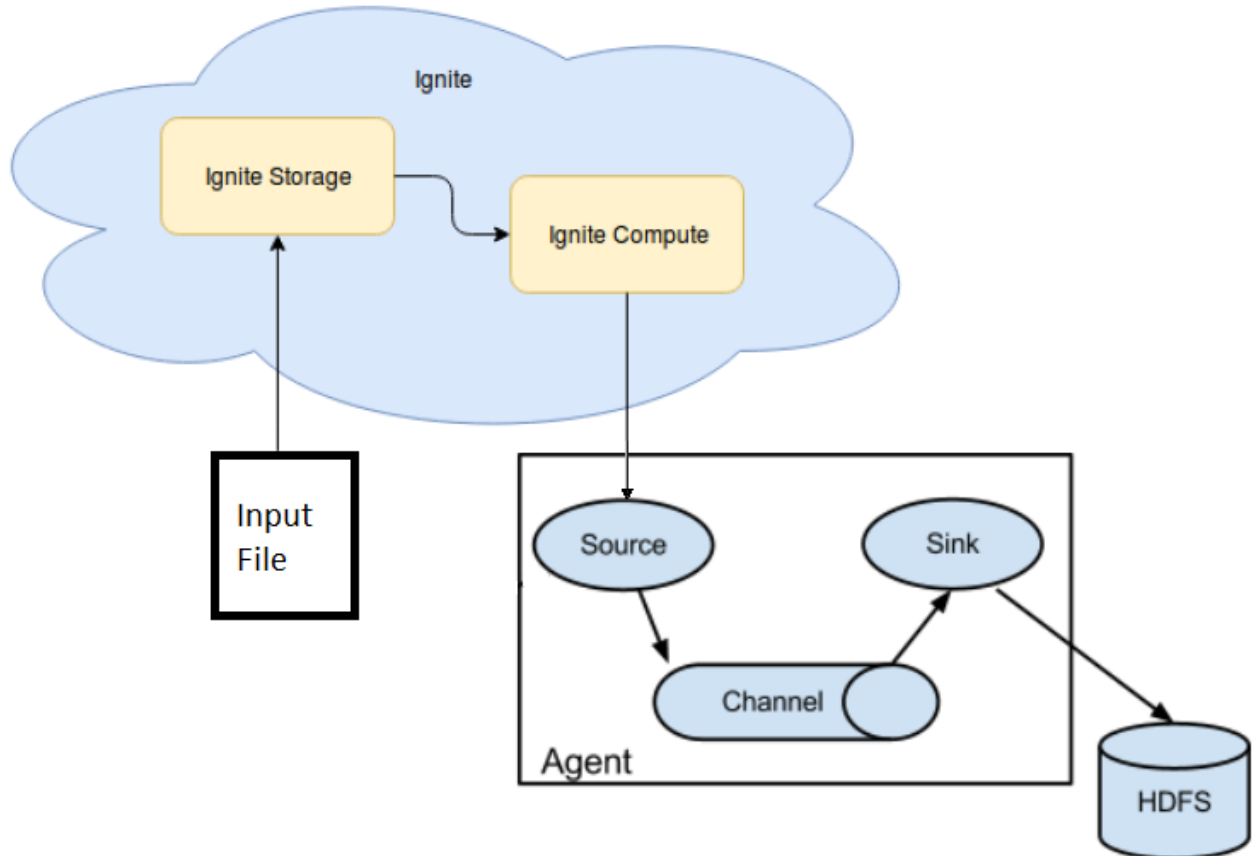
D) Computation technology – Ignite compute

Link: https://github.com/sergeyturko/DSBDA_HW2.git

Для запуска проекта необходимо установить:

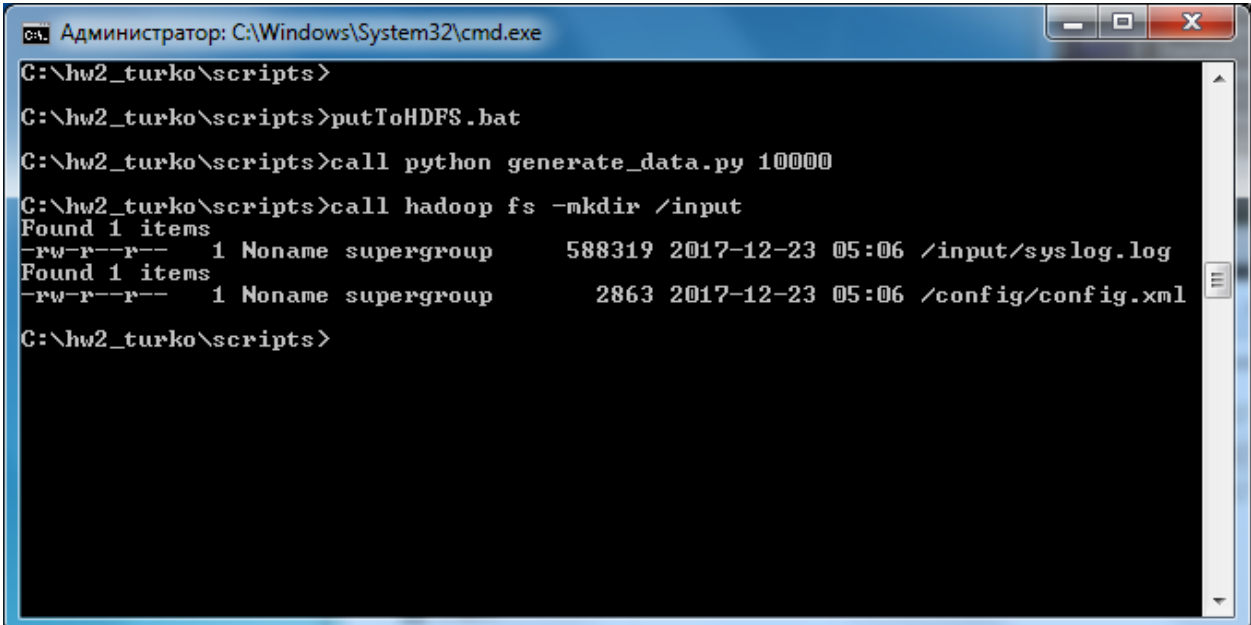
- Python 2.7;
- Java 1.8.0_152
- Apache Maven;
- Apache Hadoop 2.6.5;
- Apache Ignite 2.3.0;
- Apache Flume 1.8.0.

Схема взаимодействия:



Подготовка данных:

- Запустить Hadoop: *start-dfs.cmd*, *start-yarn.cmd*
- Запустить скрипт: *putToHDFS.bat*

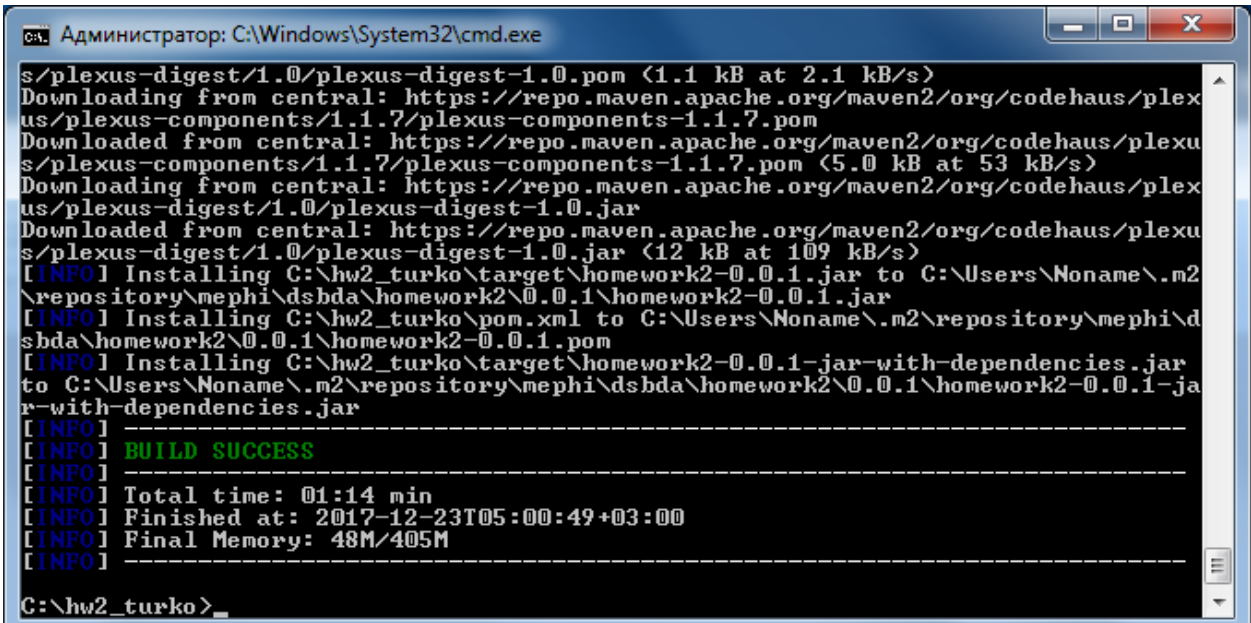


```
C:\Администратор: C:\Windows\System32\cmd.exe
C:\hw2_turko\scripts>
C:\hw2_turko\scripts>putToHDFS.bat
C:\hw2_turko\scripts>call python generate_data.py 10000
C:\hw2_turko\scripts>call hadoop fs -mkdir /input
Found 1 items
-rw-r--r-- 1 Noname supergroup      588319 2017-12-23 05:06 /input/syslog.log
Found 1 items
-rw-r--r-- 1 Noname supergroup       2863 2017-12-23 05:06 /config/config.xml
C:\hw2_turko\scripts>
```

Рисунок 1 – Копирование данных в HDFS.

Сборка проекта:

- Запустить Ignite: *ignite.bat config.xml*
- *mvn clean install*



```
C:\Администратор: C:\Windows\System32\cmd.exe
s/plexus-digest/1.0/plexus-digest-1.0.pom <1.1 kB at 2.1 kB/s>
Downloading from central: https://repo.maven.apache.org/maven2/org/codehaus/plexu
us/plexus-components/1.1.7/plexus-components-1.1.7.pom
Downloaded from central: https://repo.maven.apache.org/maven2/org/codehaus/plexu
s/plexus-components/1.1.7/plexus-components-1.1.7.pom <5.0 kB at 53 kB/s>
Downloading from central: https://repo.maven.apache.org/maven2/org/codehaus/plexu
us/plexus-digest/1.0/plexus-digest-1.0.jar
Downloaded from central: https://repo.maven.apache.org/maven2/org/codehaus/plexu
s/plexus-digest/1.0/plexus-digest-1.0.jar <12 kB at 109 kB/s>
[INFO] Installing C:\hw2_turko\target\homework2-0.0.1.jar to C:\Users\Noname\.m2
\repository\mephi\dsbda\homework2\0.0.1\homework2-0.0.1.jar
[INFO] Installing C:\hw2_turko\target\homework2-0.0.1-jar-with-dependencies.jar
to C:\Users\Noname\.m2\repository\mephi\dsbda\homework2\0.0.1\homework2-0.0.1-ja
r-with-dependencies.jar
[INFO] -----
[INFO] BUILD SUCCESS
[INFO] -----
[INFO] Total time: 01:14 min
[INFO] Finished at: 2017-12-23T05:00:49+03:00
[INFO] Final Memory: 48M/405M
[INFO] -----
C:\hw2_turko>
```

Рисунок 2 – Сборка, выполнение тестов.

Запуск:

- Запустить Flume: *flume-ng agent --conf conf --conf-file flume.conf --name hw2*
- *hadoop jar homework2-0.0.1-jar-with-dependencies.jar homework2 storage /config/config.xml /input/syslog.log*
- *hadoop jar homework2-0.0.1-jar-with-dependencies.jar homework2 compute/config/config.xml*