Отчёт. Лабораторная работа №2.

Студент: Турко С.А.
Группа: М16-522

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.
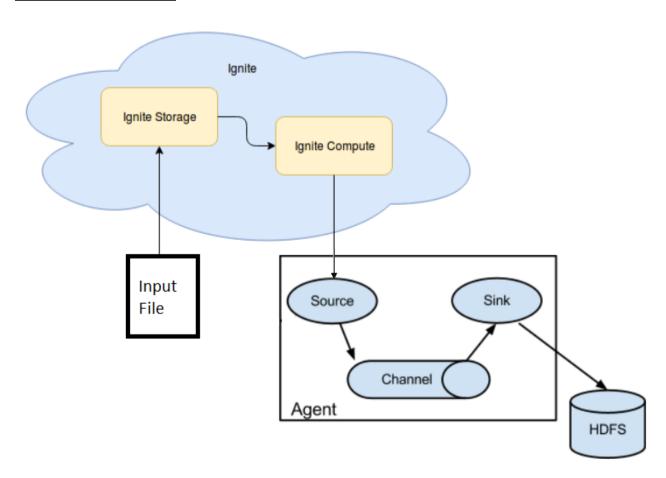B) Ingest technology - Flume from file agent
C) Storage technology - Ignite Native Persistence
D) Computation technology - Ignite compute

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: *python generate_data.py 10000*

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

Рисунок 1 – Сборка, выполнение тестов.

Запуск:

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

Рисунок 2 – Storage.

```
priority - 4 hour - 17 count - 61
priority - 4 hour - 18 count - 46
priority - 4 hour - 19 count - 61
priority - 4 hour - 20 count - 59
priority - 4 hour - 21 count - 56
priority - 4 hour - 22 count - 63
priority - 4 hour - 23 count - 60
priority - 7 hour - 0 count - 63
priority - 7 hour - 1 count - 55
priority - 7 hour - 2 count - 51
priority - 7 hour - 3 count - 63
priority - 7 hour - 4 count - 54
priority - 7 hour - 5 count - 55
priority - 7 hour - 6 count - 62
priority - 7 hour - 7 count - 55
priority - 7 hour - 8 count - 63
priority - 7 hour - 9 count - 55
priority - 7 hour - 10 count - 63
priority - 7 hour - 11 count - 53
priority - 7 hour - 12 count - 54
priority - 2 hour - 0 count - 43
priority - 7 hour - 13 count - 51
priority - 2 hour - 1 count - 47
priority - 2 hour - 2 count - 55
priority - 7 hour - 14 count - 68
priority - 2 hour - 3 count - 60
priority - 7 hour - 15 count - 66
priority - 7 hour - 16 count - 67
priority - 2 hour - 4 count - 61
priority - 7 hour - 17 count - 61
priority - 2 hour - 5 count - 59
priority - 7 hour - 18 count - 76
priority - 2 hour - 6 count - 44
priority - 7 hour - 19 count - 60
priority - 2 hour - 7 count - 51
priority - 7 hour - 20 count - 58
priority - 2 hour - 8 count - 57
priority - 2 hour - 9 count - 68
priority - 7 hour - 21 count - 59
priority - 7 hour - 22 count - 60
priority - 2 hour - 10 count - 58
priority - 2 hour - 11 count - 50
priority - 7 hour - 23 count - 60
priority - 2 hour - 12 count - 60
priority - 2 hour - 13 count - 50
priority - 2 hour - 14 count - 44
priority - 2 hour - 15 count - 57
priority - 2 hour - 16 count - 58
priority - 2 hour - 17 count - 58
priority - 2 hour - 18 count - 58
priority - 2 hour - 19 count - 66
priority - 2 hour - 20 count - 58
priority - 2 hour - 21 count - 50
priority - 2 hour - 22 count - 64
priority - 2 hour - 23 count - 67
priority - 5 hour - 0 count - 71
priority - 5 hour - 1 count - 58
priority - 5 hour - 2 count - 57
priority - 5 hour - 3 count - 51
priority - 5 hour - 4 count - 45
priority - 5 hour - 5 count - 57
priority - 5 hour - 6 count - 73
priority - 5 hour - 7 count - 61
priority - 5 hour - 8 count - 52
priority - 5 hour - 9 count - 55
priority - 5 hour - 10 count - 61
priority - 5 hour - 11 count - 53
Computing finished successfully!
```

Рисунок 3 – Compute.