

**Министерство науки и высшего образования РФ**  
Федеральное государственное автономное образовательное  
учреждение высшего образования «Национальный  
исследовательский ядерный университет «МИФИ»

Институт ИИКС  
Кафедра компьютерных систем и технологий

**Лабораторная работа №2:**

«Технологии обработки больших данных»

по дисциплине

«Наука о данных и анализ больших данных»

Выполнил:  
студент гр. М21-502  
Корнилов А. Н.

Проверил:  
Синельников Дмитрий Михайлович

Москва, 2022 г.

## Оглавление

Программные средства.....	3
Тесты.....	3
Выполненные действия .....	3
Результаты.....	6

## Программные средства

Разработка велась с использованием Apache Ignite и Apache Kafka на CentOS. Для сборки использовался maven плагин для IntelliJ IDEA.

## Тесты

Результат успешного прохождения тестов представлен на рисунке 2.

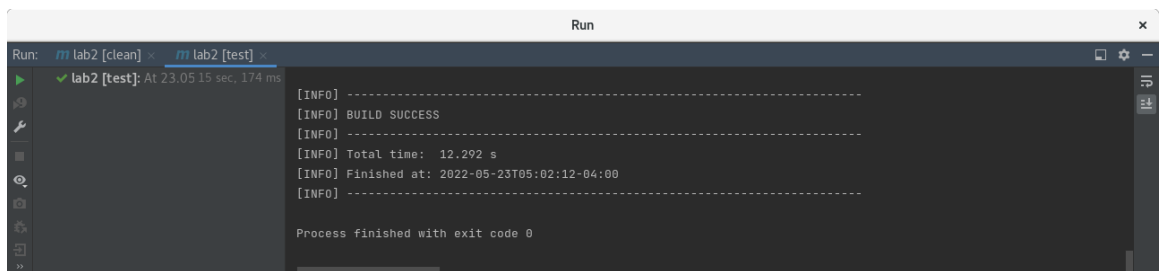


Рисунок 2 – скриншот результата успешного прохождения тестов

## Выполненные действия

Последовательность выполнения:

1. Запуск Ignite (см. рисунок 1).

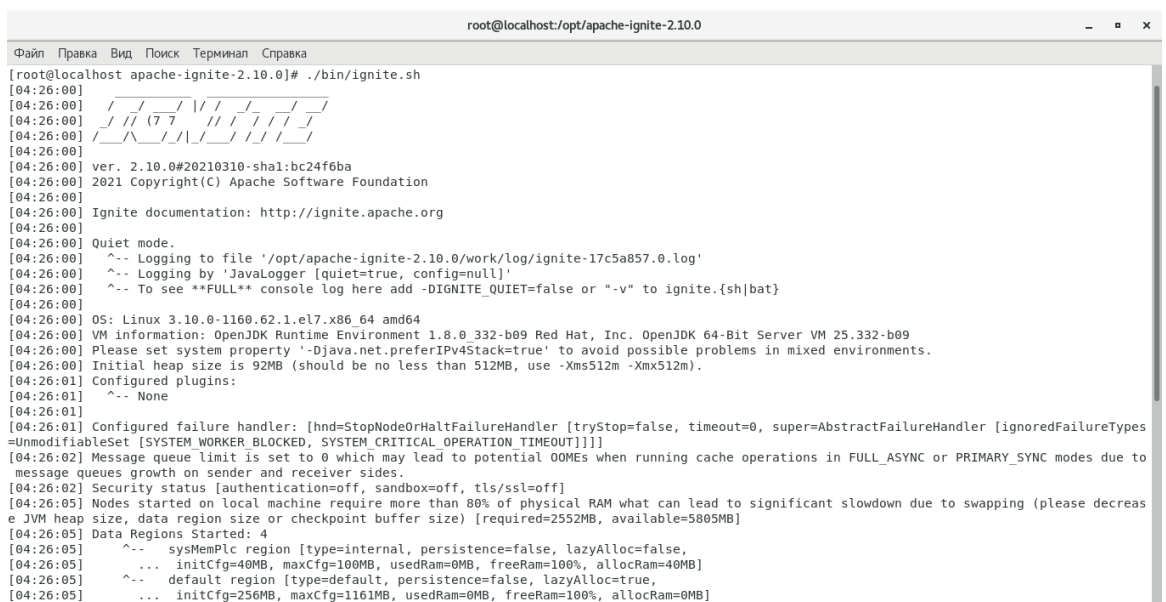


Рисунок 1 – запуск скрипта Ignite.sh

```
root@localhost:/opt/kafka-2.12-2.8.1
```

Файл Правка Вид Поиск Терминал Вкладки Справка

```
root@localhost:/opt/kafka-2.12-2.8.1 x root@localhost:/opt/kafka-2.12-2.8.1 x [icon] v
```

```
[root@localhost kafka 2.12-2.8.1]# bin/zookeeper-server-start.sh config/zookeeper.properties & kafka-2.12-2.8.1]# bin/zookeeper-server-start.sh conf
[2022-05-23 04:26:43,536] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.QuorumPeerConfig)
[2022-05-23 04:26:43,645] WARN config/zookeeper.properties is relative. Prepend . to indicate that you're sure! (org.apache.zookeeper.server.Quorum
QuorumPeerConfig)
[2022-05-23 04:26:43,690] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.QuorumPeerConfig)
[2022-05-23 04:26:43,690] INFO secureClientPort is not set (org.apache.zookeeper.server.QuorumPeerConfig)
[2022-05-23 04:26:43,694] INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DataDirCleanupManager)
[2022-05-23 04:26:43,694] INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DataDirCleanupManager)
[2022-05-23 04:26:43,694] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DataDirCleanupManager)
[2022-05-23 04:26:43,694] WARN Either no config or no quorum defined in config, running in standalone mode (org.apache.zookeeper.server.QuorumPeerMain)
[2022-05-23 04:26:43,702] INFO Log4j 1.2 jmx support found and enabled. (org.apache.zookeeper.jmx.ManagedUtil)
[2022-05-23 04:26:43,764] INFO Reading configuration from: config/zookeeper.properties (org.apache.zookeeper.server.QuorumPeerConfig)
[2022-05-23 04:26:43,754] WARN config/zookeeper.properties is relative. Prepend . to indicate that you're sure! (org.apache.zookeeper.server.Quorum
QuorumPeerConfig)
[2022-05-23 04:26:43,755] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.QuorumPeerConfig)
[2022-05-23 04:26:43,755] INFO secureClientPort is not set (org.apache.zookeeper.server.QuorumPeerConfig)
[2022-05-23 04:26:43,755] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
[2022-05-23 04:26:43,760] INFO Zookeeper.snapshot.trust.empty : false (org.apache.zookeeper.server.persistence.FileTxnSnapLog)
[2022-05-23 04:26:43,777] INFO Server environment:zookeeper.version=3.5.9-83df9301aa5c2a5d284a9940177808c01bc3cef, built on 01/06/2021 20:03 GMT (or
g.apache.zookeeper.server.ZooKeeperServer)
[2022-05-23 04:26:43,777] INFO Server environment:host.name=localhost (org.apache.zookeeper.server.ZooKeeperServer)
[2022-05-23 04:26:43,777] INFO Server environment:java.vendor=1.8.0_332 (org.apache.zookeeper.server.ZooKeeperServer)
[2022-05-23 04:26:43,777] INFO Server environment:java.version=Red Hat, Inc. (org.apache.zookeeper.server.ZooKeeperServer)
[2022-05-23 04:26:43,777] INFO Server environment:java.home=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0_332.b09-1.el7.x86_64/jre (org.apache.zookeeper.se
rver.ZooKeeperServer)
[2022-05-23 04:26:43,778] INFO Server environment:java.class.path=/lib/opt/kafka-2.12-2.8.1/bin/./:/lib/activation-1.1.1.jar:/opt/kafka-2.12-2.8.1/bin/./
./lib/catalina-repackaged-2.6.1.jar:/opt/kafka-2.12-2.8.1/bin/./lib/jargparse-0.9.jar:/opt/kafka-2.12-2.8.1/bin/./lib/audience-annotations
-0.5.0.jar:/opt/kafka-2.12-2.8.1/bin/./lib/commons-cli-1.4.jar:/opt/kafka-2.12-2.8.1/bin/./lib/commons-lang3-3.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./
./lib/connect-api-2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./lib/connect-basic-auth-extension-2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./lib/connect-file-
2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./lib/connect-json-2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./lib/connect-mirror-2.8.1.jar:/opt/kafka-2.12-2.8.1/b
in/./lib/connect-mirror-client-2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./lib/connect-runtime-2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./lib/connect-tran
sforms-2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./lib/hk2-api-2.6.1.jar:/opt/kafka-2.12-2.8.1/bin/./lib/hk2-locator-2.6.1.jar:/opt/kafka-2.12-2.8.1/b
```

```
root@localhost:/opt/kafka.2.12-2.8.1

Файл  Правка  Вид  Поиск  Терминал  Вкладки  Справка

root@localhost:/opt/kafka.2.12-2.8.1                                root@localhost:/opt/kafka.2.12-2.8.1

[root@localhost kafka-2.12-2.8.1]# bin/kafka-server-start.sh config/server.properties
[2022-05-23 04:26:44,464] INFO Registered kafka:type=kafka.Log4jControllerMBean (kafka.utils.Log4jControllerRegistrations)
[2022-05-23 04:26:46,164] INFO Setting -D jdk.tls.rejectClientInitiatedRenegotiation=true to disable client-initiated TLS renegotiation (org.apache.zookeeper.common.X509Util)
[2022-05-23 04:26:46,691] INFO Registered signal handlers for TERM, INT, HUP (org.apache.kafka.common.utils.LoggingSignalHandler)
[2022-05-23 04:26:46,699] INFO starting (kafka.server.KafkaServer)
[2022-05-23 04:26:46,760] INFO Connecting to zookeeper on localhost:2181 (kafka.server.KafkaServer)
[2022-05-23 04:26:46,761] INFO [ZooKeeperClient Kafka server] Initializing a new session to localhost:2181. (kafka.zookeeper.ZooKeeperClient)
[2022-05-23 04:26:46,768] INFO Client environment:zookeeper.version=3.5.9-83df9301ea5c2a5d284a9940177808c01bc35cef, built on 01/06/2021 20:03 GMT (org.apache.zookeeper.ZooKeeper)
[2022-05-23 04:26:46,768] INFO Client environment:host.name=localhost (org.apache.zookeeper.ZooKeeper)
[2022-05-23 04:26:46,768] INFO Client environment:java.vendor=Red Hat, Inc. (org.apache.zookeeper.ZooKeeper)
[2022-05-23 04:26:46,769] INFO Client environment:java.version=1.8.0_332 (org.apache.zookeeper.ZooKeeper)
[2022-05-23 04:26:46,769] INFO Client environment:java.home=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0_332.b09-1.el7_9.x86_64/jre (org.apache.zookeeper.ZooKeeper)
[2022-05-23 04:26:46,769] INFO Client environment:java.class.path=/opt/kafka-2.12-2.8.1/bin/./libs/activation-1.1.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/aopalliance-repackaged-2.6.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/argparse4j-0.7.0.jar:/opt/kafka-2.12-2.8.1/bin/./libs/audience-annotations-5.0.0.jar:/opt/kafka-2.12-2.8.1/bin/./libs/commons-cli-1.4.jar:/opt/kafka-2.12-2.8.1/bin/./libs/commons-lang3-3.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/connect-api-2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/connect-basic-auth-extension-2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/connect-file-2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/connect-jmx-2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/connect-mirror-2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/connect-mirror-client-2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/connect-runtime-2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/connect-transforms-2.8.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/hk2-api-2.6.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/hk2-locator-2.6.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/hk2-utils-2.6.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jackson-annotations-2.10.5.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jackson-core-2.10.5.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jackson-databind-2.10.5.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jackson-dataformat-csv-2.10.5.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jackson-datatype-jdk8-2.10.5.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jackson-jaxrs-base-2.10.5.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jackson-jaxrs-jjson-provider-2.10.5.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jackson-module-jaxb-annotations-2.10.5.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jackson-module-scala-2.12.2.10.5.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jakarta.activation-api-1.2.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jakarta.annotation-api-1.3.5.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jakarta.inject-2.6.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jakarta.validation-api-2.0.2.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jakarta.ws-rs-api-2.1.6.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jakarta.xml.bind-api-2.3.2.jar:/opt/kafka-2.12-2.8.1/bin/./libs/javassist-3.27.0-GA.jar:/opt/kafka-2.12-2.8.1/bin/./libs/javax-servlet-api-3.1.0.jar:/opt/kafka-2.12-2.8.1/bin/./libs/javax.ws-rs-api-2.1.1.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jaxb-api-2.3.0.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jersey-client-2.34.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jersey-common-2.34.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jersey-container-servlet-2.34.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jersey-container-servlet-core-2.34.jar:/opt/kafka-2.12-2.8.1/bin/./libs/jersey-hk2-2.34.jar:/opt
```

4

- 4). Программа попросит ввести число количество сгенерированных строк)

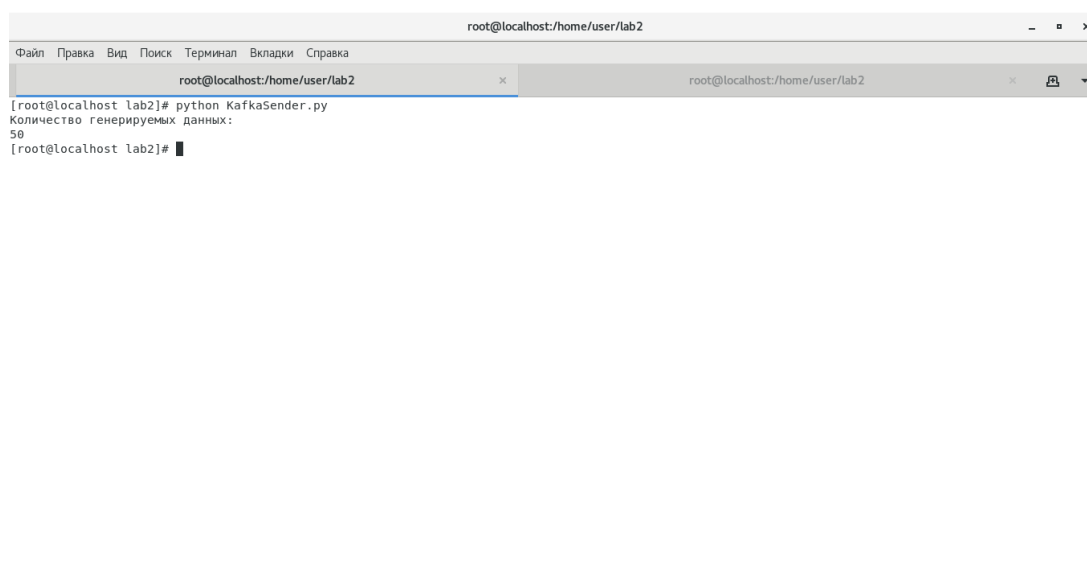


Рисунок 4 – запуск скрипта KafkaSender.py

- #### 4. Build-программы (см. рисунок 5)

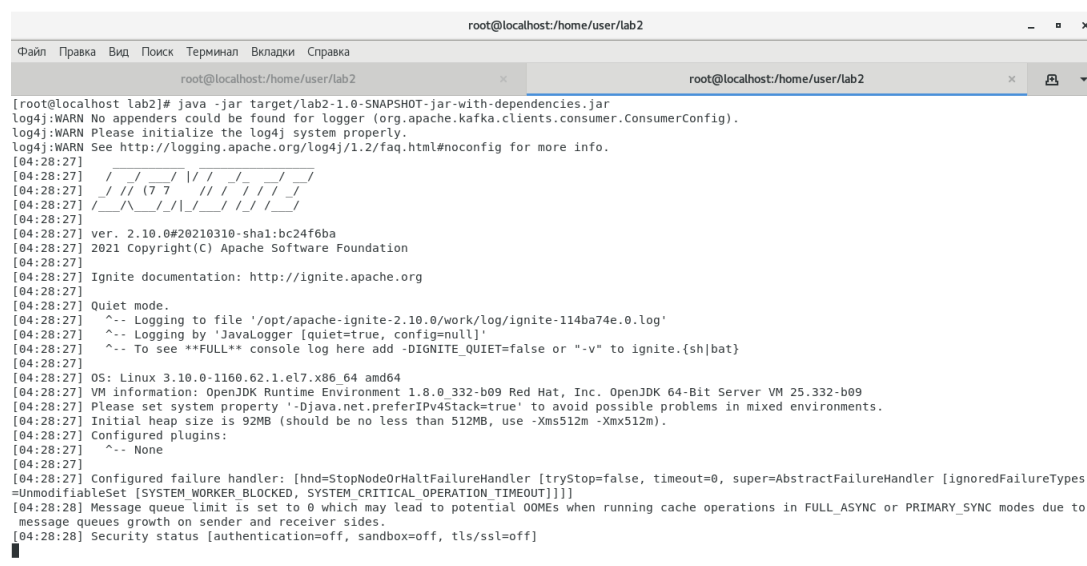
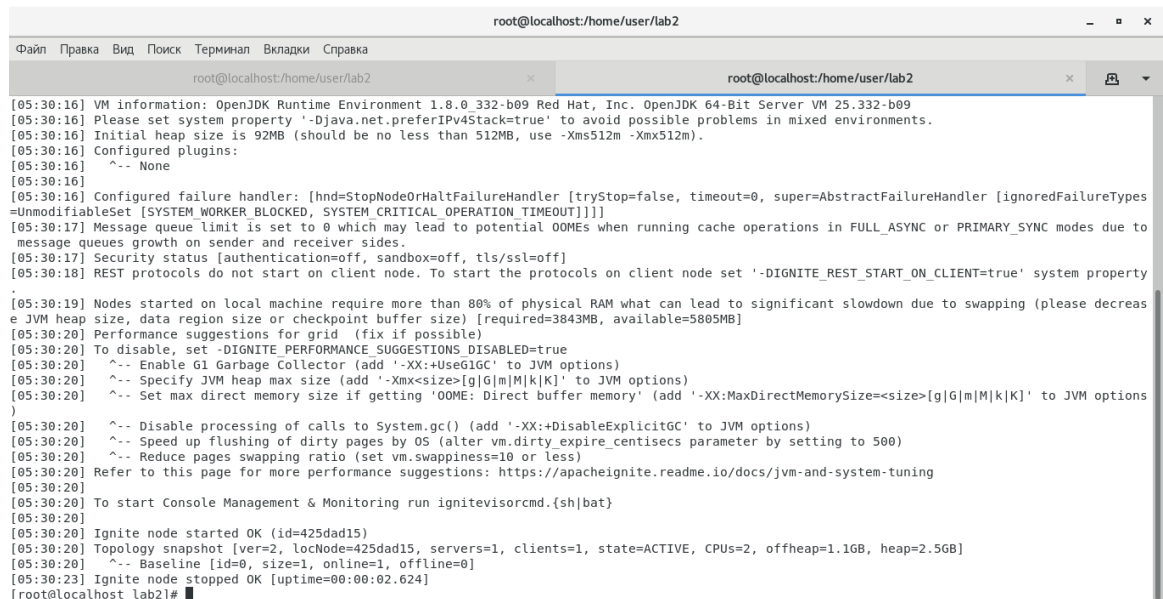


Рисунок 5 – запуск jar с зависимостями

## Результаты

Результат успешного выполнения программы представлен на рисунке 8.



```
root@localhost:/home/user/lab2
[05:30:16] VM information: OpenJDK Runtime Environment 1.8.0_332-b09 Red Hat, Inc. OpenJDK 64-Bit Server VM 25.332-b09
[05:30:16] Please set system property '-Djava.net.preferIPv4Stack=true' to avoid possible problems in mixed environments.
[05:30:16] Initial heap size is 92MB (should be no less than 512MB, use -Xms512m -Xmx512m).
[05:30:16] Configured plugins:
[05:30:16] ^-- None
[05:30:16] Configured failure handler: [hnd=StopNodeOrHaltFailureHandler [tryStop=false, timeout=0, super=AbstractFailureHandler [ignoredFailureTypes=UnmodifiableSet [SYSTEM_WORKER_BLOCKED, SYSTEM_CRITICAL_OPERATION_TIMEOUT]]]]
[05:30:17] Message queue limit is set to 0 which may lead to potential OOMs when running cache operations in FULL_ASYNC or PRIMARY_SYNC modes due to message queues growth on sender and receiver sides.
[05:30:17] Security status [authentication=off, sandbox=off, tls/ssl=off]
[05:30:18] REST protocols do not start on client node. To start the protocols on client node set '-DIGNITE_REST_START_ON_CLIENT=true' system property
[05:30:19] Nodes started on local machine require more than 80% of physical RAM what can lead to significant slowdown due to swapping (please decrease JVM heap size, data region size or checkpoint buffer size) [required=3843MB, available=5805MB]
[05:30:20] Performance suggestions for grid (fix if possible)
[05:30:20] To disable, set -DIGNITE_PERFORMANCE_SUGGESTIONS_DISABLED=true
[05:30:20] ^-- Enable G1 Garbage Collector (add '-XX:+UseG1GC' to JVM options)
[05:30:20] ^-- Specify JVM heap max size (add '-Xmx<size>[g|m|k|K]' to JVM options)
[05:30:20] ^-- Set max direct memory size if getting 'OOM: Direct buffer memory' (add '-XX:MaxDirectMemorySize=<size>[g|m|k|K]' to JVM options)
[05:30:20] ^-- Disable processing of calls to System.gc() (add '-XX:+DisableExplicitGC' to JVM options)
[05:30:20] ^-- Speed up flushing of dirty pages by OS (alter vm.dirty_expire_centisecs parameter by setting to 500)
[05:30:20] ^-- Reduce pages swapping ratio (set vm.swappiness=10 or less)
[05:30:20] Refer to this page for more performance suggestions: https://apacheignite.readme.io/docs/jvm-and-system-tuning
[05:30:20] To start Console Management & Monitoring run ignitevisorcmd.{sh|bat}
[05:30:20] Ignite node started OK (id=425dad15)
[05:30:20] Topology snapshot [ver=2, locNode=425dad15, servers=1, clients=1, state=ACTIVE, CPUs=2, offheap=1.1GB, heap=2.5GB]
[05:30:20] ^-- Baseline [id=0, size=1, online=1, offline=0]
[05:30:23] Ignite node stopped OK [uptime=00:00:02.624]
root@localhost lab2)#
```

Рисунок 8 – скриншот успешного выполнения программы

Содержимое выходного файла представлено на рисунке 9.



```
2022-08-31 20:00:00, Кипр, Испания } - 1
2022-03-29 14:00:00, Австрия, Болгария } - 1
2022-03-21 19:00:00, Германия, Венгрия } - 1
2022-02-04 07:00:00, Кипр, Испания } - 1
2022-03-01 13:00:00, Польша, Россия } - 1
2022-03-03 01:00:00, Испания, Греция } - 1
2022-01-31 14:00:00, Нидерланды, Россия } - 1
2022-07-03 01:00:00, Польша, Великобритания } - 1
2022-01-31 10:00:00, Литва, Великобритания } - 1
2022-10-21 15:00:00, Греция, Эстония } - 1
2022-01-28 19:00:00, Россия, Италия } - 1
2022-02-11 07:00:00, Хорватия, Греция } - 1
2022-03-21 06:00:00, Австрия, Испания } - 1
2022-12-24 01:00:00, Швеция, Россия } - 1
2022-09-15 21:00:00, Швейцария, Болгария } - 1
2022-05-25 15:00:00, Болгария, Великобритания } - 1
2022-06-25 09:00:00, Греция, Финляндия } - 1
2022-06-05 03:00:00, Швейцария, Великобритания } - 1
2022-10-18 21:00:00, Италия, Польша } - 1
2022-05-23 02:00:00, Болгария, Россия } - 1
2022-08-27 23:00:00, Сербия, Греция } - 1
2022-09-10 11:00:00, Италия, Литва } - 1
2022-07-31 11:00:00, Нидерланды, Польша } - 1
2022-11-07 12:00:00, Австрия, Польша } - 1
2022-02-20 22:00:00, Италия, Испания } - 1
2022-06-11 17:00:00, Великобритания, Польша } - 1
2022-04-08 23:00:00, Норвегия, Греция } - 1
2022-12-09 19:00:00, Финляндия, Австрия } - 1
2022-01-18 08:00:00, Германия, Франция } - 1
2022-10-19 13:00:00, Турция, Германия } - 1
2022-08-21 01:00:00, Сербия, Италия } - 1
2022-07-27 06:00:00, Германия, Турция } - 1
2022-07-28 13:00:00, Ирландия, Болгария } - 1
2022-09-09 23:00:00, Испания, Швеция } - 1
2022-11-01 03:00:00, Латвия, Болгария } - 1
```

Рисунок 9 – скриншот содержимого файла с результатами