

# Distributed gradient descent

Siemens internship task by Edem Ibrahimov

Проверка масштабируемости проводилась в standalone кластере поднятом локально на машине, характеристики которой изображены на последнем слайде.

На следующих слайдах приводятся результаты запуска с различным количеством worker'ов.

# Один worker



## Spark Master at spark://fr-Inspiron-5567:7077

URL: spark://fr-Inspiron-5567:7077

Alive Workers: 1

Cores in use: 1 Total, 0 Used

Memory in use: 1024.0 MB Total, 0.0 B Used

Applications: 0 Running, 1 Completed

Drivers: 0 Running, 0 Completed

Status: ALIVE

### Workers (1)

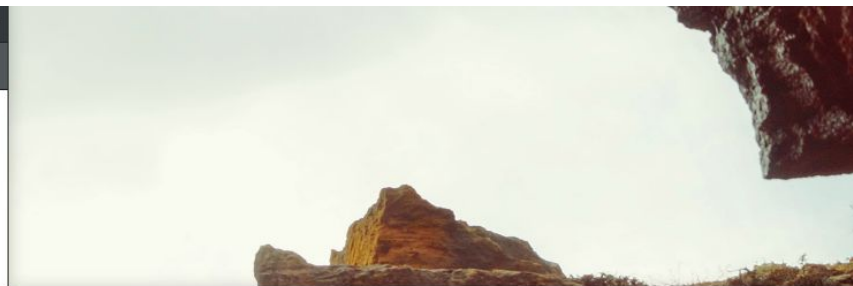
Worker Id	Address	State	Cores	Memory
worker-20200212131934-192.168.1.101-45095	192.168.1.101:45095	ALIVE	1 (0 Used)	1024.0 MB (0.0 B Used)

### Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	----------------	------	-------	----------

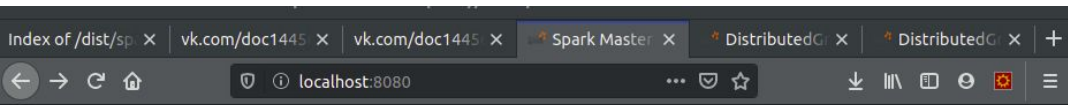
### Completed Applications (1)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
app-20200212144413-0000	DistributedGradientDescent	1	1024.0 MB	2020/02/12 14:44:13	fr	FINISHED	5,4 min



```
fr@fr-Inspiron-5567: ~  
Файл Правка Вид Поиск Терминал Справка  
point stopped!  
20/02/12 14:41:34 INFO MemoryStore: MemoryStore cleared  
20/02/12 14:41:34 INFO BlockManager: BlockManager stopped  
20/02/12 14:41:34 INFO BlockManagerMaster: BlockManagerMaster stopped  
20/02/12 14:41:34 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint:  
OutputCommitCoordinator stopped!  
20/02/12 14:41:34 INFO SparkContext: Successfully stopped SparkContext  
Trained weights:  
[1.1957390568811288, 5.49228808960139, 2.903208618008031, 12.172397039093628]  
20/02/12 14:41:34 INFO ShutdownHookManager: Shutdown hook called  
20/02/12 14:41:34 INFO ShutdownHookManager: Deleting directory /tmp/spark-01fcc0  
37-d3e5-48aa-ad3f-a260353b6ead  
20/02/12 14:41:34 INFO ShutdownHookManager: Deleting directory /tmp/spark-115c35  
8f-35bc-4d0a-af58-8bfcc68cecac  
fr@fr-Inspiron-5567:~$ jps  
10594 Master  
5020 Worker  
10671 Jps  
fr@fr-Inspiron-5567:~$ ~/spark-2.4.5-bin-hadoop2.7/bin/spark-submit --master spa  
rk://fr-Inspiron-5567:7077 siemens-1.0-SNAPSHOT.jar dataset.txt 0.0000000001 0.0  
0000000001 200 2>/dev/null  
8  
Trained weights:  
[2.957322517596875, 45.92123149822961, 20.031423546035004, 112.72008589959553]  
fr@fr-Inspiron-5567:~$
```

# Два worker'а



## Spark Master at spark://fr-Inspiron-5567:7077

URL: spark://fr-Inspiron-5567:7077

Alive Workers: 2

Cores in use: 2 Total, 0 Used

Memory in use: 2.0 GB Total, 0.0 B Used

Applications: 0 Running, 2 Completed

Drivers: 0 Running, 0 Completed

Status: ALIVE

### Workers (2)

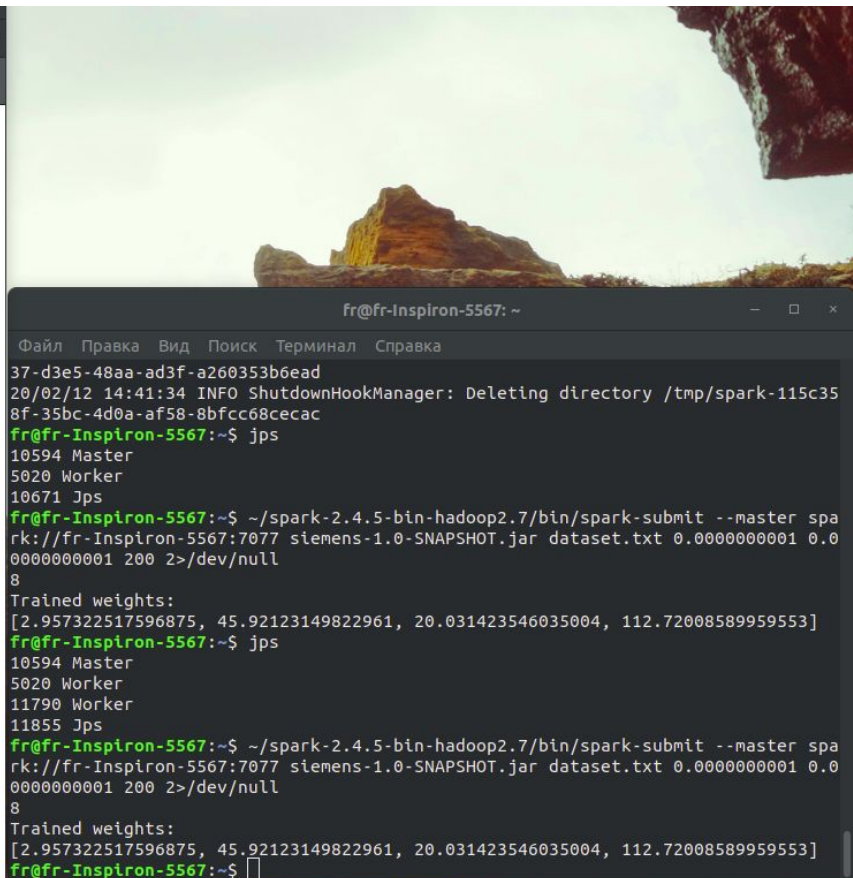
Worker Id	Address	State	Cores	Memory
worker-20200212131934-192.168.1.101-45095	192.168.1.101:45095	ALIVE	1 (0 Used)	1024.0 MB (0.0 B Used)
worker-20200212145643-192.168.1.101-42535	192.168.1.101:42535	ALIVE	1 (0 Used)	1024.0 MB (0.0 B Used)

### Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	----------------	------	-------	----------

### Completed Applications (2)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
app-20200212145727-0001	DistributedGradientDescent	2	1024.0 MB	2020/02/12 14:57:27	fr	FINISHED	3,7 min
app-20200212144413-0000	DistributedGradientDescent	1	1024.0 MB	2020/02/12 14:44:13	fr	FINISHED	5,4 min



# Три worker'a

Index of /di | vk.com/doc | vk.com/doc | Spark X | Applicat | Distribut | Distribut | Distribut | Distribut | +

localhost:8080

...

APACHE  
Spark 2.4.5

Spark Master at spark://fr-Inspiron-5567:7077

URL: spark://fr-Inspiron-5567:7077

Alive Workers: 3

Cores in use: 3 Total, 0 Used

Memory in use: 3.0 GB Total, 0.0 B Used

Applications: 0 Running, 3 Completed

Drivers: 0 Running, 0 Completed

Status: ALIVE

Workers (3)

Worker Id	Address	State	Cores	Memory
<a href="#">worker-20200212131934-192.168.1.101-45095</a>	192.168.1.101:45095	ALIVE	1 (0 Used)	1024.0 MB (0.0 B Used)
<a href="#">worker-20200212145643-192.168.1.101-42535</a>	192.168.1.101:42535	ALIVE	1 (0 Used)	1024.0 MB (0.0 B Used)
<a href="#">worker-20200212150353-192.168.1.101-37385</a>	192.168.1.101:37385	ALIVE	1 (0 Used)	1024.0 MB (0.0 B Used)

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	----------------	------	-------	----------

Completed Applications (3)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
<a href="#">app-20200212150421-0002</a>	DistributedGradientDescent	3	1024.0 MB	2020/02/12 15:04:21	fr	FINISHED	3,8 min
<a href="#">app-20200212145727-0001</a>	DistributedGradientDescent	2	1024.0 MB	2020/02/12 14:57:27	fr	FINISHED	3,7 min
<a href="#">app-20200212144413-0000</a>	DistributedGradientDescent	1	1024.0 MB	2020/02/12 14:44:13	fr	FINISHED	5,4 min

fr@fr-Inspiron-5567: ~

Файл Правка Вид Поиск Терминал Справка

```
[2.957322517596875, 45.92123149822961, 20.031423546035004, 112.72008589959553]
fr@fr-Inspiron-5567:~$ jps
10594 Master
5020 Worker
11790 Worker
11855 Jps
fr@fr-Inspiron-5567:~$ ~/spark-2.4.5-bin-hadoop2.7/bin/spark-submit --master spa
rk://fr-Inspiron-5567:7077 siemens-1.0-SNAPSHOT.jar dataset.txt 0.000000001 0.0
000000001 200 2>/dev/null
8
Trained weights:
[2.957322517596875, 45.92123149822961, 20.031423546035004, 112.72008589959553]
fr@fr-Inspiron-5567:~$ jps
10594 Master
13178 Jps
5020 Worker
13117 Worker
11790 Worker
fr@fr-Inspiron-5567:~$ ~/spark-2.4.5-bin-hadoop2.7/bin/spark-submit --master spa
rk://fr-Inspiron-5567:7077 siemens-1.0-SNAPSHOT.jar dataset.txt 0.000000001 0.0
000000001 200 2>/dev/null
8
Trained weights:
[2.957322517596875, 45.92123149822961, 20.031423546035, 112.72008589959552]
fr@fr-Inspiron-5567:~$
```



# Четыре worker'a



**Spark Master at spark://fr-Inspiron-5567:7077**

URL: spark://fr-Inspiron-5567:7077  
Alive Workers: 4  
Cores in use: 4 Total, 0 Used  
Memory in use: 4.0 GB Total, 0.0 B Used  
Applications: 0 Running, 4 Completed  
Drivers: 0 Running, 0 Completed  
Status: ALIVE

Workers (4)

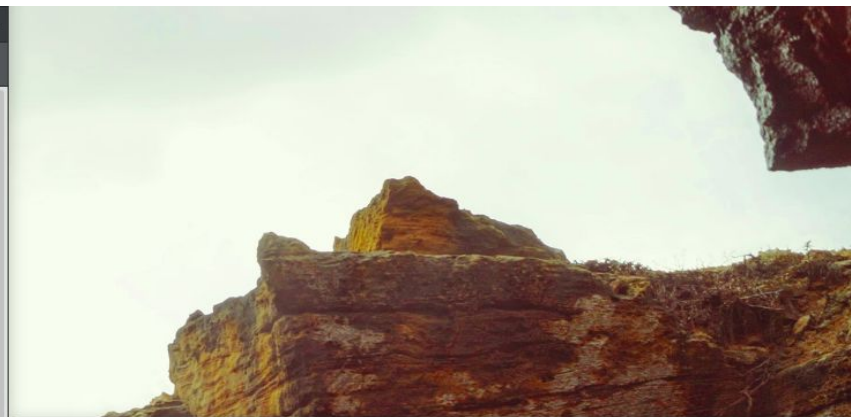
Worker Id	Address	State	Cores	Memory
worker-20200212131934-192.168.1.101-45095	192.168.1.101:45095	ALIVE	1 (0 Used)	1024.0 MB (0.0 B Used)
worker-20200212145643-192.168.1.101-42535	192.168.1.101:42535	ALIVE	1 (0 Used)	1024.0 MB (0.0 B Used)
worker-20200212150353-192.168.1.101-37385	192.168.1.101:37385	ALIVE	1 (0 Used)	1024.0 MB (0.0 B Used)
worker-20200212150952-192.168.1.101-39837	192.168.1.101:39837	ALIVE	1 (0 Used)	1024.0 MB (0.0 B Used)

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	----------------	------	-------	----------

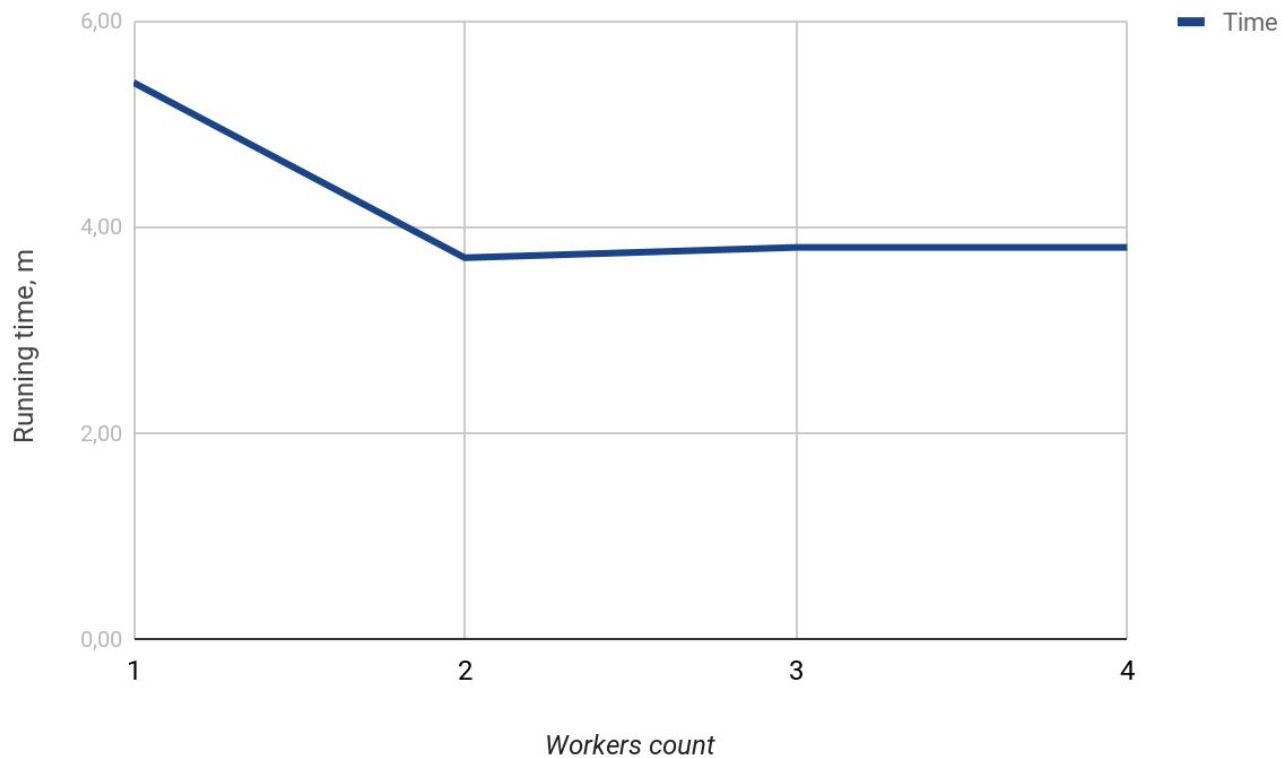
Completed Applications (4)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
app-20200212151011-0003	DistributedGradientDescent	4	1024.0 MB	2020/02/12 15:10:11	fr	FINISHED	3,8 min
app-20200212150421-0002	DistributedGradientDescent	3	1024.0 MB	2020/02/12 15:04:21	fr	FINISHED	3,8 min
app-20200212145727-0001	DistributedGradientDescent	2	1024.0 MB	2020/02/12 14:57:27	fr	FINISHED	3,7 min
app-20200212144413-0000	DistributedGradientDescent	1	1024.0 MB	2020/02/12 14:44:13	fr	FINISHED	5,4 min



```
fr@fr-Inspiron-5567: ~  
Файл Правка Вид Поиск Терминал Справка  
10594 Master  
13178 Jps  
5020 Worker  
13117 Worker  
11790 Worker  
fr@fr-Inspiron-5567:~$ ~/spark-2.4.5-bin-hadoop2.7/bin/spark-submit --master spa  
rk://fr-Inspiron-5567:7077 siemens-1.0-SNAPSHOT.jar dataset.txt 0.000000001 0.0  
000000001 200 2>/dev/null  
8  
Trained weights:  
[2.957322517596875, 45.92123149822961, 20.031423546035, 112.72008589959552]  
fr@fr-Inspiron-5567:~$ jps  
10594 Master  
14853 Jps  
14792 Worker  
5020 Worker  
13117 Worker  
11790 Worker  
fr@fr-Inspiron-5567:~$ ~/spark-2.4.5-bin-hadoop2.7/bin/spark-submit --master spa  
rk://fr-Inspiron-5567:7077 siemens-1.0-SNAPSHOT.jar dataset.txt 0.000000001 0.0  
000000001 200 2>/dev/null  
8  
Trained weights:  
[2.957322517596875, 45.92123149822961, 20.031423546035004, 112.72008589959553]  
fr@fr-Inspiron-5567:~$
```

## График зависимости времени выполнения от количества исполнителей



По графику видно, что время уменьшается при добавлении второго worker'а. При дальнейшем наращивании числа worker'ов время не уменьшается. Это связано с тем, что процессор, на котором запускался кластер, содержит лишь два ядра.

Результат работы алгоритма не зависит от количества worker'ов, что говорит о корректности разработанного распределенного алгоритма.



Информация о системе

Информация

Вид

Справка

Обновить

Создать отчёт

Копировать в буфер

- Компьютер
- Общая информация
- Операционная система
- Модули ядра
- Загрузки
- Языки
- Файловые системы
- Дисплей
- Переменные среды
- Разработка
- Пользователи
- Группы
- Устройства
- Процессор
- Память
- Устройства PCI
- Устройства USB
- Принтеры
- Батарея
- Сенсоры
- Устройства ввода
- Устройства хранения
- DMI
- Память SPD
- Ресурсы
- Сеть
- Интерфейсы
- IP Подключения
- Таблица маршрутизации
- Таблица ARP
- DNS сервера
- Статистика
- Общие каталоги
- Тесты
- CPU Blowfish
- CPU CryptoHash
- CPU Fibonacci
- CPU N-Queens

Компьютер →

- Операционная система  
Ubuntu 18.04.3 LTS
- Процессор  
Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz  
1 физический процессор; 2 ядра; 4 потока
- Оперативная память  
16294792 КиБ
- Материнская плата  
Dell Inc. 02YHJP
- Графика  
1920x1080  
(Неизвестно)  
The X.Org Foundation
- Устройства хранения  
ATA KINGSTON SA400S3  
ATA TOSHIBA MQ01ABD1  
USB DISK 2.0
- Принтеры
- Аудио  
HDA-Intel - HDA Intel PCH