



# Грузим данные в Graphite и строим отображение в Grafana

**Драль Алексей**, [study@bigdatateam.org](mailto:study@bigdatateam.org)

CEO at BigData Team, <https://bigdatateam.org>

<https://www.facebook.com/bigdatateam>



**BIGDATA  
TEAM**

# Coursera Specialization on Big Data

## Capstone Project



Hadoop, Spark

Hive, Spark

Spark ML

Real Time

**Y**andex

<https://bigdatateam.org/big-data-engineering>

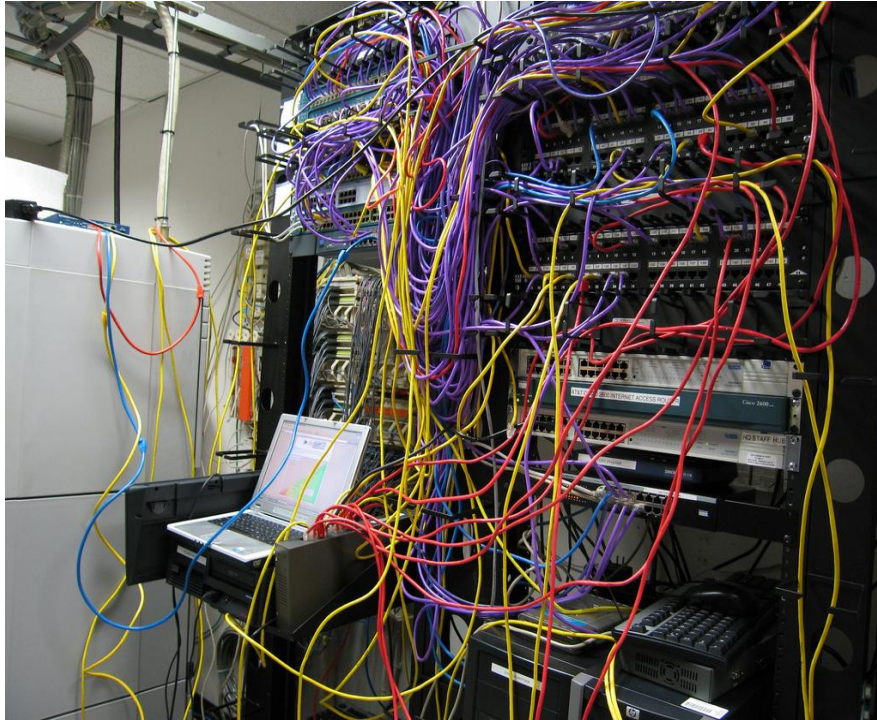


# Сколько стоит кластер?

- ▶ Namenode
- ▶ 3 Datanodes
- ▶ Client (edge node)

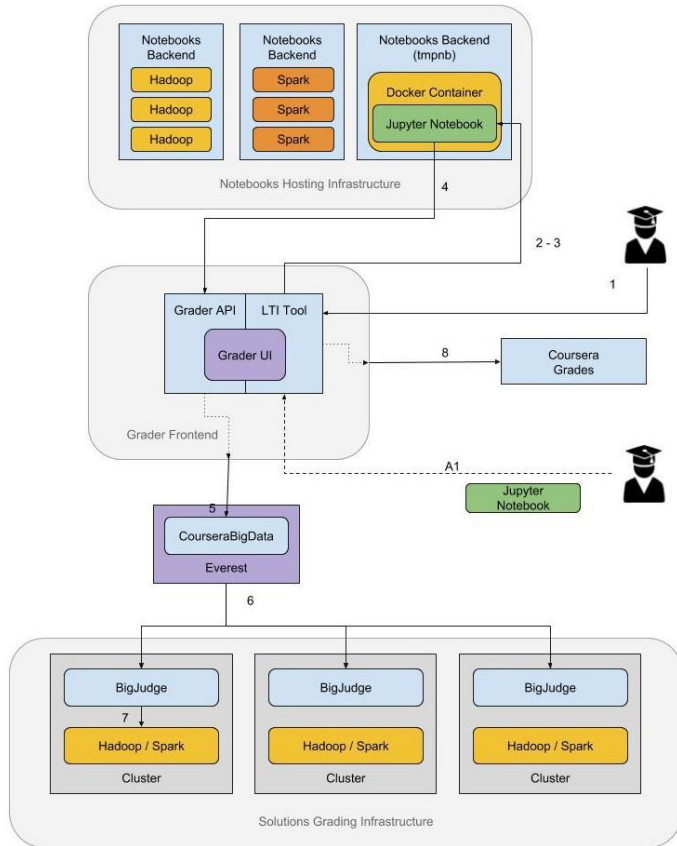
Cloud: ~1500\$/месяц

Coursera: 29-49-79\$/месяц





## (из презентации RAIF'2018)



- ▶ In-browser Big Data coding
- ▶ Zero-time environment launch
- ▶ Free Docker containers for local playground
- ▶ In-browser submission (Coursera LTI-integration)
- ▶ Hadoop cluster submission balancer
- ▶ Distributed execution and validation with 5 min SLA
- ▶ ...



**BIGDATA  
TEAM**

# Бизнес-метрики: качество обучения

feedback



Google Forms





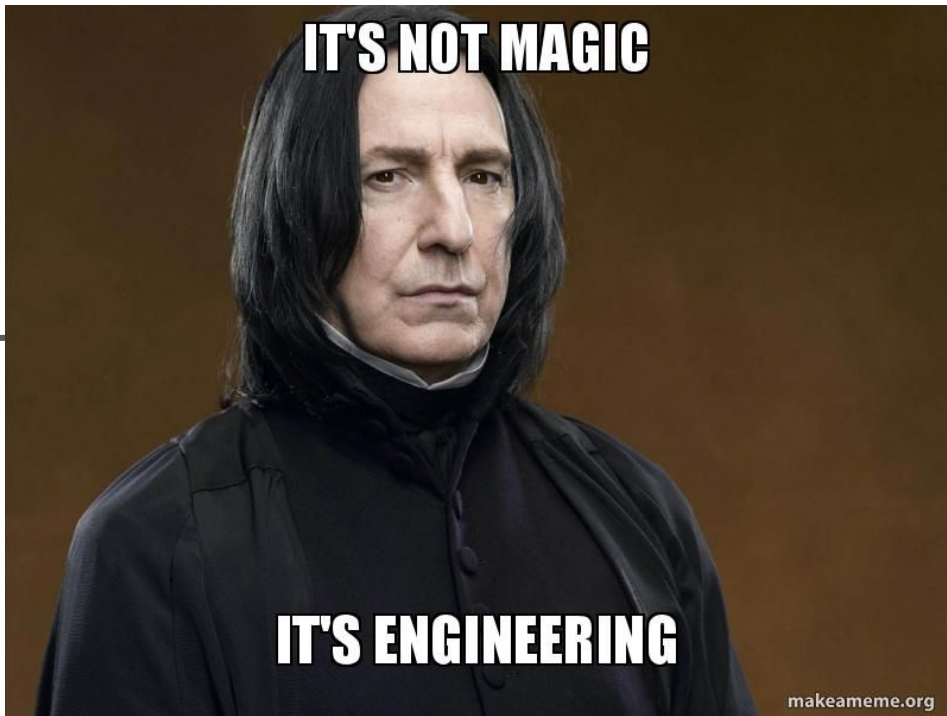
**BIGDATA  
TEAM**

# Бизнес-метрики: качество обучения

feedback



Google Forms



alerts





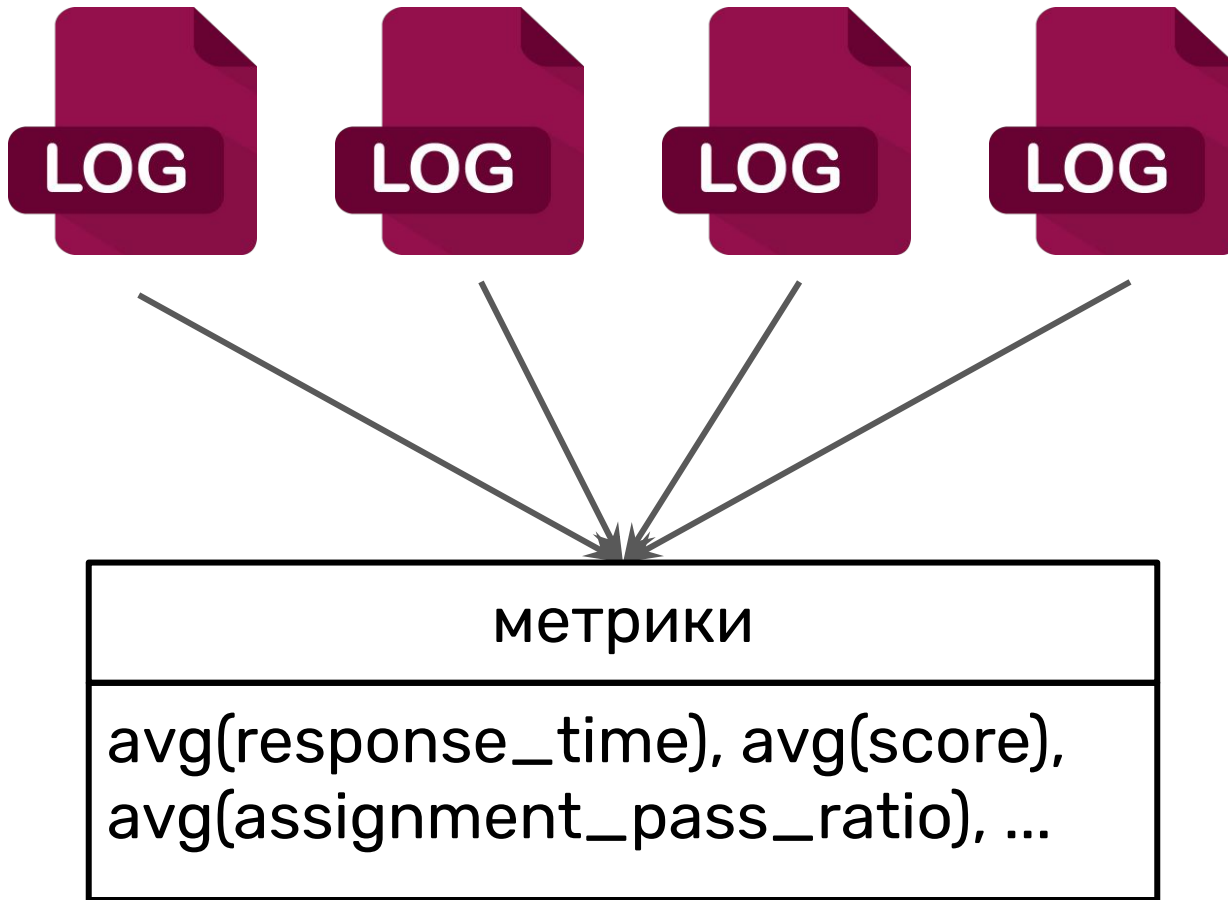
# Проверка решений и логи



...



```
task=X user=Y timestamp=Z submission=bl  
task=X user=Y timestamp=W submission=X  
...
```







`{"number": "spark.collocation", "score": 0.6, "timestamp": 1600050207}`

`{"number": "map_reduce.stackoverflow", "score": 0.0, "timestamp": 1600053047}`

`{"number": "spark.bigram", "score": 0.36000000000004, "timestamp": 1600042876}`

`{"number": "spark.collocation", "score": 0.42, "timestamp": 1600042876}`

`{"number": "spark.bigram", "score": 0.40000000000001, "timestamp": 1600118617},`

`{"number": "spark.collocation", "score": 0.48, "timestamp": 1600118617}`

...



БД	популярность*
1. InfluxDB	26.15
2. KDB+	7.66
3. Prometheus	5.75
4. Graphite	4.68
5. RRDTool	3.30
...	...

Источник: [db-engines.com](https://db-engines.com), \*[методология](#)



**BIGDATA  
TEAM**

# Эволюция мониторинга

feedback



Google Forms



alerts





feedback



Google Forms



Prometheus



Grafana

alerts



slack

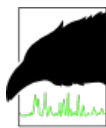




feedback



Google Forms



**MUNIN**



Prometheus



Grafana



graphite



Grafana

alerts



slack





Первоочередные вопросы:

- 1.** Частота производства данных;



Первоочередные вопросы:

- 1.** Частота производства данных;
- 2.** Минимально-достаточная точность данных;



Первоочередные вопросы:

- 1.** Частота производства данных;
- 2.** Минимально-достаточная точность данных;
- 3.** Максимальный период времени для оперативных данных;





Первоочередные вопросы:

- 1.** Частота производства данных;
- 2.** Минимально-достаточная точность данных;
- 3.** Максимальный период времени для оперативных данных;
- 4.** Наименьшая точность, которая имеет смысл;

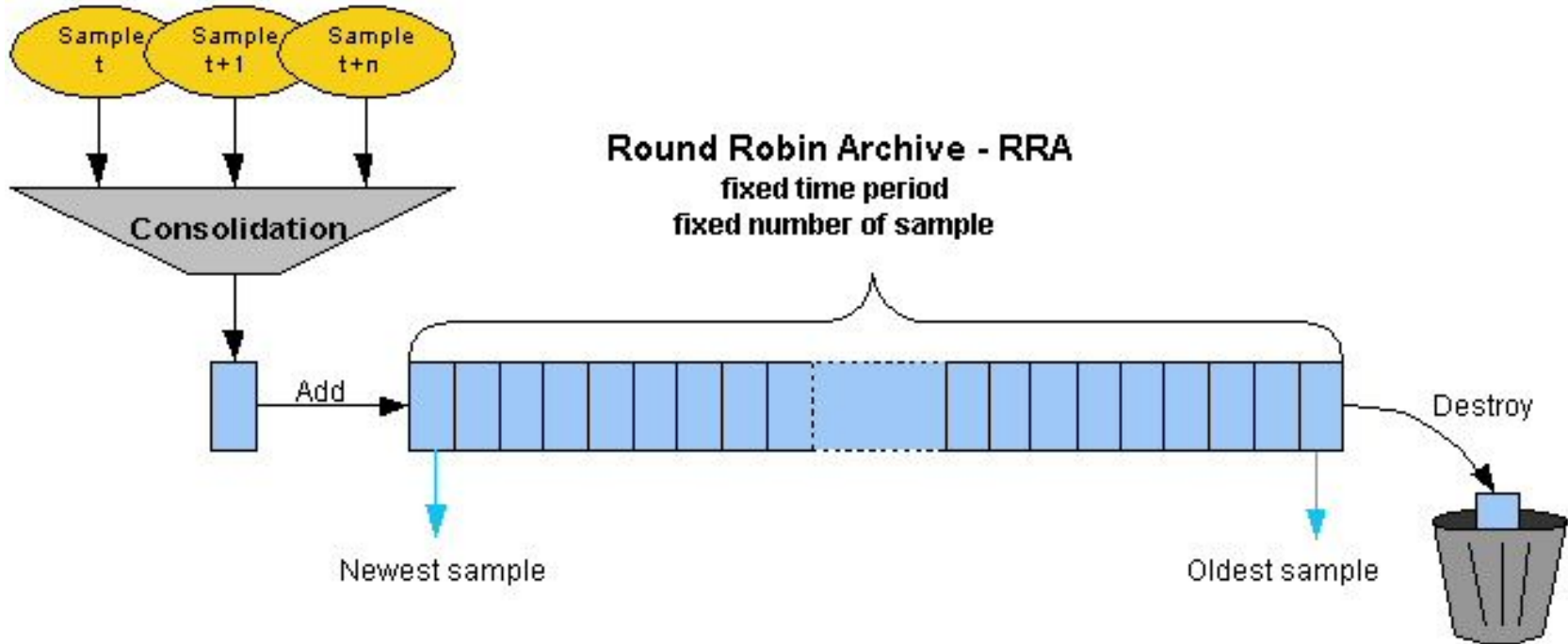


Первоочередные вопросы:

- 1.** Частота производства данных;
- 2.** Минимально-достаточная точность данных;
- 3.** Максимальный период времени для оперативных данных;
- 4.** Наименьшая точность, которая имеет смысл;
- 5.** Как далеко смотреть в прошлое.



# Консолидация метрик (пример)





```
$ tree
...
|-- map_reduce
|   |-- ids.wsp
|   `-- stackoverflow.wsp
|-- spark
|   |-- bigram.wsp
|   |-- collocation.wsp
|   |-- nested_crud.wsp
|   `-- sssp.wsp
...
```

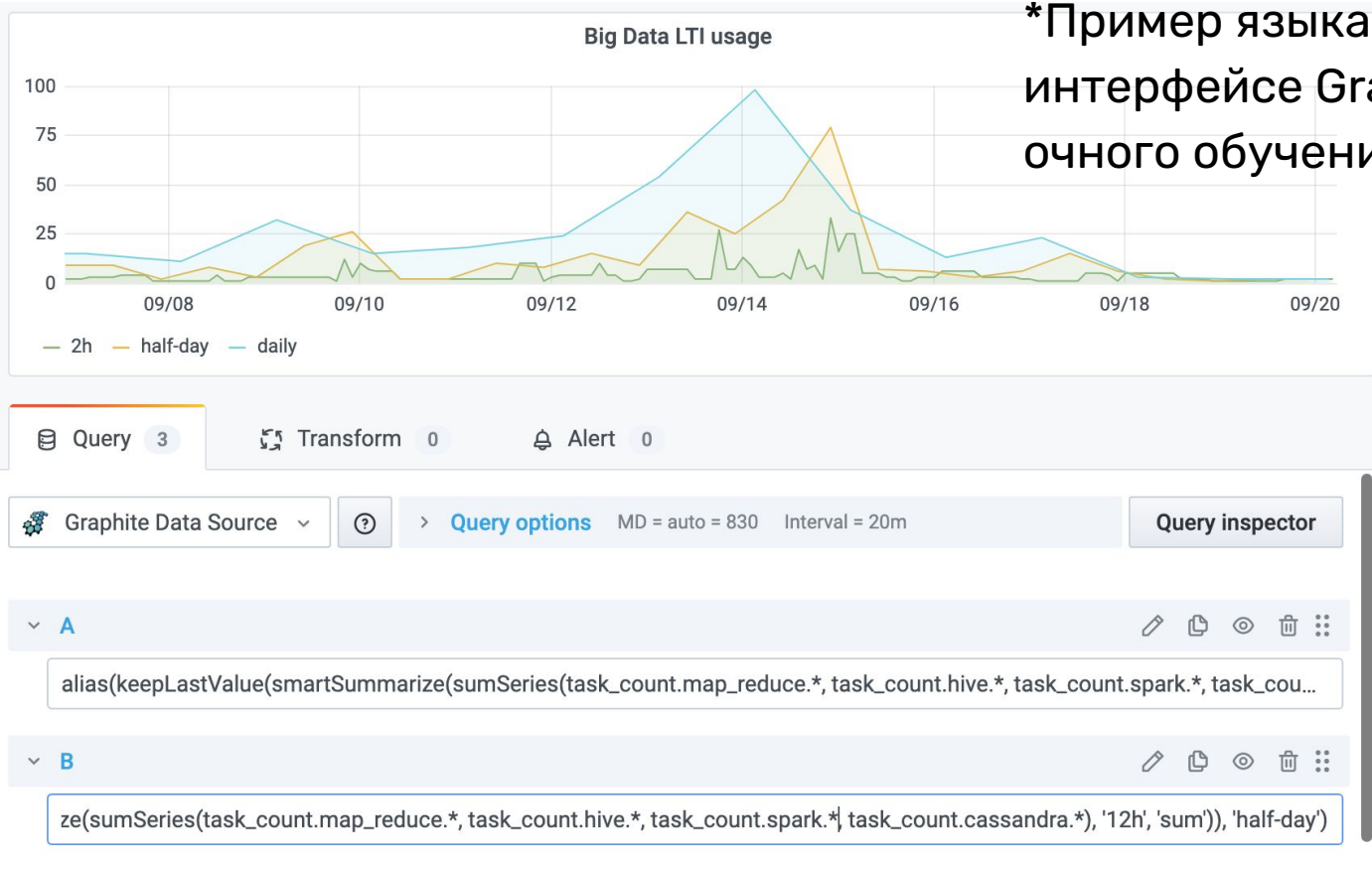


```
$ whisper-info.py /path/to/metric.wsp
...
secondsPerPoint: 300
...
$ whisper-fetch.py --from=$(date +%s -d "-60 min") metric.wsp
...
1600043400 0.420000
...
$ echo "score_result.spark.collocation 0.42 1600042876" | nc -N
localhost 2003
```

**\*Обратите внимание на точность бакета**



\*Пример языка запросов Graphite в интерфейсе Grafana для аналитики очного обучения по Big Data





- ▶ Флудить на тему time-series БД



- ▶ Флудить на тему time-series БД
- ▶ Прийти (*пока не поздно*) к бизнесу и убедиться, что все необходимые данные есть в логах