



## Резюме модуля и мини-курса

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

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

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



- ▶ Вы умеете в Hadoop и Hive Map-, Reduce-, Bucket- и SMB Join



- ▶ Вы умеете в Hadoop и Hive Map-, Reduce-, Bucket- и SMB Join
- ▶ Вы умеет пользоваться Managed и External таблицами без риска удаления production данных, а также RegexSerDe для задания произвольной схемы



- ▶ Вы умеете в Hadoop и Hive Map-, Reduce-, Bucket- и SMB Join
- ▶ Вы умеет пользоваться Managed и External таблицами без риска удаления production данных, а также RegexSerDe для задания произвольной схемы
- ▶ Вы умеете в Secondary Sort с помощью Comparator и Partitioner



## 1. HDFS, Map Reduce, Hive

1	Введение в Большие Данные (Big Data) и распределенные файловые системы
2	Hadoop экосистема, YARN и MapReduce
3	Оптимизация MapReduce вычислений
4	SQL поверх больших данных (Hive)





## 2. Spark: from zero to hero

5	Модель вычислений Spark: RDD
6	Spark DataFrames, Spark SQL
7	Оптимизация Spark вычислений

## 3. RT, NoSQL, Data Layout

8	Потоковая обработка данных (Kafka, Spark Streaming)
9	NoSQL поверх больших данных: Cassandra
10	Data Layout