



**BIGDATA  
TEAM**

# **Big Data: цель курса и его структура**

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

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

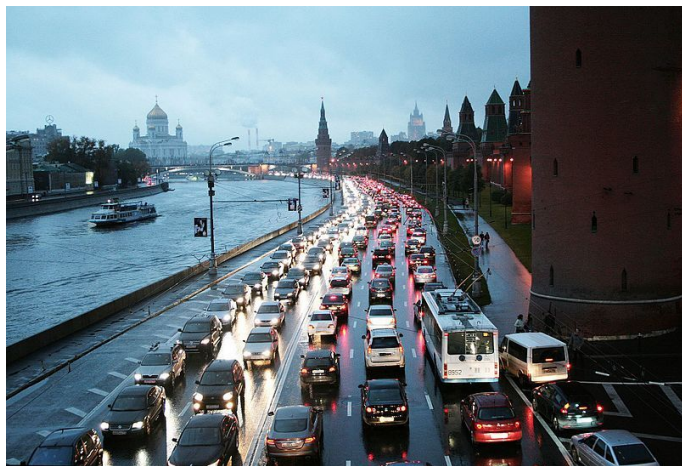
<https://www.facebook.com/bigdatateam>






**BIGDATA  
TEAM**

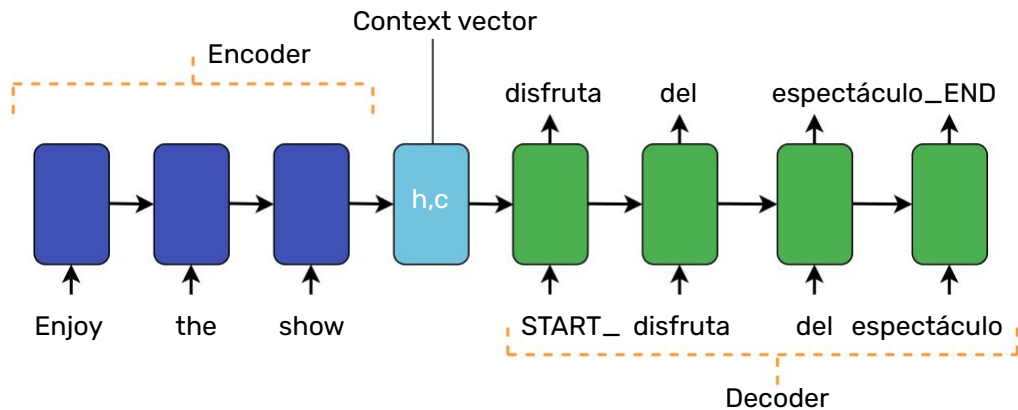
Big Data + ML = ❤️

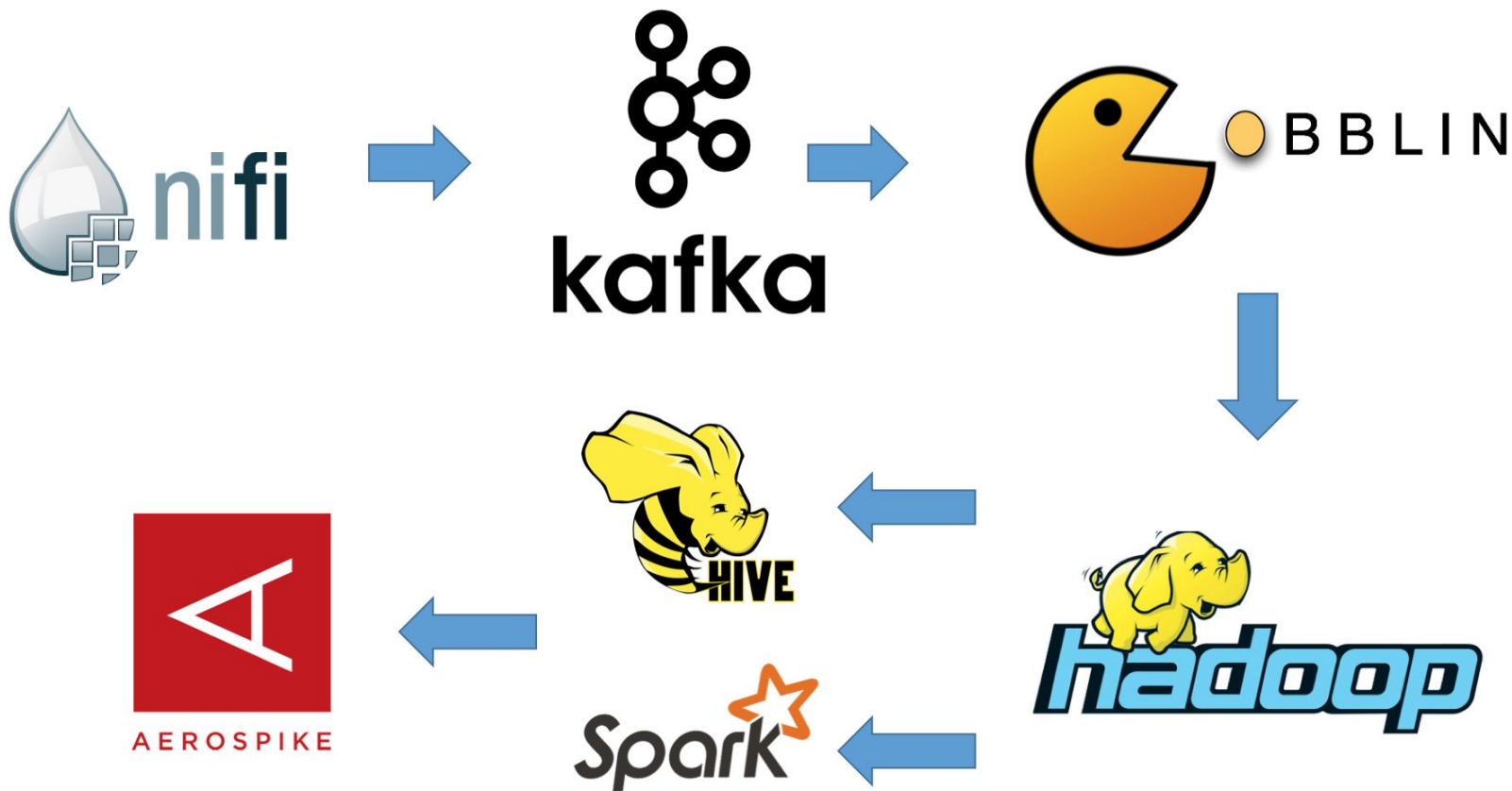




Персонализированные предложения

		
Чайник 2000 рублей	Чашка 199 рублей	Лопаточка 329 рублей







**BIGDATA  
TEAM**

# Практические навыки:



**BIGDATA  
TEAM**

# Практические навыки:

► HDFS





**BIGDATA  
TEAM**

Практические навыки:

- ▶ HDFS
- ▶ YARN





**BIGDATA  
TEAM**

## Практические навыки:

- ▶ HDFS
- ▶ YARN
- ▶ Hive







**BIGDATA  
TEAM**

## Практические навыки:

- ▶ HDFS
- ▶ YARN
- ▶ Hive
- ▶ Spark





**BIGDATA  
TEAM**

Практические навыки:

- ▶ HDFS
- ▶ YARN
- ▶ Hive
- ▶ Spark
- ▶ Kafka





**BIGDATA  
TEAM**

Практические навыки:

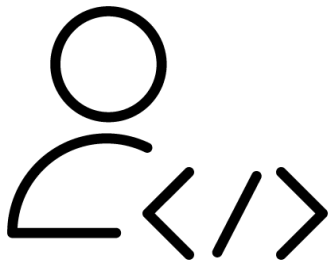
- ▶ HDFS
- ▶ YARN
- ▶ Hive
- ▶ Spark
- ▶ Kafka
- ▶ Cassandra





**BIGDATA  
TEAM**

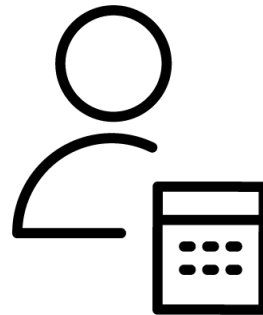
# Кому подойдет этот курс?



?



?



?



?



**BIGDATA  
TEAM**

# У кого слон больше?



**Индийский слон**



**Африканский слон**



**BIGDATA  
TEAM**

У кого слон больше?







# BIGDATA TEAM



## DS или DE ?

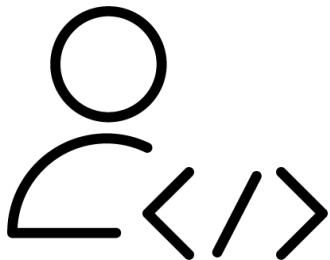


# Денежный вопрос

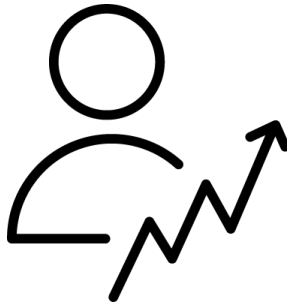
$$\frac{(n+1)!}{n! \cdot ((n+1) \rho_0 (k+12))!} = \frac{273,15 \text{ K}}{n! \cdot ((n+1) \rho_0 (k+12))!}$$



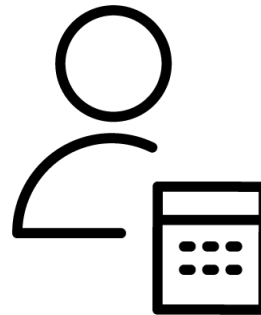
# Кому подойдет этот курс?



**разработчик**



**?**



**?**

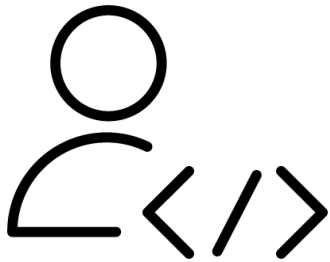


**?**

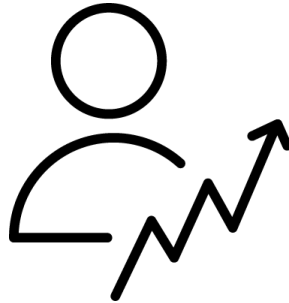




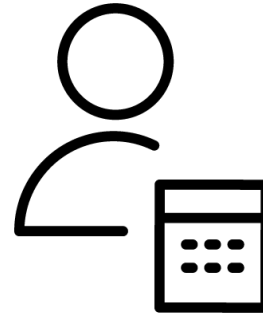
# Кому подойдет этот курс?



**разработчик**



**аналитик**



**?**

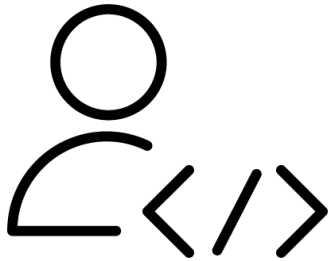


**?**



**BIGDATA  
TEAM**

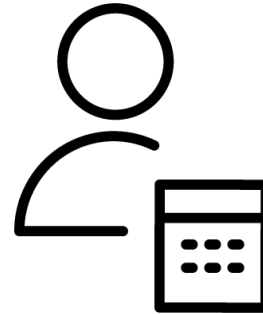
# Кому подойдет этот курс?



**разработчик**



**аналитик**



**Data  
Engineer**

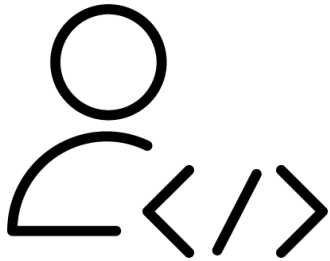


**?**



**BIGDATA  
TEAM**

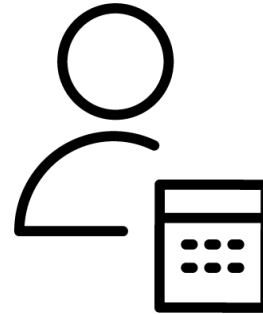
# Кому подойдет этот курс?



**разработчик**



**аналитик**



**Data  
Engineer**



**Data  
Scientist**



**BIGDATA  
TEAM**

Есть сомнения?



**Хотел стать пандой?**

**Все в твоих руках!**



**BIGDATA  
TEAM**

# Data Scientist vs Data Engineer

**indeed**

## Index jobs

My recent searches

data engineer - 34,228 new

data scientist - 9,279 new

**3-4x**

glassdoor



data engineer

United States

Jobs



Job Type

Date Posted

Salary Range

Distance

More

Data Engineer Jobs in  
United States

101,151  
Jobs



glassdoor



data scientist

United States

Jobs



Job Type

Date Posted

Salary Range

Distance

More

Data Scientist Jobs in  
United States

27,202  
Jobs





# Data Scientist vs Data Engineer

Average Base Pay

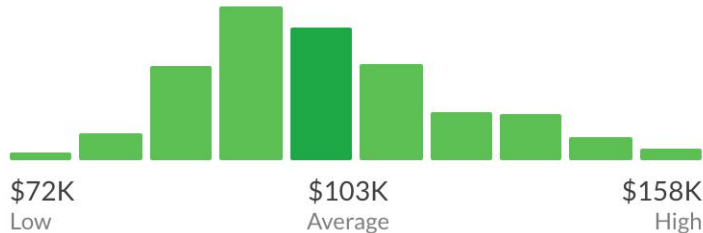
**\$113,309** /yr



**Data Scientist**

Average Base Pay

**\$102,864** /yr



**Data Engineer**



## **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)



## 1. HDFS, Map Reduce, Hive

1	Введение в Большие Данные (Big Data) и распределенные файловые системы





## **1. HDFS, Map Reduce, Hive**

1	Введение в Большие Данные (Big Data) и распределенные файловые системы
2	Hadoop экосистема, YARN и MapReduce



## **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



## **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



## 2. Spark: from zero to hero

5	Модель вычислений Spark: RDD



## 2. Spark: from zero to hero

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



## **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





3. RT, NoSQL, Data Layout	
8	Потоковая обработка данных (Kafka, Spark Streaming)



## 3. RT, NoSQL, Data Layout

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



## 3. RT, NoSQL, Data Layout

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



**BIGDATA  
TEAM**

# Coursera Specialization on Big Data

## Capstone Project



Hadoop, Spark

**Y**andex

Hive, Spark

Spark ML

Real Time

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