

Экосистема Hadoop, Spark, Hive

Экосистема Hadoop, Spark, Hive

Длительность курса: 44 академических часа

1 Экосистема Hadoop

- 1 **Hadoop (основные компоненты, дистрибутивы вендоров)**
- 2 **Архитектура HDFS (HDFS, ZooKeeper)**
- 3 **Архитектура YARN (YARN, MapReduce)**
- 4 **Форматы данных (Parquet, Orc, Avro)**

1 **Базовый синтаксис**

2 **Инструменты
разработки и
сборки (практика)**

3 **ООП**

4 **Элементы ФП**

1 **Первый проект на Apache Spark**

2 **Архитектура приложения Spark**

3 **RDD/Dataframe/Dataset**

4 **Методы оптимизации**

5 **Spark Streaming 1**

6 **Spark Streaming 2**

7 **GraphX**

8 **GeoSpark**

9 **SparkML/XGBoost**

- 1 **Обзор Hive, класса инструментов, возможностей**
 - 2 **Операции, форматы, детали, демо**
-

1 **Оркестрация
процессов
обработки данных**

2 **Мониторинг Spark-
приложений**

3 **CI/CD для Spark и
Hive**