

## Основы Hadoop

**Код:** EAS-015

**Длительность:** 24 ч.

**Описание:**

Этот тренинг дает представление об устройстве Apache Hadoop и методах разработки приложений, обрабатывающих данные на его основе. Участники познакомятся с HDFS – стандартом де-факто для долговременного надежного хранения больших объемов данных фреймворком YARN, управляющим параллельным выполнением приложений на кластерети сопутствующими проектами, составляющими экосистему Hadoop: Hive, Spark, HBase.

**Цели:**

- Понимать ключевые концепции и архитектуру Hadoop;
- Получить представление об экосистеме, сложившейся вокруг Hadoop, и ее ключевых компонентах;
- Уметь записывать и читать данные в/из HDFS;
- Понимать устройство MapReduce-программ;
- Уметь обращаться к табличным данным с использованием Hive;
- Уметь обращаться к табличным данным с использованием Spark SQL/DataFrame в пакетном режиме;
- Уметь обрабатывать потоки данных с использованием Spark Structured Streaming;
- Уметь использовать HBase для низколатентного сохранения и чтения данных.

**Целевая аудитория:**

Разработчики, архитекторы, разработчики баз данных, администраторы баз данных.

**Предварительная подготовка - общее:**

Умение работать в командной оболочке Unix/Linux (bash). Опыт работы с базами данных желателен, но не обязателен.

**Рекомендуемые дополнительные материалы, источники:**

- [Tom White. Hadoop Definitive Guide. 4th ed. O'Reilly, 2015.](#)
- Kevin Sitto, Marshall Presser. Field Guide to Hadoop. O'Reilly, 2015.

- Thilina Gunarathne. Hadoop MapReduce v2 Cookbook. 2nd ed. Packt, 2015.
- Sandeep Karanth. Mastering Hadoop. Packt, 2014.