

Таблица 1

Наименование программы	Количество часов	Форма обучения	Стоимость обучения 1 слушателя, руб.
Основы Apache Hadoop	32+4	очная	25 000

Описание программы

### **АННОТАЦИЯ**

#### **к программе повышения квалификации**

**Цель программы:** качественное изменение профессиональных компетенций, направленных на повышение профессионального уровня в рамках имеющейся у слушателей квалификации в области работы с большими данными и распараллеливанию обработки данных с помощью современных технологий.

**Результатом обучения** является приобретение **знаний и умений** по:

- Работе с данными в распределенных отказоустойчивых системах на примере дистрибутива Cloudera Hadoop
- Применимости технологий больших данных в реальных проектах
- Основам настройки и администрирования кластера Cloudera Hadoop
- Работе с распределенной файловой системой HDFS, загрузке данных из внешних файлов и баз данных
- Трансформации и анализу данных с помощью SQL-интерфейса Apache Hive
- Обзору фреймворков MapReduce, Apache Spark, Apache Kafka, Apache AirFlow и других популярных компонентов современных архитектур для работы с большими данными

**Категория слушателей:** работники профильных отраслей промышленности, высшего образования, научно-исследовательских организаций, IT-сектора (аналитики, тестировщики, разработчики, менеджеры), знакомые с основами командной строки Linux, базовым навыкам программирования и языком SQL.

**Объем программы:** 36 часов.

**Форма обучения:** очная

Таблица 2

## УЧЕБНЫЙ ПЛАН ПРОГРАММЫ ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

№ п/п	Наименование разделов	Трудоёмкость, час	Аудиторные занятия, дистанционные занятия					Аттестация		СР С, час
			Всего, час	Лекции, час	Лабораторные работы, час	Практические занятия, семинары, час	час	Форма контроля		
1	Понятие «больших данных». Обзор экосистемы Apache Hadoop и дистрибутива Cloudera Hadoop	4	4	2	2	0	0		0	
2	Платформа Apache Hadoop. Файловая система HDFS. Форматы хранения данных. Компрессия данных.	4	4	2	2	0	0		0	
3	Планировщик Apache Hadoop YARN. Управление и настройка ресурсов кластера.									
4	SQL-интерфейс к данным	4	4	2	2	0	0		0	

	Араche Hive. Архитектура, движки выполнения, особенности работы.								
5	Фреймворк Map Reduce: обзор паттернов обработки данных. Оптимизация соединений таблиц.	4	4	2	2	0	0		0
6	Обзор движка Араche Spark. Spark SQL. Преимущества и недостатки.	4	4	2	2	0	0		0
7	Обзор современной экосистемы работы с большими объемами данных: виды инструментов и примеры архитектур	4	4	4	0	0	0		0
8	Итоговая аттестация	4	4	0	0	0	4	Зачет	0
	Итого	32	32	16	12	0	4		0

Всего 36 часов