

ПРОГРАММА КУРСА «ВВЕДЕНИЕ В BIG DATA. ОСНОВЫ РАЗРАБОТКИ ВЫСОКОНАГРУЖЕННЫХ ПРИЛОЖЕНИЙ»



Общий объем тренинга: 28 академических часов (17 теории + 11 практики).

Дополнительно предполагается 30 часов самостоятельной работы.

Рекомендуемая продолжительность курса:

10 недель в календарном исчислении, по 1 трёхчасовое занятие в неделю.

Тема	Теория, часы	Практика, часы
1. Высоконагруженные приложения - Шардирование - Реплицирование - Рост нагрузки и падение серверов - Серверный стек технологий	3	0
2. Java в BigData - Collection Framework - JVM internals - Работа сборщика мусора (и его разновидности) - Работа с файлами - Как работает загрузка классов - Пулы соединений/поток	2	1
3. HDFS - Демоны HDFS - Консольные команды - Java API	2	1
4. NoSQL - Key-value хранилища - Документориентированные БД - Таблицы в оперативной памяти - Cassandra - CQL - Hbase	2	1
5. MapReduce - Основная идея - Mapper, Reducer и Combiner - Запуск MR приложения	2	1
6. Hive - SQL синтаксис для выполнения MR-заданий - Создание таблиц и их заполнение - Техники оптимизации	2	1

7. Основы Scala - Особенности Scala - Сборка и запуск проекта - Подключение зависимостей в sbt	2	1
8. Spark - Быстрые вычисления в памяти - Dataset API - Азы машинного обучения - Использование MLlib	2	1
9. Финальный экзамен (онлайн) Тест + Практические задания	0	4

К слушателям предъявляются следующие требования:

- знание основ ООП/ФП (Java, Scala)
- опыт работы с Java от полугода (наличие своего pet-project, а лучше «боевого» использования знаний на проекте)
- требуется хорошее понимание и опыт работы со след.разделами Java (Java Core, Multithreading, Collection Framework, Maven/Gradle, IO/NIO, Stream API)
- опыт работы с базами данных жизненно необходим, SQL на базовом уровне вы уже должны знать
- технический английский (на уровне чтения и понимания технической литературы)
- хорошая обучаемость, самообучаемость