

Аннотация рабочей программы дисциплины «Современные методики анализа, технологии хранения и обработки неструктурированных и больших данных»

Целью освоения дисциплины «Современные методики анализа, технологии хранения и обработки неструктурированных и больших данных» является получение широкого набора теоретических знаний и практических навыков, позволяющих обрабатывать, анализировать и извлекать полезную информацию из огромного массива сложных неструктурированных данных. Изучение дисциплины включает в себя ознакомление с такими программными продуктами, как Spark, Pig и Hive, лежащих в основе платформы Hadoop, признанной эталоном в среде анализа больших данных.

Задачи дисциплины:

- ознакомление с понятием «Большие данные»;
- изучение современных подходов к использованию распределенных вычислительных ресурсов;
- получение навыков работы с массивами неструктурированных данных.

В результате освоения дисциплины обучающиеся изучат теоретический и практический материал по следующим темам:

Введение в «Большие данные»
Платформа и фреймворк Hadoop
Аналитика в «Больших данных»
Машинное обучение и «Большие данные»
Обработка больших объемов графовых данных

Объем дисциплины 2 з.е.

Форма промежуточного контроля – *зачет*.