

## СЕРТИФИКАТ

№ 02BIGDATA-0218-7

дата выдачи: 03.02.2019 г.

подтверждает, что

# Кашин Павел Андреевич

успешно освоил(а) курс

## НАУКА О ДАННЫХ И АНАЛИТИКА БОЛЬШИХ ОБЪЕМОВ ДАННЫХ

2,5 зачетных единицы

Описание освоенного курса и достигнутых результатов обучения приведено в приложении к настоящему сертификату.

Электронная версия сертификата:

<http://open.spbstu.ru/certificate/02BIGDATA-0218-7.pdf>



проректор  
по образовательной деятельности  
Е. М. Разинкина

# Кашин Павел Андреевич

Идентификационный номер: 1205224

САНКТ ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ  
ПЕТРА ВЕЛИКОГО  
<http://www.spbstu.ru/>

КУРС: НАУКА О ДАННЫХ И АНАЛИТИКА БОЛЬШИХ ОБЪЕМОВ ДАННЫХ  
<https://openedu.ru/course/spbstu/BIGDATA/>

Период освоения курса:  
с 29 октября 2018 г. по 31 января 2019 г.

Оценка, количество часов и кредитов за курс:

Зачетные единицы	Часы		Оценка		
	Общие	Акад.	100-балльная	5-балльная	Прописью
2,5	66	90	98	5	отлично

Шкала соответствия системы оценивания:

Шкала оценивания		
100-балльная	5-балльная	Прописью
90-100	5	отлично
70-89	4	хорошо
50-69	3	удовлетворительно
0-49	2	неудовлетворительно

**ПРОГРАММА КУРСА:**

## **Тема 1. Введение в большие данные**

1. Определение больших данных и причины их появления
2. Примеры возможностей для бизнеса

## **Тема 2. Жизненный цикл аналитики данных**

1. Различие между Business Intelligence и Big Data
2. Понятие жизненного цикла аналитики данных
3. Роли, необходимые для успешного создания проекта по аналитике данных
4. Инструменты получения и обмена данными

### **Тема 3. Высокопроизводительные вычисления**

1. Распределенные вычисления на нескольких серверах, вычислительная парадигма MapReduce
2. Проект Apache Hadoop и его экосистема
3. Apache Spark и его компоненты
4. Вычисления в реальном времени, Apache Storm, Flink

### **Тема 4. Масштабирование и многоуровневое хранение данных**

1. Теорема CAP
2. Парадигма NoSQL
3. Классификация NoSQL баз данных

### **Тема 5. Визуализация данных и результатов анализа**

1. Техники визуализации данных.
2. Язык R.

### **Тема 6. Сложные методы аналитики**

1. Классификация задач анализа: Text, Data, Web, Social Mining
2. Статистические методы анализа данных. Применение машинного обучения в аналитике

### **Тема 7. Анализ текста**

1. Поисковые механизмы: Lucene, Solr, Elasticsearch
2. Алгоритм Word2Vec

### **Итоговая аттестация**

#### **РЕЗУЛЬТАТЫ ОБУЧЕНИЯ:**

- получил знания о жизненном цикле аналитики данных, технологиях и средствах распределенной обработки и хранения данных, базовых методах аналитики больших объемов данных, техниках визуализации данных;
- овладел умением использовать типовые технологии и средства аналитики данных, такие как MapReduce, Hadoop, NoSQL, язык R;
- изучил методы использования, совершенствования, разработки и внедрения современных технологий и инструментальных средств анализа и работы с большими объемами данных,

#### **НАПРАВЛЕННЫЕ НА ФОРМИРОВАНИЕ КОМПЕТЕНЦИЙ:**

- способность осваивать, разрабатывать и внедрять новые технологии обработки и анализа данных;
- способность извлекать закономерности из данных и обрабатывать неструктурированные данные;

- способность выполнять работы по созданию, модификации и сопровождению информационных систем, автоматизирующих задачи организационного управления и бизнес-процессы;
- способность анализировать потребности заинтересованных лиц и подразделений организации и подходы к исследованию больших данных;
- способность разрабатывать продукты на основе встроенной аналитики больших данных,

#### СООТВЕТСТВУЮЩИХ ФГОС ВО СЛЕДУЮЩИХ НАПРАВЛЕНИЙ ПОДГОТОВКИ:

02.00.00 Компьютерные и информационные науки

09.00.00 Информатика и вычислительная техника

27.00.00 Управление в технических системах

#### СИСТЕМА ОЦЕНИВАНИЯ (100-БАЛЛЬНАЯ)

№	Наименование оценивающего мероприятия	Набранный балл	Максимальный балл	Коэффициент
1	Промежуточные тесты	100	100	0,5
2	Экзаменационный тест	96	100	0,5
	Итоговая оценка	98	100	1

Приложение к сертификату №  
02BIGDATA-0218-7  
дата выдачи: 03.02.2019 г.



проректор  
по образовательной деятельности  
Е. М. Разинкина