

Документ подписан электронной подписью  
Информация о владельце:  
ФИО: Ашмарина Светлана Игоревна  
Должность: Ректор ФГБОУ ВО «Самарский государственный экономический университет»  
Дата подписания: 01.02.2020  
Уникальный программный ключ:  
59650034d6e3a6baac49b7bd0f8e79fea1433ff3e82f1fc7e9279a031181baba

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования

«Самарский государственный экономический университет»

**Институт** экономики предприятий  
**Кафедра** Цифровых технологий и решений

**УТВЕРЖДЕНО**  
Ученым советом Университета  
(протокол № 10 от 29 апреля 2020 г.)

**РАБОЧАЯ ПРОГРАММА**

**Наименование дисциплины** Б1.В.10 Технологии Big Data

**Основная профессиональная образовательная программа** 09.03.03 Прикладная информатика программа  
Прикладная информатика в электронной экономике

Методический отдел УМУ  
« 16 » апрель 20 20 г.  
Сахарова С.Ю. / Сахарова С.Ю.

Рассмотрено к утверждению  
на заседании кафедры Цифровых технологий и решений  
(протокол № 8 от 05.03.2020г.)  
Зав. кафедрой Е.В. Погорелова /

Научная библиотека СГЭУ  
« 16 » апрель 20 20 г.  
Е.В. Погорелова /

Квалификация (степень) выпускника бакалавр

## Содержание (рабочая программа)

Стр.

- 1 Место дисциплины в структуре ОП
- 2 Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов обучения по программе
- 3 Объем и виды учебной работы
- 4 Содержание дисциплины
- 5 Материально-техническое и учебно-методическое обеспечение дисциплины
- 6 Фонд оценочных средств по дисциплине

Целью изучения дисциплины является формирование результатов обучения, обеспечивающих достижение планируемых результатов освоения образовательной программы.

## 1. Место дисциплины в структуре ОП

Дисциплина Технологии Big Data входит в часть, формируемая участниками образовательных отношений блока Б1.Дисциплины (модули)

Предшествующие дисциплины по связям компетенций: Математика, Дискретная математика, Теория вероятностей и математическая статистика, Исследование операций и методы оптимизации

Последующие дисциплины по связям компетенций: Техничко-экономическое обоснование ИТ-проектов, Управление проектами

## 2. Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов обучения по программе

Изучение дисциплины Технологии Big Data в образовательной программе направлено на формирование у обучающихся следующих компетенций:

### Общепрофессиональные компетенции (ОПК):

ОПК-1 - Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности;

<b>Планируемые результаты обучения по дисциплине</b>			
Описание ИДК	Знать	Уметь	Владеть (иметь навыки)
ОПК-1 ИДК1 Знает основы математики, физики, вычислительной техники и программирования..	ОПК-1з1 Основы математического моделирования, теоретического и экспериментального исследования в профессиональной деятельности.	ОПК-1у1 Решать стандартные профессиональные задачи с применением теоретического и экспериментального исследования.	ОПК-1в1 Навыками теоретического и экспериментального исследования не стандартных объектов профессиональной деятельности.
ОПК-1 ИДК2 Умеет решать стандартные профессиональные задачи с применением естественнонаучных и обще-инженерных знаний, методов математического анализа и моделирования.	ОПК-1з2 Стандартные профессиональные задачи, методы математического анализа и моделирования.	ОПК-1у2 Использовать в профессиональной деятельности способности решать стандартные профессиональные задачи с применением естественнонаучных и обще-инженерных знаний.	ОПК-1в2 Приемами решения стандартных профессиональных задачи с применением естественнонаучных и обще-инженерных знаний.
ОПК-1 ИДК3 Владеет навыками теоретического и экспериментального исследования объектов профессиональной деятельности.	ОПК-1з3 Особенности теоретического и экспериментального исследования объектов алгоритмами профессиональной деятельности.	ОПК-1у3 Практически использовать.	ОПК-1в3 Приемами исследования объектов профессиональной деятельности.

### Профессиональные компетенции (ПК):

ПК УВ-2 - Способность планировать проект разработки информационной системы на основе его технико- экономического обоснования и технического задания

<b>Планируемые результаты обучения по дисциплине</b>			
Описание ИДК	Знать	Уметь	Владеть (иметь навыки)
ПК-УВ2 ИДК1 Разработка технико-экономического обоснования и технического задания.	ПК-УВ2з1 Методы, принципы и стандарты технико-экономического обоснования проектных решений и технического задания на разработку информационной системы.	ПК-УВ2у1 Составлять технико-экономическое обоснование проектных решений и техническое задание на разработку информационной системы .	ПК-УВ2в1 Навыками составления технико-экономического обоснования проектных решений и технических заданий на разработку информационной системы.
ПК-УВ2 ИДК2 Разработка иерархической структуры работ (ИСР) проекта в соответствии с техническим заданием.	ПК-УВ2з2 Стандарты разработки иерархической структуры работ (ИСР) в соответствии с техническим заданием.	ПК-УВ2у2 Составлять технико-экономическое обоснование разработки иерархической структуры работ (ИСР) в соответствии с техническим заданием.	ПК-УВ2в2 Навыками составления технико-экономического обоснования иерархической структуры работ в соответствии с техническим заданием.
ПК-УВ2 ИДК3 Подготовка текста плана управления проектом и частных планов в его составе (управления качеством, персоналом, рисками, стоимостью, содержанием, временем, субподрядчиками, закупками, изменениями, коммуникациями).	ПК-УВ2з3 Международные стандарты плана управления проектом и частных планов.	ПК-УВ2у3 Составлять планы управления качеством, персоналом, рисками, стоимостью, содержанием, временем, субподрядчиками, закупками, изменениями, коммуникациями.	ПК-УВ2в3 Навыками составления технико-экономического обоснования планов управления качеством, персоналом, рисками, стоимостью, содержанием, временем, субподрядчиками, закупками, изменениями, коммуникациями.

### 3. Объем и виды учебной работы

Учебным планом предусматриваются следующие виды учебной работы по дисциплине:

#### Очная форма обучения

Виды учебной работы	Всего час/ з.е.	
	Сем 5	Сем 6
Контактная работа, в том числе:	37.15/1.03	58.4/1.62
Занятия лекционного типа	18/0.5	28/0.78
Занятия семинарского типа, в том числе:	18/0.5	28/0.79
Лабораторные работы (лабораторный практикум)	18/0.5	14/0.39
Индивидуальная контактная работа (ИКР)	0.15/0	0.4/0.01
Групповая контактная работа (ГКР)	1/0.03	2/0.06
Самостоятельная работа, в том числе:	25.85/0.72	57.6/1.6
Промежуточная аттестация	9/0.25	28/0.78
Вид промежуточной аттестации:		
Экзамен, Зачет	Зач	Экз
Общая трудоемкость (объем части образовательной программы): Часы	72	144
Зачетные единицы	2	4

### заочная форма

Виды учебной работы	Всего час/ з.е.	
	Сем 7	Сем 8
Контактная работа, в том числе:	9.15/0.25	18.4/0.51
Занятия лекционного типа	4/0.11	8/0.22
Занятия семинарского типа	/0	4/0.11
Лабораторные работы (лабораторный практикум)	4/0.11	4/0.11
Индивидуальная контактная работа (ИКР)	0.15/0	0.4/0.01
Групповая контактная работа (ГКР)	1/0.03	2/0.06
Самостоятельная работа, в том числе:	60.85/1.69	118.6/3.29
Промежуточная аттестация	2/0.06	7/0.19
Вид промежуточной аттестации:		
Экзамен, Зачет	Зач	Экз
Общая трудоемкость (объем части образовательной программы): Часы	72	144
Зачетные единицы	2	4

### 4. Содержание дисциплины

#### 4.1. Разделы, темы дисциплины и виды занятий:

Тематический план дисциплины Технологии Big Data представлен в таблице.

#### Разделы, темы дисциплины и виды занятий Очная форма обучения

№ п/п	Наименование темы (раздела) дисциплины	Контактная работа					Самостоятельная работа	Планируемые результаты обучения в соотношении с результатами обучения по образовательной программе
		Лекции	Занятия семинарского типа		ИКР	ГКР		
			Практич. занятия	Лаборат. работы				
1.	Введение в Hadoop	18	0	18			40	ОПК-1_ИДК1 ОПК-1_ИДК2 ОПК-1_ИДК3 ПК-УВ2_ИДК1 ПК-УВ2_ИДК2 ПК-УВ2_ИДК3
2.	Технологии Big Data	28	14	14			43.45	ОПК-1_ИДК1 ОПК-1_ИДК2 ОПК-1_ИДК3 ПК-УВ2_ИДК1 ПК-УВ2_ИДК2 ПК-УВ2_ИДК3
	Контроль	37						
	<b>Итого</b>	<b>46</b>	<b>14</b>	<b>32</b>	<b>0.55</b>	<b>3</b>	<b>83.45</b>	

**заочная форма**

№ п/п	Наименование темы (раздела) дисциплины	Контактная работа					Самостоятельная работа	Планируемые результаты обучения в соотношении с результатами обучения по образовательной программе
		Лекции	Занятия семинарского типа		ИКР	ГКР		
			Практич. занятия	Лаборат. работы				
1.	Введение в Hadoop	6	2	4			89,45	ОПК-1_ИДК1 ОПК-1_ИДК2 ОПК-1_ИДК3 ПК-УВ2_ИДК1 ПК-УВ2_ИДК2 ПК-УВ2_ИДК3
2.	Технологии Big Data	6	2	4			90	ОПК-1_ИДК1 ОПК-1_ИДК2 ОПК-1_ИДК3 ПК-УВ2_ИДК1 ПК-УВ2_ИДК2 ПК-УВ2_ИДК3
	Контроль	9						
	<b>Итого</b>	<b>12</b>	<b>4</b>	<b>8</b>	<b>0.55</b>	<b>3</b>	<b>179.45</b>	

**4.2 Содержание разделов и тем**

**4.2.1 Контактная работа**

**Тематика занятий лекционного типа**

№п/п	Наименование темы (раздела) дисциплины	Вид занятия лекционного типа*	Тематика занятия лекционного типа
1.	Введение в Hadoop	лекция	Что такое Hadoop
		лекция	Распределенная файловая система HDFS
		лекция	Архитектура HDFS
		лекция	MapReduce. Введение
		лекция	Java API
		лекция	NoSQL базы данных
		лекция	Решение задач с помощью MapReduce
		лекция	Решение задач с помощью MapReduce
2.	Технологии Big Data	лекция	Проблемы MR-алгоритмов на графах
		лекция	Pig
		лекция	Основные операторы PigLatin
		лекция	Hive
		лекция	NoSQL базы данных
		лекция	Введение в HBase
		лекция	Архитектура HBase
		лекция	Cassandra
		лекция	Основные понятия Spark
		лекция	Фреймворк Spark

	лекция	YARN Введение
	лекция	Компоненты YARN
	лекция	MapReduce 2.0
	лекция	Итоговая лекция

\*лекции и иные учебные занятия, предусматривающие преимущественную передачу учебной информации педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях, обучающимся

#### Тематика занятий семинарского типа

№п/п	Наименование темы (раздела) дисциплины	Вид занятия семинарского типа**	Тематика занятия семинарского типа
1.	Введение в Hadoop	лабораторные работы	Основы Hadoop
		лабораторные работы	Распределенная файловая система HDFS
		лабораторные работы	Архитектура HDFS
		лабораторные работы	Основы MapReduce
		лабораторные работы	Основы Java API
		лабораторные работы	Основы NoSQL
		лабораторные работы	Решение задач с помощью MapReduce
		лабораторные работы	Решение задач с помощью MapReduce
		лабораторные работы	Алгоритмы на графах в MapReduce
2.	Технологии Big Data	практические занятия	Проблемы MR-алгоритмов на графах
		практические занятия	Pig
		лабораторные работы	Основные операторы PigLatin
		практические занятия	Hive
		лабораторные работы	NoSQL базы данных
		практические занятия	Введение в HBase
		лабораторные работы	Архитектура HBase
		лабораторные работы	Основы Cassandra
		лабораторные работы	Основы Spark
		лабораторные работы	Фреймворк Spark
		практические занятия	Основы YARN
		лабораторные работы	Компоненты YARN
		практические занятия	MapReduce 2.0
		практические занятия	Итоговое занятие

\*\* семинары, практические занятия, практикумы, лабораторные работы, коллоквиумы и иные аналогичные занятия

#### Иная контактная работа

При проведении учебных занятий СГЭУ обеспечивает развитие у обучающихся навыков командной работы, межличностной коммуникации, принятия решений, лидерских качеств (включая при необходимости проведение интерактивных лекций, групповых дискуссий, ролевых игр, тренингов, анализ ситуаций и имитационных моделей, преподавание дисциплин (модулей) в форме курсов, составленных на основе результатов научных исследований, проводимых организацией, в

том числе с учетом региональных особенностей профессиональной деятельности выпускников и потребностей работодателей).

Формы и методы проведения иной контактной работы приведены в Методических указаниях по основной профессиональной образовательной программе.

#### 4.2.2 Самостоятельная работа

№п/п	Наименование темы (раздела) дисциплины	Вид самостоятельной работы ***
1.	Введение в Hadoop	- подготовка доклада - тестирование
2.	Технологии Big Data	- подготовка доклада - тестирование

\*\*\* самостоятельная работа в семестре, написание курсовых работ, докладов, выполнение контрольных работ

### 5. Материально-техническое и учебно-методическое обеспечение дисциплины

#### 5.1 Литература:

##### Основная литература

Воронова Л.И. Big Data. Методы и средства анализа [Электронный ресурс]: Учебное пособие/ Воронова Л.И., Воронов В.И.— Электрон. текстовые данные.— М.: Московский технический университет связи и информатики, 2016.— 33 с.— Режим доступа: <http://www.bibliocomplectator.ru/book/?id=61463>.— «БИБЛИОКОМПЛЕКТАТОР», по паролю

##### Дополнительная литература

Воронов В.И. Data Mining - технологии обработки больших данных [Электронный ресурс]: Учебное пособие/ Воронов В.И., Воронова Л.И., Усачев В.А.— Электрон. текстовые данные.— М.: Московский технический университет связи и информатики, 2018.— 47 с.— Режим доступа: <http://www.bibliocomplectator.ru/book/?id=81324>.— «БИБЛИОКОМПЛЕКТАТОР», по паролю

#### 5.2. Перечень лицензионного программного обеспечения

1. Microsoft Windows 10 Education / Microsoft Windows 7 / Windows Vista Business
2. Office 365 ProPlus, Microsoft Office 2019, Microsoft Office 2016 Professional Plus (Word, Excel, Access, PowerPoint, Outlook, OneNote, Publisher) / Microsoft Office 2007 (Word, Excel, Access, PowerPoint)
3. R (R Licenses)
4. RStudio (AGPL v3)

#### 5.3 Современные профессиональные базы данных, к которым обеспечивается доступ обучающихся

1. Профессиональная база данных «Информационные системы Министерства экономического развития Российской Федерации в сети Интернет» (Портал «Официальная Россия» - <http://www.gov.ru/>)
2. Профессиональная база данных «Финансово-экономические показатели Российской Федерации» (Официальный сайт Министерства финансов РФ - <https://www.minfin.ru/ru/>)
3. Профессиональная база данных «Официальная статистика» (Официальный сайт Федеральной службы государственной статистики - <http://www.gks.ru/>)

#### 5.4. Информационно-справочные системы, к которым обеспечивается доступ обучающихся

1. Справочно-правовая система «Консультант Плюс»
2. Справочно-правовая система «ГАРАНТ-Максимум».

#### 5.5. Специальные помещения

Учебные аудитории для проведения занятий лекционного типа	Комплекты ученической мебели Мультимедийный проектор Доска Экран
-----------------------------------------------------------	---------------------------------------------------------------------------



Учебные аудитории для проведения практических занятий (занятий семинарского типа)	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Учебные аудитории для групповых и индивидуальных консультаций	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Учебные аудитории для текущего контроля и промежуточной аттестации	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Помещения для самостоятельной работы	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Помещения для хранения и профилактического обслуживания оборудования	Комплекты специализированной мебели для хранения оборудования

Для проведения занятий лекционного типа используются демонстрационное оборудование и учебно-наглядные пособия в виде презентационных материалов, обеспечивающих тематические иллюстрации.

### 5.6 Лаборатории и лабораторное оборудование

Лаборатория информационных технологий в профессиональной деятельности	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ Лабораторное оборудование
-----------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------

## 6. Фонд оценочных средств по дисциплине Технологии Big Data:

### 6.1. Контрольные мероприятия по дисциплине

Вид контроля	Форма контроля	Отметить нужное знаком « + »
Текущий контроль	Оценка докладов	+
	Устный/письменный опрос	-
	Тестирование	+
	Практические задачи	-
	Оценка контрольных работ (для заочной формы обучения)	-
Промежуточный контроль	Зачет	+
	Экзамен	+

Порядок проведения мероприятий текущего и промежуточного контроля определяется

**Методическими указаниями по основной профессиональной образовательной программе высшего образования, утвержденными Ученым советом ФГБОУ ВО СГЭУ №10 от 29.04.2020г.**

**6.2. Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов обучения по программе**

**Общепрофессиональные компетенции (ОПК):**

ОПК-1 - Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности;

<b>Планируемые результаты обучения по дисциплине</b>				
Описание ИДК	Уровень сформированности	Знать	Уметь	Владеть (иметь навыки)
ОПК-1 ИДК1 Знает основы математики, физики, вычислительной техники и программирования..	Пороговый	ОПК-1з1 Основы математического моделирования, теоретического и экспериментального исследования в профессиональной деятельности.	ОПК-1у1 Решать стандартные профессиональные задачи с применением теоретического и экспериментального исследования.	ОПК-1в1 Навыками теоретического и экспериментального исследования не стандартных объектов профессиональной деятельности.
ОПК-1 ИДК2 Умеет решать стандартные профессиональные задачи с применением естественнонаучных и обще-инженерных знаний, методов математического анализа и моделирования.	Базовый	ОПК-1з2 Стандартные профессиональные задачи, методы математического анализа и моделирования.	ОПК-1у2 Использовать в профессиональной деятельности способности решать стандартные профессиональные задачи с применением естественнонаучных и обще-инженерных знаний.	ОПК-1в2 Приемами решения стандартных профессиональных задачи с применением естественнонаучных и обще-инженерных знаний.
ОПК-1 ИДК3 Владеет навыками теоретического и экспериментального исследования объектов профессиональной деятельности.	Повышенный	ОПК-1з3 Особенности теоретического и экспериментального исследования объектов алгоритмами профессиональной деятельности.	ОПК-1у3 Практически использовать.	ОПК-1в3 Приемами исследования объектов профессиональной деятельности.

**Профессиональные компетенции (ПК):**

ПК УВ-2 - Способность планировать проект разработки информационной системы на основе его технико- экономического обоснования и технического задания

<b>Планируемые результаты обучения по дисциплине</b>				
Описание ИДК	Уровень сформированности	Знать	Уметь	Владеть (иметь навыки)

ПК-УВ2 ИДК1 Разработка технико-экономического обоснования и технического задания.	Пороговый	ПК-УВ2з1 Методы, принципы и стандарты технико-экономического обоснования проектных решений и технического задания на разработку информационной системы.	ПК-УВ2у1 Составлять технико-экономическое обоснование проектных решений и техническое задание на разработку информационной системы.	ПК-УВ2в1 Навыками составления технико-экономического обоснования проектных решений и технических заданий на разработку информационной системы.
ПК-УВ2 ИДК2 Разработка иерархической структуры работ (ИСР) проекта в соответствии с техническим заданием.	Базовый	ПК-УВ2з2 Стандарты разработки иерархической структуры работ (ИСР) в соответствии с техническим заданием.	ПК-УВ2у2 Составлять технико-экономическое обоснование разработки иерархической структуры работ (ИСР) в соответствии с техническим заданием.	ПК-УВ2в2 Навыками составления технико-экономического обоснования иерархической структуры работ в соответствии с техническим заданием.
ПК-УВ2 ИДК3 Подготовка текста плана управления проектом и частных планов в его составе (управления качеством, персоналом, рисками, стоимостью, содержанием, временем, субподрядчиками, закупками, изменениями, коммуникациями).	Повышенный	ПК-УВ2з3 Международные стандарты плана управления проектом и частных планов.	ПК-УВ2у3 Составлять планы управления качеством, персоналом, рисками, стоимостью, содержанием, временем, субподрядчиками, закупками, изменениями, коммуникациями.	ПК-УВ2в3 Навыками составления технико-экономического обоснования планов управления качеством, персоналом, рисками, стоимостью, содержанием, временем, субподрядчиками, закупками, изменениями, коммуникациями.

### 6.3. Паспорт оценочных материалов

№ п/п	Наименование темы (раздела) дисциплины	Контролируемые планируемые результаты обучения в соотношении с результатами обучения по программе	Вид контроля/используемые оценочные средства	
			Текущий	Промежуточный
1.	Введение в Nadoop	ОПК-1_ИДК1 ОПК-1_ИДК2 ОПК-1_ИДК3	Оценка докладов Тестирование	Зачет

		ПК-УВ2_ИДК1 ПК-УВ2_ИДК2 ПК-УВ2_ИДК3		
2.	Технологии Big Data	ОПК-1_ИДК1 ОПК-1_ИДК2 ОПК-1_ИДК3 ПК-УВ2_ИДК1 ПК-УВ2_ИДК2 ПК-УВ2_ИДК3	Оценка докладов Тестирование	Экзамен

#### 6.4.Оценочные материалы для текущего контроля

##### Примерная тематика докладов

Раздел дисциплины	Темы
Введение в Hadoop	<ol style="list-style-type: none"> <li>1. Принципы обработки больших данных</li> <li>2. Основные технологии Big Data</li> <li>3. Hadoop Common</li> <li>4. Масштабируемость Hadoop</li> <li>5. Hadoop HDFS</li> <li>6. Hadoop YARN</li> <li>7. Экосистема Hadoop</li> </ol>
Технологии Big Data	<ol style="list-style-type: none"> <li>8. Медицина и BigData</li> <li>9. BigData в социологии</li> <li>10. BigData в маркетинге</li> <li>11. BigData. Развитие, анализ и технологии</li> <li>12. Big Data и Data Science</li> <li>13. Большие свойства «маленьких данных»</li> <li>14. Возможности и проблемы, возникающие при анализе «больших данных»</li> <li>15. Большие Данные в национальной экономике</li> <li>16. Преимущества для бизнеса с BigData</li> <li>17. Технологии BigData в Образовании</li> <li>18. Кейсы использования анализа данных в операционной деятельности</li> </ol>

##### Задания для тестирования по дисциплине для оценки сформированности компетенций

<https://lms2.sseu.ru/course/index.php?categoryid=514>

Отметьте верные высказывания о принципах Hadoop:

Вертикальное масштабирование

Инкапсуляция сложности реализации распределенной системы

Горизонтальное масштабирование

Отправка кода к данным

Отказоустойчивость

Перемещение данных к коду

Какие функции выполняют вендоры дистрибутивов?

Разрабатывают дополнительные фичи в стандартных компонентах Hadoop

Обеспечивают совместимость разных компонентов Hadoop

Предоставляют дистрибутивы в различных форматах (rpm, tar.gz, образ виртуальных машин)

Исправляют ошибки в компонентах Hadoop

Как осуществляется доступ к данным на чтение в HDFS?

Произвольный доступ к нужной части файла

Последовательное чтение всего файла с данными

Как можно менять данные в файле HDFS?

Удалять данные в конце файла

Удалять произвольную часть файла

Дописывать данные в конец файла

Удалять все данные (удалять файл)

Дописывать данные в произвольное место файла (начало, середина, конец)

Для какого шаблона взаимодействия с данными был разработан HDFS?

Много раз поменяли данные / один раз прочитали

Один раз записали / один раз прочитали

Один раз записали / много раз прочитали

Много раз поменяли данные / много раз прочитали

Чем больше размер блока HDFS, тем...

Тем меньше файлов в HDFS можно хранить

Тем больше файлов в HDFS можно хранить

Как осуществляется доступ к HDFS?

Только прямой доступ (Direct Access)

Прямой доступ или через проху-сервер

Только через проху-сервер

Какие предложения характеризуют HDFS?

Данные не теряются, если выходят из строя диски/сервера

Является приложением пользовательского уровня ОС

Высокая пропускная способность вместо быстрого доступа к данным

Несколько процессов могут писать данные в один файл HDFS

Работает на кластере из суперкомпьютеров

Лучше хранить миллиард небольших файлов, чем миллион больших

Какие варианты соответствуют URI схеме и могут быть использованы в командах shell-клиента HDFS?

/tmp/output.txt

file:///home/user/

hdfs:data.txt

hdfs://localhost:9000//data/log.txt

hdfs://hdfs/ file:///

Как лучше скопировать файл file.txt размером 10Тб из папки /data/ в папку /tmp/

```
$hdfs distcp /data/file.txt /tmp/
```

```
$hdfs dfs -get /data/file.txt . ; hdfs dfs -put file.txt /tmp/
```

```
$hdfs dfs -cp /data/file.txt /tmp/
```

Как лучше скопировать файл file.txt размером 10Мб из папки /data/ в папку /tmp/

```
$hdfs distcp /data/file.txt /tmp/
```

```
$hdfs dfs -cp /data/file.txt /tmp/
```

```
$hdfs dfs -get /data/file.txt . ; hdfs dfs -put file.txt /tmp/
```

Что означает, если файл имеет такие права доступа: -rw-r-----

Владелец и пользователи из той же группы могут читать и писать в файл, другие могут только читать из файла

Все пользователи могут читать и писать в файл

Владелец может читать и писать в файл, пользователи из той же группы могут читать файл, другие ничего не могут делать с файлом

Только владелец может читать и писать в файл. Остальные пользователи не могут ни читать, ни писать в файл

Какие реализации абстрактного класса FileSystem существуют в HDFS API?

FTPFileSystem  
LocalFileSystem  
HftpFileSystem  
DistributedFileSystem

В каком виде осуществляется передача данных между Map и Reduce?

Массив байт  
Ключ / Значение  
В виде xml  
Можно самому определять формат  
Текстовые сообщения

Как определяется число mapper-процессов в MapReduce задаче?

Задается пользователем при описании MapReduce задачи  
Обычно равно числу сплитов во входных данных

Что такое data locality?

Запуск mapper на той же ноде, где хранятся нужные данные  
Процесс определения, на какой ноде расположены нужные данные  
Процесс копирования нужных данных на ноду с запущенным mapper

Как хранятся промежуточные данные между фазами Map и Reduce?

На локальном диске  
В HDFS

Что происходит, если выходит из строя диск с промежуточными данными между Map и Reduce?

Данные восстанавливаются из других реплик  
Перезапускается mapper, который создал эти данные  
Задача завершается с ошибкой, т.к. потеряны промежуточные данные

Что такое shuffle?

Определение номера редьюсера для каждой пары Ключ-Значение в промежуточных данных  
Процесс перемешивания промежуточных данных для равномерного распределения по reducer'ам  
Сортировка промежуточных данных между Map и Reduce  
Процесс передачи промежуточных данных из mapper'ов в reducer

Как определяется число reducer-процессов в MapReduce задаче?

Определяется системой, исходя из размера промежуточных данных между Map и Reduce  
Задается пользователем при описании MapReduce задачи

## 6.5. Оценочные материалы для промежуточной аттестации

### Фонд вопросов для проведения промежуточного контроля в форме зачета

Раздел дисциплины	Вопросы
Введение в Hadoop	<ol style="list-style-type: none"><li>1. Что такое Hadoop</li><li>2. Распределенная файловая система HDFS</li><li>3. Архитектура HDFS</li><li>4. Shell-команды</li><li>5. Java API</li><li>6. Основные понятия MapReduce</li><li>7. Парадигма MapReduce</li><li>8. Объясните наиболее простым способом принцип работы MapReduce.</li><li>9. Фреймворк MapReduce</li></ol>

	<ol style="list-style-type: none"> <li>10. Hadoop Streaming</li> <li>11. Алгоритмы на MapReduce</li> <li>12. Реляционные функции</li> <li>13. Расчет TF-IDF</li> <li>14. Графы в MapReduce</li> <li>15. Поиск кратчайшего пути в графе</li> <li>16. PageRank</li> <li>17. Что такое shuffle</li> <li>18. Проблемы MR-алгоритмов на графах</li> <li>19. Примеры решения задач с помощью MapReduce</li> <li>20. Алгоритмы на графах в MapReduce</li> <li>21. Два основных компонента структуры Hadoop.</li> <li>22. Как бы вы отсортировали большой массив чисел?</li> <li>23. Перед вами — большой набор данных. Как вы будете работать с выбросами? Как поступите с недостающими значениями? Что скажете о преобразованиях?</li> <li>24. mapper-процессы в MapReduce</li> <li>25. reducer-процессы в MapReduce</li> <li>26. HDFS API</li> <li>27. Права доступа HDFS</li> <li>28. Data locality</li> <li>29. NoSQL базы данных</li> <li>30. Hadoop Streaming</li> </ol>
--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Фонд вопросов для проведения промежуточного контроля в форме экзамена**

<b>Раздел дисциплины</b>	<b>Вопросы</b>
Введение в Hadoop	<ol style="list-style-type: none"> <li>1. Распределенная файловая система HDFS</li> <li>2. Архитектура HDFS</li> <li>3. Shell-команды</li> <li>4. Java API</li> <li>5. Основные понятия MapReduce</li> <li>6. Фреймворк MapReduce</li> <li>7. Hadoop Streaming</li> <li>8. Расчет TF-IDF</li> <li>9. Графы в MapReduce</li> <li>10. Проблемы MR-алгоритмов на графах</li> <li>11. Примеры решения задач с помощью MapReduce</li> <li>12. Алгоритмы на графах в MapReduce</li> <li>13. mapper-процессы в MapReduce</li> <li>14. reducer-процессы в MapReduce</li> <li>15. HDFS API</li> <li>16. Права доступа HDFS</li> <li>17. Data locality</li> <li>18. NoSQL базы данных</li> <li>19. Hadoop Streaming</li> </ol>
Технологии Big Data	<ol style="list-style-type: none"> <li>20. Pig</li> <li>21. Основные операторы PigLatin</li> <li>22. Hive</li> <li>23. Парадигма MapReduce</li> <li>24. Когда следует использовать MapReduce вместо Pig/Hive?</li> <li>25. NoSQL базы данных</li> <li>26. Введение в HBase</li> <li>27. Архитектура HBase</li> <li>28. Основы Cassandra</li> <li>29. Основы Spark</li> </ol>

	30. Фреймворк Spark 31. Spark Streaming 32. File System API в Spark 33. SparkCore 34. Основы YARN 35. Компоненты YARN 36. MapReduce 2.0
--	-----------------------------------------------------------------------------------------------------------------------------------------------------------

**6.6. Шкалы и критерии оценивания по формам текущего контроля и промежуточной аттестации**

**Шкала и критерии оценивания**

<b>Оценка</b>	<b>Критерии оценивания для мероприятий контроля с применением 2-х балльной системы</b>
<b>«зачтено»</b>	ОПК-1_ИДК1, ПК-УВ2_ИДК1
<b>«не зачтено»</b>	Результаты обучения не сформированы на пороговом уровне

<b>Оценка</b>	<b>Критерии оценивания для мероприятий контроля с применением 4-х балльной системы</b>
<b>«отлично»</b>	ОПК-1_ИДК3, ПК-УВ2_ИДК3
<b>«хорошо»</b>	ОПК-1_ИДК2, ПК-УВ2_ИДК2
<b>«удовлетворительно»</b>	ОПК-1_ИДК1, ПК-УВ2_ИДК1
<b>«неудовлетворительно»</b>	Результаты обучения не сформированы на пороговом уровне