

МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Нижегородский государственный технический университет
им. Р.Е. Алексеева» (НГТУ)

Институт радиоэлектроники и информационных технологий (ИРИТ)

(Полное и сокращенное название института, реализующего данное направление)

УТВЕРЖДАЮ:

Первый проректор – проректор по
образовательной деятельности:

_____ Ивашкин Е.Г.
подпись ФИО

“22” ИЮНЯ 2022 г.

УТВЕРЖДАЮ:

Директор института:

_____ Мякинников А.В.
подпись ФИО

“21” ИЮНЯ 2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Технологии обработки больших данных

(индекс и наименование дисциплины по учебному плану)

для подготовки магистров

Направление подготовки: 09.04.01 Информатика и вычислительная техника

Направленность: Искусственный интеллект в автоматизированных системах обработки информации и управления

Форма обучения: очная

Год начала подготовки 2022

Выпускающая кафедра ВСТ

Объем дисциплины 108 / 3
часов/з.е

Промежуточная аттестация экзамен

Автор программы:

Сухобоков А.А., к.т.н., доцент, artem.sukhobokov@yandex.ru

Ведущий преподаватель НГТУ: Киров Д.И., ассистент

Рабочая программа дисциплины: разработана в соответствии с Федеральным государственным образовательным стандартом высшего образования (ФГОС ВО 3++) по направлению подготовки 09.03.01 Информатика и вычислительная техника, утвержденного приказом МИНОБРНАУКИ РОССИИ от 19 сентября 2017 года № 918 на основании учебного плана принятого УМС НГТУ

протокол от 23.12.2021 № 6

Рабочая программа одобрена на заседании кафедры ВСТ протокол от 08.06.2022 № 8

Зав. кафедрой д.т.н, доцент, Жевнерчук Д.В. _____
(подпись)

Программа рекомендована к утверждению ученым советом института ИРИТ, Протокол от 21.06.2022 № 6

Рабочая программа зарегистрирована в УМУ, регистрационный № 09.04.01-ИИ-22

Начальник УМУ _____ Т.И. Ермакова

Оглавление

1. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ), СООТНЕСЕННЫЕ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.....	4
2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	7
3. ОБЪЕМ ДИСЦИПЛИНЫ	8
4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ, СТРУКТУРИРОВАННОЕ ПО МОДУЛЯМ С УКАЗАНИЕМ ОТВЕДЕННОГО НА НИХ КОЛИЧЕСТВА АКАДЕМИЧЕСКИХ ИЛИ АСТРОНОМИЧЕСКИХ ЧАСОВ И ВИДОВ УЧЕБНЫХ ЗАНЯТИЙ	9
5. ПЕРЕЧЕНЬ УЧЕБНО-МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ	12
6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ СТУДЕНТОВ ПО ДИСЦИПЛИНЕ	13
7. ПЕРЕЧЕНЬ ОСНОВНОЙ И ДОПОЛНИТЕЛЬНОЙ УЧЕБНОЙ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	14
8. ПЕРЕЧЕНЬ РЕСУРСОВ СЕТИ ИНТЕРНЕТ, РЕКОМЕНДУЕМЫХ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ПРИ ОСВОЕНИИ ДИСЦИПЛИНЫ.....	26
9. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ СТУДЕНТОВ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ.....	35
10. ПЕРЕЧЕНЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, ИСПОЛЬЗУЕМЫХ ПРИ ИЗУЧЕНИИ ДИСЦИПЛИНЫ, ВКЛЮЧАЯ ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ	36
11. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ИЗУЧЕНИЯ ДИСЦИПЛИНЫ	37

1.ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ), СООТНЕСЕННЫЕ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Настоящая рабочая программа дисциплины устанавливает требования к знаниям и умениям студента, а также определяет содержание и виды учебных занятий и отчетности.

Программа разработана в соответствии с:

- Федеральным государственным образовательным стандартом (ФГОС 3++) по направлению подготовки (уровень магистратуры): 09.04.01 «Информатика и вычислительная техника»;
- Основной профессиональной образовательной программой по направлению подготовки 09.04.01 «Информатика и вычислительная техника»;
- Моделью компетенций в сфере искусственного интеллекта;
- Учебным планом МГТУ им. Н.Э. Баумана по направлению подготовки 09.04.01 «Информатика и вычислительная техника».

При освоении дисциплины планируется формирование компетенций, предусмотренных ОПОП на основе ФГОС 3++ по направлению подготовки 09.04.01 «Информатика и вычислительная техника» (уровень магистратуры).

Код компетенции по ФГОС 3++	Формулировка компетенции
	Универсальные компетенции
УК-6	Способность определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки
	Профессиональные компетенции
ПК-4	Способен руководить проектами по созданию, поддержке и использованию системы искусственного интеллекта на основе нейросетевых моделей и методов

Для категорий «знать, уметь, владеть» планируется достижение следующих результатов обучения (РО), вносящих на соответствующих уровнях вклад в формирование компетенций, предусмотренных основной профессиональной образовательной программой (табл. 1).

Таблица 1. Индикаторы достижения компетенции

Компетенция: код по ФГОС 3++, формулировка	Индикаторы	Формы и методы обучения, способствующие формированию и развитию компетенции
1	2	3
УК – 6 (09.04.01) Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	ЗНАТЬ - методики самооценки, самоконтроля и саморазвития, в том числе с использованием подходов здоровьесбережения УМЕТЬ - решать задачи собственного личностного и профессионального развития, определять и реализовывать приоритеты совершенствования собственной деятельности	• Лекции • Семинары • Лабораторные работы • Самостоятельная работа Активные и интерактивные формы (методы) обучения: обсуждение практических примеров на семинарах; - совместный анализ результатов рубежных контролей и выполнения домашних заданий в форме дискуссии; - обсуждение во время лабораторных работ различных вариантов выполнения задания, сравнение и обсуждение полученных результатов.
ПК – 4 Способен руководить проектами по созданию, поддержке и использованию системы искусственного интеллекта на основе нейросетевых моделей и методов	ПК-4.1. Руководит работами по оценке и выбору моделей искусственных нейронных сетей и инструментальных средств для решения поставленной задачи ЗНАТЬ - возможности современных инструментальных средств и систем программирования для решения задач машинного обучения.	• Лекции • Семинары • Лабораторные работы • Самостоятельная работа Активные и интерактивные формы (методы) обучения: - обсуждение на лекциях и семинарах состава и функциональности проприетарных и с открытым кодом платформ обработки больших данных и дополняющих их пакетов; - обсуждение на семинарах технологической архитектуры корпоративных систем обработки данных, методов её развёртывания и требуемых аппаратных ресурсов; - обсуждение во время лабораторных работ

Компетенция: код по ФГОС 3++, формулировка	Индикаторы	Формы и методы обучения, способствующие формированию и развитию компетенции
		различных вариантов выполнения задания, сравнение и обсуждение полученных результатов.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина входит в дисциплины по выбору вариативной части блока Б1 «Дисциплины (модули)» образовательной программы магистратуры по направлению подготовки 09.04.01 «Информатика и вычислительная техника».

Изучение дисциплины предполагает предварительное освоение следующих дисциплин учебного плана:

- Теоретические основы реляционной алгебры;
- Надежность и достоверность автоматизированных систем обработки информации и управления;
- Постреляционные базы данных.

Освоение данной дисциплины необходимо как предшествующее для следующей дисциплины образовательной программы:

- Подготовка и защита ВКР.

Освоение учебной дисциплины связано с формированием компетенций с учетом матрицы компетенций ОПОП по направлению подготовки 09.04.01 «Информатика и вычислительная техника».

3. ОБЪЕМ ДИСЦИПЛИНЫ

Общий объем дисциплины составляет Зачётные единицы (з.е.), 108 ч. В том числе: 1 семестр - 3з.е. (108 ч.).

Таблица 2. Объём дисциплины по видам учебных занятий (в часах)

Виды учебной работы	Объём по семестрам, ч	
	Всего	1 семестр
Объём дисциплины	108	108
Аудиторная работа	44	44
Лекции	11	11
Лабораторные работы (ЛР)	33	33
Самостоятельная работа (СР)	64	64
Проработка материала лекций	2	2
Подготовка к лабораторным работам	14	14
Подготовка к рубежным контролям	6	6
Выполнение домашних заданий	12	12
Подготовка к экзамену	30	30
Вид промежуточной аттестации		экзамен

4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ, СТРУКТУРИРОВАННОЕ ПО МОДУЛЯМ С УКАЗАНИЕМ ОТВЕДЕННОГО НА НИХ КОЛИЧЕСТВА АКАДЕМИЧЕСКИХ ИЛИ АСТРОНОМИЧЕСКИХ ЧАСОВ И ВИДОВ УЧЕБНЫХ ЗАНЯТИЙ

Таблица 3. Содержание дисциплины

№ п/п	Тема (название) модуля	Виды занятий, часы			Активные и интерактивные формы проведения занятий		Компетенции, закрепленные за темой (код по ФГОС 3++)	Текущий контроль результатов обучения		
		Л	ЛР	СР	Форма проведения занятий	Ча- сы		Срок (неделя)	Формы	Баллы (мин/ макс)
1 семестр										
1	Базовые понятия сис- тем обработки боль- ших данных, потоковая обработка данных, интернет вещей	7	17	18	- анализ результатов выполнения домашних заданий и лабораторных работ; - обсуждение во время ла- бораторных работ различ- ных вариантов выполне- ния задания, сравнение и обсуждение полученных результатов.	5	УК - 6 ПК - 4	7	Рубежный контроль №1	6 / 10
									Защита лабораторных работ	9 / 15
									Домашнее задание №1	6 / 10
									ИТОГО	21 / 35
2	NoSQL- и SQL-систе- мы баз данных, ис- пользующие шардиро- вание, интегрирован- ные платформы обработки больших данных, озёра данных, графовые энджины	4	16	16		5		11	Рубежный контроль №2	6 / 10
									Защита лабораторных работ	9 / 15
									Домашнее задание №2	6 / 10
									ИТОГО	21 / 35
3	Экзамен	-	-	30	-	-	-	-	18/ 30	
	ИТОГО за семестр	11	33	64	-	10	-	-	-	60 / 100

Содержание дисциплины, структурированное по модулям.

№ п/п	Наименование модуля дисциплины Содержание
1.	БАЗОВЫЕ ПОНЯТИЯ СИСТЕМОБРАБОТКИ БОЛЬШИХ ДАННЫХ <i>Объем: 42 часа</i>
	Лекции <i>Объем: 7 часов</i>
Л1.1.	Роль дисциплины «Технологии обработки больших данных». Понятие Big Data. Задачи Big Data. Возникновение технологий Big Data. MPI. Распределённые файловые системы для хранения больших объёмов данных. NoSQL Data Bases. Hadoop. История разработки Hadoop. Принципы работы Hadoop. Современное состояние Hadoop. Экосистема Hadoop. Инструменты Big Data - Платформы на основе дистрибутивов Hadoop. Hortonworks Data Platform. Cloudera Enterprise. MapR. Pivotal. Инструменты Big Data - Интегрированные платформы. Возникновение и функциональность интегрированных платформ Big Data. GridGain In-Memory Data Fabric. Data Lake – основной способ организации больших данных. Архитектурные решения с использованием инструментов Big Data. Использование инструментов Big Data в интернет-компаниях. Лямбда-архитектура для BI-проектов. Подход и рекомендуемая архитектура SAP для использования инструментов Big Data в составе корпоративных системных ландшафтов. Новые направления, возникающие в результате применения и дальнейшего развития инструментария Big Data, в научных дисциплинах, использующих моделирование. Big Calculation. Big Simulation. Big Management. Big Optimal Control. - 3 часа
Л1.2.	Детальная схема работы MapReduce. YARN. HDFS. Apache Avro. Apache Pig. Apache Oozie. Apache Tez. Платформа OpenStack. OpenStack Sahara. Платформа Microsoft Azure. Платформа IBM Bluemix. Экосистема SAP. Платформа SAP HANA. Amazon EC2. Apache Spark. Apache Flink. - 2 часа
Л1.3.	Средства потоковой обработки данных. Apache Storm. Trident поверх Storm. Apache Kafka. Apache Flume. Spark Streaming. Потоковая обработка в Apache Flink. Apache Beam. MapR Streams. SAP HANA Smart Data Streaming. Декларативные языки для обработки потоковых данных. SAP HANA CCL. Интернет вещей. Архитектура приложений IoT. Apache NiFi. Microsoft Azure IoT Reference Architecture. SAP Leonardo. - 2 часа
	Лабораторные работы <i>Объем: 17 часов</i>
ЛР1.1.	Знакомство с Hadoop - 4 часа
ЛР1.2.	Знакомство со Spark. - 4 часа
ЛР1.3.	Знакомство с Kubernetes и контейнеризацией вычислений. - 5 часов
ЛР1.4.	Сложная обработка больших данных в экосистеме Hadoop. - 4 часа
	Самостоятельная работа студентов (СР) <i>Объем: 18 часов</i>
СР1.1.	Проработка материала лекций – 1 час
СР1.2.	Подготовка к лабораторным работам - 8 часов
СР1.3.	Подготовка к рубежному контролю – 3 часа
СР1.4.	Выполнение домашнего задания – 6 часов

2.	NOSQL- И SQL-СИСТЕМЫ БАЗ ДАННЫХ, ИСПОЛЬЗУЮЩИЕ ШАРДИРОВАНИЕ.Объем: 36 часов
	Лекции Объем: 4 часа
L2.1.	Базы данных NoSQL. Причины появления систем NoSQL баз данных и их роль в развитии технологий обработки больших данных. 4 этапа развития СУБД и 3 революции, произошедшие в ходе этого развития. Типы систем NoSQL баз данных и тенденции развития систем NoSQL баз данных. HBase. Hipertable. Accumulo. Cassandra. MongoDB. CouchDB какпредшественник Couchbase. RavenDB. Redis. Riak. Kudu. Введениев Search Engines. Apache Lucene. Apache Nutch и Apache Solr какразвитие Apache Lucene. Elasticsearch. Splunk.Графовые СУБД и граф-энджины. Общий обзор графовых СУБД. Neo4j. OrientDB. Titan.Общийобзор Graph Engines. Pregel. Giraph. Sufer, Signal/Collect и Mizan. Семействопродуктов GraphLab. PEGASUS. Apache Spark GraphX. СовмещениеграфовыхСУБДи Graph Engine водномпродукте. SAP HANA Graph. Microsoft Graph Engine. - 2 часа
L2.2.	SQL СУБД для работы с большими данными.История и тенденции развития SQL СУБД, способных обрабатывать большие данные.Couchbase N1QL похожий на SQL.HiveQL - почти как SQL.Couchbase. Apache Phoenix. Apache Drill. Cloudera Impala.Apache HAWQ. HP Vertica. IBM BigSQL.SAP HANA Vora.- 2 часа
	Лабораторные работыОбъем: 16 часов
LP2.1.	Знакомство сSAPHANA. - 4 часа
LР2.2.	Взаимодействие с SAPHANAчерез Web. Часть 1 - 4 часа Взаимодействие с SAPHANAчерез Web. Часть 2- 4 часа
LP2.3.	Обработка больших графов в MicrosoftGraphEngine. - 4 часа
	Самостоятельная работа студентов (СР)Объем: 16 часов
CP2.1.	Проработка материала лекций–1 час
CP2.2.	Подготовка к лабораторным работам –6 часов
CP2.3.	Подготовка к рубежному контролю – 3 часа
CP2.4.	Выполнение домашнегозадания – 6 часов
3.	ЭКЗАМЕН Объем: 30 часов
CP3.1.	Подготовка к экзамену – 30 часов

5. ПЕРЕЧЕНЬ УЧЕБНО-МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

Для обеспечения самостоятельной работы студентов по дисциплине сформирован методический комплекс, включающий следующие учебно-методических материалы.

1. Программа курса.
2. Учебники и учебные пособия [1-2].
3. Дополнительная литература.
4. Список адресов сайтов сети Интернет, содержащих актуальную информацию по теме дисциплины.

Материалы учебно-методического комплекса рассылаются студентам по электронной почте и через облачное дисковое пространство. Библиографические ссылки на учебные издания, входящие в методический комплекс, приведены в перечне основной и дополнительной учебной литературы, необходимой для освоения дисциплины (раздел 7).

К дополнительным материалам также относится перечень ресурсов сети интернет, рекомендуемых для самостоятельной работы при освоении дисциплины (раздел 8).

Студенты получают доступ к указанным материалам на первом занятии по дисциплине.

6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ СТУДЕНТОВ ПО ДИСЦИПЛИНЕ

Фонд оценочных средств (ФОС) для проведения текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине базируется на перечне компетенций с указанием этапов их формирования в процессе освоения образовательной программы (раздел 1). ФОС должен обеспечивать объективный контроль достижения всех результатов обучения, запланированных для дисциплины.

ФОС включает в себя:

- описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания;
- методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.
- типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и уровня овладения формирующимися компетенциями в процессе освоения дисциплины (перечень типовых вопросов для защиты лабораторных работ; перечни типовых тем для домашних заданий (реферат, доклад); перечень вопросов для рубежных контролей; примеры вариантов рубежных контролей, перечень экзаменационных вопросов; макет билета к экзамену).

Контроль освоения дисциплины производится в соответствии с Положением о текущем контроле успеваемости и промежуточной аттестации студентов МГТУ им. Н.Э. Баумана.

ФОС является приложением к данной программе дисциплины.

7. ПЕРЕЧЕНЬ ОСНОВНОЙ И ДОПОЛНИТЕЛЬНОЙ УЧЕБНОЙ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Основная литература

1. Лэм Ч. Hadoop в действии. – М.: ДМК Пресс, 2012. – 424 с.: ил. Режим доступа: <https://e.lanbook.com/reader/book/39997/#1> (дата обращения: 17.12.2016).
2. Карау Х., Конвински Э., Венделл П., Захария М. Изучаем Spark: молниеносный анализ данных. – М.: ДМК Пресс, 2015. – 304 с.: ил. Режим доступа: <https://e.lanbook.com/reader/book/90118/#1> (дата обращения: 17.12.2016).

Дополнительная литература

Материалы по теме «**Понятие больших данных и технологий обработки больших данных**»

1. Big Data Now. 2015-edition. O'Reilly Media, Inc., 2016. – 158 p.
2. Beyer D. The Future of Machine Intelligence. Perspectives from Leading Practitioners. – O'Reilly Media, Inc., 2016. – 78 p.
3. Overton J. Going Pro in Data Science. What It Takes to Succeed as a Professional Data Scientist. – O'Reilly Media, Inc., 2016. – 78 p.
4. Oram A. Managing the Data Lake. Moving to Big Data Analysis. – O'Reilly Media, Inc., 2015. – 24 p.
5. Russell J. Mapping Big Data. A Data-Driven Market Report. – O'Reilly Media, Inc., 2015. – 27 p.
6. Marz N., with Warren J. Big Data. Principles and best practices of scalable real-time data systems. – Manning Publications Co., 2015. – 330 p.

Материалы по теме «**Распределённые файловые системы**»

1. Dequenne F. Why HSMs still have a place at ECMWF Tape is dead! Or is it?, 32 p.
2. Duellmann D. LHC Exascale Storage Systems - Design and Implementation, 47 p.
3. Mehrotra S., Extreme capacity all-flash storage, in-memory performance. – Saratoga Speed, Inc., 2 June 2015, 9 p.
4. Пресс-релиз НОВЫЕ МАССИВЫ HITACHI ПОЛНОСТЬЮ НА ОСНОВЕ ФЛЭШ-ПАМЯТИ. – Hitachi Data Systems. 2 с.
5. Hellmold S. Emerging NVM - Enabling Next-generation Data Storage Solutions. – Western Digital Corporation, August 2014, 6 p.
6. Data-Driven Science. Advanced Storage for Genomics Workflows. – SGI, 2015, 22 p.
7. Fineberg S. Software Challenges for The Machine. – HP, June 2015, 27 p.
8. Zhang Y., Swanson S. A Study of Application Performance with Non-Volatile Main Memory, 10 p.
9. Zhang Y., Swanson S. A Study of Application Performance with Non-Volatile Main Memory, presentation, 34 p.
10. Lu Y., Shu J., Sun L. Blurred Persistence in Transactional Persistent Memory, 13 p.
11. Lu Y., Shu J., Sun L. Blurred Persistence in Transactional Persistent Memory, presentation, 26 p.
12. Sehgal P., Basu S., Srinivasan K., Voruganti K. An Empirical Study of File Systems on NVM, 14 p.
13. Sehgal P., Basu S., Srinivasan K., Voruganti K. An Empirical Study of File Systems on NVM, presentation, 35 p.
14. Lakshman A. Lessons Learned from Distributed Systems and Applied to Modern Storage Platforms. – Hedvig, 11 p.
15. Carter E. Modern Storage for Modern Business. – Hedvig, 13 p.
16. Jordan C., Kulasekaran S., Gaffney N. Wrangler: A New Generation of Data-intensive Supercomputing, 39 p.

17. Gorda B. DAOS: An Architecture for Exascale Storage, Intel, 13 p.
18. Kamber E. Parallel File Systems & Storage Architecture, 25 p.
19. Verkuil S. A Comparison of Fault-Tolerant Cloud Storage File Systems, 9 p.
20. Depardon B., Le Mahec G., S'eguín C. Analysis of Six Distributed File Systems, Research Report, 2013, 44 p.
21. Marinov M. Performance comparisson of Distributed File Systems, 28 p.
22. Donvito G., Marzulli G., Diacono D. Testing of several distributed file-systems (HDFS, Ceph and GlusterFS) for supporting the HEP experiments analysis, 2014, 8 p.
23. Donvito G. Testing of several distributed file-systems (HDFS, Ceph and GlusterFS) for supporting the HEP experiments analysis, 2014, presentation, 43 p.
24. Vijayakumari R., Kirankumar R., Gangadhara Rao K. Comparative analysis of Google File System and Hadoop Distributed File System. International Journal of Advanced Trends in Computer Science and Engineering, Vol. 3 , No.1, Pages : 553– 558 (2014)
25. Gerry J. A. High Performance Storage System. Storing Exabytes of Data Storing Exabytes of Data Across Trillions of Files, 07 May 2013, 13 p.
26. Gerry J. An HPSS Update for MSST 2015, June 2, 2015, 5 p.
27. Plimpton A. SwiftStack Filesystem Gateway Architecture, August 2014, 9 p.
28. Brim M.J., Dillow D.A., Oral S., Settlemeyer B.W. and Wang F. Asynchronous Object Storage with QoS for Scientific and Commercial Big Data, 7 p.
29. Weil S.A., Brandt S.A., Miller E.L., Long D.D.E., Maltzahn C. Ceph: A Scalable, High-Performance Distributed File System, 14 p.
30. Weil S.A., Leung A.W., Brandt S.A., Maltzahn C. RADOS: A Scalable, Reliable Storage Service for Petabyte-scale Storage Clusters, 10 p.
31. Weil S.A. CEPH: RELIABLE, SCALABLE, AND HIGH-PERFORMANCE DISTRIBUTED STORAGE. A dissertation for the degree of PhD inCOMPUTER SCIENCE. 239 p.
32. Ceph Essential, May30,2014, 151 p.
33. Comparing the Hadoop Distributed File System (HDFS) with the Cassandra File System (CFS), White Paper BY DATASTAX CORPORATION. August 2013, 13 p.

Материалы по теме«**MPI**»

1. Рахматуллин Д. Я. Введение в MPI.– ИМБИЦ УНЦ РАН, Уфа. – 47 с.
2. НемнюгинС.А. Intel® Cluster Tools. Вводный курс. Комплекс презентаций, 268 с.
3. Гергель В.П. Образовательный комплекс Введение в методы параллельного программирования. Раздел 4. Параллельное программирование на основе MPI. 35 с.+ презентации 139 с.

Материалы по теме«**Hadoop**»

1. Уайт Т. Hadoop: Подробное руководство. – СПб.: Питер, 2013. – 672 с.: ил. – (Серия «Бестселлеры O'Reilly»).
2. Лекции по Hadoop: комплекс презентаций, 238 с.
3. Shenoy A. Hadoop Explained.– Packt Publishing, 2014.– 22 p.
4. Shvachko K., Kuang H., Radia S., Chansler R. The Hadoop Distributed File System, 10 p.
5. Borthakur D. The Hadoop Distributed File System: Architecture and Design, 14 p.
6. Видеоматериалы семинара Hadoop Kitchen 27 сентября, Москва, офисMail.Ru Group
7. Karanth S. Mastering Hadoop.– Packt Publishing, 2014.– 374 p.
8. Видеоматериалы семинара. Алексей Демин. Обзор Hadoop дистрибутивов. Тюнинг «узких мест» Hadoop.
9. Arora A. Mehrotra S. Learning YARN.– Packt Publishing, 2015.– 278 p.
10. Fasale A., Kumar N. YARN Essentials.– Packt Publishing, 2015.– 285 p.
11. White T. Hadoop: The Definitive Guide. 4th Edition, Revised & Updated.– O'Reilly Media, Inc., 2016. – 805 p.
12. Turkington G., Modena G. Learning Hadoop 2.– Packt Publishing, 2015.– 382 p.

13. Sitto K., Presser M. Field Guide to Hadoop.– O'Reilly Media, Inc., 2015. – 132 p.

Материалы по теме «**OpenStack**»

1. Kumar A., Shelley D. OpenStack Trove. – Apress, 2015.– 326 p.
2. Roadmap: OpenStack Style, October 2015, A presentation by the Product WG, 38 p.
3. Видеоматериалы семинара Дорога в облака: OpenStack&BigData, 29 апреля 2014, Москва, Yandex
4. Александр Игнатов Hadoop on OpenStack
5. Руслан Камалдинов OpenStack – development and testing process

Материалы по теме «**Apache Pig**»

1. Bhandarkar M. Practical Problem Solving with Hadoop and Pig. Middleware, 2009.–199 p.
2. Gates A. Programming Pig. –Yahoo!, 2011, Published by O'Reilly Media, Inc.– 222 p.
3. Olsten C., Reed B., Srivastava U., Kumar R., Tomkins A. Pig Latin: A Not-So-Foreign Language for Data Processing, 14 p.
4. Pasupuleti P. Pig Design Patterns.– Packt Publishing, 2014.– 310 p.
5. Olston C., Chopra S, Srivastava U. Generating Example Data For Dataflow Programs, Yahoo! Research, 41 p.
6. Kawa A. Introduction to Apache Pig. The 3rd Meeting of WHUG June 21, 2012, Warsaw.– 81 p.

Материалы по теме «**Apache Oozie**»

1. Oozie / coreservlets.com and Dima May, 2012, 22 p.
2. Islam M.K., Srinivasan A. Apache Oozie.– O'Reilly Media, Inc., 2015. – 271 p.

Материалы по теме «**Apache Tez**»

1. Shah H. Apache Tez. Hortonworks Inc., 2011.– 29 p.
2. Heger D.A., A Brief Introduction to Apache Tez.– DHTechnologies & Data Nubes. – 5 p.

Материалы по теме «**Apache Spark**»

1. Sankar K., Karau H. Fast Data Processing with Spark. 2nd Edition.– Packt Publishing, 2015.– 184 p.
2. Guller M. Big Data Analytics with Spark. A Practitioner's Guide to Using Spark for Large-Scale Data Processing, Machine Learning, and Graph Analytics, and High-Velocity Data Stream Processing. □ Apress, 2015.– 290 p.
3. Ramamonjison R. Apache Spark Graph Processing. Build, process, and analyze large-scale graphs with Spark.– Packt Publishing, 2015.– 148 p.
4. Ryza S., Laserson U., Owen S., Wills J. Advanced Analytics with Spark.– O'Reilly Media, Inc., 2015. – 276 p.
5. Frampton M. Mastering Apache Spark. Gain expertise in processing and storing data by using advanced techniques with Apache Spark.– Packt Publishing, 2015.– 318 p.
6. Scott J.A. Getting Started with Apache Spark.– MapR Technologies, Inc., 2015.– 88 p.
7. Spark Core Programming.– Tutorials Point (I) Pvt. Ltd., 2015.– 36 p.
8. Spark Tutorial @ DAO.– Databricks, 2015.– 115 p.
9. Intro to Apache Spark.– Databricks, 2014.– 194 p.
10. Nathan P. Intro to Apache Spark. GOTO Chicago, 2015-05-14.– Databricks, 2015.– 188 p.
11. Barr T. Enterprise Data Storage and Analysis on Spark.– CRAY, 2015.– 34 p.
12. Zaharia M., in collaboration with Chowdhury M., Das T., Dave A., Engle C., Franklin M., Li H., Lupper A., Ma J., McCauley M., Shenker S., Stoica I., Xin R. Spark and Shark. High-Speed In-Memory Analytics over Hadoop and Hive Data.– UC Berkeley, 2012.– 36 p.
13. Scott J. An Overview of Apache Spark.– MapR Technologies, Inc., July 10, 2014.– 36 p.
14. Maniyam S. Apache Spark: Fast and Easy Data Processing. 2015 SNIA Analytics & Big data Summit. – Elephant Scale LLC, 2015.– 31 p.
15. Видеоматериалы семинара. Алексей Демин. «Почему Spark отнюдь не так хорош»

Материалы по теме«**SAPHANA**»

1. Singh V. Real Time Analytics with SAP HANA.– Packt Publishing, 2015.– 227 p.
2. Mankala C., Mahadevan V.G. SAP HANA Cookbook.– Packt Publishing, 2013.– 284 p.
3. Sikka V., Färber F., Lehner W., Cha S.K., Peh T., Bornhövd C. Efficient Transaction Processing in SAP HANA Database – The End of a Column Store Myth.– SAP.– 11 p.
4. Baeuerle S. SAP HANA Platform for BigData & IoT. Guide For Applications.– SAP, March, 2016.– 43 p.

Материалы по теме«**IBMBluemix**»

1. IBM Bluemix: платформа для разработки и разработчиков.– IBM, 2015.– 26 с.
2. Программа академической инициативы IBM в облаке.– IBM, 2015.– 4 с.
3. Thomas P. IBM Bluemix.– IBM, 2016.– 35 p.
4. Полунин А. IBM Bluemix: Использование в университетских курсах.– IBM, 2015.– 19 с.
5. Zikopoulos P.C., deRoos D., Parasuraman K., Deutsch T., Corrigan D., Giles J. Harness the Power of Big Data: The IBM Big Data Platform.– The McGraw-Hill Companies, 2013.–281 p.

Материалы по теме«**AmazonEC2**»

1. Reddy S., Sarkar A. Amazon EC2 Cookbook.– Packt Publishing, 2015.– 195 p.

Материалы по теме«**MicrosoftAzure**»

1. Nadipalli R. HDInsight Essentials. 2nd Edition.– Packt Publishing, 2015.– 179 p.
2. Chauhan A., Fontana V., Hart M., Tok W.H., Woody B. Introducing Microsoft Azure HDInsight: Technical Overview.– Microsoft Corporation, 2014.– 130 p.
3. Sarkar D. Pro Microsoft HDInsight: Hadoop on Windows. – Apress, 2014.– 258 p.

Материалы по теме«**Средства потоковой обработки данных**»

1. Betts R., Hugg J. FastData: Smart and at Scale. Design Patterns and Recipes.– O'Reilly Media, Inc., 2016.– 51 p.
2. Kleppmann M. Making Sense of Stream Processing: The Philosophy Behind Apache Kafka and Scalable Stream Data Platforms.– O'Reilly Media, Inc., 2016.– 182 p.
3. Amakobe M. A comparison between Apache Samza and Storm.– Colorado Tech University, 2016.– 9 p.
4. Bockermann C. A Survey of the Stream Processing Landscape. Technical Report.– Technische Universität Dortmund.– 47 p.
5. Namiot D. On Big Data Stream Processing. International Journal of Open Information Technologies vol. 3, no. 8, 2015, pp.48-51

Материалы по теме«**Storm**»

1. Leibusky J., Eisbruch G., Simonassi D. Getting Started with Storm.– O'Reilly Media, Inc., 2012.– 105 p.
2. Jain A., Nalya A. Learning Storm.– Packt Publishing, 2014.– 252 p.
3. Saxena S. Real-time Analytics with Storm and Cassandra.– Packt Publishing, 2015.– 220 p.
4. Allen S. T. Jankowski M., Pathirana P. Storm Applied. Strategies for real-time event processing.– Manning Publications Co., 2015.– 280 p.
5. Goetz P.T., O'Neill B. Storm Blueprints: Patterns for Distributed Real-time Computation.– Packt Publishing, 2014.– 336 p.
6. Anderson Q. Storm Real-time Processing Cookbook.– Packt Publishing, 2013.– 254 p.

Материалы по теме«**ApacheKafka**»

1. Minni S. Apache Kafka Cookbook.– Packt Publishing, 2015.– 420 p.
2. Garg N. Apache Kafka.– Packt Publishing, 2013.– 88 p.
3. Garg N. Learning Apache Kafka. Second Edition.– Packt Publishing, 2015.– 210 p.

Материалы по теме «**Apache Flume**»

1. Hoffman S. Apache Flume: Distributed Log Collection for Hadoop.– Packt Publishing, 2013.– 108 p.
2. Shreedharan H. Using Flume.– O'Reilly Media, Inc., 2015. –269 p.

Материалы по теме «**Spark Streaming**»

1. Das T. Streaming items through a cluster with Spark Streaming.– Databricks, 2015.– 41 p.

Материалы по теме «**Apache Samza**»

1. Kleppmann M. Apache Samza: Taking stream processing to the next level.– LinkedIn, 2014.– 41 p.
2. Narkhede N. Stream Processing at LinkedIn: Apache Kafka & Apache Samza.– LinkedIn.– 45 p.

Материалы по теме «**Apache Flink**»

1. Baltagi S. Apache Flink: What, How, Why, Who, Where? .– New York City (NYC) Apache Flink Meetup, February 2nd, 2016.– 116 p.
2. Fóra G. Real-time data processing with Apache Flink.– Swedish ICT.– 28 p.
3. Hueske F. Apache Flink: Fast and Reliable Large-Scale Data Processing.– 39 p.
4. Friedman E., Tzoumas K. Introduction to Apache Flink: Stream Processing for Real Time and Beyond.– O'Reilly Media, Inc., 2016. –108 p.
5. Tzoumas K. Apache Flink: Next-gen data analysis.– 39 p.

Материалы по теме «**MapR Streams**»

1. Dunning T., Friedman E. Streaming Architecture: New Designs Using Apache Kafka and MapR Streams.– O'Reilly Media, Inc., 2016. –117 p.
2. Leone M. MapR Converged Data Platform with MapR Streams.– The Enterprise Strategy Group, Inc., 2016. –7 p.
3. Bloor R., Jozwiak R. MapR Provides an Architectural Foundation for Data-Driven Enterprises.– The Bloor Group, 2016. –7 p.

Материалы по теме «**SAP HANA Smart Data Streaming**»

1. SAP HANA Smart Data Streaming: CCL Reference. SAP HANA Platform SPS 12. Document Version: 1.0 – 2016-10-11.– SAP SE, 2016.– 340 p.
2. SAP HANA Smart Data Streaming: Developer Guide. SAP HANA Platform SPS 12. Document Version: 1.0 – 2016-10-11.– SAP SE, 2016.– 360 p.

Материалы по теме «**NoSQL системы баз данных**»

1. Vaish G. Getting Started with NoSQL.– Packt Publishing, 2013.– 142 p.
2. McCreary D., Kelly A. Making Sense of NoSQL: A guide for managers and the rest of us.– Manning Publications Co., 2014. – 314 p.
3. Benchmarking Top NoSQL Databases Apache Cassandra, Couchbase, HBase, and MongoDB.– End Point Corporation.– Originally Published: April 13, 2015 Revised: May 27, 2015.– 20 p.
4. Kriha W. NoSQL Databases. Lecture Selected Topics on Software - Technology Ultra-Large Scale Sites. Course of Studies Computer Science and Media (CSM). – University Hochschule der Medien, Stuttgart (Stuttgart Media University).–149 p.
5. Redmond E., Wilson J.R. Seven Databases in Seven Weeks.–The Pragmatic Bookshelf, 2012.–347 p.

Материалы по теме «**Apache Zookeeper**»

1. Щитинин Д. Apache ZooKeeper Курс «Базы данных.– Computer Science Center, 11 ноября 2013 г. –54 с.
2. Haloi S. Apache ZooKeeper Essentials.– Packt Publishing, 2015.– 169 p.
3. Junqueira F., Reed B. ZooKeeper.– O'Reilly Media, Inc., 2014. –238 p.

Материалы по теме «**Apache Hbase**»

1. Garg N. HBase Essentials.– Packt Publishing, 2014.– 164 p.
2. Dimiduk N., Khwarana A. HBase in Action.– Manning Publications Co., 2013. – 362 p.
3. Kerzner M., Maniyam S. HBase Design Patterns.– Packt Publishing, 2014.– 150 p.
4. George L. HBase: The Definitive Guide.– O'Reilly Media, Inc., 2011. –554 p.
5. Spaggiari J.-M., O'Dell K. Architecting HBase Applications.– O'Reilly Media, Inc., 2015. – 41 p.
6. Jiang Y. HBase Administration Cookbook.– Packt Publishing, 2012.– 332 p.

Материалы по теме «**Hipertable**»

1. Judd D. Hypertable.– Zvents, Inc., 2015. – 39 p.

Материалы по теме «**Apache Accumulo**»

1. Cordova A., Rinaldi B., Wall M. Accumulo: Application Development, Table Design, and Best Practices.– O'Reilly Media, Inc., 2015. – 552 p.
2. Halldórsson G.J. Apache Accumulo for Developers.– Packt Publishing, 2013.– 120 p.
3. Apache Accumulo.– Tetra Concepts training.– 110 p.

Материалы по теме «**Apache Cassandra**»

1. Mishra V. Beginning Apache Cassandra Development. – Apress, 2014.– 235 p.
2. Kan C.Y. Cassandra Data Modeling and Analysis.– Packt Publishing, 2014.– 196 p.
3. Sharma S. Cassandra Design Patterns.– Packt Publishing, 2014.– 88 p.
4. Strickland R. Cassandra High Availability.– Packt Publishing, 2014.– 248 p.
5. Hewitt E. Cassandra: The Definitive Guide.– O'Reilly Media, Inc., 2011. – 330 p.
6. Parthasarathy V. Learning Cassandra for Administrators.– Packt Publishing, 2013.– 120 p.
7. Neeraj N. Mastering Apache Cassandra.– Packt Publishing, 2013.– 340 p.
8. Bradberry R., Lubow E. Practical Cassandra: A Developer's Approach.– Pearson Education, Inc., 2014.– 193 p.
9. Saxena S. Real-time Analytics with Storm and Cassandra.– Packt Publishing, 2015.– 220 p.

Материалы по теме «**MongoDB**»

1. Banker K. MongoDB in Action.– Manning Publications Co., 2012. – 311 p.
2. Банкер К. MongoDB в действии. □ М.: ДМК Пресс, 2012. □ 394 с.: ил.
3. Batra A., Pahwa A., Sharma D. Mongo DB - NoSql Database: A Survey.– Advances in Computer Science and Information Technology, Volume 2, Number 5; April-June, 2015 pp. 495-499
4. Leonard A. Pro Hibernate and MongoDB. – Apress, 2013.– 377 p.
5. Dayley B. Node.js, MongoDB and AngularJS Web Development.– Pearson Education, Inc., 2014.– 1276 p.
6. Membrey P., Hows D., Plugge E. MongoDB Basics: a quick introduction to MongoDB. – Apress, 2014.– 133 p.
7. Hows D., Membrey P., Plugge E., Hawkins T. The Definitive Guide to MongoDB. A complete guide to dealing with Big Data using MongoDB. Third Edition. – Apress, 2015.– 319 p.
8. Holmes S. Getting MEAN with Mongo, Express, Angular, and Node.– Manning Publications Co., 2014. – 442 p.
9. Chodorow K. 50 Tips and Tricks for MongoDB Developers.– O'Reilly Media, Inc., 2011. – 66 p.
10. Chodorow K. Scaling MongoDB.– O'Reilly Media, Inc., 2011. – 59 p.
11. Chodorow K. MongoDB: The Definitive Guide. 2nd Edition.– O'Reilly Media, Inc., 2013. – 432 p.
12. Wilson M. Building Node Applications with MongoDB and Backbone.– O'Reilly Media, Inc., 2013. – 203 p.
13. Сегуин К. Маленькая книга о MongoDB.– 34 с.

14. O'Higgins N. MongoDB and Python.– O'Reilly Media, Inc., 2013. – 66 p.
15. Borland B. Pentaho Analytics for MongoDB.– Packt Publishing, 2014.– 146 p.
16. Copeland R. MongoDB Applied Design Patterns.– O'Reilly Media, Inc., 2013. – 175 p.
17. Islam R. PHP and MongoDB Web Development: Beginner's Guide.– Packt Publishing, 2011.– 292 p.
18. Gupta Edward S.G., Sabharwal N. Practical MongoDB: Architecting, Developing, and Administering MongoDB. – Apress, 2015.– 263 p.
19. Francia S. MongoDB and PHP.– O'Reilly Media, Inc., 2012. – 76 p.
20. Hoberman S. Data Modeling for MongoDB: Building Well-Designed and Supportable MongoDB Databases.– Technics Publications, LLC, 2014. – 252 p.
21. França W.da R. MongoDB Data Modeling.– Packt Publishing, 2015.– 202 p.

Материалы по теме «**CouchDB**»

1. Anderson J.C., Lehnardt J., Slate N. CouchDB: The Definitive Guide.– O'Reilly Media, Inc., 2010. – 272 p.
2. Holt B. Scaling CouchDB: Replication, Clustering, and Administration.– O'Reilly Media, Inc., 2011. – 72 p.
3. Lennon J. Beginning CouchDB. – Apress, 2009.– 317 p.
4. Brown MC Getting Started with CouchDB.– O'Reilly Media, Inc., 2012. – 82 p.
5. Thompson M. Getting Started with GEO, CouchDB, and Node.js.– O'Reilly Media, Inc., 2011. – 66 p.
6. Juravich T. CouchDB and PHP Web Development Beginner's Guide.– Packt Publishing, 2012.– 304 p.
7. Holt B. Writing and Querying MapReduce Views in CouchDB.– O'Reilly Media, Inc., 2011. – 75 p.

Материалы по теме «**CouchBase**»

1. Zablocki J. Couchbase Essentials.– Packt Publishing, 2015.– 313p.
2. Brown MC Getting Started with Couchbase Server.– O'Reilly Media, Inc., 2012. – 90 p.
3. Brown MC Developing with Couchbase Server.– O'Reilly Media, Inc., 2013. – 87 p.
4. Potsangbam H. Learning Couchbase.– Packt Publishing, 2015.– 248p.
5. Vohra D. Pro Couchbase Development. – Apress, 2015.– 338p.
6. Ostrovsky D., Rodenski Y. Pro Couchbase Server. – Apress, 2014.– 329p.

Материалы по теме «**RavenDB**»

1. Ritchie B. RavenDB High Performance.– Packt Publishing, 2013.– 124p.

Материалы по теме «**Redis**»

1. Сегуин К. Маленькая книга о Redis.– 32 с.
2. Das V. Learning Redis.– Packt Publishing, 2015.– 318 p.
3. Johanan J. Building Scalable Apps with Redis and Node.js.– Packt Publishing, 2014.– 316 p.
4. Macedo T., Oliveira F. Redis Cookbook.– O'Reilly Media, Inc., 2011. – 72 p.
5. Da Silva M.D., Tavares H.L. Redis Essentials.– Packt Publishing, 2015.– 230 p.
6. Carlson J.L. Redis in Action.– Manning Publications Co., 2013. – 322 p.

Материалы по теме «**Kudu**»

1. Lipcon T., Alves D., Burkert D., Cryans J.-D., Dembo A., Percy M., Rus S., Wang D., Bertozzi M., McCabe C.P., Wang A. Kudu: Storage for Fast Analytics on Fast Data.– Cloudera, inc., 2015. – 13 p.
2. Baranowski Z. Cloudera Kudu Introduction. – 26 p.

Материалы по теме «**Общий взгляд на развитие СУБД**»

1. Harrison G. Next Generation Databases: NoSQL, NewSQL, and Big Data. – Apress, 2015.– 244p.

Материалы по теме«**Apache Hive**»

1. Du D. Apache Hive Essentials.– Packt Publishing, 2015.– 313 p.
2. Capriolo E., Wampler D., Rutherglen J. Programming Hive.– O'Reilly Media, Inc., 2012. – 350 p.

Материалы по теме«**Apache Phoenix**»

1. Taylor J. Apache Phoenix. – 77 p.
2. Phoenix SQL[®] Reference Guide.– OPW Dover Company, 2016. – 95 p.

Материалы по теме«**Cloudera Impala**»

1. Russell J. Getting Started with Impala.– O'Reilly Media, Inc., 2015. – 112 p.
2. Chauhan A. Learning Cloudera Impala.– Packt Publishing, 2013.– 150 p.

Материалы по теме«**Apache HAWQ (Pivotal HDB)**»

1. Chang L., Wang Z., Ma T., Jian L., Ma L., Goldshuv A., Lonergan L., Cohen J., Welton C., Sherry G., Bhandarkar M. HAWQ: A Massively Parallel Processing SQL Engine in Hadoop.– Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, 2014 pp.1223-1234
2. Pivotal HAWQ and Hortonworks Data Platform: Modern Data Architecture for IT Transformation.– PivotalSoftwareInc., 2015. – 25 p.
3. Grishchenko A. Pivotal HAWQ.– HadoopKitchen, Mail.ru, 27 Sep 2014. – 46 p.
4. Radhakrishnan P., Banarse A., Plotkin J. PIVOTAL NY – BIG DATA & ANALYTICS MEETUP.– Pivotal Labs, April 7th, 2015. – 75 p.
5. Pivotal HAWQ 1.3.1 Documentation. Pivotal HAWQ User's Guide.– PivotalSoftwareInc., 2015. – 840p.

Материалы по теме«**HP Vertica**»

1. Agrawal R. HP Vertica Essentials.– Packt Publishing, 2014.– 106 p.
2. LaPlante A. The Big Data Transformation: Understanding Why Change Is Actually Good for Your Business.– O'Reilly Media, Inc., 2016. – 54 p.
3. Vertica Documentation. HPE Vertica Analytics Platform. Software Version: 7.2.x.– Hewlett Packard Enterprise Development LP, 2016. – 5309 p.

Материалы по теме«**IBM Big SQL**»

1. IBM Big SQL –Vendor Landscape.– IBM Corporation, 2014. – 35 p.
2. Saracco C.M. Big SQL: A Technical Introduction.– IBM Corporation, 2014. – 39 p.
3. Hoge W. Big SQL 3.0: Fast and easy SQL on Hadoop.– IBM Corporation, 2014. – 24 p.
4. Khatri A., Schewenger M.N., Finkelstein N. Big SQL 3.0 Functionality: A comprehensive approach to SQL-on-Hadoop.– International Journal of Computer and Information Technology. Volume 03 – Issue 06, November 2014, pp.1121-1128

Материалы по теме«**SAP HANA Vora**»

1. SAP HANA Vora[™]: Gain Contextual Awareness for a Smarter Digital Enterprise.– SAPSE, 2015. – 8p.
2. SAP HANA Vora - Delivering Contextual Awareness in the Digital Enterprise.– SAPSE, 2015. – 25p.

Материалы по теме«**Элементы теории графов**»

1. Носов В.И., Бернштейн Т.В., Носкова Н.В., Храмова Т.В. Элементы теории графов. Учебное пособие. – Новосибирск, 2008. – 107с.
2. Кристофидес Н. Теория графов. Алгоритмический подход.– М.: Мир, 1978. – 432с.
3. Brath R., Jonker D. Graph Analysis and Visualization: Discovering Business Opportunity in Linked Data.– John Wiley & Sons, Inc., 2015. – 539 p.

Материалы по теме«**Введение в графовые СУБД**»

1. Angles R. A Comparison of Current Graph Database Models.– 3rd Int. Workshop on Graph Data Management: Techniques and applications (GDM 2012) 5 April, Washington DC, USA. – 24 p.
2. Jones C.W. Intro to Graph Databases: Using Tinkerpop, TitanDB, and Gremlin.– 69 p.

Материалы по теме «**Neo4j**»

1. Hunger M., Boyd R., Lyon W. The Definitive Guide to Graph Databases for the RDBMS Developer.– Neo Technology, Inc., 2014. – 35 p.
2. IRobinson I., Webber J., Eifrem E. Graph Databases. 2nd Edition.– O'Reilly Media, Inc., 2015. – 237 p.
3. Van Bruggen R. Learning Neo4j.– Packt Publishing, 2014.– 222 p.
4. Goel A. Neo4j Cookbook.– Packt Publishing, 2015.– 226 p.
5. Vukotic A., Watt N. with Abedrabbo T., Fox D., Partner J. Neo4j in Action.– Manning Publications Co., 2015. – 305 p.
6. Jordan G. Practical Neo4j. – Apress, 2014.– 393 p.

Материалы по теме «**OrientDB**»

1. Tesoriero C. Getting Started with OrientDB.– Packt Publishing, 2013.– 138 p.
2. OrientDB[®] Distributed DBMS.– Orient Technologies LTD.– 23 p.

Материалы по теме «**Titan**»

1. Broecheler M. Titan Big Graph Data with Cassandra.– AURELIUS.– 68 p.

Материалы по теме «**Pregel**»

1. Malewicz G., Austern M.H., Bik A.J.C., Dehnert J.C., Horn J., Leiser N., Czajkowski G. Pregel: A System for Large-Scale Graph Processing.– Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. Indianapolis, Indiana, USA — June 06 - 10, 2010, pp. 135-146

Материалы по теме «**Apache Giraph**»

1. Martella C., Shaposhnik R., Logothetis D. Practical Graph Analytics with Apache Giraph. – Apress, 2014.– 320 p.

Материалы по теме «**Surfer**»

1. Chen R., Weng X., He B., Yang M., Choi B., Li X. On the Efficiency and Programmability of Large Graph Processing in the Cloud.–Microsoft Research Asia, Nanyang Technological University, Hong Kong Baptist University, Beijing University, 2009.– 16 p.

Материалы по теме «**GBase**»

1. Kang U, Tong H., Sun J., Lin C.-Y., Faloutsos C. GBASE: A Scalable and General Graph Management System.– Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. San Diego, California, USA — August 21 - 24, 2011, pp. 1091-1099

Материалы по теме «**Signal / Collect**»

1. Stutz P., Bernstein A., Cohen W. Signal/Collect: Graph Algorithms for the (Semantic) Web.–ISWC'10 Proceedings of the 9th international semantic web conference on The semantic web/ Shanghai, China — November 07 - 11, 2010. Volume Part I, pp. 764-780
2. Stutz P., Strebel D., Bernstein A. Signal/Collect: Processing Large Graphs in Seconds.– International Semantic Web Conference ISWC 2010.– 29 p.

Материалы по теме «**Mizan**»

1. Khayyat Z., Awara K., Alonazi A., Jamjoom H., Williams D., Kalnis P. Mizan: A system for Dynamic Load Balancing in Large-Scale Graph Processing.–EuroSys '13 Proceedings of the 8th ACM European Conference on Computer Systems. Prague, Czech Republic — April 15 - 17, 2013, pp 169-182

2. Khayyat Z., Awara K., Alonazi A., Jamjoom H., Williams D., Kalnis P. Mizan: A system for Dynamic Load Balancing in Large-Scale Graph Processing.–EuroSys '13 Proceedings of the 8th ACM European Conference on Computer Systems. Prague, Czech Republic — April 15 - 17, 2013, (presentation 45 p.)
3. Khayyat Z., Awara K., Alonazi A., Jamjoom H., Williams D., Kalnis P. Mizan: A system for Dynamic Load Balancing in Large-Scale Graph Processing.– IBM, 2013.– 35 p.

Материалы по теме «**GraphLab**»

1. Drumond L.R. Big Data Analytics. GraphLab.– Information Systems and Machine Learning Lab (ISMLL) Institute of Computer Science University of Hildesheim, Germany.– 43 p.
2. Low Y., Gonzalez J., Kyrola A., Bickson D., Guestrin C., Hellerstein J. GraphLab: A New Framework For Parallel Machine Learning.– Carnegie Mellon University. School of Computer Science, 2010.– 10 p.
3. Low Y., Gonzalez J., Kyrola A., Bickson D., Guestrin C., Hellerstein J. GraphLab: A New Framework For Parallel Machine Learning.– CarnegieMellonUniversity. School of Computer Science, 2010.– presentation, 24 p.
4. Gonzalez J. A New Parallel Framework for Machine Learning.– CarnegieMellonUniversity.– presentation, 74 p.
5. Low Y., Gonzalez J., Kyrola A., Bickson D., Guestrin C., Hellerstein J. Distributed GraphLab: A Framework for Machine Learning and Data Mining in the Cloud.– CarnegieMellonUniversity, 2012.– 12 p.
6. Gonzalez J.E., Low Y., Gu H., Bickson D., Guestrin C. PowerGraph: Distributed Graph-Parallel Computation on Natural Graphs.– 10th USENIX Symposium on Operating Systems Design and Implementation (OSDI '12), 2012.– pp.17-30
7. GraphLab and PowerGraph.– Computer Science & Engineering at the University of Washington, 2013.– 5 p.
8. Sun S. Large-Scale Graph Machine Learning.– Graduate School of Computer Science Saarland University.– May 29, 2015.– presentation, 60p.
9. Low Y. GraphLab: A Distributed Abstraction for Large Scale Machine Learning.– PhD Thesis, Machine Learning Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, June 2013.– 126 p.
10. Barnawil A., Batarfil O., Behteshi S.-M.-R., Elshaw R., Fayoumi A., Nouri R., Sakr S. On Characterizing the Performance of Distributed Graph Computation Platforms.– Performance Characterization and Benchmarking. Traditional to Big Data. 6th TPC Technology Conference, TPCTC 2014 Hangzhou, China, September 1–5, 2014. Revised Selected Papers.– pp. 29-43

Материалы по теме «**PEGASUS**»

1. Kang U, Meeder B., Faloutsos C. Spectral Analysis for Billion-Scale Graphs: Discoveries and Implementation.– Carnegie Mellon University, School of Computer Science, 2011.–12 p.
2. Kang U, Tsourakakis C.E., Faloutsos C. PEGASUS: A Peta-Scale Graph Mining System - Implementation and Observations.– IEEE International Conference On Data Mining 2009, Miami, Florida, USA.– Carnegie Mellon University, School of Computer Science, 2009.– 10 p.
3. PEGASUS User's Guide / Kang U., Chau D.H., Faloutsos C.– Carnegie Mellon University. ProjectPegasu, Sep. 6th, 2010.– 22 p.

Материалы по теме «**Apache Spark GraphX**»

1. Ryza S., Laserson U., Owen S., Wills J. Advanced Analytics with Spark.– O'Reilly Media, Inc., 2015. – 276 p.
2. Ramamonjison R. Apache Spark Graph Processing. Build, process, and analyze large-scale graphs with Spark.– Packt Publishing, 2015.– 148 p.

2. Guller M. Big Data Analytics with Spark. A Practitioner's Guide to Using Spark for Large-Scale Data Processing, Machine Learning, and Graph Analytics, and High-Velocity Data Stream Processing. □ Apress, 2015.– 290 p.
3. Raj P., Chelladurai J.S., Singh V. Learning Docker.– Packt Publishing, 2015.– 240 p.
4. Frampton M. Mastering Apache Spark.– Packt Publishing, 2015.– 318 p.

Материалы по теме «**SAP HANA Graph Engine**»

1. SAP HANA Graph Engine.– SAP Startup Focus. HANA 1.0 SPS 9, H1 2015.– 31 p.
2. Chen M.P. SAP HANA Graph Use Cases.– SAP SE, March 2015.– 14 p.
3. SAP HANA Graph Reference. SAP HANA Platform SPS 12. Document Version: 1.1 – 2016-09-13.– SAP SE, 2016.– 28 p.

Материалы по теме «**Microsoft Graph Engine**»

1. Han W., Miao Y., Li K., Wu M., Yang F., Zhou L., Prabhakaran V., Chen W., Chen E. Chronos: A Graph Engine for Temporal Graph Analysis.–Tsinghua University, University of Science and Technology of China, Microsoft Research, 2014.– 43 p.
2. Sarwat M., Elnikety S. He Y., Klot G. Horton: Online Query Execution Engine For Large Distributed Graphs.– Dept. of Computer Science and Engineering University of Minnesota, Microsoft Research, 2012.– 4 p.
3. Shao B., Wang H., Li Y. Trinity: A Distributed Graph Engine on a Memory Cloud.– Microsoft Research, Hong Kong University of Science and Technology, 2013.– 12 p.
5. Zeng K., Yang J., Wang H., Shao B., Wang Z. A Distributed Graph Engine for Web Scale RDF Data.–University of California, Los Angeles, Columbia University, Microsoft Research Asia, Renmin University of China, 2013.– 12 p.

Материалы по теме «**Apache Nutch и Apache Solr как развитие Apache Lucene**»

1. Ng E., Mohan V. Lucene 4 Cookbook.– Packt Publishing, 2015.– 220 p.
2. McCandless M., Hatcher E., Gospodnetić O. Lucene in Action, Second Edition.– Manning Publications Co., 2010. – 528 p.
3. Laliwala Z., Shaikh A. Web Crawling and Data Mining with Apache Nutch.– Packt Publishing, 2013.– 136 p.
4. Smiley D., Pugh E. Parisa K., Mitchell M. Apache Solr Enterprise Search Server, 3rd Edition.– Packt Publishing, 2015.– 432 p.
6. Kuć R. Apache Solr 4 Cookbook, 2nd Edition.– Packt Publishing, 2013.– 328 p.
7. Mohan S. Adminstrating Solr.– Packt Publishing, 2013.– 120 p.
8. Gazzarini A. Apache Solr Essentials.– Packt Publishing, 2015.– 338 p.
9. Mohan S. Apache Solr High Performance.– Packt Publishing, 2014.– 124 p.
10. Kumar J. Apache Solr Search Patterns.– Packt Publishing, 2015.– 316 p.
11. Shahi D. Apache Solr: A Practical Approach to Enterprise Search. □ Apress, 2015.– 316 p.
12. Karambelkar H. Scaling Big Data with Hadoop and Solr.– Packt Publishing, 2013.– 144 p.
13. Grainger T., Potter T. Solr in Action.– Manning Publications Co., 2014. – 666 p.

Материалы по теме «**ElasticSearch**»

1. Mohan V. Elasticsearch Blueprints.– Packt Publishing, 2015.– 192 p.
2. Paro A. Elasticsearch Cookbook, 2nd Edition.– Packt Publishing, 2015.– 472 p.
3. Gheorghe R., Hinman M.L., Russo R. Elasticsearch in Action.– Manning Publications Co., 2015. – 498 p.
4. Akdoğan H. Elasticsearch Indexing.– Packt Publishing, 2015.– 176 p.
5. Kuć R., Rogoziński M. Mastering ElasticSearch.– Packt Publishing, 2013.– 386 p.

Материалы по теме «**Splunk**»

1. Zadrozny P., Kodali R. Big Data Analytics Using Splunk. □ Apress, 2013.– 362 p.
2. Bumgarner V. Implementing Splunk: Big Data Reporting and Development for Operational Intelligence.– Packt Publishing, 2013.– 448 p.
4. Miller J. Mastering Splunk.– Packt Publishing, 2014.– 344 p.

5. Smith K. Splunk Developer's Guide, 2nd Edition.— Packt Publishing, 2016.— 190 p.
6. Sigman B.P. Splunk Essentials.— Packt Publishing, 2015.— 156 p.
7. Diakun J., Johnson P.R., Mock D. Splunk Operational Intelligence Cookbook.— Packt Publishing, 2014.— 414 p.

8. ПЕРЕЧЕНЬ РЕСУРСОВ СЕТИ ИНТЕРНЕТ, РЕКОМЕНДУЕМЫХ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ПРИ ОСВОЕНИИ ДИСЦИПЛИНЫ

1. Сайт библиотеки МГТУ им. Н.Э. Баумана: <http://library.bmstu.ru>.
2. Сайты кафедры ИУ5 «Системы обработки информации и управления»:
 - a. http://e-learning.bmstu.ru/portal_iu5/
 - b. <http://iu5.bmstu.ru>
3. Сайт веб-консорциума: <https://www.w3.org/>
4. Ресурсы по теме «Понятие больших данных и технологий обработки больших данных»:
 - http://bigdatawg.nist.gov/uploadfiles/M0055_v1_7606723276.pdf
 - http://sk.ru/cfs-file.ashx/_key/Skl-Entities-Files/d89a55d4_2D00_5dd8_2D00_44e5_2D00_ad36_2D00_7e0f99226573/2013_2D00_11_2D00_08_2D00_Tabakov.pdf
 - <http://infrastructurearchitecture.blogspot.ru/2012/11/8v-spider-big-data-assessment-model.html>
 - <https://www.m-brain.com/home/technology/big-data-with-8-vs/>
 - <http://igel.wpengine.netdna-cdn.com/wp-content/uploads/2013/11/David-Parker.pdf>
 - <http://www.datameer.com/solutions/industries/financial-services.html>
 - <http://sp.parallels.com/fileadmin/media/hcap/events/summit/2014/documents/Summit2014-keynote-SergueiBelousov.pdf>
 - <http://ihealthtran.com/images/Infographic-the-body-as-a-source-of-big-data-HealthIT-Infographic-NetApp-Infographic.pdf>
 - <http://d1c25a6gwz7q5e.cloudfront.net/reports/2014-09-12-Sustainability-in-the-Age-of-Big-Data.pdf>
 - https://fbcdn-dragon-a.akamaihd.net/hphotos-ak-prn1/851575_520797877991079_393255490_n.pdf
 - <https://about.twitter.com/company>
 - <http://www.techrepublic.com/article/why-the-worlds-largest-hadoop-installation-may-soon-become-the-norm/>
 - <http://www.datameer.com/product/industries-use-cases/retail/>
 - <http://www.datameer.com/solutions/industries/telecommunications.html>
 - <https://datafloq.com/read/big-data-will-revolutionize-learning/206>
 - <http://www.socialnomics.net/2014/03/05/big-data-in-education-big-potential-or-big-mistake/>
 - <http://queue.acm.org/detail.cfm?id=988408>
 - <http://static.googleusercontent.com/media/research.google.com/en/us/archive/gfs-sosp2003.pdf>
 - <http://cacs.usc.edu/education/cs653/Dean-MapReduce-CACM08.pdf>
 - <https://developer.yahoo.com/blogs/hadoop/yahoo-launches-world-largest-hadoop-production-application-398.html>
 - <http://www.computerra.ru/cio/5598>
 - http://static.usenix.org/event/osdi06/tech/full_papers/weil/weil.pdf
 - <http://www.pdsw.org/pdsw13/papers/p7-pdsw13-brim.pdf>
 - https://hal.inria.fr/hal-00789086/PDF/a_survey_of_dfs.pdf
 - http://iopscience.iop.org/1742-6596/513/4/042014/pdf/1742-6596_513_4_042014.pdf
 - <https://gigaom.com/2012/07/11/because-hadoop-isnt-perfect-8-ways-to-replace-hdfs/>
 - http://hadoopilluminated.com/hadoop_illuminated/Bigdata_Ecosystem.html
 - <http://scn.sap.com/community/business-intelligence/blog/2013/12/29/big-data-in-an-sap-landscape-asug-webcast-part-1>
 - <http://hadoopecosystemtable.github.io/>
 - <http://technomag.bmstu.ru/doc/761354.html>
 - <http://www.slideshare.net/SAPMENA/hana-27077351>
 - http://help.sap.com/hana/SAP_HANA_Master_Guide_en.pdf

<http://download.sap.com/download.epd?context=E67AFF9FD4CFC6693FC443EB965A7B4FE599BA88ED6C9F622486D0E5BEB350EB7DAD751393D16A2D916A3C9EA834D1CB2A8138467760590E>
https://hcp.sap.com/content/dam/website/saphana/en_us/Technology%20Documents/SAP%20HANA%20Architecture%20and%20Landscape%20Options.pdf
<https://lenovopress.com/redpxref.pdf>
<http://saphanatutorial.com/trendz/december-2015/what-is-sap-hana-vora-why-has-it-become-important/>
<http://www.friendlyfunction.com/ru/apache-spark-introduction/>
<http://hortonworks.com/why-hortonworks/>
<http://hortonworks.com/hdp/>
<http://hortonworks.com/hdf/>
<http://www.cloudera.com/products.html>
<http://www.pcweek.ru/infrastructure/blog/infrastructure/8168.php>
<https://www.gartner.com/doc/reprints?id=1-2YEVU8U&ct=160210&st=sb>
http://help.prognoz.com/en/help.htm#mergedProjects/Intro/architecture/inf_delivery_analysis.htm#Modelling
<https://spark-summit.org/2014/wp-content/uploads/2014/07/Sparks-Role-in-the-Big-Data-Ecosystem-Matei-Zaharia1.pdf>
<http://www.slideshare.net/databricks/introduction-to-spark-intern-event-presentation>
<http://searchaws.techtarget.com/definition/data-lake>
[http://searchdatamanagement.techtarget.com/feature/Swim-fast-with-a-Hadoop-data-lake-architecture-or-sink?utm_medium=EM&asrc=EM_ERU_53261581&utm_campaign=20160210_ERU%20Transmission%20for%2002/10/2016%20\(UserUniverse:%201940550\)_myka-reports@techtarget.com&utm_source=ERU&src=5479321](http://searchdatamanagement.techtarget.com/feature/Swim-fast-with-a-Hadoop-data-lake-architecture-or-sink?utm_medium=EM&asrc=EM_ERU_53261581&utm_campaign=20160210_ERU%20Transmission%20for%2002/10/2016%20(UserUniverse:%201940550)_myka-reports@techtarget.com&utm_source=ERU&src=5479321)
<http://searchbusinessanalytics.techtarget.com/feature/Hadoop-clusters-provide-single-spot-for-spreadmarts-analytics>
<http://searchdatamanagement.techtarget.com/feature/Hadoop-data-lakes-must-get-more-efficient-less-messy-to-oust-EDWs>
<http://searchdatamanagement.techtarget.com/feature/When-building-Hadoop-data-lakes-dont-plunge-in-without-a-plan>
<http://searchdatamanagement.techtarget.com/feature/Dont-throw-out-design-principles-when-jumping-in-Hadoop-data-lake>
http://www.st.ewi.tudelft.nl/~hauff/BDP-Lectures/5_filesystem_gfs_hdfs.pdf
<http://blog.trentonsystems.com/apache-hadoop-the-open-source-elephant-of-big-data/>
<http://nosql.mypopescu.com/post/30815314471/life-of-data-at-facebook>
<http://yahooohadoop.tumblr.com/post/98213421641/storm-and-spark-at-yahoo-why-chose-one-over-the>
<https://developer.yahoo.com/blogs/hadoop/apache-hbase-yahoo-multi-tenancy-helm-again-171710422.html#more-id>
<http://www.odbms.org/blog/2014/09/interview-mithun-radhakrishnan/>
<http://yahooohadoop.tumblr.com/post/98751512631/the-evolution-of-storm-at-yahoo-and-apache>
http://www.youtube.com/watch?v=QAejaeTvj_M
<http://lambda-architecture.net/>
<http://hortonworks.com/wp-content/uploads/2013/09/CIO.Guide.How.to.Use.Hadoop.with.Your.SAP.Software.Landscape.pdf>
<http://scn.sap.com/community/business-intelligence/blog/2013/12/29/using-hana-and-hadoop-key-scenarios>

<http://scn.sap.com/community/business-intelligence/blog/2013/12/29/fitting-hadoop-in-an-sap-software-landscape-part-3-asug-webcast>
<https://blogs.saphana.com/2014/10/21/what-is-new-in-sap-hana-sps-09/>
<https://blogs.saphana.com/2014/03/05/guinness-world-record-largest-data-warehouse/>
<http://www.slideshare.net/bluespiremarketing/analytics-and-big-data-and-roioh-my-trendlab-webinar>
<http://www.mathnet.ru/links/e09eed4f2e8ea54ad33b6efc32c376a2/da161.pdf>
https://smartmanufacturingcoalition.org/sites/default/files/optimization_with_grid_computing.pdf
<http://cis.rosnou.ru/UniversysDWNL/Library/D9CA647C-7F45-4069-9623-7018580F554B/D9CA647C-7F45-4069-9623-7018580F554B.pdf>
http://iipo.tu-bryansk.ru/fileadmin/user_upload/content_admins/Journal_PrI/ZHurnal_Programmnaja_Inzhenerija_2011_No_3.pdf
<http://lambda.uta.edu/mrql-bsp.pdf>
<http://www.dissercat.com/content/issledovanie-i-razrabotka-modelei-i-arkhitektury-sredstv-kontrollinga-dlya-mezhregionalnykh-#ixzz3O5emDlst>
<http://www.dissercat.com/content/intellektualnaya-sistema-dlya-resheniya-zadach-optimalnogo-upravleniya-s-vychislitelnymi-oso#ixzz3O5bV01s9>

Ресурсы по MPI:

http://www.opennet.ru/docs/RUS/MPI_intro/
https://parallel.ru/tech/tech_dev/mpi.html
<http://cluster.linux-ekb.info/mpi.php>
<http://www.ccas.ru/mmes/educat/lab04k/01/basics.html>
<http://pro-spo.ru/parallel/2151--mpi>
https://en.wikipedia.org/wiki/Message_Passing_Interface
http://matem.anrb.ru/sites/default/files/publications/2006_rahmatullin_d.ya_vvedenie_v_mpi_ufa.pdf
<http://www.hpcc.unn.ru/mskurs/RUS/DOC/ppr04.pdf>

Учебник при поддержке Microsoft:

http://www.hpcc.unn.ru/mskurs/RUS/HTML/cs338_ppr_course.htm
http://www.hpcc.unn.ru/mskurs/RUS/PPT/ppr04_1.pdf
http://www.hpcc.unn.ru/mskurs/RUS/PPT/ppr04_2.pdf
http://www.hpcc.unn.ru/mskurs/RUS/PPT/ppr04_3.pdf

Учебник при поддержке Intel: С.А. Немнюгин, Введение в программирование на кластерах

Материал по MPI содержится в следующих разделах:

Лекция 2 <http://www.intuit.ru/studies/courses/4448/984/lecture/14935>
<http://www.intuit.ru/studies/courses/4448/984/lecture/14935?page=2>
<http://www.intuit.ru/studies/courses/4448/984/lecture/14935?page=3>
 Лекция 3 <http://www.intuit.ru/studies/courses/4448/984/lecture/14937>
<http://www.intuit.ru/studies/courses/4448/984/lecture/14937?page=2>
<http://www.intuit.ru/studies/courses/4448/984/lecture/14937?page=3>
 Лекция 4 <http://www.intuit.ru/studies/courses/4448/984/lecture/14939>
<http://www.intuit.ru/studies/courses/4448/984/lecture/14939?page=2>
 Лекция 5 <http://www.intuit.ru/studies/courses/4448/984/lecture/14941>
<http://www.intuit.ru/studies/courses/4448/984/lecture/14941?page=2>
<http://www.intuit.ru/studies/courses/4448/984/lecture/14941?page=3>
 Лекция 6 <http://www.intuit.ru/studies/courses/4448/984/lecture/14943>
<http://www.intuit.ru/studies/courses/4448/984/lecture/14943?page=2>
<http://www.intuit.ru/studies/courses/4448/984/lecture/14943?page=3>
<http://www.intuit.ru/studies/courses/4448/984/lecture/14943?page=4>

Лекция 7 <http://www.intuit.ru/studies/courses/4448/984/lecture/14945>
<http://www.intuit.ru/studies/courses/4448/984/lecture/14945?page=2>

Ресурсы по **Apache Avro:**

<http://www.tutorialspoint.com/avro/index.htm>

Ресурсы по **Apache Pig:**

<http://www.ibm.com/developerworks/ru/library/l-apachepigdataquery/index.html>

Ресурсы по **Apache Oozie:**

<http://www.ibm.com/developerworks/ru/library/bd-ooziehadoop/index.html>

<https://azure.microsoft.com/ru-ru/documentation/articles/hdinsight-use-oozie-coordinator-time/>

<http://hortonworks.com/hadoop/oozie/>

<https://www.mapr.com/products/product-overview/apache-oozie>

Ресурсы по **Apache Tez:**

<http://hortonworks.com/blog/apache-tez-a-new-chapter-in-hadoop-data-processing/>

<http://hortonworks.com/blog/expressing-data-processing-in-apache-tez/>

<http://hortonworks.com/blog/task-api-apache-tez/>

<http://hortonworks.com/blog/writing-a-tez-inputprocessoroutput-2/>

<http://hortonworks.com/blog/apache-tez-dynamic-graph-reconfiguration/>

<http://hortonworks.com/blog/re-using-containers-in-apache-tez/>

<http://hortonworks.com/blog/introducing-tez-sessions/>

<http://www.infoq.com/articles/apache-tez-saha-murthy>

<http://hadooptutorial.info/apache-tez-successor-mapreduce-framework/>

<https://www.youtube.com/watch?v=9wFL3Hm6dfw>

<https://www.youtube.com/watch?v=cPSfA1bhgVA>

<https://www.youtube.com/watch?v=9ZLLzlsz7h8>

<https://www.youtube.com/watch?v=J5j1FOFJl2U>

<https://www.youtube.com/watch?v=AaDEetR74lc>

<https://www.youtube.com/watch?v=-7YhVwqky6M>

Ресурсы по **Apache Spark:**

http://www.youtube.com/watch?v=oeoirw_SRPw

<http://sparkdeveloper.com/>

<http://sparkdeveloper.com/2015/03/19/apache-spark-quick-guide/>

Ресурсы по **Apache Samza:**

<http://samza.apache.org/>

У Samza очень хорошая документация проекта:

<http://samza.apache.org/archive/index.html#latest> и вложенные страницы документации:

<http://thenewstack.io/apache-samza-linkedins-framework-for-stream-processing/>

Ресурсы по **Apache Flink:**

<https://ci.apache.org/projects/flink/flink-docs-master/>

<https://ci.apache.org/projects/flink/flink-docs-master/concepts/concepts.html>

<http://dataartisans.github.io/flink-training/index.html> и вложенные страницы (лекции,

упражнения, документы)

<http://hortonworks.com/blog/overview-of-apache-flink-the-4g-of-big-data-analytics-frameworks/> и вложенные страницы

<http://www.infoworld.com/article/3042159/hadoop/apache-flink-10-takes-on-spark-in-hadoop-processing.html> и вложенные страницы

https://en.wikipedia.org/wiki/Apache_Flink (много ссылок)

Ресурсы, содержащие **сравнение нескольких систем обработки потоковых данных:**

<https://tsicilian.wordpress.com/2015/02/16/streaming-big-data-storm-spark-and-samza/>

<http://www.infoq.com/news/2015/12/yahoo-flink-spark-storm>
http://www.slideshare.net/arankejariwal/real-time-analytics-algorithms-and-systems?imm_mid=0e3906&cmp=em-data-na-na-newsltr_20160511

Ресурсы, содержащие теорию NoSQL-систем:

Виктор Смирнов: «Дизайн продвинутых нереляционных схем для Big Data»

<http://www.youtube.com/watch?v=oLm0hnM8p5k>

NoSQL in a Hadoop World <http://blog.cloudera.com/blog/2014/11/nosql-in-a-hadoop-world-2/>

<http://searchdatamanagement.techtarget.com/feature/NoSQL-Hadoop-data-engines-shifted-into-new-gears-in-2015>

<http://www.tomsitpro.com/articles/open-source-cloud-computing-software,2-754-8.html>

Ресурсы, содержащие сопоставления и обзоры NoSQL-систем:

Сергей Сверчков: «Оцениваем решения NoSQL: какая база данных подходит для вашей системы» <http://www.youtube.com/watch?v=f1qVn22qfXc>

<https://www.trustradius.com/nosql-databases>

<http://kkovacs.eu/cassandra-vs-mongodb-vs-couchdb-vs-redis>

<http://www.infoivy.com/2013/07/nosql-database-comparison-chart-only.html>

<http://db-engines.com/en/ranking>

Ресурсы по Apache Zookeeper:

<https://habrahabr.ru/post/144708/>

http://ru.bmstu.wiki/Apache_ZooKeeper

<http://blog.cloudera.com/blog/2013/02/how-to-use-apache-zookeeper-to-build-distributed-apps-and-why/>

<http://www.confluent.io/blog/distributed-consensus-reloaded-apache-zookeeper-and-replication-in-kafka>

Ресурсы по Hipertable:

<http://www.hypertable.com/>

http://www.hypertable.com/why_hypertable

<http://archive.is/PfIBR>

http://www.hypertable.com/why_hypertable/hypertable_vs_hbase_2/

<http://www.hypertable.com/documentation/>

Ресурсы по Apache Cassandra:

<https://habrahabr.ru/post/155115/>

<http://www.ibm.com/developerworks/ru/library/os-apache-cassandra/index.html>

<http://profyclub.ru/docs/172>

http://ru.bmstu.wiki/Apache_Cassandra

<http://www.tadviser.ru/index.php/%D0%9F%D1%80%D0%BE%D0%B4%D1%83%D0%BA%D1%82:Cassandra>

<http://www.slideshare.net/tivelkov/part-one-18057392>

<http://cassandra.apache.org/doc/cql3/CQL.html>

Ресурсы по MongoDB:

<http://metanit.com/nosql/mongodb/>

<https://docs.mongodb.org/manual/>

https://university.mongodb.com/?_ga=1.117299650.511697163.1461714698

Ресурсы по CouchDB – Couchbase:

CouchDB – первоначальная реализация системы, формально не относящаяся к BigData:

<https://habrahabr.ru/post/101251/>

<https://ru.wikipedia.org/wiki/CouchDB>

<http://dic.academic.ru/dic.nsf/ruwiki/1351934>

<http://anton.shevchuk.name/php/couchdb-for-developers/>

Couchbase - BigData система после редизайна:

https://ru.wikipedia.org/wiki/Couchbase_Server

<http://www.couchbase.com/>

<http://blog.dtulyakov.ru/2015/10/couchbase-server-40-couchdb-memcached.html>

Ресурсы по **Redis**:

<https://ru.wikipedia.org/wiki/Redis>

<http://eax.me/redis/>

<http://profyclub.ru/docs/169>

<http://redis.io/documentation>

Ссылка про Решардинг Redis "наживую" приведена, поскольку показывает внутренние механизмы, в последних версиях системы процесс реализован в полуавтоматическом режиме.

Ресурсы по **Riak**:

<https://habrahabr.ru/post/75202/>

<http://www.ibm.com/developerworks/ru/library/os-riak1/index.html>

<http://www.ibm.com/developerworks/ru/library/os-riak2/index.html>

<http://basho.com/products/#riak>

Ресурсы по **Apache Kudu**:

<https://habrahabr.ru/post/272267/>

<http://kudu.incubator.apache.org/>

Блог в трёх частях про ClouderaKudu:

<http://www.dbms2.com/2015/09/28/introduction-to-cloudera-kudu/>

<http://www.dbms2.com/2015/09/28/cloudera-kudu-deep-dive/>

<http://www.dbms2.com/2015/09/28/the-potential-significance-of-cloudera-kudu/>

<http://venturebeat.com/2015/09/24/cloudera-kudu/>

<https://sdsblog.com/2015/10/12/kudu/>

<https://www.mapr.com/blog/much-ado-about-kudu>

<http://www.datanami.com/2015/09/28/cloudera-unveils-kudu-a-fast-new-storage-option-for-hadoop/>

<http://www.slideshare.net/cloudera/apache-kudu-incubating-new-hadoop-storage-for-fast-analytics-on-fast-data-by-todd-lipcon-software-engineer-cloudera-kudu-founder>

<http://www.slideshare.net/CasertaConcepts/introducing-kudu-big-data-warehousing-meetup>

<http://www.dbta.com/Columns/Big-Data-Notes/Apache-Kudu-Bridges-NoSQL-Analytics-and-OLTP-107928.aspx>

<http://www.marktorr.com/the-cloudera-scoop-one-platform-initiative-recordservice-kudu/>

Ресурсы по теме **SQL vs NoSQL Databases**:

<http://www.sitepoint.com/sql-vs-nosql-differences/>

<http://www.thegeekstuff.com/2014/01/sql-vs-nosql-db/>

Ресурсы по теме **Couchbase N1QL** похожий на SQL:

<http://developer.couchbase.com/documentation/server/4.1/n1ql/n1ql-intro/data-access-using-n1ql.html>

<http://developer.couchbase.com/documentation/server/4.1/n1ql/n1ql-intro/n1ql-sql-differences.html>

Ресурсы по **Apache Phoenix**:

<http://phoenix.apache.org/>

<https://dzone.com/articles/apache-phoenix-sql-driver>

http://www.slideshare.net/Hadoop_Summit/w-145p230-ataylorv2
<https://bighadoop.wordpress.com/2014/05/17/apache-phoenix-an-sql-driver-for-hbase/>
<http://www.javaworld.com/article/3030214/big-data/open-source-java-projects-apache-phoenix.html>
<http://hadooptutorial.info/apache-phoenix-hbase-an-sql-layer-on-hbase/>

Ресурсы по **Apache Drill**:

<http://drill.apache.org/>
<http://drill.apache.org/docs/why-drill/>
<http://drill.apache.org/docs/architecture-introduction/>
<http://drill.apache.org/docs/drill-query-execution/>
<http://drill.apache.org/docs/core-modules/>
<http://drill.apache.org/docs/performance/>

Ресурсы по **Apache HAWQ**:

<http://hawq.incubator.apache.org/>
<http://pivotal.io/big-data/pivotal-hdb>
<http://pivotal.io/big-data/white-paper/a-true-sql-engine-for-hadoop-pivotal-hd-hawq>
<http://hdb.docs.pivotal.io/20/overview/HAWQOverview.html>
<http://hdb.docs.pivotal.io/hdb20/releasenotes/HAWQ20ReleaseNotes.html>
<http://www.slideshare.net/AGrishchenko/apache-hawq-architecture>

Ресурсы по **HP Vertica**:

<http://www.sql.ru/forum/1063820/faq-hp-vertica>
http://www.tadviser.ru/index.php/%D0%9F%D1%80%D0%BE%D0%B4%D1%83%D0%BA%D1%82:HP_Vertica_%D0%A1%D0%A3%D0%91%D0%94
<http://ascrus.blogspot.ru/>
<https://habrahabr.ru/post/190740/>
https://my.vertica.com/docs/7.2.x/PDF/HP_Vertica_7.2.x_Complete_Documentation.pdf

Ресурсы по **IBM BigSQL**:

<http://www.ibm.com/developerworks/ru/library/bd-bigsq/index.html>

Ресурсы по **SAP HANA Vora**:

<http://scn.sap.com/community/hana-in-memory/blog/2016/04/11/introducing-sap-hana-vora12>
<http://scn.sap.com/community/hana-in-memory/blog/2016/04/27/vora-12-modeling-tool>
<http://scn.sap.com/community/developer-center/hana/blog/2015/10/05/sap-hana-vora-hadoop>
<http://www.slideshare.net/SparkSummit/spark-summit-presentation-by-ken-tsai>

Ресурсы, содержащие **Введение в графовые СУБД**:

http://db-engines.com/en/blog_post/43
<http://db-engines.com/en/ranking/graph+dbms>
http://db-engines.com/en/blog_post/53
http://db-engines.com/en/ranking_categories
https://en.wikipedia.org/wiki/Graph_database (первая половина страницы)

Ресурсы по **Neo4j**:

<http://neo4j.com/>
<http://neo4j.com/developer/graph-database/>
<http://neo4j.com/developer/graph-db-vs-rdbms/>
<http://neo4j.com/developer/graph-db-vs-nosql/>
<http://neo4j.com/docs/stable/introduction.html>
<http://neo4j.com/docs/stable/graphdb-neo4j.html>
<http://neo4j.com/docs/stable/tutorial-comparing-models.html>
<http://neo4j.com/docs/stable/tutorials.html>

<http://neo4j.com/docs/stable/cypher-intro-patterns.html>
<http://neo4j.com/docs/stable/cypherdoc-patterns-in-practice.html>
<http://neo4j.com/docs/stable/cypherdoc-getting-the-results-you-want.html>
<http://neo4j.com/docs/stable/cypherdoc-how-to-compose-large-statements.html>
<http://neo4j.com/docs/stable/cypherdoc-labels-constraints-and-indexes.html>
<http://neo4j.com/docs/stable/cypherdoc-cypher-vs-sql.html>
<http://neo4j.com/docs/stable/cypher-query-lang.html>

Ресурсы по **OrientDB**:

<http://orientdb.com/why-orientdb/>
<http://orientdb.com/docs/latest/index.html>
<http://orientdb.com/docs/latest/Tutorial-Working-with-graphs.html>
<http://orientdb.com/docs/latest/Tutorial-Using-schema-with-graphs.html>
<http://orientdb.com/docs/latest/Graph-Consistency.html>

Ресурсы по **Titan**:

<http://thinkarelius.github.io/titan/>
<http://s3.thinkarelius.com/docs/titan/1.0.0/>
<http://s3.thinkarelius.com/docs/titan/1.0.0/benefits.html>
<http://s3.thinkarelius.com/docs/titan/1.0.0/arch-overview.html>
<http://s3.thinkarelius.com/docs/titan/1.0.0/getting-started.html>
<http://s3.thinkarelius.com/docs/titan/1.0.0/schema.html>
<http://s3.thinkarelius.com/docs/titan/1.0.0/gremlin.html>
<http://s3.thinkarelius.com/docs/titan/1.0.0/advanced-schema.html>
<http://s3.thinkarelius.com/docs/titan/1.0.0/graph-partitioning.html>
<http://www.allthingsdistributed.com/2015/08/titan-graphdb-integration-in-dynamodb.html>
<https://thinkarelius.com/blog/>

Ресурсы, содержащие **Введение в GraphEngines**:

<http://www.ibm.com/developerworks/ru/library/os-giraph/index.html>
https://en.wikipedia.org/wiki/Graph_database
<http://blog.octo.com/en/introduction-to-large-scale-graph-processing/>

Ресурсы по **Signal/Collect**:

<http://uzh.github.io/signal-collect/>
<http://uzh.github.io/signal-collect/documentation.html>

Ресурсы по **GraphLab**:

<https://en.wikipedia.org/wiki/GraphLab>
https://dato.com/products/create/open_source.html
<http://www.powergraph.ru/http://forum.dato.com/discussion/992/what-is-the-difference-between-graphlab-and-powergraph>
<http://www.analyticsvidhya.com/blog/2015/12/started-graphlab-python/>
<https://dato.com/solutions/machine-learning-algorithms.html>
<https://dato.com/learn/userguide/index.html>

Ресурсы по **PEGASUS**:

<http://www.cs.cmu.edu/~pegasus/>
<http://www.cs.cmu.edu/~pegasus/what%20is%20pegasus.htm>

Ресурсы, содержащие **Введение в графовые СУБД и GraphEngine в одном продукте**:

https://en.wikipedia.org/wiki/Graph_database (вторая половина страницы)

Ресурсы по **Microsoft Graph Engine**:

<http://research.microsoft.com/en-us/projects/graphengine/>

<http://www.graphengine.io/>
<http://www.graphengine.io/docs/manual/index.html>
<http://www.graphengine.io/docs/manual/basics.html>
<http://www.graphengine.io/docs/manual/TSL/index.html>
<http://www.graphengine.io/docs/manual/TSL/tsl-basics.html>
<http://www.graphengine.io/docs/manual/DataAccess/LINQ.html>

Ресурсы, содержащие **Введение в Search Engines:**

<http://db-engines.com/en/ranking/search+engine>
<http://db-engines.com/en/article/Search+Engines>
http://db-engines.com/en/blog_post/55

Ресурсы по **ApacheLucene:**

<http://lucene.apache.org/>
<http://wiki.apache.org/lucene-java/FrontPage?action=show&redirect=FrontPageEN>

Ресурсы по **ApacheNutch:**

<http://nutch.apache.org/>
<https://wiki.apache.org/nutch/>

Ресурсы по **ApacheSolr:**

<http://lucene.apache.org/solr/>
<http://lucene.apache.org/solr/resources.html#documentation>
https://ru.wikipedia.org/wiki/Apache_Solr
<https://www.drupal.org/project/apachesolr>

Ресурсы по **ElasticSearch:**

<https://www.elastic.co/products/elasticsearch>
<http://www.datanami.com/2015/01/22/solr-elasticsearch-question/>
<https://www.elastic.co/guide/index.html>
<http://solr-vs-elasticsearch.com/>

Ресурсы по **Splunk:**

<http://docs.splunk.com/Documentation/Splunk/6.2.0/Installation/Splunkarchitectureandwhatgetsininstalled>
<http://www.splunk.com/view/SP-CAAABF9>
<http://www.learnsplunk.com/splunk-architecture.html>
<https://bighadoop.wordpress.com/tag/splunk/>
<https://blog.rootshell.be/2012/12/22/howto-distributed-splunk-architecture/>
<http://www.slideshare.net/Splunk/taking-splunk-to-the-next-level-architecture-56251324>
<http://www.searchtechnologies.com/log-analytics-tools-open-source-vs-commercial>

9. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ СТУДЕНТОВ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Приступая к работе, каждый студент должен принимать во внимание следующие положения.

Дисциплина построена по модульному принципу, каждый модуль представляет собой логически завершённый раздел курса. Дисциплина делится на три модуля, включая экзамен.

На первом занятии каждый студент получает в электронном виде полный комплекс учебно-методических материалов по дисциплине.

Лекционные занятия посвящены рассмотрению ключевых, базовых положений курса и разъяснению учебных заданий, выносимых на самостоятельную проработку.

Лабораторные работы предназначены для приобретения опыта практической реализации основной профессиональной образовательной программы. Техническая литература по теме лабораторной работы прорабатывается студентами во время самостоятельной подготовки. Необходимый уровень подготовки контролируется перед проведением лабораторных работ.

Самостоятельная работа студентов включает проработку лекционного курса, подготовку к лабораторным работам, подготовку к рубежным контролям, выполнение домашних заданий, подготовку к экзамену.

Текущий контроль проводится в течение каждого модуля, его итоговые результаты складываются из оценок по следующим видам контрольных мероприятий:

- домашние задания;
- защита лабораторных работ;
- рубежные контроли.

Освоение дисциплины, ее успешное завершение на стадии промежуточной аттестации возможно только при регулярной работе во время семестра и планомерном прохождении текущего контроля.

Для завершения работы в семестре студент должен выполнить все контрольные мероприятия.

Промежуточная аттестация по результатам семестра по дисциплине проходит в форме экзамена, контролирующего освоение ключевых, базовых положений дисциплины, составляющих основу остаточных знаний по ней.

Методика оценки по рейтингу

Студент, выполнивший все предусмотренные учебным планом задания и сдавший все контрольные мероприятия, получает итоговую оценку по дисциплине за семестр в соответствии со шкалой:

Рейтинг	Оценка на экзамене
85 – 100	отлично
71 – 84	хорошо
60 – 70	удовлетворительно
0 – 59	неудовлетворительно

Оценивание дисциплины ведется в соответствии с Положением о текущем контроле успеваемости и промежуточной аттестации студентов МГТУ им. Н.Э. Баумана.

10. ПЕРЕЧЕНЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, ИСПОЛЬЗУЕМЫХ ПРИ ИЗУЧЕНИИ ДИСЦИПЛИНЫ, ВКЛЮЧАЯ ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ

В процессе преподавания дисциплины используются следующие методы, средства и программное обеспечение информационных технологий:

- e-mail преподавателя для оперативной связи: artem.sukhobokov@yandex.ru
- пакеты программ и облачные ресурсы, используемые в ходе выполнения лабораторных работ;
- электронные учебно-методические материалы для обеспечения самостоятельной работы студентов;
- презентации в среде PowerPoint и видео сюжеты по теме дисциплины;
- список ссылок в Интернет для поиска научно-технической информации по разделам дисциплины;
- офисный пакет приложений – Microsoft Office.

11. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ИЗУЧЕНИЯ ДИСЦИПЛИНЫ

Перечень материально-технического обеспечения дисциплины

№ п/п	Вид занятий	Вид и наименование оборудования
1.	Лекционные занятия	Аудитория с доской и проектором
2.	Лабораторные работы	Аудитория, оснащенная компьютерами с доступом к сети Интернет, пакеты прикладных программ
3.	Самостоятельная работа	Библиотека, имеющая рабочие места для студентов; Аудитории, оснащенные компьютерами с доступом к сети Интернет. Социокультурное пространство университета, позволяющее студенту качественно выполнять самостоятельную работу.

ЛИСТ ВНЕСЕННЫХ ИЗМЕНЕНИЙ И ДОПОЛНЕНИЙ

Номер изменения, дата внесения изменения, номер страницы для внесения изменений	
БЫЛО:	СТАЛО:
Основание:	
Подпись лица, ответственного за внесение изменений	