

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Локтионова Оксана Геннадьевна

Должность: проректор по учебной работе

Дата подписания: 28.03.2024 10:22:44

Уникальный программный ключ:

0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4851fda56d089

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное
учреждение высшего образования

«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра вычислительной техники

УТВЕРЖДАЮ

Проректор по учебной работе

 О.Г. Локтионова

« 26 » 03 2024 г.



СИСТЕМЫ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ

Методические указания по выполнению самостоятельной работы
для студентов направления подготовки 09.04.01

Курск 2024

УДК 004

Составитель: В.С. Панищев

Рецензент

Кандидат технических наук, доцент *Т.Н. Конаныхина*

Системы обработки больших данных: методические указания по выполнению самостоятельной работы для студентов направления подготовки 09.04.01 / Юго-Зап. гос. ун-т; сост.: В.С. Панищев. – Курск, 2024. –13 с.: Библиогр.: с. 13.

Методические указания соответствуют требованиям рабочей программе по дисциплине «Системы обработки больших данных» и разработанным оценочным средствам.

Предназначены для студентов направления подготовки 09.04.01 Информатика и вычислительная техника.

Содержат основные сведения об организации самостоятельной работы студентов. Описаны основные виды самостоятельной работы.

Текст печатается в авторской редакции

Подписано в печать . Форма 60x84 1/16.

Усл. печ. л. 0,76. Уч.-изд.л. 0,68. Тираж 50 экз. Заказ. ⁴⁶ Бесплатно

Юго-Западный государственный университет.

305040, г. Курск, ул. 50 лет Октября, 94.

Содержание

Введение	4
1 Организация самостоятельной работы студентов	5
2 Учебно-методическое обеспечение самостоятельной работы	6
3 Рекомендации по выполнению самостоятельной работы	7
4 Запланированные виды самостоятельной работы по дисциплине «Системы обработки больших данных»	9
5 Примеры заданий для текущего контроля	10
Список использованных источников	13

Введение

Самостоятельная работа – планируемая учебная, учебно-исследовательская, научно-исследовательская работа студентов, выполняемая во внеаудиторное время по заданию и при методическом руководстве преподавателя, но без его непосредственного участия.

Самостоятельная работа студентов (далее СРС) является важным видом учебной и научной деятельности студента. Самостоятельная работа студентов играет значительную роль в рейтинговой технологии обучения.

Формы самостоятельной работы студентов разнообразны. Они включают в себя:

- изучение лекционного материала по конспекту с использованием рекомендованной литературы;
- изучение учебной, научной и методической литературы, материалов периодических изданий с привлечением электронных средств официальной, статистической, периодической и научной информации с использованием информационно-поисковых систем, в частности глобальной сети «Интернет»;
- подготовку к собеседованию;
- подготовку к практическим работам;
- подготовку к выполнению лабораторных работ;
- выполнения индивидуального задания (решение задач, подготовка сообщений, докладов, исследовательские работы и т.п.);
- работу над творческими заданиями;
- подготовку кратких сообщений, докладов, рефератов, самостоятельное составление задач по изучаемой теме (по указанию преподавателя);
- работу над выполнением наглядных пособий (схем, таблиц и т.п.).

Самостоятельная работа приобщает студентов к научному творчеству, поиску и решению актуальных современных проблем.

1 Организация самостоятельной работы студентов

При организации СРС важным и необходимым условием становятся формирование умения самостоятельной работы для приобретения знаний, навыков и возможности организации учебной и научной деятельности.

Назначение самостоятельной работы студентов.

– **овладение знаниями**, что достигается чтением текста (учебника, первоисточника, дополнительной литературы), составлением плана текста, графическим структурированием текста, конспектированием текста, выписками из текста, работой со словарями и справочниками, ознакомлением с нормативными документами, выполнением учебно-исследовательской работы, поиском информации в сети Интернет и т.п.;

– **закрепление знаний**, что достигается работой с конспектом лекций, обработкой текста, повторной работой над учебным материалом (учебником, первоисточником, дополнительной литературой), оставлением плана, составлением таблиц для систематизации учебного материала, ответами на контрольные вопросы, заполнением рабочей тетради, аналитической обработкой текста (аннотирование, рецензирование, реферирование, конспект-анализ и др.), подготовкой мультимедиа сообщений/докладов к выступлению на семинаре (конференции), подготовкой реферата, составлением библиографии и т.п.;

– **формирование навыков и умений**, что достигается решением задач и упражнений по образцу, решением вариативных задач, выполнением чертежей, схем, выполнением расчетов (графических работ), решением ситуационных (профессиональных) задач, подготовкой к деловым играм, проектированием и моделированием разных видов и компонентов профессиональной деятельности, опытно экспериментальной работой и т.п.

Самостоятельная работа может осуществляться индивидуально или группами студентов в зависимости от цели, объема, конкретной тематики самостоятельной работы, уровня сложности, уровня умений студентов.

Контроль результатов внеаудиторной самостоятельной работы студентов может осуществляться в пределах времени, отведенного на

обязательные учебные занятия по дисциплине и внеаудиторную самостоятельную работу студентов по дисциплине, может проходить в письменной, устной или смешанной форме.

Текущий контроль качества выполнения самостоятельной работы может осуществляться с помощью:

- контрольного опроса;
- собеседования;
- компьютерного и/или бланкового тестирования.

2 Учебно-методическое обеспечение самостоятельной работы

Студенты могут при самостоятельном изучении отдельных тем и вопросов дисциплины пользоваться учебно-наглядными пособиями, учебным оборудованием в лабораториях и методическими разработками кафедры вычислительной техники в рабочее время, установленное Правилами внутреннего распорядка работников.

Учебно-методическое обеспечение самостоятельной работы обучающихся по данной дисциплине организуется:

библиотекой университета:

- библиотечный фонд укомплектован учебной, методической, периодической, справочной литературой в соответствии с нормативными документами по дисциплине;

- имеется доступ к основным информационным образовательным ресурсам, информационной базе данных, в том числе библиографической, возможность выхода в Интернет.

кафедрой:

- путем обеспечения доступности всего необходимого учебно-методического и справочного материала, обеспечение возможности выхода в Интернет;

- путем предоставления сведений о наличии учебно-методической литературы, современных программных средств.

- путем разработки:

- 1) заданий для самостоятельной работы;

- 2) вопросов к зачету;

3) методических указаний к выполнению практических работ и т.д.

типографией университета:

– помощь авторам в подготовке и издании научной, учебной и методической литературы;

– удовлетворение потребности в тиражировании научной, учебной и методической литературы.

3 Рекомендации по выполнению самостоятельной работы

Приступая к изучению новой учебной дисциплины, студенты должны ознакомиться с учебной программой, учебной, научной и методической литературой, имеющейся в библиотеке, получить в библиотеке рекомендованные учебники и учебно-методические пособия.

При подготовке к практическим занятиям и зачету следует в полной мере использовать курсы учебников, рекомендованных преподавателем, т.к. они дают более углубленное представление о проблемах, получивших систематическое изложение в учебнике.

Для освоения дисциплины в полном объеме студенту необходимо посещать все аудиторные занятия и самостоятельно прорабатывать полученный материал. Изучение теоретической части дисциплин способствует углублению и закреплению знаний, полученных во время аудиторных занятий, а также развивает у студентов творческие навыки, инициативы и умение организовать свое время.

При изучении дисциплины «Системы обработки больших данных» студентам рекомендуется самостоятельно готовиться по вопросам к собеседованию. Данные виды интеллектуальной практической деятельности способствуют закреплению навыков и знаний по дисциплине.

Собеседование – это вид самостоятельной работы студентов, заключающийся в освоении студентами темы на основе изучения литературы и подготовки развернутого ответа в соответствии с заданными вопросами.

Отличительными признаками подготовки к собеседованию являются:

- получение навыков передачи информации в устной форме;
- четкие формулировки;
- умение в сжатой форме изложить ключевые положения исследуемого вопроса и сделать выводы.

Материал, законспектированный в течение лекций, необходимо регулярно прорабатывать и дополнять сведениями из других источников литературы, представленных не только в программе дисциплины, но и в периодических изданиях. При самостоятельном изучении дисциплины, подготовке к аудиторным занятиям и выполнении домашних заданий студенты должны использовать рекомендованную учебную литературу и учебно-методические указания. Источники информации доступны на сайте кафедры. При освоении дисциплины сначала необходимо по каждой теме изучить рекомендованную литературу и составить краткий конспект основных положений, терминов, сведений, требующих запоминания и являющихся основополагающими в этой теме для освоения последующих тем курса. Для расширения знания по дисциплине рекомендуется использовать Интернет-ресурсы; проводить поиски в различных системах и использовать материалы сайтов, рекомендованных преподавателем. По требованию преподавателя конспект лекций предоставляется ему для проверки. Замеченные недостатки и внесенные замечания и предложения следует отработать в приемлемые сроки.

Подготовка к занятиям. Лабораторные и практические занятия углубляют, конкретизируют и расширяют знания, полученные на лекциях, помогают овладеть ими на более высоком уровне репродукции и трансформации. Эти виды учебного процесса способствуют закреплению умений и навыков самостоятельной работы, полученных в процессе работы над лекцией.

При подготовке к самостоятельной работе студент должен изучить соответствующие методические указания, а также подготовить вспомогательные материалы, необходимые для ее выполнения.

К практическим занятиям студент допускается только после инструктажа по технике безопасности. Положения техники безопасности изложены в инструкциях, которые имеются в аудиториях.

Подготовка к промежуточной аттестации. Основная функция зачета – обучающая, и только потом оценочная и воспитательная.

Серьезная и методически грамотно организованная работа по подготовке к практическим занятиям, написанию докладов и рефератов значительно облегчит подготовку к промежуточной аттестации.

4 Запланированные виды самостоятельной работы по дисциплине «Системы обработки больших данных»

Названия, содержание и объём тем (разделов) изучаемой дисциплины, а также перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины, приведены соответственно в рабочей программы дисциплины. Рекомендации по подготовке к занятиям приведены в методических указаниях по их выполнению (п.8.3 рабочей программы дисциплины) и содержат требования к видам и объёму самостоятельной работы при подготовке, выполнении, оформлении отчетов и защите работ на занятиях.

Содержание самостоятельной работы по дисциплине приведено в таблице 1.

Таблица 1 – Самостоятельная работа студентов по дисциплине «Системы обработки больших данных»

№ раздела	Название раздела (темы) дисциплины	Срок выполнения
1	Фреймворки обработки больших данных	8 неделя
2	Обработка потоковых данных	16 неделя

Текущий контроль знаний, основанный на определении качества выполненной самостоятельной работы студентов, производится по указанным в рабочих программах дисциплины

неделям семестра (таблица 4.1.2 РПД по дисциплине) и предусматривает собеседование (С) в четырёх контрольных точках и тестирование (Т).

5 Примеры заданий для текущего контроля

1. Четвертая промышленная революция. Концепция «Индустрии 4.0». Значение и роль данных в современном мире.
2. Понятие больших данных (Big Data). Модель mV для больших данных. Источники больших данных. Применение больших данных.
3. Методы анализа больших данных. Подходы к анализу больших данных. Проблемы, связанные с обработкой больших данных традиционными инструментами. Требования к системам BigData.
4. Понятие технологии MapReduce, основные принципы, лежащие в ее основе. Характеристики этапов map и reduce. Примеры задач MapReduce.
5. Фреймворк Hadoop как распределенная система обработки больших данных. Экосистема Hadoop. Масштабируемость и производительность Hadoop. Принцип определения доступности узлов. Понятие Heartbeat.
6. Понятие HDFS. Организация HDFS. Архитектура HDFS. NameNode, SecondaryNameNode и DataNode: понятие и функции. Обеспечение надежности хранения данных. Понятие репликации. Характеристики и ключевые особенности HDFS как файловой системы.
7. Процесс чтения и записи данных в HDFS. Консольный клиент.
8. Понятие и назначение YARN. Архитектура YARN. ResourceManager и NodeManager: понятие и функции. Понятие контейнеров (container) YARN. Отличия MapReduce v1 от MapReduce v2.
9. Процесс запуска и исполнения приложения под управлением YARN. Понятие ApplicationMaster. Взаимодействие компонентов YARN. Понятие локальности данных. Организация отказоустойчивости YARN.
10. Основные планировщики YARN: FIFO, Capacity, Fair. Их назначение и ключевые особенности.
11. Процесс выполнения приложения в Hadoop MapReduce. Понятие задания (job), задачи (task), контекста (context). Конфигурация параметров Hadoop, примеры. Поток данных в Hadoop.
12. Подготовка данных для этапа Map в Hadoop. Класс InputFormat, его наследники. Split и RecordReader, их взаимосвязь. Поток данных до этапа Map.

13. Этап Map в Hadoop. Класс Mapper. Методы map(), setup(), cleanup(). Пример реализации Map. Класс Partitioner.
14. Обработка промежуточных записей. Кольцевой буфер и операция spill. Этапы sort и combine. Перемешивание данных (shuffle) и его влияние на производительность.
15. Операция Reduce в Hadoop. Класс Reducer. Методы reduce(), setup(), cleanup(). Пример реализации Reduce. Сохранение полученных результатов. Класс OutputFormat. Класс RecordWriter.
16. Обмен данными между основными этапами в MapReduce Hadoop. Понятие сериализации. Интерфейсы Writable и WritableComparable. Основные Writable-классы в Hadoop.
17. Фреймворк обработки больших данных Apache Spark, его назначение, функции и отличия от Hadoop MapReduce. Экосистема Spark. API фреймворка.
18. Архитектура приложения Apache Spark. Модели запуска Spark-приложений (YARN, Mesos, Standalone, Kubernetes). Понятие драйвера (driver) и исполнителей (executors). Процесс запуска Spark-приложения под управлением YARN.
19. Понятие устойчивого распределенного набора данных (RDD). Понятие раздела RDD (partition). Способы создания RDD. Трансформации (transformations) и действия (actions). Примеры. Внутренние свойства RDD.
20. RDD и PairRDD: понятие, назначение. Основные трансформации (transformations) и действия (actions) над ними. Понятие кортежа (tuple2), связь между JavaPairRDD и JavaRDD посредством кортежа.
21. Реализация концепции MapReduce в фреймворке Spark. Методы map, flatMap, mapValues, flatMapValues, mapPartitions, mapToPair, filter.
22. Реализация концепции MapReduce в Spark. Особенности реализации Reduce, требования к функции редукции. Методы reduce и reduceByKey.
23. Управление распределением данных. Класс Partitioner и его наследники. Способы распределения по разделам. Сохранение раздела при трансформациях. Влияние распределения по разделам на производительность. Примеры.
24. Понятие базы данных и хранилища данных (warehouse). ELT и ETL. Архитектура хранилищ данных. Примеры.

25. Понятие озера данных. Основные элементы озера данных. Архитектура озера данных. Примеры технологий для построения озера. Отличия озера данных от хранилища данных. Перспективы развития озер данных.
26. Понятие распределенной системы, проблемы построения распределенных систем. Частичные сбои. CAP-теорема и BASE-принцип. Строгая (strict) согласованность и согласованность в конечном счете (eventual).
27. Понятие NoSQL баз данных. Особенности NoSQL БД и отличия их от реляционных БД. Типы NoSQL баз данных. Их особенности, преимущества и недостатки. Примеры.

Полностью оценочные средства представлены в учебно-методическом комплексе дисциплины.

Промежуточная аттестация по дисциплине проводится в форме зачета, который проводится в форме тестирования (бланкового и/или компьютерного). Для тестирования используются контрольно-измерительные материалы (КИМ) – задания в тестовой форме, составляющие банк тестовых заданий (БТЗ) по дисциплине, утвержденный в установленном в университете порядке.

Проверяемыми на промежуточной аттестации элементами содержания являются темы дисциплины, указанные в разделе 4 рабочей программы дисциплины.

Для проверки знаний используются вопросы и задания в различных формах:

- закрытой (с выбором правильного ответа),
- открытой,
- на установление правильной последовательности,
- на установление соответствия.

Умения, навыки и компетенции контролируются в ходе выполнения и защиты лабораторных работ и практических занятий, выполнения внеаудиторной контрольной работы.

В каждый вариант КИМ включаются задания по каждому проверяемому элементу содержания во всех перечисленных формах и разного уровня сложности. Такой формат КИМ позволяет объективно определить качество освоения обучающимися основных элементов содержания дисциплины и уровень сформированности компетенций.

Список использованных источников

1. Алханов, А. Самостоятельная работа студентов / А.Алханов // Высшее образование в России. – 2005. – №11. – С.86-89.
2. Гладышева М.М., Тутарова В.Д., Польшиков А.В. Формирование исследовательских компетенций студентов в процессе самостоятельной учебной работы в техническом вузе // Высшее образование сегодня. – 2010. – № 3. – С. 24-26.
3. Измайлова М.А. Организация внеаудиторной самостоятельной работы студентов: Методическое пособие. – М.: Издательско-торговая корпорация «Дашков и К°», 2008. – 64 с.
4. Росина, Н. Организация СРС в контексте инновационного образования / Н. Росина // Высшее образование в России. – 2006. – №7. – С.109-114.