

Министерство науки и высшего образования РФ

Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Хакасский технический институт – филиал федерального государственного
автономного образовательного учреждения высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине (модулю) Б1.В. 13 Технологии BigData
индекс и наименование дисциплины или практики в соответствии с ФГОС ВО и учебным планом

Направление подготовки 09.03.03 Прикладная информатика
код и наименование направления подготовки

Направленность 09.03.03.04 Прикладная информатика в государственном и муниципальном
управлении
код и наименование направленности

1 Перечень компетенций с указанием индикаторов их достижения, соотнесенных с результатами обучения по дисциплине (модулю), практики и оценочными средствами

Семестр	Код и содержание индикатора компетенции	Результаты обучения	Оценочные средства
Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач (УК-1)			
7	УК-1.1. Знает принципы сбора, отбора и обобщения информации, методики системного подхода для решения профессиональных задач.	Знать принципы сбора, отбора и обобщения информации.	ОС-1
7	УК-1.2. Умеет анализировать и систематизировать разнородные данные, оценивать эффективность процедур анализа проблем и принятия решений в профессиональной деятельности.	Уметь анализировать и систематизировать разнородные данные.	ОС-2, Вопросы к зачету
7	УК-1.3. Владеет навыками научного поиска и практической работы с информационными источниками; методами принятия решений.	Владеть навыками систематизации наблюдаемых данных.	ОС-2, Вопросы к зачету
Способен осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач (ПК-9)			
7	ПК-9.1. Знает инструменты и методы проектирования и верификации структуры базы данных ИС.	Знать теоретические основы организации распределенных вычислений.	ОС-2, Вопросы к зачету
7	ПК-9.2. Умеет разрабатывать структуру баз данных.	Уметь реализовывать параллельные алгоритмы обработки данных на высокоуровневых языках программирования с использованием библиотек. Владеть навыками использования программных средств, поддерживающих технологии BigData, для решения практических задач в предметной области.	ОС-2, ОС-3, Вопросы к зачету

2 Типовые оценочные средства или иные материалы, с описанием шкал оценивания и методическими материалами, определяющими процедуру проведения и оценивания достижения результатов обучения

2.1 Оценочные средства для текущего контроля.

Текущий контроль знаний необходим для проверки усвоения учебного материала и его закрепления. Контроль следует проводить на протяжении всего периода изучения дисциплины. Текущий контроль осуществляется на контрольной неделе и на практических занятиях.

Оценочное средство 1 – Тест 1«Анализ данных» (ОС-1).

Оценка этапа сформированности компетенции производится на 1 контрольной неделе. Тест выполняется вне аудитории. На выполнение теста отводится 40 минут. Основная задача теста – оценить знания студентов по основным понятиям анализа данных.

Вариант тестового задания:

1. Эффективный поиск в больших массивах данных требует распараллеливанияВыберите один или несколько ответов:

- 1) процессов ввода/вывода (+)
- 2) процессов соединений
- 3) процессов вычислений (+)

2. Что понимается под термином «наложение рутины на вершину»?Выберите один ответ:

- 1) одно событие влечет за собой другое событие

2) использование описания одной рутины для одинаковых по функционированию периферийных узлов (+)

3) действия, описанные в событийной секции, выполняются мгновенно

3. Какое название получила технология использования глобальной сети компьютеров для решения сложной задачи? Выберите один ответ:

1) Grid (+)

2) Site

3) Web

4. Под степенью отношения понимается... Выберите один ответ:

1) количество атрибутов (+)

2) количество кортежей

3) количество строк

5. С помощью какого алгоритма производится поиск идентификатора с наибольшей оценкой? Выберите один ответ:

1) децентрализованного алгоритма (+)

2) алгоритма смещения

3) алгоритма Тарри

6. В каком слое отсутствует описание связей между узлами? Выберите один ответ:

1) сообщений

2) структуры

3) алгоритмов (+)

7. Подсистема подготовки и принятия решений использует информацию, получаемую из... Выберите один ответ:

1) подсистемы мониторинга (+)

2) хранилища РРИСО

3) подсистемы внутреннего документооборота

8. В каком случае возникает вертикальная фрагментация? Выберите один ответ:

1) по различным сайтам распределена информация о различных сущностях предметной области

2) источники данных и пользователи привязаны к сайтам, находящимся в различных точках физического пространства, а реляционные схемы, которыми они пользуются, могут быть одинаковы

3) об одних и тех же сущностях на разных сайтах имеется информация разного рода (+)

9. Для чего необходимо вести централизованную базу данных РРИСО? Выберите один ответ:

1) для обеспечения управления системой (+)

2) для формирования типовых документов

3) для обеспечения единого интерфейса пользователя

10. К репликации можно отнести... Выберите один или несколько ответов:

1) создание копий некоторых фрагментов отношений (+)

2) одновременное хранение нескольких копий на разных сайтах (+)

3) повышение надежности хранения данных

11. Транзакция – это ... Выберите один ответ:

1) некоторый набор операций над базой данных, который рассматривается как единственное завершённое, с точки зрения пользователя, действие над некоторой информацией, обычно связанное с обращением к базе данных (+)

2) разновидность систем хранения, ориентированная на поддержку процесса анализа данных, обеспечивающая целостность, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов

3) высокоуровневые средства отражения информационной модели и описания структуры данных

4) установление зависимости дискретной выходной переменной от входных переменных

12. Укажите правильное соответствие между типами архитектур хранилищ данных и их определениями

Архитектура	Определение
1) Глобальное хранилище данных (Global data warehouse) или хранилище данных масштаба организации (С)	А) управляются рабочими группами, отделами или направлениями бизнеса, и разрабатываются исключительно для реализации аналитических потребностей последних
2) Независимые киоски данных (Stand-alone Data Marts) (А)	В) управляются рабочими группами, отделами или направлениями бизнеса, но разрабатываются в рамках единой для организации схемы удовлетворения информационных и аналитических потребностей
3) Связанные киоски данных (В)	С) это такое ХД, в котором будут поддерживаться все, или большая часть, данных организации

13. Укажите правильное соответствие между вариантами реализации хранилища данных и их характеристиками

Варианты реализации	Характеристики
1) Виртуальное хранилище данных (С)	А) Архитектура представляет собой единый источник интегрированных данных организации
2) Глобальное хранилище данных (А)	В) Архитектура является разновидностью глобального ХД, в которую технологически реализуются три уровня (На первом уровне располагается корпоративное ХД организации. На втором уровне поддерживаются связанные киоски данных тематической направленности на основе многомерной СУБД. На третьем уровне находятся клиентские приложения пользователей с установленными на них средствами анализа данных)
3) Хранилища данных с многоуровневой (в основном трехзвенной) архитектурой или корпоративные ХД (В)	С) Архитектура обеспечивает доступ к "живым" данным в режиме реального времени через программное обеспечение промежуточного слоя. В основе такого решения лежит репозиторий метаданных, который описывает источники данных, процедуры их предварительной обработки и форматы представления информации конечному пользователю. Недостаток такого решения - интенсивный сетевой трафик, снижение производительности несущей системы, угроза нарушения целостности данных в случае неудачных действий пользователей ХД
4) Встроенные (комбинированные) хранилища данных (Е)	Д) В этой архитектуре ХД не является единым физическим репозиторием (в отличие от СIF). Это "виртуальное" ХД, представляющее коллекцию витрин данных, каждая из которых имеет архитектуру типа "звезда"
5) Корпоративная информационная фабрика (Corporate Information Factory, CIF) (F)	Е) Архитектура представляет собой ХД, которые органически встраиваются в виртуальное предприятие (Enterprise Information Factory, EIF) или используются

	как компонент аналитической поддержки в информационной реализации бизнес-функций
6) Хранилище данных с архитектурой шины данных (DataWarehouseBus) (D)	F) Архитектура является развитием архитектуры корпоративного ХД (enterprisedatawarehouse, EDW). Использование данной архитектуры предполагает скоординированное извлечение данных из источников, загрузку их в реляционную БД со структурой в третьей нормальной форме, использование построенного ХД для наполнения дополнительных репозиториях презентационных данных
7) Объединенное (федеративное) ХД (H)	G) Архитектура представляет собой облегченный вариант ХД тематической направленности
8) Киоскиданных (G)	H) В этой архитектуре ХД состоит из ряда экземпляров ХД, которые функционируют на полуавтономной основе и, как правило, организационно или географически разнесены, однако могут рассматриваться и управляться как одно большое ХД

14. Какие команды относятся к категории команд DML (язык манипулирования данными)? Выберите один ответ:

- 1) CREATE, ALTER, DROP
- 2) INSERT, UPDATE, DELETE (+)
- 3) SELECT, WHERE, ORDER
- 4) TABLE, COLUMN, DATABASE

Критерии оценивания:

- «**ЗАЧТЕНО**» выставляется обучающемуся, если он выполнил 80 % и более тестовых заданий верно.

- «**НЕ ЗАЧТЕНО**» выставляется обучающемуся, если он выполнил менее 80 % тестовых заданий верно.

В случае выполнения тестовых заданий на оценку «не зачтено», необходимо выполнить повторную диагностику.

Оценочное средство 2 – Практико-ориентированные задания(ОС-2)

Оценка этапа сформированности компетенции производится на каждом лабораторном занятии при выполнении определенного раздела дисциплины. Задания лабораторных работ выполняются в аудитории. Основная задача задания – оценить навыки использования программных средств, поддерживающих технологии BigData, для решения практических задач.

Лабораторная работа «Введение в СУБД MongoDB»

Задание. Используя СУБД MongoDB, разработайте базу данных, предназначенную для хранения логов веб-сервера. Лог включает в себя следующие поля: адрес ресурса (URL), IP-адрес пользовательского компьютера, отметка времени начала просмотра ресурса, длительность просмотра ресурса.

Лабораторная работа «Работа с БД в СУБД MongoDB»

Задание. Разработайте и протестируйте запросы на выборку данных из созданных коллекций.

Лабораторная работа «Основы OpenMP»

Задание 1. Входные данные:

N – размер массивов, x – вводимый параметр, для параллельной программы; k – количество потоков.

Задание:

- инициализировать массивы А и В в соответствии с выражением (см. вариант);
- выполнить вычисление массива С и переопределение массива В;
- по требованию выполнить вывод всех массивов на экран (в файл);
- предусмотреть возможность исполнения некоторых параллельных секций числом потоков, задаваемых переменной k;
- при входе в параллельную секцию на экран выводить краткое описание секции – число исполняющих нитей.

Задание 2.

- Проанализировать программу Задания 1 и определить, какие фрагменты программы целесообразно реализовать в виде секции секций.
- Распараллелить программу с использованием директивы `#pragma omp parallel sections`. Кроме указанной директивы допускается использование функций библиотеки Open MP для работы с переменными окружения.

- Убедиться, что прежняя и новая версии выдают одинаковые результаты.
- Зафиксировать время выполнения программы при различных размерах обрабатываемых массивов.

Задание 3.

- Проанализировать программу Задания 1 и определить, какие фрагменты программы целесообразно/возможно распараллелить, используя директиву `for`

- Распараллелить программу с использованием директивы `#pragma omp for` (статическое планирование исполнения)

- Убедиться, что прежняя и новая версии выдают одинаковые результаты.
- Зафиксировать время выполнения программы при различных размерах обрабатываемых массивов.

- Изменить способ планирования выполнения циклов на любой другой. Повторить эксперимент.

Лабораторная работа «Инструменты работы с большими данными: платформа Python, библиотеки работы с табличными данными - Numpy и Pandas»

Задание.

1. Загрузить набор данных (находится в задании);
2. Привести данные к одному виду (можно использовать `StandardScaler` или `MinMaxScaler`).
3. Случайно разделить выборку на тренировочную и тестирующую (размер соотношения студент устанавливает по желанию, например, 7:3).
4. На тренировочной выборке построить несколько моделей, предсказывающих цену квартиры. При построении модели выполнить поиск гиперпараметров с помощью поиска по сетке (`GridSearchCV`) или случайного поиска (`RandomizedSearchCV`).
5. Сравнить модели по ошибке на тестовом наборе данных и выбрать итоговую модель.

Критерии оценивания практических заданий:

- «**ОТЛИЧНО**» (84-100 баллов) выставляется обучающемуся, если он полностью выполнил задание.

- «**ХОРОШО**» (67-83 балла) выставляется обучающемуся, если он полностью выполнил задание, но есть замечания.

- «**УДОВЛЕТВОРИТЕЛЬНО**» (50-66 баллов) выставляется обучающемуся, если он выполнил 50 % задания.

- «**НЕ УДОВЛЕТВОРИТЕЛЬНО**» выставляется обучающемуся, если он выполнил менее 50% задания.

До конца учебного семестра должны быть выполнены все практические задания на оценку «УДОВЛЕТВОРИТЕЛЬНО» или выше для достижения этапа формирования компетенции.

Оценочное средство 3 – Контрольные вопросы к лабораторным работам(ОС-3)

Оценка этапа сформированности компетенции производится на лабораторном занятии.

Лабораторная работа «Введение в СУБД MongoDB»

1. Какую модель данных поддерживает СУБД MongoDB?
2. Какой формат данных поддерживается в MongoDB?
3. Из чего состоит база данных в MongoDB?
4. Имеет ли БД в MongoDB жесткую структуру?

Лабораторная работа «Работа с БД в СУБД MongoDB»

1. Приведите синтаксис метода «Find».
2. Возможно, ли производить поиск по нескольким полям?
3. Каким образом можно произвести сортировку по нескольким полям? Приведите пример запроса.
4. Можно ли смешивать включения и исключения полей при составлении запроса?

Лабораторная работа «Основы OpenMP»

1. Что такое OpenMP?
2. Из каких компонент состоит OpenMP?
3. Какова структура OpenMP программы?
4. Каким образом создаются нити в OpenMP программе?
5. Каким образом выполняется распределение вычислений в OpenMP?
6. Каким образом происходит в OpenMP взаимодействие между нитями?
7. Какие классы данных фигурируют в параллельной области?
8. Чем определяется количество нитей в параллельной области?
9. Каким образом можно изменить количество нитей в параллельной области?
10. Каким образом выполняется синхронизация нитей?
11. Назовите важнейшие переменные окружения OpenMP.
12. Назначение встроенных функций в OpenMP.

Лабораторная работа «Инструменты работы с большими данными: платформа Python, библиотеки работы с табличными данными - NumPy и Pandas»

1. Как можно обнаружить выбросы в наборе данных?
2. Какие способы визуализации данных существуют?
3. Что такое тренд?
4. Чем сезонность отличается от цикличности?
5. Как можно заполнить пропуски в данных?
6. Как импортируются данные в формате .csv?
7. Какие параметры, аргументы и методы у класса KMeans?

Критерии оценивания ответов на вопросы:

- «**ЗАЧТЕНО**» выставляется обучающемуся, если он ответил на 2 вопроса из предлагаемого списка.
- «**НЕ ЗАЧТЕНО**» выставляется обучающемуся, если он ответил менее, чем на 2 вопроса из предлагаемого списка.

2.2 Оценочные средства для промежуточной аттестации

Учебным планом изучения дисциплины предусмотрен зачет в 7 семестре.

1. ВОПРОСЫ К ЗАЧЕТУ

1. Проблемы применения и узкие места BigData. Точки роста технологий BigData.
2. Конвергенция BigData и классических технологий высокопроизводительных вычислений.
3. Недостатки Map/Reduce.
4. Настройки над Hadoop.
5. Обзор технологий стека Apache BigData.
6. Рассмотрение элементов стека Apache с примерами на языках высокого уровня

7. Распределенная файловая система Hadoop. Принцип доставки вычислений к данным. Метод Map/Reduce. Примеры реализации на языках программирования

Для получения зачета необходимо выполнить все виды оцениваемой учебной деятельности: лабораторные работы, решить тестовые задания. В случае невыполнения какого-либо вида работы студенту необходимо выполнить одно из практических заданий и ответить на один теоретический вопрос из предложенного перечня.

Критерии для выставления зачета

- «**ЗАЧТЕНО**»выставляется обучающемуся, если он выполнил все виды оцениваемой учебной деятельности .

- «**НЕ ЗАЧТЕНО**»выставляется обучающемуся, если он не выполнил хотя бы один из видов оцениваемой учебной деятельности.

2. ПРОЦЕДУРА ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Сдача зачета производится в последнюю неделю обучения. Ведущим преподавателем может быть проведена промежуточная аттестация студента по результатам обучения без дополнительной сдачи зачета по вопросам.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических возможностей (подбираются индивидуально в зависимости от возможностей здоровья студента):

Категории студентов	Виды оценочных средств	Форма контроля и оценки результатов обучения
С нарушением слуха	Контрольные вопросы для зачета	Преимущественно письменная проверка
С нарушением зрения	Контрольные вопросы для зачета	Преимущественно устная проверка (индивидуально)
С нарушением опорно-двигательного аппарата	Контрольные вопросы для зачета	Письменная проверка

Разработчик: _____ / М. А. Буреева