Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования

Московский государственный университет имени М.В. Ломоносова

Механико-математический факультет

Кафедра теоретической информатики

УТВЕРЖДАЮ

Заведующий кафедрой

\_\_\_\_\_\_\_\_\_\_\_\_\_\_/А.В.Михалев /

«22» января 2019 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**Наименование дисциплины :**

**Базы данных: дополнительные главы**

**Уровень высшего образования:**

**магистратура**

**Направление подготовки (специальность):**

**02.04.01 Математика и компьютерные науки**

**Направленность (профиль) ОПОП:**

**Математика**

Форма обучения:

очная

Рабочая программа рассмотрена и одобрена

на заседании кафедры Теоретической информатики

(протокол №2а, «22» января 2019 года)

Москва 2019

Рабочая программа дисциплины разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по специальности «Математика и компьютерные науки» ( программымагистратуры) в редакции приказа МГУ от 30 декабря 2016 г.

Год (годы) приема на обучение\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**1.** Место дисциплины в структуре ОПОП ВО. **Вариативная часть ОПОП ВО***.*

**2.** Входные требования для освоения дисциплины , предварительные условия : знание основ статистики, линейной алгебры и программирования.

**3.** Результаты обучения по дисциплине, соотнесенные с требуемыми компетенциями выпускников*.*

|  |  |
| --- | --- |
| **Компетенции выпускников (коды)** | **Планируемые результаты обучения по дисциплине , соотнесенные с компетенциями** |
| *ПК-1, ПК-5* | ***Знать:*** *современные методы эффективного представления и обработки данных системами управления базами данных (СУБД) в памяти, NoSQL модели данных, технологии распределенного реестра, блокчейн.*  ***Уметь*** *ориентироваться в современных методах и алгоритмах компьютерной математики, уметь решать стандартные задачи по изученным темам.*  ***Владеть:*** *прикладными методами решения практических задач по изученным темам.* |
| *ОПК-1, ОПК-2, ОПК-4* | ***Уметь*** *решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий*  ***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***  ***Иметь опыт*** *использования на практике математических алгоритмов, в том числе с применением современных вычислительных систем*  *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_* |
| *УК-1* | ***Уметь*** *формулировать научно обоснованные гипотезы, создавать теоретические модели явлений и процессов, применять методологию научного познания в профессиональной деятельности*  ***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***  ***Владеть*** *специальными разделами фундаментальной математики, методами анализа и решения задач*  *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_* |

**4.** Формат обучения очный.

**5.** Объем дисциплины составляет **4**  з.е., в том числе **50** академических часов, отведенных на контактную работу обучающихся с преподавателем, **30**  академических часов на самостоятельную работу обучающихся.

**6.** Содержание дисциплины , структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Наименование и краткое содержание разделов и тем дисциплины ,**  **Форма промежуточной аттестации по дисциплине** | **Всего**  **(часы**) | | В том числе | | | |
| **Контактная работа  (работа во взаимодействии с преподавателем)**  **Виды контактной работы, часы** | | | **Самостоятельная работа обучающегося,**  **часы** |
| Занятия лекционного типа\* | Занятия семинарского типа\* | **Всего** |  |
| Тема 1. Прошлое, настоящее и будущее корпоративных приложений. | 4 | | 2 |  | 2 | 2 |
| Тема 2. Методы хранения баз данных, словарное кодирование, сжатие данных. | 4 | | 2 |  | 2 | 2 |
| Тема 3. Размещение данных в оперативной памяти и секционирование. | 6 | | 4 |  | 4 | 2 |
| Тема 4. Архивирование и старение данных, мульти-температурное управление данными, фактические и исторические данные. | 4 | | 2 |  | 2 | 2 |
| Тема 5. Операторы манипулирования данными (DML). | 6 | | 4 |  | 4 | 2 |
| Тема 6. Стратегии материализации и алгоритмы соединения таблиц. | 4 | | 2 |  | 2 | 2 |
| Текущий контроль успеваемости | 4 | |  | 2 | 2 | 2 |
| Тема 7. Архитектура базы данных в памяти. Индексы, кеширование для агрегатов, резервные копии и восстановление. | 6 | | 4 |  | 4 | 2 |
| Тема 8. Горячее резервирование и k-безопасность, реплики, бизнес-логика в базе данных. | 4 | | 2 |  | 2 | 2 |
| Тема 9. Параллельная обработка данных, закон Амдала. Конфликты и изолированность транзакций. | 4 | | 2 |  | 2 | 2 |
| Тема 10. Прогнозирование и интеллектуальный анализ текстов в базах данных в памяти. | 4 | | 2 |  | 2 | 2 |
| Тема 11. Базы данных NoSQL, теорема CAP, размещение данных. | 6 | | 4 |  | 4 | 2 |
| Тема 12. Базы данных и технология распределенного реестра, консенсус модели. | 6 | | 4 |  | 4 | 2 |
| Консультации | 4 | | | | 2 | 2 |
| Экзамен | 4 | | | | 4 |  |
| Промежуточная аттестация – контрольная работа | 10 | | | | 8 | 2 |
| **Итого** | *80* | *50* | | | | 30 |

7. Фонд оценочных средств (ФОС) для оценивания результатов обучения по дисциплине

7.1. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости.

1. Дана таблица R с атрибутом City размером 50 Byte.

Количество записей в таблице – 14 000 000.

Количество уникальных значений атрибута City – 3000.

Вычислить объем данных в случае:

– построчного хранения;

– поколоночного хранения со словарным кодированием;

– поколоночного хранения со словарным кодированием и со сжатием вектора атрибута с помощью кодирования длин серий (Run Length Encoding).

2. Таблица, содержащая информацию о запасах продукта, имеет следующие атрибуты:

Warehouse (4 byte); Product Id (4 byte); Product Name Short (20 byte); Product Name Long (40 byte); Self Production (1 byte); Production Plant (4 byte); Product group (4 byte); Sector (4 byte); Stock Volume (8 byte); Unit of Measure (3 byte); Price (8 byte); Currency (3 byte); Total Stock Value (8 byte); Stock Currency (3 byte).

Размер строки кэша равен 64 байта. Скорость сканирования ядра процессора составляет 4 МБ в миллисекунду.

Подсчитать время реконструкции кортежа при построчном и поколоночном хранении.

3. Каково максимальное теоретическое ускорение программы:

– состоящей из 50% последовательной части и

– выполняемой четырьмя отдельными процессорами?

4. Сколько бит минимально необходимо для представления 1140 различных значений в векторе атрибута при использовании словарного кодирования?

7.2. Типовые контрольные задания или иные материалы для проведения промежуточной аттестации.

1. Корпоративные приложения: современные требования, проблемы в аппаратном и программном обеспечении и способы их решения, характеристики современных корпоративных приложений, OLTP и OLAP системы

2. Словарное кодирование в In-Memory базе данных, понятие энтропии

3. Сжатие данных в In-Memory базе данных: префиксное кодирование (Prefix Encoding), разреженное кодирование (Sparse encoding), непрямое кодирование (Indirect encoding)

4. Сжатие данных в In-Memory базе данных: кодирование длин серий (Run-length encoding), кластерное кодирование (Cluster Encoding), дельта-кодирование (Delta encoding)

5. Методы хранения баз данных: секционирование, мульти-температурное управление данными, фактические и исторические данные

6. Операторы манипулирования данными в In-Memory базе данных: вставка, обновление, удаление, " insert only"-подход

7. Оператор выборки в In-Memory базе данных: стратегии материализации, агрегатные функции

8. Построчное и поколоночное размещение данных: сравнение подходов при реконструкции кортежей

9. Построчное и поколоночное размещение данных: сравнение подходов при полном сканировании таблицы, шаговом доступе и полном сканировании столбца

10. In-Memory СУБД: архитектура, понятие дифференциального буфера и процесс слияния, кеш для агрегатов

11. Индексы в In-Memory базе данных, инвертированный индекс, эффективность инвертированного индекса при словарном кодировании

12. Параллелизм на уровне аппаратного обеспечения, параллелизм на уровне программного обеспечения: закон Амдала

13. NoSQL: идея NoSQL, теорема CAP, размещение данных

14. Базы данных и технология распределенного реестра, блокчейн, дерево Меркла

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине (модулю)** | | | | |
| Оценка  РО и соответствующие виды оценочных средств | 2 | 3 | 4 | 5 |
| **Знания**  *(виды оценочных средств: устные и письменные опросы и контрольные работы, тесты, и т.п. )* | Отсутствие знаний | Фрагментарные знания | Общие, но не структурированные знания | Сформированные систематические знания |
| **Умения**  *(виды оценочных средств: практические контрольные задания, написание и защита рефератов на заданную тему и т.п.)* | Отсутствие умений | В целом успешное, но не систематическое умение | В целом успешное, но содержащее отдельные пробелы умение (допускает неточности непринципиального характера) | Успешное и систематическое умение |
| **Навыки  (владения, опыт деятельности)**  *(виды оценочных средств: выполнение и защита курсовой работы, отчет по практике, отчет по НИР и т.п.)* | Отсутствие навыков (владений, опыта) | Наличие отдельных навыков (наличие фрагментарного опыта) | В целом, сформированные навыки (владения), но используемые не в активной форме | Сформированные навыки (владения), применяемые при решении задач |

8. Ресурсное обеспечение:

* Перечень основной и дополнительной литературы,
* А. Основная литература
* 1. Hasso Plattner. A Course in In-Memory Data Management: The Inner Mechanics of In-Memory Databases. Second Edition. Springer-Verlag Berlin Heidelberg, 2014.
* 2. Jeffrey Word. SAP HANA Essentials. 5th Edition. Epistemy Press, 2013.
* 3. Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller, Steven Goldfeder. Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction. Princeton University Press, 2016.
* Б. Дополнительная литература
* 1. Hasso Plattner, Bernd Leukert. The In-Memory Revolution: How SAP HANA Enables Business of the Future. Springer International Publishing, 2016.
* 2. Guy Harrison. Next Generation Databases: NoSQL, NewSQL and Big Data. Apress, 2015.
* 3. Christoph Meinel, Tatiana Gayvoronskaya, Maxim Schnjakin. Blockchain: Hype or Innovation. Universitätsverlag Potsdam, 2018.
* 4. Mark Needham and Amy E. Hodler. Graph Algorithms: Practical Examples in Apache Spark and Neo4j. O’Reilly Media, 2019.

Описание материально-технического обеспечения.

Компьютерный класс

9. Язык преподавания.

Русский

10. Преподаватели.

Главацкий С.Т., Бурыкин И.Г.

11. Авторы программы.

Главацкий С.Т., Бурыкин И.Г.