

Формализация информации и Биг Дата (Big Data)

02.03.03 -Математическое обеспечение и администрирование информационных систем,
направленность (профиль) -разработка и администрирование информационных систем

<http://vikchas.ru>

Вопросы 3-ей контрольной точки 24.11.2025 -
28.11.2025

Часовских Виктор Петрович
доктор технических наук, профессор кафедры
ШИиКМ, ФГБОУ ВО «Уральский государственный
экономический университет

Екатеринбург 2025

РАЗДЕЛ 1

Как реализуется подсистема ввода данных?

- с помощью технологии Data Mining;
- с помощью базы данных;
- с помощью СУБД;
- с помощью хранилища данных;
- с помощью витрины данных.

Какие данные могут храниться в системе анализа?

- детализированные;
- обобщенные;
- детализированные и обобщенные.

Могут ли в системе анализа храниться данные в разных форматах?

- могут;
- не могут;
- могут только данные в текстовых форматах.

Какие данные могут храниться в системе анализа?

- нормализованные;
- избыточные (денормализованные);
- частично нормализованные.

Как называются структуры данных, предназначенные для ввода, модификации и поиска?

- оперативные источники данных;
- хранилища данных;
- базы данных;
- файлы данных.

Как называются структуры данных, предназначенные для анализа?

- оперативные источники данных;
- хранилища данных;
- базы данных;
- витрины данных.

Предметно-ориентированный, интегрированный, неизменчивый, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений – это

- оперативный источник данных;
- хранилище данных;
- база данных;
- файл данных.

Какие данные отражают сущность события?

- измерения;
- метаданные;
- факты.

Какие данные описывают события?

измерения;
метаданные;
факты.

РАЗДЕЛ 2

Как называются числовые фактические данные, которые могут быть просуммированы по всем измерениям?

аддитивные;
полуаддитивные;
неаддитивные.

Как называются числовые фактические данные, которые могут быть просуммированы только по определенным измерениям?

аддитивные;
полуаддитивные;
неаддитивные.

Как называются фактические данные, которые не могут быть просуммированы ни по одному измерению?

аддитивные;
полуаддитивные;
неаддитивные.

На какие вопросы должны отвечать метаданные?

что, кто, где, как, когда, почему;
что, кто, зачем, как, когда, почему;
что, кто, где, как, по какой причине, почему;
что, кто, где, как, когда, сколько.

Репозиторий – это

словарь терминов;
хранилище метаданных;
каталог с файлами.

РАЗДЕЛ 3

Если поток образуется данными, копируемыми из ОИД, то он называется

входным потоком;
потоком обобщения;
архивным потоком;
потоком метаданных;
обратным потоком.

Если поток образуется агрегированными данными, то он называется

входным потоком;
потоком обобщения;
архивным потоком;
потоком метаданных;
обратным потоком.

Если поток образуется детальными данными, количество обращений к которым снизилось, то он называется

входным потоком;
потоком обобщения;
архивным потоком;
потоком метаданных;
обратным потоком.

Если поток образуется данными, переносимыми в репозиторий, то он называется

входным потоком;
потоком обобщения;
архивным потоком;
потоком метаданных;
обратным потоком.

Если поток образуется очищенными данными, записываемыми в ОИД, то он называется

входным потоком;
потоком обобщения;
архивным потоком;
потоком метаданных;
обратным потоком.

РАЗДЕЛ

Многомерный анализ – это:

одновременный анализ по нескольким измерениям;
одновременный анализ по нескольким параметрам;
одновременный анализ по нескольким данным.

Формирование подмножества многомерного массива данных, соответствующего единственному значению одного или нескольких элементов измерений, не входящих в это подмножество, называется

гиперкубом;
срезом гиперкуба;
базой данных;
витриной данных.

Технология оперативной аналитической обработки данных, использующая методы и средства для сбора, хранения и анализа многомерных данных в целях поддержки процессов принятия решений, называется

OLTP;
OLAP;
Data Mining.

Если для реализации многомерной модели используют многомерные базы данных, то способ реализации гиперкуба называется

MOLAP;
ROLAP;
HOLAP;
DOLAP;
JOLAP.

Если для реализации многомерной модели используют реляционные базы данных, то способ реализации гиперкуба называется

MOLAP;
ROLAP;
HOLAP;
DOLAP;
JOLAP.

Если для реализации многомерной модели используют и многомерные, и реляционные базы данных, то способ реализации гиперкуба называется

MOLAP;
ROLAP;
HOLAP;
DOLAP;
JOLAP.

Настольная OLAP, предназначенная для локального анализа и представления данных, называется

MOLAP;
ROLAP;
HOLAP;
DOLAP;
JOLAP.

OLAP, предназначенная для создания и управления данными и метаданными, называется

MOLAP;
ROLAP;
HOLAP;
DOLAP;
JOLAP.

В каком отношении находятся таблица фактов и таблица измерений?

«один-к-одному»;
«один-ко-многим»;
«многие-ко-многим».

Исследование и обнаружение машиной (алгоритмами, средствами искусственного интеллекта) в сырых данных скрытых знаний, которые: ранее не были известны, нетривиальны, практически полезны, доступны для интерпретации человеком, называется

OLTP;
хранилищем данных;

OLAP;
Data Mining.

Какие операции над данными включены в ETL-процесс?

ввод, модификация, вывод;
чтение, изменение, запись;
извлечение, преобразование, загрузка;
получение, хранение, анализ.

Многомерный просмотр данных основан на ...

многомерной базе данных;
технологии мультимедиа;
многослойной базе;
сетевой технологии;
гипертекстовой технологии.

Какую технологию используют большинство хранилища данных?

концептуальную БД;
реляционную БД;
иерархическую БД;
физическую БД.

Какое конструирование у хранилищ данных?

физическое;
логическое;
логическое и физическое;
иерархическое.

Концептуальную модель хранилища данных можно представить в виде...

таблицы;
графического рисунка;
геометрических фигур;
схемы.

РАЗДЕЛ 4

К классу описательных задач относятся:

кластеризация и классификация;
кластеризация и поиск ассоциативных правил;
классификация и регрессия;
классификация и поиск ассоциативных правил.

К классу предсказательных задач относятся:

кластеризация и классификация;
кластеризация и поиск ассоциативных правил;
классификация и регрессия;
классификация и поиск ассоциативных правил.

К классу задач supervised learning (обучение с учителем) относятся:

кластеризация и классификация;
кластеризация и поиск ассоциативных правил;
классификация и регрессия;
классификация и поиск ассоциативных правил.

К классу задач unsupervised learning (обучение без учителя) относятся:

кластеризация и классификация;
кластеризация и поиск ассоциативных правил;
классификация и регрессия;
классификация и поиск ассоциативных правил.

Задача классификации сводится к ...

нахождения частых зависимостей между объектами или событиями;
определения класса объекта по его характеристиками;
определение по известным характеристикам объекта значение некоторого его параметра;
поиска независимых групп и их характеристик в всем множестве анализируемых данных.

Задача регрессии сводится к ...

нахождения частых зависимостей между объектами или событиями;
определения класса объекта по его характеристиками;
определение по известным характеристикам объекта значение некоторого его параметра;
поиска независимых групп и их характеристик в всем множестве анализируемых данных.

Задача кластеризации заключается в ...

нахождения частых зависимостей между объектами или событиями;
определения класса объекта по его характеристиками;
определение по известным характеристикам объекта значение некоторого его параметра;
поиска независимых групп и их характеристик в всем множестве анализируемых данных.

Целью поиска ассоциативных правил является ...

нахождения частых зависимостей между объектами или событиями;
определения класса объекта по его характеристиками;
определение по известным характеристикам объекта значение некоторого его параметра;
поиска независимых групп и их характеристик в всем множестве анализируемых данных.

В описательных моделях относятся следующие модели данных:

модели классификации и последовательностей;
регрессивные, кластеризации, исключений, итоговые и ассоциации;
классификации, кластеризации, исключений, итоговые и ассоциации;
модели классификации, последовательностей и исключений.

Модели классификации описывают ...

правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

Модели последовательностей описывают ...

правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

Регрессивные модели описывают ...

правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

Какова основная цель создания и использования хранилищ данных:

анализ данных для принятия управленческих решений;
надежное хранение, накопленных данных;
резервное копирование данных.

Какие из вариантов ответов являются характерными требованиями к хранению данных для принятия решений в хранилищах данных?

данные ориентированы на приложения;
данные управляются транзакциями;
данные обобщены либо очищены.

Перечислите основные этапы работы с хранилищами данных?

этап очистки данных;
этап обновления;
этап нормализации.

Что называют кубом OLAP?

структуру, в которой хранятся совокупности данных, полученные путем всех возможных сочетаний измерений в таблице измерений;
структуру, в которой хранятся совокупности данных, полученные путем всех возможных сочетаний измерений в таблице фактов;
таблицу размерностей.

Информационные хранилища созданы для удобства ...

руководителей всех уровней для принятия решений;
стратегического планирования;
реорганизации бизнеса;
предметных приложений;
редактирования данных.

Информационные хранилища размещаются на ...

библиотеках-автоматах;

сетевых серверах;
мейнфреймах;
серверах и кластерах серверов;
файл-серверах.

При слиянии данных в информационное хранилище из внутренних и внешних источников обеспечивается ...

предметная ориентация данных;
выбор требуемых сведений из предметных приложений по наименованиям;
гипертекстовый просмотр данных;
согласование данных по наименованию;
хранение данных по предметным областям.

Интеллектуальный выбор данных из информационного хранилища – это ...

реализация методов искусственного интеллекта;
выбор по заданному алгоритму;
реализация самообучающихся систем;
реализация экономико-статистических методов.

РАЗДЕЛ 5

В процессе погружения в информационное хранилище данные ...

очищаются от ненужной для анализа информации;
агрегируются;
преобразуются из разных типов данных предметных приложений в единую структуру хранения;
индексируются;
синхронизируются.

Какова основная цель создания и использования хранилищ данных:

анализ данных для принятия управленческих решений;
надежное хранение накопленных данных;
резервное копирование данных.

Классификация — ...

некоторый набор операций над базой данных, который рассматривается как единственное завершённое, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных;
разновидность систем хранения, ориентированная на поддержку процесса анализа данных, обеспечивающая непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов;
высокоуровневые средства отражения информационной модели и описания структуры данных;
это установление зависимости дискретной выходной переменной от входных переменных.

Регрессия — ...

это установление зависимости непрерывной выходной переменной от входных переменных;
это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;

выявление закономерностей между связанными событиями;
это установление зависимости дискретной выходной переменной от входных переменных.

Кластеризация — ...

это установление зависимости непрерывной выходной переменной от входных переменных;
это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;
выявление закономерностей между связанными событиями;
это установление зависимости дискретной выходной переменной от входных переменных.

Ассоциация — ...

это установление зависимости непрерывной выходной переменной от входных переменных;
это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;
выявление закономерностей между связанными событиями;
это установление зависимости дискретной выходной переменной от входных переменных.

Машинное обучение — ...

специализированное программное решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных;
это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;
набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной параметр, которому отвечает правильный выходной результат;
подразделение искусственного интеллекта, изучающее методы построения алгоритмов, способных обучаться на данных.

Аналитическая платформа — ...

специализированное программное решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных;
это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;
набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной параметр, которому отвечает правильный выходной результат;
подразделение искусственного интеллекта, изучающее методы построения алгоритмов, способных обучаться на данных.

Обучающая выборка — ...

это группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;
набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной параметр и соответствующий ему правильный выходной результат;
набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной параметр, которому отвечает правильный выходной результат;
выявление в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Ошибка обучения — ...

это ошибка, допущенная моделью на учебном множестве;

это ошибка, полученная на тестовых примерах, то есть вычисляется по тем же формулам, но для тестового множества;

имена, типы, метки и назначения полей исходной выборки данных;

набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной параметр, которому отвечает правильный выходной результат.

Ошибка обобщения — ...

это ошибка, допущенная моделью на учебном множестве;

это ошибка, полученная на тестовых примерах, то есть вычисляется по тем же формулам, но для тестового множества;

имена, типы, метки и назначения полей исходной выборки данных;

набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной параметр, которому отвечает правильный выходной результат.

ЗАДАНИЕ. Выбрать из каждого раздела номер 1,3,4,5 по 3 вопроса, из 2 раздела 1 вопрос. Указать правильное значение и подробную интерпретацию.