

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ ХРАНИЛИЩА ДАННЫХ

Автор: Стефановский Дмитрий Владимирович

Код и наименование направления подготовки, профиля: 38.04.01 Экономика («Системы больших данных в экономике»)

Квалификация (степень) выпускника: Магистр

Форма обучения: очная

Цель освоения дисциплины:

Сформировать компетенции в сфере аналитической работы в области экономики и финансов, обработки и анализа данных, а также научно-исследовательской работы

План курса:

Тема 1. Введение и общие понятия.

Хранилища данных. Концепция систем складирования данных. Концепция хранилищ данных. Типы хранилищ данных. Архитектура хранилищ данных. Основные типы программно-аппаратной архитектуры хранилища данных. Подходы в организации работ по созданию хранилища данных. Характеристика решений ведущих производителей. Типовые программно-аппаратные решения (технологические решения). Корпоративные информационные фабрики. Хранилища данных с архитектурой шины данных Объединенное (федеративное) ХД

Тема 2. Модель типового проекта создания хранилища данных.

Жизненный цикл разработки хранилища данных. Модель жизненного цикла хранилища данных. Бизнес-модель типового проекта создания хранилища данных. Системы деловой осведомленности (Business Intelligence Systems) и хранилища данных. Определение систем деловой осведомленности. Информационная безопасность систем бизнес-аналитики. Технологии обработки данных. Архитектура систем бизнес-уровня. Построение систем бизнес-аналитики: проблемы и решения. Хранилища данных и системы бизнес-аналитики.

Тема 3. Метод моделирования «сущность-связь».

Понятие предметной области и архитектура данных. Моделирование методом «сущность-связь». Моделирование темпоральных (временных) данных в хранилищах данных. Темпоральные данные и базы данных. Введение в моделирование темпоральных данных. Основные приемы моделирования темпоральных данных. Метод многомерного моделирования. Основные понятия метода многомерного моделирования. Многомерная модель. Факты. Ключи в таблицах фактов. Таблицы фактов. Измерения. Основные схемы многомерной модели. Моделирование таблиц фактов. Моделирование таблиц измерений.

Тема 4. Физическая модель хранилища данных.

Создание физической модели хранилища данных. Объекты физической модели данных. Моделирование объектов физической модели хранилища данных. Разработка скрипта для создания объектов физической модели хранилища данных. Метаданные в хранилищах данных. Метаданные. Функции метаданных в хранилище данных. Состав

метаданных в хранилище данных. Логическая структура метаданных хранилища данных. Стандарты метаданных. Выбор метамодели при проектировании хранилища данных. Проектирование логической модели метаданных хранилища данных.

Тема 5. Проектирование и разработка процесса ETL.

Введение. Подходы к реализации ETL-процесса. Разработка ETL-процесса. Элементы ETL-процесса. Проектирование процессов преобразования данных с помощью CASE-инструментов. Создание модели хранилища данных на основе корпоративной модели данных. Корпоративная модель данных. Алгоритм преобразования корпоративной модели данных в модель хранилища данных. Использование CASE-инструментов для создания многомерной модели на основе корпоративной модели данных. Секционирование таблиц многомерной модели данных в PowerDesigner.

Тема 6. Метод моделирования «Свод данных».

Метод моделирования «Свод данных» (Data Vault). Пример проектирования модели «Свод данных». Заполнение данными объектов «Свода данных». Физическая модель хранилища данных: учет влияния транзакций, денормализация таблиц. Определение транзакций базы данных. Денормализация. Методы разбиения таблиц. Денормализация колонок.

Тема 7. Проектирование производительности.

Повышение производительности запросов: индексы. Повышение производительности запросов: секционирование. Повышение производительности запросов: кластеры. SQL в хранилищах данных: агрегация и суммирование. Введение. Оператор SELECT и схема «звезда». Расширение оператора SELECT для обработки данных. Функция GROUPING. Управление иерархией в предложениях ROLLUP и CUBE.

Тема 8. SQL в хранилищах данных.

SQL в хранилищах данных: аналитическая обработка данных. SQL для анализа данных. Агрегатные и статистические функции. Медианы. Оконные функции. Функции ранжирования. Функции, генерирующие отчеты. CASE-выражения и создание гистограмм. Статистические функции в других диалектах SQL. Проектирование кубов данных. OLAP на клиенте и на сервере. Технические аспекты многомерного хранения данных. Основные понятия OLAP. Проектирование кубов данных с использованием CASE-инструментов. Генерация куба данных.

Аудиторные часы: 144

Формы текущего контроля и промежуточной аттестации: опросы, контрольная работа, зачет

Основная литература:

1. Туманов В.Е. Проектирование хранилищ данных для систем бизнес-аналитики: учебное пособие / В.Е. Туманов. — М.: Интернет-Университет Информационных Технологий: БИНОМ. Лаборатория знаний, 2010. — 615 с: ил., табл. — (Основы информационных технологий).