

Проектирование хранилищ данных, 24

ФИО преподавателя: Смирнов Михаил Вячеславович

e-mail: smirnovmgupi@gmail.com

Лекция 1

Проектирование хранилищ данных. Вводная лекция.

Составитель курса

- ▶ Смирнов Михаил Вячеславович, к.э.н.
- ▶ Научные интересы и предметная область: математическая статистика, машинное обучение, технологии хранения данных, проектирование баз данных, администрирование баз данных
- ▶ Контакты:
 - ▶ Электронная почта: smirnovmgupi@gmail.com
 - ▶ Образовательный сайт: <http://msuniversity.ru>
 - ▶ Образовательный канал в YT: [тык сюда!](#)
 - ▶ Группа Telegram:



Вопросы лекции

- базовые требования к слушателям курса
- план лекционных и практических занятий
- программное и методическое обеспечение курса
- предложения по научной и методической работе студентов в ходе курса
- практические результаты курса и их дальнейшее применение

Базовые требования к слушателям курса

- знание основ реляционной модели хранения данных (ACID, IC, нормализация);
- понимание реляционных типов данных и ограничений;
- базовый уровень языка SQL (DDL, DML);
- навыки работы с оболочками СУБД (создание и исполнение запросов, прямой и обратный инжиниринг реляционных моделей данных).

Что почитать, если есть проблемы?

- Крёнке Д. - Теория и практика построения баз данных, 9 изд. русский
- Крёнке Д. - Теория и практика построения баз данных, 15 изд. Английский
- Молина, Ульман, Уидом – Системы баз данных. Полный курс.
- Смирнов М.В. – Управление данными, учебно-методическое пособие, РТУ-МИРЭА, 2020.
- Смирнов М.В., Смирнов С.В. – Проектирование баз данных. Практикум. РТУ-МИРЭА, 2020.

Темы лекционных занятий

- Тема 1. Определение и компоненты хранилища данных (особенности, витрины, архитектурные типы)
- Тема 2. Аспекты реализации проекта хранилища данных
- Тема 3. Архитектурные решения классических и big data хранилищ данных.
- Тема 4. Процедуры ETL (extract, transform, load).
- Тема 5. Описание актуальных технологий хранилищ данных для BigData.
- Тема 6. Документная модель MongoDB.
- Тема 7. Шардированные кластеры документного хранилища MongoDB.

Темы практических занятий

- Тема 1. Рабочая область ms sql server, установка учебного хранилища данных AdventureWorksDW. Изучение схемы хранилища данных на примере учебной базы.
- Тема 2. Физический словарь хранилища данных на примере учебной базы.
- Тема 3. Практика построения хранилища данных типа «звезда».
- Тема 4. Импорт и экспорт данных средствами ms sql server.
- Тема 5. Проектирование потоков управления SSIS.
- Тема 6. Проектирование потоков данных SSIS.
- Тема 7. Рабочая область mongoDB. Импорт данных в документы и коллекции.
- Тема 8. Настройка шардирования в mongoDB.

Программное обеспечение, используемое в курсе (подразумевается демонстрация)

- Oracle VirtualBox (виртуальная машина ОС Linux с предустановленными MS SQL Server Developer Edition и учебной базой данных AdventureWorksDW);
- MS SQL Server Express и файлы учебных баз данных AdventureWorks и AdventureWorksDW;
- MS Visual Studio 2019 с пакетом SSIS (SQL Server Integration Services);
- Оболочка управления СУБД DBeaver (обратный инжиниринг).
- Оболочка управления Compass (mongoDB)

Методическое обеспечение, используемое в курсе

- Сарка, Лах, Йеркич – Microsoft SQL Server 2012. Реализация хранилищ данных;
- Д. Петкович – Microsoft SQL Server 2012. Руководство для начинающих;
- P. Ponniah – Data warehousing. Fundamentals for IT Professionals;
- Мартишин, Симонов, Храпченко – Базы данных. Практическое применение СУБД SQL и noSQL типа для проектирования ИС.

Предложения по научной и методической работе.

- Образцы оформления отчетной документации по практическим работам в ходе курса. Результат – комплект эскизов технических документов.
- Рабочая тетрадь по практическим заданиям дисциплины «Проектирование хранилищ данных».
- Научные статьи по тематике «Проектирование хранилища данных в определенной предметной области». В первую очередь для докладов на конференциях.

Практические результаты курса и их дальнейшее применение

- Результат курса – физическая модель хранилища данных типа «Звезда» и «Снежинка», построенная учащимся в ПО MS SQL Server (или по согласованию в другом доступном ПО) в рамках определенной им (темой диссертационного исследования) предметной области, комплект документации по разработанной им модели (физический словарь данных, описание представлений и витрин данных, другое).
- Применение результатов курса – одно из значимых предложений для магистерской диссертации (пункт 3, предложения по разработке или модернизации для выбранной организации или предметной области), научные доклады по разработанной модели.

Спасибо за внимание!