

1. Введение в базы данных

к.т.н., доцент кафедры ИиСП

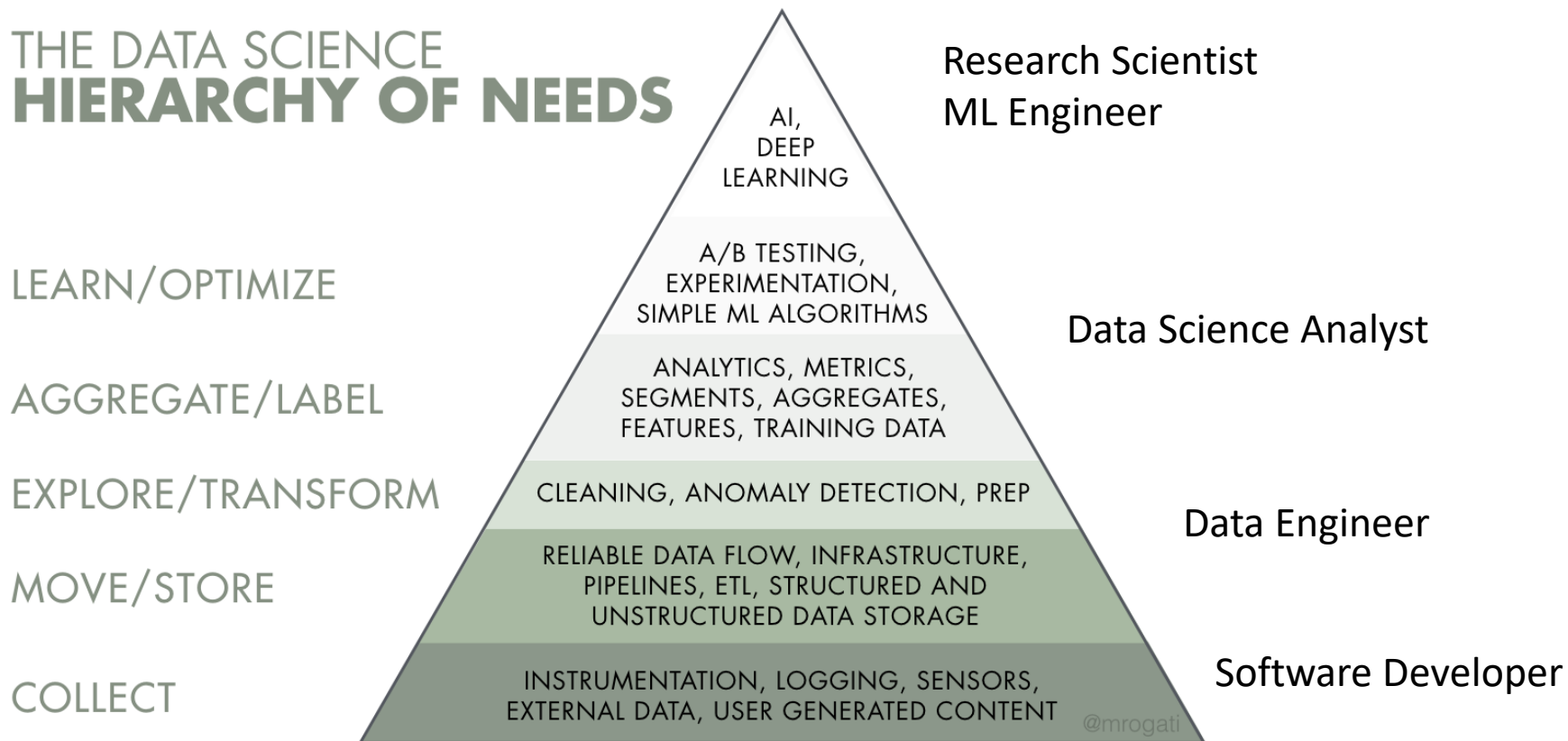
Лучинин
Захар Сергеевич

План курса

- Проектирование баз данных
- Реляционные СУБД
- SQL. Оптимизация запросов
- Масштабирование базы данных
- Транзакции
- ORM
- NoSQL базы данных

Иерархия потребностей

THE DATA SCIENCE **HIERARCHY OF NEEDS**



[Source](#): Monica Rogati's fantastic Medium post "The AI Hierarchy of Needs"

Почему важно изучать БД?

- Практически каждое приложение использует БД
- Почти каждая вакансия с упоминанием SQL
- Базы данных в петабайты. Анализ данных откроет новые знания
- Отсутствие готовых решений. Вызов для инженеров

Понятие базы данных

- **Данные** – поддающееся многократной интерпретации представление информации в формализованном виде, пригодном для передачи, связи, или обработки.

[ISO/IEC 2382:2015 Information technology — Vocabulary]

- **База данных** – совокупность данных, хранимых в соответствии со схемой данных, манипулирование которыми выполняют в соответствии с правилами средств моделирования данных.

[ISO/IEC TR 10032:2003 Information technology – Reference model of data management]

- **База данных** – совокупность данных, организованных в соответствии с концептуальной структурой, описывающей характеристики этих данных и взаимоотношения между ними, причём такое собрание данных, которое поддерживает одну или более областей применения.

[ISO/IEC 2382:2015 Information technology — Vocabulary]

Понятие СУБД

- Система управления базами данных (СУБД) – совокупность программных и лингвистических средств общего или специального назначения, обеспечивающих управление созданием и использованием баз данных.

[ISO/IEC TR 10032:2003 Information technology – Reference model of data management]

Какие задачи решает СУБД?

- Надежное хранилище больших данных
- Эффективный доступ
- Многопользовательский доступ
- Разграничение прав доступа
- Удобная работа с данными (язык программирования для доступа к данным)
- Обеспечение доступа (по сети)

Классификация БД

- По модели данных
 - Иерархическая
 - Реляционная
 - Документо-ориентированная
- По среде постоянного хранения
 - В оперативной памяти
 - Во вторичной памяти (SSD, жёсткий диск)
- По степени распределённости
 - Централизованная
 - Распределённая БД

Классификация СУБД

- По модели данных
 - Иерархическая
 - Реляционная
 - Документо-ориентированная
- По степени распределённости
 - Локальные СУБД
 - Распределённые СУБД
- По способу доступа к БД
 - Файл-серверные
 - Клиент-серверные
 - Встраиваемые

Проектирование баз данных

- Проектирование базы данных — создание **эффективной структуры данных**, обеспечивающее хранение требуемой информации

Основные задачи проектирования

- Обеспечение хранения в БД всей необходимой информации
- Обеспечение возможности получения данных по всем необходимым запросам
- Сокращение избыточности и дублирования данных
- Обеспечение целостности данных (правильности их содержания): исключение противоречий в содержании данных, исключение их потери и т.д.

Критерии оценки схемы данных

Критерий	Описание
Структурная достоверность	Соответствие способу определения и организации информации
Простота	Удобство изучения модели как профессионалами в области разработки информационных систем, так и обычными пользователями
Выразительность	Способность представлять различия между данными, связи между данными и ограничения
Отсутствие избыточности	Исключение излишней информации, т.е. любая часть данных должна быть представлена только один раз

Критерии оценки схемы данных

Критерий	Описание
Способность к совместному использованию	Отсутствие принадлежности к какому-то особому приложению или технологии и, следовательно, возможность использования модели во многих приложениях и технологиях
Расширяемость	Способность развиваться и включать новые требования с минимальным воздействием на работу уже существующих приложений
Целостность	Согласованность со способом использования и управления информацией внутри предприятия
Схематическое представление	Возможность представления модели с помощью наглядных схематических обозначений

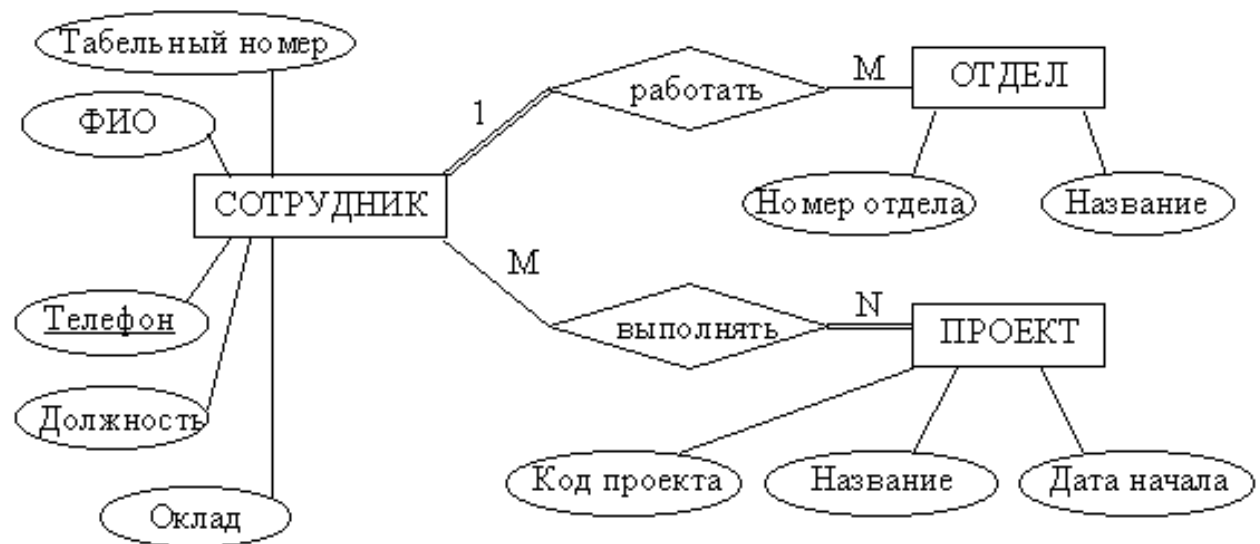
Уровни проектирования

1. **Концептуальное
(инфологическое) проектирование**
2. Логическое (дatalogическое)
проектирование
3. Физическое проектирование

Концептуальное проектирование

Концептуальное (инфологическое) проектирование — построение семантической модели предметной области, то есть информационной модели наиболее высокого уровня абстракции

- сущности
- атрибуты
- СВЯЗИ



Процедуры концептуального проектирования

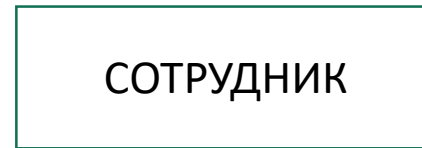
- определение сущностей и их документирование
- создание семантической модели предметной области
- определение связей между сущностями и их документирование
- обсуждение модели данных

Особенности концептуального проектирования

- Обследование предметной области, изучение ее информационной структуры
- Выявление всех фрагментов, каждый из которых характеризуется пользовательским представлением, информационными объектами и связями между ними, процессами над информационными объектами
- Моделирование и интеграция всех представлений

Составляющие концептуальной модели

- Сущность

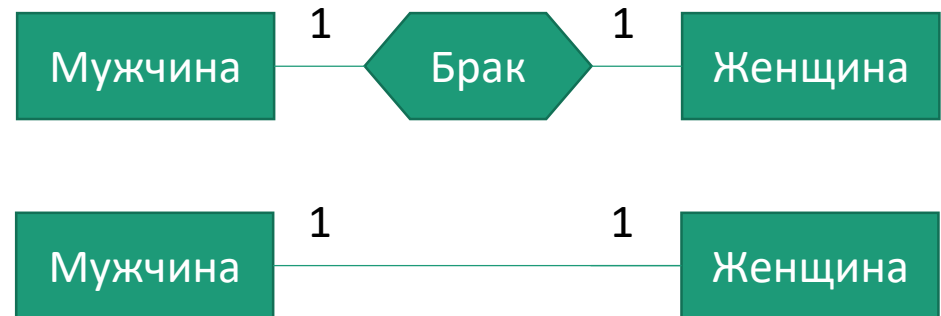


- Атрибут



- Связь

- ОДИН-К-ОДНОМУ
- ОДИН-КО-МНОГИМ
- МНОГИЕ-КО-МНОГИМ

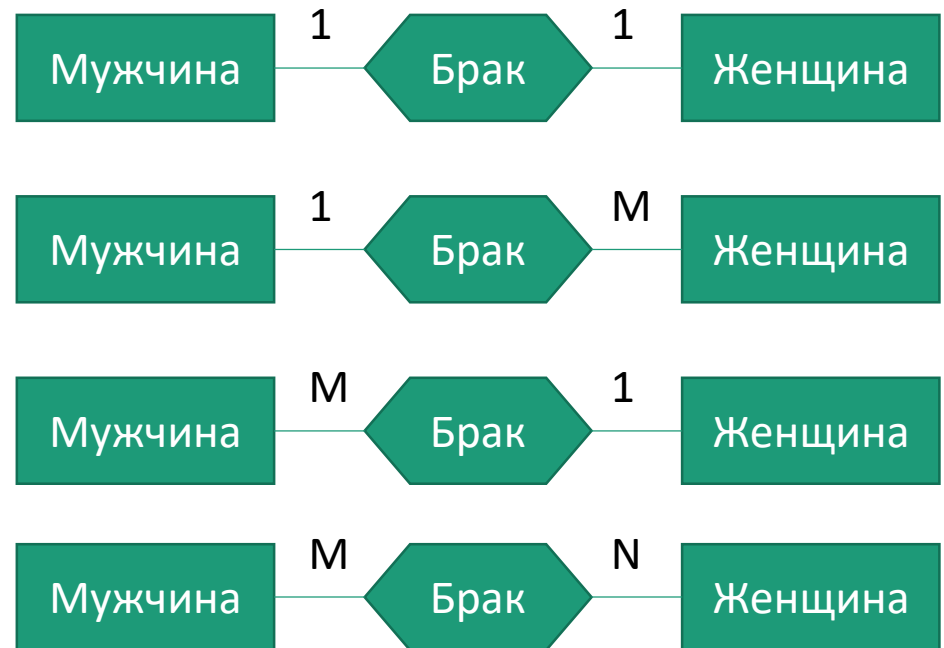


Выделение сущностей

- Работа эксперта предметной области, Product Manager'а, системного аналитика
- Выяснение потребностей и представления предметной области для каждой роли пользователя
- Выделяются независимые типы сущностей
- Отбрасываются “ненужные” сущности
- Именованение сущностей

Определение связей

- Типы связей:
 - Один ко многим
 - Многие к одному
 - Многие ко многим
 - Один к одному
- Кратность связей
 - 1
 - 0..1
 - 0..n
 - 1..n



Определение атрибутов

- Выделяются атрибуты типов сущностей
- Выделяются атрибуты (описание) связей
- Простые и составные атрибуты
- Именованние атрибутов

Проверка

- Наличие связей один к одному (сведение к одному)
- Наличие избыточных связей (нет прямой связи между сущностями)
- Проверка выполнимости пользовательских операций - мысленный эксперимент
- Обсуждение с пользователями

Уровни проектирования

1. Концептуальное (инфологическое) проектирование
- 2. Логическое (дatalogическое) проектирование**
3. Физическое проектирование

Логическое проектирование БД

- Преобразование **концептуальной модели** на основе выбранной модели данных в **логическую модель**
- Представление программиста/инженера данных
- Учитывается специфика конкретной модели данных, но не учитывается специфика конкретной СУБД

Процедуры логического проектирования

1. Выбор модели данных
2. Определение набора таблиц
3. Нормализация таблиц
4. Проверка логической модели данных на предмет возможности выполнения всех транзакций, предусмотренных пользователями
5. Определение требований поддержки целостности данных
6. Создание окончательного варианта логической модели данных и обсуждение его с пользователями

Литература

- Дейт К.Дж. Введение в системы баз данных
- Документация к СУБД
- Статьи на информационных ресурсах