当代数据管理系统

目的：使用数据管理系统来构建应用（APP）的能力 –> 要存什么数据、怎么存、怎么访问？

展开：通过数据管理系统的接口来使用它（CRUD\SQL\Transaction）、数据库设计（概念设计、结构设计、物理设计）、数据管理系统的工作原理（存储结构、查询处理、数据正确性的实现方法\事务处理）

具体的知识要点：

1. 什么是系统？（在一个应用程序里面能够独立运行的模块、模块化（information hiding））什么是数据管理系统？

2. 文档数据库（mongoDB）

（1）文档数据库的接口（CRUD）；

（2）文档数据库的设计：面向对象的设计方式（概念设计和结构设计）、物理设计（怎么使用索引）；

（3）文档数据的存储功能和索引（实现原理之一）：分页、索引结构（B+树）；

（4）CRUD操作的原子性（实现原理之二）：面向故障（日志机制），面向并发（并发控制、两阶段锁）；

（5）文档数据库的高可用机制（实现原理之三）：可用性（Availability）、一致性协议（Raft）；

3. 关系数据库（SQL数据库）

（1）关系数据库的接口：关系模型、关系代数（选择、投影、连接、聚集）、SQL（单表查询、连接查询、嵌套查询、聚集查询）、声明式查询语言的概念（好的模块化的方式）、关系数据库DDL中的约束（外键、唯一键等）、事务处理的接口；

（2）关系数据库的设计：ER图（概念设计），从ER图到关系模式（结构设计）、索引的使用、约束的使用、事务设计（如何设计短小、高效的事务）；

（3）关系的存储（实现原理之一）；

（4）SQL查询的执行方法（实现原理之二）：查询优化、查询的执行、每一个关系代数的操作的算法（投影的算法（多种）、连接的算法：散列连接、排序合并、嵌套循环、索引连接）、以I/O复杂度为中心的算法设计（数据访问的局部性）；

（5）事务处理的实现方法（实现原理之三）：对操作原子性的扩展、ACID的概念；

（6）数据备份和恢复（backup）：备份功能与日志和高可用功能的区别。

4. 文档数据库和关系数据库优缺点的对比（从概念上有所理解）。

5. 分布式数据库（面向云计算平台的数据管理系统的实现原理）

（1）云计算平台的基本概念：弹性（计算、存储）；

（2）应用（APP）的扩展：web server的扩展（复制）、DB server的扩展（数据分片）；

（3）数据管理系统的两种扩展方式：分库分表（应用层的扩展方式）、分布式数据库

（4）分布式数据库的架构：shared-everything, shared-disk(storage), shared-nothing(重点)；

（5）数据分片的方式以及各自的优缺点；

（6）如何在分片的数据上进行简单的SQL查询：算子下推（将在整个数据上进行的计算操作分解成在局部分片上进行的操作）、半连接（semi-join）的方法；

（7）如何进行分布式的事务处理：把一个事务分成多个子事务，交给不同节点去运行，但需要特殊机制保证整个事务的原子性 🡪 多个子事务要么都提交，要么都撤销 🡪 两阶段提交协议（阻塞的问题）；

（8）高可用的机制：可用性（Availability）、一致性协议（Raft）、CAP原理。