

第七章作业

姓名：袁昊男 学号：2018091618008

一、单选题

1. 以下哪一项工作通常不是数据库系统面临的挑战？（D）
A. 数据库高并发读写需求 B. 海量数据的高效存储和处理
C. 数据库高扩展性 D. 编写数据库应用程序
2. 以下哪一项不是 NoSQL 的共同特征？（D）
A. 分区 B. 异步复制 C. BASE D. CAP
3. HBase 是哪一种存储模型的 NoSQL 数据库？（A）
A. 列存储 B. 文档存储 C. 键值对存储 D. 图形存储
4. MongoDB 是哪一种存储模型的 NoSQL 数据库？（B）
A. 列存储 B. 文档存储 C. 键值对存储 D. 图形存储
5. Neo4j 是哪一种存储模型的 NoSQL 数据库？（D）
A. 列存储 B. 文档存储 C. 键值对存储 D. 图形存储

二、判断题

1. CAP 是在分布式的环境下设计和部署系统时的 3 个核心的需求。（√）
2. Redis 复制主要包括 RDB 复制和 AOF 复制。（√）
3. HBase 中的区域和表的关系是随着数据的增加而动态变化的。（√）
4. MongoDB 的分片是将一个集合的数据分别存储在不同的结点上减轻单机压力。（√）
5. Neo4j 的数据物理存储主要分为结点、关系、结点或关系上的属性这 3 类数据存储。（X）

三、填空题

1. 事务 ACID 特性包括原子性、一致性、隔离性和（持久性）。
2. 在数据库中，BASE 模型包含 BA（Basically Available），基本可用；S（Soft State），软状态/柔性事务；（最终一致性）。
3. Neo4j 的 CQL 的全称是（Cypher Query Language）。
4. Redis 安装完成后，默认数据库有（16）个。
5. MongoDB 中基本的概念是文档、（集合）、数据库。

四、简答题

1. NoSQL 的特征是什么？

答：NoSQL 没有明确的范围和定义，普遍存在的共同特征如下。

- (1) 不用预定义模式：不需要事先预定义表结构；数据中的每条记录都可能有不同的属性和格式，插入数据时不需要预先定义它们的模式。
- (2) 无共享架构：与将所有数据存储在网络中的全共享架构不同，NoSQL 将数据划分后存储在各个本地服务器上。
- (3) 弹性可扩展：在系统运行的时候，动态增加或者删除结点，不需要停机维护，数据

可以自动迁移。

- (4) 分区: NoSQL 数据库将数据进行分区, 将记录分散在多个结点上, 分区的同时还要做复制。这样既提高了并行性能, 又保证没有单点失效的问题。
- (5) 异步复制: NoSQL 中采用基于日志的异步复制, 数据被尽快地写入一个结点, 缺点是并不总能保证一致性, 在出现故障时, 可能会丢失少量的数据。
- (6) BASE: NoSQL 数据库保证的是 BASE 特性, 保证事务的最终一致性和软事务。

2. Redis 支持的数据类型有哪些? 在这些类型上有哪些操作?

答: Redis 数据库中的所有数据都是键值对, 不管放入数据库中的是什么类型的数据, 放在底层的都以二进制字节数组的格式存放, 客户端使用时需要自己来转换。Redis 键值是二进制安全的, 用任何二进制序列作为键, 空字符串也是有效键。键取值太长将导致查找键的计算成本高取值, 而太短时, 可读性较差。每个数据库的大小不能超过内存, 它完全在内存中操作数据, 数据类型丰富。Redis 支持的值类型有 5 种, 即字符串类型 (string)、哈希表类型 (hash)、链表类型 (list)、集合类型 (set)、有序集合类型 (ordered set)。

(1) 字符串

字符串是最基本的类型, 它是二进制安全的, 可以包含任何数据, 如 JPG 图片或者序列化的对象, 其字符串值最多是 512MB。字符串是最常用的一种数据类型, 可应用于普通的键值对存储, 具有定时持久化、操作日志及复制等功能。

(2) 哈希表

哈希是一个键值对集合, 一个字符串类型的域 (field) 和值 (value) 的映射表, 适用于存储对象。例如, 存储用户信息对象数据用户 ID 为键, 存储的值是用户对象, 包含姓名、年龄、生日、专业等信息, 用哈希表内部存储的值为一个哈希映射 (HashMap)。

(3) 链表

链表是简单的字符串列表, 是重要的数据结构之一。它的实现为一个双向链表, 支持反向查找和遍历, 可用于发送缓冲队列等功能。

(4) 集合

Redis 中的集合是一个无序的、去重的元素集合, 元素是字符串类型, 最多包含 $2^{32}-1$ 个元素。集合是通过哈希表实现的, 所以添加、删除、查找的复杂度都是 $O(1)$ 。Redis 的集合对外提供与链表类似的功能, 集合的概念就是一堆不重复值的组合。集合的内部实现是一个值永远为 null 的哈希映射, 是通过计算哈希的方式来快速去重的。

(5) 有序集合

有序集合的操作类似集合, 有序的、去重的、元素是字符串类型、不允许有重复的成员, 每一个元素都关联着一个浮点数分值 (Score), 按照分值从小到大的顺序排列集合中的元素。分值可以相同, 最多包含 $2^{32}-1$ 个元素, 成员是唯一的但分数 (Score) 可以重复。有序集合的使用场景与集合类似, 区别是集合不是自动有序的, 有序集合通过用户额外提供一个优先级 (Score) 的参数来为成员排序, 是插入有序的即自动排序。有序集合的内部使用哈希映射和跳跃表 (SkipList) 来保证数

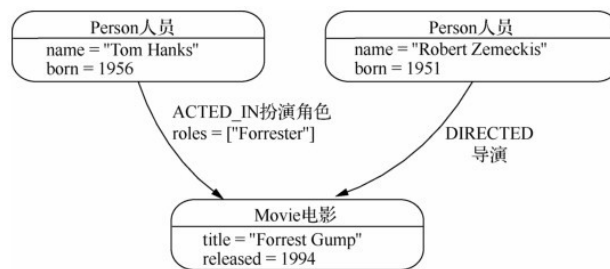
据的存储和有序。

3. MongoDB 的存储架构是什么？

答：MongoDB 分布式集群能够对数据进行备份，提高数据安全性，以及提高集群读/写服务的能力和数据存储能力。系统主要通过副本集(Replica)对数据进行备份，通过分片(Sharding)对大的数据进行分割，分布存储在不同结点上。MongoDB 目前支持 3 种集群模式：主从集群，副本集集群，分片集群。

4. Neo4j 的数据模型是什么？请举例说明结点、关系、属性、标签的含义。

答：Neo4j 图形数据库是基于属性图模型来描述数据的，用属性图模型表示结点、关系和属性中的数据；结点和关系都包含属性；关系连接结点；属性是键值对；结点用圆圈表示，关系用方向键表示；关系具有方向——单向和双向；每个关系包含“开始结点”(或“从结点”)和“到结点”(或“结束结点”)；在属性图数据模型中，关系是定向的。Neo4j 图形数据库将其所有数据存储存储在结点和关系中。



如上图，存在三个结点（“Tom Hanks”、“Robert Zemeckis”、“Forrest Gump”）和两个关系（“ACTED_IN”、“DIRECTED”）。Person 和 Movie 是标签，name、title、roles 等是结点和关系的属性。实体包括结点和关系，结点有标签和属性，关系是有向的，连接两个结点，具有属性和关系类型。