

4-5-cn

1

我们学习了数据存储系统，在这节课我们讨论了四种主要类型的 NoSQL 数据库。

2

大数据计算系统可以概括为三个部分：数据存储系统、数据处理系统、数据应用系统这里我们仍然在数据存储系统的范围内。

在数据存储系统中，有 4 个部分完成不同的任务，它们是数据收集和建模、分布式文件系统、分布式数据库/数据仓库和统一数据访问接口。

这里我们还是在 NoSQL 数据库中。主流的 No SQL 数据库有 4 种类型，下面我们看下这四种类型的数据库

3

NoSQL 数据库主要分为四种类型:键值对、面向列的、基于图的和面向文档的。

每种类别都有其独特的属性和局限性。上述指定的数据库都不能更好地解决所有问题。

用户应该根据其产品需求选择相应的数据库。

4

键值数据库基于键值对(键、值)来存储和查询数据。

数据以键/值对的形式存储。它被设计成处理大量数据和高负载的方式。

其基本思想是，数据值由键查询，键可以是字符串类型，值可以是任何类型的数据，如整数、字符、数组、列表、集合、JSON、BLOB(二进制大对象)等。

底层结构可以使用哈希表对键进行索引和管理，以支持快速查询。

但是键值数据库不支持基于数据值的查询

它是最基本的 NoSQL 数据库示例之一。

这种 NoSQL 数据库被用作集合、字典、关联数组等。键值存储帮助开发人员存储无模式数据。它们最适合用于购物车中的内容。

5

面向列的数据库在列上工作，并基于谷歌编写的 BigTable 文件。

列存储数据库采用列存储结构，也称为 DSM(分解存储模型)，这意味着根据范围(列)检索、存储和管理数据。

列存储仍然使用键，但键指向列。列存储数据库支持高压缩比。

对大规模数据下数据字段的查询效率很高，但不适合实时删除或更新整个记录，也不支持数据表 join 操作。

每一列都单独处理。单列数据库的值是连续存储的。

基于列的 NoSQL 数据库提供高性能的聚合查询，如 SUM, COUNT, AVG, MIN 等，因为数据在列中随时可用。

基于列的 NoSQL 数据库被广泛用于管理数据仓库、商业智能、客户关系管理、图书馆卡片目录。

6

面向文档的 NoSQL DB 以键值对的形式存储和检索数据，但值部分以文档的形式存储。

文档以 JSON 或 XML 格式存储。该值被 DB 理解，可查询。

文档数据库是围绕一系列语义上自包含的文档组织起来的，以组织数据管理。

文档没有模式，这意味着文档不需要特定的结构。这使得它具有灵活性。

文档数据库实际上是文档的集合。文档实际上是一种数据记录。

该记录可以“自我描述”所包含数据的类型和内容。

XML 文档、HTML 文档和 JSON 文档都属于这一类。

每个文档中包含的数据记录是一系列数据项的集合。

每个数据项都有一个名称和相应的值。值可以是简单的数据类型，如字符串、数字和日期，也可以是复杂的类型，如有序列表和关联对象。

文档数据库特别适合管理面向文档的数据或类似的半结构化数据，例如后台有大量读写操作的网站，使用 JSON 数据结构或使用嵌套结构化非标准化数据的应用程序

关系 V.S 文档

在左边的图表中，您可以看到我们有行和列，在右边，我们有一个文档数据库，它的结构与 JSON 类似。现在对于关系数据库，您必须知道有哪些列等等。

但是，对于文档数据库，您有 JSON 对象这样的数据存储。

文档类型主要用于 CMS 内容管理系统、博客平台、实时分析和电子商务应用。

它不应该用于需要对不同聚合结构进行多个操作或查询的复杂事务

7

图形数据库

每个图将社会关系和其他数据描述为点(顶点)和边(边缘)及其属性，可以看作是一个结构化数据。

这里的点表示实体，如个人、公司、帐户或任何其他数据项，类似于关系数据库中的数据记录或文档数据库中的文件;

“边”表示点之间的关系;它是用户关注的点相关特征。

关系数据库表示数据范围内不同数据之间的关联关系。而在图形数据库中，这种关系直接定义并存储在数据库中，

遍历关系非常快，因为它们已经被捕获到 DB 中，不需要计算它们。

因此，图数据库善于处理高度相关的数据，适用于社交网络、模式识别、依赖分析、推荐系统和寻路等可以用关系图表示的问题。

8

文档数据库包括 mongoDB、CouchDB 等。

图形数据库包括 Neo4j、infiniteGraph 等。

键值数据库包括 Redis、DynamoDB 等。

列数据库包括 Hbase Cassandra 等。

9

今天的内容就学习到这里，谢谢大家