

大数据管理方法与应用实验报告

学院名称：	数字经济与管理学院
专业名称：	大数据管理与应用
班 级：	数据管理 2302
作 者：	林子晗
学 号：	2023113223

一、 概述

本实验报告总结了基于三国人物关系图谱和监控系统的设计与实现过程。实验主要包含以下几个部分的内容与结果：

1. 三国人物关系图谱的构建

通过使用 Neo4j，并且基于三国历史人物与战役的关系，成功构建了一个包含多个节点和关系的图数据库。节点包括人物、战役、地理位置，关系涵盖了人物之间参与战役、人物关系等多种类型。

2. 监控系统的设计与实现

通过 Grafana 和 InfluxDB 搭建了一个监控系统，主要用于实时监控系统运行情况。通过采集和展示关键指标，监控系统帮助实时掌握系统性能和资源利用率。

二、 数据管理方法综述

数据库 类型	适用场景	数据 类型	数据关系	优势特点
MariaDB	关系型数据库，适用于传统业务系统和事务性应用	关系数据	一对一、一对多、多对多的关系	SQL 支持，高可靠性
MongoDB	面向文档的 NoSQL 数据库，适用于大数据量、高并发场景	JSON	无固定关系	高扩展性，结构灵活，支持大数据量和快速查询
Neo4j	图数据库，适用于复杂网络关系的数据分析	图数据	节点与节点之间通过关系连接	图数据存储和查询，适合社交网络等关系密集型应用
InfluxDB	时序数据库，适用于时间序列数据的存储与查询	时间序列数据	无关系，可通过时间戳进行查询	高效存储和查询时序数据，适用于监控、物联网、日志分析等

1. 适用场景和数据关系分析

- **MariaDB:**

- 1) **适用数据:** MariaDB 适用于存储结构化数据, 常见的应用场景包括传统的业务系统、银行业务、企业应用等。数据是表格结构, 并且需要维护数据的一致性。
- 2) **数据关系:** MariaDB 支持多种类型的关系模型, 包括一对一、一对多和多对多的关系, 可以通过外键和联合查询来管理关系。适用于复杂的业务逻辑和查询需求。

- **MongoDB:**

- 1) **适用数据:** MongoDB 适用于存储非结构化和半结构化的数据, 例如日志、产品目录、用户信息。由于其不强制要求固定的数据结构, 因此可以灵活处理不同格式的数据。
- 2) **数据关系:** MongoDB 通常用于处理关系较弱的数据库, 适合于无复杂关联关系的应用。

- **Neo4j:**

- 1) **适用数据:** Neo4j 为图数据库, 适合存储和查询复杂的网络关系数据, 如社交网络、推荐系统、路由优化等。
- 2) **数据关系:** 图数据库优势在于能处理大量相互关联的数据, 通过节点与关系的方式组织数据。Neo4j 通过图结构表示数据关系, 便于复杂关系的查询和分析。

- **InfluxDB:**

- 1) **适用数据:** InfluxDB 是时序数据库, 专门用于存储和查询时间序列数据。例如监控记录、传感器数据、日志数据等。
- 2) **数据关系:** 时序数据库通过时间戳来标识数据项, 因此不需要复杂的结构或关系。InfluxDB 适合处理按时间顺序排列的数据, 查询时以时间为主要维度。

2. 数据库特点总结

- **MariaDB** 适合关系型、结构化数据, 适合高事务性和复杂查询支持的应用。

- **MongoDB** 适合灵活、非结构化或半结构化数据，尤其当数据结构不稳定时，它的文档模型允许快速迭代开发。
- **Neo4j** 适用于复杂的关系网络数据，适合社交关系、推荐系统等高度连接性的场景。
- **InfluxDB** 适用于时间序列数据，适合物联网、日志监控等数据量大且基于时间的数据分析。

三、知识图谱

1. 小组成员分工

组员	学号	贡献比例	主要内容
赵泽亚	2023110983	25%	关系结构的构建
林子晗	2023113223	25%	节点结构的构建
赵学强	2023110914	25%	查阅资料，确定图谱内容
唐煜承	2023111456	25%	检查图谱是否重复，完善图谱

2. 三国图谱设计

选用三国人物关系的数据集进行导入，并且手动导入了部分三国中发生过的战役及战役发生地点。

图谱包含了以下节点：

- **人物节点 (Hero)**：每个节点代表一位三国人物。
- **战役节点 (Battle)**：每个节点代表一场历史战役，主要属性包括战役名称、战役年份、战役结果等
- **地理节点 (Location)**：每个节点代表一个地理位置。

图谱包含了以下关系：

- **参与 (参与战役)**：连接人物和战役，表示某个人物参与了某场战役。比如“刘备”与“赤壁之战”之间的关系可以表示刘备在赤壁之战中的参与。
- **战死 (死亡)**：表示人物在特定战役中死亡，链接人物和战役节点。比如“关羽”与“襄樊之战”之间的关系可以表示关羽在襄樊之战中战死。

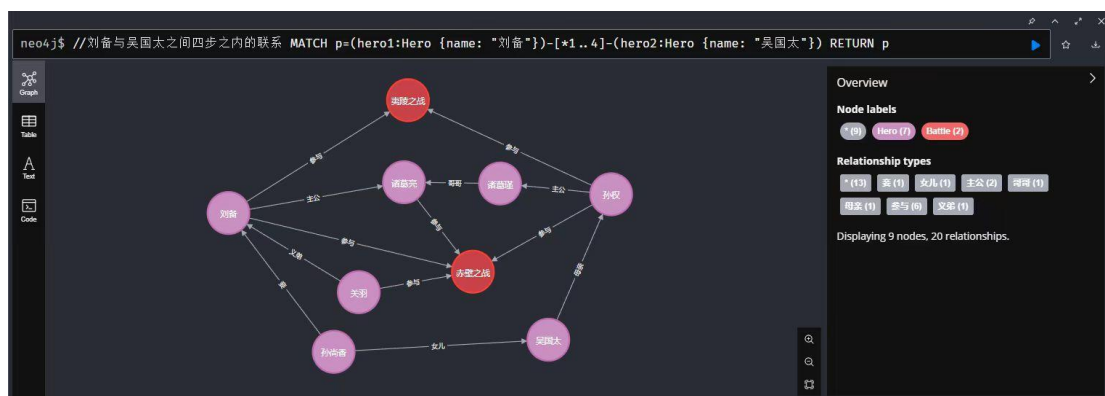


图 3 刘备与吴国太关联性

- 查询刘备与吕布的关联性。(如图 4 所示)

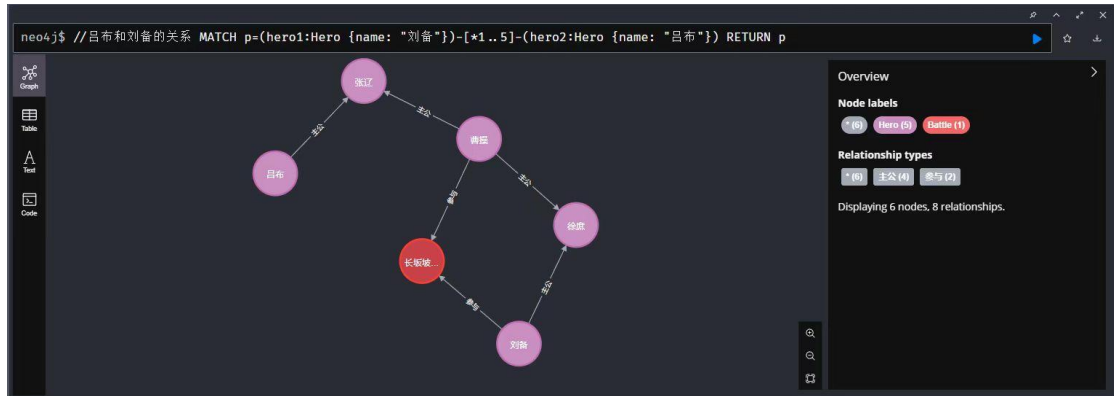


图 4 刘备与吕布关联性

- 按时间递增，发生过的所有战役。(如图 5 所示)

The image shows a Neo4j interface with a Cypher query: `MATCH (b:Battle) RETURN b.name AS Battle, b.year AS Year ORDER BY b.year;`. The table displays the following data:

Battle	Year
"源郡之战"	193
"平原之战"	193
"白马之战"	200
"华容道之战"	200
"官渡之战"	200
"南郡之战"	208

图 5 按时间递增的所有战役

4. 小结

通过使用 Neo4j，我们构建了一个基于三国人物、战役和地理位置的关系图谱。通过设计并创建包括人物、战役、地点等多个节点类型，通过相应的关系（如“参与”、“战死”等）连接节点。同时通过 Cypher 查询，高效探索和分析了人物之间的互动与历史事件的背景。

四、 InfluxDB 时序

1. 设计思路

- 实验内容

本实验利用 InfluxDB 作为时间序列数据库，结合 Telegraf 作为数据采集工具监测并收集电脑的各项使用情况数据，包括 CPU 使用率、内存与虚拟内存

空间的使用情况、磁盘容量以及磁盘 I/O 性能指标等，并通过 Grafana 进行可视化，选择合适的图表展示数据，便于后续的观察与分析。

● 整体数据流向

利用 Telegraf 从目标电脑收集性能指标数据；将收集到的数据通过 HTTP、UDP 协议传输到 InfluxDB 服务器；InfluxDB 接收并存储这些数据；使用 Grafana 创建图表，直观展示系统性能指标

2. DashBoard 设计与实现

● Overview of Server Resources

此仪表盘（如图 6）提供了服务器资源使用情况的全面概览，展示了关键的性能指标，可以获得的指标包括：机器运行时间、CPU 核心数、内存大小等等。

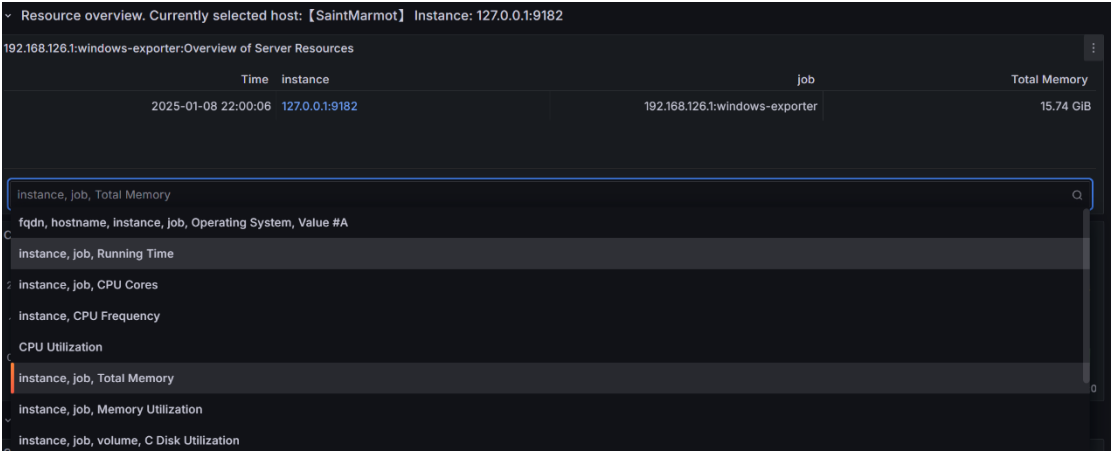


图 6 服务器资源使用情况

● CPU Usage by Host

此仪表盘（如图 7）展示了每个主机的 CPU 使用情况，旨在帮助用户实时监控服务器性能，并识别可能存在的性能瓶颈。通过对不同主机的 CPU 使用率进行展示，用户可以清楚地看到各主机的资源消耗情况，从而做出相应的优化和调整。

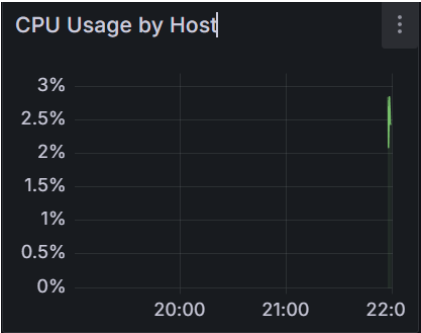


图 7 CPU 使用情况

- **Network details of the maximum traffic network card of each host**

此仪表盘（如图 8）展示了每台主机网络卡的详细流量信息，特别是最大流量的网卡。通过监控每台主机最活跃的网络接口，用户可以深入了解网络资源的使用情况。

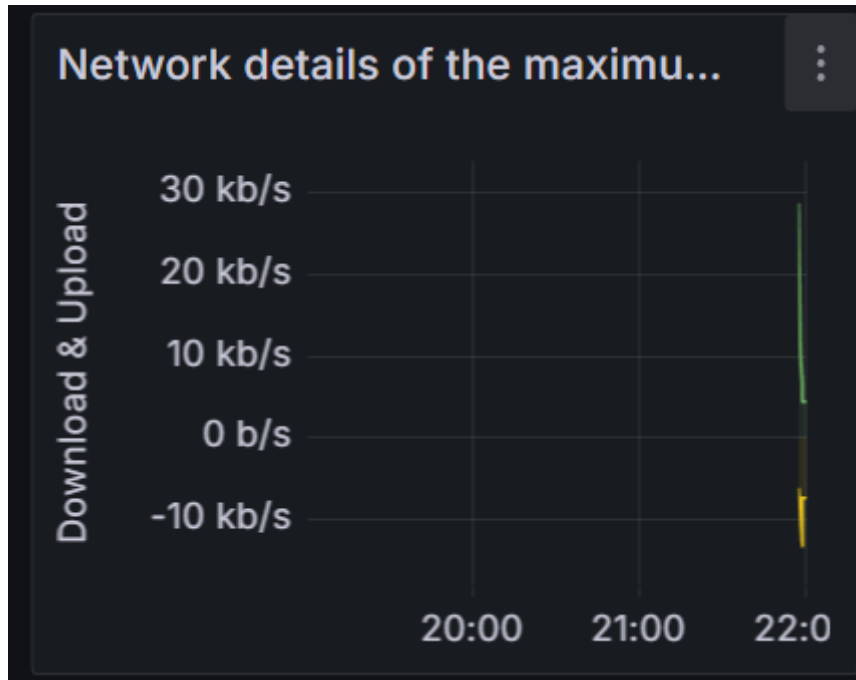


图 8 网络最大流量

五、 其他数据库

1. MySQL 在 Nextcloud 中的作用

我通过 MySQL 搭建了 Nextcloud 存储系统(140.143.210.11)，其中，MySQL 主要用于存储和管理所有与用户数据和系统配置相关的元数据。作为一个高效的关系型数据库，MySQL 提供了数据存储、查询和事务处理的能力，支持 Nextcloud 系统的正常运行和高效的资源管理。

2. 数据库的核心作用

- **存储用户和文件元数据**

MySQL 数据库是 Nextcloud 的后端数据库，用于存储用户数据相关的元数据。每个上传的文件和目录都会在数据库中创建对应的记录，包括文件名、文件路径、文件大小、上传时间、最后修改时间等信息。数据库确保了对文

件元数据的高效管理，使得 Nextcloud 可以快速查询、检索文件信息。

● 管理用户权限和共享设置

MySQL 数据库也用于存储用户权限、共享设置和访问控制信息。在 Nextcloud 中，用户可以创建共享文件夹、设置文件权限和控制文件的访问权限。所有这些信息都存储在 MySQL 数据库中，确保系统能够正确识别每个用户的权限并根据设置进行访问控制。

● 支持大规模数据管理

随着数据量的增加，MySQL 提供了高效的查询处理能力。通过优化索引和表结构，Nextcloud 能够高效地管理成千上万的用戶文件和相关的元数据，支持大规模数据存储和快速的检索。无论是通过文件名、用户、上传时间，还是通过文件的标签，用户都能通过 MySQL 实现快速的检索操作。

● 支持高并发操作

在多个用户同时操作时，MySQL 能够高效地处理并发请求。无论是文件的上传、下载，还是文件共享、权限修改等操作，MySQL 都能够通过锁机制和查询优化，支持高并发访问，确保系统的稳定性和响应速度。

数据流向与 MySQL 在 Nextcloud 中的工作流程

在 Nextcloud 系统中，用户上传文件时，系统会将文件内容存储到磁盘或对象存储服务中，而与文件相关的元数据（如文件路径、文件大小、修改时间等）则被存储到 MySQL 数据库中。每次对文件或文件夹进行操作时（如删除、修改或共享），相应的元数据会被实时更新到数据库中。

数据库中的数据表主要包括：

- 用户表：存储用户的基本信息（如用户名、密码、电子邮件、角色等）。
- 文件元数据表：存储每个文件或目录的相关信息（如文件名、文件路径、上传者、创建时间等）。
- 共享表：存储文件或目录的共享设置和权限信息。
- 会话表：存储用户登录状态和会话信息。

3. MySQL 与 Nextcloud 的集成

通过与 MySQL 的集成，Nextcloud 可以实现高效的文件管理、权限控制和数据安全。在部署 Nextcloud 时，系统管理员需要设置 MySQL 数据库并配置

其连接，确保数据库与 Nextcloud 的无缝对接。MySQL 提供了强大的数据存储和管理能力，而 Nextcloud 则通过其丰富的用户界面和功能模块，充分利用数据库的优势，为用户提供稳定、灵活的云存储服务。

六、总结

其实更认真做可以做得好很多但是我好累。。。。。。。。