

1. 2023-07-17 Day1

1.1. 今日小结

1. 今日上午入职。了解实习任务。
2. 初步整理了mongoDB 的相关资料，开始学习mongoDB的使用方式。
3. 学习 mongoDB 相关的基本概念并整理成册。

1.2. TODO

1. 去了解MongoDB 所在的生态位,去了解相关竞品.去思考mongoDB有什么替代品和优势.类似与postgresql
2. 了解一个 mongoDB 集群能接受的 QPS 有多大.
3. 了解布局一个 mongoDB 的成本有多大.
4. 去思考monngoDB 对于公司有什么意义.
5. 构造一个和公司业务有关的数据集.

目录

1. 2023-07-17 Day1	1
1.1. 今日小结	1
1.2. TODO	1
2. mongoDB 简介	2
3. 资料收集	3
3.1. 官方资料	3
3.2. 外文书籍	3
3.2.1. 实体书	3
3.2.2. 第三方资料	3
3.3. 中文资料	3
3.3.1. 实体书	3
3.3.2. 第三方资料	3
4. 零碎的内容	4
4.1. mongoDB 在易方达中的预期使用场景	4
4.2. 有关 benchmark	4
4.3. 租户 tenant	4
4.4. 其他	4
5. Day1 附录	5
5.1. 实习安排	5

2. mongoDB 简介

了解一个产品，最直接的方式是通过官网进行了解。除此之外，还可以关注社区论坛。

- 国内中文社区：<https://www.mongoing.com/>
- 国外社区：

以及可以通过一些第三方评测机构所出的对比，例如 DB-Engine 所展示的数据库热度榜，该热度榜综合考虑了一个数据库的搜索指数、就业情况、产品规模进行排序，可以快速了解一个产品的在产业内的地位。（国内与之相似的有 墨天轮，可以查看国内数据库的排名。）

419 systems in ranking, July 2023									
Rank			DBMS	Database Model	Score				
Jul 2023	Jun 2023	Jul 2022			Jul 2023	Jun 2023	Jul 2022		
1.	1.	1.	Oracle	Relational, Multi-model	1256.01	+24.54	-24.28		
2.	2.	2.	MySQL	Relational, Multi-model	1150.35	-13.59	-44.53		
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	921.60	-8.47	-20.53		
4.	4.	4.	PostgreSQL	Relational, Multi-model	617.83	+5.01	+1.96		
5.	5.	5.	MongoDB	Document, Multi-model	435.49	+10.13	-37.49		
6.	6.	6.	Redis	Key-value, Multi-model	163.76	-3.59	-9.86		
7.	7.	7.	IBM Db2	Relational, Multi-model	139.81	-5.07	-21.40		
8.	8.	8.	Elasticsearch	Search engine, Multi-model	139.59	-4.16	-14.74		
9.	9.	9.	Microsoft Access	Relational	130.72	-3.73	-14.37		
10.	10.	10.	SQLite	Relational	130.20	-1.02	-6.48		

Figure 1: 2023 年 7 月数据库热度排行榜

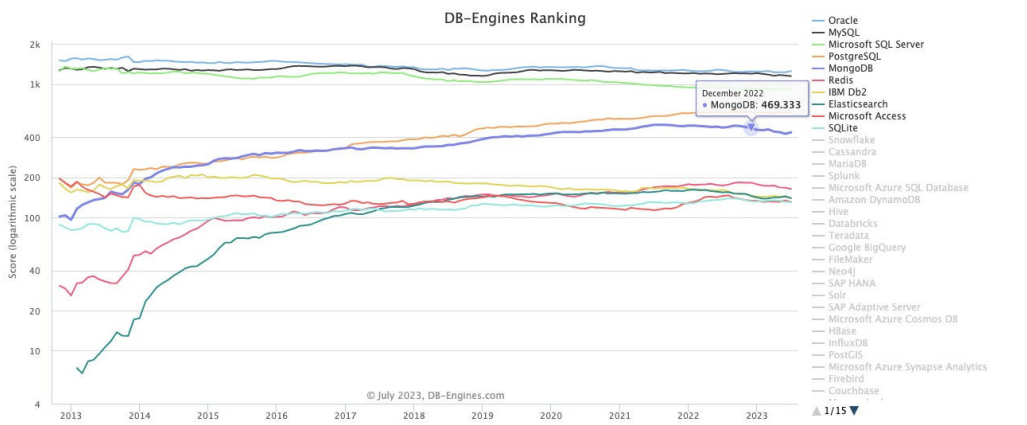


Figure 2: 2023 年 7 月数据库热度趋势图

可见在世界范围内各种类型的数据库相比，mongoDB 仍能位列全球第五，也是位列第一的非关系型数据库。

57 systems in ranking, July 2023									
Rank			DBMS	Database Model	Score				
Jul 2023	Jun 2023	Jul 2022			Jul 2023	Jun 2023	Jul 2022		
1.	1.	1.	MongoDB	Document, Multi-model	435.49	+10.13	-37.49		
2.	2.	2.	Amazon DynamoDB	Multi-model	78.81	-1.10	-5.13		
3.	3.	3.	Databricks	Multi-model	68.47	+2.65	+17.25		
4.	4.	4.	Microsoft Azure Cosmos DB	Multi-model	36.49	-0.08	-3.59		
5.	5.	5.	Couchbase	Document, Multi-model	25.14	-0.20	-2.77		
6.	6.	6.	Firebase Realtime Database	Document	18.73	+0.15	+0.39		
7.	7.	7.	CouchDB	Document, Multi-model	14.40	-1.02	-2.30		
8.	8.	↑10.	Google Cloud Firestore	Document	10.36	-0.52	+1.27		
9.	9.	9.	MarkLogic	Multi-model	8.59	-0.05	-1.19		
10.	10.	↓8.	Realm	Document	8.24	-0.08	-1.67		

Figure 3: 2023 年 7 月文档型数据库热度排行榜

单独观察文档型数据库的排行榜，可见mongoDB的热度远超任何一款NoSQL产品，是综合能力最强的一款产品。

3. 资料收集

本节列写了一些能够系统性地获取资料的页面。

3.1. 官方资料

mongoDB 官方提供了较为详尽的资源，也是学习 mongoDB 最主要的资料。主要包括下述三个方面。

1. 官方文档
2. 官方白皮书
3. 官方博客

3.2. 外文书籍

3.2.1. 实体书

目前，能在 zlibrary 上公开获取的主要包括如下书籍：

1. MongoDB The Definitive Guide Powerful and Scalable Data Storage 3rd Edition
2. MongoDB Performance Tuning Optimizing MongoDB Databases and their Applications
3. MongoDB Topology Design Scalability, Security, and Compliance on a Global Scale

第一本书是最为权威的 MongoDB 指南。第二本书讲述了调优角度。第三本书从网络拓扑角度出发描述了各个方面

3.2.2. 第三方资料

1. Youtube 等视频平台。

3.3. 中文资料

3.3.1. 实体书

在各种购书平台能收集到的书籍有：

1. MongoDB 权威指南
2. MongoDB 性能调优实战

这两本书是英文实体书的翻译。此外，还有部分国人所写的书籍，内容大同小异，都是入门书籍。此处不表。

3.3.2. 第三方资料

国内中文社区 列写了部分资深运维的所写的技术博客，是国内为数不多的中文资源。

- mongoDB 的 Documents 基本上复用了 javascript 的对象 (JSON) .只是讲存储格式修改成为了 BSON，而不再是 ascii。

去了解什么index 的建立是有效的,什么是无效的.

p47 可以在 array 上建立索引.

目前能了解到的,基本上只是各种优化的方法.

4. 零碎的内容

4.1. mongoDB 在易方达中的预期使用场景

和斌哥沟通,了解到当前企业内部使用 mongoDB 的原因是,部分产品会使用到 mongoDB,目前的使用方式在在应用本地启动一个 mongoDB.

接下来的改进方案是:

1. 通过建立一个同意的集群,来给各个应用做支撑.
2. 一个产品使用一个集群.

如果是第二种式样方式,我们需要一个尽可能方便的部署、测试、监控发布方式。

4.2. 有关 benchmark

- <https://www.mongodb.com/scale/mongodb-benchmark>

是官方提供的benchmark,使用的是亚马逊提供的工具,和 couchDB 进行了对比。

- <https://github.com/mongodb/mongo-perf>

是官方提供的,和mongoDB 集群之间对比性能的工具。

4.3. 租户 tenant

在mongoDB的企业版atlas上是存在 tenant 机制的。但是在社区版本中并没有。需要通过其他方式解决用户使用量的问题。

4.4. 其他

- 管理视角和运维视角之间的区别。
- 什么叫做用 apache? apache 又不是一个具体的产品吗?
- mongoDB 中三个特殊的DB: admin, local, config。这些是新版本中存在的 db。
- 总结 mongoDB Documents 之于 json 的区别。

5. Day1 附录

5.1. 实习安排

当前，实习计划初步安排如下：

项目名称：MongoDB集群运维

项目简介：本项目旨在让实习生通过学习MongoDB的基本概念和操作，深入了解MongoDB集群的架构、配置和运维，并能够完成一些常见的MongoDB运维任务。

项目内容：

1. MongoDB基础知识学习（约2天）
 - MongoDB的基本概念和原理
 - MongoDB的安装和配置
 - MongoDB的数据模型和查询语言
 - MongoDB的索引和性能优化
2. MongoDB集群架构与配置（约2天）
 - MongoDB副本集和分片集群的架构
 - MongoDB集群的配置和部署：学习如何在多个节点上部署和配置MongoDB，以实现高可用性和容错性。了解如何使用副本集和分片来提高数据库的性能和伸缩性。
 - MongoDB安全性和权限管理：设置MongoDB的安全性和权限管理，在数据库中实施访问控制策略，以确保只有授权用户才能访问敏感数据。
 - MongoDB集群的数据备份和恢复：设置自动化的数据备份和恢复策略，以确保数据在意外故障或数据丢失时得到保护。
 - MongoDB集群的监控和日志：设置监控系统，以跟踪MongoDB数据库的健康状况和性能指标，并设置日志记录系统以进行错误排查和故障排除。
 - MongoDB性能调优：深入了解MongoDB的性能调优技巧，如索引优化、查询优化、内存管理和磁盘IO优化等。能够使用工具对MongoDB进行性能调整，以优化查询性能、减少响应时间和提高吞吐量。
 - MongoDB自动化运维：了解如何使用自动化工具（如Ansible、Puppet或Chef）来部署、配置和管理MongoDB集群。
3. MongoDB运维实践（约4天）
 - 使用MongoDB Shell进行基本操作
 - 在MongoDB集群中添加和删除节点
 - 在MongoDB集群中创建和管理用户
 - 在MongoDB集群中进行数据迁移和故障恢复
4. MongoDB应用实践（2天）
 - 学习MongoDB数据库的使用场景，开发完成一个小项目
5. 实习总结和报告（约2天）
 - 撰写实习总结和报告，总结本次实习的收获和体验。

项目要求：

- 熟悉Linux操作系统和基本命令。
- 对数据库技术有一定的了解，了解SQL语言。
- 有良好的学习能力和沟通能力，能够承担相应的工作任务。
- 有一定的自我管理和组织能力，能够按时完成实习任务和撰写实习报告。

参考资料：

- MongoDB官方文档: <https://docs.mongodb.com/>
- MongoDB中文社区: <https://www.mongoing.com/>