

数据筛选

股权穿透图谱

- 1. 股东信息:
 - 股东类型
 - 投资人
 - 认缴金额, 实缴金额
 - 控股比例
- 2. 对外投资:
 - 对外投资企业名称
 - 投资年度
- 3. 分支机构:
 - 分支结构名称
 - 分支负责人
 - 注册时间
 - 状态
- 4. 变更记录:
 - 变更事项
 - 变更前内容
 - 变更后内容
 - 变更日期

有一些看似相关但是由于缺乏专业知识不太清楚的数据: 股权出质

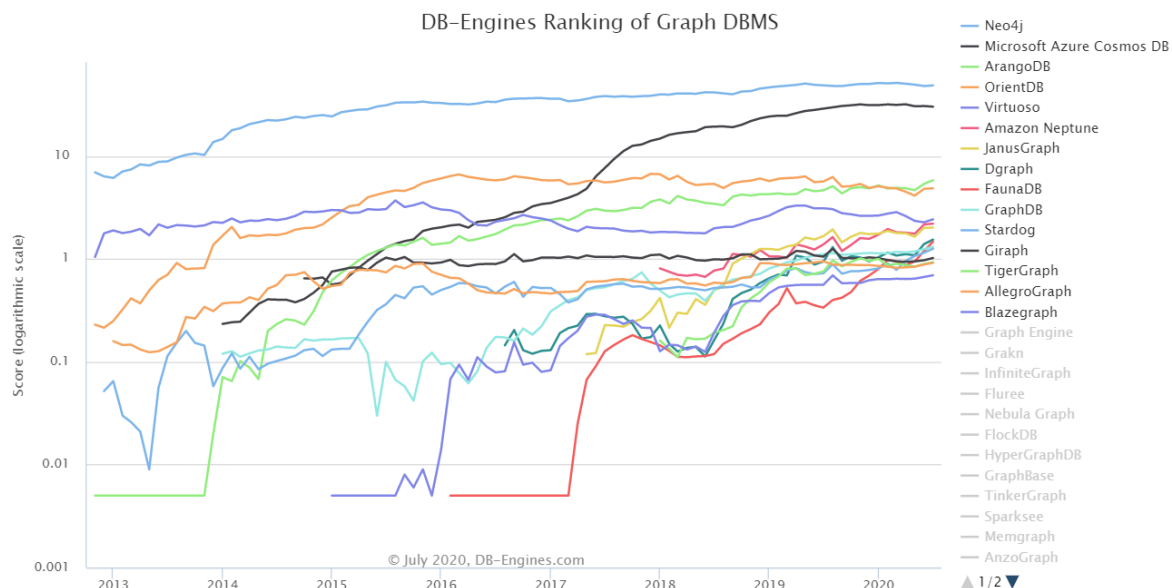
企业最终受益人

所需信息基本和上面相同.

Graph Database

图数据库概览

关于图数据库的对比, 网上搜索量比较高的是 [DB-ENGINES](#), [图数据库排名](#), 与之对应的还有一张图数据库的发展趋势图。



关于这个网站对数据库打分的依据，主要有以下几点：

1. 该图数据库在搜索引擎中的搜索结果数。主要以[Google](#)，[Bing](#) 还有 [Yandex](#) 的搜索次数作为衡量标准。
2. 该图数据库在搜索引擎中被搜索的频率。主要以[Google Trends](#)的数据作为依据。
3. 该图数据库在著名的技术博客网站中的讨论次数。主要以[Stack Overflow](#) 和 [DBA Stack Exchange](#) 的数据作为依据。
4. 有关该图数据库的就业分析。将主流的求职搜索引擎的上相关的offer数量作为主要依据，如[LinkedIn](#) 和 [Upwork](#)。
5. 在职业网络中个人主页的被提及次数。使用国际上最流行的职业网络LinkedIn和Upwork。
6. 社交网络中的火热程度。计算提及该图数据库的[Twitter](#) tweet的数量。

因此，从图中可以得到，Neo4j依旧是目前最活跃和主流的图数据库，其次是微软的Azure Cosmos DB。然而由于该网站收集数据的来源大多来自国外的互联网，因此这样的结果仍然是相对片面的。接下来，我会同时结合国内一些有关图数据库对比的技术博客做全面地概览和分析。

然而在做全面的分析之前，我们需要了解一个完善的图数据库需要哪些功能。

图数据库组件

一个完善的图数据系统应该至少包括数据库存储引擎，数据导入导出，管理运维，查询和计算，商业化产品需要有高可用及容灾备份。

1. **数据库存储引擎**：图存储是所有图数据库最重要的功能之一。此功能允许数据库用户以图的形式存储信息。数据库引擎为快速存储、查询、索引和检索提供处理和索引功能。具有高级索引功能的图数据库允许用户从大型数据库中快速检索图信息。
2. **数据导入导出**：数据从外界到图存储的导入导出能力，如从外界的json、csv、rdf等数据形式导入到图数据库中，或将图数据库中的数据导出来。
3. **管理运维**：管理运维则包含系统的监控，配置及可视化能力。
4. **查询和计算**：主要指提供查询语言供用户进行图的查询遍历等操作。

Neo4j

老牌的图数据代表。其功能强大，性能优秀，单节点的服务器可承载上亿级的节点和关系，单节点性能不够时也可进行分布式集群部署。

Neo4j有自己的后端存储，不必如同JanusGraph等一样还要依赖另外的数据库存储。Neo4j在每个节点中存储了每个边的指针，因而遍历时效率相当高。

Neo4j分为社区版和企业版，社区版功能受限，另外其提供可视化的客户端感觉很不错。

据neo4j的中国合作方的社区中描述，主要区别如下：

1. **容量**：社区版最多支持 320 亿个节点、320 亿个关系和 640 亿个属性，而企业版没有这个限制。
2. **并发**：社区版只能部署成单实例，不能做集群。而企业版可以部署成高可用集群或因果集群，从而可以解决高并发量的问题。
3. **容灾**：由于企业版支持集群，部分实例出故障不会影响整个系统正常运行。
4. **热备**：社区版只支持冷备份，即需要停止服务后才能进行备份，而企业版支持热备，第一次是全量备份，后续是增量备份。
5. **性能**：社区版最多用到 4 个内核，而企业能用到全部内核，且对性能做了精心的优化。
6. **支持**：企业版客户能得到 5X10 电话支持（Neo4j 美国电话、邮件，微云数聚电话、微信、邮件）。

社区版和企业版的对比

特性	企业版	社区版
数据库集群	支持	无支持
数据库热备份	支持	手工备份，运行时需要停止服务
在线存储空间再使用	回收并再使用数据删除释放的空间	无支持
LDAP/AD继承	与企业级用户身份认证和访问控制系统集成	无支持
访问控制	为数据库管理员提供指定的用户角色功能	无支持
数据库监控	提供监控数据库运行状况的整套指标	无支持
JMX日志和监控	包含GraphiteJMX客户端	无支持
更全面的日志	包括http，刚才， security日志	无支持
节点、关系、属性限制	无限制	节点344亿，关系344亿
节点键	使用一个或者多个属性来表示唯一的节点（类似主键）	无支持
运营管理工作	管理员命令行界面	无支持
写入锁、执行过程优化、硬件支持	支持超过4核心	最多4核心

OrientDB

OrientDB是第二代分布式图数据库,以混合数据模型为特点，它包括可以在最复杂的场景中使用复制和分片，并以Apache2许可证提供开放源代码。

ORIENTDB工作速度快，能够在最常见的硬件上每秒存储220000条记录，并且支持无模式、完整和混合模式，可以使用SQL作为查询语言之一。

OrientDB为确保更好的性能，最近引入了节点的快速重新同步，即使处理数十亿条记录，遍历速度也不会受到影响。

Teleporter功能是OrientDB的一个新工具，它允许用户与现有的关系数据库同步，并能够将它们完全迁移到OrientDB的图数据库。为了确保全程不会丢失任何数据，OrientDB还提供了备份功能。

OrientDB Community Edition可供用户出于任何目的免费使用。这为那些正在寻找低成本图数据库的用户提供了不错的选择。

和Neo4j的对比

1. OrientDB据描述性能可以达到Neo4j的数倍，但也有测试表明在遍历时磁盘空间增加，以空间换时间，遍历性能不高，但计算最短路径等性能高。
2. Neo4j和OrientDB在插入数据时候都会默认建立索引，索引的不同也造成了其不同操作的性能差异；
3. Neo4j：擅长遍历图及不存在大量关系的节点的图计算；OrientDB：侧重文档数据库，主要还是SB树索引导致，空间浪费比较大；插入节点与neo4j差不多，但是在插入节点关系即边时无优化；在图论算法上性能高，但遍历性能低。
4. OrientDB也有社区版及企业版，但是其基于Apache2.0协议，这个更友好

JanusGraph

JanusGraph是一个Linux基金会下的开源分布式图数据库。JanusGraph提供Apache2.0软件许可证。该项目由IBM、Google、Hortonworks支持。JanusGraph是由TitanDB 图数据库修改而来，TitanDB从2012年开始开发。目前最新版本为0.3.1。

JanusGraph支持多种储存后端（包括Apache Cassandra、Apache HBase、Bigtable、Berkeley DB）。JanusGraph的可扩展性取决于与JanusGraph一起使用的基础技术。例如，通过使用Apache Cassandra作为存储后端，可以将JanusGraph简单地扩展到多个数据中心。

JanusGraph通过与大数据平台（Apache Spark，Apache Giraph，Apache Hadoop）集成，支持全局图数据的分析、报告和ETL。

JanusGraph通过外部索引存储（Elasticsearch，Solr，Lucene）支持地理、数字范围和全文搜索。

和Neo4j的对比

特性 数据库名称	Neo4j	JanusGraph
是否开源	社区版开源，企业版收费	完全开源
第一版发行时间	2007年	2016年12月【基于TitanDB[2012]】
技术特点	一站式服务、工具齐全	需要开发者自行组合，因此技术门槛高
查询语言	Cypher	Gremlin
开发语言	Java	Java
集群	企业版支持，社区版不支持	支持
量级	轻量级	重量级
额外组件依赖	官方提供组件，可以与ES、MongoDB、Cassandra等NoSqlDb进行交互	数据存储服务可以选用Cassandra、HBase或Berkeley DB服务。数据索引可以选用Es、Solr或Lucene服务

HugeGraph

HugeGraph是一款面向分析型，支持批量操作的图数据库系统，它能够与大数据平台无缝集成，有效解决海量图数据的存储、查询和关联分析需求。HugeGraph支持HBase和Cassandra等常见的分布式系统作为其存储引擎来实现水平扩展。HugeGraph可以与Spark GraphX进行链接，借助Spark GraphX图分析算法（如PageRank、Connected Components、Triangle Count等）对HugeGraph的数据进行分析挖掘。

HugeGraph的主要特点包括：

1. 基于TinkerPop 3 API实现，支持Gremlin图查询语言。
2. 拥有完善的周边工具链和相关功能组件，可以满足图数据库开发的基本需求，提供易用高效的使用体验。
3. 具备独立的Schema管理模块，丰富完善的Schema校验机制，确保图数据库中的数据完整性和一致性。
4. 支持数据的备份和还原，可以在不同的后端存储之间转换。
5. 多种ID生成策略应对不同业务场景，拥有完善的索引管理机制，支持多种索引查询操作。
6. 可以实现与Hadoop、Spark、HBase、ES等大数据系统集成，支持多种Bulk Load操作，实现海量数据快速插入。

除上述特定之外，HugeGraph还针对图数据库的高频应用（例如：ShortestPath、k-out、k-neighbor等）做了特定性能优化，并且为用户提供更为高效的使用体验

综述

名称	Neo4j	OrientDB	JanusGraph	HugeGraph
可查文档	很多，技术成熟	很多	很少	只有国内文档
是否开源	仅社区版开源	开源	开源	开源
是否收费	仅企业版收费	WebUI的管理模块收费	不收费	
初始发行时间	2007	2010	2017	2013
实施语言	Java, scala	Java	Java	C++
服务器操作系统	Linux, OSX, Windows	所有带有Java JDK的操作系统	Linux, OSX, Windows	Linux
支持的编程语言	.Net, Java, Javascript, Groovy, Clojure, PHP, Python, Ruby, Scala	C, C#, C++, Java, Javascript, PHP, Python, Ruby, Scala	Clojure, Java, Python	.Net, C++, Java, PHP, Python
是否专注于图	是	否	是	是
SQL支持	是	否	是	是
是否有mapreduce	没有		有	有
是否支持集群	社区版不支持	支持	支持	支持
是否支持在线备份	社区版不支持	主-主备份	支持（后端框架支持）	支持（后端框架支持）
存储方式	原生数据库	原生数据库	Cassandra, Hbase	Cassandra, Hbase
是否支持大文件	不支持		支持	支持
图管理	社区版不支持		可以在配置文件中配置多图	无法配置多图
可插入节点数	320亿个节点		亿级	千万级
是否支持无向边	不支持	不支持	不支持	支持