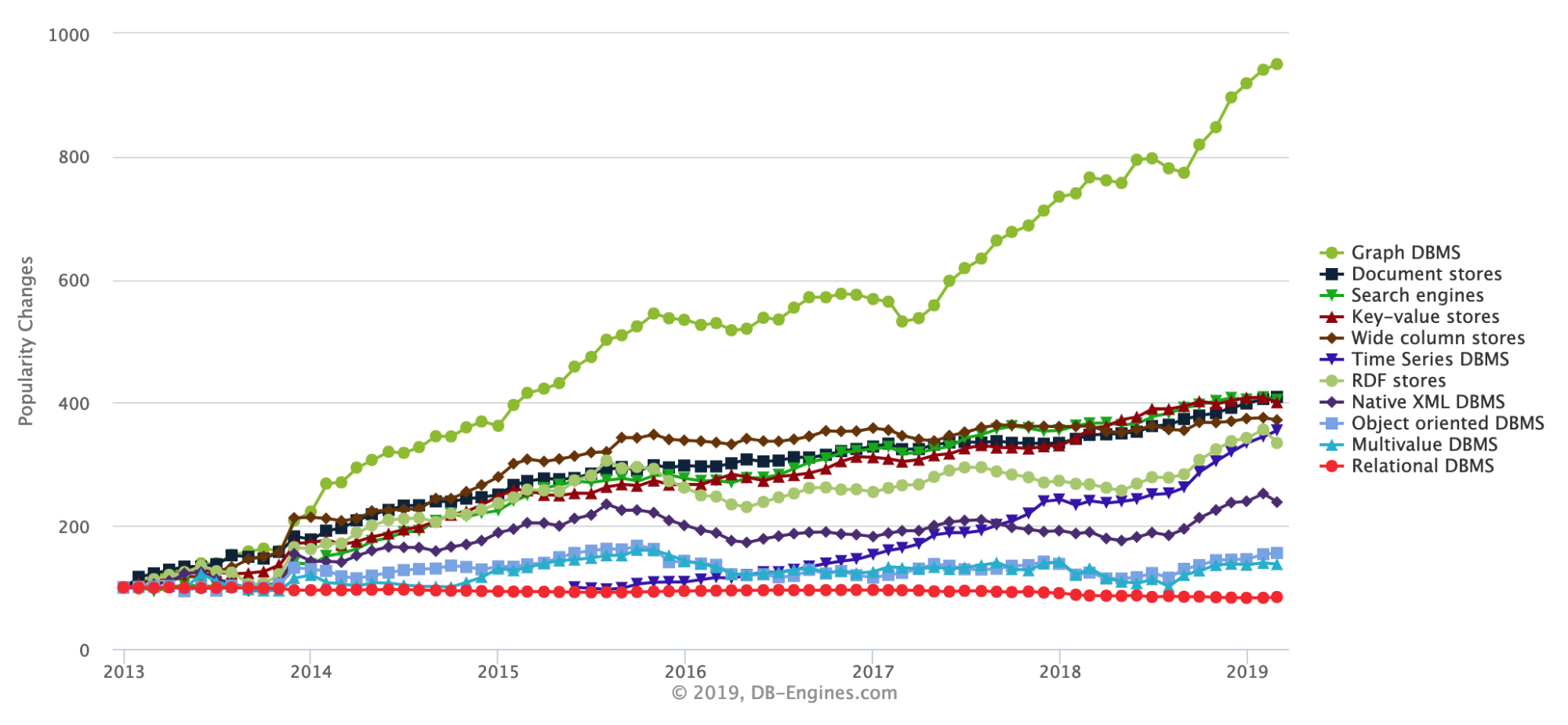
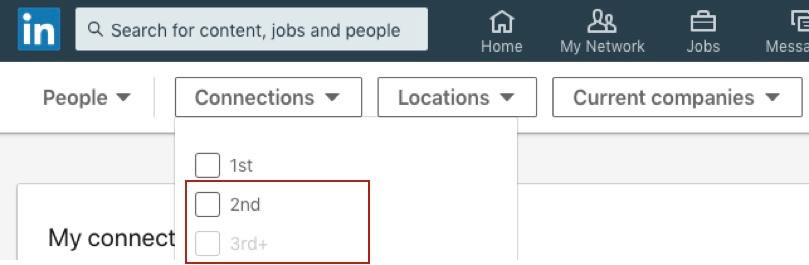
Atitit 数据库的类别图数据库 kv数据库 关系数据库 nosql数据库 对象数据库存储

## **背景**

如果你关注过db-engines.com，你很可能会发现有一种类别的数据库在最近几年正以火箭般的速度迅速流行起来，那就是Graph DBMS。阿里云于2019年3月重磅发布了数据库新产品——图数据库（Graph Database，简称GDB），本文将从图数据库的基本背景知识说起，带你深入了解阿里云Cloud-Native图数据库GDB。  
  


## **图数据库 vs 传统关系型数据库**

所谓术业有专攻，相对于传统的关系型数据，图数据库更擅长高密互连的数据处理。让我们先从Linkedin类似查询为例来管中窥豹看看图数据库的价值。



比如，用户UserId-xxx发起“\_查询自己2度好友中名字叫James的人\_”这样一个查询请求。以传统关系数据库的做法，你可能会写出如下查询语句：

select \* from t as t3 where t3.uid in(

select friend\_uid from t as t2 where t2.uid in (

select friend\_uid from t as t1 where t1.uid='UserId-xxx'

)

) and t3.name='James'

抛开查询语句的复杂度，该类查询语句在关系型数据库中执行起来是非常低效的，往往要秒级才能正常返回。如果进一步变为3度查询，则更慢了。那么，要是采用图数据库呢？以Gremlin为例，你可以很简单的写出这样一段查询语句：

g.V('UserId-xxx').repeat(both().simplePath()).times(2).dedup().has('name', 'James')

更为关键的是，这种2度，3度查询请求在图数据库中处理起来是非常高效的，性能往往能够轻松的达到传统关系型数据库的十倍乃至几十倍。这种性能的差异并非简单的调优问题，而是更深层次的数据库建模以及内核层面决定的。可以说，图数据库在基因层面更适合高度连接数据库的处理。

下面的表格将从更多层面对二者进行对比：

| **分类** | **关系型数据库** | **图数据库** |
| --- | --- | --- |
| 模型 | 表结构 | 图结构 |
| 存储信息 | 高度结构化数据 | 结构化/半结构化数据 |
| 2度查询 | 低效 | 高效 |
| 3度查询 | 低效/不支持 | 高效 |
| 空间占用 | 中 | 高 |

## **图模型**

图的概念源自图论，大家如果有所了解，可能听得最多的会是无向图，有向图这两种概念。

| **图分类** | **描述** | **示例** |
| --- | --- | --- |
| 无向图 | 一个图中所有的边都是无方向的，比较典型例子就是facebook社交网络中的好友关系。 | IMG_257 |
| 有向图 | 图中的边是有方向的，比较典型的例子就是Twitter这种社交网络中的关注关系。 | IMG_258 |

随着理论和技术的不断演进，当前在图数据库的应用当中，最为流行的是下面两种图模型：

| **图模型** | **描述** | **示例** |
| --- | --- | --- |
| 属性图 | 属于有向图，进一步细化了点和边的属性特点（部分进一步增加了Label概念），是当前图数据库采用的主流模型。 | IMG_259 |
| RDF图 | 核心是一种资源描述方式，每条描述都由主谓宾三元组构成。RDF作为W3C标准大量应用于知识图谱等相关应用。 | IMG_260 |

RDF是当前W3C的一个推荐标准，同时有标准查询语言SPARQL配合，在早期有大量RDF Stores产品出现。大家可以看到db-engines.com到目前还保留着RDF Stores这样一个分类。后来，属性图模型被大量图数据库产品所采用，属性图相对RDF更为灵活，新的图数据库也逐渐侵蚀部分RDF Stores的市场，更有不少产品在属性图基础之上做RDF及SPARQL的支持，以便无缝的承接RDF及SPARQL之上的各类应用及生态。

## **图数据库GDB整体介绍**

阿里云图数据库（Graph Database, 简称GDB）是一种支持属性图模型，用于处理高度连接数据查询与存储的实时可靠的在线数据库，支持 Apache TinkerPop Gremlin 查询语言，可以帮助用户快速构建基于高度连接的数据集的应用程序。阿里云图数据库GDB具有如下特点：

* 开放图查询语言：支持属性图模型，高度兼容Gremlin图查询语言
* 服务高可用：支持高可用实例，节点故障迅速转移，保障业务连续性
* 高度优化的自研引擎：高度优化的自研图计算层和存储层，云盘多副本保障数据超高可靠，支持ACID事务
* 易运维：提供备份恢复，自动升级，监控告警，故障切换等丰富的运维功能，大幅降低运维成本

### **开放图查询语言**

阿里云图数据库GDB支持属性图模型，公测版对外开放[Apache TinkerPop Gremlin](https://tinkerpop.apache.org/gremlin.html" \t "https://yq.aliyun.com/articles/_blank)图查询语言。



Gremlin是一由Apache TinkerPop提供支持的图查询语言，License为开源的Apache License Version 2.0。Gremlin和传统SQL有明显区别，是一种函数式、面向数据流的查询语言，使用户可以更为直接地控制和表达图查询的复杂逻辑。Gremlin包含一个宽松的DSL规范描述，以及基于Java/Groovy的开源实现。

阿里云图数据库GDB实现了Gremlin的高度兼容，支持直接使用开源的3.3.3版本以3.4版本开源客户端连接GDB进行操作，具体包括Gremlin Console，Java SDK，Python SDK，.Net SDK，并且支持Rest接口。如果希望进一步了解具体兼容性，可以参考：[GDB Gremlin实现的兼容性](https://help.aliyun.com/document_detail/102883.html?spm=a2c4g.11186623.6.564.50e76cdfNlwFZQ" \t "https://yq.aliyun.com/articles/_blank)

## NoSQL 数据库分类

|  |  |  |
| --- | --- | --- |
| 类型 | 部分代表 | 特点 |
| 列存储 | Hbase  Cassandra  Hypertable | 顾名思义，是按列存储数据的。最大的特点是方便存储结构化和半结构化数据，方便做数据压缩，对针对某一列或者某几列的查询有非常大的IO优势。 |
| 文档存储 | MongoDB  CouchDB | 文档存储一般用类似json的格式存储，存储的内容是文档型的。这样也就有机会对某些字段建立索引，实现关系数据库的某些功能。 |
| key-value存储 | Tokyo Cabinet / Tyrant  Berkeley DB  MemcacheDB  Redis | 可以通过key快速查询到其value。一般来说，存储不管value的格式，照单全收。（Redis包含了其他功能） |
| 图存储 | Neo4J  FlockDB | 图形关系的最佳存储。使用传统关系数据库来解决的话性能低下，而且设计使用不方便。 |
| 对象存储 | db4o  Versant | 通过类似面向对象语言的语法操作数据库，通过对象的方式存取数据。 |
| xml数据库 | Berkeley DB XML  BaseX | 高效的存储XML数据，并支持XML的内部查询语法，比如XQuery,Xpath。 |