**专场19 图数据库技术与实践**

船头尺 *2019.05.15 16:23*

1. HugeGraph 开源图数据库技术

演讲简介：

在 NoSQL 领域，从最近十年的表现来看图数据库已经成为关注度最高，发展趋势最明显的数据库类型。HugeGraph 是百度自主研发的一款图数据库，在百度安全事业部的反欺诈、黑产打击业务中发挥了重要作用。并以此为基础，逐步扩展完善成为可以支持广泛图数据库需求场景的通用图数据库，于2018年开源。本报告将结合过去几年的实践经验，分享 HugeGraph 图数据库的典型应用场景、高效使用方法、技术选型、技术架构、技术挑战与难点等方面内容。

李章梅 百度 HugeGraph 图数据库技术负责人

嘉宾介绍：

现任百度安全事业部资深研发工程师，HugeGraph 图数据库技术负责人；曾任职于华三通信、EasyStack 从事云计算与分布式存储相关研发工作；2012年毕业于河北科技大学，2015年开始持续贡献开源社区，包括 OpenStack、Ceph、ScyllaDB、TinkerPop 等社区，个人 GitHub 主页地址：[https://github.com/javeme。](https://github.com/javeme%E3%80%82)

2. 探索互联数据的奥秘 - 图数据库 GDB

演讲简介：

阿里图数据库 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、可扩展的在线数据库服务。基于图数据库 GDB，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，高效地构建社交网络、推荐引擎、欺诈检测、知识图谱等应用。

夜炎（王炎）阿里云智能数据库事业部高级技术专家

嘉宾介绍：

阿里云智能数据库事业部高级技术专家；在分布式系统、存储领域以及 NoSQL 系统研发领域经验丰富；18年加入阿里巴巴集团，现在阿里负责图数据库系统相关研发工作。

3. 分布式图计算的快速故障恢复技术

演讲简介：

分布式图处理系统需要不断增加节点数目，来应对数据规模不断增加所带来的计算挑战。计算节点的增加在加强系统算力的同时，增加了系统故障的概率，因此，急需一种有效的故障恢复策略。该报告探讨提出一套基于检查点+日志的快速容错方案，用以加快“以顶点为中心”的分布式图处理系统的错误恢复速度。在该方案中，无需重新计算所有数据，只需对分布在故障节点上的数据重新分区，分配到集群节点上进行并行恢复即可，从而加快系统的故障恢复速度。另外，为进一步加快故障恢复速度，探讨对提出的方案进行了进一步优化，如根据 cost model 进行分区、采用数据压缩减小通信和I/O代价等。实验表明，在运行在40个计算节点的分布式图处理系统上，本项目提出的基于图数据自适应迁移算法、基于数据特征的图划分最优算法的故障恢复策略，能比现有的工作平均快32倍。相关工作发表在数据库顶级会议和期刊：VLDB‘2015 和 TKDE‘2018。  
大纲：1、分布式图计算故障恢复的研究背景；  
2、分布式图计算故障恢复问题的定义；  
3、分布式图计算的总体架构与故障恢复技术实现。

卢卫 中国人民大学 中国人民大学副教授

嘉宾介绍：

卢卫博士拥有多年海外留学和工作经历，先后于2007年在澳大利亚昆士兰大学数据与知识工程研究（DKE）团队和2010年在新加坡国立大学数据库研究团队进行学术访问；2011年在中国人民大学获得博士学位后，于2011年7月至2014年9月在新加坡国立大学计算机学院从事博士后研究工作；是2012年中国人民大学优秀博士学位论文奖和 CCF 优秀博士学位论文奖的获得者之一，并在 CCF A 类发表论文十余篇。

4. 图数据库在智能问答和聊天机器人中的应用

演讲简介：

本次演讲主要介绍图数据在自然语言处理场景中的应用，这里包括利用非结构化数据来搭建图数据库，利用搭建好的图数据库来赋能教育领域客服机器人，智能医疗机器人以及订餐机器人。

李文哲 北京贪心科技 创始人兼 CEO

嘉宾介绍：

人工智能、知识图谱领域专家，北京贪心科技创始人兼 CEO，美国南加州大学博士，荷兰阿姆斯特丹大学访问学者；曾任凡普金科集团（爱钱进）首席数据科学家、北京会牛科技公司首席科学家，美国亚马逊高级工程师，美国高盛工程师等职位；在大数据、机器学习、自然语言处理、知识图谱等领域有丰富的研究和实践经验；在 AAAI, KDD 等顶会上发表过15篇左右论文。

5. Raft - 基于共识的分布式数据库协同算法及其在 Neo4j 集群中的实现

演讲简介：

分布式数据库高可用最佳实践：  
1. 分布式数据库的主要挑战：CAP Theorum  
；2. 分布式数据库协议：Paxos 和 Raft   
；3. Raft 协议的特点和设计思路；  
4. Raft 协议的重要内容：分布式日志，共识算法，Leader 及其选举，有效期（Term）；  
5. Raft 的主要过程：Leader 的选举；日志复制和事务提交；处理 Leader 失联；数据安全 / 一致性；  
6. Neo4j 的因果集群：实现的特性； 部署实例探讨；应用开发。

俞方桦 Neo4j 亚太地区技术专家

嘉宾介绍：

Neo4j 亚太地区技术专家，IEEE 和澳大利亚计算机学会会员；二十多年IT从业历史，职责包括开发、架构、咨询顾问、技术管理，行业经验跨越政府、零售、金融、科技、制造，专注领域涉及大数据、商业智能、AI和机器学习、云计算、软件开发以及儿童编程教育；现定居悉尼，业余时间喜爱读书、旅游、美食、网球和任何新颖的科技产品，联系方式：https://www.linkedin.com/in/joshuayu。

13：30 - 14：20 李章梅\_HugeGraph 应用案例与存储原理.pdf

15：30 - 16：20 卢卫\_分布式图计算的快速故障恢复技术.pdf

14：20 - 15：10 夜炎（王炎）\_探索互联数据的奥秘 - 图数据库GDB.pdf

16：20 - 17：10 李文哲\_图数据库在智能问答和聊天机器人中的应用.pdf

17：10 - 18：00 俞方桦\_Raft - 基于共识的分布式数据库协同算法及其在 Neo4j 集群中的实现.pdf