

本报记者 齐旭

随着全球各大科技巨头的竞相加入,开源软件技术已经活跃在各个信息技术领域当中。其中,大数据生态成为开源技术的直接受益者。开源技术适用于庞杂的数据管理系统,带来敏捷、高效、可扩展以及可自控的管理能力,并帮助企业降低IT建设及维护成本。2018年双11当天,阿里云原生数据库PolarDB轻松应对了0点0分0秒瞬时提升122倍的数据洪峰。Netflix也采用自研开源架构Metacat将海量数据集合成一个“单一”的数据仓库,大幅提升管理能效。

当前,我国有越来越多的企业、人才加入到开源社区,贡献力也“后来居上”,共同推进开源项目、开源生态的繁荣和可持续发展。

# 供需两侧均受益 开源成软件技术创新生态抓手

## 大数据生态成为开源技术重大“受益者”

近年来,在互联网服务、多媒体以及科学研究等多个领域,都可见到大数据的身影。在大数据时代,不断增长的数据量、快速处理数据的需求以及数据类型、结构和来源的多样性给数据库敏捷、高效、可扩展性以及个性化管理带来了全新挑战。

开源技术赋能了大数据生态的高质量发展。赛迪智库信息化和软件服务业所博士蒲松涛表示,经过了数十年的发展,开源软件和开源工具已经应用到了大数据产业发展的各个环节,基于开源软件,企业可以快速构建大数据应用平台,提供丰富的大数据开发和应用工具。

## 开源社区供需“双赢” 中国力量已崛起

开源成为一种加速技术研发速度、节约成本、效益最大化以及推动技术创新和产业发展的有效路径。阿里云智能数据库事业部负责人、达摩院数据库首席科学家李飞飞在接受《中国电子报》记者专访时表示,不同于传统的闭源商业软件,开源软件的创新不局限于一个企业或某些团队,而是无国界的,是群智开发的典型实践。

开源社区的建立为推动开源软件发展、构建行业竞争优势做出突出贡献,队伍的壮大需要每一位使用者持续不断的贡献智慧,以实现真正的“共赢”。开源的发展历程中,极客、大公司、商业颠覆者轮番登场,开源技术的诉求也从商业驱动向生态驱动发展。中国工程院院士廖湘科指出,开源是软件创新技术的主要



不断增长的数据量、快速处理数据的需求以及数据类型、结构、来源的多样性给数据库管理带来新挑战。

当前,几乎各种规模的企业都在使用开源软件和工具做大数据处理和基于数据的预测分析。开源界也涌现出了Hadoop、OpenStack、OpenShift、MapReduce、docker等引领行业技术创新方向的重量级开源项目。

华泰人寿基于OpenShift架构打造易于管理的新IT系统,以提升企业竞争力,实现业务数字化转型。在基础设施上,引入红帽OpenShift容器解决方案和红帽Ceph分布式存储。通过将保险业务上docker云,实现华泰人寿业务的弹性伸缩和快速上线,加速其互联网保险项目快速落地。

来源,是生态抓手,而非赢利的切入点。

开源软件的“共享和贡献”机制吸引了众多开发者的参与,给了每一位开发者“颠覆游戏规则”的权利。有了这种生态的加持,信息技术将被快速推进,各个参与者将持续获利。对此,李飞飞表示,开源生态的受益者是开源技术的需求侧和供给侧双方。从供给侧角度来看,参与的人越多,思维碰撞而引发的迭代演进就会越快;从需求侧角度来看,各个企业不仅可以免除被闭源系统“技术绑定”,还可以在开源社区实现数据库技术迁移,企业还可针对企业技术特征进行数据库的个性化定制,实现大量的应用和代码的改造且系统间互相兼容。

中国开源软件推进联盟副主席兼秘书长刘澎在PostgreSQL.CN 2019上表

## 阿里云数据库全球第三 中国开源社区生态未来看好

的产品中,还在SIGMOD、ICDE、VLDB等国际顶尖数据库技术交流会议上宣讲,侧面证明阿里云数据库与国际知名云数据库亚马逊AWS、微软SQL Server站到了同一高度。

记者:阿里为何自研云原生数据库PolarDB,未来将有怎样的技术布局?

李飞飞:随着移动互联网、物联网的发展,数据量剧增,企业上云大势所趋。传统商业数据库因昂贵、扩展性差、技术复杂,迭代慢饱受诟病。云原生数据库天然拥有云计算的弹性能力,兼具开源数据库的易用、开放特点,及传统数据库的管理和处理性能优势,是云时代下企业数据库的最佳选择。

不同于传统商业数据库的单节点和计算、存储绑定式架构,PolarDB具有的显著运行优势为整个开源数据库做出重要贡献。PolarDB能实现计算与存储分离、采用分布式共享存储,大幅提升了数据库的存储容量和数据吞吐量,巧妙解决了分布式事务处理等技术难题。此外,PolarDB还具有分钟级备份恢复、秒级弹性扩容能力,充分兼容MySQL、Oracle、PostgreSQL等架构。

客观来讲,目前PolarDB的架构也还有改进空间。由于分布式共享存储需要利用RDMA远程快速访问网络来实现,这就

### 多年实战经验催生自研

### 阿里云数据库技术领跑全球市场

记者:阿里云数据库近两年已进入世界各大研究机构榜单的前列,请问你是如何看待阿里云数据库在国际市场定位的?

李飞飞:从Gartner发布的全球云数据库的市场份额榜单来看,阿里云已经位列第三,超越了Oracle、IBM和谷歌。阿里云数据库技术不仅运用到阿里自身



美国知名在线影片租赁提供商Netflix也采用了大数据发现服务的开源框架Metacat。由于Netflix的数据仓库由许多大型的数据集组成,为了确保数据平台能够横跨这些数据集成为一个“单一”的数据仓库,Netflix开发的元数据服务Metacat,能让数据的发生、发现、处理和管理变得更加快捷高效、处理精度大幅提升;同时还可兼容Spark、Presto、Pig和Hive架构。Netflix软件架构师Ajay Majumdar指出,开放开源是身为技术公司的竞争战略,既能够将自己的解决方案建立为行业标准和最佳实践,又能建立Netflix的技术品牌,还能从共享

生态中获得反馈输入并受益。

事实上,推动大数据应用高质量发展的主流开源平台还有很多,例如Spark、Shark、Bagel等。蒲松涛表示,这些开源平台大幅降低应用门槛,有效帮助企业实现工业级应用,进而带动各行业大规模部署。此外,大数据还涌现出了一批开源支线平台。其中,Storm完全摆脱了经典的MapReduce架构,重新设计了一个适用于流式计算的架构,以数据流为驱动触发计算,计算时效性高,适应有向无环图计算拓扑的设计,计算方式较为灵活,在业界得到了一定的部署应用。

开源软件的“共享和贡献”机制吸引了众多开发者参与,给了每一位开发者“颠覆游戏规则”的权利。

力象限,这是该榜单首次出现中国公司。近日,Gartner发布的全球云数据库市场份额榜单中,阿里云位居第三,超越了Oracle、IBM和谷歌。5月21日,阿里云提供传统数据库一键迁移上云能力,可以帮助企业将线下的MySQL、PostgreSQL和Oracle等数据库轻松上云,最快数小时内迁移完成。李飞飞表示,阿里云自研的PolarDB云原生数据库的分布式存储架构具有一写多度、计算与存储分离等优势,帮助淘宝交易平台应对了双11当天瞬时提升122倍的数据洪峰。

此外,国内还有包括百度、浪潮、瀚高等在内的众多企业积极参与并贡献到开源社区当中。人工智能、自动驾驶等新兴信息技术也成为开源项目的重要应用领域。

导致该网络能够连接的节点数存在上限,在承受巨大并发量时容易出现問題。

完善后的架构将能更好地支持大体量、集中化的事务处理,实现多写多读、跨节点查询,事务处理的复杂程度大幅降低。

赋能新兴信息技术应用  
看好中国开源社区生态

记者:开源技术对人工智能、大数据、云计算、自动驾驶的发展有何推动作用?

李飞飞:当前,人工智能、大数据、自动驾驶等新兴技术和系统已经发展到了非常复杂的程度,如果没有开源生态的加持,我很难想象这些新兴技术能发展到今天的高度。

首先,从供给侧角度来看,开源式数据库的体系架构较封闭式的数据库系统能吸引更多有才华的开发者。参与的人越多,思维碰撞而引发的迭代演进就会越快。对于人工智能、大数据、自动驾驶这些极其复杂的新兴技术和系统来讲,任何一家闭源数据库公司的资源和开发力量都相对有限,技术需要越来越多的开源人才贡献力量。

### IT观察

## 美国叫停人脸识别和Libra 释放什么信息?

李佳师

最近美国频频爆出对新技术应用收紧的消息。先是在波士顿附近的萨默维尔市,议会于当地时间6月27日以11:0的投票通过了禁止该市相关机构在公共场所使用面部识别技术,这是今年5月旧金山出台法令禁止警察和其他政府机构使用人脸识别技术后,美国第二个禁止使用该技术的城市。紧接着7月2日下午,美国众议院财政委员会向扎克伯格等Facebook高管致函,要求立刻停止数字货币/钱包项目Libra/Calibra的所有相关工作。美国对AI与区块链等新技术的应用采取更为审慎的态度,释放什么信息?

未来学家凯文·凯利在1994年所写的图书《失控》中这样写到:机械与生命体之间的重叠在一年年增加,人造物表现得越来越像生命体,生命体变得越来越工程化。“当生命的力量释放到我们所创造的机器中的同时,我们也丧失了对它的控制。之后我们所有的造物主都必须面对我们不再完全拥有我们最得意的创造物。人造的世界就像天然的世界一样,很快具有自治力、适应力和创造力,也随着失去我们的控制。”凯文·凯利表示,人造世界失去控制,依然是个美妙的结局。

凯文·凯利写的是书,书里的世界失去控制,但如果说我们的创造物正脱离我们的控制,甚至颠覆我们的现实世界的时候,我们依然能够说“这依然是美妙的结局”吗?显然不能。

就像美国众议院财政委员会给Facebook高管的信函中所言:“Facebook发布的白皮书对于Libra/Calibra项目的意图、所扮演的角色、潜在用途和安全性方面提供的信息极少,暴露出了项目存在大规模风险并且缺乏明确的监管保护。如果这些产品和服务没有得到充分监管,甚至受到不当管理,它会带来危及美国 and 全球金融稳定的系统性风险,引发金融海啸。”每一次金融海啸、金融危机所引发的企业倒闭、经济衰退、经济萧条,我们仍历历在目,我们不能因为数字化推进的莽撞,而让全球经济陷入泥潭。

人脸识别技术尚不完全成熟,如果在不成熟的时候被滥用,同时又缺乏监管将会造成很大的风险。2018年7月,美国公民自由联盟(ACLU)使用亚马逊的Rekognition人脸识别产品将美国的国会535名议员与25000张警方公开的罪犯面部照片进行匹配,该软件错误的将28名议员识别为罪犯,这事曾引发了巨大的争议。2018年6月福布斯网站对亚马逊面部识别产品“Rekognition”进行了一次测试,证明了任何人都可以使用极其廉价工具来追踪他人的行踪。

应该说,旧金山是美国高科技企业汇聚的地区,是全球最顶尖AI技术的聚集区,照理说,这样的城市更应该对高科技的应用持“乐观”态度,他们的谨慎事实上给我们很多启示,我们对科技的采用需要更综合地进行多维度考量。如果风险大于获益,我们应该放慢推进的速度,寻求更可控的方式推进。目前在美国不仅仅是旧金山和萨默维尔市,正在有越来越多的城市考虑将禁止使用人脸识别技术。在欧洲,欧洲人工智能高级别专家组(HLEG)也在不久前发布报告,提议“欧洲应该禁止AI进行大规模监视和社会信用评分”。呼吁“政府应承诺只部署和采购值得信赖的人工智能系统,其设计宗旨是尊重法律和基本权利,符合伦理原则,社会技术健全。”

科技同样存在两面性。“当负责任地使用时,它可能是一个很好的公共安全工具。但如果人脸识别被滥用、缺乏监督和技术的不成熟,就会造成很大的风险。”赛迪智库产业政策研究所所长栾群表示,人工智能与其他技术不一样。

应用人工智能应谨慎,但不应该“一刀切”。赛迪顾问人工智能研究中心分析师李振表示,美国这两个城市考虑限制人脸识别软件的使用范围,是考虑到该软件对区域内居民个人隐私、种族歧视等问题的不良影响,这也是人工智能、物联网以及大数据等产业发展本身所涉及的问题所在。但新技术应用在伦理、法律等方面需要不断改进,不应简单粗暴地限制其使用,尤其在安防领域,人脸识别技术具有较强的技术优势,对维护社会稳定具有重要作用。以人脸识别等技术作为重要支撑点的人工智能产业是中国经济结构转型升级的良好发展契机,中国牢牢抓住本轮科技革命的发展机遇,积极布局,不应简单粗暴一刀切限制适用范围,可在法律监管等方面予以相应措施进行逐步规范。

赛迪智库政策法规所所长栾群认为,人工智能的治理日益成为国际共识。美国人的权利和法律意识很强,以伦理的眼光看待人工智能和提出治理规范,欧盟却走在了世界前列。2019年4月,欧盟发布《可信人工智能伦理准则》,提出可信人工智能伦理的7个关键条件——人的能动性、监督能力、安全性、隐私数据管理、透明度、包容性、社会福祉和问责机制,以确保人工智能足够安全可靠。与此同时,很多机构也发布有关人工智能原则,如IEEE提出的《合伦理设计的一般人工智能准则》,日本也提出过《以人类为中心的人工智能的社会原则》等,这些伦理准则表现出人类对人工智能技术发展的担忧——要利用人工智能技术实现生产效率提高和社会进步,但这一切都要建立在对风险的了解和预防的基础上。目前,中国正在积极融入人工智能全球治理体系,对于人工智能的治理,下一步中国应该实施国家规划,凝聚伦理共识,拟定评价规则,履行社会责任,管好重点行业,让所有主体行为都在法治的轨道内运行。