

## 官宣！Apache Doris 从 Apache 孵化器毕业，正式成为 Apache 顶级项目！

全球最大的开源软件基金会 Apache 软件基金会（以下简称 Apache）于美国时间 2022 年 6 月 16 日宣布，Apache Doris 成功从 Apache 孵化器毕业，正式成为 Apache 顶级项目（Top-Level Project，TLP）。

以下内容译自 Apache Doris 官网（<https://doris.apache.org/>）。Apache Doris 是一个基于 MPP 的现代化、高性能、实时的分析型数据库，以极速易用的特点被人们所熟知，仅需亚秒级响应时间即可返回海量数据下的查询结果，不仅可以支持高并发的点查询场景，也能支持高吞吐的复杂分析场景。基于此，Apache Doris 在多维报表、用户画像、即席查询、实时大屏等诸多业务领域都能得到很好应用。Apache Doris 最早是诞生于百度内部广告报表业务的 Palo 项目，2017 年正式对外开源，2018 年 7 月由百度捐赠给 Apache 基金会进行孵化，之后在 Apache 导师的指导下由孵化器项目管理委员会成员进行孵化和运营。“我们很自豪 Doris 能够顺利从 Apache 孵化器毕业，这是一个重要的里程碑。在整个孵化阶段，依靠 Apache 理念的指导和孵化器导师的帮助，我们学会了如何以 Apache 的方式去发展我们的项目与社区，也在这一进程中获得了巨大的成长。” Apache Doris VP 陈明雨说道。目前 Apache Doris 社区已经聚集了来自不同行业近百家企业的 300 余位贡献者，并且每月活跃贡献者人数也接近 100 位。在孵化期间，Apache Doris 一共发布了 8 个重要版本，完成了包括存储引擎升级、向量化执行引擎等诸多重大功能，并正式发布了 1.0 版本。正是依靠这些来自开源贡献者的力量，才使得 Apache Doris 取得了今天的成就。

与此同时，Apache Doris 如今在中国乃至全球范围内都拥有着广泛的用户群体，截止目前，Apache Doris 已经在全球超过 500 家企业的生产环境中得到应用，在中国市值或估值排行前 50 的互联网公司中，有超过 80% 的公司长期使用 Apache Doris，包括百度、美团、小米、京东、字节跳动、腾讯、快手、网易、微博、新浪、360 等知名公司。同时在一些传统行业如金融、能源、制造、电信等领域也有着丰富的应用。

“你可以基于 Apache Doris 快速构建一个简单易用并且性能强大的数据分析平台，非常易于上手，所需要付出的学习成本非常低。并且 Apache Doris 的分布式架构非常简洁，可以极大降低系统运维的工作量，这也是越来越多用户选择 Apache Doris 的关键因素。”

作为一款成熟的分析型数据库项目，Apache Doris 有以下优势：

- 性能优异：  
自带高效的列式存储引擎，减少数据扫描量的同时还实现了超高的数据压缩比。同时 Doris 还提供了丰富的索引结构来加速数据读取与过滤，利用分区分桶裁剪功能，Doris 可以支持在线服务业务的超高并发，单节点最高可支持上千 QPS。更进一步，Apache Doris 结合了向量化执行引擎来充分发挥现代化 CPU 并行计算能力，辅以智能物化视图技术实现预聚合加速，并可以通过查询优化器同时进行基于规划和基于代价的查询优化。通

过上述多种方式，实现了极致的查询性能。

- 简单易用：支持标准 ANSI SQL 语法，包括单表聚合、排序、过滤和多表 Join、子查询等，还支持窗口函数、Grouping Set 等复杂 SQL 语法，同时用户可以通过 UDF 和 UDAF 等自定义函数来拓展系统功能。除此以外，Apache Doris 还实现了 MySQL 协议兼容，用户可以通过各类客户端工具来访问 Doris，并支持与 BI 工具的无缝对接。
- 架构精简：系统只有两个 Frontend (FE) 和 Backend (BE) 两个模块，其中 FE 节点负责用户请求的接入、查询计划的解析、元数据存储及集群管理等工作，BE 节点负责数据存储和查询计划的执行，自身就是一个完备的分布式数据库管理系统，用户无需安装任何第三方管控组件即可运行起 Apache Doris 集群，并且部署和升级过程都非常简易。同时，任一模块都可以支持横向拓展，集群最高可以拓展到数百个节点，支持存储超过 10PB 的超大规模数据。
- 稳定可靠：  
支持数据多副本存储，集群具备自愈功能，自身的分布式管理框架可以自动管理数据副本的分布、修复和均衡，副本损坏时系统可以自动感知并进行修复。节点扩容时，仅需一条 SQL 命令即可完成，数据分片会自动在节点间均衡，无需人工干预或操作。无论是扩容、缩容、单节点故障还是在升级过程中，系统都无需停止运行，可正常提供稳定可靠的在线服务。
- 生态丰富：  
提供丰富的数据同步方式，支持快速加载来自本地、Hadoop、Flink、Spark、Kafka、SeaTunnel 等系统中的数据，也可以直接访问 MySQL、PostgreSQL、Oracle、S3、Hive、Iceberg、Elasticsearch 等系统中的数据而无需数据复制。同时存储在 Doris 中的数据也可以被 Spark、Flink 读取，并且可以输出给上游数据应用进行展示分析。

“毕业不是最终目标，它是新征程的起点。”陈明雨说到。“在过去，我们发起 Doris 的目标是为更多人提供体验更佳的数据分析工具、解决他们数据分析的难题。成为 Apache 顶级项目一方面是对 Apache Doris 社区过去所有贡献者一直以来辛勤工作的肯定，另一方面也意味着我们在 Apache Way 的指引下建立了一个强大的、繁荣的、可持续发展的开源社区。未来我们将会继续以 Apache 方式运作社区，相信会吸引到更多优秀的开源贡献者参与社区中来，社区也会在所有贡献者的帮助下得到进一步成长。”

“Apache Doris 后续将开展更多富有挑战且有意义的工作，包括新的查询优化器、对湖仓一体化的支持，以及面向云上基础设施的架构演进等等。欢迎更多的开源技术爱好者加入 Apache Doris 的社区，携手共成长。”

“我们再次由衷地感谢所有参与建设 Apache Doris 社区的贡献者们，以及所有使用 Apache Doris 并不断提出改进建议的用户们。同时也感谢一路走来，不断鼓励、支持和帮助过我们的孵化器导师、IPMC 成员以及各个开源项目社区的朋友们。”

**本博客文章除特别声明，全部都是原创！**  
**原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。**  
**本文链接: 【】（）**