

Hadoop的原生组件 满足企业需求

灵活性

对HDFS和HBase中存储的任何数据进行交互式分析

原生性

与Hadoop组件共享工作负载管理、元数据、安全性、SQL语法和用户界面

经济高效

减少数据移动、建模和存储的统一存储和分析

快速

行业领先的多用户性能，以及Hadoop可扩展性

用户友好

用预先构建的分析功能和现有BI工具集成来轻松探索和查询数据

熟悉

标准以及供应商专属SQL功能，利用现有的SQL技能

Cloudera Impala: Apache Hadoop的开源分析型数据库

Cloudera Impala是运行于Apache Hadoop中的开源分析型数据库。Impala结合了其他原生Hadoop框架的所有好处，包括灵活性、扩展性、经济高效，以及传统数据库所需要的性能、耐用性和SQL功能。Impala是Cloudera企业数据平台的一个集成部分 --- 享有像MapReduce、Apache Hive、Apache Pig等组件同样灵活的文件和数据格式、元数据、安全性以及资源管理。它还与广受欢迎的第三方工具无缝集成（比如IBM Cognos, Microsoft BI, Microstrategy, SAP BusinessObjects, SAS, Tableau, 以及更多）--- 使得企业可以继续利用现有的投资。

兼而有之的优势

自从Impala问世以后，它已经成为行业标准的SQL接口，有着超过100万的开源下载次数和几百个生产部署。作为Hadoop生态系统的一个原生组件，相比于传统的数据库和其他的Hadoop框架来说，Impala提供了关键的益处。它将数据开放给更多的用户以进行多样化分析；同时仍然可以满足企业的需求，提供了兼而有之的优势。

Hadoop的益处	同时满足企业要求
<p>灵活 - 可以用 schema-on-read 或 schema-on-write 来轻松处理结构化、半结构化和非结构化数据，来满足Hadoop所有数据的多样化分析需求。</p> <p>可扩展性 - 保持低延迟，在整个Hadoop规模内，通过直接查询Hadoop上的数据来实现互动性能。</p> <p>经济高效 - 通过一个统一的系统来支持高度的用户并发性，从而无需独立的或者多个系统。</p> <p>综合视角 - 在统一的数据存储中查询所有的原始历史数据，同时不会因为聚合或者遵循固定数据集格式所带来的信息流失。</p> <p>开放 - 开源的Apache许可和多个供应商支持，消除了供应商锁定。</p>	<p>极致性能 - 行业领先的用户并发性 - 比同类产品快近30倍。</p> <p>适用性 - 通过自动的基于成本的查询最优化与利用现有商业智能工具的能力，实现用户友好型的数据开发和分析。</p> <p>兼容性 - 强大的SQL合规性可以利用熟悉的SQL技术，同时额外的专属供应商扩展超越了标准SQL。</p> <p>安全性 - 与Hadoop和Cloudera Security的集成，实现简单的Kerberos授权和Apache Sentry的统一的、基于角色的授权 - 都可以轻松地通过Cloudera Manager来配置。</p>

为Hadoop带来分析型数据库

你可以用Impala来编写SQL语言，或者用你最喜欢的商业智能工具来交互式地查询Hadoop数据，如同传统的平行关系型数据库。对于那些选择SQL的人而言，Impala代表着强大的SQL合规性和必需的核心SQL语言功能，包括：

- > 分析性/窗口功能
- > 基于磁盘的查询处理
- > 在WHERE和EXISTS子句中的子查询（关联的和非关联的）
- > 传统数据库的附加专属供应商扩展

为了最大程度减少延迟，Impala直接在HDFS上运行数据（而不是通过一个中介，比如说Spark、Tez或者MapReduce），利用HDFS的缓存，优化了查询时的内存数据传输。其结果是将近30倍的速度提升与更高效的硬件使用率。

作为Hadoop生态系统的一个原生组件，Impala像该平台的其他组件一样，共享同样的存储（HDFS和HBase），元数据，和资源管理（YARN）。至于安全方面，则与Kerberos进行集成来实现授权（通过Cloudera Manager自动实现），以及与Apache Sentry进行集成来实现高精粒度的、基于角色的授权。此外，Impala可以在所有共同的原生Hadoop文件格式中进行查询。你可以查询文本文件、SequenceFiles、RCFiles和Apache Avro上的原始数据，同时实现最佳性能和Parquet列格式的空间效率。这些文件格式都是免费开源的，而且可以在Hadoop中被用来进行任何其他查询或处理，比如MapReduce、Hive或Pig。

Impala：量身打造的框架

Impala创建之初就以易于部署和使用为目的。为了保证与现行Hive集成组件的兼容性，Impala被设计用来利用、延展和提升Hive的关键基础架构 - 元数据、数据模式、安全性、驱动软件和SQL语法 - 实现最大化重复使用和集成而不需要增加设计工作、集成和培训。如果你理解了Hive，那么转化到Impala来获得显著的多用户性能、兼容性和适用性提升是很容易的。

Impala和Hive都在Hadoop生态系统中扮演很重要的角色。基于MapReduce和Spark框架的Hive，最适合长时间运行批次处理和转化（比如，抽取、转化和加载 - 或者ETL）。Impala是专门设计成一个原生的Hadoop分析型数据库，具备极致的性能、适用性和扩展SQL兼容性 - 同时可以查询统一的Hadoop存储。

Impala的常见用例包括：

>运营数据商店 - 轻松集合多个来源的数据，并且为内外部使用快速呈现操作面板。

>数据发掘工具 - 大规模快速探索数据，通过熟悉而强大的SQL能力，以及与流行的可视化工具的完美集成。

>专属构建应用的分析后台 - 通过一个集成平台，以更低的成本将数据和系统向更多的用户开放。

Cloudera企业数据平台中的Impala

Impala是Cloudera企业数据平台的一个关键组件：一个让存储和分析所有数据和元数据的统一解决方案，同时提供合规的安全性和监管，以及端对端的系统管理。Impala作为企业数据平台的一部分，全球的领先组织可以从更多、更多种数据中快速获得价值。

Cloudera 企业版订阅

通过订阅Cloud企业版来最大程度获取Impala的价值。Cloudera企业版提供全面的技术支持和关键任务应用的必要特性，以最大化Impala所创造的价值。



下载Cloudera Search
cloudera.com/downloads

请试用Cloudera Live
cloudera.com/live