

基于HBase的分布式 NewSQL数据库

HBase的优点及缺陷

HBase是Hadoop生态体系中最著名的分布式NoSQL数据库之一，其设计理念来源于Google的Bigtable。HBase主要组件包括HMaster和HRegionserver，为用户提供表格类型的数据模型，表按主键范围划分为多个region，HMaster负责管理和分配region，HRegionserver负责region数据的读写。HBase的主要优点有：

- **集群可扩展性**：HBase具有良好的可扩展性，集群服务能力随着数据节点扩容可以得到准线性提升。在实际应用中，HBase集群可以扩展到千台机器，处理PB级别的数据。
- **数据多副本**：HBase底层存储依赖于HDFS，写入HBase的数据通过存储多副本保证数据高可靠。
- **服务高可靠性**：HBase的架构中不存在单点，当集群中出现机器故障时系统能自动进行容错处理，恢复服务。
- **优秀的主键查询性能**：HBase能够提供毫秒级的主键查询性能，20MB/节点的顺序扫描性能。
- **稳定的插入、更新性能**：HBase底层存储采用log base模型，能够在任何数据规模下保证稳定的插入、更新性能。

HBase的以上优点使其成为目前应用最广泛的分布式NoSQL数据库之一，但随着越来越多的应用尝试迁移到HBase上，HBase的缺陷也越来越被暴露出来，主要包括：

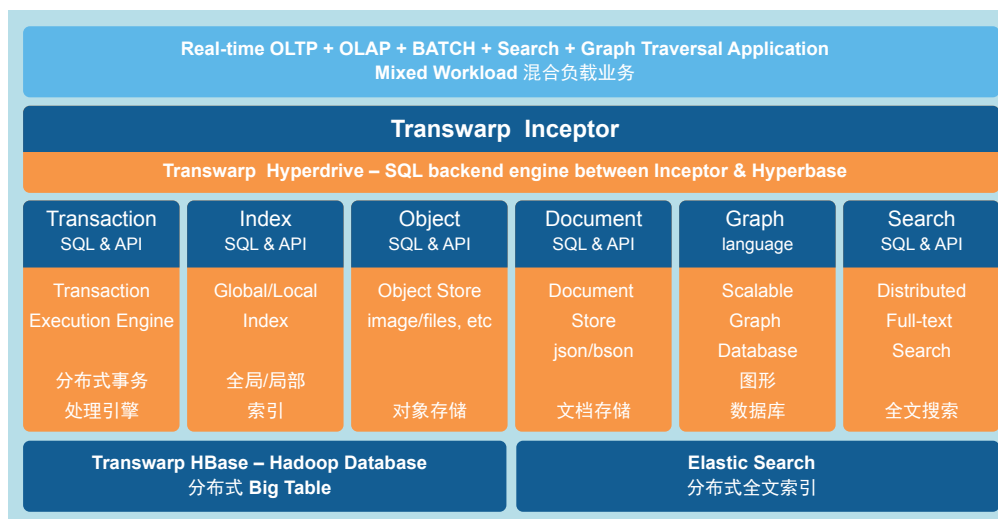
- **使用成本高**：用户需要通过API编程访问HBase，对于复杂应用使用成本过高；不支持标准JDBC/ODBC接口，ETL过程十分复杂。使用成本过高直接导致许多较为复杂的应用无法使用HBase。
- **无法有效的支持非主键查询**：在实际应用中，用户往往需要进行多维度的查询，HBase无法有效的支持非主键查询。
- **只支持单行事务**：在实际应用中，事务往往涉及到多表中的多行数据，HBase提供的单行事务无法满足应用需求。
- **无法有效支持半结构化数据**：HBase的数据模型是完全结构化的，无法有效支持半结构化形式数据（如JSON）。
- **无法有效支持图片存储**：在公安，交通等领域中，用户往往需要存储大量的图片数据，典型的一张图片大小在500K到2MB之间，实践证明HBase无法有效的满足图片类型的存储。
- **可用性较低**：HBase的每个region同时只在一台HRegionserver上提供服务，当出现HRegionserver故障宕机时，该HRegionserver上所有region对应的数据会暂时不可用，直到容错机制将region重新分配到其他HRegionserver上，这使得HBase的可用性不足以满足一般在线业务的需求。



Transwarp Hyperbase

Hyperbase技术特点

星环科技推出的Hyperbase有效的解决了HBase中存在的问题，Hyperbase是基于HBase开发的分布式NewSQL数据库，其架构如下：



Transwarp Hyperbase NewSQL数据库架构

Hyperbase的主要特点包括：

- **支持SQL，支持JDBC/ODBC：**Hyperbase通过星环自研的交互式分析引擎Inceptor支持SQL，用户可以通过编写SQL在Hyperbase上实现复杂的业务逻辑，极大的降低了使用成本；支持JDBC/ODBC接口，极大的简化了ETL过程。
- **支持二级索引，高效解决非主键查询需求：**Hyperbase允许用户根据具体的业务逻辑灵活的建立二级索引，在实际应用中用户往往会建立多个二级索引，在使用时Hyperbase会根据查询条件动态计算使用索引的代价，自动选择最合适的索引。
- **支持分布式事务：**Hyperbase支持跨行，跨表的分布式事务，支持完整的ACID事务语义，并且重点优化了高并发情况下的事务操作，可以满足大部分的OLTP应用。
- **支持半结构化数据：**Hyperbase支持JSON数据格式，用户可以直接将JSON格式的数据存储在Hyperbase之中，并对JSON的任意字段进行查询，创建索引以及删改。
- **支持全文检索：**Hyperbase通过Elasticsearch支持分布式全文检索，用户可以为自己的表创建全文索引，并在SQL中使用全文检索语法进行搜索。
- **支持高效的图片存储：**Hyperbase提供LOB存储，LOB可以高效的满足单条数据大小在几百K到10M的二进制存储需求，用户可以通过LOB满足图片存储需求。
- **高可用性：**Hyperbase允许为region同时维护多个副本，region多副本机制可以保证在HRegionserver宕机时完全不影响读服务，写服务秒级恢复，有效的提升可用性，满足在线业务的需求。

📍 地址：上海市徐汇区桂平路481号18幢3层301室（漕河泾新兴技术开发区）

✉ 邮编：200233

☎ 电话：4008-079-976

🌐 网址：www.transwarp.io

