拓尔思 TRS
语义智能 300229

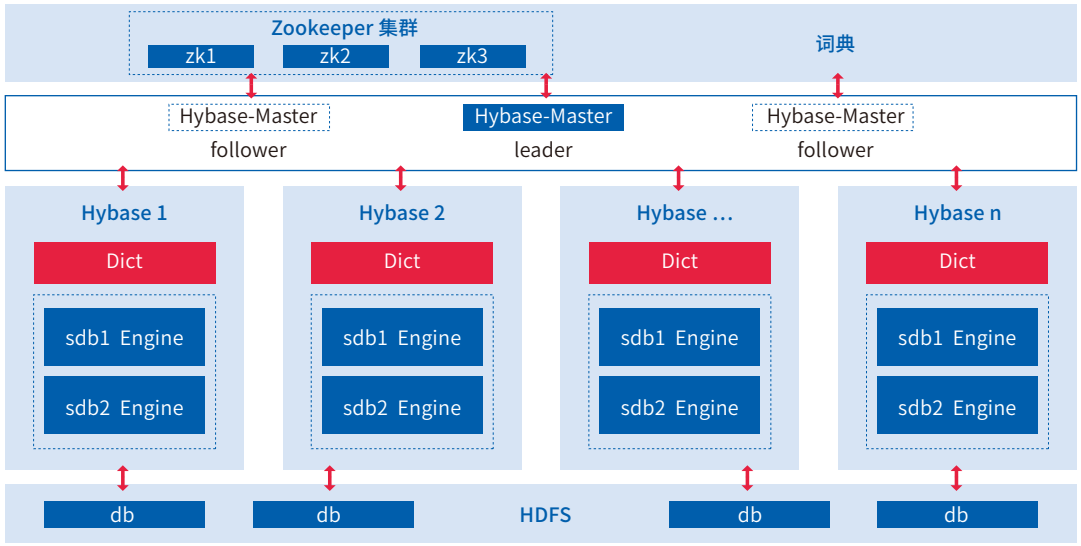
TRS Hybase 海贝大数据管理系统

以存储、检索和统计为核心的大数据时代检索引擎

Hybase产品概述

TRS Hybase海贝大数据管理系统定位为大数据应用支撑软件，为大数据应用提供高效的数据存储、全文检索、分类统计等数据管理服务，支持几乎所有类型的数据，包括文本、数字、地理空间、图片等各种结构化、半结构化和非结构化数据。系统从设计之初就考虑到海量数据的特点，支持SSD、HDD、HDFS混合存储，引入了多副本、列存储和内存索引等机制，充分利用现代计算机多CPU、大内存的特点，实现分布式并行计算，满足用户对海量数据交互式处理的需求。

产品架构（第三代）



产品设计理念

简单易用

开箱即用，降低使用门槛，一体化管理
自动部署、更新

高性能

充分释放硬件资源，多核、大内存、SSD

安全可靠

权限管理、黑白名单，HTTPS、加密存储

存储、检索、统计

全息索引，任意维度组合查询、图片索引、地图索引、二进制索引、分类统计、多维下钻

弹性扩展

单集群支持PB级别存储量、单集群支持万亿级别数据量、支持在线扩容

生态对接

TRS ETL、TRS 水晶球、Hadoop/Spark

Hybase技术优势

核心技术

内存索引
随机读写->顺序读写

全文索引
支持任意维度的组合查询

列存储
排序加速、统计加速

自然语言处理
语义化检索

多副本
数据安全、在线扩容

索引分片
分布式并行计算

产品优势

安全
用户IP黑白名单控制
用户逻辑隔离和物理隔离
数据加密存储、传输
权限管理

高效
实时索引、离线索引、
读写分离、二级分区、
列存储、冷热数据分离
全文索引

智能
全语种数据统一处理
字词混合分词
词典的统一维护，无需重构索引
自然语言处理

可靠
扁平化设计，弹性扩展
PB级海量数据支撑
基于HDFS多副本机制
异常感知

功能
功能一体化系统管理台
支持自动化部署和自动化更新
高效技术支持服务
企业级服务

混合存储
支持SSD、非SSD、HDFS等
多种文件系统统一管理
海量存储

Hybase 新功能特性

图文检索 引入深度学习引擎技术支持“以文搜图、以图搜图、图文一起搜索”等功能。

以文搜图：利用输入的关键词查找图片所在文档的文本或图片中的文字属性，从而搜索到相关图片。

用户在提交图片的同时，加入常用的关键词描述，能够大大提高检索的效率和准确率。

以图搜图：通过对提交图片的特征进行抽取，并与库中其它图片特征进行比较，搜索得到相似图片。

镜像视图 数据方面实现读写分离、访问隔离变得更加智能化。

读写分离

避免数据同时入库+检索时的锁竞争；节点分离，避免系统CPU、IO、内存等资源的竞争

大小库

多个镜像视图，不同配置
镜像视图1 — 一个月的数据
镜像视图2 — 近半年的数据

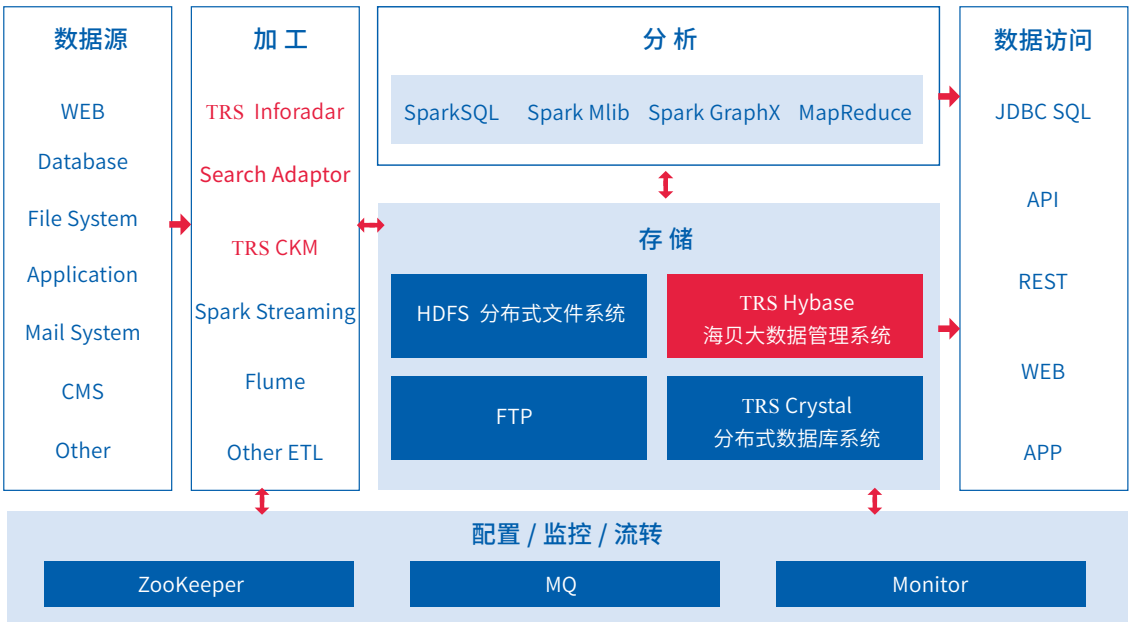
访问隔离

多个镜像视图，不同配置
镜像视图1 — 自动分析程序
镜像视图2 — 在线用户访问

Hybase大数据生态

TRS Hybase海贝大数据管理系统不仅仅是一个简单的检索引擎，拓尔思以海贝大数据管理系统为支撑，构建了一套从数据采集（TRS InfoRadar）、数据加工（TRS Search Adapter、DL-CKM）、数据存储、数据分析（TRS水晶球）、数据展现等多款产品融合交互的完整大数据生态流程体系。

TRS Hybase海贝大数据管理系统生态流程图



Hybase应用概况

TRS Hybase海贝大数据管理系统作为一款基础搜索平台，依托拓尔思自身在中文检索领域的专业优势，在保障数据安全性的基础上已为政府、媒体、公安、军工等客户提供服务：1993年服务于新华社，后续为国家知识产权出版社、国保局、专利出版社等提供优质的企业搜索服务。

产品历程

- **新闻资料检索系统** — 1985年
- **TRS Database Server** — 服务于新华社 — 1993年
- **TRS Database Cluster** — 服务于国家知识产权出版社 — 2004年

服务领域

公安

公共安全

某公安、安全部各地省厅

媒体

媒体融合

新华社、浙报、重报

政府

政务公开

海关总署、工商总局
专利局、商标局

信创

信创产业

TRS Hybase V7.0

服务于某部云搜索
国保局等项目

2012年

TRS Hybase V8.0

服务于专利出版社

2017年

TRS Hybase V9.0

引入深度图像检索
服务于某互联网监管单位等

2020年

案例1：信用中国

2018年底“信用中国”网站从某云迁移部署到国家电子政务外网机房，将原来的搜索服务迁移到TRS Hybase大数据管理平台。“信用中国”系统，自检索服务迁移至TRS Hybase平台后，利用6个节点可支撑供每秒5000+次的高并发检索，且运行稳定，国家公共信用信息中心给予了高度认可和赞赏。

案例2：中国知识产权大数据与智慧服务系统

TRS Hybase针对1.3亿件全球专利，采用7台高配置PC服务器成功搭建“中国知识产权大数据与智慧服务系统”的后台大数据管理系统。该系统充分利用TRS Hybase在自然语言处理方面的优势，利用单一分词器处理全语种数据，大大简化了系统的构建，提供跨语种检索服务。

赋能数字经济

拓尔思信息技术股份有限公司
TRS Information Technology Co.,Ltd.

总部地址：北京市朝阳区大屯路风林西奥中心B座16层　邮编：100101
电话：010-64848899　传真：010-64879084　服务热线：4006 300229
E-mail：trs@trs.com.cn　网址：www.trs.com.cn

扫一扫 关注拓尔思