

星环科技大数据产品和技术介绍



星环科技

www.transwarp.io

2017年7月5日

星环科技公司简介

TRANSWARP

星 环 科 技

全球最具有前瞻性的大数据厂商

——Gartner 2016年数据仓库及数据管理解决方案魔力象限远见者领域

- 中国最早自主研发的大数据基础软件
- 最早入选中央政府采购网的Hadoop发行版产品
- 率先支持银行业复杂关键应用的大数据平台
- 国内落地应用案例最多的大数据厂商
- Gartner认可的全球六大（中国唯一）Hadoop发行版厂商之一

Gartner.

G00273381

Market Guide for Hadoop Distributions

Published: 15 December 2015

Analyst(s): Nick Heudecker, Merv Adrian, Ankush Jain

Representative Vendors

The vendors listed in this Market Guide do not imply an exhaustive list. This section is intended to provide more understanding of the market and its offerings.

Selected Hadoop Software Distribution Vendors

Transwarp

Headquartered in Shanghai, China, Transwarp offers Transwarp Data Hub, a full suite of Hadoop components supplemented by its own NoSQL search engine, batch analytics, stream processing and data mining componentry. It supports Docker and Kubernetes and offers a unique compiler for Oracle PL/SQL. Its operation today is almost entirely in China.

在十月份的webinar Hadoop 2016: Moving Into Mainstream中，Gartner宣布了市场中的新成员——星环科技的加入。而在最新发布的研究报告Market Guide for Hadoop Distributions中，Gartner将星环，Cloudera, Hortonworks, IBM，MapR 和Pivotal一同被列为Representative Vendors（最具代表性厂商）中的主流Hadoop发行版厂商。Gartner对星环及其产品的概括如下：

星环

总部设在中国上海。星环科技的产品名为Transwarp Data Hub，它包含一整套Hadoop组件，星环科技还为Transwarp Data Hub配备了星环独有的NoSQL搜索引擎、批量分析组件、流处理组件和数据挖掘组件。Transwarp Data Hub支持Docker和Kubernetes，还提供独特的PL/SQL编译器。目前，星环科技大多数业务在中国。



近日，国际知名咨询机构Gartner针对当前数据仓库及数据管理解决方案市场，在其魔力象限1.中对全球21家厂商进行了对比分析，其中，Oracle、Teradata、Oracle和Microsoft三家公司包揽了前三甲，而在**大数据核心技术最领先的公司**却是来自中国的公司——星环科技（Transwarp），这也是**该领域的魔力象限中第一次出现中国公司**。

优势

星环科技虽然年轻，但是已经在中国市场中颇具影响力——星环科技在18个月内赢得了200个客户。

星环科技的产品有其独特的功能，例如它的SQL引擎Inceptor，基于Apache Spark，兼容Oracle SQL和PL/SQL，支持事务处理的CRUD(CREATE, READ, UPDATE, DELETE)并能保证ACID。Inceptor在被调查的用户中受到了非常高的评价。

被调查的用户表示他们对星环的产品、支持以及提供的培训都非常满意。

注意

目前，星环仅在中国有业务。虽然如此，中国庞大的市场以及中国市场特有的要求给星环的发展空间巨大。

虽然暂时还没有推出云上的解决方案，但是星环科技的云解决方案已经在计划中。

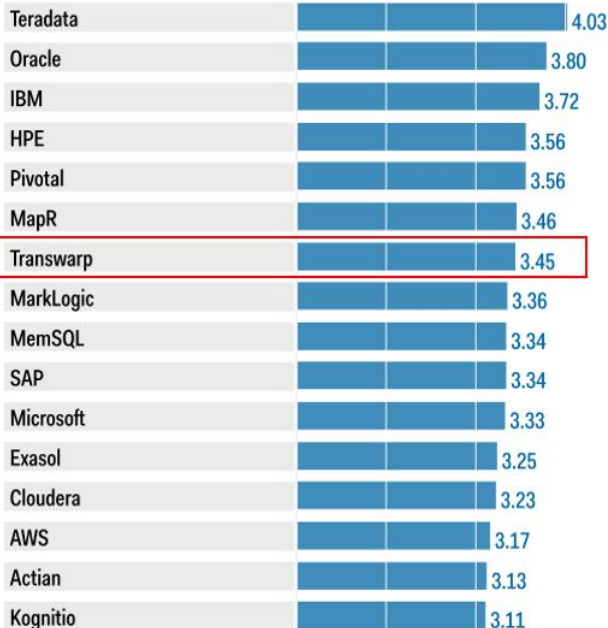
被调查的客户指出星环的产品还有一些功能的缺失，尤其在产品的管理功能方面。被调客户还指出市场中对口人才的稀缺。即使如此，被调客户对星环科技各方面的评价都持平报告中的平均水平。

HPE = Hewlett Packard Enterprise

Source: Gartner (February 2016)

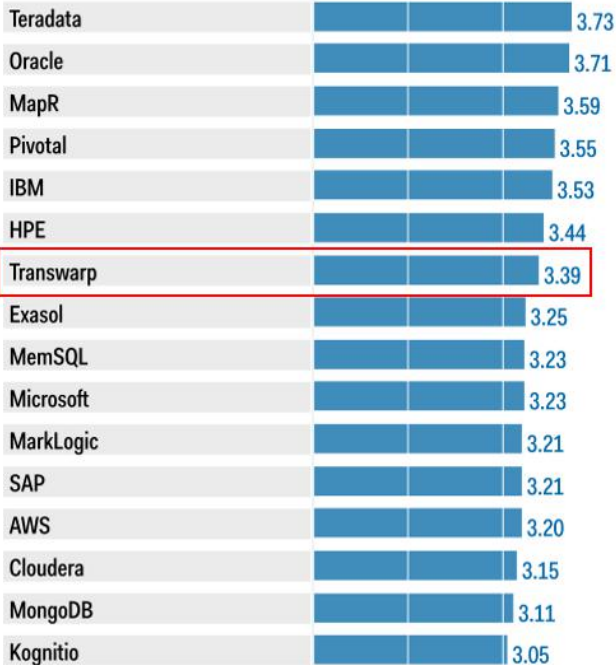
传统数据仓库排名第七位

Product or Service Scores for Traditional Data Warehouse



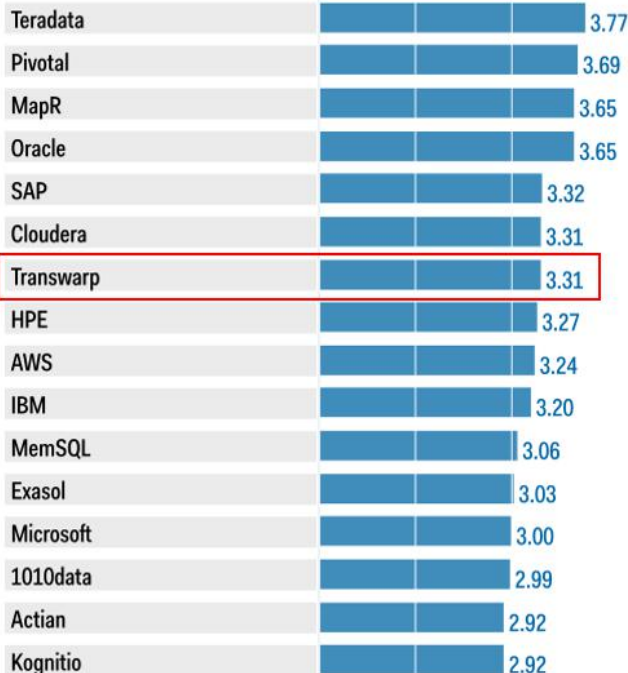
逻辑数据仓库排名第七位

Product or Service Scores for Logical Data Warehouse



机器学习及数据探索类数据仓库排名并列第六位

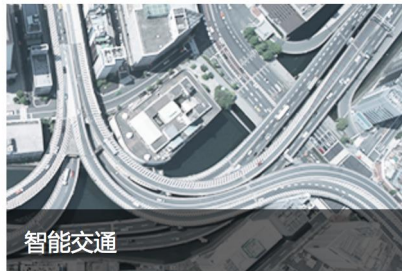
Product or Service Scores for Context-Independent Data Warehouse



我们的合作伙伴 ~ 100家



星环科技典型案例 (落地案例最多)



电信运营商

✓ 移动、联通、电信

交通公安

✓ 山东、辽宁、浙江等

金融证券

✓ 银行+证券

能源

✓ 国网+南网

互联网

✓ 电商+CDN

政府

✓ 工商+税务

物流快递

✓ EMS

广播电视

✓ 华数+卫视

我们的部分客户 >200个

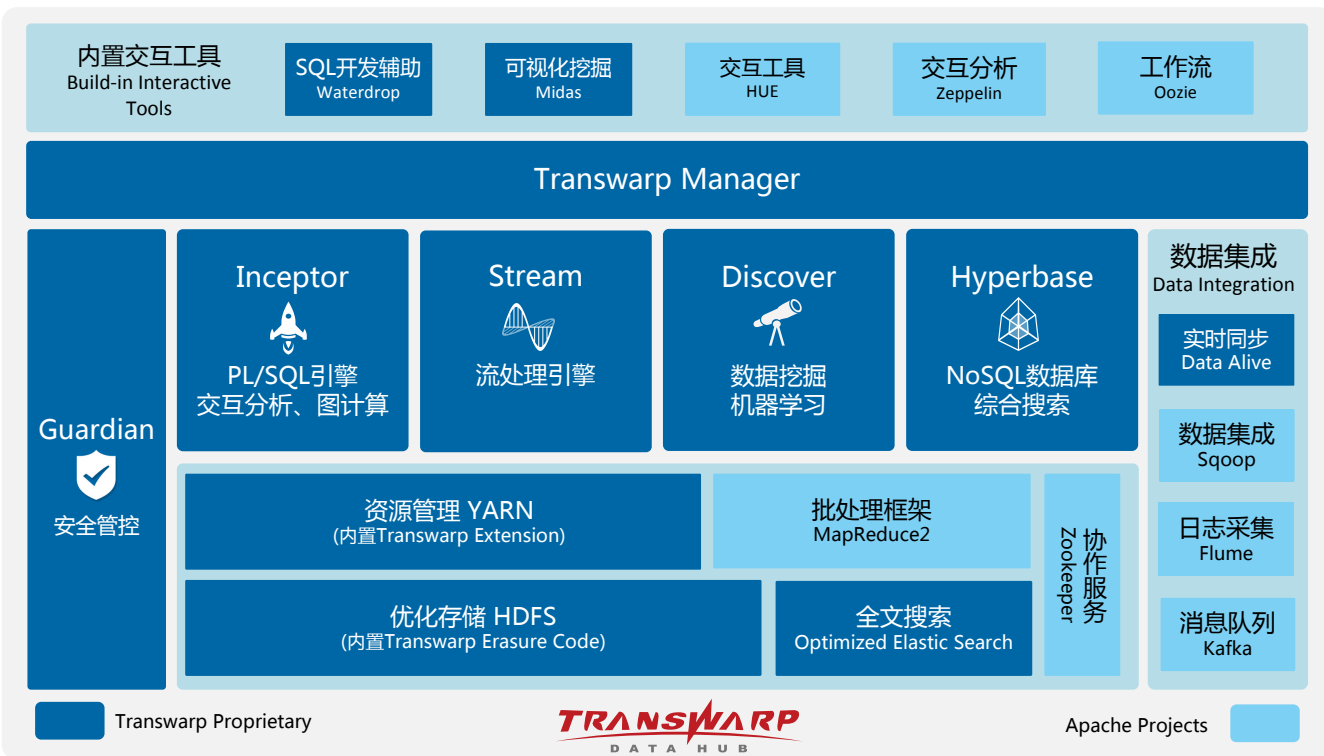


华通银行(筹)



Transwarp Data Hub产品概述

Transwarp Data Hub架构图



最完整的SQL支持

99%的SQL 2003支持, **唯一完整**支持PL/SQL的SQL on Hadoop引擎(98%), **唯一完整**支持ACID分布式事务的SQL引擎; 定位数据库和数据集市市场, 可用于补充或替代Oracle、DB2、MPP等分析用数据库。

高效交互分析----内存/SSD计算

第一个支持SSD的基于Hadoop的高效计算引擎, 可比硬盘快一个数量级; 可用于建立各种数据集市, 对接多种主流报表工具。

最完整的分布式机器学习算法库

支持**最全**(超过50余种)的分布式统计算法和机器学习算法, 同时整合超过5000个R语言算法包。适合风险控制、反欺诈、文本分析、精准营销等应用。

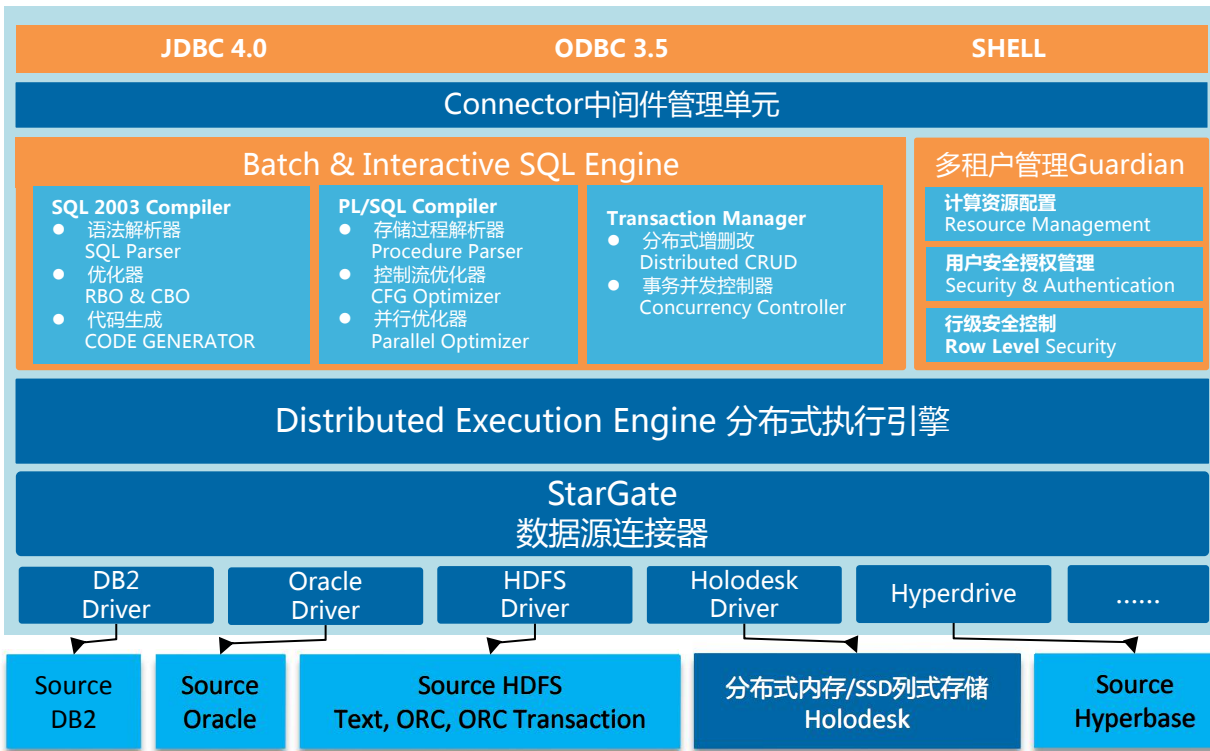
支持最完整SQL和索引的NoSQL数据库

支持SQL2003、索引、全文索引, 支持图数据库和图算法, 支持非结构化数据存储
支持高并发查询

最健壮和功能丰富的流处理框架

支持所有组件的高可用(HA)
支持流式SQL和流式机器学习

交互式分析引擎Inceptor - 完整SQL支持



Apache Spark

基于内存的Map/Reduce计算引擎，即将成为新一代主流计算框架。处理大数据像“光速”一样快，比Hadoop Map/Reduce快10x倍。

Holodesk

跨内存/闪存/磁盘等介质的分布式混合列式存储，常用于缓存数据供Spark高速访问。Holodesk内建内存索引，可提供比开源Spark更高的交互式统计性能；结合使用低成本的内存/SSD混合存储方案，可接近全内存存储的分析性能。

Most complete SQL support

兼容>95% ANSI SQL 2003, HiveQL和>90% PL/SQL语法，支持数据仓库、数据集市等分析系统中常用的复杂分析型语法，方便应用迁移。

SQL引擎

高度优化的高速SQL引擎，可运行在Spark或Map/Reduce上，可高速处理缓存在Holodesk上的列式数据。

丰富的工具支持

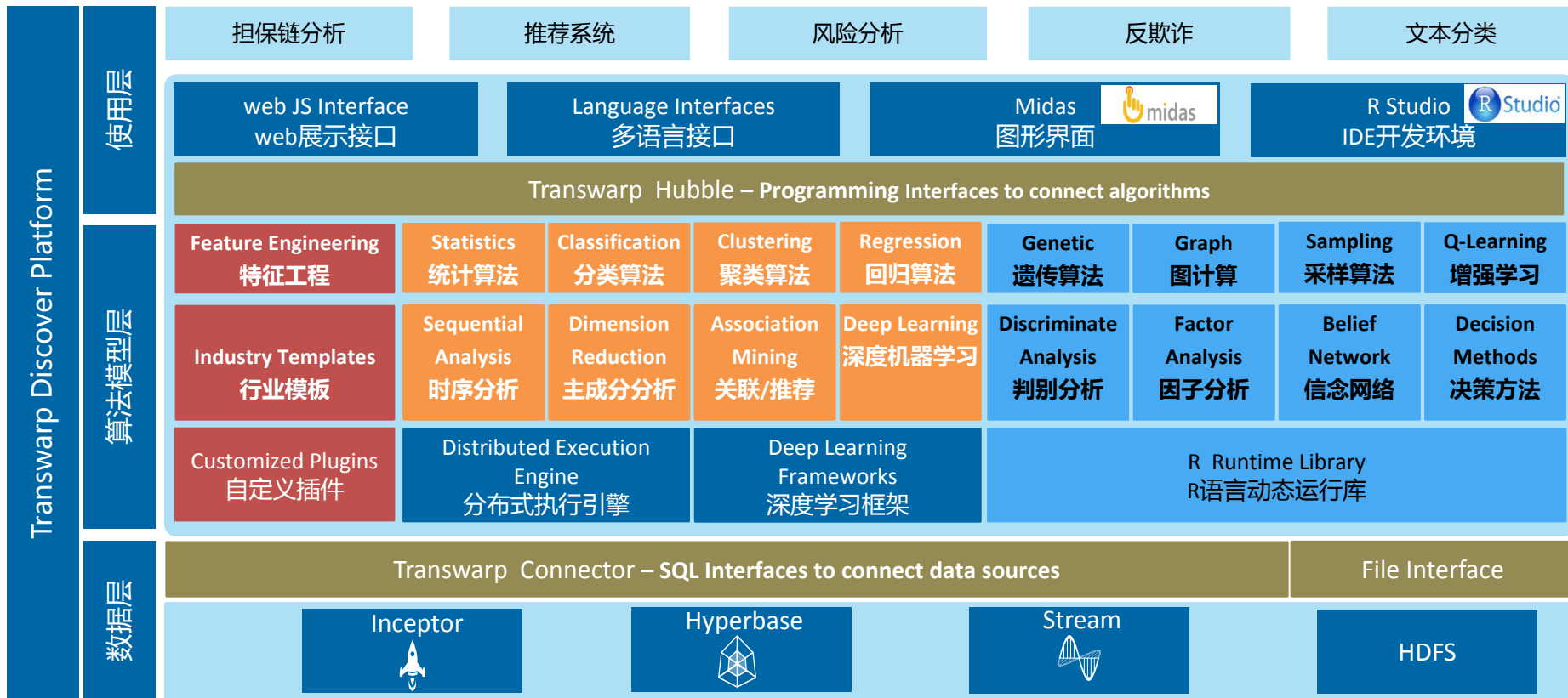
支持主流可视化和BI/挖掘工具，包括Tableau, IBM Cognos, SAP BO, Oracle BI, SAS等。支持Informatica, Pentaho/Kettle等ETL工具。

Data Federation

具备对多种关系数据库和Hadoop数据源进行交叉查询，聚合，以及关联操作等能力

驱动层实现数据访问
计算算子下推以减少数据传输
执行计划仍采用分布式计算

Inceptor 不同版本	功能描述
标准版	SQL2003, 分布式事务
专业版	基础版+内存/SSD数据库功能
企业版	专业版+PL/SQL



- Discover R语言开发 => 算法工程师, 数据科学家
- Midas交互式挖掘 => 业务分析师, 数据科学家

数据预览

预处理

特征工程

模型训练

模型上生产

- R和Midas中可以连接TDH中数据表做数据预览
- 可以对列做tag/feature的管理



- 通过内置的分布式统计算法完成相关的预处理与数据分析
- 支持标准化, 归一化, 正则化, 缺失值填充, 数据分箱等
- 支持通过Inceptor SQL进行数据ETL处理



- 结合业务领域专家知识, 以及相关算法降维, 选择特征指标与维度
- 利用深度学习神经网络算法, 通过升维降低特征工程维度选取难度

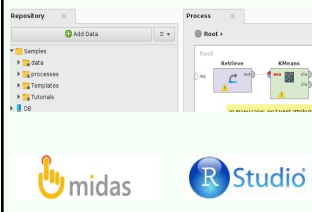
Dimension Reduction
主成分分析

Linear Regression
线性回归

Deep Learning
深度学习

.....

- 用户通过GUI选择算法开发训练模型
- 模型编译成为DAG, 由Hubble组件来调度任务
- 支持单机R算法和分布式算法训练模型



- 导出模型为PMML
- PMML模型可以转换成生产系统的代码, 部署到实际业务中

```
<model key="decision_tree">  
  <parameter key="FeaturesCol" value="features"/>  
  <parameter key="labelCol" value="indexed_label"/>  
  <parameter key="predictionCol" value="prediction"/>  
  <parameter key="probabilityCol" value="probability"/>  
  <parameter key="rawPredictionCol" value="rawPrediction"/>  
</model>  
<tree>  
  <name>decision_tree</name>  
  <root>  
    <name>Internal</name>  
    <count>3.0</count>  
    <count>3.0</count>
```

Discover

Stream

重要组成:

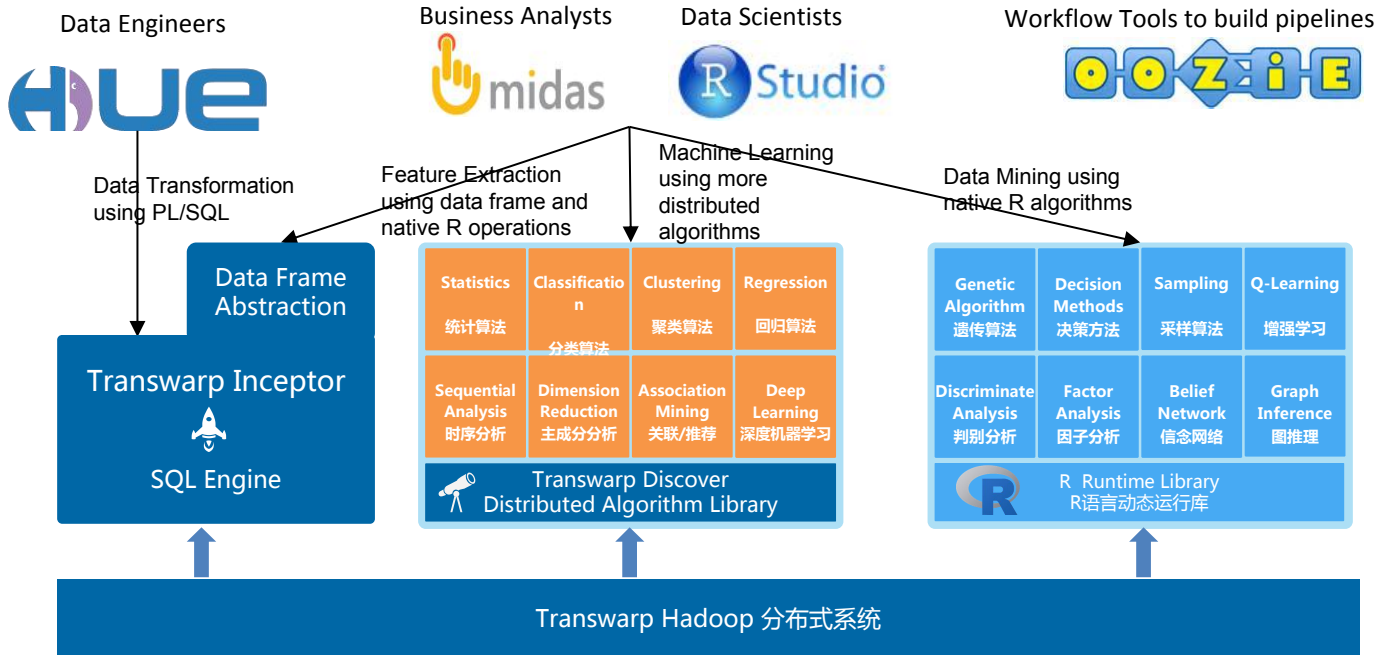
特征抽取

Feature Engineering:
Extract value from dark data

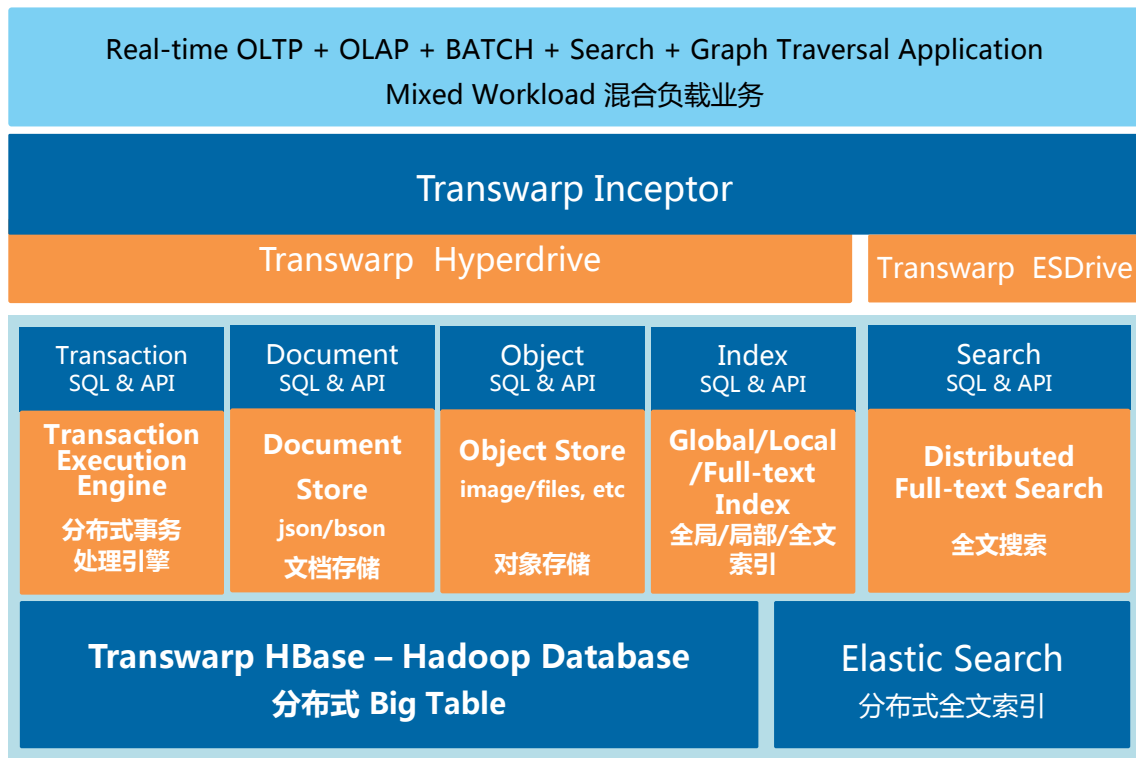
分布式机器学习算法
More Distributed Machine Learning Algorithms

易用的工具

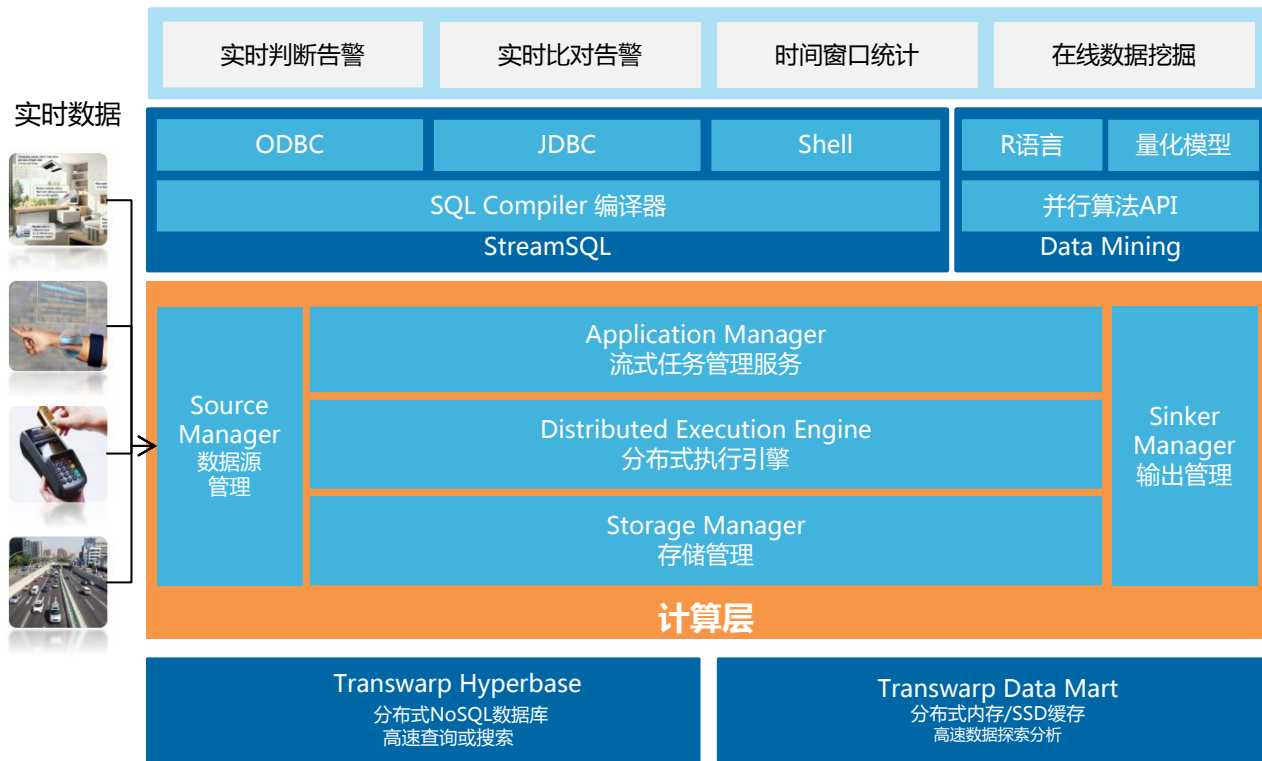
Graphical tools for data scientists, e.g., workbook, notebook; and a workflow tool to build the pipeline



实时NewSQL数据库Hyperbase



Hyperbase 不同版本	功能描述
标准版	<p>OLTP 支持高并发毫秒级数据插入/修改/查询/删除(CRUD)。结合Inceptor SQL引擎, 可以支持通过SQL进行高并发的CRUD。 支持分布式事务处理。 支持常见数据类型, 可更高效的存取数据</p> <p>OLAP 支持多种索引 (global/local/high-dimensional index)。 结合Inceptor, 可进行行列存储转换, 进行秒级高效分析。 支持复杂查询条件, 自动利用索引加速数据检索, 无需指定索引。</p>
专业版	<p>基础版+ 内嵌搜索引擎 实时同步创建索引 实现秒级关键字搜索</p>
企业版	<p>专业版+ 图数据库 支持高并发图遍历和检索 多类型支持 结构化记录 半结构化文档 (JSON/BSON) 非结构化数据 (图片、音频、二进制文档等) 支持混合结构数据的存储、搜索、统计、分析 支持SQL访问关系表和层次化文档</p>



StreamSQL

- 99%的ANSI SQL 2003的支持率
- 强大的优化器提升性能
- 支持按时间切分滑动窗口和滑动步长

多数据源支持

- Socket
- 文件
- Kafka
-

多种输出方式支持

- Hyperbase
- Holodesk
- HDFS
-

企业数据总线支持

- 从一个流读入数据，再将其输入另一个流，如常见的企业数据总线以Kafka为存储中间件

在线数据挖掘支持

- 离线模型，在线预测
- 时间窗口数据进行实时挖掘分析

Stream 不同版本	功能描述
标准版	基本流处理框架，支持流式聚合
专业版	基础版+SQL
企业版	专业版+机器学习

统一权限管控 Transwarp Guardian



Accounting 用户管理	Build-in LDAP HA supported	Existing AD/LDAP with Kerberos 使用既有用户管理 与认证系统	
Authentication 用户认证	Build-in Kerberos HA supported		

Audit 审计	Grant/Revoke Audit	Access Control Audit
	Resource Usage Audit	Policy & Alert

RBAC控制	<p>Data Authorization 数据权限管理</p> <table border="1"> <tr> <td> <p>Inceptor</p> <ul style="list-style-type: none"> 用户安全授权管理 Grant/Revoke via SQL 行/列级安全控制 Row/Column Level </td> <td> <p>Stream</p> <ul style="list-style-type: none"> 流处理作业授权管理 StreamApp/Job Grant/Revoke via SQL 主题级安全控制 Kafka Topic Level Security </td> <td> <p>Discover</p> <ul style="list-style-type: none"> 用户安全授权管理 Authentication with Inceptor and HDFS 数据交互安全控制 Data Transfer & Communication </td> <td> <p>Hyperbase</p> <ul style="list-style-type: none"> 用户安全授权管理 Grant/Revoke via SQL (with Inceptor) 单元格安全控制 Cell Level </td> </tr> <tr> <td colspan="4"> <p>分布式文件系统HDFS</p> <ul style="list-style-type: none"> Dir/File access control Dir/File ACL </td> </tr> </table>	<p>Inceptor</p> <ul style="list-style-type: none"> 用户安全授权管理 Grant/Revoke via SQL 行/列级安全控制 Row/Column Level 	<p>Stream</p> <ul style="list-style-type: none"> 流处理作业授权管理 StreamApp/Job Grant/Revoke via SQL 主题级安全控制 Kafka Topic Level Security 	<p>Discover</p> <ul style="list-style-type: none"> 用户安全授权管理 Authentication with Inceptor and HDFS 数据交互安全控制 Data Transfer & Communication 	<p>Hyperbase</p> <ul style="list-style-type: none"> 用户安全授权管理 Grant/Revoke via SQL (with Inceptor) 单元格安全控制 Cell Level 	<p>分布式文件系统HDFS</p> <ul style="list-style-type: none"> Dir/File access control Dir/File ACL 				<p>Resource Authorization 资源管理控制</p> <p>计算资源</p> <table border="1"> <tr> <td>Inceptor Pool Control</td> <td>SLA Control</td> </tr> <tr> <td colspan="2">计算资源管理调度YARN Queue Use/Admin Control</td> </tr> </table> <p>存储资源</p> <table border="1"> <tr> <td colspan="2">数据库/仓库</td> </tr> <tr> <td>Static Quota</td> <td>Temp Quota</td> </tr> <tr> <td colspan="2">文件系统 Usage Quota Control</td> </tr> </table>	Inceptor Pool Control	SLA Control	计算资源管理调度YARN Queue Use/Admin Control		数据库/仓库		Static Quota	Temp Quota	文件系统 Usage Quota Control	
	<p>Inceptor</p> <ul style="list-style-type: none"> 用户安全授权管理 Grant/Revoke via SQL 行/列级安全控制 Row/Column Level 	<p>Stream</p> <ul style="list-style-type: none"> 流处理作业授权管理 StreamApp/Job Grant/Revoke via SQL 主题级安全控制 Kafka Topic Level Security 	<p>Discover</p> <ul style="list-style-type: none"> 用户安全授权管理 Authentication with Inceptor and HDFS 数据交互安全控制 Data Transfer & Communication 	<p>Hyperbase</p> <ul style="list-style-type: none"> 用户安全授权管理 Grant/Revoke via SQL (with Inceptor) 单元格安全控制 Cell Level 																
<p>分布式文件系统HDFS</p> <ul style="list-style-type: none"> Dir/File access control Dir/File ACL 																				
Inceptor Pool Control	SLA Control																			
计算资源管理调度YARN Queue Use/Admin Control																				
数据库/仓库																				
Static Quota	Temp Quota																			
文件系统 Usage Quota Control																				

- 星环科技技术论坛：

<http://support.transwarp.cn/>

- 星环科技产品手册：

<https://docs.transwarp.cn/>



The image features a stylized red logo for 'TRANSWARP' centered over a background of a planet's horizon. The planet's surface is visible as a thin blue and green line, with a bright yellow sun rising behind it, creating a lens flare effect. The logo consists of the word 'TRANSWARP' in a bold, italicized, sans-serif font. A red underline is positioned below the letters 'R', 'A', 'R', and 'P'. The letter 'W' is uniquely designed with a sharp, upward-pointing triangular shape extending from its top, which overlaps the underline. The entire scene is set against a dark, black background.

TRANSWARP