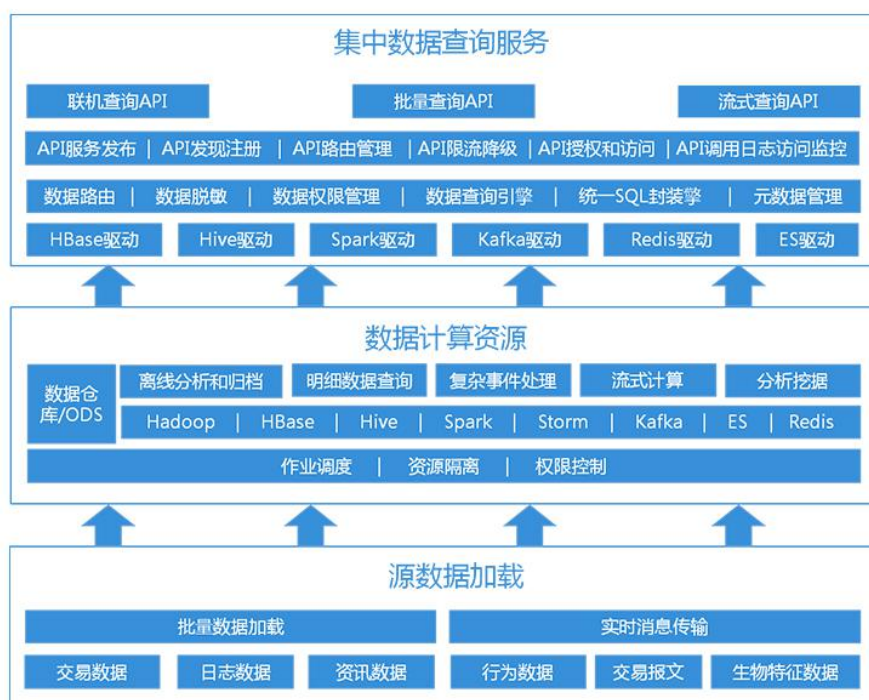


我们的产品

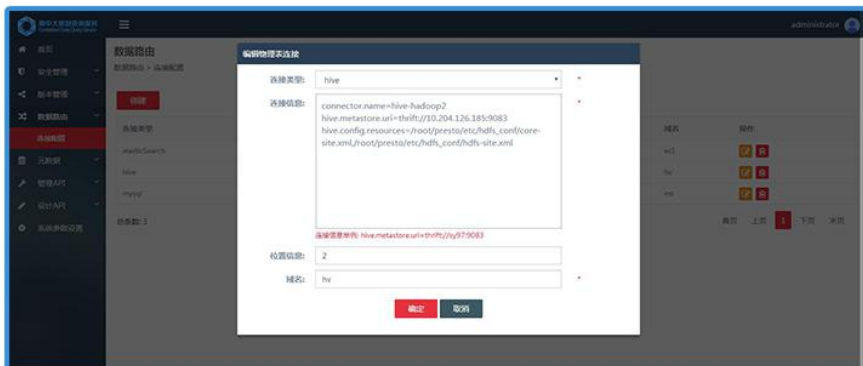
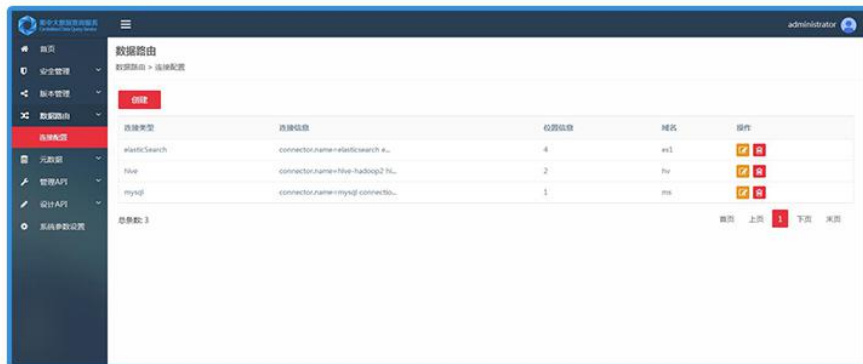
- 全面支持ANSI SQL语法
- 跨库统一查询接口
- GB到PB级交互式查询
- 跳过大量的ETL工作
- 至少10倍于Hive的查询效率
- 支持高访问量和高并发
- 准实时在线查询
- 接口灰度变更发布
- 接口流量控制和状态采集监控
- 接口授权和访问控制



我们提供

大数据平台统一标准SQL查询

针对多种大数据和关系型数据库组件支持标准ANSI SQL语法查询，支持的查询对象包括Hive, HBase, Elasticsearch, Spark, Kafka, Redis和关系型数据库集群。



提供多种接口查询方式

针对查询和开发的不同需求，提供多种查询途径，包括JavaSDK开发包的API调用，RESTful接口查询和JDBC直连。

跨库多数据源结果合并，即时或分批次查询

一条查询可以合并多个数据源数据，通过分布式查询引擎，满足交互式的分析查询，准实时返回查询结果。

查询接口多版本控制

支持多版本同时在线，支持灰度发布接口版本。版本切换回退不需要重启平台相关系统，业务不感知。

数据查询的访问日志调用监控

对流经数据查询服务的访问调用，进行日志收集和存储和查询；提供API调用成功率、失败率、吞吐量、饱和度、平均相应时长等指标的监控呈现（柱状图/饼图/曲线图等）。针对监控指标配置预警阈值，如超过阈值，发送预警短信或邮件。

元数据和数据查询路由配置

通过数据路由配置，数据源可以进行动态加载，新增连接时能自动生成该连接下的逻辑表结构，当数据源进行迁移时，只需修改对应的连接配置就可瞬间切换数据源。

The image displays two screenshots of the '元数据' (Metadata) configuration interface in the Centralized Data Query Service. The interface is divided into a left sidebar with navigation options and a main content area for configuring a specific data source.

Top Screenshot: bank data source configuration

逻辑表名称	类型	注释	逻辑表字段名
bank_id	bigint		bank_id
count	double		count
opt_type	varchar		opt_type
remark	varchar		remark
src_type	varchar		src_type
id	bigint		id
opt_datetime	timestamp		opt_datetime
card_id	bigint		card_id

Bottom Screenshot: weibo data source configuration

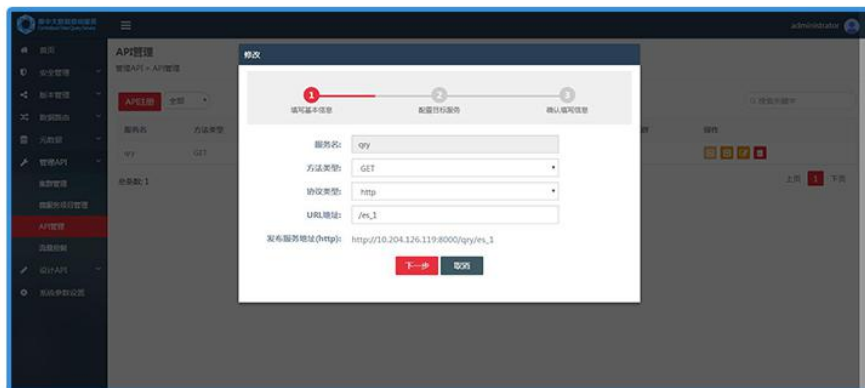
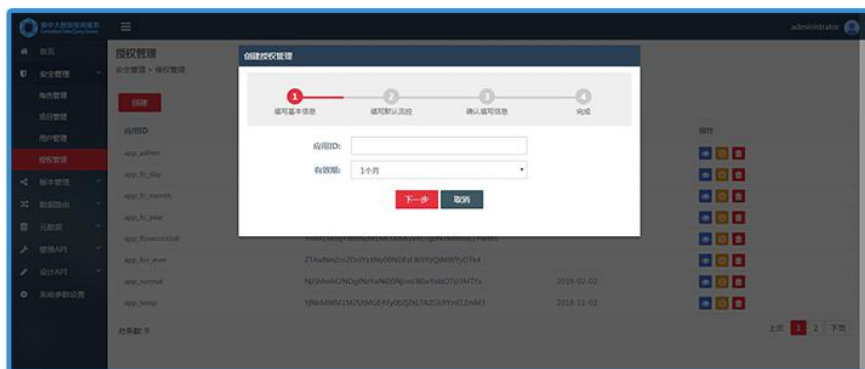
逻辑表名称	类型	注释	逻辑表字段名
permission_denied	varchar		permission_denied
state	varchar		state
retweeted_uid	varchar		retweeted_uid
text	varchar		text
image	varchar		image
deleted_last_seen	varchar		deleted_last_seen
timestamp	varchar		timestamp
uid	varchar		uid

数据查询结果的动态脱敏

动态脱敏支持对某些敏感信息通过脱敏规则进行数据变形，系统预置脱敏数据域，支持自定义脱敏函数和脱敏策略的配置，实现敏感隐私数据的可靠保护。

统一数据查询权限控制

对查询服务，接口权限，查询数据访问权限进行统一的权限控制。包括凭证创建、凭证续期、根据凭证进行服务限流。凭证管理粒度可细化到用户、用户下面的项目、单个API三个维度（默认只支持到用户，即该用户创建的凭证可以访问其管辖项目下所有的API）。凭证可以是永久凭证或有期限的凭证，凭证到期后可续期。



技术场景

应用场景

为了应对大数据时代数据爆炸式的增长，以及组织内部越来越多的数据分析需求，集中大数据查询服务（CDQS）已经开始崭露头角，并逐步被更多企业所重视和实施。那么大家自然会想到一个问题：CDQS与传统的数据仓库、ODS有何不同？我们理解，传统数据仓库和ODS更侧重于数据存储，靠ETL过程将数据搜集、处理并集中存储，这种架构存在两个缺陷：一是传统技术擅长处理结构化数据，而无法满足组织对非结构化数据处理的需求；二是时效性低，无法获取最新的数据，导致分析结果与应用场景分离。

CDQS更侧重于互联互通，这是与传统数据仓库最根本的区别。CDQS是更广义上的概念，它具有如下特征：

- 数据是否持久化决定了CDQS的技术框架和效率，数据持久化的CDQS就类似ODS，但效率会降低。
- 无论数据是否持久化，都需要一个逻辑上的通用数据模型，实现业务语义的描述。元数据格外重要，方便数据访问。
- 数据类型越来越丰富，行为信息、位置信息、生物特征等非结构化在信息共享中的占比越来越多，新的数据处理技术Hadoop、NoSQL数据库被广泛使用；

- CDQS实时的数据传输技术逐步替代批量数据加载技术，大大提高数据的时效性和访问效率。
- CDQS成为数据治理的监控站和控制点，数据质量、数据安全、访问控制、数据脱敏等成为重点，但数据生命周期的控制已经不需要从技术角度考虑，因为新技术在存储、处理能力方面都得到了提高。

综上，以下几种场景更需要使用集中数据查询服务：

- 急需提升PB级数据查询返回速度
- 需要在多个数据源查询数据并对结果进行合并
- 需要对数据接口访问权限控制
- 数据查询权限控制
- 快速构建基于大数据组件的数据市场，提供外部访问统一入口
- 降低大数据组件的使用成本

产品技术

分布式的数据存储和计算

融合内外部数据、结构化与非结构化数据，提供分布式的存储能力和分布式的计算能力，满足数据处理能力的需求

实时的数据连接和传输

借助数据连接技术与T+0的实时数据传输技术，使数据可以随时随地按需获取和使用，提高数据时效性

完整的业务数据视图

搭建通用的数据模型和元数据管理，使企业内全部数据资产一目了然，为用户提供完整、清晰的数据地图

全面的接口和数据安全

通过接口的版本控制、鉴权、单位时间查询次数、流量限制，数据查询的访问日志调用监控和查询数据结果的动态脱敏等，全面保证接口和数据的安全

为什么选择我们



领先的设计理念、 开放的技术路线

产品基于先进的分布式架构设计，保证了系统的高可用、高并发，易与第三方系统集成的特性。紧跟最前沿的开源技术保证系统组件的开放性和包容性。



聚焦的产品形态

着眼于银行、证券、保险、政府、企业等领域客户迫切解决的问题，针对各行业大数据应用的特点，提供全方位的集中数据查询服务能力支撑，屏蔽底层组件，提升查询效率，控制数据访问权限和全方位的数据治理。



完全的接口支持与 开放

系统提供多种标准查询接口，基于内存的可扩展式并行计算作业，方便不同角色的岗位人员开发、测试和对接发布版本。有效提升团队工作效率，降低组件差异带来的学习和变更成本。



专家级的团队

产品设计和系统开发人员均来自于银行、保险、证券、云计算和通讯、互联网企业的大数据领域资深专家级人才。对分布式系统架构、大型企业系统设计、云计算、大数据和运维技术，有多年的从业经验和独到见解，充分了解行业客户的真实需求，研发出行业用户真正需要的产品。



服务可用性 $\geq 99.99\%$



7*24热线支持



平均故障恢复时间 ≤ 60 分钟



故障分析报告时限24小时