

腾讯云数据工坊

产品简介

产品文档



腾讯云

【版权声明】

©2013-2017 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

【商标声明】



及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。

【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或模式的承诺或保证。

文档目录

文档声明.....	2
产品简介.....	4
什么是数据工坊	4
应用场景	6
功能模块	10
术语解析	11
用户权限	13
服务等级协议.....	14

产品简介

什么是数据工坊

什么是数据工坊

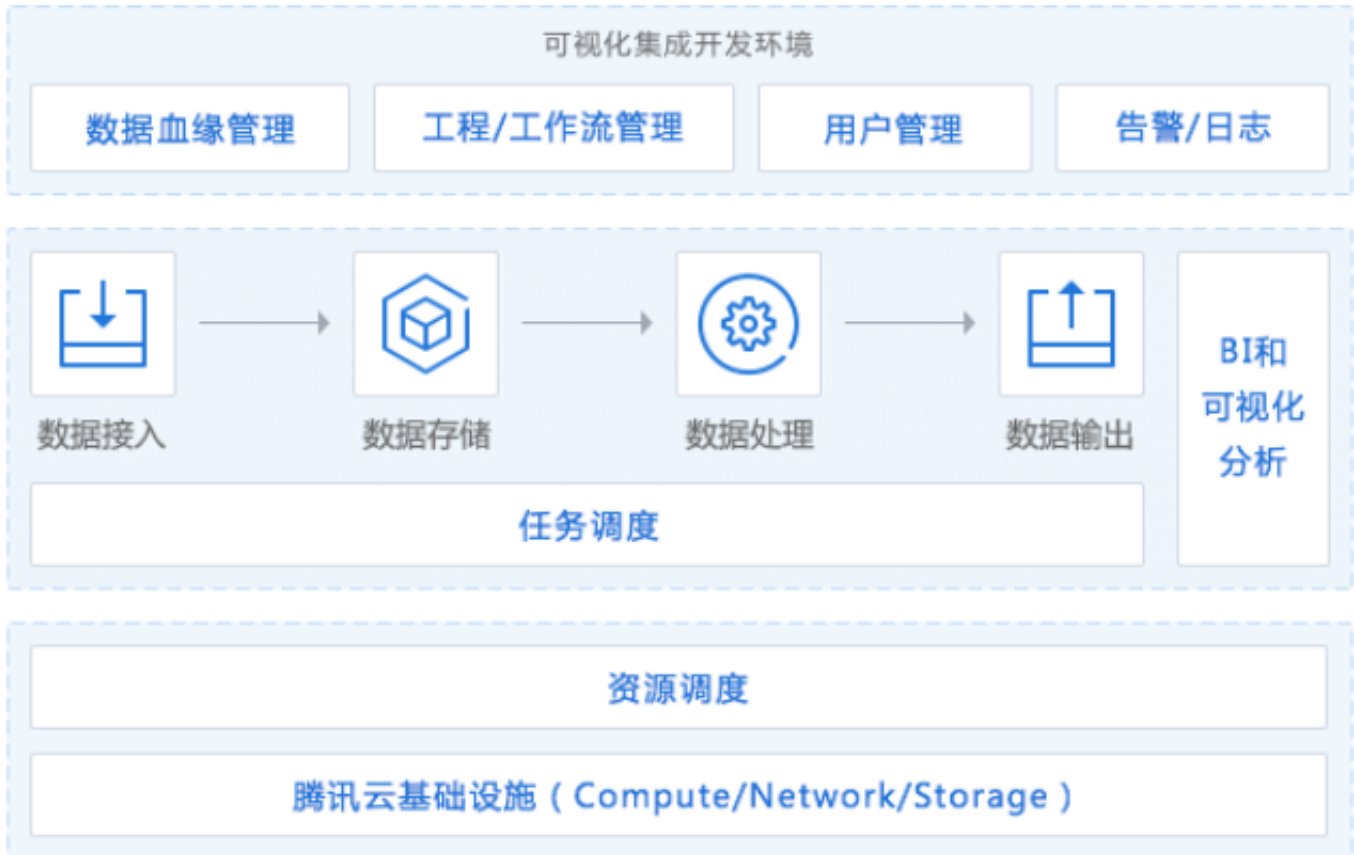
腾讯云大数据数据工坊（Tencent Data Factory 简称：TDF）依托腾讯海量数据处理经验和方法论积累，为用户提供轻量级的大数开发应用平台。通过数据集成开发环境、数据开发、数据管理、运维中心、工程管理等组件，以公有云服务的形式，帮助用户快速完成大数据开发、管理、运维的可视化操作，深入挖掘数据价值，灵活应对业务需求。

TDF 整合了腾讯大数据产品能力和公有云方案优势，帮助用户快速具备大数据开发和展示能力，将精力集中在数据开发本身，而无需考虑底层物理机器和运维细节。

此产品文档从基本概念入手，通过示例和功能说明，详细介绍腾讯云大数据 TDF 的使用方法，适合于需要了解 TDF 功能和具体操作方法的用户使用。如果需要快速对整个 TDF 的操作流程有大致了解，可在了解基本概念后，重点阅读【快速开始】部分。

TDF 架构图

TDF 架构图如下所示：



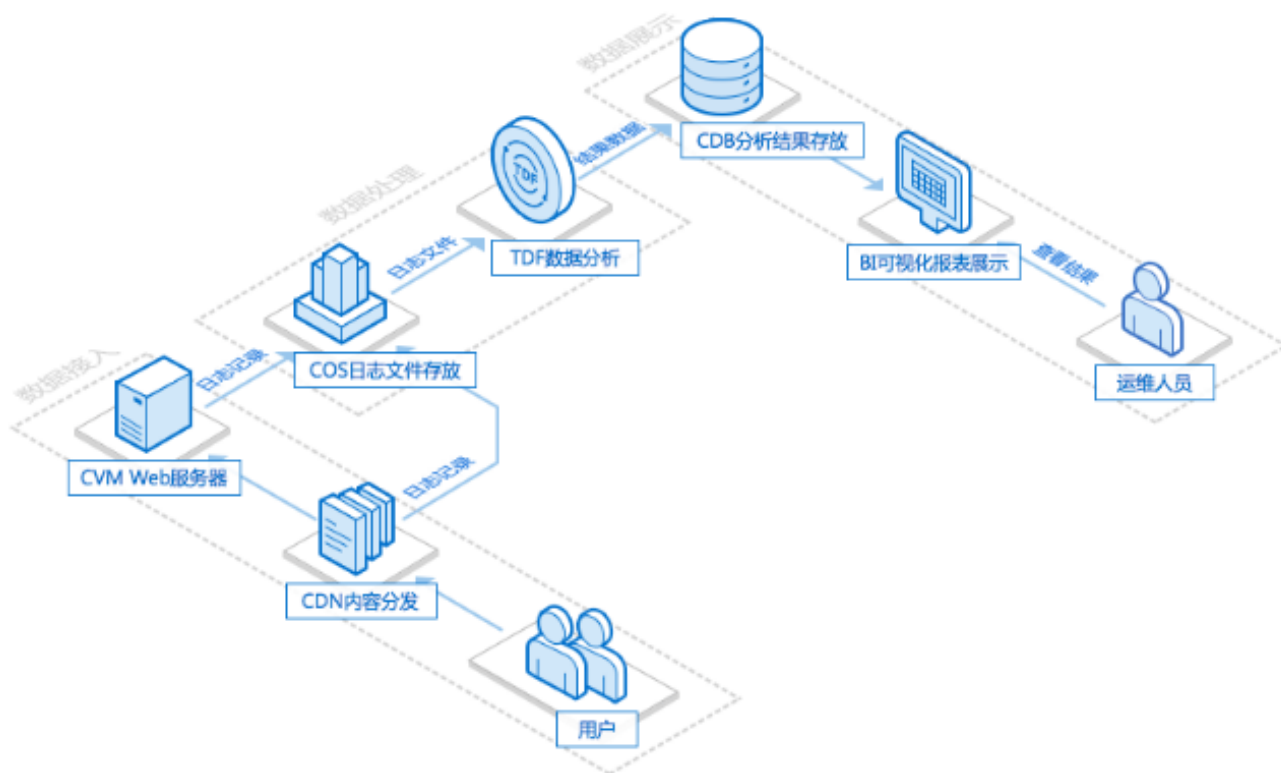
应用场景

Web 应用日志分析

无论是PC端还是移动端网页浏览，都会产生大量的日志信息；这些日志可以为界面交互设计改进和业务逻辑优化提供重要的信息。使用TDF可以为大规模日志数据处理提供强大灵活的引擎，并做到按需开通，在计算完成后，即刻释放计算资源，满足海量日志计算对灵活性的要求。

在下图中，数据接入模块将最终用户访问记录，从 CDN 节点以及 Web 应用服务器以文件形式存放到 COS 节点。随后将 COS 节点上的日志文件导入 TDF 中，并进行查询分析。TDF处理结果可周期性存放至 CDB，并借助BI组件进行可视化分析。

Web 应用日志分析架构如下图所示：

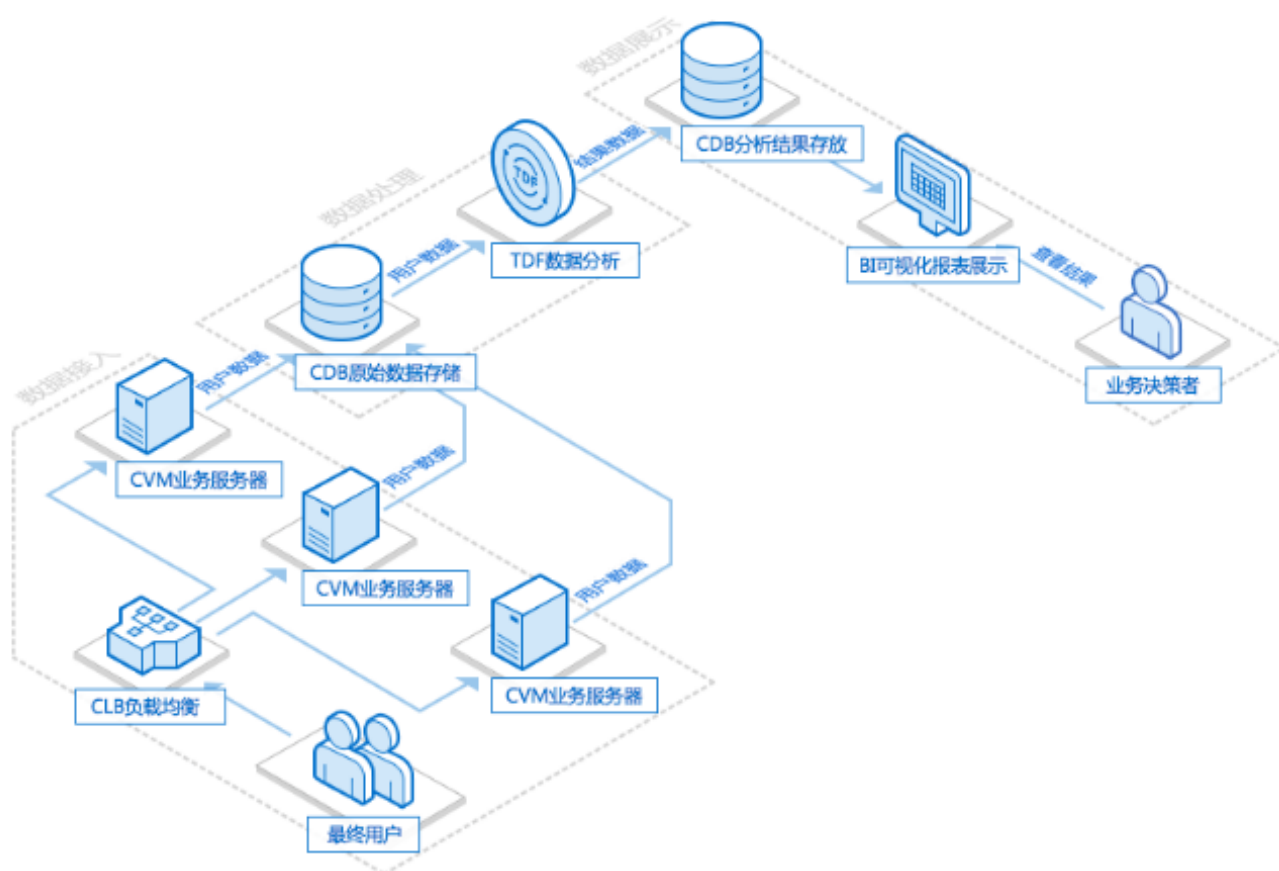


游戏用户数据分析

从千万级用户数据中往往可以提取出宝贵信息以指导业务决策，包括游戏数值设定、转化率优化方案等等。TDF 可以为游戏开发商提供 HIVE 等成熟技术框架用于用户数据分析服务，实现海量用户数据分析。

在下图数据接入部分，最终用户在访问业务服务器过程中，产生大量的行为数据和自身属性数据，这些数据由各业务服务器产生并收集。这些原始数据可存放在 CDB 数据库，随后周期性提交给 TDF，进行数据分析。最终，TDF 处理结果导出至 CDB，并由 BI 组件进行可视化分析，业务决策者可由此分析判断进行决策。

游戏用户数据分析架构图如下所示：



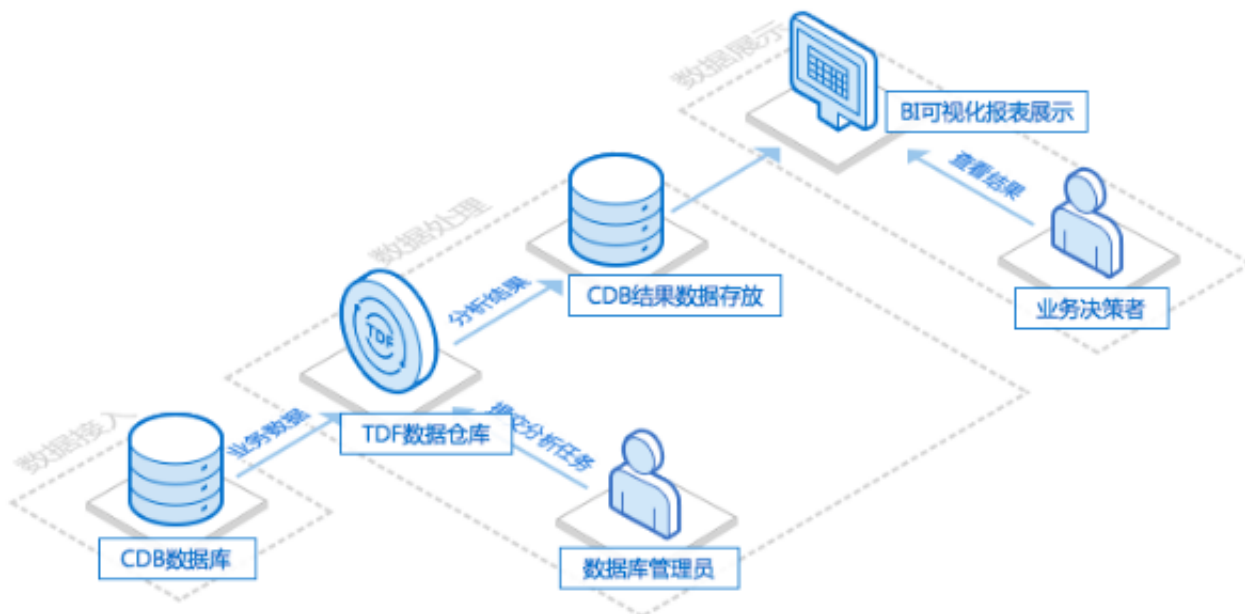
PB 级数据仓库应用

传统 OLTP 数据库在超大规模数据分析上往往捉襟见肘，TDF 可以为用户提供高达 P 级容量的强大数据仓库查询和分析能力，帮助用户以极低成本实现海量数据 OLAP 处理过程。

如下图，在数据接入部分，CDB 数据库可直接运行 OLTP 业务，产生的数据将作为原始数据待分析，在数据处理部分，数据库管理员可以直接通过 TDF 管理控制台提交 OLAP 数据仓库应用作业，作业将被 TDF 调度并计算，最后 TDF 计算结果可被导出至 CDB 存放，并由 BI

组件进行可视化分析，业务决策者可由此分析判断进行决策。

PB 级数据仓库应用架构如下图所示：



金融领域数据仓库分析

无论是银行、保险、证券等传统金融机构，还是互联网金融企业，都需要灵活快速对已有数据进行分析，发现业务规律，实现精细化管理。

传统数据仓库分析往往需要使用复杂的数据库应用软件和商业智能套件。IT技术团队需要耗费大量的时间完成技术设施搭建和运行维护工作，并保障系统的可用性、安全性和稳定性。

采用腾讯云 TDF 解决方案，可通过公有云服务方式，对用户已有海量业务数据和客户信息进行分析处理，在包括但不限于如下具体业务领域实现快速应用交付：

- 个人金融产品精准营销：通过 TDF 对个人信用卡消费时间、频次、金额、类别等详细交易信息进行分析，确定推广目标人群，分析其年龄阶段、消费偏好、收入状况、支付能力等属性，并对消费行为建模分析,还可借助TDF 内置机器学习算法，预测消费行为，并有针对性进行金融产品精准营销。

- 个人或小微企业贷款风控：根据海量交易数据，在 TDF 中通过机器学习算法进行风险计算，提升风险评估的精确性和处理速度。

功能模块

TDF 以腾讯云管理控制台为核心界面，同时提供丰富 API（待发布），供用户灵活调用。其中模块划分如下：

数据开发

以工作流开发为核心，完成数据接入、数据计算、数据导出任务的创建和编辑过程。

数据管理

从表总览和表管理两个方面，对和用户相关的表进行管理，配置不同用户对库表的操作权限。

运维中心

从任务运维管理角度，对数据库开发 IDE 创建后正在运行的工作流任务进行管理，包括监控告警管理。

工程管理

配置工程基本属性，包括数据源和计算引擎，以及对工程相关成员进行管理。

数据可视化

提供用于直接呈现计算结果的数据可视化功能，通过拖拽和自定义报表，实现交互式数据可视化（通过独立产品实现，待发布）。

术语解析

本部分是为了更有效的使用 TDF 而介绍了一些所使用到的部分重要术语，这些 TDF 解析采用了行业通用术语且兼顾了腾讯云已有规范。

用户

管理、使用TDF的逻辑实体，单个用户对应于腾讯云账号。

角色

TDF中不同权限集对应的逻辑实体，角色的权限为系统固定配置；单个用户可以拥有多种角色；当前系统配置角色包括：主帐号拥有者（协作者）、工程管理员、开发工程师、运维工程师；具体权限参见后文。

组织

主帐号拥有者所处的全局工作空间，单个组织对应唯一的主帐号拥有者，以及一个或多个协作者；用户的所有操作均在该组织内进行。

工程

TDF的工作空间，单个工程内，可以有多个工作流；所有的工程均处于当前组织内。

工作流

任务和其相互关系所组成的集合，可以以有向无环图（DAG）的形式或独立任务的形式存在；工作流可根据调度设置自动或手动运行，运行后即生成对应工作流实例。

工作流实例

工作流在运行时由系统调度生成的具体运行实例；工作流执行完毕代表其中所有任务均被执行完毕；工作流可根据调度设置自动或手动运行，如果工作流内有多个任务，则在保持上下游依赖关系的前提下，按照调度设置

运行。

任务

可被调度的最小节点；任务当前可以分为如下类型：数据计算（HQL），数据接入（腾讯云 CDB MySQL -> Hive），数据导出（Hive -> 腾讯云 CDB MySQL）。如果设置工作流调度，则任务按照工作流调度信息和上下游依赖关系；任务类型根据开发进度将持续丰富。

数据源

工程所对应的数据源逻辑实体，该实体通常关联至实际的数据来源，例如某个腾讯云 CDB MySQL实例中的具体表，单个工程可对应多个数据源；未来将支持本地上传和其他数据源等。

表

TDF处理数据的标准单元，数据表需提前创建，并指定格式，具体数据通常由数据源导入；表同用户、工程关联；除非申请授权，否则表仅为表创建者可见。

计算引擎

用于大数据处理的、按照CU（Compute Unit）度量的计算资源；典型的单个CU一般对应为4GB RAM、1vCPU的计算能力；计算引擎同时作为计费依据；在初期版本中，用户暂不可创建计算引擎。

用户权限

用户角色和权限

如下为系统默认角色和对应权限，为客户在复杂工程中实现权限划分提供灵活的设置。

其中主帐号拥有者（及协作者）为全局概念，其余角色同特定工程相关。

【注意】：为了上线用户第一次使用方便，默认值为：当前用户（主帐号拥有者），一旦创建工程（默认赋予工程管理员权限），将同时具备开发、运维角色。用户可根据实际情况修改。

	主帐号拥有者 (及协作者)	工程管理员	开发工程师	运维工程师
添加全局用户， 以及指定协作者	√	X	X	X
创建工程	√	X	X	X
指定计算资源	√	X	X	X
工程表指定/修改 所有者以及权限 审批	X	√(对工程内所有 表)	√(对所拥有的表)	X
工程表结构修改	X	X	√(对个人创建表)	X
工程表读写	X	√	√	√
工程数据源管理	X	X	√	X
工程用户管理	X	√	X	X
任务/工作流查看	X	√	√	√
任务/工作流编辑 /删除	X	X	√	X
任务工作流运维	X	X	√(可查看和运维自 己作为责任人的任 务工作流)	√(即可进入运维界面 查看和运维相关工程全 部任务工作流)

服务等级协议

内测期暂不提供服务等级协议说明。