

腾讯云数据工坊

常见问题

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
常见问题.....	4
基本概念	4
开通和计费.....	6
数据接入和导出	7
数据计算	8
权限.....	9

常见问题

基本概念

什么是TDF?

数据工坊 (TDF , Tencent Data

Factory) 是基于腾讯云数智大数据套件的轻量级云上大数据产品，为用户提供 Hive 等常见大数据计算框架，适合于需要动态灵活获取大数据计算能力的客户。

TDF 能帮我完成什么工作？

当前，TDF 支持 Hive

计算框架，因此TDF特别适用于大数据集的批处理作业，例如网络日志分析，用户行为分析等等。

TDF 不适用于处理哪些工作？

TDF不适用如下场景：

1. 实时或超低延迟的查询。由于 Hive 本身不具备处理实时查询的功能，不建议使用 TDF 处理对延迟要求较高的数据分析应用，未来我们将发布对应云端解决方案覆盖该场景。
2. 大量数据存放本地且不易迁移。TDF 为公有云化服务，即用户需将数据上传至TDF数据表进行处理。TDF 未来将提供数据接口，支持本地上传，但对于本地有大量数据场景，建议通过 TBDS 大数据套件私有部署方式支持。
3. 复杂ELT操作。当前 TDF 所支持的 Hive 框架可通过类 SQL的 HiveQL语言实现数据操作，主要用于结构化数据的查询。根据 Hive 本身特点，并不擅长处理数据的改写和添加，或非结构化至结构化数据的处理。未来我们将发布对应云端解决方案覆盖该场景。

和线下方案相比，云上大数据产品 TDF 有什么好处？

云上方案相对于线下部署，有如下特点：

1. 成本低：用户无需承担硬件和软件一次性采购和部署成本，而是按需付费。
2. 动态扩缩容：根据实际需要，动态订购计算引擎。
3. 免运维：计算框架所依赖的计算节点和存储资源对用户透明，用户仅关心具体任务或 workflow 即可。
4. 易于对接云上数据源：如果用户在云上数据库如腾讯云 CDB 已经存放大量数据，这些数据可通过数据中心内部网络直接接入至 DF（例如 Hive 表）进行计算，快速高效。

TDF 和其他大数据产品组件的区别和联系？

TDF 是轻量级线上大数据服务，其技术原理和核心架构来源于 TBDS，即线下部署的大数据套件。相对于 TBDS，TDF 更适合于以下几类场景：

1. 数据已经存放于云上。
2. 对计算能力动态扩缩容的应用场景。
3. 可以用 Hive 计算框架完成。

如果用户需要进行私有部署，且使用 Spark 等其他计算框架，则可考虑使用 TBDS 大数据套件。

开通和计费

如何开通？

当前 TDF 正处于有限测试阶段，用户需提交开通申请，我们会根据当前测试环境预留和用户申请数量评估开通。一般提交申请后，5个工作日后将有结果。成功开通后我们将主动联系用户，如果没有开通恕不另行通知。

如何计费？

当前有限测试阶段无需任何费用。

正式上线后，会根据用户所购买的计算引擎*使用时长计费。计算引擎即以 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。

另，如果涉及数据的导入导出，用户有可能需根据云端数据传输或数据库计费策略对传入或传出至TDF的数据计费，详情请阅对应产品计费说明。

数据接入和导出

我能够接入哪些类型的数据？支持什么数据源？

当前 TDF 支持在 Hive 框架下处理数据。因此任何可以写入为 Hive 表的结构化数据均可处理。目前我们提供对腾讯云数据库 CDB MySQL 数据源的直接支持，未来将增加其他数据源。

计算完成后的数据存放在哪里？

数据计算后，可在任务节点中选择离线导出节点。数据即可按照配置中所提供的数据库目标地址（当前支持腾讯云 CDB MySQL）进行存放。

何时支持其他数据源？

我们计划在 17 年第一季度增加腾讯云面向对象存储 COS 等存储服务的支持。此外，我们还计划发布独立的数据集成产品，以方便用户在不同的数据源之间转移数据。具体时间以实际发布为准，请留意相关发布信息。

数据计算

当前TDF支持什么计算框架？

当前 TDF 支持 Hive SQL。

何时支持其他计算框架，例如流式计算？

对于大数据处理的其他常见框架，例如流式计算，我们将尽快提供支持，发布日期待定。

计算的耗时如何估算？

计算时间同任务占用的计算引擎，以及所使用的计算框架执行复杂度相关。按照当前 HiveSQL 的执行效率，通常处理 GB 级的查询，会有数分钟的耗时。

什么是 workflow 调度？

workflow 调度为需要周期性执行，或任务之间存在先后关联的复杂业务形态提供了直观、图形化的组织形式。在配置 workflow 调度后，workflow 内的所有任务将按照 workflow 调度信息和先后依赖关系执行。

权限

为什么我操作系统提示没有权限？

为了提供灵活的权限管理，在系统设置，我们默认提供了不同角色的权限设定，单个用户可以拥有多种角色。

当前系统配置角色包括：主账号拥有者（协作者）、工程管理员、数据开发人员、运维人员。如果系统提示操作没有权限，可检查用户是否具备对应角色，并联系工程管理员修改调整。

为什么我看不到所需的表或 workflow？

整个 TDF 的运行逻辑是以工程为应用逻辑分隔，即推荐用户针对不同的业务，配置不同的工程。用户默认将无法查看和自己无关的工程表项和 workflow/任务。因此，首先需要检查自己是否同工程相关（即在工程中获得某种角色），如果没有，需联系工程管理员添加至工程。

用户也可自行申请对某个表的权限，申请后，表的拥有者可以拒绝或允许该申请。