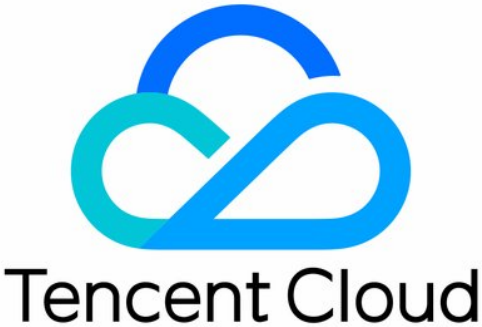
# Tencent Sparkling Data Warehouse Suite

# Product Introduction

## Product Documentation




©2013-2019 Tencent Cloud. All rights reserved.

Copyright Notice

©2013-2019 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

Trademark Notice

Tencent Cloud

All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

# Contents

Product Introduction
  Product Overview
  Product Strengths
  Usage Scenarios

©2013-2019 Tencent Cloud. All rights reserved.

# Product Introduction

# Product Overview

Last updated：2019-05-30 17:20:59

## Overview

Tencent Sparkling Data Warehouse Suite (Sparkling) is based on industry-leading Apache Spark framework, and offers a set of fully-managed, easy-to-use and high-performance petabyte-level cloud data warehousing solution. Sparkling provides a one-stop solution for big data development and data computing. Sparkling features cross-data source conjoint analysis, which simplifies cloud data analysis, and allows companies to focus on data value mining and exploration.

Sparkling features elastic scaling, enabling you to adjust the cluster resource size based on your business development needs. It supports the import of heterogeneous data from multiple sources and job scheduling, allowing you to implement aggregated analysis of different data sources. It offers a consolidated, graphical interaction way that helps you perform data development and analysis operations easily and efficiently.

## Features

### Cluster Management

A Sparkling cluster is the core Sparkling Data Warehouse Suite component, composing of master nodes and worker nodes. Worker nodes can be divided into core nodes and elastic compute nodes. The former provides data storage capability and computing power, while the latter provides computing power. The size of the Sparkling cluster determines the upper limit of the storage capacity and computing power that Sparkling can provide.
With Sparkling, you can create highly available clusters to ensure high service availability. In high availability mode, the master node will still be available even in the event of rack-level hardware failures.

- Sparkling has a very simple and easy-to-use cluster management feature that allows you to specify the size of a Sparkling cluster that meets your business needs. In addition, with the growth of your business and increase in storage and computing needs, you can easily scale the cluster. As more nodes are added, the storage capacity and computing performance of the entire cluster will increase linearly.

©2013-2019 Tencent Cloud. All rights reserved.

- When you no longer need to use a cluster, you can terminate it. The terminated cluster cannot be recovered, and the data in the cluster will become inaccessible eventually.
- In addition to creating, terminating, and scaling a Sparkling cluster, you can view various information of the cluster such as the current running status, resource information, regions, and node model.

## Data Integration

Sparkling is capable of accessing and integrating a wide range of heterogeneous data sources. Data from traditional relational databases, COS, and Kafka streaming data storage can be extracted, transformed and loaded into the storage of Sparkling using the Data Studio console.
Sparkling offers a wealth of features for you to customize the data ingestion process. You can:

- **Crop rows**: Set filter criteria to crop rows of the imported data.
- **Import partial columns**: Delete some columns of the imported data and only import specified columns into Sparkling.
- **Adjust column order**: Adjust the order of the imported columns.
- **Set partitions**: Set partitions on the specified columns to improve the efficiency of subsequent data queries.
- **Import data in multiple formats**: Import data from data storage files in multiple formats.
- **Manage data sources**: Save and manage data sources for easy setup of new data import tasks and data traceability.
- **Import into existing data tables**: Import data into an existing data table in Sparkling.
- **Set mappings**: Set the mapping between the source data table and target data table to make data import more flexible.
- **Preview data**: Preview the imported data during data import.
- **Schedule imports**: Perform a single import or set up a scheduled task for periodic import.
- **Perform full/incremental imports**: Import all or incremental data.
- **Customize syntax for incremental import**: Add custom syntax to incremental import criteria for additional flexibility.

## Data Development

Sparkling is equipped with a Notebook-based online interactive environment. You can analyze and process data in Sparkling by running code in Notebook.

Supported languages include SQL, Python, and Spark. Sparkling SQL is a structured query language with syntax similar to MySQL, Oracle, and Hive SQL. It is compatible with ANSI SQL 2003, the industry SQL standard. Users who are familiar with traditional databases or Hive can easily get started with it. In addition to standard SQL operations, it has a rich set of embedded advanced functions that cover common mathematical operations, statistical analysis, and time and date computation.

©2013-2019 Tencent Cloud. All rights reserved.

You can also run Spark and PySpark programs in Sparkling Notebook, which facilitates the development of more flexible data analysis programs.

Sparkling Notebook offers data visualization tools. By dragging and dropping components, you can visualize data in Notebook in a variety of ways, such as pie charts and scatter plots. By combining interactive programming and data visualization, you can easily analyze and debug your data. You can also display the data analysis results in the form of reports and export and download them to a local file system.

Sparkling also offers many auxiliary features to help boost productivity. For example, you can organize your own notebooks by project and view data tables in an SQL IDE.

### Task Management

Sparkling supports setting periodic data import and Notebook tasks runs for continuous data updates. The schedule scale ranges from hour to month. Sparkling comes with reliable periodic scheduling and supports backfill task scheduling.
In addition to basic data import and scheduled Notebook tasks, you can mix and match data import and Notebook to form a DAG workflow task, which Sparkling will schedule by dependency. In complex data analysis pipelines and data science scenarios, this feature is especially useful.
Data import and scheduled Notebook tasks can be viewed and managed via Sparkling's consolidated task management interface. You can perform various operations such as viewing task status and history information, temporarily triggering tasks, or terminating tasks.

### Elastic Scaling

Sparkling comes with powerful elastic scalability. Computation and storage are separated and worker nodes of the cluster are divided into core nodes and elastic compute nodes. Manual and/or automated scale-out of high numbers of nodes and scale-up/down of computing and storing devices can be quickly achieved using Data Studio or Cloud API. Automated elastic scale-in is available to elastic compute nodes to meet changing business developments.

### Data Management

Sparkling has a metadata management module that enables registration, import, storage, retrieval, export and release of technical, management and business metadata, while providing a rich set of data management capabilities such as data maps, data dictionaries, data lineage and impact analysis, metadata version management, metadata statistical analysis and data quality reports.

### Project Management

Sparkling features a project management module that enables you to create project spaces by your organization's internal product lines, teams and projects and manage project personnel and notebooks.

©2013-2019 Tencent Cloud. All rights reserved.

# Product Strengths

Last updated：2019-04-23 17:33:13

Sparkling has the following features and advantages.

## One-stop Creation

To create a Sparkling service, you only need to select the parameter metrics of the product in the Tencent Cloud terminal interface. The specific creation process is completely invisible to you, as the backend takes care of the application for resources (eg. CVM, TencentDB, CLB) and other creation tasks such as Sparkling cluster building, service pulling, and monitoring.

## Consolidated Interaction Method

Sparkling offers a consolidated interaction method, so you can use data development platform for interactive data processing. It also offers JDBC/ODBC APIs which enable you to interact with Sparkling through programming.

## Professional Cluster Management

The exclusive usage mode of Sparkling provides cluster management and monitoring modules that support cluster creation, automatic scaling, cluster configuration, start/stop and intelligent resource monitoring and alarming. Sparkling clusters are fully managed, so you do not need to worry about the underlying cluster architecture or stress over OPS needs.

## Rich Heterogeneous Data Integration

Sparkling is capable of accessing and integrating a wide range of heterogeneous data sources. Data from traditional relational database RDBMS, COS, and CKafka can be extracted, transformed and loaded into the storage of Sparkling using the Data Studio console. Plus, Sparkling provides a rich set of extraction criteria and task scheduling features to meet your different data import demands.

## Comprehensive Data Management

Sparkling has a metadata management module that enables registration, import, storage, retrieval, export and release of technical, management and business metadata, while providing a rich set of data management capabilities such as data maps, data dictionaries, data lineage and impact analysis, metadata version management, metadata statistical analysis and data quality reports.

## Enterprise-grade Job Scheduling

Sparkling features an enterprise-grade job scheduling engine that helps you build complete workflows based on complex dependencies. It also offers a task scheduling and management module that supports

©2013-2019 Tencent Cloud. All rights reserved.

time-driven and event-driven DAG task scheduling. Meanwhile, it features comprehensive task monitoring to facilitate ETL and processing of OPS data.

### Complete Project Management and Account Service

Sparkling features a project management module that enables you to create project spaces by your organization's internal product lines, teams and projects and manage project personnel and notebooks. It has a complete set of account service-related schemes which ensure smooth resource sharing and permission management between the root account and sub-accounts.

### High Performance, Availability, and Scalability

Relying on the IaaS services provided by Tencent Cloud and its own component capabilities, Sparkling provides high-performance, high-availability, and high-scalability data warehousing services.

### Data Security

Sparkling protects your data security with the aid of dedicated networks, permission management, and multi-tenant isolation.

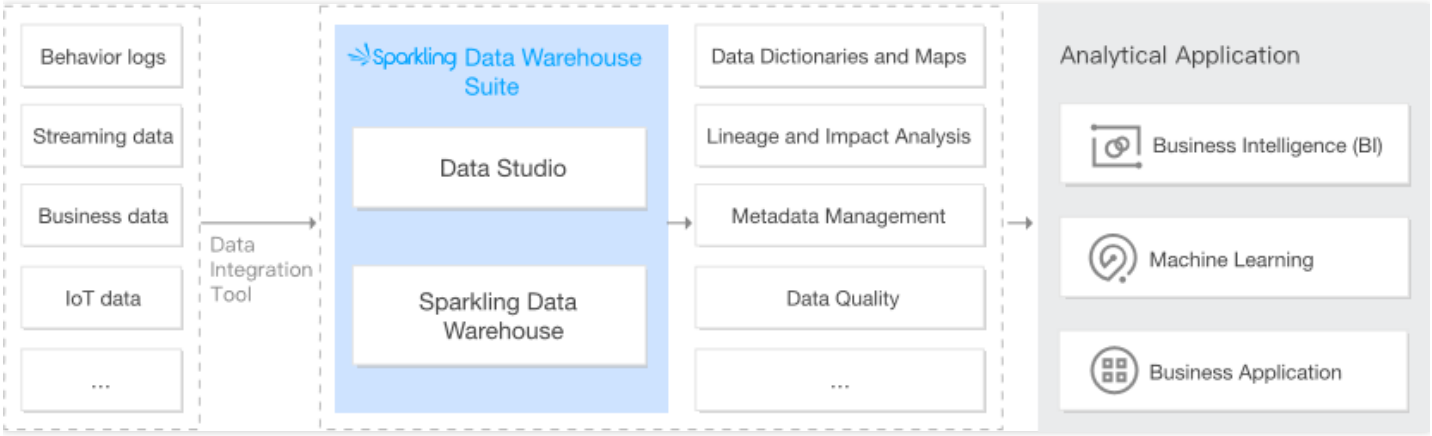
©2013-2019 Tencent Cloud. All rights reserved.

# Usage Scenarios

Last updated：2019-01-31 16:35:37

## Global Data Asset Management

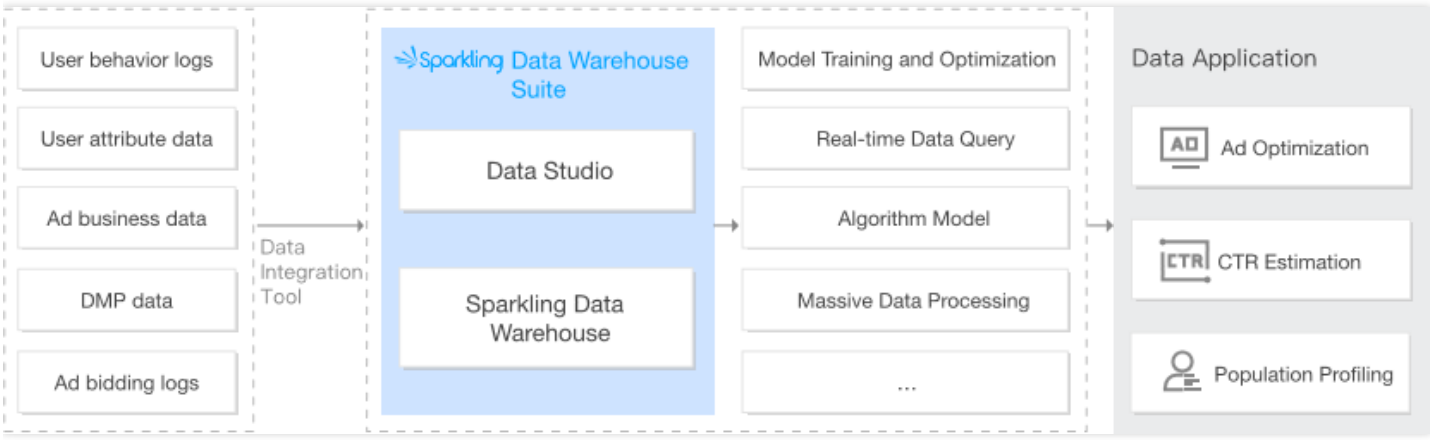
Sparkling effectively meets the urgent needs of industries such as gaming, finance, retail and industrial engineering by providing a tool to centrally manage and analyze management and business data of user behaviors, staffing, procurement, sales, assets and supply chain, so that a comprehensive view of global data can be generated to help understand overall operational conditions and make rapid and accurate decisions. By aggregating data from various scattered sources such as traditional databases, business systems and servers through its data integration tool, Sparkling enables the construction of an enterprise-level data map portal which leverages high-performance heterogeneous data source aggregation and analysis capabilities for data management and business value mining.




## Analysis of Large Volumes of Logs and Targeted Marketing

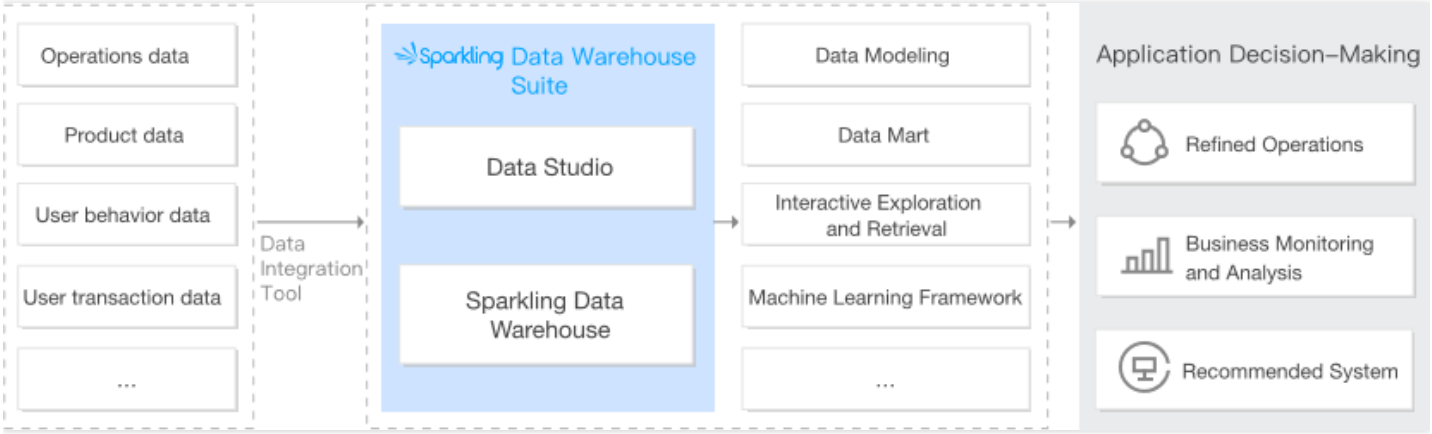
Featuring a log standardization and normalization mechanism, Sparkling enables you to conveniently analyze petabytes of structured or semi-structured data such as user behavior and system logs, generate cookie-based consumer profiles and personalize recommendations to users, helping significantly improve the efficiency of targeted marketing. Moreover, it supports real-time data ingestion and in-depth integration with COS.

©2013-2019 Tencent Cloud. All rights reserved.




## Efficient Data-based Decision-making

With the aid of its easy-to-use machine learning framework, interactive collaborative programming environment and real-time data query and analysis capabilities, Sparkling provides data scientists with powerful tools for data modeling. Plus, it enables business managers to refine corporate operations and helps them enhance business insight capabilities.




©2013-2019 Tencent Cloud. All rights reserved.