ALIBABA CLOUD

Alibaba Cloud

Realtime Compute

Introduction

Document Version: 20211229

Alibaba Cloud

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.

2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company or individual in any form or by any means without the prior written consent of Alibaba Cloud.

3. The content of this document may be changed because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.

4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides this document based on the "status quo", "being defective", and "existing functions" of its products and services. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, special, or punitive damages, including lost profits arising from the use or trust in this document (even if Alibaba Cloud has been notified of the possibility of such a loss).

5. By law, all the contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.

6. Please directly contact Alibaba Cloud for any errors of this document.

Document conventions

Style	Description	Example
Danger	A danger notice indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	**Danger:** Resetting will result in the loss of user configuration data.
Warning	A warning notice indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	**Warning:** Restarting will cause business interruption. About 10 minutes are required to restart an instance.
Notice	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	**Notice:** If the weight is set to 0, the server no longer receives new requests.
Note	A note indicates supplemental instructions, best practices, tips, and other content.	**Note:** You can use Ctrl + A to select all files.
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click **Settings**> **Network**> **Set network type**.
Bold	Bold formatting is used for buttons , menus, page names, and other UI elements.	Click **OK**.
`Courier font`	Courier font is used for commands	Run the `cd /d C:/window` command to enter the Windows system folder.
Italic	Italic formatting is used for parameters and variables.	`bae log list --instanceid` *Instance_ID*
[] or [a|b]	This format is used for an optional value, where only one item can be selected.	`ipconfig [-all|-t]`
{} or {a|b}	This format is used for a required value, where only one item can be selected.	`switch {active|stand}`

Table of Contents

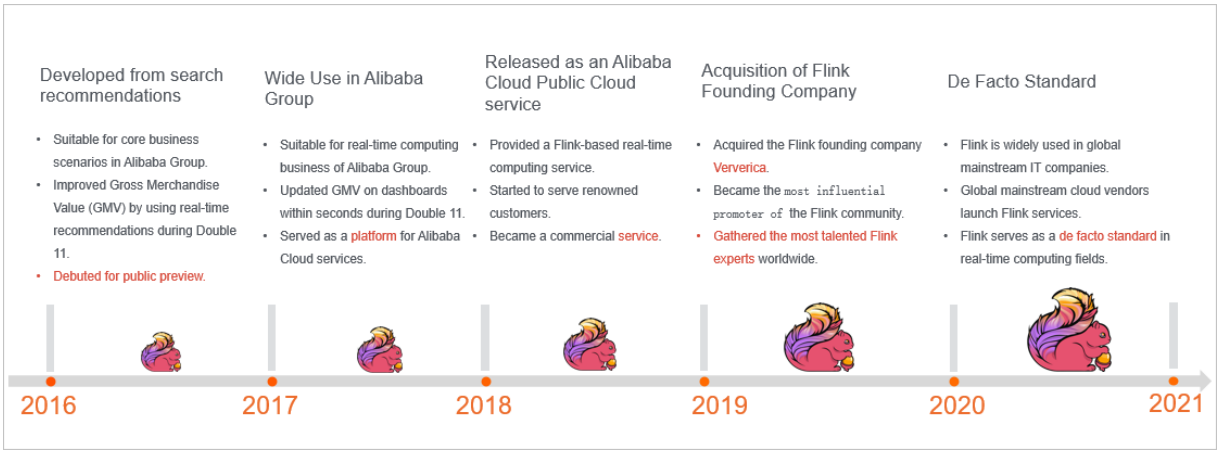
1.What is Alibaba Cloud Realtime Compute for Apache Flink?

Alibaba Cloud Realtime Compute for Apache Flink is an end-to-end real-time big data analytics platform that is built on Apache Flink. Alibaba Cloud Realtime Compute for Apache Flink provides end-to-end real-time data analysis capabilities with subsecond data processing latency. It simplifies the business development process by using standard SQL statements to help enterprises transform their business into real-time and intelligent big data computing business.

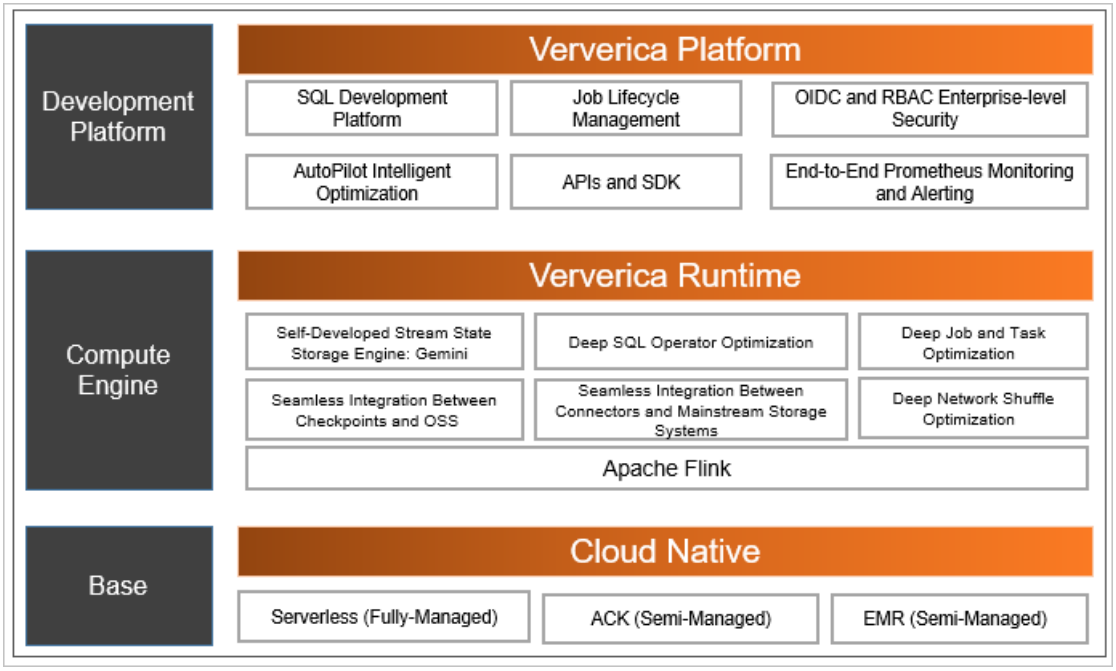
History

In 2017, Alibaba Group integrated Blink with Galaxy and JStorm. Blink was used as a unified real-time computing V2.0 product due to its excellent performance and provided real-time computing services for all business units (BUs) in Alibaba Group. Blink is a branch of Apache Flink. In the past four years, Alibaba Group made in-depth optimizations and improvements to enable Blink to support ultra-large-scale business scenarios of Alibaba Group, such as search and recommendation.

In January 2019, Alibaba Group acquired the Flink founding company Data Artisans. The Blink technical team and the Flink founding team jointly built a globally unified Flink Enterprise Edition platform, which is called Ververica Platform (VVP). This development leads real-time computing to Era 3.0.



Architecture



Benefits

- Superior performance: A single CPU core can process hundreds of thousands of data records per second, with subsecond data processing latency between ports. Tens of thousands of ultra-large-scale real-time computing tasks can run in parallel.
- Powerful features: Realtime Compute for Apache Flink is an end-to-end SQL-based development and O&M platform that provides intelligent diagnosis and automatic configuration optimization. Realtime Compute for Apache Flink can seamlessly connect to mainstream data services of Alibaba Cloud.
- Cost-effectiveness: The hourly computing fees per CPU core are low. Auto scaling is implemented based on the workload and the pay-as-you-go billing method is supported. The total cost of ownership (TCO) of Realtime Compute for Apache Flink is significantly lower than the TCO of self-managed Flink in data centers.
- Guaranteed stability and reliability: The service level agreement (SLA) guarantees 99.9% availability. End-to-end metric monitoring and alerting are supported. Realtime Compute for Apache Flink provides high stability and reliability in large-scale deployment scenarios such as Double 11.
- Compatibility with self-managed Flink: Realtime Compute for Apache Flink is fully compatible with self-managed Flink. Therefore, you can smoothly migrate the business of self-managed Flink to the cloud. Realtime Compute for Apache Flink can be seamlessly connected to mainstream open source big data ecosystems.
- Outstanding branding: Realtime Compute for Apache Flink is officially released by the founding team of Apache Flink and certified by the China Academy of Information and Communications Technology (CAICT). Realtime Compute for Apache Flink is the only real-time stream processing product that is recognized in the Forrester Wave.

Comparison between Realtime Compute for Apache Flink and self-managed Flink

Realtime Compute for Apache Flink takes more advantages in terms of functionality and stability over self-managed Flink. In addition to O&M advantages, Realtime Compute for Apache Flink provides the out-of-the-box feature for ease of use. The following table describes the advantages of Realtime Compute for Apache Flink.

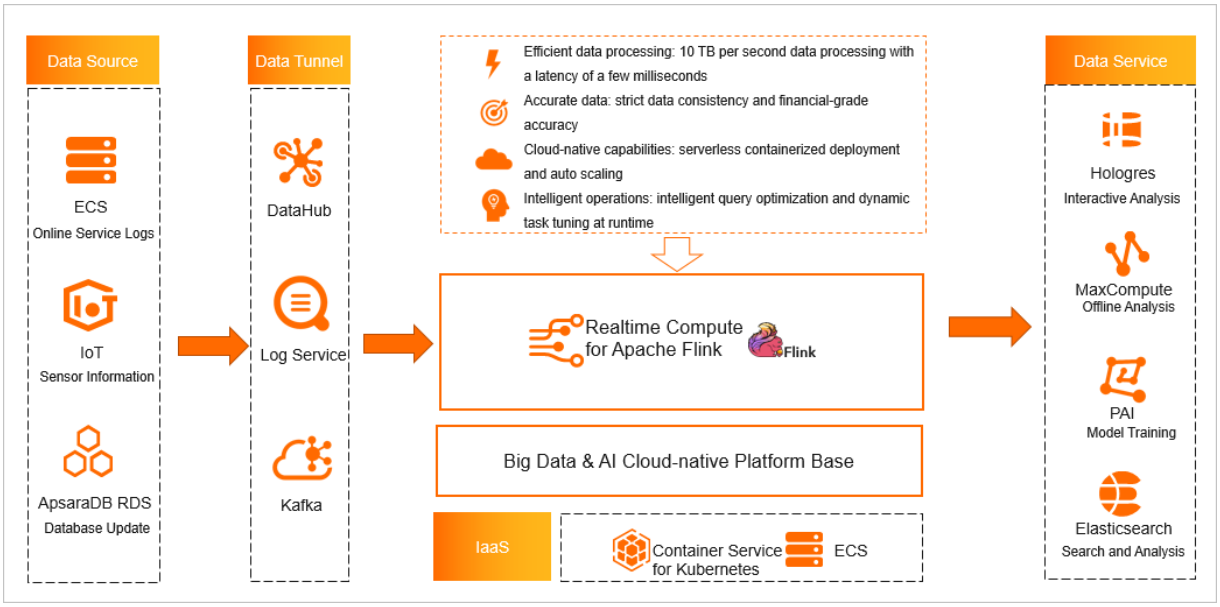
Category	Feature	Description
Development	Data connection	Fully managed Flink can be seamlessly integrated with mainstream data services of Alibaba Cloud including mainstream databases, Message Queue, and Log Service.
		You can access various external storage systems from fully managed Flink by using custom connectors.
	Task development	Programming languages: Realtime Compute for Apache Flink provides an end-to-end development and management platform, which supports various programming languages, including SQL, Java, Scala, and Python.
		Metadata: Realtime Compute for Apache Flink provides a unified metadata management system and can be seamlessly connected to external metadata systems, such as MySQL and Hive.
		Function libraries: Realtime Compute for Apache Flink supports multiple built-in function libraries of different fields, such as Analytics Zoo Cluster Serving, and allows you to use user-defined functions (UDFs) based on your business requirements.
	Code debugging	Test data management: Realtime Compute for Apache Flink supports online sampling and management of mock testing data to help you build a test process.
		Fast running and debugging: Realtime Compute for Apache Flink allows you to start or stop jobs in session clusters within seconds. This makes job debugging more efficient.
		Development and production isolation: Realtime Compute for Apache Flink isolates development from production. This way, jobs and data in the production environment are not affected during the debugging process.
O&M	Monitoring and alerting	Realtime Compute for Apache Flink supports end-to-end monitoring and alerting. When you run a job, Realtime Compute for Apache Flink automatically reports alerts if issues such as data delay, data skew, and backpressure occur. Realtime Compute for Apache Flink can also monitor metrics and aggregate dimensions to help you troubleshoot issues, such as job delays, data skew, and backpressure.
		Alert notifications can be sent by using DingTalk, emails, and text messages in a timely manner. You can also connect Realtime Compute for Apache Flink to an internal unified alerting system, such as Prometheus or Graphite.
	Intelligent diagnosis and Autopilot	Intelligent diagnosis: identifies job issues in a timely manner and provides suggestions for troubleshooting.
		Autopilot: automatically monitors and adjusts job resource allocation in unattended mode to manage traffic surges.

Category	Feature	Description
	Fine-grained resource management	Realtime Compute for Apache Flink supports fine-grained resource configuration at the operator level. You can configure CPU cores and memory for each operator of each job. This significantly improves resource utilization and service stability and reduces costs and the probability of out of memory (OOM) errors.
	High availability	The maintenance service delivers SLA-guaranteed service availability of up to 99.9%. In addition, end-to-end automated fault tolerance can ensure system stability.
Cost	Billing method	The subscription and pay-as-you-go billing methods are supported. You can select a billing method that suits your business requirements.
	Core performance	The Nexmark benchmark test result shows that the stream computing performance of Realtime Compute for Apache Flink is about three times the performance of self-managed Flink. The strong R&D team of Alibaba Group optimizes Realtime Compute for Apache Flink based on the practices that have been accumulated in core internal business scenarios. This highlights the core advantages of Flink and reduces the basic cost of the service.
	Auto scaling	Realtime Compute for Apache Flink has cloud-native auto scaling capabilities. It can perform automatic scale-out or scale-in operations based on the workload. This ensures the timeliness of business and improves the resource utilization. Cloud computing costs are optimized to improve system performance and reduce the TCO.
Security	Isolation	Tenant-level and project-level resource isolation and code isolation are supported to allow different teams to collaborate on projects. Containerized task isolation is used to improve user experience.
	Access control	Realtime Compute for Apache Flink uses the Alibaba Cloud account system to support the OpenID Connect (OIDC) protocol and role-based access control (RBAC). You can seamlessly manage the security of your services by using your Alibaba Cloud account. This significantly improves the security of your business.

Solution

As a real-time stream computing engine, Flink can process a variety of real-time data, including online service logs of Elastic Compute Service (ECS) instances and sensor data in IoT scenarios. You can also subscribe to updates of binary logs in relational databases, such ApsaraDB RDS and PolarDB. Then, you can use DataHub, Log Service, and Message Queue to subscribe to real-time data. After Realtime Compute for Apache Flink reads the real-time data, it analyzes and processes the data in real time. The analysis results are written to different data services, such as MaxCompute, Hologres, Machine Learning Platform for Artificial Intelligence (PAI), and Elasticsearch. You can select an ideal data service based on your business requirements to improve data utilization.

Realtime Compute for Apache Flink is mainly used to subscribe to, process, and analyze data from various real-time data sources in real time, and write the analysis results to other online storage for subsequent use. Realtime Compute for Apache Flink is a comprehensive enterprise-class service, which is developed based on the cloud-native architecture and provides fast, accurate, and intelligent data computing. Realtime Compute for Apache Flink runs on Infrastructure as a Service (IaaS) services of Alibaba Cloud, such as Container Service for Kubernetes and ECS. Realtime Compute for Apache Flink can connect to various Alibaba Cloud services.



2.Service types

Before you purchase Realtime Compute for Apache Flink, we recommend that you are familiar with the features of each service type so that you can select suitable instances for your business.

Features

Realtime Compute for Apache Flink has five service types. The following table describes the features of each service type.

Service type	Product line	Release status	Deployment mode	Compute engine	Development platform
Fully-managed Flink	New product line	Commercialized	Kubernetes-based deployment	Flink	VVP
Semi-managed Flink (ACK-based)	New product line	Phased-out	Kubernetes-based deployment	Flink	VVP
Semi-managed Flink (EMR-based) **Note** A Realtime Compute for Apache Flink cluster of this service type is a dataflow cluster that runs on Alibaba Cloud E-MapReduce.	New product line	Commercialized	YARN-based deployment	Flink	User-defined
Blink exclusive cluster (original product line)	Original product line	Phased-out	YARN-based deployment	Blink	Bayes
Blink shared cluster (original product line)	Original product line	Phased-out	YARN-based deployment	Blink	Bayes

Selection recommendations

- **Fully-managed Flink**: If your company or team wants to focus on business development without spending extra time on cluster O&M, we recommend that you use this service type.
- **Semi-managed Flink (ACK-based)**: This service type reached its End of Service (EOS) on April 28, 2021. If you want to purchase Realtime Compute for Apache Flink, we recommend that you purchase the fully managed Flink service type.
- **Semi-managed Flink (EMR-based)**: If your company or team wants to gain full control over all cluster resources during business development and is familiar with YARN or Alibaba Cloud E-MapReduce, we recommend that you use this service type.
- **Blink exclusive cluster (original product line)**: This service type reached its EOS on April 28, 2021.

If you want to purchase Realtime Compute for Apache Flink, we recommend that you purchase the fully managed Flink service type.

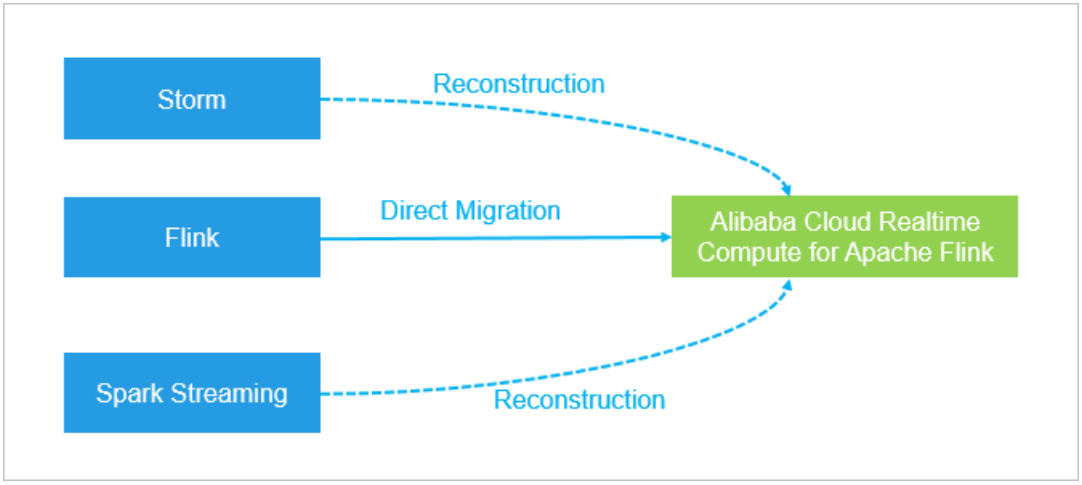
- **Blink shared cluster (original product line)**: This service type is phased out. You can scale in or out only the existing Blink shared clusters. We recommend that you do not use this service type.

3.Scenarios

Realtime Compute for Apache Flink is widely used for real-time big data computing. This topic describes how to apply Realtime Compute for Apache Flink to different enterprise departments and technologies.

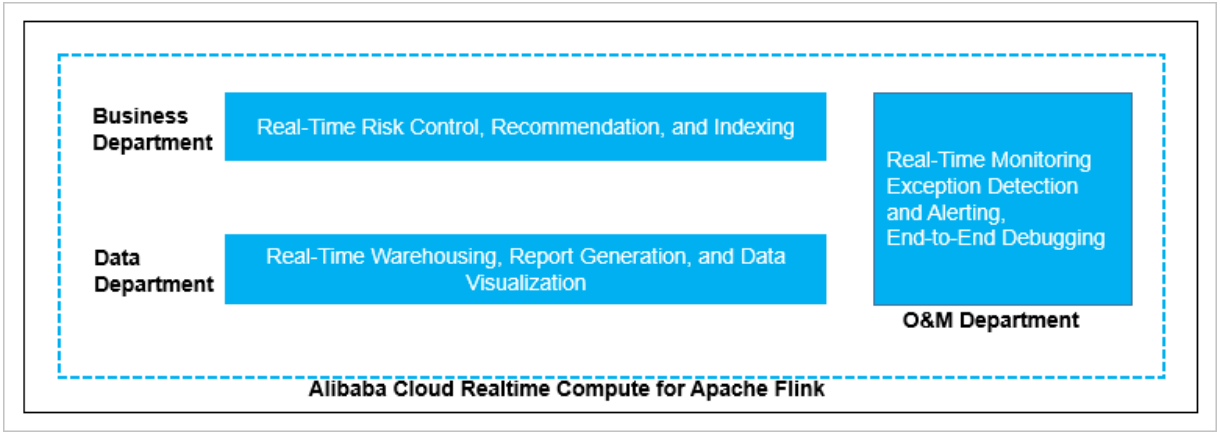
Existing stream processing system

If you have installed the Flink system on your local server, you can directly migrate it to Realtime Compute for Apache Flink. If you have installed the Storm or Spark Streaming system on your local server, you can rebuild and then migrate it to Realtime Compute for Apache Flink.



Enterprise departments

Realtime Compute for Apache Flink provides different features for the following departments:



- Business department: real-time risk control, real-time recommendation, and real-time indexing of search engines.
- Data department: real-time data warehousing, real-time reports, and real-time dashboards
- O&M department: real-time monitoring, real-time exception detection and alerting, and end-to-end debugging.

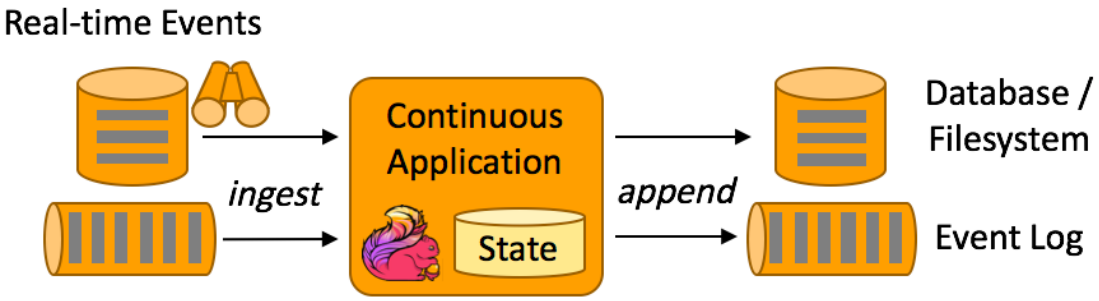
Technologies

From the technical perspective, Realtime Compute for Apache Flink is suitable for the following

scenarios:

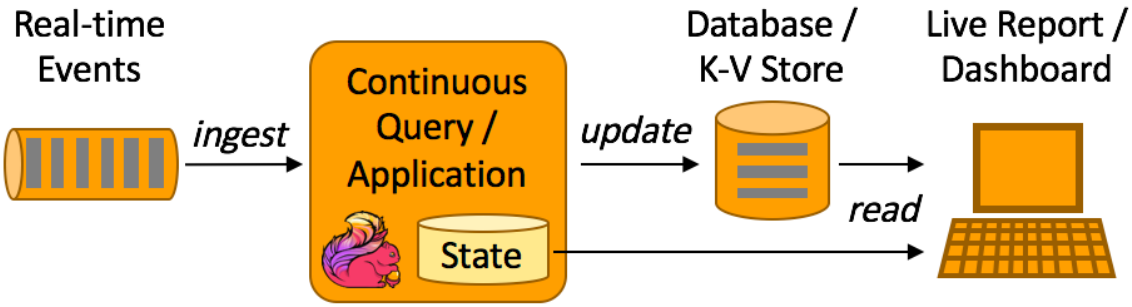
- Real-time extract, transform, load (ETL) and data streams

  Data is delivered from point A to point B by using the real-time ETL procedure and data streams. During data delivery, data cleansing and integration may be required, such as real-time indexing in the search system and ETL procedure in real-time data warehousing.



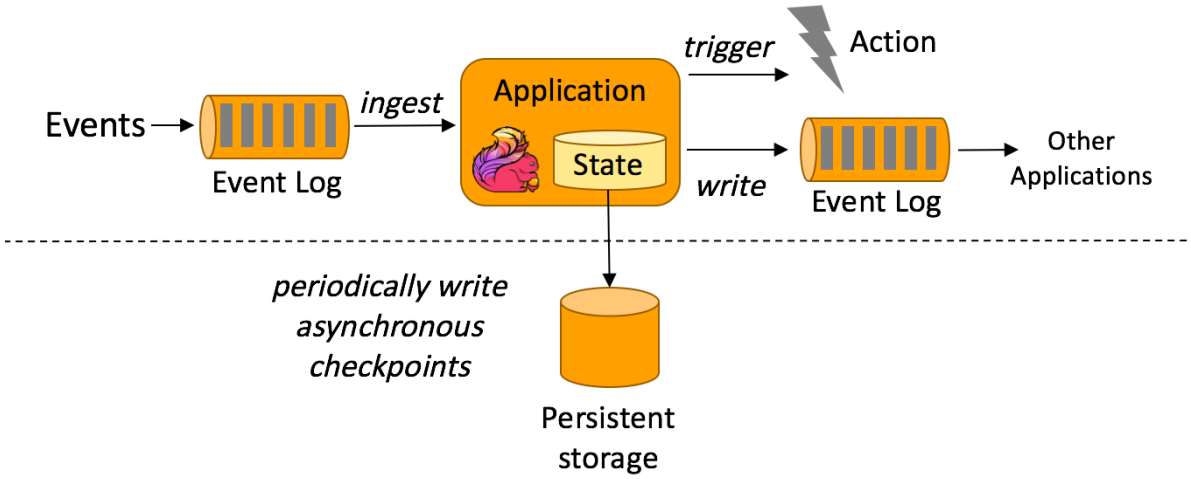
- Real-time data analysis

  Data analysis is a process to extract and integrate required information from raw data to achieve your business objectives. For example, you can view the top 10 products sold per day, the average turn-around time in the warehouse, the average document click rate, and the open rate for push notifications. Real-time data analysis allows you to view real-time reports or dashboards.



- Event-driven applications

An event-driven application is a system that processes or reacts to subscription events. Event-driven applications depend on internal states and respond to suspicious events detected during fraud detection or in the risk control system or O&M exception detection system. If the behavior of a user triggers a risk control point, the system captures the event and analyzes the current and previous user behavior to determine whether to perform risk control over the user.



Note The preceding technology flowcharts are obtained from Apache Flink official website.

4.Technical support

This topic describes how to obtain technical support when you encounter issues about Realtime Compute for Apache Flink.

How to obtain technical support

- Submit a ticket

  > **Notice** If you use the credentials of a RAM user, you cannot submit a ticket. In this case, you can use an Alibaba Cloud account to attach the **AliyunSupportFullAccess** policy to your RAM user and grant the ticket management permissions to the RAM user. For more information, see Use RAM to manage permissions of O&M engineers.

  i. Got to the New Ticket page of the ticket system.

  ii. Click the icon under **More Products and Services**.

  iii. Click **Realtime Compute** under **Base Service**.

  iv. On the **New Ticket** page, click New Ticket. In the New Ticket dialog box, enter related information.

  v. Click **Submit**.

- Join a DingTalk group

  You can scan the following QR code to join the DingTalk group of Realtime Compute for Apache Flink. You can communicate with other users in the DingTalk group.

  > **Note** The DingTalk group is used only for communication. If you have business issues or product issues, submit a ticket.

Service SLA

For more information about the supported service SLA, see Service level comparison.