Batch Compute

Product Introduction

Product Documentation

Contents

Product Introduction
What's Batch?

Last updated：2024-01-13 11:19:28

BatchCompute (Batch) is a distributed computing platform used to handle massive batch processing jobs. It is designed to provide cost-effective and easy-to-use computing services for enterprises and R&D institutions that require big data computing. It can intelligently manage resource allocation and job execution for batch processing jobs of any size, enabling you to focus on analyzing and processing data. With these advantages, it is suitable for fields such as gene sequencing, medical analysis, scientific research, video rendering, financial big data, and targeted advertising.

As a completely secure hosted service, BatchCompute enables developers, scientists, and engineers to execute batch processing jobs of any size. By using BatchCompute, you only need to provide an environment for computing implementation, commands, and the storage path for input/output data. BatchCompute can get Tencent Cloud's elastic resources based on the workload and automatically schedule the job execution process to complete high volumes of batch processing jobs. It can help reduce the operation complexity, save time, and minimize costs to allow developers, scientists, and engineers to easily execute batch processing jobs in Tencent Cloud, relieving enterprises and R&D institutions of the need to make a huge upfront investment in building a batch processing data center.

Scenarios

Last updated：2024-01-13 11:19:28

Gene Sequencing

Bioinformatics companies or laboratories use sequencers to obtain the raw files of gene sequences, upload the information to cloud storage systems such as COS after completing the preliminary analysis of the sequences, and then further process the information using BatchCompute.
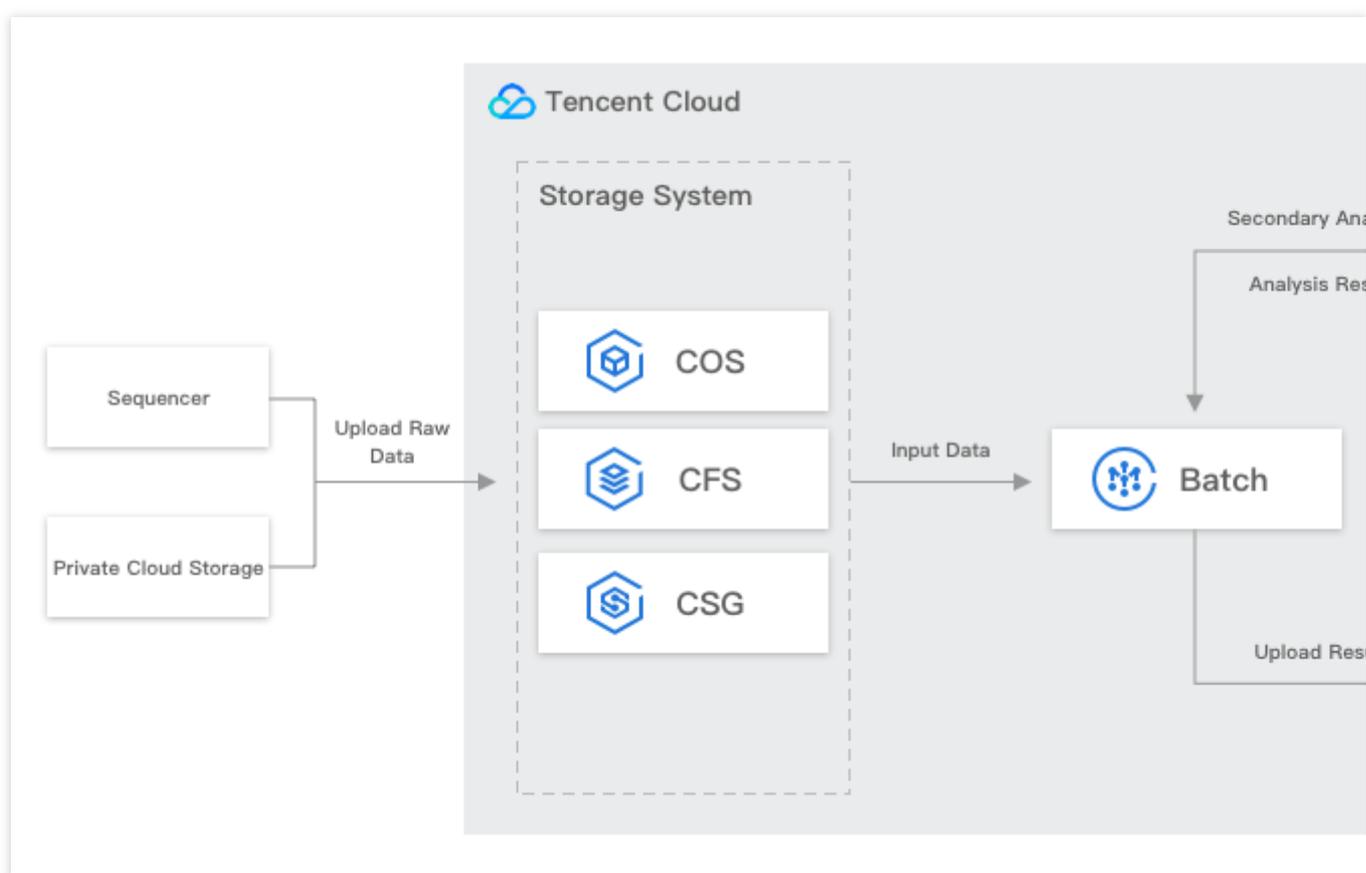
General Steps

1. The bioinformatics expert gets the raw information from the sequencer or private cloud storage and uploads it to Tencent Cloud storage services such as COS and CFS.
2. The user defines the BatchCompute job used to analyze the information and then submits the job. The storage configuration of the job is associated with the raw information uploaded in the previous step.
3. BatchCompute automatically schedules resources, deploys the user-uploaded custom analysis image to the scheduled CVM instance, and schedules the job to analyze the raw information.

4. After the computation on CVM is completed, BatchCompute will automatically upload the analysis results to the user-specified location;

5. The raw information to be analyzed will be uploaded to COS.

Movie, Video, and Design Sketch Rendering

In the visual creation industries such as movie, television, advertising, and architectural planning, content creators and post-production companies need to use a large number of machines to complete the rendering work related to visual effects, 3D animations, and design sketches. BatchCompute enables users to automate content rendering workflows. Users can build their own rendering-dependent processes while leveraging BatchCompute's abilities to schedule massive volumes of resources and jobs to efficiently complete visual creation work.
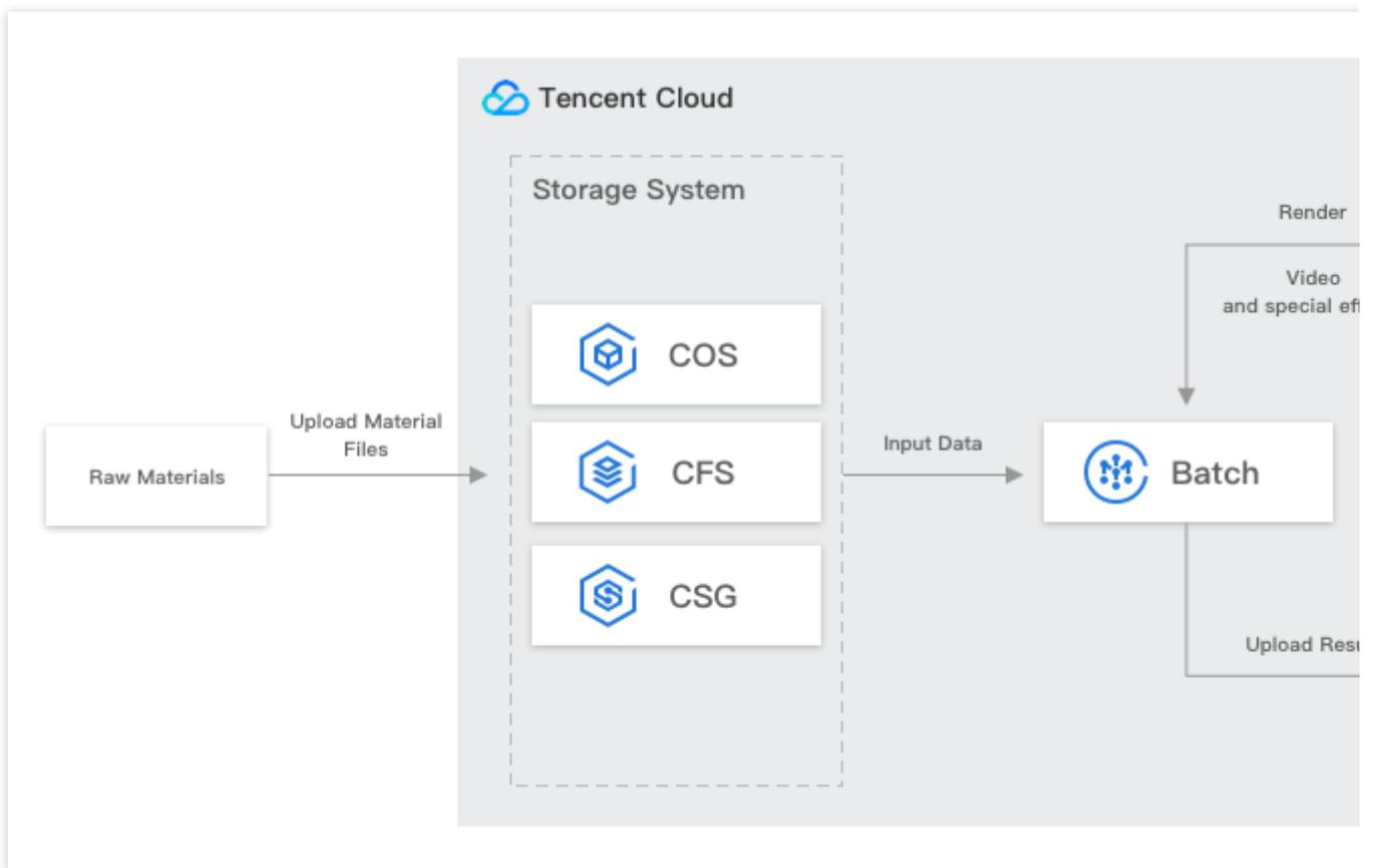
General Steps

1. The user prepares the raw rendering material files needed and uploads them to Tencent Cloud storage services such as COS and CFS;

2. The user defines the BatchCompute job used to analyze the information and then submits the job. The storage configuration of the job is associated with the raw information uploaded in the previous step.

3. BatchCompute automatically schedules resources, deploys the user-uploaded custom rendering image to the scheduled CVM instance, and schedules the job to render the videos or design sketches.

4. After the rendering on CVM is completed, BatchCompute will upload the generated videos or sketches to the user-specified location.

Advantages

Last updated：2024-01-13 11:19:28

| Advantage	Tencent Cloud BatchCompute	Self-Built Compute Cluster
Low costs	**Zero upfront costs** You can call a large number of CVM resources for use by BatchCompute free of charge when needed and only pay for the CVM instances you actually use. Further, you don't have to concern over building and maintaining the infrastructure, saving costs and time.	**Costly construction** A self-built infrastructure, whether purchased or leased, incurs very high costs. It takes at least millions of dollars to build a large-scale compute cluster. Building a compute cluster can also be very time-consuming with unknown costs and risks, making it difficult for you to focus on the businesses.
Elastic resources	**On-demand resource scaling** BatchCompute can be used fully on demand, so you can use it to get the necessary resources only when you have a computation job.	**Wasted or insufficient computing power** No matter how you carefully plan the size of a self-built resource cluster, waste or insufficiency are inevitable in the face of unpredictable business needs. The risk of cost audits caused by idle resources and negative impact on business development caused by lack of resources are common problems related to self-built clusters.
Intelligent scheduling	**Fully managed solution** BatchCompute can fully manage and schedule resources, which is the most challenging part of batch computing, so you only need to define and submit computing jobs and then focus on result analysis and issue solving.	**Manual support** Assigning resources and scheduling processes manually or with self-developed scheduling software is very cumbersome and error-prone, especially when in the face of massive amounts of resources and computing operations. Plus, it is difficult to estimate how many manpower and resources will be wasted.
Functionality and ease of use	**Rich and convenient features** In addition to all the common features for batch processing, BatchCompute also comes with a suite of complementary features that help you define and submit computing jobs, such as defining a variety of implementation methods, orchestrating the computing process, and monitoring running state.	**Self-developed or third-party tools** Developing a massive batch processing system in house leads to a large amount of work which is not related to your main business, while most of third-party tools lack timely technical support, requiring you to explore how to use them on your own. In both cases, it is impossible for you to focus on business development.

| | In addition, it can work in tandem with other relevant Tencent Cloud services such as COS. It provides one-stop closed-loop services for computing material acquisition, computing job submission and execution, and computing result upload. | |

Glossary

Last updated：2024-01-13 11:19:28

Job

A job is the smallest unit in which a user submits a batch processing work and consists of one or multiple tasks with dependencies. Dependencies among multiple batch processing tasks can be set through the very easy-to-use DAG syntax to combine such tasks into one single job, and then the tasks can be executed in sequence till all of them are completed, that is, the job is completed. Dependencies among tasks can be specified only when the job is submitted and cannot be modified after the submission.
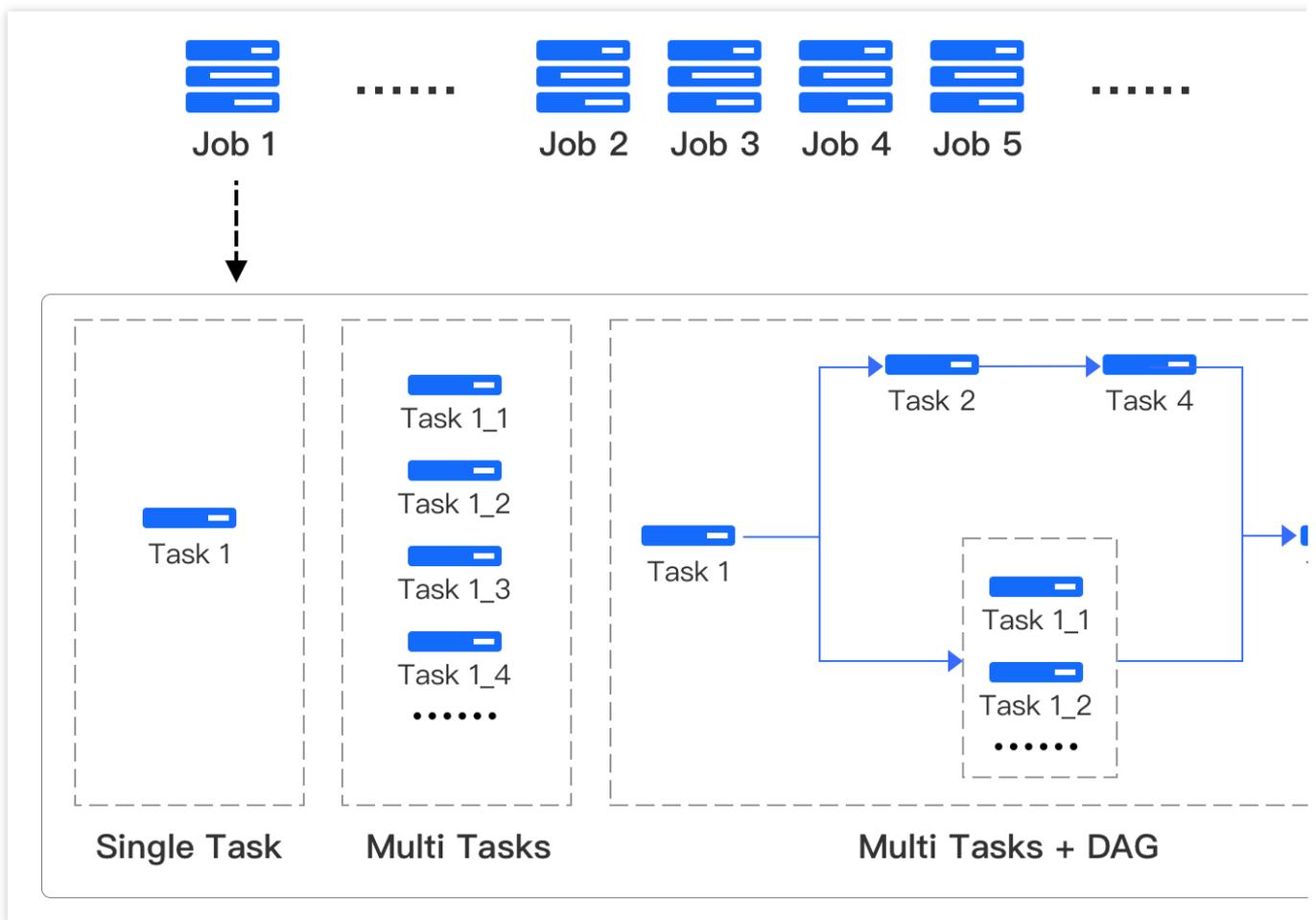
Task

A task is a basic component of a job and contains the information about the application actually executed on a CVM instance. BatchCompute's scheduling system automatically creates CVM instances and executes applications based on the configuration submitted by the user. The task cannot be submitted directly. Instead, it must be placed in a job for submission and execution. One or more jobs can be placed in one job.

The core configurable attributes of a task are as follows:

`CVM instance configuration` : the task is executed on CVM. You need to configure the type and specification of the CVM instance according to the characteristics of your computing task, such as selecting a compute instance (C2) or a high IO instance (IO2), memory and disk size, and the VPC where the instance resides.

`Execution environment` : it contains the image and command configuration. The image is usually a custom image containing your applications and the environment they depends on, and the command specifies how to launch these applications for computation.

`Remote storage mapping` : a remote storage address can be mapped to a local file system address. Currently, this feature supports COS. For more information, please see the separate definitions below.

`Standard output` : you can configure the mapping address (COS) of the standard output. The information output to `stdout` and `stderr` from the application will be uploaded to the corresponding address after the task is completed, so that the computing process can be traced back after the task is completed.

Task Template

You can make a task template for common tasks based on which to customize different tasks and thus submit jobs more quickly.

Remote Storage Mapping

Storage mapping maps the remote storage address (COS or CFS) to the CVM instance's local file system, so that the remote storage can be read and written in the same way just like a local file system.

Task Instance

Task instance is the smallest unit of scheduling and execution in BatchCompute. One task contains one or more task instances, each of which runs on a CVM instance for executing the corresponding computation task. You can set the number of instances that need to be executed concurrently in the task configuration. BatchCompute will schedule the specified number of instances with the same configuration when the task is started. You can distinguish among these instances by environment variables in the executor.

A typical application scenario of multiple instances is that the input data can be split and then processed concurrently, so that the advantages of elastic cloud resources can be fully leveraged to compute concurrently and improve work efficiency.

Image

An image is configured in a task and used for instance creation. It must be a standard or custom CVM image. You need to prepare the compute environment and application and make a custom image containing them. You can view your custom images in the Image Console. For more information on how to make a custom image, please see Creating Custom Image .

Compute Environment

A compute environment (ComputeEnv) is a compute cluster consisting of one or more CVM instances and can be scaled up or down based on business needs. If the task configuration specifies the compute environment, the task instance will be scheduled to the nodes in the specified compute environment for execution; otherwise, BatchCompute will create a CVM instance for executing the task instance and terminate the CVM instance after the task instance is completed by default.

Quota & Limits

Last updated：2024-01-13 11:19:28

Quota Limit	Description
CVM	BatchCompute may use CVMs (pay-as-you-go instances and spot CVM instances) while executing a job. There are limits on the number of CVM instances that a user can hold in different availability zones. For more information, see the "Purchase Limits for Pay-as-You-Go CVM Instances" section in Purchase Limits.
Job	Up to 1,000 jobs per user
Task template	Up to 50 task templates per user
Compute environment	Up to 10 compute environments per user Up to 2,000 CVM nodes per environment Up to 500 BM nodes per environment

Data Security

Last updated：2024-01-13 11:19:28

Data Encryption

Batch Compute is a computing service that encrypts sensitive data in computing environment, jobs and task templates and stores Ciphertext persistently in TencentDB.

Data Storage

Batch Compute provides compute capability and performs task scheduling. However, when using Batch Compute, the generated data will not be kept in the service but stored in Tencent Cloud storage products, such as Cloud Block Storage, Cloud Message Queue, Cloud Object Storage and Cloud File Storage.