

# 批量计算

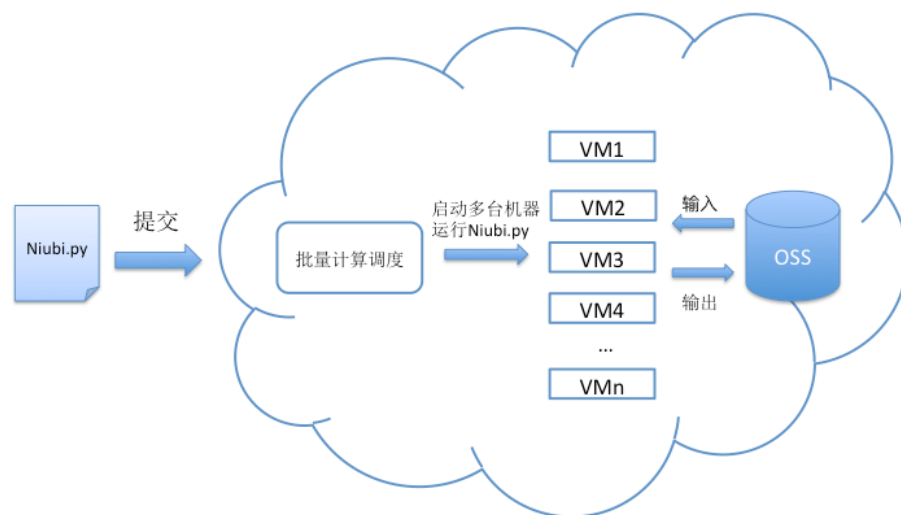
## 产品简介

# 产品简介

## 什么是批量计算

批量计算（BatchCompute）是一种适用于大规模并行批处理作业的分布式云服务。BatchCompute可支持海量作业并发规模，系统自动完成资源管理、作业调度和数据加载，并按实际使用量计费。

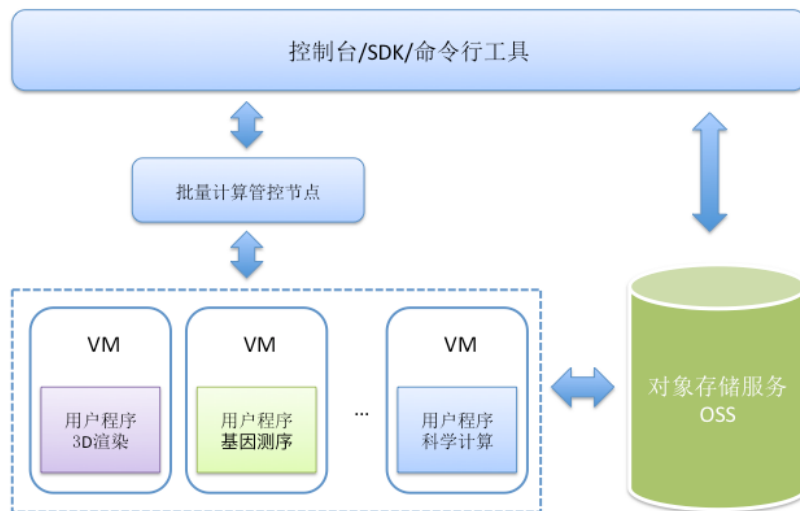
通俗的讲，就是您提交一个程序（任意的计算机程序），让它在阿里云机房中的多台机器上同时运行，然后把结果写入到指定的位置（如：OSS，阿里云对象存储服务），您可以去指定位置查看结果。



BatchCompute整体架构，包括：

### BatchCompute入口

- 批量计算服务提供基于HTTP的API。在API之上还封装了一些工具，比如SDK，命令行工具，控制台等。
- 您可以使用这些工具向BatchCompute提交作业，查询作业运行状态，并管理作业生命周期（比如停止已提交的作业，释放已完成的作业等）。



#### - 运行环境

BatchCompute通过虚拟化技术构建沙箱，并在其中运行用户程序，确保用户环境和用户数据的安全性。您可以自行定制其虚拟机镜像。

#### OSS持久化存储

- OSS 有完善的用户管理和安全机制。
- BatchCompute使用 OSS 作为持久化存储。
- 用户的程序、自定义Docker镜像、输入输出数据都存储在 OSS 中。

## 作业

用户的每个特定的计算需求在BatchCompute中被描述为一个作业（Job）。在BatchCompute系统中，用户可以通过BatchCompute管理控制台或SDK提交、管理和查询作业。

## 任务

一个作业由一组任务（Task）及其依赖关系组成。批量计算支持能以有向无环图（directed acyline graph，DAG）形式描述的作业。任务间的依赖关系只能在作业提交时指定，提交完成后不能修改。

## 实例

每个任务可以有一个或多个执行实例（Instance）。同一任务的各个实例并行处理各自的输入数据。实例是

BatchCompute调度与执行的最小单元，这些实例会动态的运行在后台分布式系统的各节点上。

## 集群

一系列的VM实例组在BatchCompute中被定义为一个集群（Cluster）。一个集群可以由多个实例组（Group）组成，一个组中的实例共享同样的实例类型定义（实例类型，CPU，Memory等），多个组之间共享镜像（Image）。

## 镜像

镜像（Image）是集群资源创建或者作业运行环境的模板。它是一个标准的或者自定义的ECS镜像，需要安装BatchCompute Agent，然后授权给BatchCompute服务使用。

## Docker镜像

使用官方提供的ubuntu镜像(内置Docker容器支持)，可以支持自定义Docker镜像。自定义Docker镜像像普通镜像一样，可以自己安装运行作业的各种程序。具体方法请看[这里](#): 使用Docker。

## 1. 提交作业

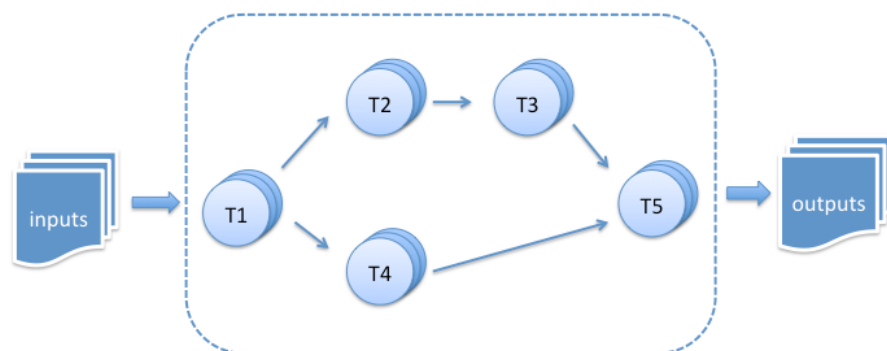
用户使用工具（如SDK，命令行工具等）向BatchCompute提交作业，BatchCompute使用用户指定的镜像（如：ubuntu）启动虚拟机(VM)，在虚拟机中运行用户程序，运行完成后释放虚拟机(VM)。

- BatchCompute中使用OSS作为持久化存储。您可以在程序运行完成时将结果数据保存到OSS中。在批量计算中，也可以通过文件接口的方式访问OSS上的数据，请参阅[OSS挂载功能](#)。
- BatchCompute程序默认运行在VM中，也可以支持Docker容器。也就是说，您可以自定义ECS镜像或者使用Docker，在镜像中安装自己需要的任何软件，用来运行您的任何程序。

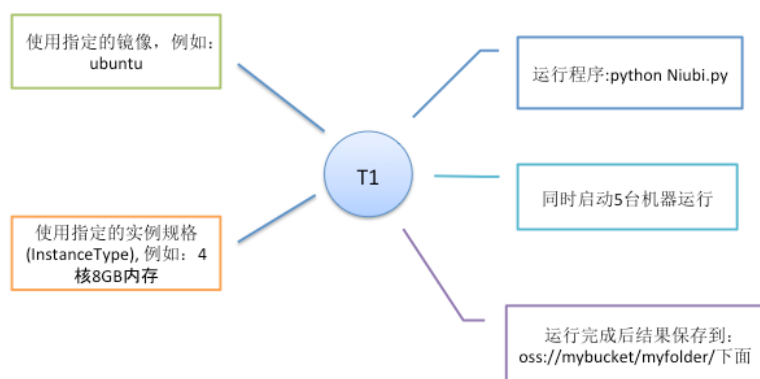
### 作业描述

用户需要提交一个作业(Job)描述JSON文件到批量计算服务，该JSON文件中详细描述了需要执行哪些程序(可以多个程序)，运行哪些程序需要启动多少台机器，机器的规格（内存和CPU等），运行日志打印到哪里，完成后结果输出到哪里等。

一个作业(Job)包含多个任务(Task)，按照您指定DAG描述的顺序执行。



每个任务定义了使用哪个镜像，使用什么实例规格，运行哪个程序，需要多少台机器运行，还有结果存储在哪里等。



## 2. 管理我的作业

您可以使用工具（控制台，命令行工具等），查看我提交的作业，可以停止，重启，或删除作业。查看各个任务的情况，各个实例（VM实例）的情况和日志。下图是控制台的作业管理界面：

11

- 支持自定义镜像。
- 支持Open API易于集成。
- 多租户隔离。

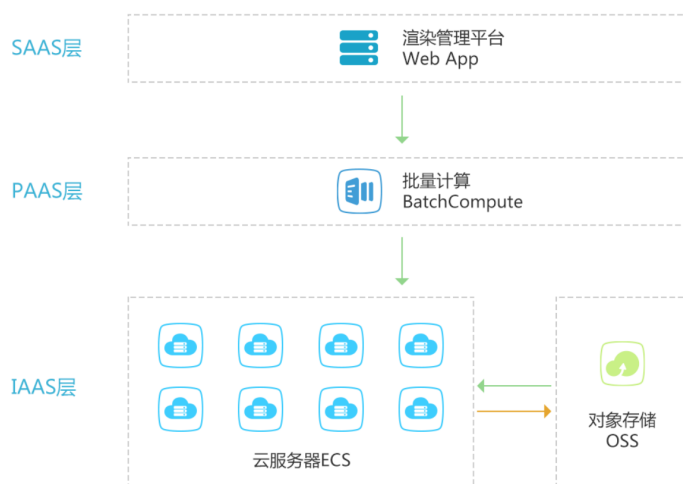
按量付费：按照计算资源实际使用量付费。

BatchCompute广泛应用于电影动画渲染、生物数据分析、多媒体转码、金融保险分析等领域。

## 典型案例

### 电影动画渲染

阿里云渲染云解决方案基于BatchCompute搭建。国产动画片如小门神、昆塔等，利用BatchCompute调用阿里云日夜不停进行影片视效渲染，极大提高了渲染效率。



### 生物数据分析

生物基因企业利用BatchCompute完成大规模基因组测序分析。

