# 机器学习PAI

## 产品定价

# 产品定价

## 购买开通流程

登录阿里云账号，进入机器学习产品页面，单击**立即购买**按钮。


### 购买并开通


PAI目前有两个购买菜单，**PAI后付费**包含PAI-STUDIO(可视化建模)、PAI-DSW(Notebook建模)、PAI-EAS(模型在线服务)的一键式开通。**PAI-DSW预付费**指的是DSW按照GPU卡数包年包月购买，新用户建议选择**PAI后付费**。

详细价格请参见云产品定价。

购买后单击下图中的**管理控制台**。


进入如下页面：

# 创建PAI-STUDIO项目

在左侧列表选择**可视化建模**进入PAI-STUDIO控制台，单击**创建项目**。


右侧出现项目创建弹窗，需要按照指示依次完成账号实名认证、Accesskey创建以及MaxCompute的购买和开通。其中MaxCompute的开通建议选择按量付费，与PAI付费模式保持一致。


创建完项目之后，单击**进入机器学习**即可开始使用。


# 创建PAI-DSW项目

在左侧列表选择**Notebook建模**，单击**创建实例**。


选择付费类型以及卡的个数，即可创建实例。

注：pai.medium.2xp100，表示两张p100 GPU卡。预付费需要购买预付费对应的卡数才可以建立。


# 产品模块定价与计费

# PAI-Studio计费

PAI-Studio可视化建模平台提供百余种常规机器学习算法组件，包含数据预处理、特征工程、统计分析、机器学习、深度学习、时间序列、文本分析、网络分析、工具、金融板块等种类。

收费Region：

- 目前针对常规机器学习算法组件，全Region收费。

- 深度学习组件只支持华北2、华东2两个区域。

计费方式：

- 按量后付费：账单价格= 计算时*单价

计算时概念：

- 计算消耗的CPU core数量及内存量。
- 1个计算时的单位为1 CPU Core +4GB 内存。
- 计算时的计算方法为 Max(CPU*时长*，*内存*时长/4)。例如一个用户使用了2 Core ，5GB运行1 小时，计算时为=Max(2*1,5*1/4)=2。如果一个用户使用了1 Core，5GB运行1小时，计算时为=Max(1*1,5*1/4)=1.25

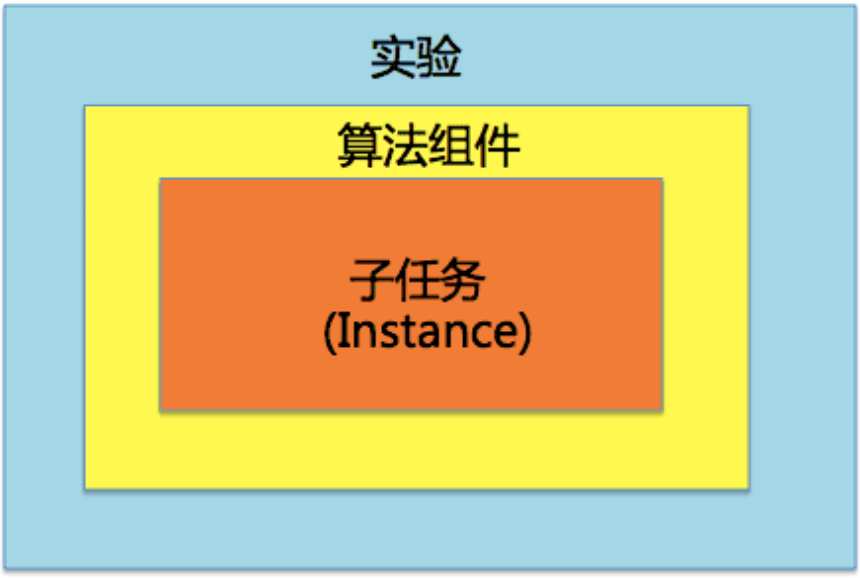
注意：SQL脚本、JOIN、UNION、过滤与映射四个组件底层执行单元为MaxCompute SQL，使用过程中可能会产生MaxCompute账单，详细请参考MaxCompute关于SQL的计费：MaxCompute计费详情

# 单价列表

| 账单种类 | 包含的算法模块 | 价格 | 支持区域 |
| --- | --- | --- | --- |
| 数据预处理 | 数据预处理、特征工程 | 1元/计算时 | 全区域 |
| 数据分析 | 统计分析、机器学习、时间序列、网络分析、金融板块 | 1.3元/计算时 | 全区域 |
| 文本分析 | 文本分析 | 1.7元/计算时 | 全区域 |
| 深度学习 | Tensorflow、Caffe、MXNet相关组件 | 华北2：12元/卡/小时<br>华东2：8.4元/卡/小时 | 华北2、华东2 |
| default | Notebook、TensorBoard费用及其它未成功配置价格的算法费用（配置失败按照最低价格1元/计算时收费） | 1元/计算时 | 全区域 |

# 费用具体计算举例

通常PAI-Studio中的机器学习实验是由多个算法组件组合而成，每个算法组件又是由多个子计算任务（sub Instance）组合而成，如下图所示。




因此当我们需要计算某个实验的运算费用时，需要分别计算该实验中各个组件的费用。而计算各个组件的费用，则拆解为计算该组件下运行的各个子任务的费用，并累计总和。

下面我们就以新闻分类实验中的PLDA组件为例，介绍如何计算该组件的费用（计算整个实验的费用只需要把实验中各个组件的费用按同样方式计算累加即可，由于组件较多，在此不一一示范）

## 步骤一：定位算法组件类别和单价

首先我们要找到PLDA组件的组件类型，每个组件都可以在左边的组件列表中找到它的具体类别并定位到计价模块。

例如，下图中的PLDA组件属于文本分析类别。


随后我们可以根据这个组件类型，通过**单价列表**查看对应的组件单价。

例如，PLDA组件属于**文本分析**这个账单类型，所以价格是1.7元/计算时。于是PLDA组件及其所包含的子任务的计费单价都是1.7元/计算时。

# 步骤二：计算子任务费用并求和

上面我们查到，PLDA组件的单价是1.7元/计算时。

每次PLDA产生的总费用是由下图中所有子计算任务（sub Instance）产生的费用之和组成（日志中每一个蓝色的logview，对应的是一个sub instance子任务），且每个sub Instance的单价都是1.7元/计算时。


下面我们要做的就是找到各个子任务的计算时，然后按照：

费用= 计算时*单价

的公式进行计算，并累加各个子任务的费用。

## 1、计算时概念

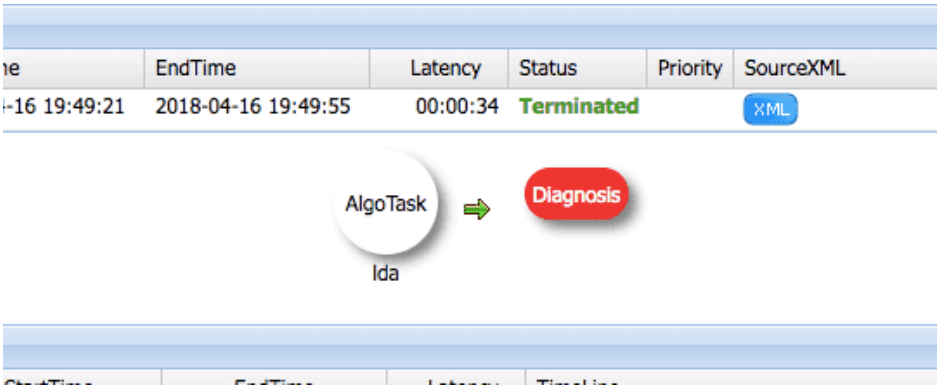
- 计算消耗的CPU core数量及内存量。
- 1个计算时的单位为1 CPU Core+4GB内存。
- 计算时的计算方法为：Max(CPU×时长，内存×时长/4)。
  例如，一个用户使用了2 Core，5GB运行1小时，计算时为Max(2×1,5×1/4)=2。如果一个用户使用了1 Core，5GB运行1小时，计算时为Max(1×1,5×1/4)=1.25。

## 2、查看子任务计算时

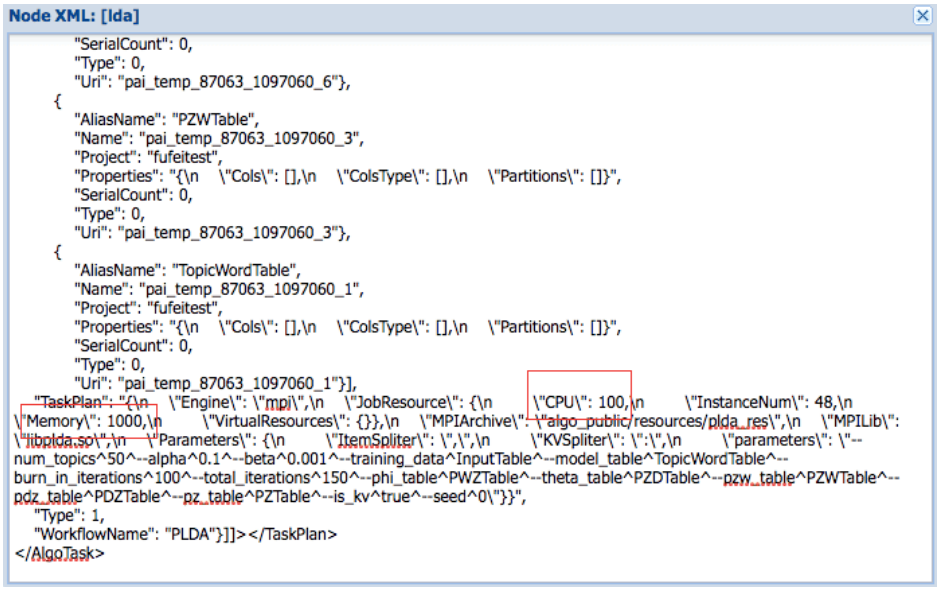
（1）查看PLDA组件的子任务，即sub instance组成。

右键单击PLDA组件，选择**查看日志**。日志中每一个蓝色的logview，对应一个sub instance。

（2）随机单击PLDA日志中的一个logview蓝色链接，本文以最后一个logview为例，单击**AlgoTask**，如下图所示。

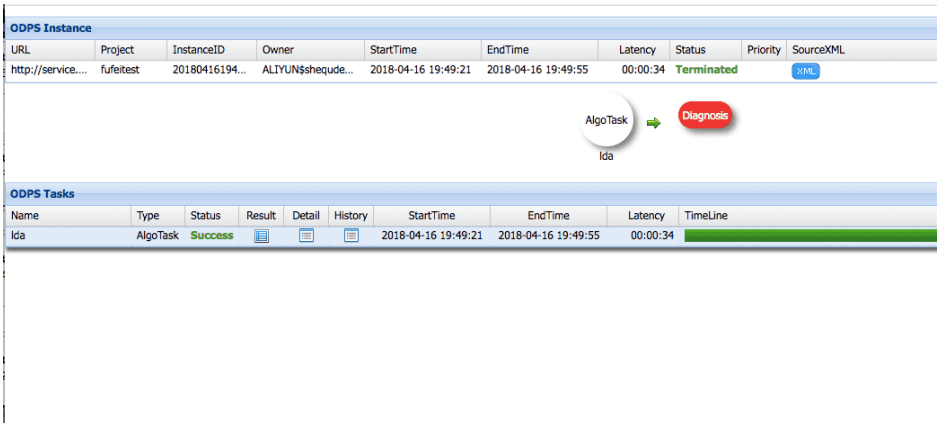



（3）在**TaskPlan**下找到**CPU**和**Memory**两个字段，如下图所示。

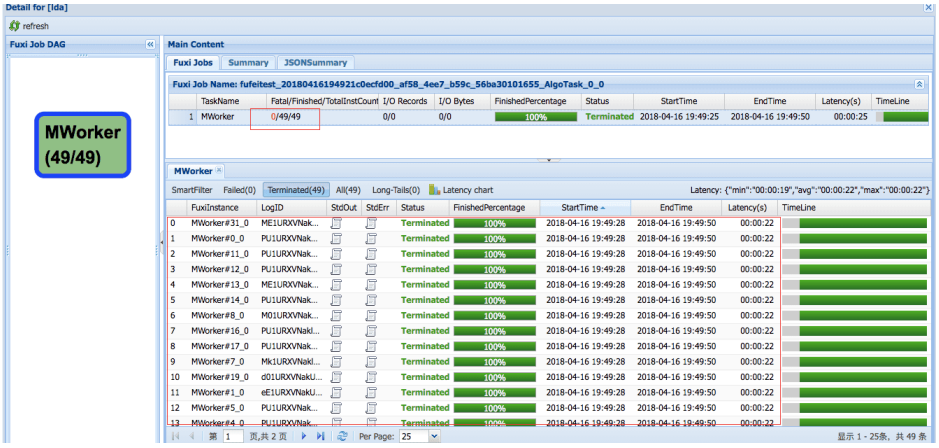



CPU数量除以100是每个计算instance分配的CPU core，Memory以MB为单位。

（4）双击下图中**ODPS Task**中的对象。



（5）双击**MWorker**对象，单击**Terminated**打开如下图所示界面，**Latency**表示每个worker的运行时间。




### 3、计算子任务计算时

- 根据上文（3）计算得到每个worker占用cpu为1core，占用内存memory约为1G。
- 根据上文（5）得到该sub instance一共有49个worker，每个worker时间约等于22s。

根据**计算时**概念计算出这个sub instance的计算时约为：
Max(CPU×时长，内存×时长/4)= Max(49×1×22,49×1×22/4)=1078计算秒=1078/3600计算时

### 3、计算子任务费用

当前sub instance总费用为：(1078/3600)计算时×1.7计算时/元=0.51元

### 4、累计求和各个子任务费用

上面我们求出了该PLDA组件下面的单个子任务的费用，累计PLDA组件下面所有子任务的费用就是此算法组件的总消费。进一步的，累计当前实验下的所有组件费用，就是当前实验的总消费。

## 其它

如果需要手动计算实验所产生的费用，可以通过导出账单明细来得到具体的资源消耗，具体请参考**账单明细查询**和**计价页面**。

# PAI-DSW计费

PAI-DSW提供云端开发环境，用户可以在云端实现代码开发以及自定义安装所需要的开源算法包。DSW提供资源可视化、内置案例、强大的计算引擎以及丰富的建模服务。

收费Region：

- 华东1、华东2、华北2、华南1、新加坡

区域选择：

- 预付费可用区域：华东2（NVidia Tesla M40 GPU ）、华北2（NVidia Tesla P100 GPU）
- 按量付费可用区域：华东1、华东2、华北2、华南1、新加坡

计费方式：

- 账单价格=计算时*单价

计算时概念：

卡数*使用小时数=计算时，

比如用户使用8卡作业运行0.5小时，那么总计算时为8*0.5=4计算时

注意：当DSW实例处于Running状态及开始计费，请及时停止不需要的实例，以免造成不必要的费用。

# 价格列表

## GPU价格

| GPU算力卡类型 | 计费类型 | 价格 | 可用区域 |
| --- | --- | --- | --- |
| V100 | 计算时（按量付费） | 29.1元/计算时 | 华东1、华东2、华北2、华南1 |
| P100 | 包年包月（预付费） | 2000元/卡/月 | 华北2 |
| P100 | 计算时（按量付费） | 12元/计算时（限时优惠） | 全区域 |
| T4 | 计算时（按量付费） | 15.4元/计算时 | 全区域 |
| M40 | 包年包月（预付费） | 500元/卡/月（限时优惠） | 华东2 |
| M40 | 计算时（按量付费） | 8.4元/计算时 | 华东2 |

## CPU价格

| CPU资源类型 | 计费类型 | 价格 | 资源规格 | 可用区域 |
| --- | --- | --- | --- | --- |
| pai.large.2core 4g | 计算时（按量付费） | 0.70元 | CPU2核4G | 全区域 |
| pai.xlarge.4core 8g | 计算时（按量付费） | 1.38元 | CPU4核8G | 全区域 |

| | | | | |
|---|---|---|---|---|
| pai.2xlarge.8core16g | 计算时（按量付费） | 2.76元 | CPU8核16G | 全区域 |
| pai.4xlarge.16core32g | 计算时（按量付费） | 5.48元 | CPU16核32G | 全区域 |
| pai.6xlarge.24core48g | 计算时（按量付费） | 8.22元 | CPU24核48G | 全区域 |

# PAI-EAS计费

PAI-EAS在线预测服务提供将常规机器学习模型/深度学习模型一键发布为Restful API的功能，支持用户通过http请求调用模型服务做实时预测，并且提供蓝绿部署、模型版本管理、在线调试等服务功能。

## 一、概述

使用EAS进行部署服务，首先要考虑的就是将服务部署在什么样的计算资源上。PAI-EAS提供将模型服务部署在公共资源组和专属资源组上两种方式。具体两种资源组的区别详情见文档：资源组使用介绍

对应于这两种部署使用的资源不同，EAS的计费方式也有不同，详见下表：

| 部署使用资源 | 计费主体 | 付费形式 | 计费规则 |
|---|---|---|---|
| 公共资源组 | 模型服务运行时长 | 后付费 | 该种计费适用于模型服务直接部署在公共资源组上的情况，此时该模型服务占用多少公共计算资源多长时间，就产生多少按量后付费费用。 |
| 专属资源组 | 资源组运行时长 | 预付费 | 该种情况适用于用户将模型服务部署在专属资源组上的情况，由于专属资源组是提前下单购买的，此时用户只需要为资源组付费即可，部署在上面的模型服务不再产生计费。资源组购买提供包年包月预付费和按量使用后付费两种购买形式 |
| | | 后付费 | |

如果以上两种部署方式对应的资源都使用了，那么EAS产生的总费用为所有费用之和。下面介绍两种部署产生费用的具体计算方式。

# 二、公共资源组部署计费

## 1、基本规则

按照上述表格的计费规则介绍，使用公共资源组部署情况下，费用计量是按照部署的模型服务的运行时长（即占用公共资源的时长）来计量的。具体计算方式为：

每个模型服务费用 = 部属资源数量 * 部属资源单价 * 使用时长

- 模型一旦部署并处于running状态就会开始计费，请切记及时停止无用的模型服务，以免造成不必要的费用开销
- 计费时间起点为模型（即对应占用资源）开始运行（状态转为Running）时间，计费时间终点为模型（即对应占用资源）停止（状态转为Stopped）时间。
- 模型扩容后，对应的新资源使用时长从扩容成功时刻开始计算。模型缩容后，被释放的资源从释放成功后停止计费，剩余资源继续计费。
- 使用时长精确到分钟，不足1分钟的时长舍弃不计费。

其中部署资源单价可以见下表，表中1Quota=1核+4GB内存：

| 部署资源类型 | 单价 | 区域 |
| --- | --- | --- |
| CPU服务 | 0.4元/Quota/小时 | 华东1、华东2、华北2、华南1、新加坡 |
| GPU服务（P4卡） | 8元/P4Quota/小时 | 华东1、华东2、华北2、华南1、新加坡 |

注：上表中公共资源组GPU(P4)资源即将下线，请合理安排使用。

## 2、计费举例

假设A用户在华东1Region部署了一个模型服务，初始占用资源为2Quota(2核+8GB内存)，9:00完成部署并服务进入Running状态，10:00用户完成缩容，占用资源减少为1Quota(1核+4GB内存)，11:00用户完成扩容，占用资源增加为4Quota(4核+16GB内存)，12:00用户完成停止服务并状态变为Stopped态。那么在这种情况下，用户使用产生的费用为：

部属资源数量 * 部属资源单价 * 使用时长 = 2(Quota) * 0.4(元) * 1(h) + 1(Quota) * 0.4(元) * 1(h) +4(Quota) * 0.4(元) * 1(h) = 2.8 元

# 三、专属资源组部署计费

用户将模型服务部署在专属资源组上时，由于专属资源组是提前下单购买的，此时用户只需要为资源组付费即可，部署在上面的模型服务不再产生计费。资源组购买提供包年包月预付费（预付费专属资源组）和按量使用后付费（后付费专属资源组）两种购买形式。

# 1、预付费专属资源组

## （1）基本规则

每个资源组费用 = 购买资源数量 * 资源单价 * 购买时长

- 购买时长范围可以选择：1月～12月
- 购买时长从购买次日开始计起（即购买当日为赠送免费使用时间），次日起往后推30天为一个月。例：用户7月31日下单购买一个月，则从8月1日开始计算日子，8月31日00:00资源组到期。
- 部署资源单价见下表，其中“国内Region”包括：华东1、华东2、华北2、华南1。其中部分机器资源在部分区域可能存在短期无货无法购买。

| 机器型号 | GPU配置 | CPU配置 | 国内Region单价(每机器每月) | 新加坡Region单价(每机器每月) |
| --- | --- | --- | --- | --- |
| ecs.c5.6xlarge | / | 24核48G | 2360元 | 3850元 |
| ecs.g5.6xlarge | / | 24核96G | 3200元 | 4810元 |
| ecs.gn5i-c4g1.xlarge | 1 * Nvidia Tesla P4卡 | 4 核16G | 2920元 | 无 |
| ecs.gn5i-c8g1.2xlarge | 1 * Nvidia Tesla P4卡 | 8 核32G | 3510元 | 无 |
| ecs.gn6i-c4g1.xlarge | 1 * Tesla T4卡 | 4 核15G | 3683元 | 4697元 |
| ecs.gn6i-c8g1.2xlarge | 1 * Tesla T4卡 | 8 核31G | 4435元 | 5570元 |
| ecs.gn6i-c16g1.4xlarge | 1 * Tesla T4卡 | 16核62G | 5198元 | 7317元 |
| ecs.gn6i-c24g1.6xlarge | 1 * Tesla T4卡 | 24核93G | 5445元 | 9172元 |
| ecs.gn5-c4g1.xlarge | 1 * NVIDIA P100卡 | 4 核30G | 4049元 | 6646元 |
| ecs.gn5-c8g1.2xlarge | 1 * NVIDIA P100卡 | 8 核60G | 4876元 | 8004元 |
| ecs.gn5-c28g1.7xlarge | 1 * NVIDIA P100卡 | 28核112G | 7565元 | 11517元 |
| ecs.gn6v-c8g1.2xlarge | 1 * NVIDIA V100卡 | 8 核32G | 8382元 | 无 |

## （2）计费举例

假设用户A购买了华东1Region的4核15G GPU T4卡2张，购买时长为3个月，则购买费用为：

购买资源数量 * 资源单价 * 购买时长 = 2张 * 3683元 * 3个月 = 22098 元

# 2、后付费专属资源组

## （1）基本规则

每个资源组费用 = 购买资源数量 * 资源单价 * 实际使用时长

- 资源组一旦创建成功并处于running状态就会开始计费，请切记及时停止无用的资源组，以免造成不必要的费用开销
- 计费时间起点为资源组开始运行（状态转为运行中）时间，计费时间终点为资源组中所有机器被缩容释放掉后的（状态转为无机器）时间。
- 资源组扩容后，对应的新资源使用时长从扩容成功时刻开始计算。资源组缩容后，被释放的资源从释放成功后停止计费，剩余资源继续计费。
- 使用时长精确到分钟，不足1分钟的时长舍弃不计费。
- 各类资源单价见下表，其中“国内Region”包括：华东1、华东2、华北2、华南1。
  注1：其中部分机器资源在部分区域可能存在短期无货无法购买。
  注2：为方便用户直观查看，下表的单价为小时价。实际计费为分钟价计费，由于单位转换会存在略微差价，请以实际出账为准。

| 机器型号 | GPU配置 | CPU配置 | 国内Region单价(每机器) | 新加坡Region单价(每机器) |
| --- | --- | --- | --- | --- |
| ecs.c5.6xlarge | / | 24核48G | 4.4元/小时 | 7.13元/小时 |
| ecs.g5.6xlarge | / | 24核96G | 6.6元/小时 | 10.69元/小时 |
| ecs.gn5i-c4g1.xlarge | 1 * Nvidia Tesla P4卡 | 4 核16G | 10.66元/小时 | 无 |
| ecs.gn5i-c8g1.2xlarge | 1 * Nvidia Tesla P4卡 | 8 核32G | 12.84元/小时 | 无 |
| ecs.gn6i-c4g1.xlarge | 1 * Tesla T4卡 | 4 核15G | 12.79元/小时 | 9.79元/小时 |
| ecs.gn6i-c8g1.2xlarge | 1 * Tesla T4卡 | 8 核31G | 15.40元/小时 | 11.61元/小时 |
| ecs.gn6i-c16g1.4xlarge | 1 * Tesla T4卡 | 16核62G | 18.05元/小时 | 15.25元/小时 |
| ecs.gn6i-c24g1.6xlarge | 1 * Tesla T4卡 | 24核93G | 18.91元/小时 | 19.11元/小时 |
| ecs.gn5-c4g1.xlarge | 1 * NVIDIA P100卡 | 4 核30G | 14.06元/小时 | 13.84元/小时 |
| ecs.gn5-c8g1.2xlarge | 1 * NVIDIA P100卡 | 8 核60G | 16.93元/小时 | 16.67元/小时 |
| ecs.gn5-c28g1.7xlarge | 1 * NVIDIA P100卡 | 28核112G | 26.27元/小时 | 23.99元/小时 |
| ecs.gn6v-c8g1.2xlarge | 1 * NVIDIA V100卡 | 8 核32G | 29.11元/小时 | 无 |

### （2）计费举例

假设用户A购买了华东1Region的24核96G CPU 2台，使用时长为45分钟，则使用费用为：

购买资源数量 * 资源单价 * 实际使用时长 = 2台 *（6.6 / 60）元/分钟 * 45分钟 = 9.9 元

# PAI-AutoLearning计费

当前PAI-AutoLearning处于免费公测阶段，无价格列表。

注意：PAI-AutoLearning训练完成模型后，若想要部署为EAS在线服务才会产生费用，费用价格请参考PAI-EAS价格列表。

# 查看账单

阿里云机器学习商业化账单可以通过以下步骤查看。在使用机器学习期间所产生的费用，以账单最终呈现方式为准。账单产生规则为t+1天，每天会统计当日消费，然后在下一日的账单中显示前一日的具体消费明细。具体的账单查看方法如下：

登陆阿里云官网，单击用户名。

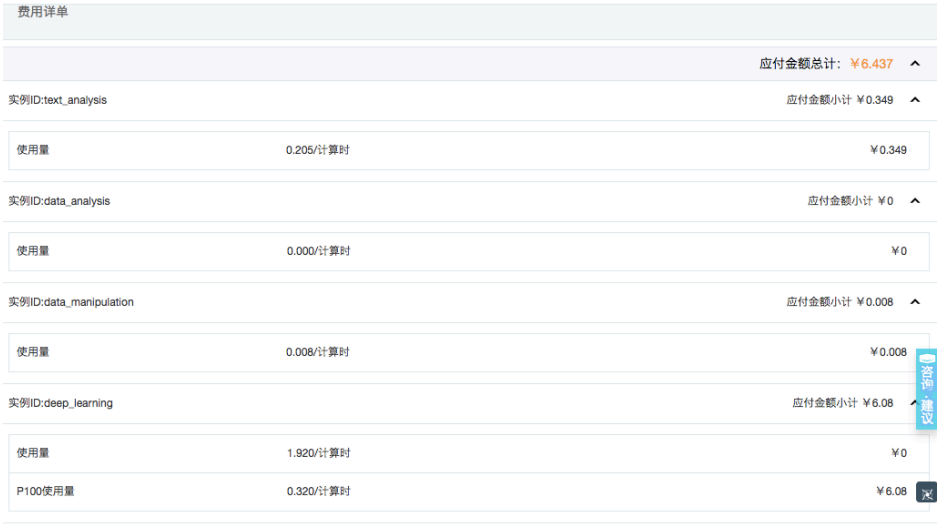

单击界面上方的**费用**。


在左侧列表选择**消费记录>消费明细**，在右端产品中选择**机器学习**，并根据实际情况选择**支付状态**和**账期**，单击**查询**，即可看到具体的消费明细。

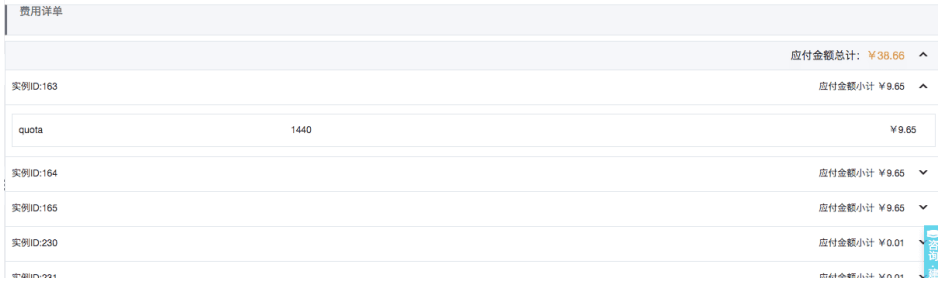

单击某一条消费记录的**详情**，查看消费明细。


在消费明细中可以看到用户在机器学习平台不同功能模块下的当日消费汇总，具体计算方式请参考阿里云机器学习价格计算方法。

模型训练账单详情：




在线预测服务账单详情：




- 实例ID：对应模型在线部署的模型ID号。
- quota数值为模型消耗的quota数量乘以分钟，比如模型A部署占用3quota，部署了20分钟，则quota项数值为3*20=60。

# 查看消费明细

## 查看模型训练消费明细

本节为您具体介绍每一个实例ID的费用是如何产生的。

付费购买使用阿里云机器学习产品后，每日都会产生账单，账单金额为当日累计用量产生的费用，账

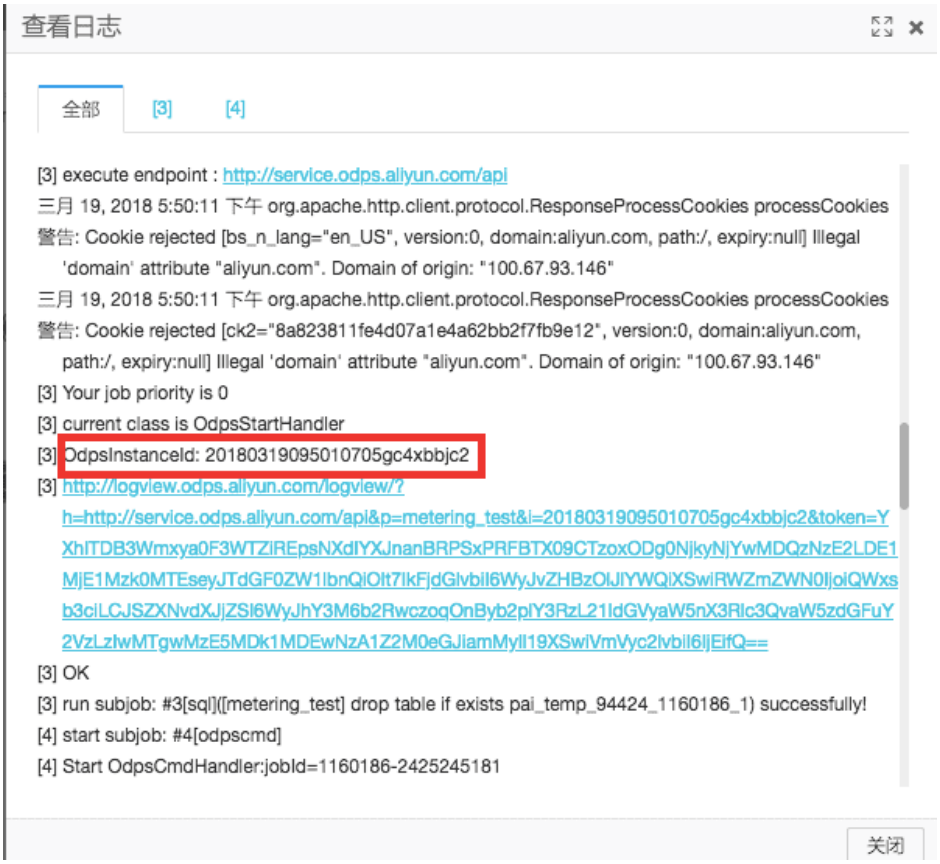

单示例图如下。

机器学习的使用记录可以通过费用中心的以下路径下载。


通过以上页面导出存储计量明细的CSV文件，文件格式如下图所示。

```
"UserId","ProductId","MeteringId","CUUsage","XflowName","XflowCategory","M40GPUUsage","P100GPUUsage"
"1884692660043716","PAI","20180301015514303g7qwwzh8","6894","tensorflow_ext121_allreduce","deep_learning",,"1149"
"1884692660043716","PAI","20180312064929353g0b0w2i8","924","PLDA","text_analysis",,
```

其中，**MeteringId**对应的是机器学习实验运行中产生的InstanceID，可以在组件运行日志中查看：




通过**xflowCategory**区分作业类型。

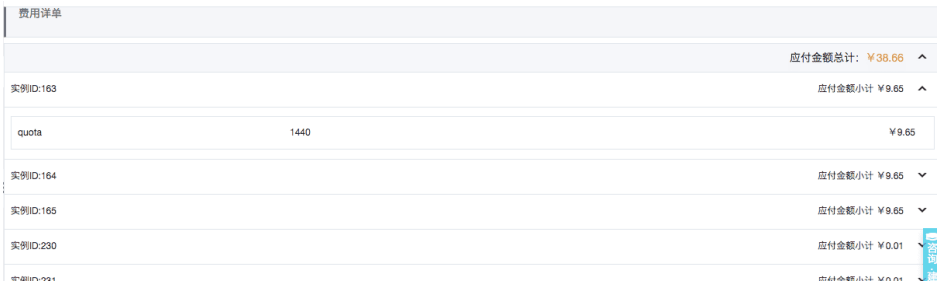
- 如果是deep_learning，则通过**P100GPUUsage**或**M40GPUUsage**除以3600计算资源消耗单元。
- 否则通过**CUUsage**除以3600计算资源消耗单元。

具体价格计算方式：

- 示例CSV文件中的第二行是深度学习作业，P100GPUUsage对应的是华北2的深度学习服务，那么计费价格是1149/3600×单价×折扣。
- 示例CSV文件中的第三行是机器学习作业，是text_analysis，也就是文本算法组件，那么计费价格是924/3600×单价×折扣。

# 查看在线预测消费明细

在线预测账单详情：



在线预测账单详情下载：


收支明细
消费记录
消费总览
消费明细
使用记录
实例消费明细
月度成本消耗
导出记录
存储到OSS
账单分析
产品账单分析
导出说明：
1. 导出文件格式为CSV，您可以使用Excel等工具查看。
2. 如果导出文件中有错误提示，请按照提示重新操作。
3. 如果导出记录过大，文件可能会被截断，请修改导出条件并重试。
产品：
PAI(EAS)
使用期间：
2018-10-22
至
2018-10-23
计量粒度：
小时
验证码：
看不清楚，换一张
导出CSV