

解决方案实践

华为云 D-Plan AI 解决方案实践

文档版本 1.0
发布日期 2023-12-07



版权所有 © 华为技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

<https://www.huawei.com/cn/psirt/vul-response-process>

如企业客户须获取漏洞信息，请参见如下网址：

<https://securitybulletin.huawei.com/enterprise/cn/security-advisory>

目录

1 方案概述	1
2 资源和成本规划	3
3 准备工作	4
4 实施步骤	5
4.1 开始使用.....	5
4.2 快速卸载.....	9
4.3 定时调用.....	9
5 附录	11
6 修订记录	12

1 方案概述

应用场景

华为云销量预测-社区团购电商销量预测提供了一个开箱即用的是社区团购类电商的工业化销量预测解决方案。社区团购类电商的模式为：消费者今日线上下单，明日便可线下自提，为合理安排时间进行商品加工与物流配送，如何准确且高效得预估当天销量并指导配送与加工是其中一个重要的问题。

该解决方案为企业业务人员提供了一个可视化的操作界面来扩展对机器学习（ML）的访问，只需简单配置，就能够自行生成准确的销量预测模型，而无需编写任何代码；而对于初阶的算法工程师，该方案还提供了对应的高阶功能，允许进行模型调参，从而达到更佳模型效果。

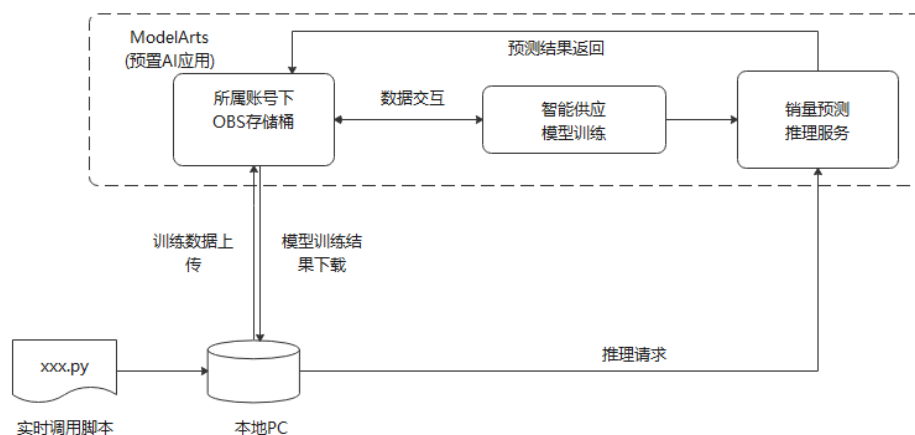
该解决方案适合如下场景的销量预测：

- 生鲜销量预测
- 线上超市销量预测

方案架构

该解决方案基于华为云ModelArts搭建了一套端到端的销量预测 workflow，考虑企业级客户的工业化部署使用，该解决方案支持定时调用功能与资源统一管理功能，整体架构如图1所示。

图 1-1 方案架构图



该解决方案会部署如下资源：

- 创建用于上传训练数据和结果数据导出的OBS桶，企业业务系统定时向该桶上传更新数据和下载结果数据；
- 函数工作流：用于实现调用销量预测服务的业务逻辑，完成模型的自动部署。
- 销量预测服务：提供分时销量预测服务，可灵活调整预测时间点，根据历史销量、商品属性、促销活动等基础信息训练得到准确的预测模型。

方案优势

行业化建模经验

- 内置社区团购类销量预测行业化建模经验，有效提高模型预测准确率。

降本增效

- 支持多个时间点的算法滚动预测，提前感知销量，提升整体时效，有效指导加工配送。
- 算法稳定，辅助人工预测，降低人力投入。

简单易用

- Workflow流水线式交付，只需简单的操作，就可以自行对模型做迭代，并方便对多模型进行统一管理，让模型在各种场景上都具有扩展性。
- 定时预测：支持公有云部署模型部署方式，您可以在本地服务器定时调用，实现模型定时训练更新与定时预测。
- 支持二次开发。

约束与限制

- 该方案当前仅支持华北-北京四区域。在开始之前，请确认您已经拥有一个可以访问该区域的华为云账号，并完成实名认证，帐号不能处于欠费或冻结状态。
- 该方案仅支持CSV格式数据。
- 数据要求:必须提供ds时间列和target预测目标列，其中时间列需符合标准的"%Y-%m-%d-%H-%M-%S" 时间列格式，数据中异常值需提前处理，其余特征列列名需映射到解决方案支持的标准列名。
- 卸载解决方案前，请先确保OBS桶中无数据，否则解决方案将卸载失败。

该解决方案暂时不支持OBS上传KMS加密文件。

2 资源和成本规划

表 2-1 资源和成本规划

华为云服务	计费说明	数量	每月费用（元）
销量预测服务	模型训练：按需计费 推理服务：按需计费	1	按需计费
对象存储服务	数据存储（多AZ存储）：按需计费 请求费用：按需计费	1	预计每月新增1GB数据量，按需计费。该方案存储费用消耗较低，详细请参考每月账单。
函数工作流	请求次数0-100万次/月：免费 计量时间： 0-400,000 GB/秒： 免费 节点执行次数： 0-5000次：免费	1	0
总计		1	按需计费

3 准备工作

- [注册华为云帐号](#)并完成[实名认证](#)。
- 使用华为云帐号登录[ModelArts控制台](#)，选择“华南-广州”区域，[配置访问授权](#)。

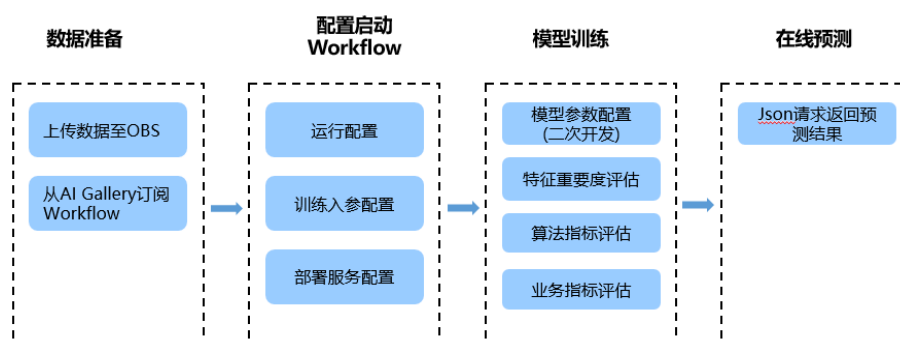
登录[OBS管理控制台](#)，选择“华南-广州”区域，[创建OBS桶](#)用于存储数据。

4 实施步骤

- 4.1 开始使用
- 4.2 快速卸载
- 4.3 定时调用

4.1 开始使用

图 4-1 图示 1



步骤1 准备数据

使用自己的数据集，直接将数据上传至OBS文件夹中。具体操作，请参见[上传文件](#)。

步骤2 订阅销量预测 workflow

1. 打开[Workflow社区团购销量预测 workflow](#)页面，单击“订阅”。
2. 订阅完成后，单击“已订阅”旁的“运行”。
3. 在弹出的“从AI Gallery导入 workflow”对话框中配置一下参数。
 - “资产版本”：选择最新版本。
 - “云服务区域”：选择训练数据对应region区域。

📖 说明

确保导入页面URL（“https://console.huaweicloud.com/modelarts/?region=cn-north-4#/ai-workflow/subscription/6fc22001-c778-46a7-99f9-154ebb7a9579”）中的region取值为cn-south-1，若为其他区域，请手动修改为cn-south-1，刷新页面后再进行导入。

4. 单击“导入”。
页面自动跳转至ModelArts管理控制台Workflow运行总览页面。

步骤3 配置Workflow

1. 登录**ModelArts管理控制台**，在控制台顶部区域下拉框中，选择“华南-广州”。
 2. 在左侧导航栏，选择“Workflow”。
 3. 在Workflow列表中，单击Workflow所在行操作列的“配置”。
- Workflow配置中各个节点的详细配置参数信息，请参见下表：

表 4-1 Workflow 详细配置参数信息

配置节点	配置项
运行配置	<ul style="list-style-type: none">- 输出目录：选择一个OBS空目录存储训练输出的模型。
资源配置	<p>各个训练节点资源可参见如下配置：</p> <ul style="list-style-type: none">- 选择CPU：2 核 8GB
消息通知	<p>订阅消息使用消息通知服务，在事件列表中选择需要监控的节点和Workflow状态，在事件发生时发送消息通知。您可以根据实际情况设置是否打开开关，如果开启订阅消息，请根据实际情况填写如下参数。</p> <ul style="list-style-type: none">- 主题名：订阅消息主题名称。您可以单击创建主题，在消息通知服务中创建主题。- 订阅对象：单击“增加订阅消息”，选择你需要的“订阅对象”和“订阅事件”。- 使用订阅消息服务会产生相关服务费用，详细信息请参见资费说明。- 如果您订阅了节点事件，同时也订阅了Workflow事件，消息通知会重复发送。

配置节点	配置项
模型训练	<p>节点中各参数取值如下：</p> <ul style="list-style-type: none">- pred_hour：范围：[14,16,18]，- combined_params_list：参数集合名称,格式为{"params": [pred_hour_1, ...pred_hour_n]}参数示例：{"params": [14,16]}- 参数说明：combined_params_list 为json类型，key为"params"，内容为list，若想同时跑多个时间点数据，即可传入多个list，若各个时间点数据不同，可在对应时间点后指定具体路径。 <p>注意</p> <ol style="list-style-type: none">1. 若想新增pred_hour，需在 config.py中进行相应配置；2. combined_params_list模型为空，若“combined_params_list”：设置了值，则pred_hour不生效；若未设置，则运行pred_hour对应时间点模型；
模型训练	<p>若需修改模型参数重新训练模型，可在对应的数据目录（data_url）下，上传config.py文件，修改其中对应参数。</p>
模型注册	<p>model_name：模型参数名称。</p>
部署服务	<p>model_name：模型参数名称，和模型注册部分的需保持一致， “service_name”：在线推理服务名称。</p>

4. 完成以上节点参数配置后，在Workflow页面右上角单击“保存配置”。

步骤4 启动Workflow

1. 在Workflow页面右上角单击“启动”，自动跳转至 workflow 运行总览页面。
2. 工作流开始从模型训练节点运行，您可以在运行状态区域，查看节点的运行状态，节点运行成功后自动运行下一个训练节点。
3. 部署服务的配置要在所有节点运行完成后，才能进行配置。“创建模型”节点运行成功后，单击“部署服务”节点，在部署服务区域进行参数配置。
 - 选择模型版本：选择模型对应的版本。
 - 节点规格：选择“CPU: 2核 8GB”。
 - 分流（%）：设置当前AI版本流量占比，服务调用请求根据该比例分配当前版本上。
 - 计算节点个数：设置当前AI版本应用的实例个数。
 - 环境变量：您可以根据业务需求增加环境变量。

- 是否自动启停：打开此开关，在线服务的运行时间将在您选择的时间点后，自动停止，同时服务计费停止。

4. 部署服务节点参数配置完成后，运行部署服务。

步骤5 在线预测

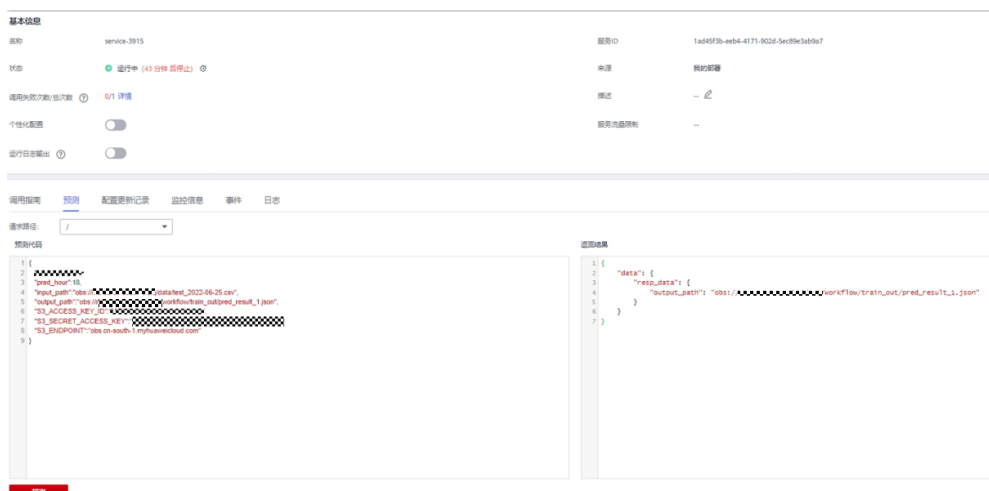
1. 部署服务运行成功后，单击“部署服务”区域右侧的“详情”，跳转至在线服务页面。

ModelArts支持代码调试和添加文件测试，根据AI应用定义的输入请求不同（JSON文本或文件），测试服务包括如下两种方式：JSON文本预测和文件预测。本Workflow以JSON文本预测的方式进行预测，更多信息，请参见[测试服务](#)。

2. 填写所需json在左侧文本框中，单击下方预测按钮，单击“预测”即可进行服务的预测。

3. 本Workflow预测结果以返回文件的形式自动返回到对应输出obs路径中(output_path)参数，若预测成功，将会在在“预测结果显示”区域显示返回路径。

图 4-2 图示 2



步骤6 模型评估

节点训练完成后，可以单击对应的训练节点，单击节点详情，查看对应的训练日志。您可以根据模型评估指标情况，在Workflow配置中修改模型节点参数，重新启动Workflow进行模型部署。

- 模型评估指标

本Workflow的模型评估指标使用的是MAPE，具体会打印出原始的xgboost和lightgbm的mape值，以及模型后处理过后和模型融合过后的模型评估MAPE指标，具体见下方截图。

- 业务评估指标

['时间', 'ptitle总数', '准确率', 'qty总数', '偏大数量', '偏大最大', '偏大总数', '偏小数量', '偏小最大', '偏小总数', '偏大比例', 'gmv损失比例', '偏小比例']

图 4-3 业务评估指标

```
[2022-08-21 19:14:50] [INFO] [*****evaluate*****]
[2022-08-21 19:14:50] [INFO] [lgb ori mae: 7.385826770078741 lgb mape: 0.23791722060135306
lgb ori mae: 8.527559055118111 xgb mape: 0.1626652778751398
lgb mae: 29.566929133858267 lgb mape: 0.7140512001231711
xgb mae: 36.40944881889764 xgb mape: 0.7665249839118135
mae: 5.377952755905512 mape: 0.17091785676611793
]
[2022-08-21 19:14:50] [INFO] [xgb_rresl_max,min,mean 5.0 -293.0 -10.866141730706662]
[2022-08-21 19:14:50] [INFO] [lgb_rresl_max,min,mean 97.0 -38.0 -4.023622045669292]
[2022-08-21 19:14:50] [INFO] [pred_rresl_max,min,mean 19.0 -76.0 -3.5196850393700787]
[2022-08-21 19:14:50] [INFO] [['时间', 'prctile总数', '预测数', 'qty总数', '最大数量', '最大最大', '最大总数', '最小数量', '最小最大', '最小总数', '最大比例', 'gm预测比例', '最小比例']]
[2022-08-21 19:14:50] [INFO] [['2022-06-26', 64, 0.7985399150217953, 2889, 36, 44.0, 240.0, 23, -9.0, -85.0, 0.08307372793354102, -0.042454914125340755, -0.029421945309795777]]
```

----结束

4.2 快速卸载

- 注册华为云帐号并完成实名认证。
- 使用华为云帐号登录ModelArts控制台，选择“华南-广州”区域，配置访问授权。

登录OBS管理控制台，选择“华南-广州”区域，创建OBS桶用于存储数据。

4.3 定时调用

- 定时预测

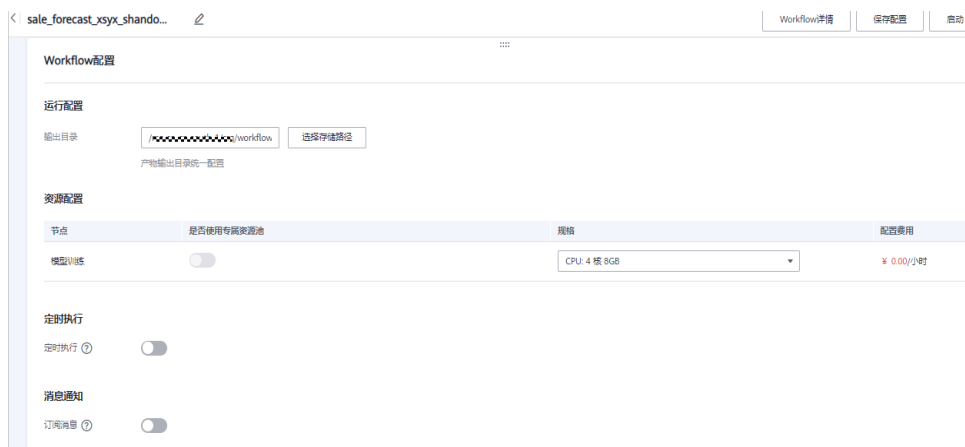
从console界面上获取到deploy_apipath，body体请求示例如下：

```
data_dict = {
    "pred_hour":18,
    "input_path":"obs://obs_path/test_xxx.csv", # 自定义，指定到csv
    "output_path":"obs://obs_path/pred_result.json",# 自定义，输出格式为json
    "S3_ACCESS_KEY_ID":"xxx",# ak
    "S3_SECRET_ACCESS_KEY":"xxx",# sk
    "S3_ENDPOINT":"obs.{region_name}.myhuaweicloud.com" # 对应region的endpoint
}
def infer(header_config, deploy_apipath, data_dict):
    body = json.dumps(data_dict)
    response = requests.post(deploy_apipath,
    headers=header_config, data=body, verify=False)
    print(response.text.replace("\n", "").replace('/', '/'))
    rep_json = response.json()
    return rep_json
infer(header_config, deploy_apipath, data_dict)
```

- 定时训练

为防止因数据分布引起模型精度下降，建议每天更新模型，只需在界面上将workflow配置建模的定时启动打开即可。

图 4-4 图示 3



5 附录

名词解释

- 基本概念、云服务简介、专有名词解释：
社区团购模式生鲜电商
- **对象存储服务**OBS：对象存储服务（Object Storage Service，OBS）是一个基于对象的海量存储服务，为客户提供海量、安全、高可靠、低成本的数据存储能力。
- **函数工作流**FunctionGraph：FunctionGraph是一项基于事件驱动的函数托管计算服务。使用FunctionGraph函数，只需编写业务函数代码并设置运行的条件，无需配置和管理服务器等基础设施，函数以弹性、免运维、高可靠的方式运行。此外，按函数实际执行资源计费，不执行不产生费用。

6 修订记录

表 6-1 修订记录

发布日期	修订记录
2022-09-07	第一次正式发布。