

解决方案实践

基于 TTS 快速实现文字转语音

文档版本 1.0.0
发布日期 2024-01-30



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 实施步骤	5
3.1 准备工作.....	5
3.2 快速部署.....	12
3.3 开始使用.....	18
3.4 快速卸载.....	21
4 附录	24
5 修订记录	26

1 方案概述

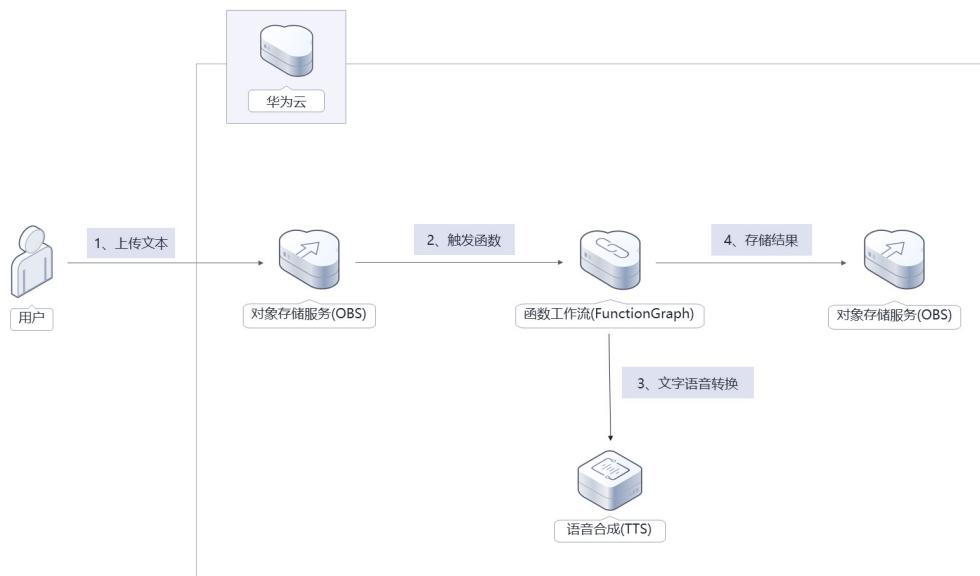
应用场景

该解决方案基于华为云**语音合成 TTS**构建，快速帮助用户在华为云控制台上完成语音合成功能的部署。TTS是指“Text-to-Speech”，即将文本转换为语音的技术。它使用先进的语音合成技术，将文本转换为可听的人工语音。通过TTS技术，计算机可以模拟人类的语音，实现自然、流畅、准确的语音输出。TTS技术广泛应用于语音助手、机器人、自动语音应答系统、有声读物等领域。

方案架构

该解决方案部署架构如下图所示：

图 1-1 方案架构图



该解决方案将会部署如下资源：

- 创建两个**对象存储服务 OBS**，一个用于上传需要合成语音的文本文件，另一个存放生成的音频文件。

- 创建一个[函数工作流 FunctionGraph](#)，用于调用TTS服务。
- 创建一个语音合成 TTS，用于实现文本转语音功能。

方案优势

- **快速部署**
一键轻松部署，即可完成函数工作流 FunctionGraph，对象存储服务 OBS资源的，语音合成 TTS的创建与配置。
- **多语种多音色**
华为云语音合成 TTS提供男女声、童声等38种音色，支持中英文、方言自由切换。
- **操作简单**
用户只需上传文本文件到指定的OBS桶，即可实现语音合成功能，得到相应的音频文件。

约束与限制

- 部署该解决方案之前，您需注册华为云账户，完成实名认证，且帐号不能处于欠费或冻结状态，请根据资源和成本规划中预估价格，确保余额充足。
- 请确保在部署解决方案前，使用的华为云帐号有IAM的足够权限，具体请参考[3.1-创建rf_admin_trust委托](#)。
- 确认租户配额充足：在“资源 > 我的配额”中查看以下配额是否充足，如配额不够，请提前提[工单](#)申请增加配额。

2 资源和成本规划

该解决方案主要部署如下资源，不同产品的花费仅供参考，具体请参考华为云官网[价格详情](#)，实际收费以账单为准。

表 2-1 资源和成本规划（按需计费）

华为云服务	配置示例	每月预估花费
语音合成 TTS	<ul style="list-style-type: none">按需计费：2元/千次（大于10千次/月）区域：华北-北京四计费模式：按需计费购买量：1	接口调用次数费用
函数 workflow FunctionGraph	<ul style="list-style-type: none">按需计费：1.33元/100万次（调用大于100万次）区域：华北-北京四计费模式：按需计费购买量：1	函数调用次数费用
合计	-	接口调用费用+函数调用次数费用

表 2-2 资源和成本规划（包年包月）

华为云服务	配置示例	每月预估花费
语音合成 TTS	<ul style="list-style-type: none">按需计费：18元/万次/年区域：华北-北京四计费模式：包年包月购买量：1	接口调用次数费用

华为云服务	配置示例	每月预估花费
函数工作流 FunctionGraph	<ul style="list-style-type: none">● 按需计费：1.33元/100万次（调用大于100万次）● 区域：华北-北京四● 计费模式：包年包月● 购买量：1	函数调用次数费用
合计	-	接口调用费用+函数调用次数费用

3 实施步骤

- 3.1 准备工作
- 3.2 快速部署
- 3.3 开始使用
- 3.4 快速卸载

3.1 准备工作

创建 rf_admin_trust 委托（可选）

步骤1 进入华为云官网，打开[控制台管理](#)界面，鼠标移动至个人帐号处，打开“统一身份认证”菜单。

图 3-1 控制台管理界面



图 3-2 统一身份认证菜单



步骤2 进入“委托”菜单，搜索“rf_admin_trust”委托。

图 3-3 委托列表



- 如果委托存在，则不用执行接下来的创建委托的步骤。
- 如果委托不存在时执行接下来的步骤创建委托。

步骤3 单击步骤2界面中的“创建委托”按钮，在委托名称中输入“rf_admin_trust”，委托类型选择“云服务”，选择“RFS”，单击“下一步”。

图 3-4 创建委托

委托 / 创建委托

* 委托名称

* 委托类型 普通帐号
将帐号内资源的操作权限委托给其他华为云帐号。
 云服务
将帐号内资源的操作权限委托给华为云服务。

* 云服务

* 持续时间

描述

0/255

步骤4 在搜索框中输入“Tenant Administrator”权限，并勾选搜索结果，单击“下一步”。

图 3-5 选择策略

委托“rf_admin_trust”将拥有所有策略

策略已选(1) 从其他区域或目录复制策略

名称	类型
<input checked="" type="checkbox"/> Tenant Administrator 全局云服务管理员 (全局IAM管理权限)	系统角色

步骤5 选择“所有资源”，并单击下一步完成配置。

图 3-6 设置授权范围

1) 选择策略 2) 设置最小授权范围 3) 完成

1 您需当前所选择的策略。系统提供以下授权范围方案。适用于您最小化授权。了解详情请点击应用列表中的授权范围方案

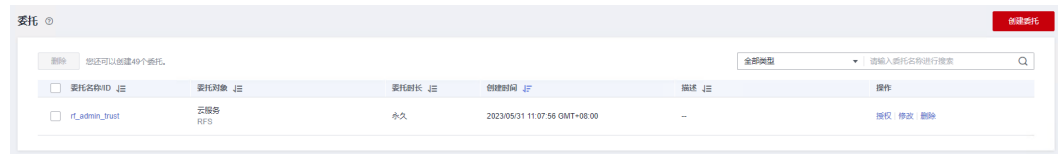
选择授权范围方案

所有资源
授权后，IAM用户可以管理权限范围内所有资源，包括企业项目、区域项目和全局服务资源。

[展开其他方案](#)

步骤6 “委托”列表中出现“rf_admin_trust”委托则创建成功。

图 3-7 委托列表

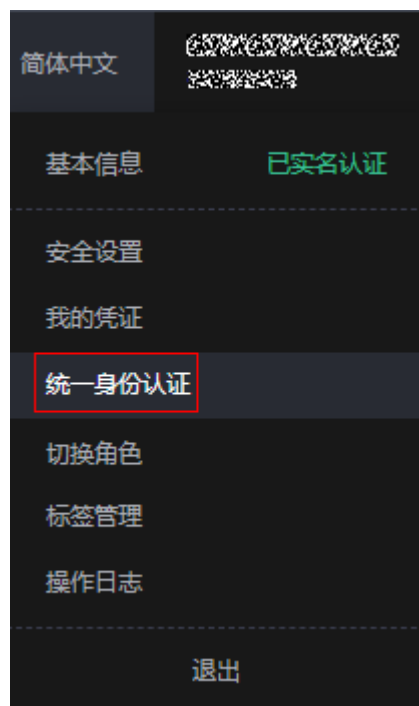


----结束

创建 IAM Agency Management FullAcces 策略（可选）

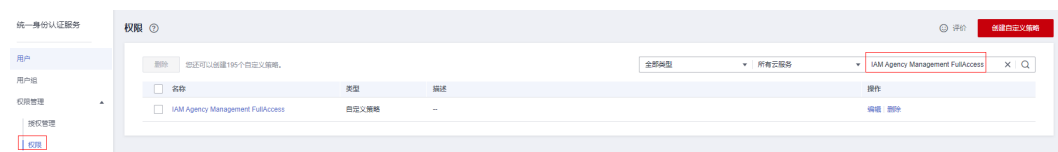
步骤1 打开“统一身份认证”菜单

图 3-8 统一身份认证菜单



步骤2 进入“权限管理”->“权限”菜单，在搜索框输入“IAM Agency Management FullAcces”当前帐号是否存在IAM委托管理权限

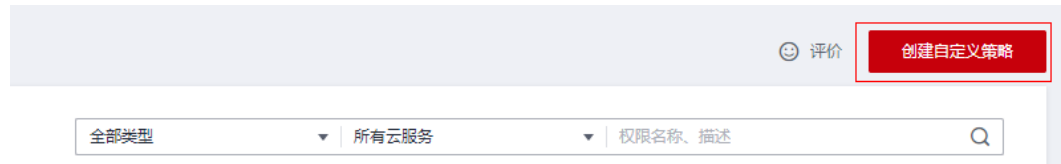
图 3-9 权限列表



- 如果搜索结果不为空，则当前帐号已经存在IAM委托管理权限，不需要重复创建
- 如果过搜索结果为空，则继续创建“IAM Agency Management FullAcces”权限

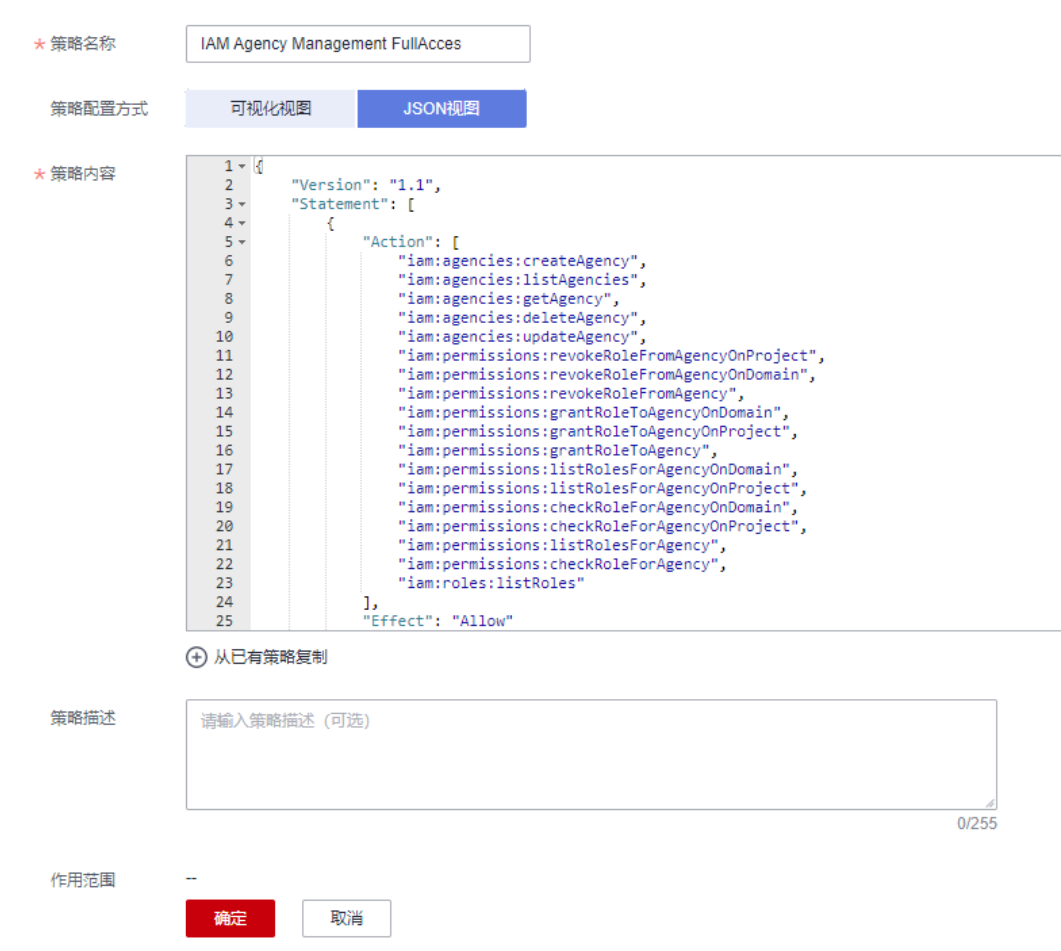
步骤3 单击“创建自定义策略”按钮

图 3-10 创建自定义策略



步骤4 输入策略名称为“IAM Agency Management FullAcces”，选择“JSON视图”，在策略内容中输入如下JSON代码，单击确认按钮

图 3-11 创建自定义策略



```
{  
  "Version": "1.1",  
  "Statement": [  
    {  
      "Action": [  
        "iam:agencies:createAgency",  
        "iam:agencies:listAgencies",  
        "iam:agencies:getAgency",  
        "iam:agencies:deleteAgency",  
        "iam:agencies:updateAgency",  
        "iam:permissions:revokeRoleFromAgencyOnProject",  
        "iam:permissions:revokeRoleFromAgencyOnDomain",  
        "iam:permissions:revokeRoleFromAgency",  
        "iam:permissions:grantRoleToAgencyOnDomain",  
        "iam:permissions:grantRoleToAgencyOnProject",  
        "iam:permissions:grantRoleToAgency",  
        "iam:permissions:listRolesForAgencyOnDomain",  
        "iam:permissions:listRolesForAgencyOnProject",  
        "iam:permissions:checkRoleForAgencyOnDomain",  
        "iam:permissions:checkRoleForAgencyOnProject",  
        "iam:permissions:listRolesForAgency",  
        "iam:permissions:checkRoleForAgency",  
        "iam:roles:listRoles"  
      ],  
      "Effect": "Allow"  
    }  
  ],  
}
```

```
        "iam:permissions:listRolesForAgencyOnDomain",  
        "iam:permissions:listRolesForAgencyOnProject",  
        "iam:permissions:checkRoleForAgencyOnDomain",  
        "iam:permissions:checkRoleForAgencyOnProject",  
        "iam:permissions:listRolesForAgency",  
        "iam:permissions:checkRoleForAgency",  
        "iam:roles:listRoles"  
    ],  
    "Effect": "Allow"  
}  
]  
}
```

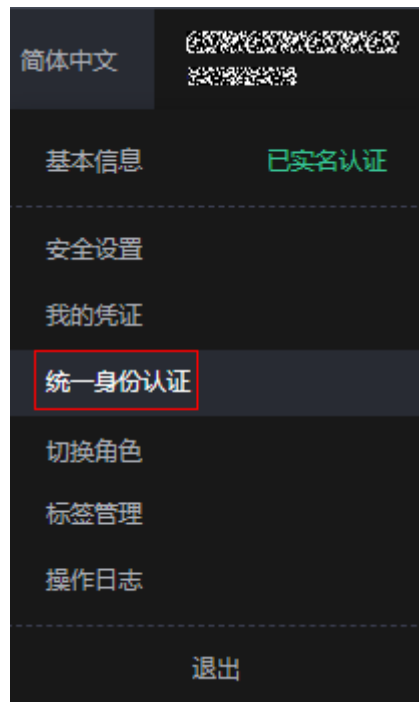
步骤5 界面无报错，则成功创建IAM Agency Management FullAcces权限

----结束

给 rf_admin_trust 委托添加 IAM Agency Management FullAcces 权限（可选）

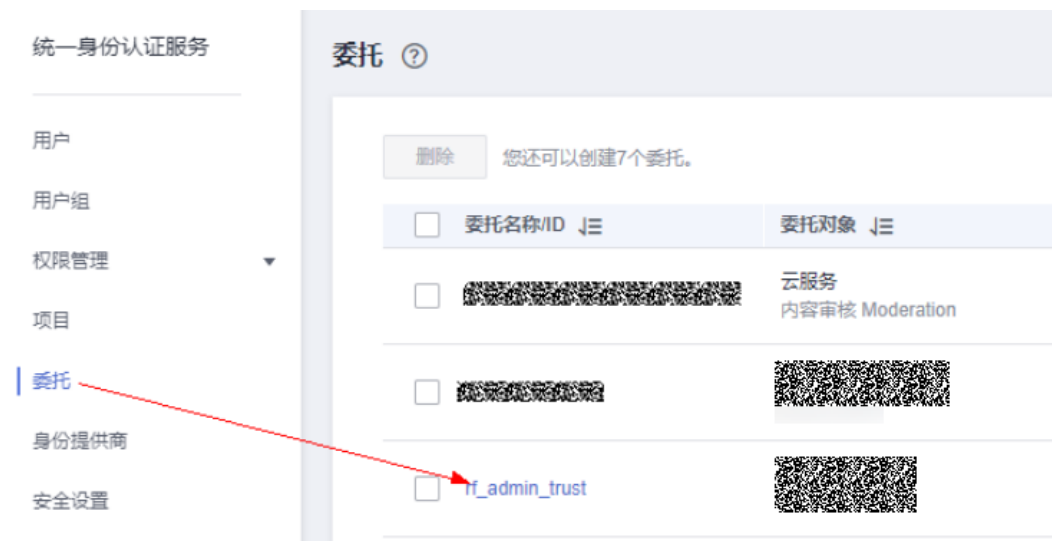
步骤1 打开“统一身份认证”菜单

图 3-12 统一身份认证菜单



步骤2 进入“委托”菜单，选择rf_admin_trust委托

图 3-13 委托列表



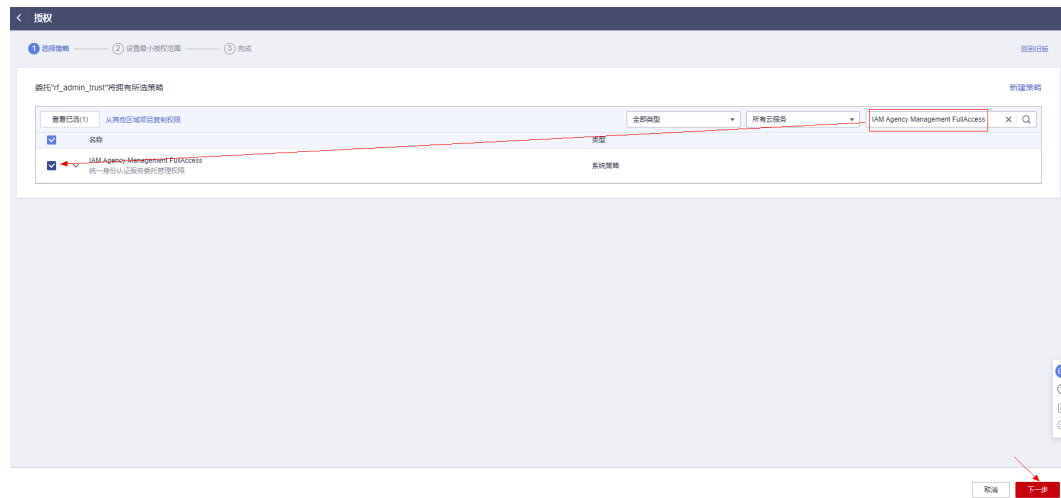
步骤3 进入“授权记录”菜单，单击“授权”按钮

图 3-14 授权记录



步骤4 在搜索框输入IAM Agency Management FullAcces，勾选过滤出来的记录，单击下一步，并确认完成权限的配置

图 3-15 配置 IAM Agency Management FullAccess 策略



步骤5 配置好后的情况：rf_admin_trust委托拥有Tenant Administrator和IAM Agency Management FullAccess权限

图 3-16 授权记录列表



----结束

3.2 快速部署

本章节主要帮助用户快速部署“基于TTS快速实现文字转语音”解决方案。

表 3-1 参数说明

参数名称	类型	是否可选	参数解释	默认值
functiongraph_name	string	必填	函数名称，用于定义创建函数，不支持重名。取值范围：2-57个字符，可包含字母、数字、_（下划线）和-（中划线），以字母开头，以字母或数字结尾。	text_to_speech-demo
text_bucket_name	string	必填	对象存储服务 OBS桶名称，全局唯一，用于上传文本文件。取值范围：3-63个字符，以小写字母或数字开头，仅支持小写字母、数字、中划线（-）、英文句号（.）。	空
result_bucket_name	string	必填	对象存储服务 OBS桶名称，全局唯一，用于存放结果音频文件。取值范围：3-63个字符，以小写字母或数字开头，仅支持小写字母、数字、中划线（-）、英文句号（.）。	空

参数名称	类型	是否可选	参数解释	默认值
timbre	string	必填	音色类型，支持选择38种音色。后缀不带普通字段的选项均为精品发音人，精品发音人不支持音高调节；带16k字段的选项仅支持16000赫兹的采样率。	标准女声_客服_普通
audio_format	string	必填	音频格式，支持wav、mp3、pcm三种格式。	mp3
sample_rate	string	必填	合成语音的采样率，支持16000赫兹和8000赫兹。	16000
speed	string	必填	合成语音的语速，取值范围-500到500。	0
pitch	string	必填	合成语音的音高，取值范围-500到+500。精品发音人不支持调节音高。	0
volume	string	必填	合成语音的音量，取值范围0到100。	50

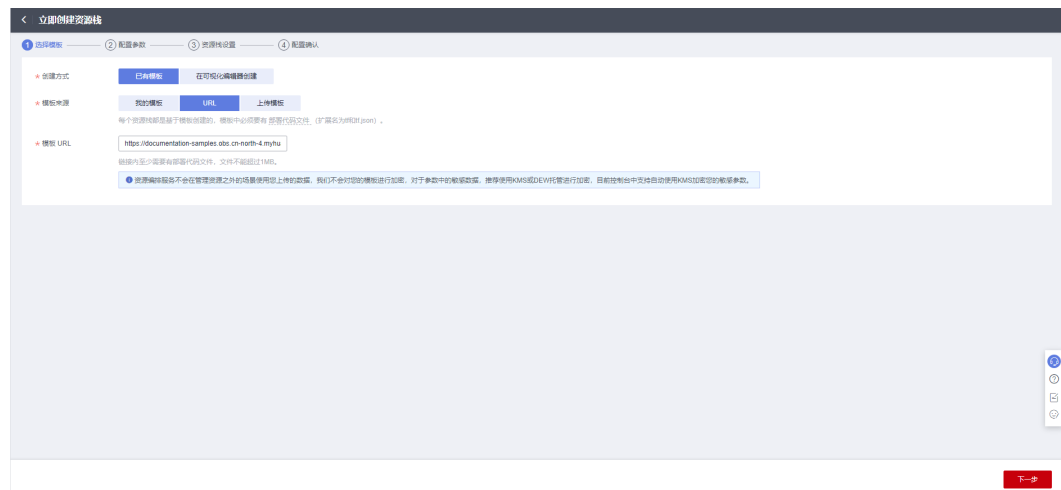
步骤1 登录华为云解决方案实践，选择“基于TTS快速实现文字转语音方案”。

图 3-17 解决方案实施库



步骤2 单击“一键部署”，跳转至该解决方案创建资源栈部署界面。

图 3-18 一键部署



步骤3 单击“下一步”，参考表1和附录-参数填写参考完成自定义参数填写。
text_bucket_name和result_bucket_name不能同名，否则可能造成方案功能出现问题。

图 3-21 创建执行计划



步骤6 单击“部署”，并且在弹出的执行计划确认框中单击“执行”。

图 3-22 部署执行计划



图 3-23 执行计划确认



步骤7 待出现“Apply required resource success”，表示该解决方案所有资源已经部署完成。

图 3-24 执行完成



----结束

3.3 开始使用

步骤1 在华为云控制台打开[对象存储服务 OBS](#)，打开上传文本的OBS（本案例中为text-to-speech-in，实际OBS名称以部署指定参数为准）。

图 3-25 打开上传文本的 OBS



步骤2 上传需要合成的文本文件（文件编码格式为utf-8，建议使用txt文件，文件名称只支持包含数字、英文、中划线-、下划线_，文本内容长度在500个字符以内）。

图 3-26 上传文本文件



步骤3 打开存放结果的OBS（本案例中为text-to-speech-result，实际桶名称以部署指定参数为准）。

图 3-27 打开存放结果的 OBS



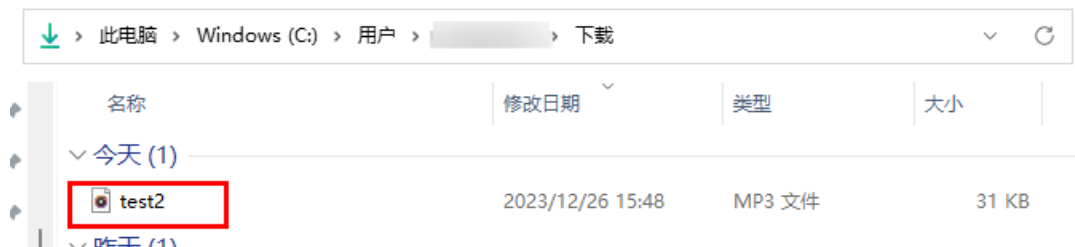
步骤4 下载已经合成完成的音频即可使用。

图 3-28 下载音频



步骤5 播放音频文件。

图 3-29 打开音频文件



----结束

改变合成音频配置

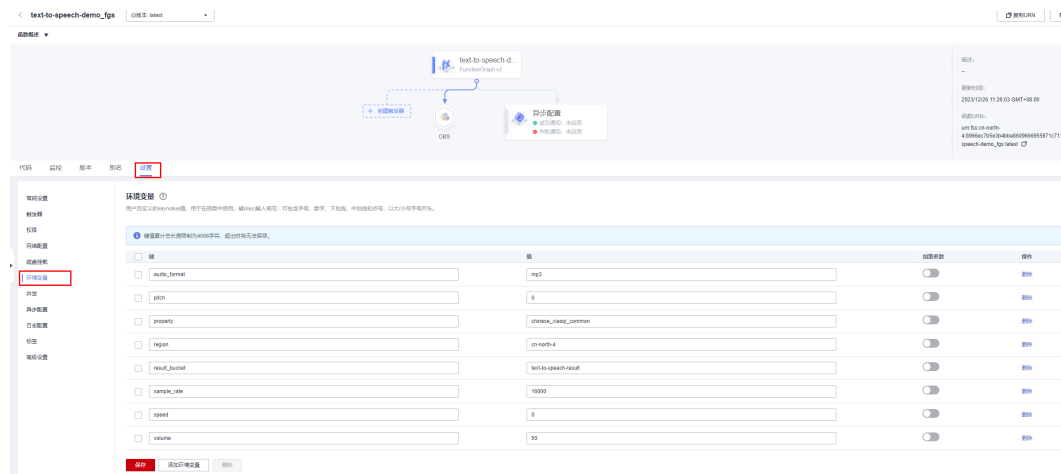
步骤1 在华为云控制台打开[函数工作流 FunctionGraph](#)，找到方案部署的函数。

图 3-30 打开指定函数



步骤2 打开函数后，在设置-环境变量下改变参数。参数填写的取值请参考[附录-参数填写参考](#)。

图 3-31 改变参数



步骤3 改变后使用该方案合成的音频会根据最新配置生成。

----结束

3.4 快速卸载

步骤1 进入上传文本的OBS和存放结果的OBS中，删除所有存放在里面的文件。

----结束

图 3-32 删除文本上传的 OBS 中的文件



图 3-33 删除存放结果的 OBS 中的文件



步骤1 登录[资源编排服务 RFS](#)资源栈，找到该解决方案创建的资源栈，单击资源栈名称最右侧“删除”按钮，在弹出的“删除资源栈”提示框输入“Delete”，单击“确定”进行解决方案卸载。

图 3-34 解决方案删除

资源名称ID	状态	图标	物理时间	更新时间	操作
stack_20231226_1125_7zmk	部署成功	-	2023/12/26 11:25:53 GMT+08:00	2023/12/26 11:26:10 GMT+08:00	删除

图 3-35 删除确认

删除资源栈

您确定要删除该资源栈及资源栈内资源吗？**删除后不能恢复，请谨慎操作**

资源栈名称	状态	创建时间
stack_20231226_1125_7zmk	部署成功	2023/12/26 11:25:53 GMT+08:00

资源列表 (5)

云产品名称	物理资源名称/ID	资源状态
函数工作流	text-to-speech-demo_fgs urn:fss:cn-north-4	生成完成
函数工作流		生成完成
统一身份认证服务	text-to-speech-demo_agency	生成完成
对象存储服务	text-to-speech-result	生成完成
对象存储服务	text-to-speech-in	生成完成

删除方式 删除资源 保留资源

如您确定要删除资源栈或其资源，请输入Delete以确认删除

----结束

4 附录

名词解释

基本概念、云服务简介、专有名词解释

- 函数工作流 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。
- 对象存储服务 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。
- 语音合成 TTS：是一种将文本转换为语音的技术。它使用先进的语音合成技术，将文本转换为可听的人工语音。主要提供如下集中关键的特性：
 - 语音合成
计算机可以模拟人类的语音，实现自然、流畅、准确的语音输出。
 - 语音模型训练
TTS 带有预训练的模型，是衡量数据集质量的工具，旨在在易于训练、速度和质量之间实现最佳权衡。

参数填写参考

- 音频格式 audio_format：wav、mp3、pcm
- 采样率 sample_rate：16000、8000
- 音色 timbre：
 - 普通发音人：标准女声_客服_普通、柔美女声_客服_普通、标准男声_电销_普通、热情女声_电销_普通、温柔女声_文学_普通、成熟女声_文学_普通、童声_普通、呆萌童声_普通、激昂男声_新闻播报_普通、标准女声_新闻播报_普通、英文女声_英文_普通。
 - 精品发音人：热情女声_电销、嗲柔女声_电销、严厉女声_电销、朝气男声_电销、利落男声_客服、柔美女声_客服、温柔女声_客服、朝气男声_客服、知性女声_新闻播报、嘹亮女声_新闻播报、成熟男声_新闻播报、知性女声_中英混合、中国台湾女声_方言、闽南女声_方言、舒缓女声_文学、武侠女生_文学_16k、悬疑男声_文学_16k、朝气男声_中英混合_16k、柔美女声_中英混合、知性女声_中英混合_16k、沉稳男声_中英混合、温柔女声_中英混合、温柔女声_中英混合_16k、朝气女声_中英混合_16k、成熟女声_中英混合_16k、成熟男声_纯英文、成熟女声_纯英文。

- 语速speed: 取值范围-500~500, 默认值为“0”。当取值为“0”时, 表示一个成年人正常的语速, 约为250字/分钟。设置该值时, 语速和数值没有绝对的映射关系。
- 音高pitch: 取值范围-500~500, 默认值为“0”。精品发音人不支持调节音高。
- 音量volume: 取值范围0~100, 默认值为“50”。

5 修订记录

表 5-1 修订记录

发布日期	修订记录
2024-1-30	第一次正式发布。