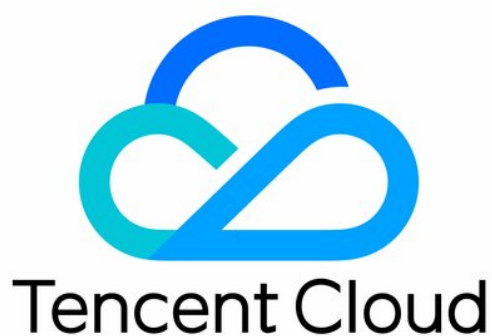


Tencent HealthCare Omics Platform

Getting Started

Product Documentation



Copyright Notice

©2013-2024 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

Trademark Notice



All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

Contents

Getting Started

Run an Analysis Workflow

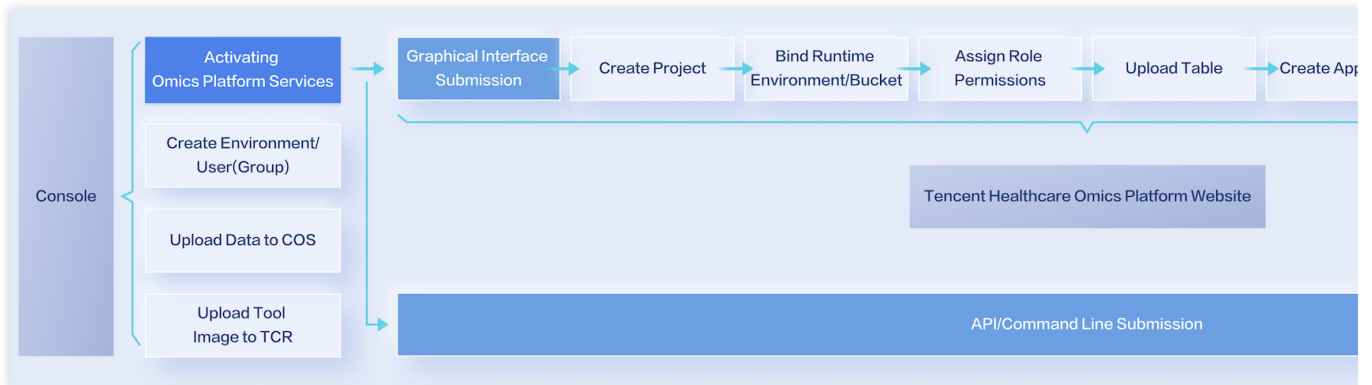
Example: Rapid Application Execution in 5 Minutes

Getting Started

Run an Analysis Workflow

Last updated : 2024-10-22 15:26:45

The common usage process of Tencent Omics Platform is shown in the diagram below:



Step 1: Apply for the Tencent HealthCare Omics Platform Service

Register and apply for a Tencent Cloud account and fill in [Tencent HealthCare Omics Platform activation application](#) to apply for a Tencent HealthCare Omics Platform account.

Step 2: Create Environment/User/User Group in Console

Create an environment in [Tencent Cloud Console - Tencent HealthCare Omics Platform module](#), configure the storage, computing, and other environmental resources of the underlying cluster (such as Tencent Kubernetes Engine TKE Serverless, CFS, and CVM) with one click. After the environment is created, create user/user group accounts as needed to log in to [Tencent HealthCare Omics Platform website](#).

Step 3: Upload Data

Create a Tencent COS bucket. The omics platform supports Beijing, Shanghai, and Guangzhou. It is recommended that a bucket be created in the region closest to the mentioned above. The analysis data is uploaded by the user to the COS bucket he built. For specific operations, see [File Management](#).

Step 4: Upload Tool Image (Optional)

Initialize Tencent Cloud TCR and push the tool image. For details, see [How to Upload Docker Images to Tencent Container Registry \(TCR\)?](#)

Step 5: Create a Project and Bind the Environment/Bucket

After successfully creating the environment and completing the resource configuration in the console, you can log in to [Tencent HealthCare Omics Platform website](#) to officially run an analysis workflow.

First, you can create project based on the business type and bind the environment and the bucket you need to use.

Step 6: Assign Role Permissions

The platform supports project-centered role-based permission management and allows managers to assign project permissions to users in the console. It also supports project owners to assign role permissions in project settings. For specific operations, see [Permission Management](#).

Step 7: Upload Tables

For batch task submission scenes, you need to upload a table. Create a table template and upload the table after its generation. It also supports uploading existing csv format tables. For specific operations, see [Table Management](#).

Step 8: Create a New Application

Create a WDL application. The online editor supports editing, and batch local uploads are also available. You can manage application versions and publish official versions. For detailed operations, see [Application Editing \(Code Editing\)](#).

Step 9: Submit and run the Workflow

After the user completes the application task settings and parameter settings, the running is submitted, and the platform will automatically perform verification and run.

Step 10: View Running

Check the running status and completion progress through the running group list and details page.

Step 11: Obtain Results

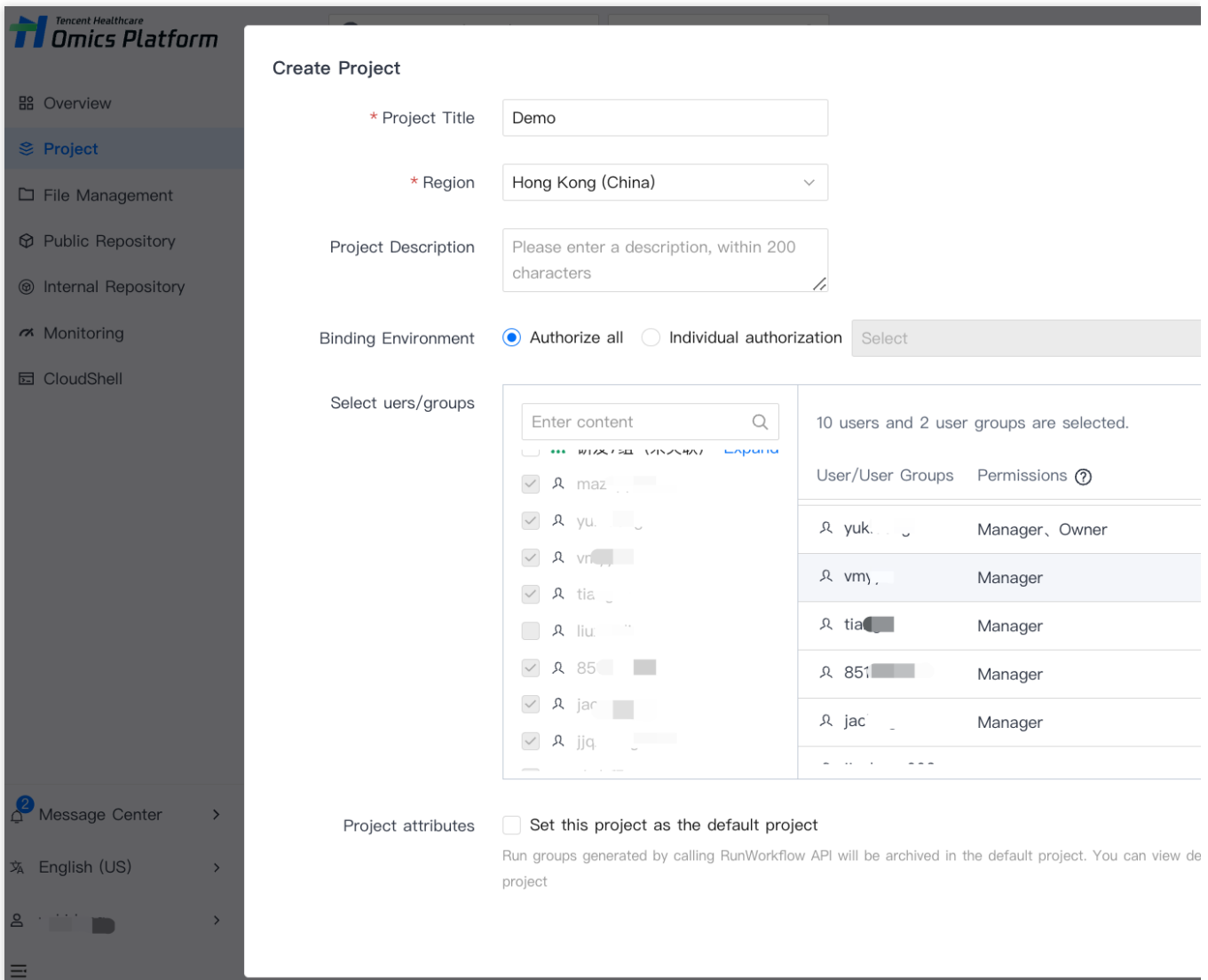
After the running is completed, the analysis results are automatically archived in the specified directory in the COS bucket. Users can directly download and view the analysis results in COS.

Example: Rapid Application Execution in 5 Minutes

Last updated : 2024-10-22 15:26:45

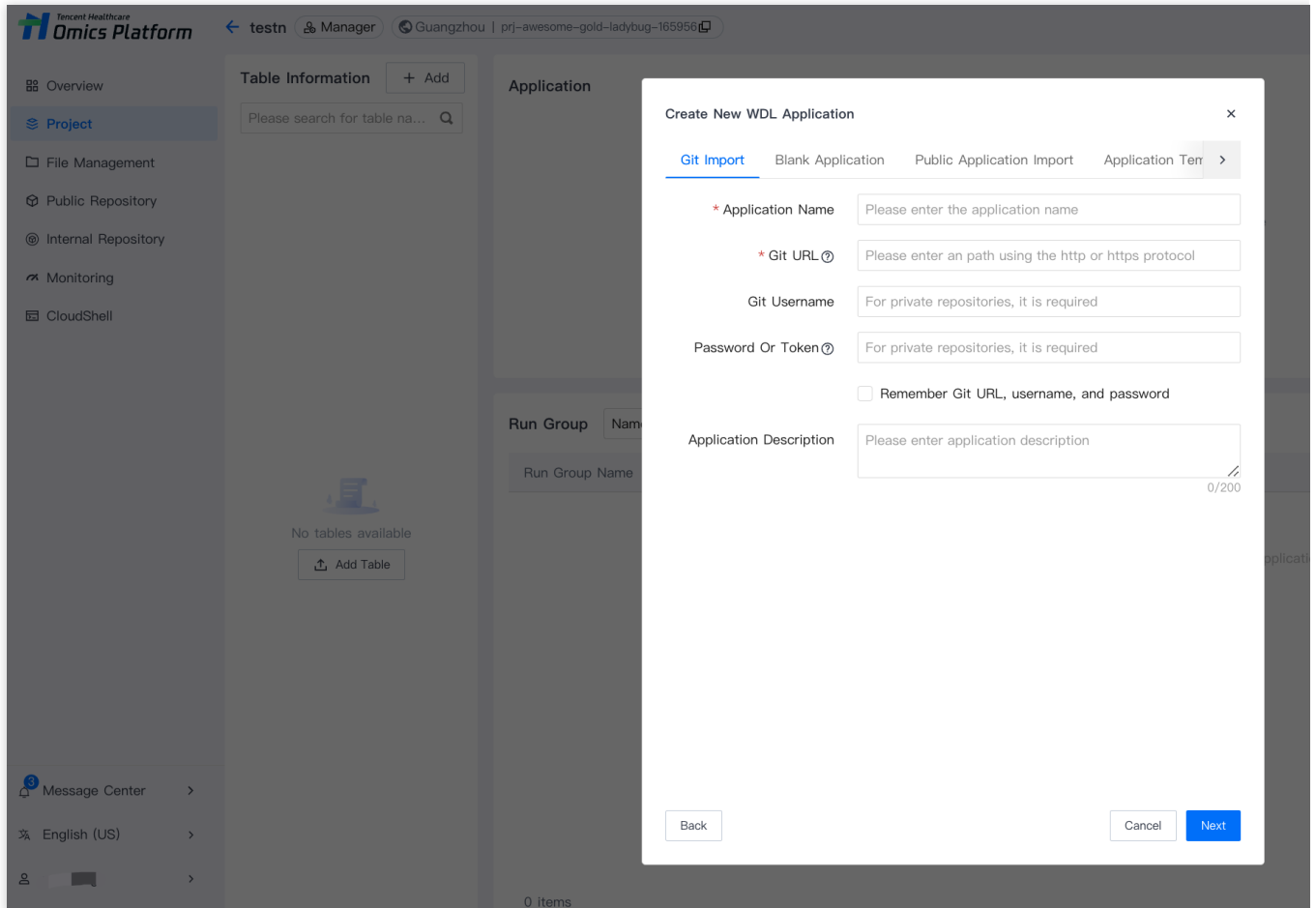
Step 1: Create a project

On the [Project page](#), select **Create**, and enter the necessary information (the headline, region, and associated bucket) to create a new project.



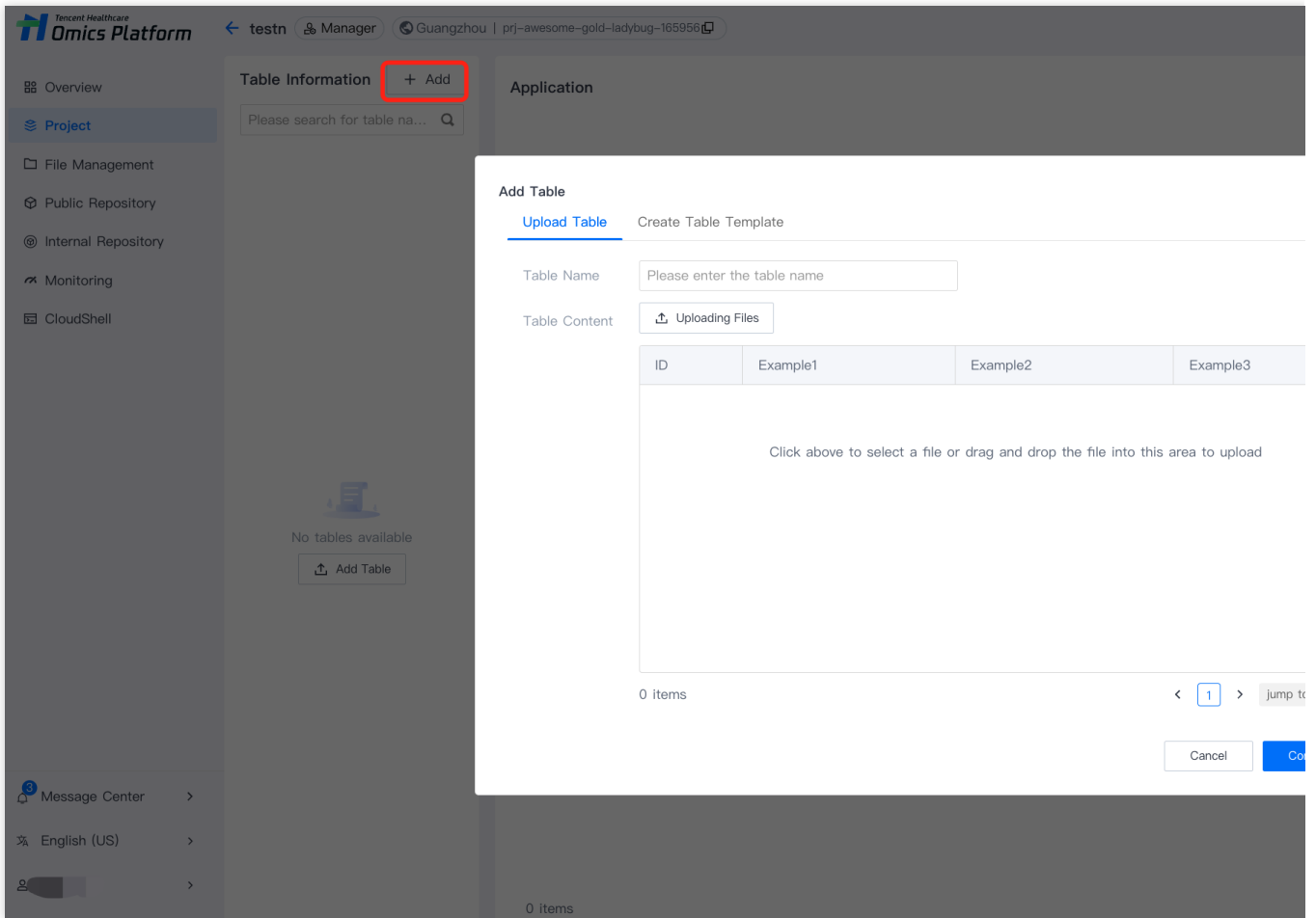
Step 1: Create an application

On the selected project page, select **Add** at the top right corner of the application column, enter the necessary information (application name, type, etc.), and submit to create a new application.



Step 3: Upload table delivery task

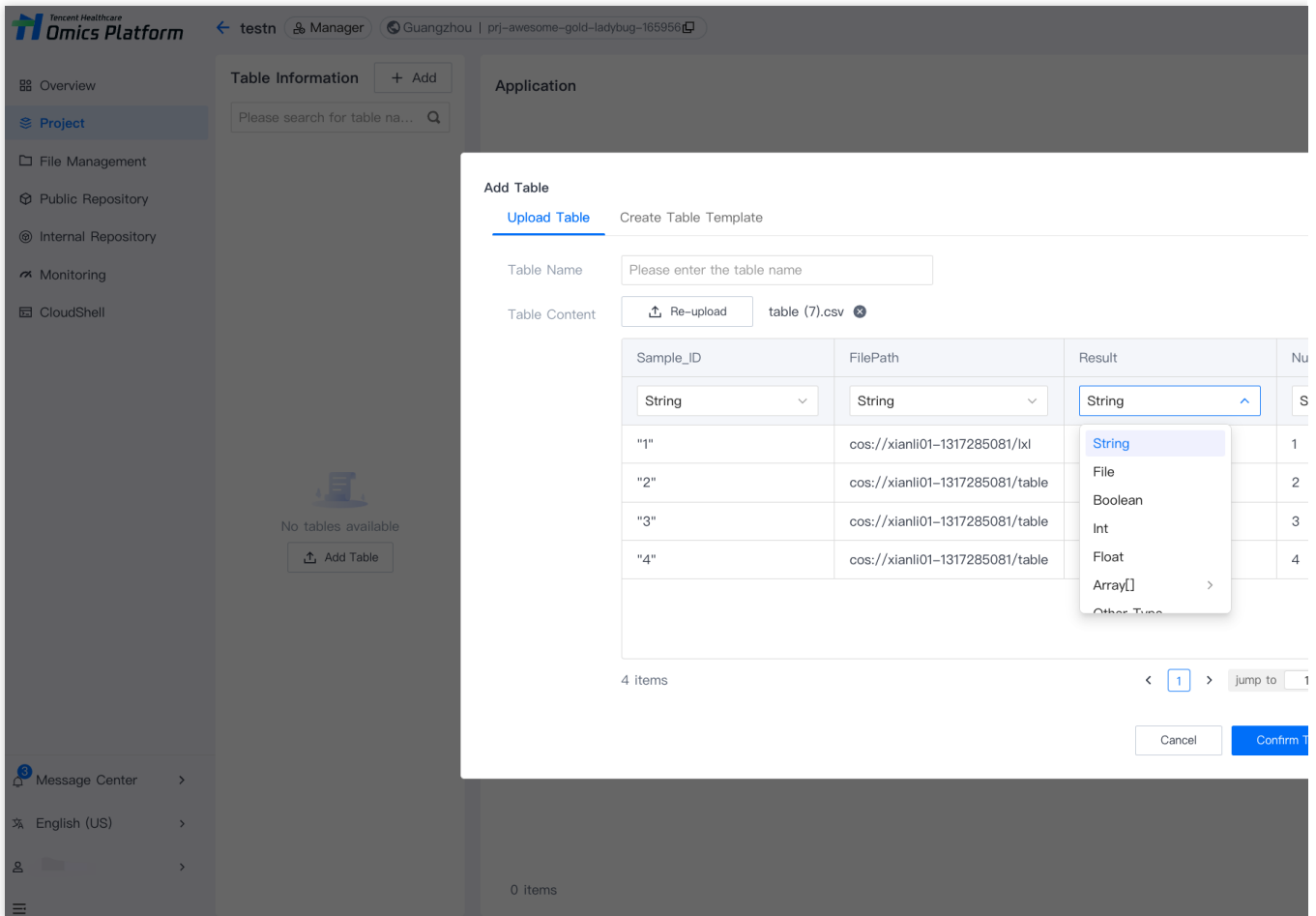
1. Add a new table. For details, see documentation: [Add a New Table](#).



2. Upload a table. For details, see documentation: [Add a New Table](#).

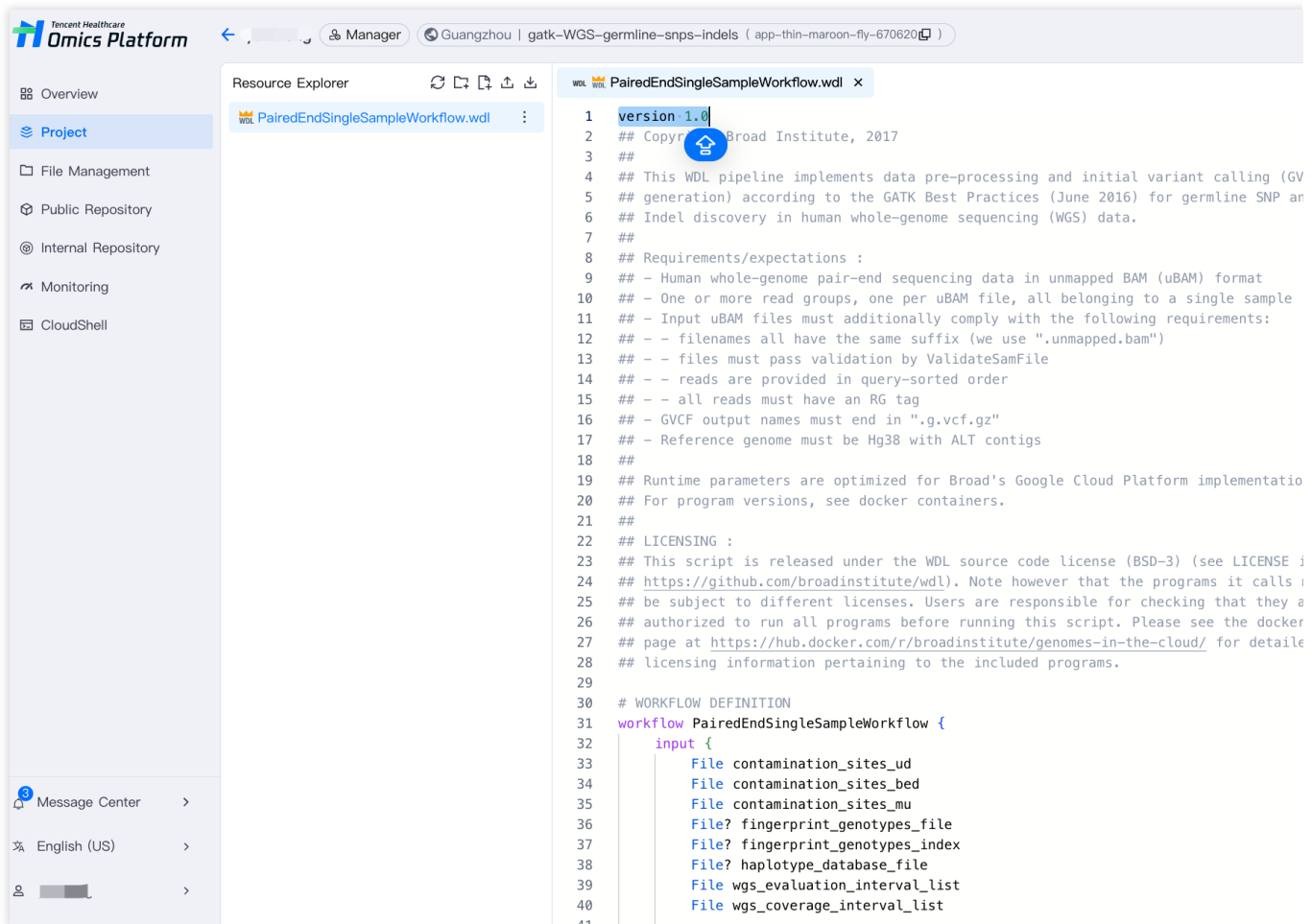
The screenshot displays the Tencent HealthCare Omics Platform interface. The main navigation menu on the left includes: Overview, Project (selected), File Management, Public Repository, Internal Repository, Monitoring, and CloudShell. The 'Table Information' section shows a search bar and a '+ Add' button. The 'Application' section is visible at the top right. A modal dialog titled 'Add Table' is open, featuring a red-bordered 'Upload Table' button. Below this, there is a 'Table Name' input field and a 'Table Content' section with an 'Uploading Files' button. A table with columns 'ID', 'Example1', 'Example2', and 'Example3' is shown, with a message 'Click above to select a file or drag and drop the file into this area to upload'. At the bottom of the dialog, it shows '0 items', a page indicator '< 1 >', and 'jump to' buttons, along with 'Cancel' and 'Con' buttons.

3. Select parameter type. For details, see documentation: [Add a New Table](#).



Step 4: Edit the application using WDL language

Select the parameter type. For details, see documentation: [Edit Application](#).

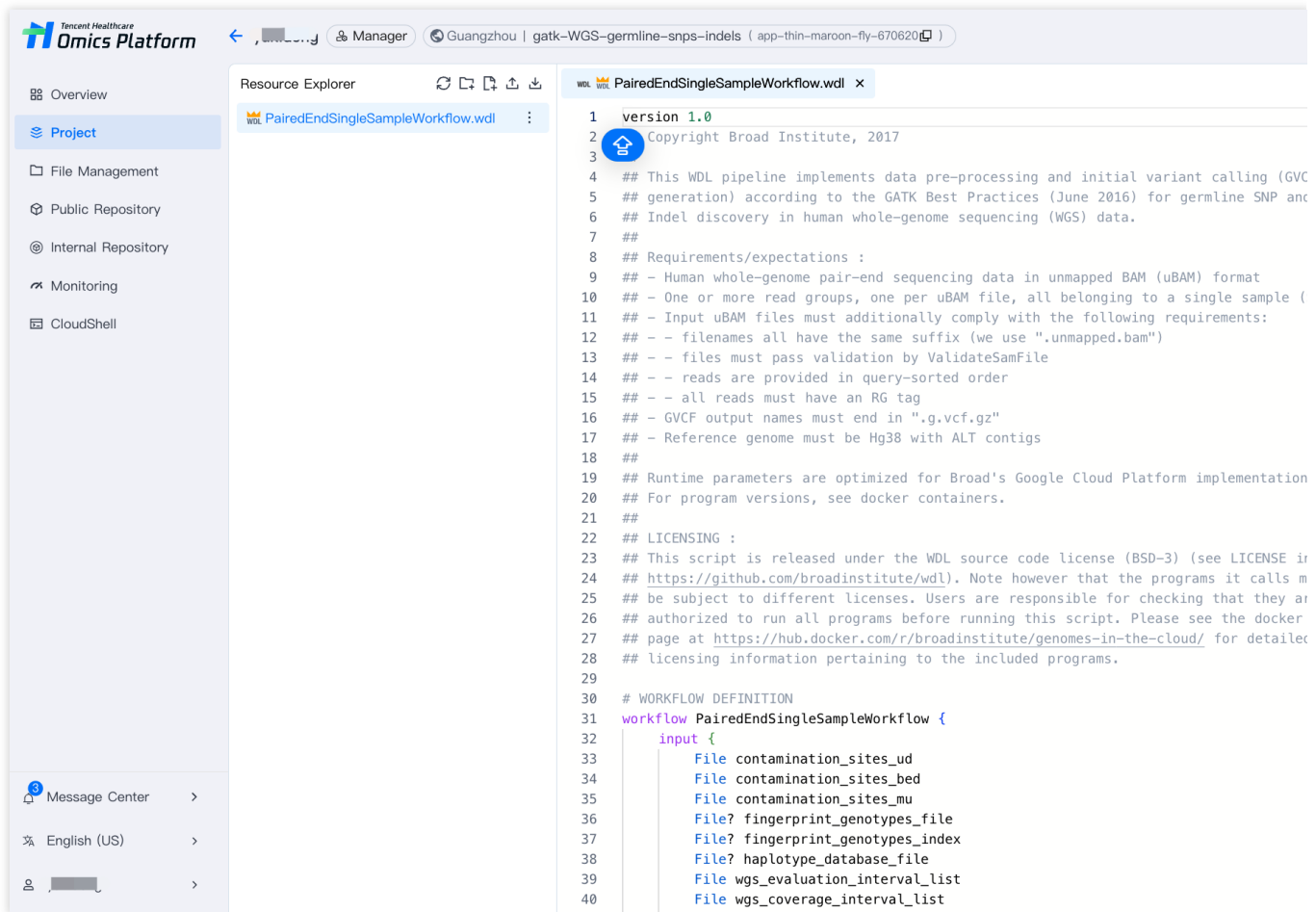


The screenshot displays the Tencent HealthCare Omics Platform interface. The left sidebar contains navigation options: Overview, Project (selected), File Management, Public Repository, Internal Repository, Monitoring, and CloudShell. At the bottom of the sidebar are Message Center, English (US), and a user profile icon. The main area is titled 'Resource Explorer' and shows a file named 'PairedEndSingleSampleWorkflow.wdl'. The right pane displays the WDL code for this workflow, starting with 'version 1.0' and including copyright information for Broad Institute, 2017. The code describes a pipeline for germline SNP and Indel discovery in human whole-genome sequencing (WGS) data. It lists requirements for input uBAM files, including file naming conventions, validation, and sorting. The workflow definition section defines the 'PairedEndSingleSampleWorkflow' with an 'input' block containing several file parameters: 'contamination_sites_ud', 'contamination_sites_bed', 'contamination_sites_mu', 'fingerprint_genotypes_file', 'fingerprint_genotypes_index', 'haplotype_database_file', 'wgs_evaluation_interval_list', and 'wgs_coverage_interval_list'.

```
1 version 1.0
2 ## Copyright 2017 Broad Institute, 2017
3 ##
4 ## This WDL pipeline implements data pre-processing and initial variant calling (GV
5 ## generation) according to the GATK Best Practices (June 2016) for germline SNP an
6 ## Indel discovery in human whole-genome sequencing (WGS) data.
7 ##
8 ## Requirements/expectations :
9 ## - Human whole-genome pair-end sequencing data in unmapped BAM (uBAM) format
10 ## - One or more read groups, one per uBAM file, all belonging to a single sample
11 ## - Input uBAM files must additionally comply with the following requirements:
12 ## -- filenames all have the same suffix (we use ".unmapped.bam")
13 ## -- files must pass validation by ValidateSamFile
14 ## -- reads are provided in query-sorted order
15 ## -- all reads must have an RG tag
16 ## - GVCF output names must end in ".g.vcf.gz"
17 ## - Reference genome must be Hg38 with ALT contigs
18 ##
19 ## Runtime parameters are optimized for Broad's Google Cloud Platform implementatio
20 ## For program versions, see docker containers.
21 ##
22 ## LICENSING :
23 ## This script is released under the WDL source code license (BSD-3) (see LICENSE i
24 ## https://github.com/broadinstitute/wdl). Note however that the programs it calls i
25 ## be subject to different licenses. Users are responsible for checking that they a
26 ## authorized to run all programs before running this script. Please see the docker
27 ## page at https://hub.docker.com/r/broadinstitute/genomes-in-the-cloud/ for detaile
28 ## licensing information pertaining to the included programs.
29 ##
30 # WORKFLOW DEFINITION
31 workflow PairedEndSingleSampleWorkflow {
32   input {
33     File contamination_sites_ud
34     File contamination_sites_bed
35     File contamination_sites_mu
36     File? fingerprint_genotypes_file
37     File? fingerprint_genotypes_index
38     File? haplotype_database_file
39     File wgs_evaluation_interval_list
40     File wgs_coverage_interval_list
41
```

Step 5: Submit and run

Click **Run** application to run.



The screenshot displays the Tencent HealthCare Omics Platform interface. On the left, a navigation sidebar includes 'Overview', 'Project', 'File Management', 'Public Repository', 'Internal Repository', 'Monitoring', and 'CloudShell'. The main area is split into a 'Resource Explorer' on the left and a code editor on the right. The code editor shows a WDL file named 'PairedEndSingleSampleWorkflow.wdl' with the following content:

```
1 version 1.0
2 Copyright Broad Institute, 2017
3
4 ## This WDL pipeline implements data pre-processing and initial variant calling (GVC
5 ## generation) according to the GATK Best Practices (June 2016) for germline SNP and
6 ## Indel discovery in human whole-genome sequencing (WGS) data.
7 ##
8 ## Requirements/expectations :
9 ## - Human whole-genome pair-end sequencing data in unmapped BAM (uBAM) format
10 ## - One or more read groups, one per uBAM file, all belonging to a single sample (
11 ## - Input uBAM files must additionally comply with the following requirements:
12 ## - - filenames all have the same suffix (we use ".unmapped.bam")
13 ## - - files must pass validation by ValidateSamFile
14 ## - - reads are provided in query-sorted order
15 ## - - all reads must have an RG tag
16 ## - GVCF output names must end in ".g.vcf.gz"
17 ## - Reference genome must be Hg38 with ALT contigs
18 ##
19 ## Runtime parameters are optimized for Broad's Google Cloud Platform implementation
20 ## For program versions, see docker containers.
21 ##
22 ## LICENSING :
23 ## This script is released under the WDL source code license (BSD-3) (see LICENSE ir
24 ## https://github.com/broadinstitute/wdl). Note however that the programs it calls m
25 ## be subject to different licenses. Users are responsible for checking that they ar
26 ## authorized to run all programs before running this script. Please see the docker
27 ## page at https://hub.docker.com/r/broadinstitute/genomes-in-the-cloud/ for detailec
28 ## licensing information pertaining to the included programs.
29
30 # WORKFLOW DEFINITION
31 workflow PairedEndSingleSampleWorkflow {
32   input {
33     File contamination_sites_ud
34     File contamination_sites_bed
35     File contamination_sites_mu
36     File? fingerprint_genotypes_file
37     File? fingerprint_genotypes_index
38     File? haplotype_database_file
39     File wgs_evaluation_interval_list
40     File wgs_coverage_interval_list
```

Step 6: Set running and parameters for running

1. Select the parameter type. For details, see documentation: [Set running](#).
2. Complete the running settings and running parameter configuration, then click **start** at the top right corner.

Run Settings

- * Name: gatK-WGS-germline-snps-indels-20240903105322-0
- Description: Enter the run description
- Table: Select sample (Select data)
- Options: Use Call-Caching feature Use relative directory output
- * Failure Mode: NoNewCalls
- * Cache Retention Period: 24 hours 3 days 7 days 30 days Clean up immediately Indefinite
- * Environment: 2024/08/19 (menv-tbd3iagw)
- * Select Volume: 默认缓存卷 (menv-tbd3iagw) (vol-iiry15fl)
- * Output Directory: please enter Add run group name and run id to output p
- * Result Notification: Internal Message Internal Message + Email

Run Parameter Settings

Input Settings

Variable Name	Type
*PairedEndSingleSampleWorkflow.contamination_sites_ud	File
*PairedEndSingleSampleWorkflow.contamination_sites_bed	File
*PairedEndSingleSampleWorkflow.contamination_sites_mu	File
*PairedEndSingleSampleWorkflow.wgs_evaluation_interval_list	File
*PairedEndSingleSampleWorkflow.wgs_coverage_interval_list	File
*PairedEndSingleSampleWorkflow.sample_name	String

Step 7: Obtain Results

After the running of the task, it outputs the results to COS where users can download them directly. Click **Details** of the run group in project list page to obtain run group details.

The screenshot displays the Tencent HealthCare Omics Platform interface. On the left is a navigation sidebar with options: Overview, Project (selected), File Management, Public Repository, Internal Repository, Monitoring, and CloudShell. At the top, the user is logged in as 'yukideng' in the 'Manager' role, located in 'Guangzhou' with project ID 'prj-daft-turquoise-mongrel-219754'. The main content area is divided into three sections:

- Table Information:** A search bar with the text 'Please search for table na...'. Below it, a table lists a single entry: 'test' with user 'yukideng' and timestamp '2024/08/07 21:40:18'.
- Application:** Three application cards are shown:
 - testx:** app-ashamed-tomato-flamingo-296453, Graphical Editing, 2024/08/09 15:32:23.
 - test:** app-mature-violet-ibis-643112, Graphical Editing, 2024/08/09 14:54:58.
 - fastp:** app-lucky-pale, A multi-feature files, featuring C, Coding, 2024/08/09 14:54:58.
- Run Group:** A table with columns: Run Group Name, Application, Status, and a progress indicator. All entries show 'Completed' status with a green checkmark and the number '1'.

Run Group Name	Application	Status	Progress
gatk-WGS-germline-snps-indels-...	gatk-WGS-ger...	Completed	1
fastp-20240809111945-0	fastp (fastp_co...	Completed	1
gatk-WGS-germline-snps-indels-...	gatk-WGS-ger...	Completed	1
gatk-WGS-germline-snps-indels-...	gatk-WGS-ger...	Completed	1
gatk-WGS-germline-snps-indels-...	gatk-WGS-ger...	Completed	1

At the bottom left, there are links for 'Message Center' and 'English (US)'.

In run group details page, Click **Directory** to go to COS path where output files were archived.

Guangzhou | "gatk-WGS-germline-snp-indels-20240809112150-0" Run Details

- Overview
- Project
- File Management
- Public Repository
- Internal Repository
- Monitoring
- CloudShell

Completed

0
Analyzing

0
Initializing

0
Preprocessing

0
Computing

0
Archiving

1
Completed

0
Failed

0
Terminated

Name	gatk-WGS-germline-snp-indels-20240809112150-0	Volume	vol-umuraugi
ID	run-hilarious-mauve-goat-553638	Volume ID	Indefinite
Application type	Private Application	Cache Duration	
AccessMode			
Description			
Environment			
Environment ID	menv-1yur561e		
Options	Use Call-Caching feature		
Failure Mode	NoNewCalls		
Directory	cos://[redacted]-1317285081/test		

List of Jobs of run-hilarious-mauve-goat-553638 | ✔ Completed | 2024/08/09 11:23:49 ~ 2024/08/09 11:26:03

Job name	Status	Start Time	D
PairedEndSingleSampleWorkflow.ApplyBQSR	✔ Completed	2024/08/09 11:24:54	11
PairedEndSingleSampleWorkflow.BaseRecalibrator	✔ Completed	2024/08/09 11:24:34	10
PairedEndSingleSampleWorkflow.CalculateReadGroupChecksum	✔ Completed	2024/08/09 11:25:15	8:
PairedEndSingleSampleWorkflow.CheckContamination	✔ Completed	2024/08/09 11:24:34	10
PairedEndSingleSampleWorkflow.CheckPreValidation	✔ Completed	2024/08/09 11:25:33	11
PairedEndSingleSampleWorkflow.CollectAggregationMetrics	✔ Completed	2024/08/09 11:25:15	17
PairedEndSingleSampleWorkflow.CollectGvcfCallingMetrics	✔ Completed	2024/08/09 11:25:54	8:
PairedEndSingleSampleWorkflow.CollectQualityYieldMetrics	✔ Completed	2024/08/09 11:23:55	7: