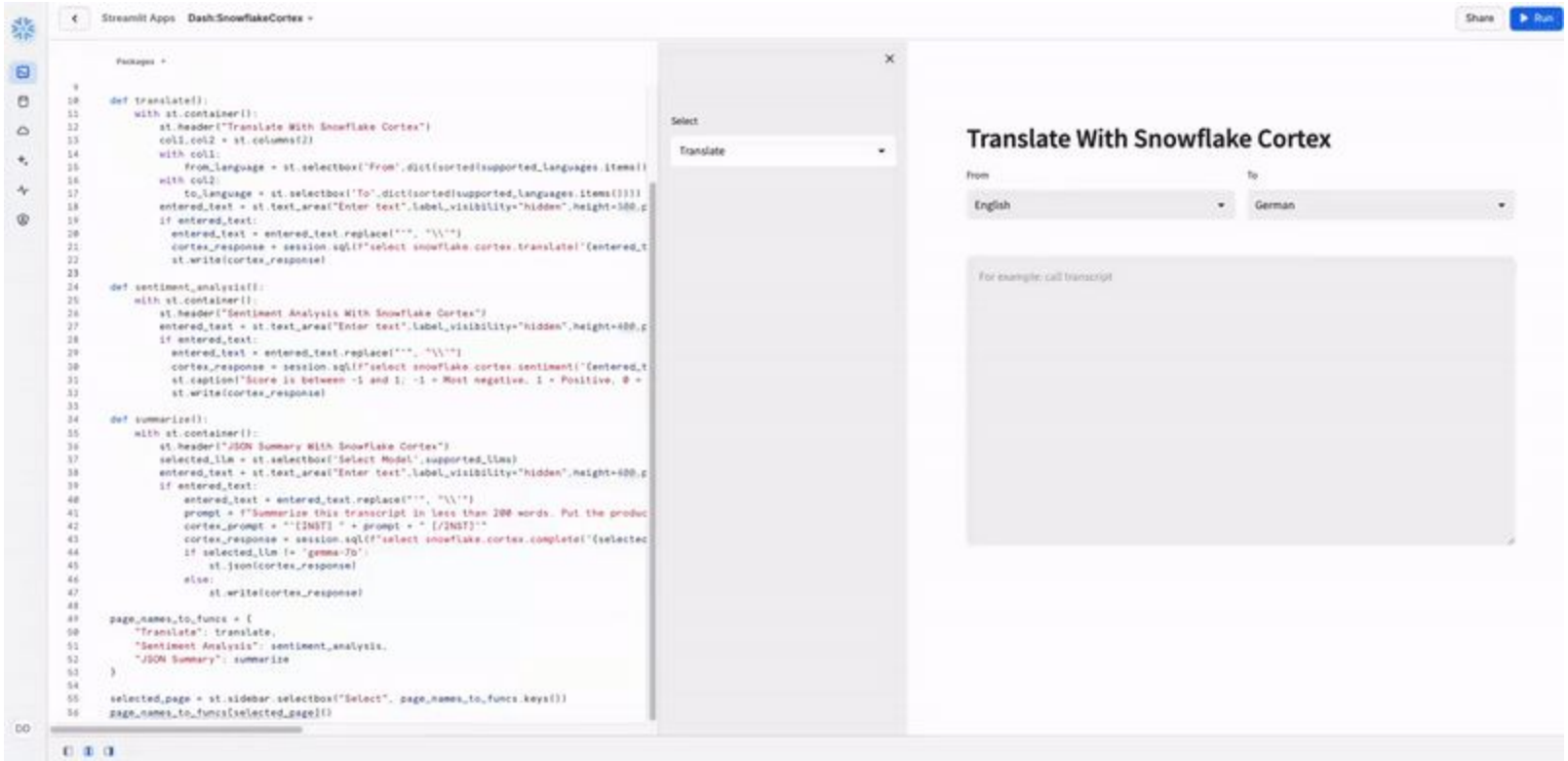
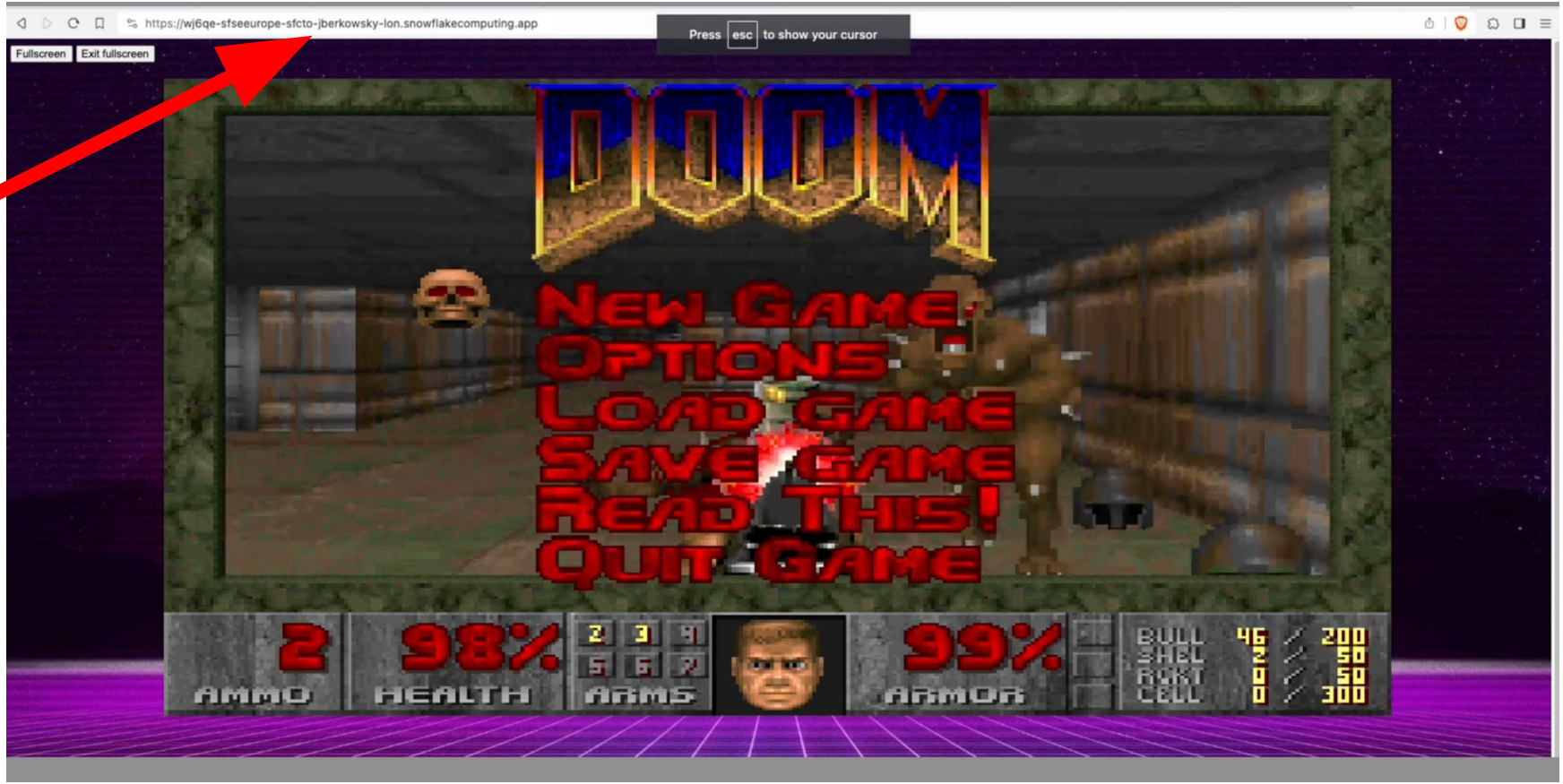
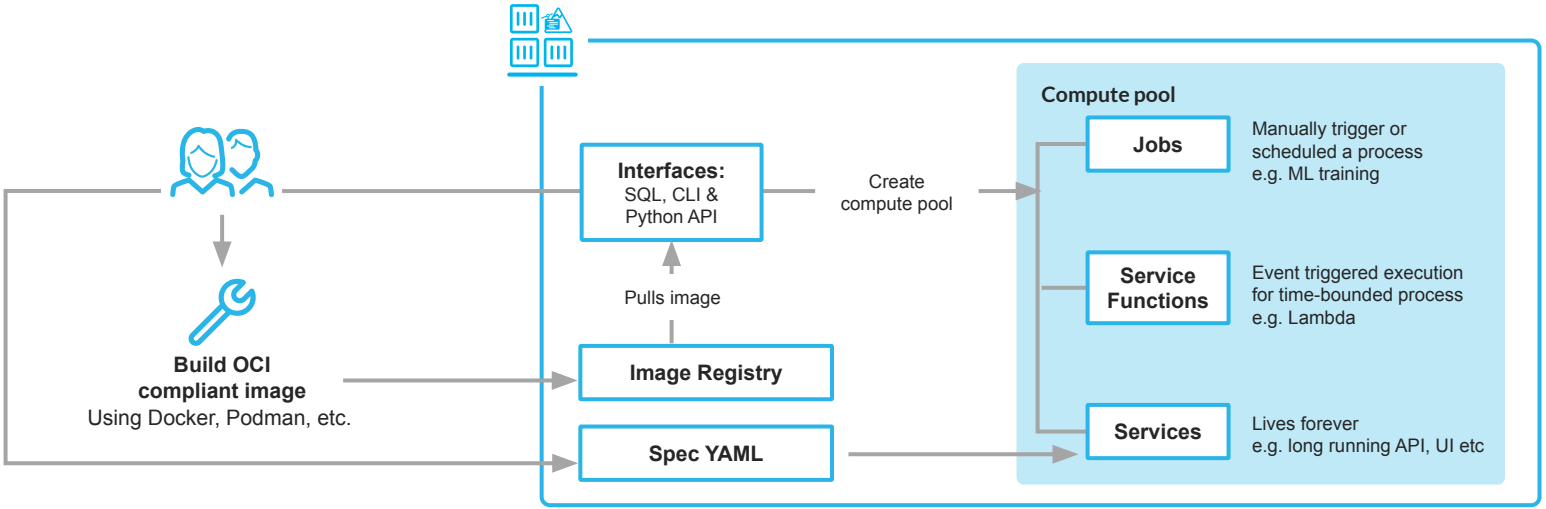




Snowpark Container Services



Snowpark Container Services

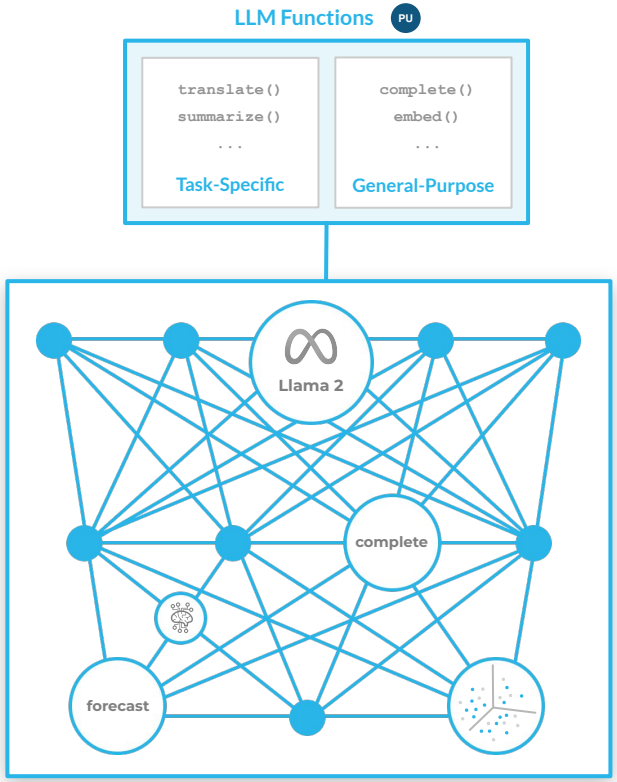
```
CREATE COMPUTE POOL my_cpl
      MIN_NODES = 1
      MAX_NODES = 3
      INSTANCE_FAMILY = GPU

CREATE SERVICE [ IF NOT EXISTS ] inference point
      MIN_INSTANCES = 1
      MAX_INSTANCES = 3
      COMPUTE_POOL = my_cpl
      SPEC = @my_service_stage/path/to/inference_point.yaml
```

Define compute pool size and instance type

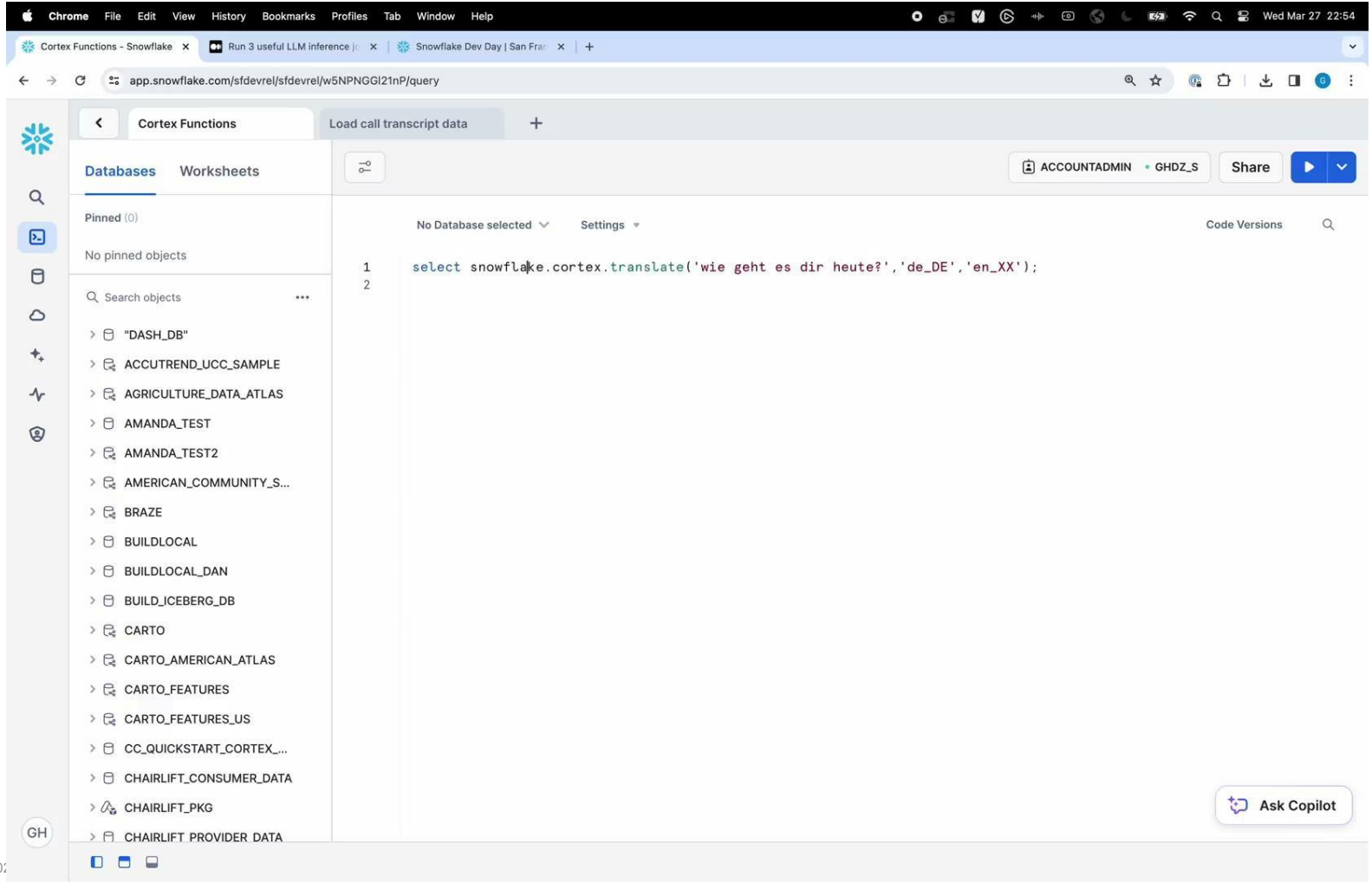
Choose and define execution as service, function or job

Point to spec file with service definition

Snowflake Cortex



PU Public Preview


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Wed Mar 27 22:54
Cortex Functions - Snowflake
Run 3 useful LLM inference
Snowflake Dev Day | San Fran
app.snowflake.com/sfdevrel/sfdevrel/w5NPNGGI21nP/query
Cortex Functions
Load call transcript data
Databases
Worksheets
Pinned (0)
No pinned objects
Search objects
"DASH_DB"
ACCUTREND_UCC_SAMPLE
AGRICULTURE_DATA_ATLAS
AMANDA_TEST
AMANDA_TEST2
AMERICAN_COMMUNITY_S...
BRAZE
BUILDLOCAL
BUILDLOCAL_DAN
BUILD_ICEBERG_DB
CARTO
CARTO_AMERICAN_ATLAS
CARTO_FEATURES
CARTO_FEATURES_US
CC_QUICKSTART_CORTEX_...
CHAIRLIFT_CONSUMER_DATA
CHAIRLIFT_PKG
CHAIRLIFT_PROVIDER_DATA
GH
ACCOUNTADMIN • GHDZ_S
Share
No Database selected
Settings
Code Versions
1 select snowflake.cortex.translate('wie geht es dir heute?','de_DE','en_XX');
2
Ask Copilot

Run 3 useful LLM inference jobs in minutes with Snowflake Cortex

Dash Desai · Follow
Published in Snowflake · 5 min read · Feb 28, 2024

Overview

Getting started with AI on enterprise data can seem overwhelming, between getting familiar with LLMs, how to perform custom prompt engineering, and how to get a wide range of LLMs deployed/integrated to run multiple tests all while keeping that valuable enterprise data secure. Well, a lot of these complexities are being abstracted away for you in Snowflake Cortex — currently in Public Preview.