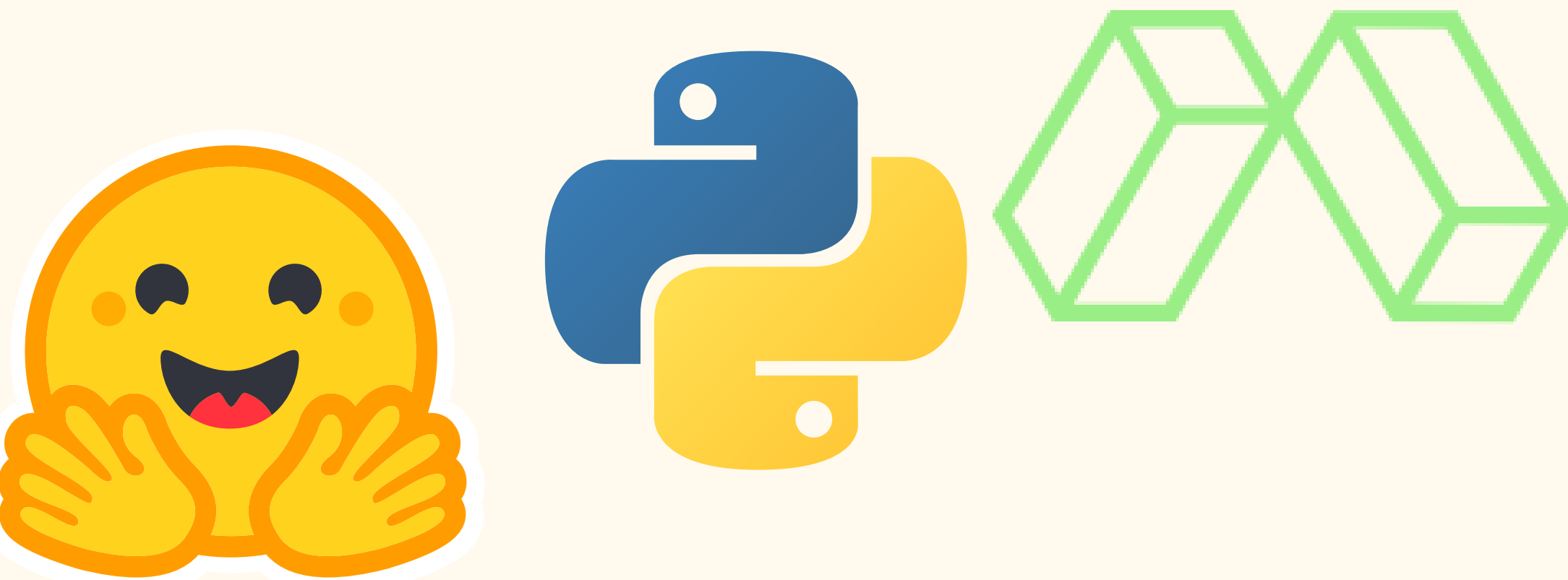


STREAMLINING ML DEPLOYMENT IN CLOUD WITH MODAL



**SPEED THROUGH POC
TO DEPLOYING YOUR
MODEL WITH MODAL**



CHALLENGE SOLVED : DEPLOYING AI IDEAS TO CLOUD

- **THESE DAYS AI IDEAS INCLUDES NOT JUST CODE. IT INCLUDES LARGE MODELS, AND THE HW CONFIGURATION REQUIRED RUN THESE MODELS**
- **CREATING CLOUD INSTANCES CAN BE SIMPLE FOR POC. WHEN IT COMES TO DEPLOYMENT THE WORKFLOW IS LABORIOUS**
- **UNDERSTANDING EACH CLOUD SERVICE PROVIDER WILL REQUIRE ADDITIONAL TIME, AND DEVELOPMENT COST.**

THIS IS WHERE MODAL COMES IN...

- **WE HAVE MANY OPENSOURCE MODELS THAT CAN HELP IN AUTOMATING MANY TASKS RANGING FROM CLASSIFICATION TO TEXT GENERATION. WITH MODAL WE CAN DEPLOY THEM AND USE IT**
- **ALL OF IT IS DONE USING PYTHON CODE !!! YEP ITS COMPLETELY ABSTRACTED**

MODAL IN 4 STEPS

1. IT IS AN END-TO-END CLOUD COMPUTE SERVICE

2. CAN DO : MODEL INFERENCE, BATCH JOBS, TASK QUEUES, WEB APPS AND MORE. ALL WITHOUT YOUR OWN INFRASTRUCTURE.

3. LIBRARY : PIP INSTALL MODAL-CLIENT

4. API : MODAL TOKEN NEW

MODAL CODE

```
IMPORT MODAL

STUB = MODAL.STUB("EXAMPLE-GET-STARTED")

@STUB.FUNCTION()

DEF SQUARE(X):
    PRINT("THIS CODE IS RUNNING ON A REMOTE WORKER!")

    RETURN X**2

@STUB.LOCAL_ENTRYPOINT()

DEF MAIN():
    PRINT("THE SQUARE IS", SQUARE.CALL(42))
```

RAW PYTHON CODE

```
DEF SQUARE(X):
    PRINT("THIS CODE IS RUNNING ON LOCAL MACHINE!")
    RETURN X**2

DEF MAIN():
    PRINT("THE SQUARE IS", SQUARE(42))

IF __NAME__ == __MAIN__:
    MAIN()
```

MODALS & HF MODELS

- WILL LOOK AT SIMPLE PYTHON PROGRAM EXECUTED INSIDE MODAL
- IMPLEMENT THE TEXT CLASSIFICATION TASK USING OPEN SOURCE MODEL FROM HUGGING FACE
- IMPLEMENT TEXT GENERATION TASK USING THE GPT2 MODEL FROM HUGGING FACE
- DISCUSS ON USING THE BIGGER MODELS WITH GPU OPTIONS INSIDE MODAL

insightbuilder/
python_de_learners_data



Repo contains the code, data and supporting documents including presentations, playbooks and additional documents to support learning

2 Contributors 0 Issues 57 Stars 29 Forks



python_de_learners_data/gpt2_generate.py at main · insightbuilder/python_de_learners_data

Repo contains the code, data and supporting documents including presentations, playbooks and additional documents to support learning - python_de_learners_data/gpt2_generate.py at main · insightbui...

GitHub

insightbuilder/
python_de_learners_data



Repo contains the code, data and supporting documents including presentations, playbooks and additional documents to support learning

2 Contributors 0 Issues 57 Stars 29 Forks

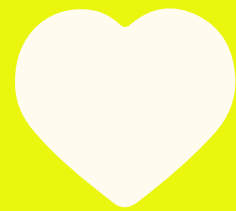


python_de_learners_data/get_started.py at main · insightbuilder/python_de_learners_data

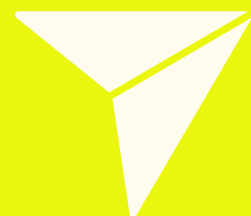
Repo contains the code, data and supporting documents including presentations, playbooks and additional documents to support learning - python_de_learners_data/get_started.py at main · insightbuild...

GitHub

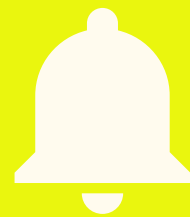
THANKS FOR WATCHING



LIKE



SHARE



SUBSCRIBE