

In [2]: `!pip install sagemaker`

Requirement already satisfied: sagemaker in /opt/conda/lib/python3.12/site-packages (2.242.0)

Requirement already satisfied: attrs<24,>=23.1.0 in /opt/conda/lib/python3.12/site-packages (from sagemaker) (23.2.0)

Requirement already satisfied: boto3<2.0,>=1.35.75 in /opt/conda/lib/python3.12/site-packages (from sagemaker) (1.37.1)

Requirement already satisfied: cloudpickle>=2.2.1 in /opt/conda/lib/python3.12/site-packages (from sagemaker) (3.1.1)

Requirement already satisfied: docker in /opt/conda/lib/python3.12/site-packages (from sagemaker) (7.1.0)

Requirement already satisfied: fastapi in /opt/conda/lib/python3.12/site-packages (from sagemaker) (0.115.12)

Requirement already satisfied: google-pasta in /opt/conda/lib/python3.12/site-packages (from sagemaker) (0.2.0)

Requirement already satisfied: importlib-metadata<7.0,>=1.4.0 in /opt/conda/lib/python3.12/site-packages (from sagemaker) (6.10.0)

Requirement already satisfied: jsonschema in /opt/conda/lib/python3.12/site-packages (from sagemaker) (4.23.0)

Requirement already satisfied: numpy<2.0,>=1.9.0 in /opt/conda/lib/python3.12/site-packages (from sagemaker) (1.26.4)

Requirement already satisfied: omegaconf<=2.3,>=2.2 in /opt/conda/lib/python3.12/site-packages (from sagemaker) (2.3.0)

Requirement already satisfied: packaging>=20.0 in /opt/conda/lib/python3.12/site-packages (from sagemaker) (24.2)

Requirement already satisfied: pandas in /opt/conda/lib/python3.12/site-packages (from sagemaker) (2.2.3)

Requirement already satisfied: pathos in /opt/conda/lib/python3.12/site-packages (from sagemaker) (0.3.4)

Requirement already satisfied: platformdirs in /opt/conda/lib/python3.12/site-packages (from sagemaker) (4.3.7)

Requirement already satisfied: protobuf<6.0,>=3.12 in /opt/conda/lib/python3.12/site-packages (from sagemaker) (5.28.3)

Requirement already satisfied: psutil in /opt/conda/lib/python3.12/site-packages (from sagemaker) (5.9.8)

Requirement already satisfied: pyyaml~=6.0 in /opt/conda/lib/python3.12/site-packages (from sagemaker) (6.0.2)

Requirement already satisfied: requests in /opt/conda/lib/python3.12/site-packages (from sagemaker) (2.32.3)

Requirement already satisfied: sagemaker-core<2.0.0,>=1.0.17 in /opt/conda/lib/python3.12/site-packages (from sagemaker) (1.0.31)

Requirement already satisfied: schema in /opt/conda/lib/python3.12/site-packages (from sagemaker) (0.7.7)

Requirement already satisfied: smdebug-rulesconfig==1.0.1 in /opt/conda/lib/python3.12/site-packages (from sagemaker) (1.0.1)

Requirement already satisfied: tblib<4,>=1.7.0 in /opt/conda/lib/python3.12/site-packages (from sagemaker) (3.1.0)

Requirement already satisfied: tqdm in /opt/conda/lib/python3.12/site-packages (from sagemaker) (4.67.1)

Requirement already satisfied: urllib3<3.0.0,>=1.26.8 in /opt/conda/lib/python3.12/site-packages (from sagemaker) (2.4.0)

Requirement already satisfied: uvicorn in /opt/conda/lib/python3.12/site-packages (from sagemaker) (0.34.2)

Requirement already satisfied: botocore<1.38.0,>=1.37.1 in /opt/conda/lib/python3.12/site-packages (from boto3<2.0,>=1.35.75->sagemaker) (1.37.1)

Requirement already satisfied: jmespath<2.0.0,>=0.7.1 in /opt/conda/lib/python3.12/site-packages (from boto3<2.0,>=1.35.75->sagemaker) (1.0.1)

Requirement already satisfied: s3transfer<0.12.0,>=0.11.0 in /opt/conda/lib/python3.12/site-packages (from boto3<2.0,>=1.35.75->sagemaker) (0.11.3)

Requirement already satisfied: zipp>=0.5 in /opt/conda/lib/python3.12/site-packages (from importlib-metadata<7.0,>=1.4.0->sagemaker) (3.21.0)

Requirement already satisfied: antlr4-python3-runtime==4.9.\* in /opt/conda/lib/python3.12/site-packages (from omegaconf<=2.3,>=2.2->sagemaker) (4.9.3)

Requirement already satisfied: pydantic<3.0.0,>=2.0.0 in /opt/conda/lib/python3.12/site-packages (from sagemaker-core<2.0.0,>=1.0.17->sagemaker) (2.11.3)

Requirement already satisfied: rich<14.0.0,>=13.0.0 in /opt/conda/lib/python3.12/site-packages (from sagemaker-core<2.0.0,>=1.0.17->sagemaker) (13.9.4)

Requirement already satisfied: mock<5.0,>4.0 in /opt/conda/lib/python3.12/site-packages (from sagemaker-core<2.0.0,>=1.0.17->sagemaker) (4.0.3)

Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /opt/conda/lib/python3.12/site-packages (from jsonschema->sagemaker) (2025.4.1)

Requirement already satisfied: referencing>=0.28.4 in /opt/conda/lib/python3.12/site-packages (from jsonschema->sagemaker) (0.36.2)

Requirement already satisfied: rpds-py>=0.7.1 in /opt/conda/lib/python3.12/site-packages (from jsonschema->sagemaker) (0.24.0)

Requirement already satisfied: charset-normalizer<4,>=2 in /opt/conda/lib/python3.12/site-packages (from requests->sagemaker) (3.4.2)

Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.12/site-packages (from requests->sagemaker) (3.10)

Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/lib/python3.12/site-packages (from requests->sagemaker) (2025.1.31)

Requirement already satisfied: starlette<0.47.0,>=0.40.0 in /opt/conda/lib/python3.12/site-packages (from fastapi->sagemaker) (0.46.2)

Requirement already satisfied: typing-extensions>=4.8.0 in /opt/conda/lib/python3.12/site-packages (from fastapi->sagemaker) (4.13.2)

Requirement already satisfied: six in /opt/conda/lib/python3.12/site-packages (from google-pasta->sagemaker) (1.17.0)

Requirement already satisfied: python-dateutil>=2.8.2 in /opt/conda/lib/python3.12/site-packages (from pandas->sagemaker) (2.9.0.post0)

Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.12/site-packages (from pandas->sagemaker) (2024.2)

Requirement already satisfied: tzdata>=2022.7 in /opt/conda/lib/python3.12/site-packages (from pandas->sagemaker) (2025.2)

Requirement already satisfied: ppft>=1.7.7 in /opt/conda/lib/python3.12/site-packages (from pathos->sagemaker) (1.7.7)

Requirement already satisfied: dill>=0.4.0 in /opt/conda/lib/python3.12/site-packages (from pathos->sagemaker) (0.4.0)

Requirement already satisfied: pox>=0.3.6 in /opt/conda/lib/python3.12/site-packages (from pathos->sagemaker) (0.3.6)

Requirement already satisfied: multiprocessing>=0.70.18 in /opt/conda/lib/python3.12/site-packages (from pathos->sagemaker) (0.70.18)

Requirement already satisfied: click>=7.0 in /opt/conda/lib/python3.12/site-packages (from uvicorn->sagemaker) (8.1.8)

Requirement already satisfied: h11>=0.8 in /opt/conda/lib/python3.12/site-packages (from uvicorn->sagemaker) (0.16.0)

Requirement already satisfied: annotated-types>=0.6.0 in /opt/conda/lib/python3.12/site-packages (from pydantic<3.0.0,>=2.0.0->sagemaker-core<2.0.0,>=1.0.17->sagemaker) (0.7.0)

Requirement already satisfied: pydantic-core==2.33.1 in /opt/conda/lib/python3.12/site-packages (from pydantic<3.0.0,>=2.0.0->sagemaker-core<2.0.0,>=1.0.17->sagemaker) (2.33.1)

Requirement already satisfied: typing-inspection>=0.4.0 in /opt/conda/lib/python3.12/site-packages (from pydantic<3.0.0,>=2.0.0->sagemaker-core<2.0.0,>=1.0.17->sagemaker) (0.4.0)

Requirement already satisfied: markdown-it-py>=2.2.0 in /opt/conda/lib/python3.12/site-packages (from rich<14.0.0,>=13.0.0->sagemaker-core<2.0.0,>=1.0.17->sagemaker) (3.0.0)

Requirement already satisfied: pygments<3.0.0,>=2.13.0 in /opt/conda/lib/python3.12/site-packages (from rich<14.0.0,>=13.0.0->sagemaker-core<2.0.0,>=1.0.17->sagemaker) (2.19.1)

Requirement already satisfied: anyio<5,>=3.6.2 in /opt/conda/lib/python3.12/site-

packages (from starlette<0.47.0,>=0.40.0->fastapi->sagemaker) (4.9.0)  
 Requirement already satisfied: sniffio>=1.1 in /opt/conda/lib/python3.12/site-packages (from anyio<5,>=3.6.2->starlette<0.47.0,>=0.40.0->fastapi->sagemaker) (1.3.1)  
 Requirement already satisfied: mdurl~=0.1 in /opt/conda/lib/python3.12/site-packages (from markdown-it-py>=2.2.0->rich<14.0.0,>=13.0.0->sagemaker-core<2.0.0,>=1.0.17->sagemaker) (0.1.2)

```
In [5]: import sagemaker
import boto3
```

- sagemaker session bucket -> used for uploading data, models and logs
- sagemaker will automatically will create this bucket if not exists

```
In [7]: session = sagemaker.Session()
```

```
In [8]: sagemaker_session_bucket = None
```

```
In [9]: if sagemaker_session_bucket is None and session is not None:
sagemaker_session_bucket = session.default_bucket()
```

## Role management

```
In [10]: try:
role = sagemaker.get_execution_role()
except ValueError as e:
iam = boto3.client("iam")
role = iam.get_role(RoleName = 'sagemaker_execution_role')['Role']['Arn']
print()
```

```
In [12]: session = sagemaker.Session(default_bucket = sagemaker_session_bucket)
```

```
In [13]: print(f'sagemaker role ARN: ', role)
print("sagemaker serrion region", session.boto_region_name)
```

sagemaker role ARN: arn:aws:iam::368603112002:role/service-role/AmazonSageMaker-ExecutionRole-20250527T121519  
 sagemaker serrion region us-east-1

```
In [14]: from sagemaker.huggingface.model import HuggingFaceModel

hub = {
    'HF_MODEL_ID': 'distilbert-base-uncased-distilled-squad',
    'HF_TASK': 'question-answering'
}
```

```
In [15]: hugging_face_model = HuggingFaceModel(
env=hub,
role=role,
transformers_version="4.26",
pytorch_version="1.13",
py_version="py39"
)
```

```
In [16]: predictor = hugging_face_model.deploy(
initial_instance_count=1,
```

```
instance_type="ml.m5.xlarge"  
)
```

-----!

```
In [18]: data = {  
    "inputs": {  
        "question": "What is used for inference? ",  
        "context" : "My name is shravan k subrahmanya and i live in thirthahalli"  
    }  
}
```

```
In [19]: predictor.predict(data)
```

```
Out[19]: {'score': 0.9980034232139587, 'start': 85, 'end': 94, 'answer': 'sagemaker'}
```

```
In [ ]: predictor.predict()
```