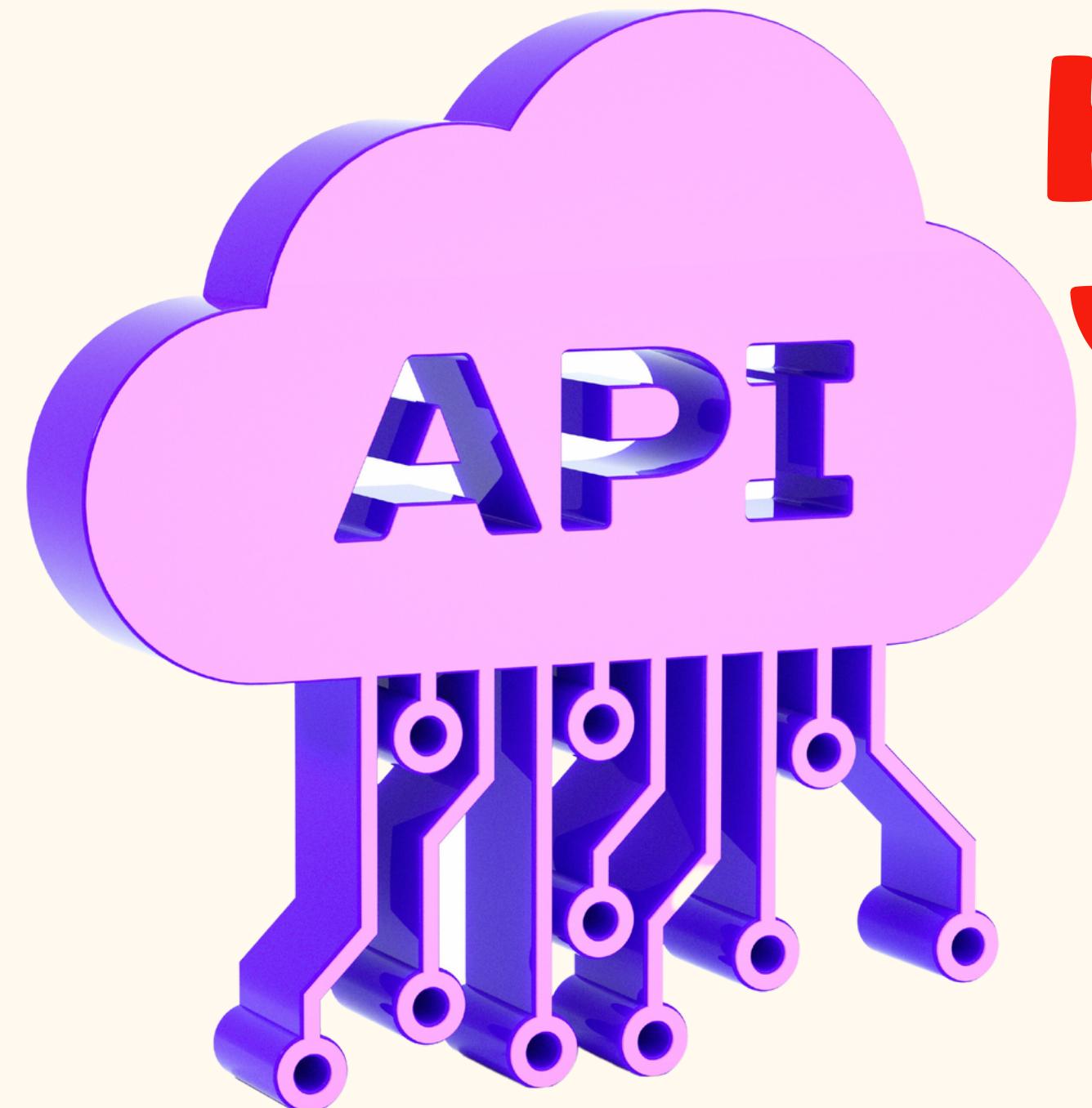


HOSTING LLMS ON YOUR API SERVER



A STEP-BY-STEP
GUIDE FOR COLAB
NOTEBOOK

[HTTPS://GITHUB.COM/INSIGHTBUILDER](https://github.com/insightbuilder)



INSIGHT BUILDER



Kamalraj M M

@insightbuilder 59 subscribers 88 videos

[More about this channel >](#)

[Customize channel](#)

[Manage videos](#)

HOME

VIDEOS

PLAYLISTS

COMMUNITY

CHANNELS

ABOUT



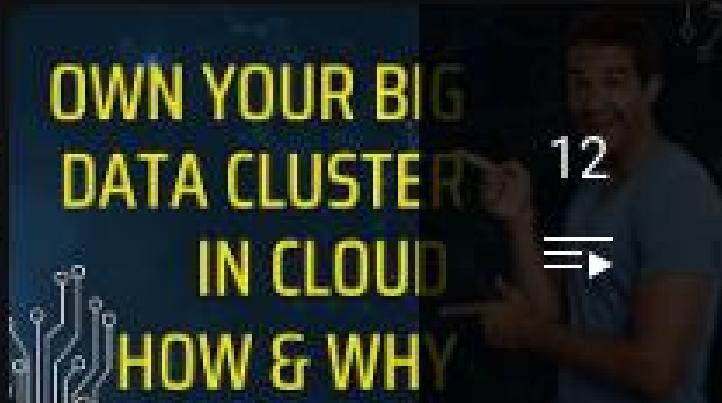
>

All playlists ~

Created playlists



Liked videos



Build and Own Your Big Data
ETL, ELT, ...
[View details](#)



Exploring Top Challenges
That Linux Solves: A...
[View details](#)



SQL Mastery : Absolute
Basics to Complex Data...
[View details](#)



Learn about AI Language
Models and Reinforcement...
[View details](#)

HOW TO GET YOUR OWN LLMS LIKE TEXT-DAVINCI



ALPACA-7B

ON COLLAB

PRACTICAL ON COLLAB

STEPS TO HOST LLM

1) API SERVER MUST HAVE

NVIDIA P100 SERIES GPU WITH MINIMUM 8GB RAM

2) PYTHON WITH PYNGROK, FASTAPI LIBRARY INSTALLED

**3) HUGGINGFACE LIBRARY WITH LORALIB, BITSANDBYTES &
PEFT MODULES**

**4) HUGGING FACE DECAPODA-RESEARCH/LLAMA-7B-HF &
TLOEN/ALPACA-LORA-7B MODEL PRESENT LOCALLY**

5) LANGCHAIN LIBRARY TO RUN INFERENCE

GOOGLE COLLAB

EASIEST AND MOST

REPEATABLE WAY OF GETTING

THE LLM TO PLAY WITH IS

GOOGLE COLLAB

THANKS FOR WATCHING

 **LIKE**

 **SHARE**

 **SUBSCRIBE**