Promotori di PyData Venice #17

Promotori di PyData Venice #17

Proposte

Agenda

Speech
Prossimi incontri
Networking

Agenda

Speech
Prossimi incontri
Networking

Hugging Face Overview

Alessandra Bilardi - Data & Automation Specialist @ Corley Cloud

@PyDataVenice #17 #Meetup #PyData

Agenda

Hugging Face in numbers
Hugging Face packages and third party tools
The platform

Hugging Face in numbers

🤗 Hugging Face history

Event	When	Notes
Foundation	2016	Clément Delangue (CEO), Julien Chaumond (CTO), Thomas Wolf (CSO)
BLOOM (**B**ig-science **L**arge **O**pen-science **O**pen-access **M**ultilingual language model)	2021 - 2022	Open LLM with research groups 176 billion parameter
Private Hub	August 2022	Support SaaS or on-prem deployment
Gradio	December 2022	The fastest way to demo your ML model with a friendly web interface
Partnerships	2023 - 2024	AWS, Google, Nvidia, AMD, Intel, IBM, ..

https://en.wikipedia.org/wiki/Hugging_Face

may

open source

with you

🤗 Hugging Face sections

Name section	Hugging Face	Kaggle	Name section
Datasets	230K	389K (298K)	Datasets
Models	1M	9200 (2300)	Models
Spaces (Gradio & Streamlit)	over 500K	1.2M (1M)	Code
Posts, Blog, Articles, Discussions and Learn		Discussions and Learn	
Hardware for spaces		Hardware for notebooks	
Enterprise Hub / Competitions		Competitions	

Hugging Face Hub

🤗 Hugging Face resources

Resource	Quantity
huggingface_hub[cli] (manager of spaces)	datasets and files of spaces
huggingface_hub python library	all in Hugging Face Hub
Spaces	infinity
Web interfaces for spaces (docker, gradio or streamlit)	free
Hardware for spaces (with 2 vCPU 16 GB or ZeroGPU)	free
Inference API (serverless)	1000 requests / day
Organization	free

https://huggingface.co/docs

Hugging Face packages and third party tools

🤗 Hugging Face packages on pypi.org
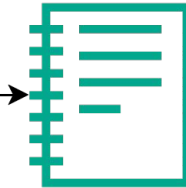
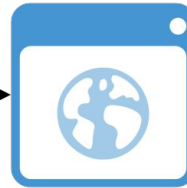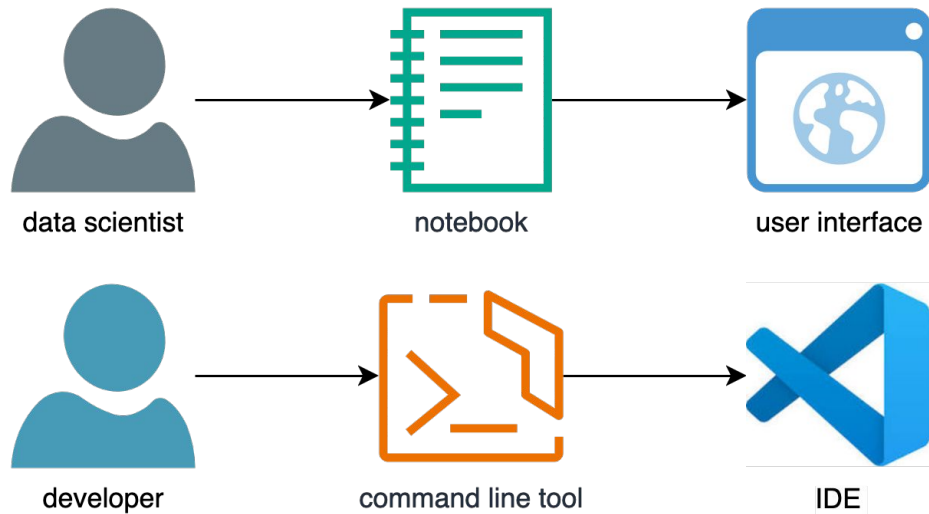| Resource	Notes
accelerate, optimum, peft	Hardware / software optimization
datasets, tokenizers, evaluate	Basic packages
diffusers, transformers	https://github.com/huggingface/transformers.js
gradio	The fastest way to demo your ML model with WI
huggingface_hub	https://github.com/huggingface/huggingface.js

🤗 Client use cases
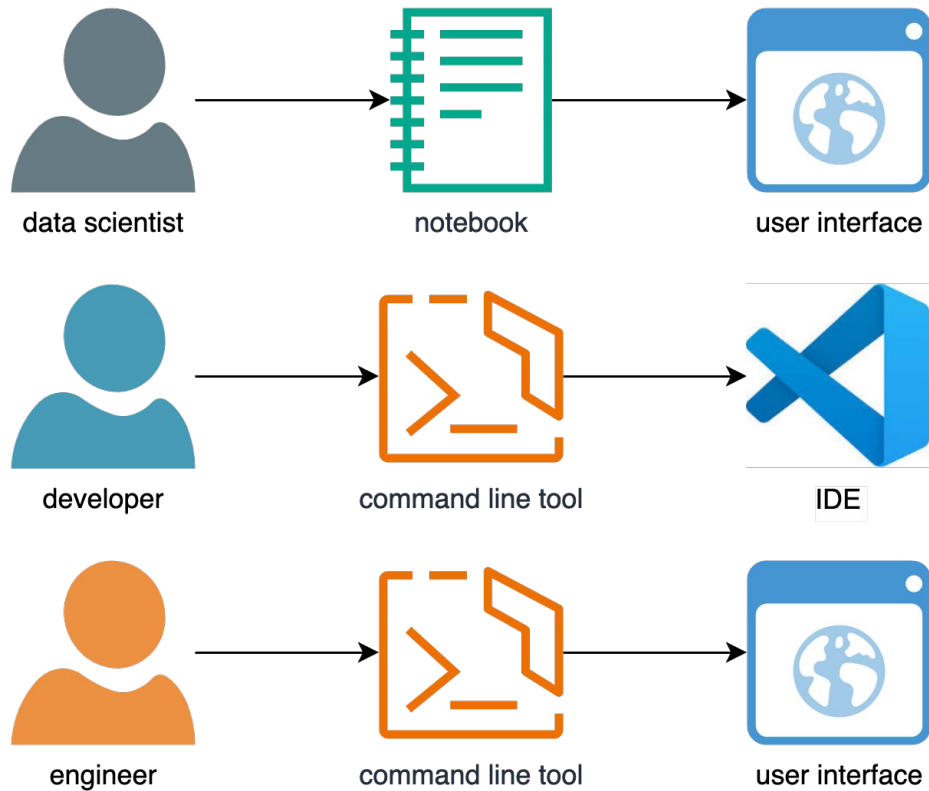

data scientist → notebook → user interface

🤗 Client use cases

data scientist → notebook → user interface

developer → command line tool → IDE

🤗 Client use cases

| data scientist | → | notebook | → | user interface |

| developer | → | command line tool | → | IDE |

| engineer | → | command line tool | → | user interface |

🤗 Use cases tools

Command line tool (GitHub contributors)

- ggerganov/llama.cpp (915)
- vllm-project/vllm (601)
- ggerganov/whisper.cpp (419)

User Interface (GitHub contributors)

- oobabooga/text-generation-webui (332)
- ollama/ollama (316)
- nomic-ai/gpt4all (115)

https://github.com/ggerganov/llama.cpp

🤗 Use cases tools

Command line tool (GitHub contributors)

- ggerganov/llama.cpp (915)
- vllm-project/vllm (601)
- ggerganov/whisper.cpp (419)

Packages (GitHub contributors)

- 🤗 /transformers (2799), 🤗 /diffusers (816)
- gradio-app/gradio (339), streamlit (239)
- 🤗 /huggingface_hub (206)

User Interface (GitHub contributors)

- oobabooga/text-generation-webui (332)
- ollama/ollama (316)
- open-webui/open-webui (236)

Frameworks (GitHub contributors)

- tensorflow/tensorflow (3604)
- langchain-ai/langchain (3177)
- run-llama/llama_index (1275)

The platform

Questions ?

@PyDataVenice #17 #Meetup #PyData

Questions ?

@PyDataVenice #17 #Meetup #PyData

Excursus

- 10 👥 21/09/19, PyDataVE #0
- 9 👥 13/12/19, PyDataVE #1 - Introduzione ai Notebook
- 23 👥 31/03/20, PyDataVE #2 - DataViz (geo e non)
- ..
- 26 👥 19/04/24, PyDataVE #15 - OpenCV Use Cases
- 66 👥 27/06/24, PyDataVE #16 - ComputerVision & DecisionMaking
- 59 👥 24/10/24, PyDataVE #17 - HuggingFace and libraries

YouTube: @pydatavenice/videos & @pydatavenice/streams

@PyDataVenice #17 #Meetup #PyData

Prossimi incontri

- 3 - 5 dicembre, https://pydata.org/global2024

Prossimi incontri

- 3 - 5 dicembre, https://pydata.org/global2024
- 5 - 6 dicembre, https://pydata.org/global2024/impact-program

Prossimi incontri

- 3 - 5 dicembre, https://pydata.org/global2024
- 5 - 6 dicembre, https://pydata.org/global2024/impact-program
- YouTube: @PyDataTV/playlists

Prossimi incontri

- 3 - 5 dicembre, https://pydata.org/global2024
- 5 - 6 dicembre, https://pydata.org/global2024/impact-program
- YouTube: @PyDataTV/playlists

- giovedì 12 dicembre ore 19:00

Prossimi incontri

- 3 - 5 dicembre, https://pydata.org/global2024
- 5 - 6 dicembre, https://pydata.org/global2024/impact-program
- YouTube: @PyDataTV/playlists

- giovedì 12 dicembre ore 19:00

English

Proposte

Thanks for listening.

@PyDataVenice #17 #Meetup #PyData