

# Docling IBM

<https://medium.com/ai-software-engineer/i-tested-ibms-ai-python-library-that-turns-messy-pdfs-into-perfect-data-you-need-this-0bb89cde25ef>

## 1. Install Macbook Airr

```
pip install -U "huggingface_hub[cli]"
docling
```

```
import time
from docling.document_converter import DocumentConverter

# This can be a local file path or a URL
source = "https://assets.accessible-digital-documents.com/uploads/2017/01/sample-tables.pdf" # The Docling technical paper itself
# source = "https://www.w3.org/WAI/WCAG20/Techniques/working-examples/PDF20/table.pdf"
# source = "https://drive.google.com/file/d/1cswCefl5i-mf8qpv_NjLov5niYmkt3l-/view?usp=sharing&a=1"
# PDF
# source = "https://docs.google.com/presentation/d/1ktn3-xE47NXPkokTOnB0VmXBQGwBk8nB/edit?usp=sharing&oid=107038111603680757101&rtpof=true&sd=true"
# source = "https://docs.google.com/spreadsheets/d/12B733LSKBIDX2HeX3yGM4h7WIARVBtdt/edit?usp=sharing&oid=107038111603680757101&rtpof=true&sd=true"
converter = DocumentConverter()

start_time = time.time()
result = converter.convert(source)
end_time = time.time()
# Export to markdown - clean, structured, beautiful
print(result.document.export_to_markdown())
print(f"🕒 Thời gian chạy: {end_time - start_time:.2f} giây")
```

## 2. Install Colab

- [https://colab.research.google.com/drive/1p6ugNtfseDk\\_ZV0tlzTQENa9OTw4ftuY#scrollTo=1P9pm9uBES2y](https://colab.research.google.com/drive/1p6ugNtfseDk_ZV0tlzTQENa9OTw4ftuY#scrollTo=1P9pm9uBES2y)

## 3. Install ThueGPU A6000

```
pip install notebook
jupyter notebook --ip 0.0.0.0 --allow-root
```

```
# Install docker
echo "----- 🟢 Start install docker -----"
sudo apt update
sudo apt install -y apt-transport-https ca-certificates curl software-properties-common
curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo apt-key add -
sudo add-apt-repository -y "deb [arch=amd64] https://download.docker.com/linux/ubuntu focal stable"
apt-cache policy docker-ce
sudo apt install -y docker-ce
echo "----- 🟡 Finish install docker -----"
echo "----- 🟢 Start install nvidia support 4 docker -----"
curl -fsSL https://nvidia.github.io/libnvidia-container/gpgkey | sudo gpg --dearmor -o /usr/share/keyrings/nvidia-container-toolkit-keyring.gpg
```

```
curl -s -L https://nvidia.github.io/libnvidia-container/stable/deb/nvidia-container-toolkit.list \
| sed 's#deb https://#deb [signed-by=/usr/share/keyrings/nvidia-container-toolkit-keyring.gpg] https://#g' \
| sudo tee /etc/apt/sources.list.d/nvidia-container-toolkit.list
sudo apt-get update
sudo apt-get install -y nvidia-container-toolkit
sudo nvidia-ctk runtime configure --runtime=docker
sudo systemctl restart docker
echo "----- 🟡 Finish install nvidia support 4 docker -----"
```

```
docker run --gpus all -p 5001:5001 ghcr.io/docling-project/docling-serve-cu126
```