aws machine learning

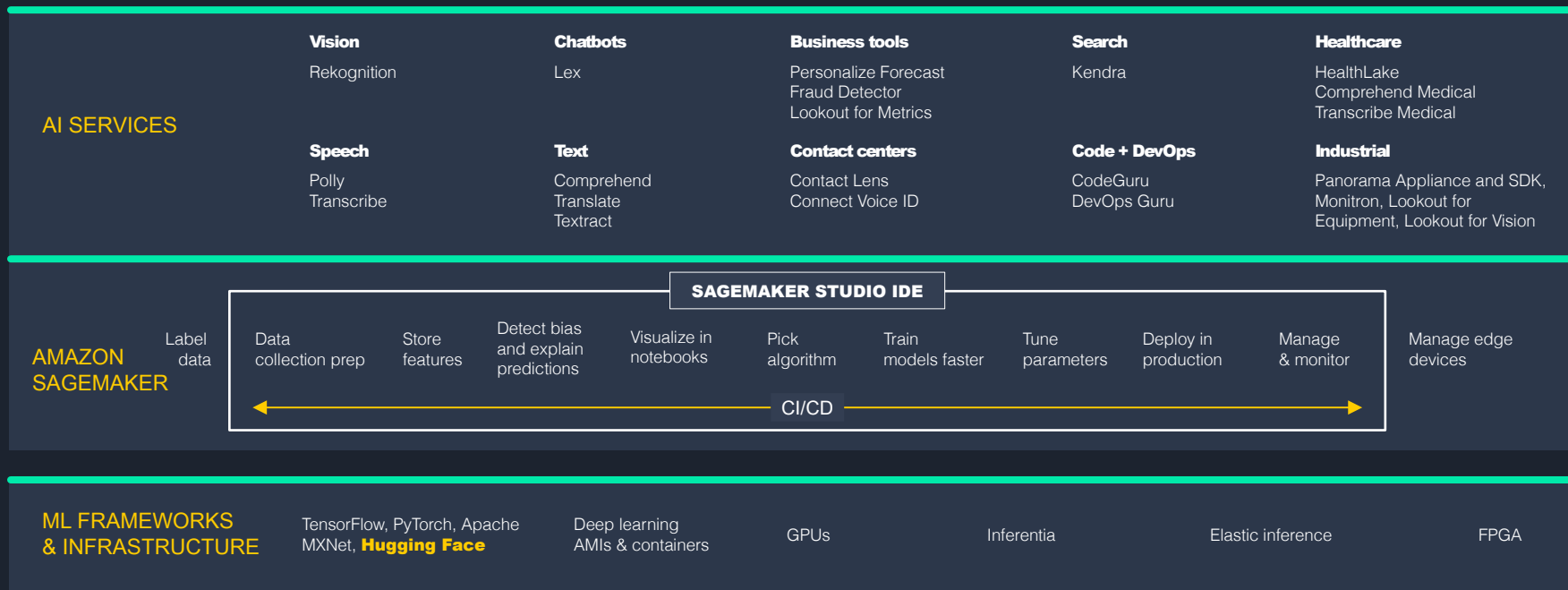# Hugging Face on Amazon SageMaker

Scale and accelerate your NLP projects with Hugging Face models

# AWS ML Vision

Make machine learning accessible to all customers (developers, data scientists, business users)

aws machine learning

# The AWS ML stack

## Broadest and most complete set of machine learning capabilities

### AI SERVICES

**Vision**
Rekognition

**Chatbots**
Lex

**Business tools**
Personalize Forecast
Fraud Detector
Lookout for Metrics

**Search**
Kendra

**Healthcare**
HealthLake
Comprehend Medical
Transcribe Medical

**Speech**
Polly
Transcribe

**Text**
Comprehend
Translate
Textract

**Contact centers**
Contact Lens
Connect Voice ID

**Code + DevOps**
CodeGuru
DevOps Guru

**Industrial**
Panorama Appliance and SDK,
Monitron, Lookout for
Equipment, Lookout for Vision

### AMAZON SAGEMAKER

**SAGEMAKER STUDIO IDE**

| Label data | Data collection prep | Store features | Detect bias and explain predictions | Visualize in notebooks | Pick algorithm | Train models faster | Tune parameters | Deploy in production | Manage & monitor | Manage edge devices |
|---|---|---|---|---|---|---|---|---|---|---|

CI/CD

### ML FRAMEWORKS & INFRASTRUCTURE

TensorFlow, PyTorch, Apache MXNet, **Hugging Face**

Deep learning AMIs & containers

GPUs

Inferentia

Elastic inference

FPGA

aws

# Customers building NLP solutions on AWS

# Hugging Face

- Startup founded in 2016

- Based in Paris and New York

- Open source solutions for Natural Language Processing

- $40 million Series B rounds a few days ago

aws machine learning

# Hugging Face libraries

**Open-source**
Datasets, Tokenizers and Transformers

**Popular**
52k+ GitHub stars (March 2021), 1M+ downloads per month

**Intuitive**
NLP-specific Python frontends based on PyTorch or TensorFlow

**State of the art**
Transformer-based models are state-of-the-art, enable transferlearning and scale

**Comprehensive**
Model zoo with 6000+ model architectures, 160+ languages

aws machine learning

# What are the Hugging Face libraries?

**Open-source**
`Datasets`, `Tokenizers` and `Transformers`

**Popular**
52k+ GitHub stars (March 2021), 1M+ downloads per month

**Intuitive**
NLP-specific Python frontends based on PyTorch or TensorFlow

**State of the art**
Transformer-based models are state-of-the-art, enable transferlearning and scale

**Comprehensive**
Model zoo with 6000+ model architectures, 160+ languages

aws machine learning

# Introducing a new Hugging Face experience in Amazon SageMaker

**A new Deep Learning Container (DLC) for SageMaker,** developed with HF (training only for now). `Datasets`, `Transformers` and `Tokenizers`, with SageMaker integration such as SageMaker Debugger and the SageMaker distributed training libraries

**A Hugging Face Estimator in the SageMaker SDK** to launch NLP scripts on scalable, cost-effective SageMaker training jobs with minimal code changes (script mode), and without any Docker and infrastructure work

**An example gallery** with samples of Hugging Face scripts on Amazon SageMaker

**Supported and maintained by AWS**

aws machine learning

# What are the benefits of running Hugging Face in Amazon SageMaker?

**Cost-effective** - SageMaker optimizes scale, performance, and costs

**MLOps-ready** – Monitoring, profiling and debugging (SageMaker Debugger & Profiler), experiment management (SageMaker Experiments), etc.
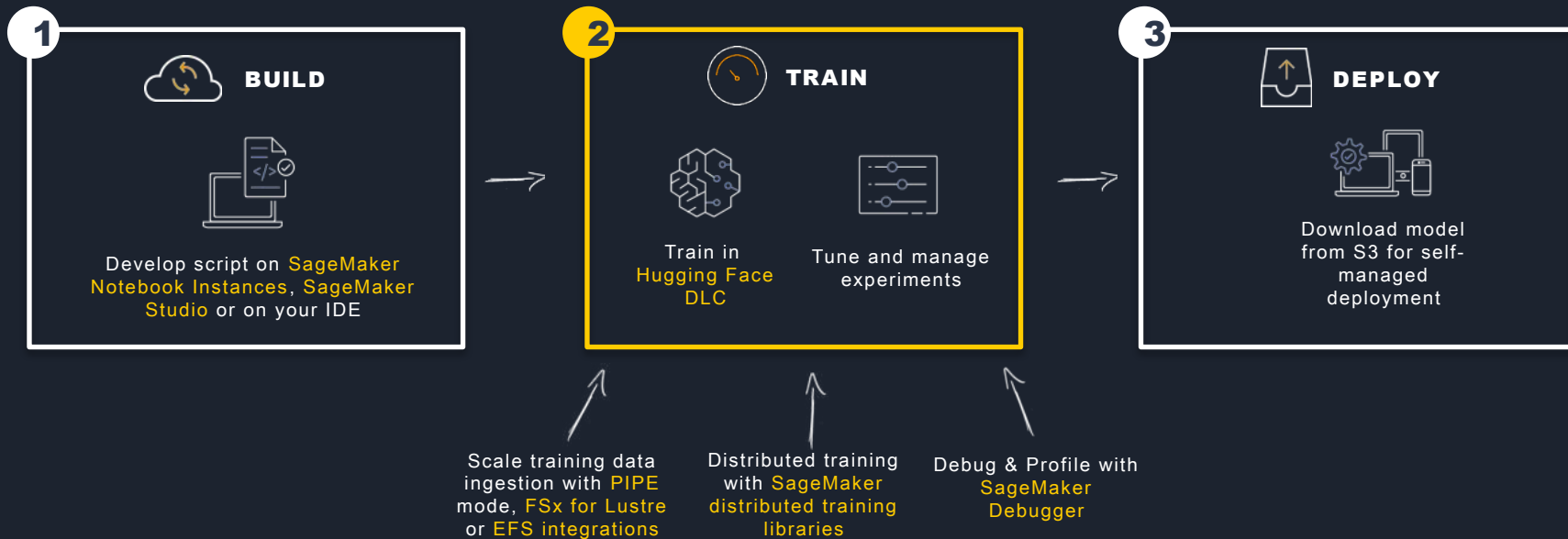
**Scalable** – Able to run on clusters of GPUs, with efficient data-parallel and model-parallel distribution provided by Amazon SageMaker. Ability to launch several concurrent jobs at the same time with the async mode of the API

**Secure** – High bar on security, with available mechanisms including encryption at rest and in transit, VPC connectivity and fine-grained IAM permissions

aws machine learning

# Integrated workflow with Amazon SageMaker

**1**

**BUILD**

Develop script on SageMaker Notebook Instances, SageMaker Studio or on your IDE

**2**

**TRAIN**

Train in Hugging Face DLC

Tune and manage experiments

**3**

**DEPLOY**

Download model from S3 for self-managed deployment

Scale training data ingestion with PIPE mode, FSx for Lustre or EFS integrations

Distributed training with SageMaker distributed training libraries

Debug & Profile with SageMaker Debugger

aws machine learning

# Customers using Hugging Face on SageMaker

## Quantum Health

Quantum Health is on a mission to make healthcare navigation smarter, simpler, and more cost-effective for everyone. They use Hugging Face and Amazon SageMaker for use cases like text classification, text summarization, and Q&A to help agents and members.

" *We use Hugging Face and Amazon SageMaker a lot for many NLP use cases such as text classification, text summarization, and Q&A with the goal of helping our agents and members. For some use cases, we just use the Hugging Face models directly and for others we fine tune them on SageMaker. We are excited about the integration of Hugging Face Transformers into Amazon SageMaker to make use of the distributed libraries during training to shorten the training time for our larger datasets*

Jorge Grisman, NLP Data Scientist, Quantum Health

aws

# Customers using Hugging Face on SageMaker

## Kustomer

Kustomer is a customer service CRM platform for managing high support volume with ease. They use machine learning models to help customers contextualize conversations, remove time-consuming tasks, and deflect repetitive questions.

> *In our business, we use machine learning models to help customers contextualize conversations, remove time-consuming tasks, and deflect repetitive questions. We use Hugging Face and Amazon SageMaker extensively, and we are excited about the integration of Hugging Face Transformers into SageMaker since it will simplify the way we fine tune machine learning models for text classification and semantic search*

Victor Peinado, ML Software Engineering Manager, Kustomer

aws

Get started with
# Hugging Face on Amazon SageMaker

- Launch blog post on the AWS Machine Learning blog

- SageMaker docs: https://docs.aws.amazon.com/sagemaker

- SageMaker SDK docs: https://sagemaker.readthedocs.io/

- Hugging Face examples: https://github.com/huggingface/notebooks/

aws