AWS re:Invent

GPSTEC335

Power your machine learning with Amazon SageMaker & Fast.ai

Amit Mukherjee
Partner Solutions Architect
Amazon Web Services

Kris Skrinak
Partner Solutions Architect, Global Machine Learning Segment Lead
Amazon Web Services

AWS re:Invent

aws

THE AWS ML STACK

Broadest and deepest set of capabilities

AI services

VISION			SPEECH		LANGUAGE		CHATBOTS	FORECASTING	RECOMMENDATIONS
Amazon Rekognition Image	Amazon Rekognition Video	Amazon Textract	Amazon Polly	Amazon Transcribe	Amazon Translate	Amazon Comprehend & Amazon Comprehend Medical	Amazon Lex	Amazon Forecast	Amazon Personalize

ML services

Amazon SageMaker

Ground Truth	Notebooks	Algorithms + Marketplace	Reinforcement Learning	Training	Optimization	Deployment	Hosting

ML frameworks + infrastructure

FRAMEWORKS	INTERFACES
TensorFlow, mxnet, PYTORCH, Chainer	GLUON, Keras

INFRASTRUCTURE							
EC2 P3 & P3DN	EC2 G4	EC2 C5	FPGAs	DL Containers & AMIs	AWS IoT Greengrass	Amazon Elastic Inference	AWS Inferentia

Amazon SageMaker: Build, train, and deploy ML models at scale

Prebuilt notebooks for common problems	Built-in, high-performance algorithms	One-click training on the highest performing infrastructure	Model optimization	One-click deployment	Fully managed with auto scaling for 75% less
Collect and prepare training data	Choose and optimize your ML algorithm	Set up and manage environments for training	Train and tune ML models	Deploy models in production	Scale and manage the production environment

Intuit

NFL

tinder

KIA

CONVOY

SIEMENS

Dow Jones

Snapchat

Sony

Amazon SageMaker training options

- Matrix factorization
- Regression
- Principal component analysis
- K-means clustering
- Gradient boosted trees
- And more!

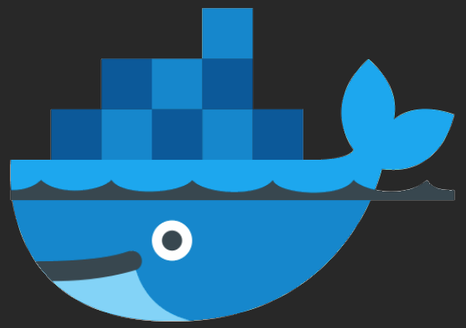
Built-in algorithms

Bring your own script
(Amazon SageMaker managed container)

Bring your own algorithm
(you build the Docker container)

Subscribe to Algorithms and Model Packages on AWS Marketplace

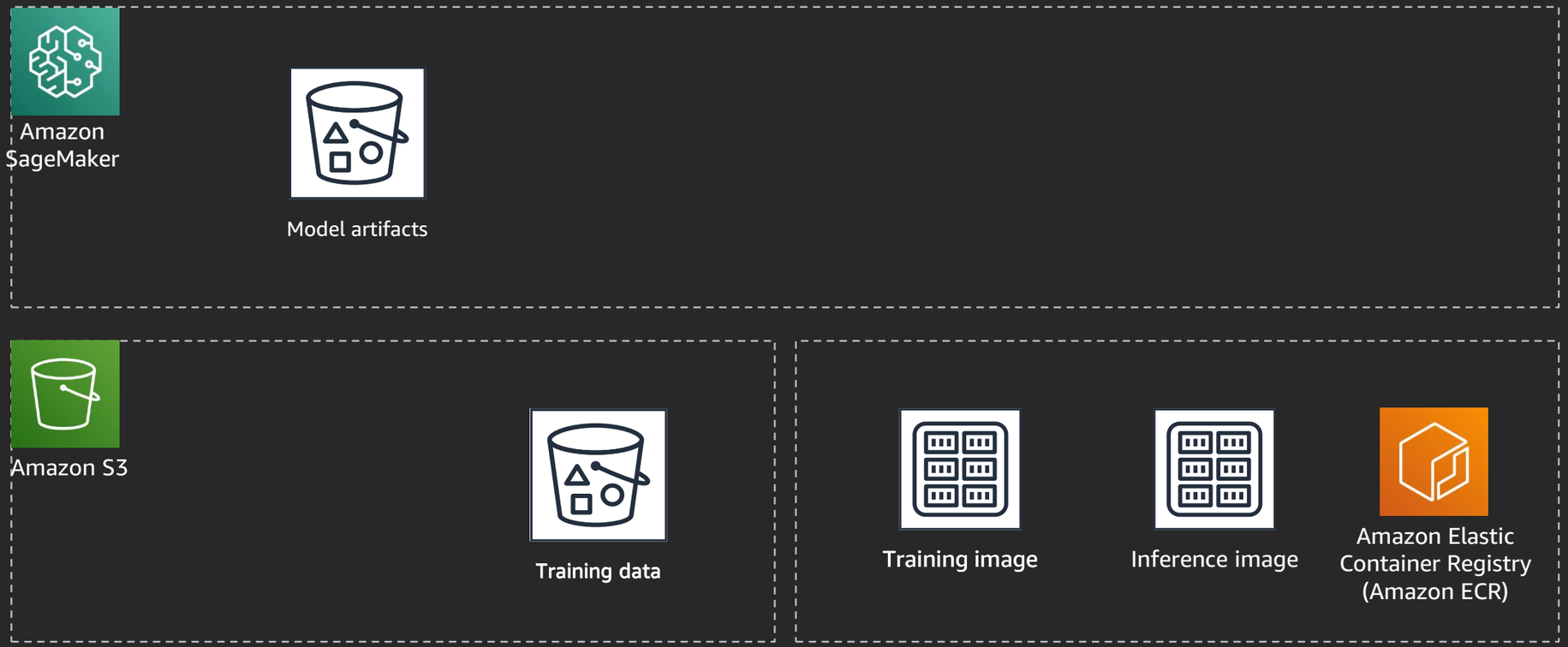
Script mode



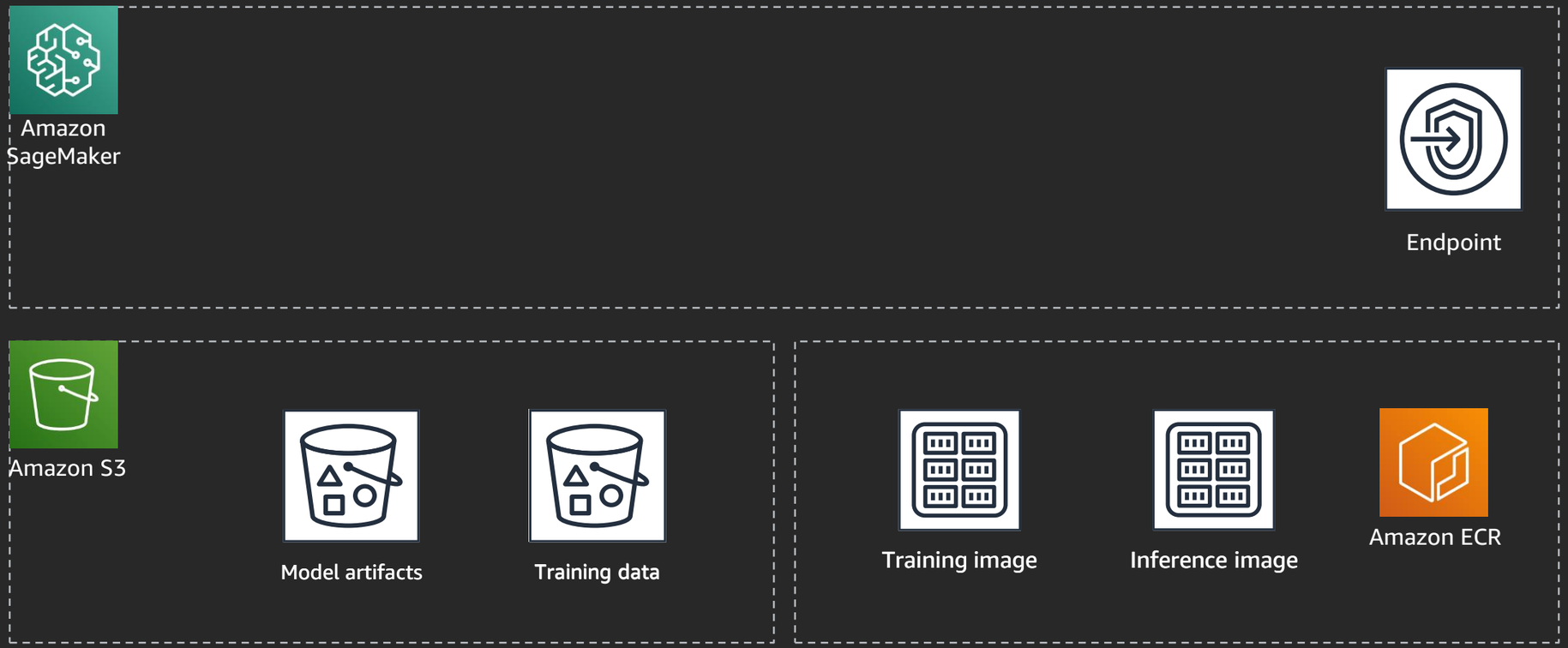
AWS
Managed

1. Point to the AWS-managed container of your choice
2. Write your model as a bundle of files
3. Specify the entry point in the Amazon SageMaker Estimator
4. Include any extra libraries with requirements.txt
5. Use our web server for inference

Amazon SageMaker training service



Amazon SageMaker hosting service



Fast.ai: Making neural nets uncool again

Deep learning for coders

- MOOC taught by Jeremy Howard
- Part 1 and Part 2 both have new v3

Fast.ai library

- Deep learning framework built on PyTorch
- SOTA best practices made easy
- Easily extensible

Founded by Jeremy Howard and Rachel Thomas

Dogs vs. cats

Arch: Resnet34*	**Fast.ai**	**Keras**
Lines of code (excluding imports)	5	31
Stage 1 error	0.70%	2.05%
Stage 2 error	0.50%	0.80%
Test time augmentation (TTA) error	0.30%	N/A*
Stage 1 time	4:56	8:30
Stage 2 time	6:44	17:38

*Keras does not provide resnet 34 or TTA

Source: https://www.fast.ai/2018/10/02/fastai-ai/

Why Fast.ai with Amazon SageMaker?

Fast.ai sample training job in Amazon SageMaker

```
pets_estimator = PyTorch(entry_point='source/pets.py',
                         base_job_name='fastai-pets',
                         role=role,
                         framework_version='1.0.0',
                         train_instance_count=1,
                         train_instance_type='ml.p3.2xlarge')

pets_estimator.fit(data_location)
```

```
print('Creating DataBunch object')
    data = ImageDataBunch.from_name_re(path_img, fnames, pat,
                                       ds_tfms=get_transforms(),
                                       size=args.image_size,
                                       bs=args.batch_size).normalize(imagenet_stats)

    # create the CNN model
    print('Create CNN model from model zoo')
    print(f'Model architecture is {args.model_arch}')
    arch = getattr(models, args.model_arch)
    print("Creating pretrained conv net")
    learn = create_cnn(data, arch, metrics=error_rate)
    print('Fit for 4 cycles')
    learn.fit_one_cycle(4)
    learn.unfreeze()
    print('Unfreeze and fit for another 2 cycles')
    learn.fit_one_cycle(2, max_lr=slice(1e-6,1e-4))
    print('Finished Training')
```

Fast.ai model deployment in Amazon SageMaker

```
pets_model=PyTorchModel(model_data=pets_estimator.model_data,
                        name=pets_estimator._current_job_name,
                        role=role,
                        framework_version=pets_estimator.framework_version,
                        entry_point=pets_estimator.entry_point,
                        predictor_cls=ImagePredictor)

pets_predictor = pets_model.deploy(initial_instance_count=1,
                                   instance_type='ml.c5.large')
```

```
    print('Creating DataBunch object')
    empty_data = ImageDataBunch.load_empty(path)
    arch_name = os.path.splitext(os.path.split(glob.glob(f'{model_dir}/resnet*.pth')[0])[1])[0]
    print(f'Model architecture is: {arch_name}')
    arch = getattr(models, arch_name)
    learn = create_cnn(empty_data, arch, pretrained=False).load(path/f'{arch_name}')
    return learn
```

Demo

Getting started

Fast.ai online course – http://course.fast.ai/

Fast.ai documentation – https://docs.fast.ai/

Fast.ai library – https://github.com/fastai/fastai

Amazon SageMaker PyTorch container with Fast.ai v1.0.39 –
https://github.com/aws/sagemaker-pytorch-container

Amazon SageMaker Python SDK – https://sagemaker.readthedocs.io/en/stable/

Amazon SageMaker & Fast.ai blog –
https://aws.amazon.com/blogs/machine-learning/building-training-and-deploying-fastai-models-with-amazon-sagemaker/

FAQ

1. Can I train Fast.ai models in Amazon EC2 instead of Amazon SageMaker?
2. When choosing the deep learning AMI, do I need to install a CUDA driver separately?
3. Can I use Spot Instance to train Fast.ai models in Amazon SageMaker to optimize cost?
4. Does Fast.ai have any AWS Marketplace algorithm or model packages?
5. What version of Fast.ai library is attached to AWS-managed PyTorch container?
6. How I can use the Fast.ai higher version of V1.0.39 in Amazon SageMaker?
7. What option do I have if I don’t have Amazon SageMaker console access to train Fast.ai models from a local laptop?
8. I have already trained Fast.ai models on-premises; can I host the same in Amazon SageMaker?
9. Does Fast.ai library support distributed training?
10. What is TTA in Fast.ai?
11. Can I deploy Fast.ai models in AWS Lambda?

Thank you!

Amit Mukherjee
amitmukh@amazon.com

Kris Skrinak
skrinak@amazon.com

AWS re:Invent

aws

Please complete the session survey in the mobile app.