aws INNOVATE
DATA AND AI/ML EDITION

March 9th, 2023

aws INNOVATE

DATA AND AI/ML EDITION

MLT04

Train ML models at scale with Amazon SageMaker

Ayman Salama

Sr. Partner Solution Architect, AWS

Agenda

Large Scale Training Challenges

Amazon SageMaker End-to-End Solution

Types of Distributed Training

Distributed Training features

Demo

Why do we need distributed training?

Models grow faster than hardware, leading to bottlenecks

INCREASING COMPLEXITY

- Businesses need higher precision in their model predictions
- Results in larger and more complex models
- Requires frequent retraining of models

INCREASING COSTS

- Increasing compute power required for frequent training of larger models drives up cost to train
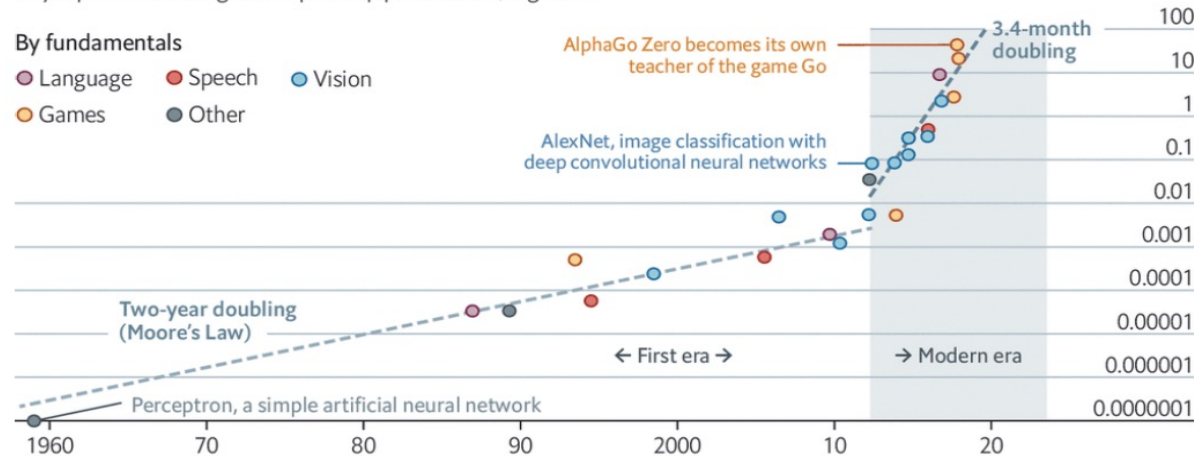- Becomes a barrier for innovation and growth
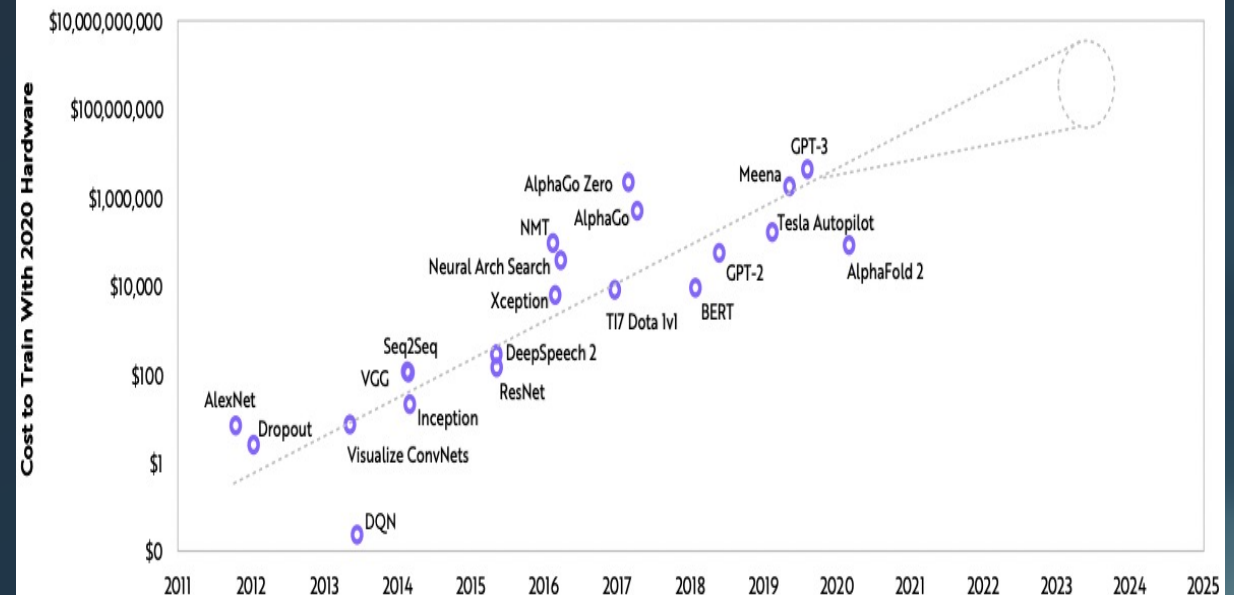

Compute power ~ 2x every 3.4 months

Model size increase ~ 10x/ year, Cost of Training increase ~ 100x by 2025

*According to Open AI as of 2018
*ARK Invest Big Ideas 2021

Amazon SageMaker ephemeral training clusters

Amazon S3

Write trained model to Amazon S3

Sends your data

SageMaker Instance

SageMaker Instance

Docker Container

docker run image train

Stores metadata with hyperparameters

SageMaker control plane

Streams logs to CloudWatch

Amazon CloudWatch

model.fit()

SageMaker Notebook

SageMaker Volume

Downloads your image

Amazon ECR

Fully managed by SageMaker
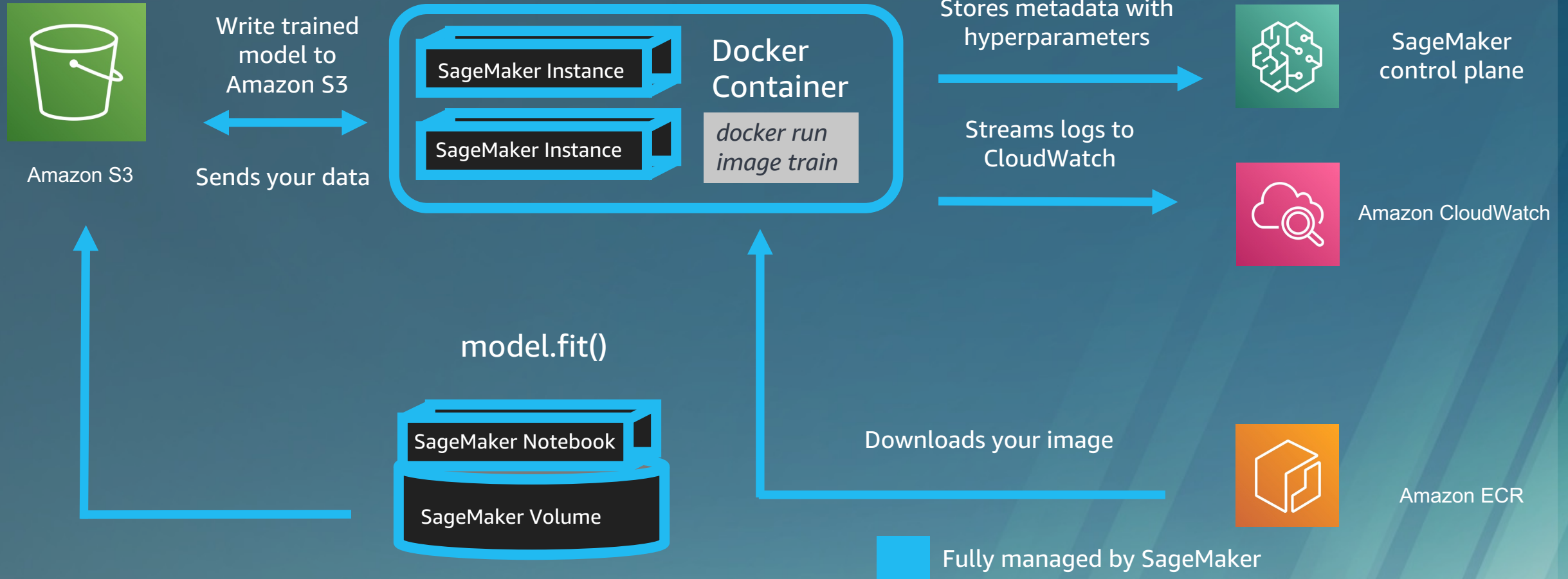
ML.EC2 INSTANCES FOR DISTRIBUTED TRAINING

Broadest and most complete set of Distributed Training Infrastructure choices

|	**G4dn**			**P3**			**P3dn**			**P4d**		
Ideal for	Accelerated training of small to medium sized models with less than 100M parameters			Training medium to large models with 100M to 300M parameters Good for single node distributed training			Training large models with more than 300M parameters Spot Training may provide better price-performance than P4d Good for multi-node distributed training			Customer looking for best training performance on the cloud Training large models with more than 300M parameters Good for multi-node distributed training		
Key features	16 GB/GPU	PCIe only	25 - 50 Gbps networking and 100 Gbps on bare-metal	16 GB/GPU	200 - 300 GB/s NVLink (4, 8 GPUs)	10-25 Gbps networking	32 GB/GPU	300 GB/s NVLink (8 GPUs)	100 Gbps networking	40 GB/GPU	600 GB/s NVLink (8 GPUs)	400 Gbps networking
GPU Config	1, 4, or 8 NVIDIA Tesla T4s			1, 4, or 8 NVIDIA Tesla V100s			8 NVIDIA Tesla V100s			8 NVIDIA Tesla A100s (latest)		

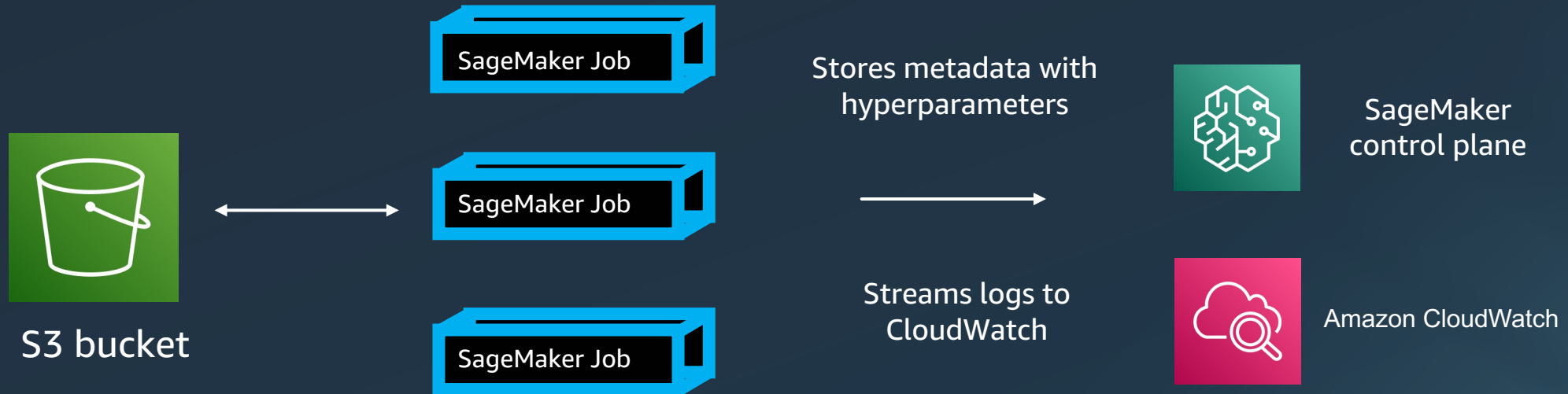# There are many kinds of distributed training
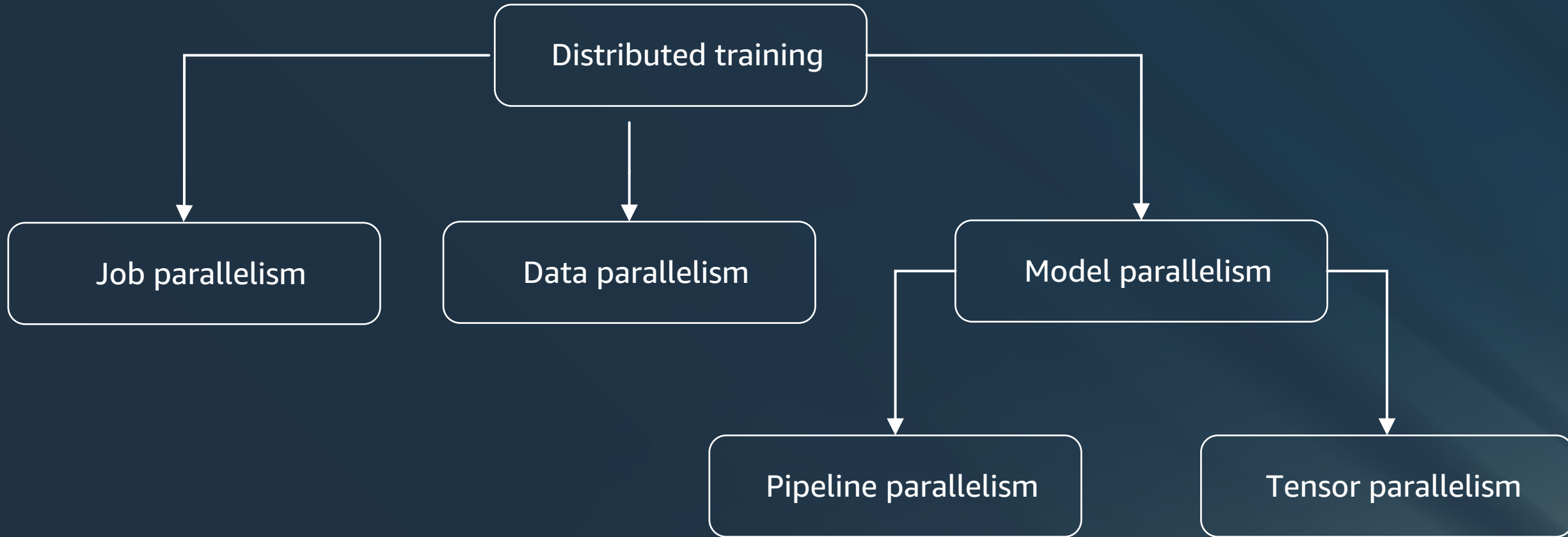
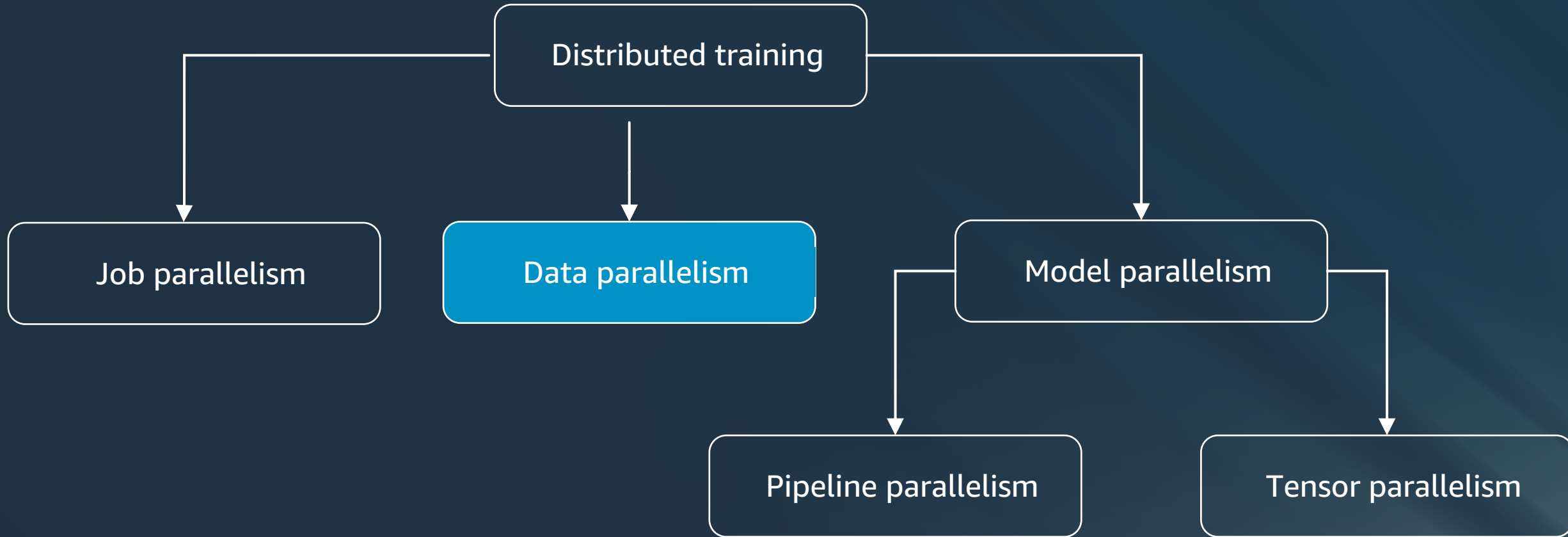# There are many kinds of distributed training
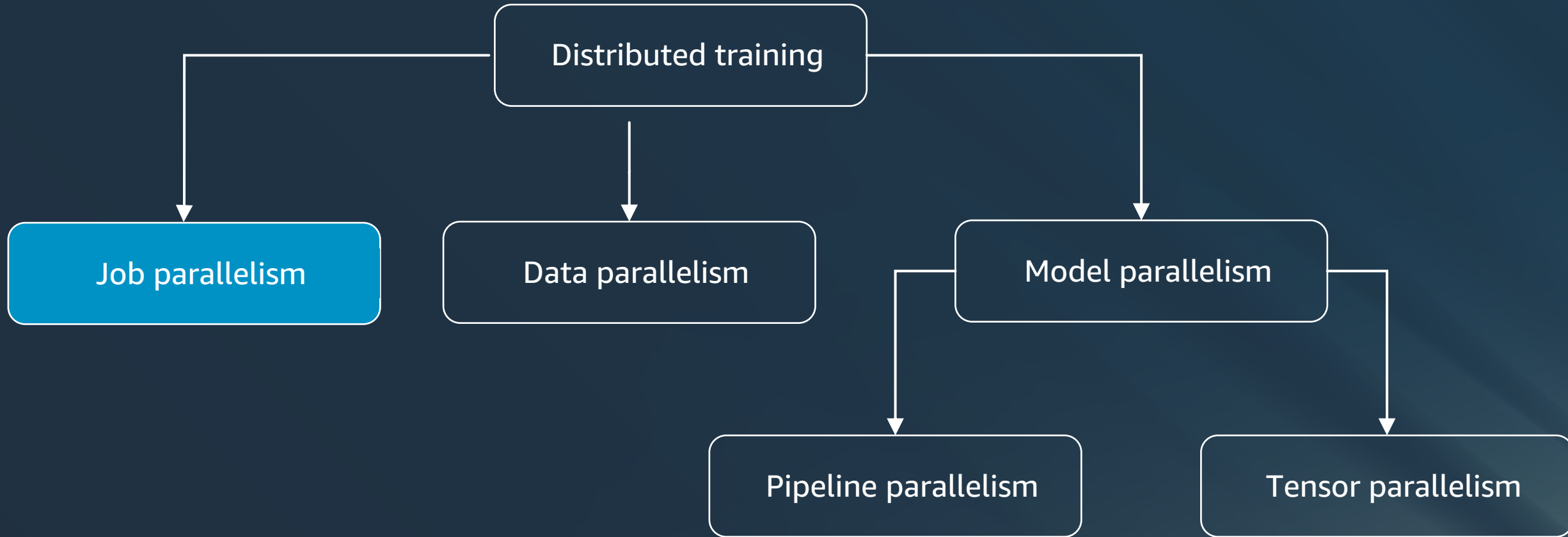
Train with parallel jobs at high frequency

1. Each job can train as many models as you need.

2. You can use *warm pools* to retrain as quickly as possible

```
for model in list_of_models:
    s3_input = get_data(model)
    s3_output = get_location(model)
    estimator = get_estimator(model, s3_output)
    estimator.fit(s3_input, wait=False)
```


S3 bucket

SageMaker Job

SageMaker Job

SageMaker Job

Stores metadata with hyperparameters

Streams logs to CloudWatch

SageMaker control plane

Amazon CloudWatch

There are many kinds of distributed training

```
                        ┌─────────────────────┐
                        │ Distributed training │
                        └─────────────────────┘
        ┌───────────────────────┼───────────────────────┐
        ▼                       ▼                       ▼
┌─────────────────┐   ┌─────────────────┐     ┌─────────────────┐
│ Job parallelism │   │ Data parallelism│     │ Model parallelism│
└─────────────────┘   └─────────────────┘     └─────────────────┘
                                         ┌─────────┴─────────┐
                                         ▼                   ▼
                               ┌─────────────────┐  ┌─────────────────┐
                               │Pipeline parallelism│ │ Tensor parallelism│
                               └─────────────────┘  └─────────────────┘
```
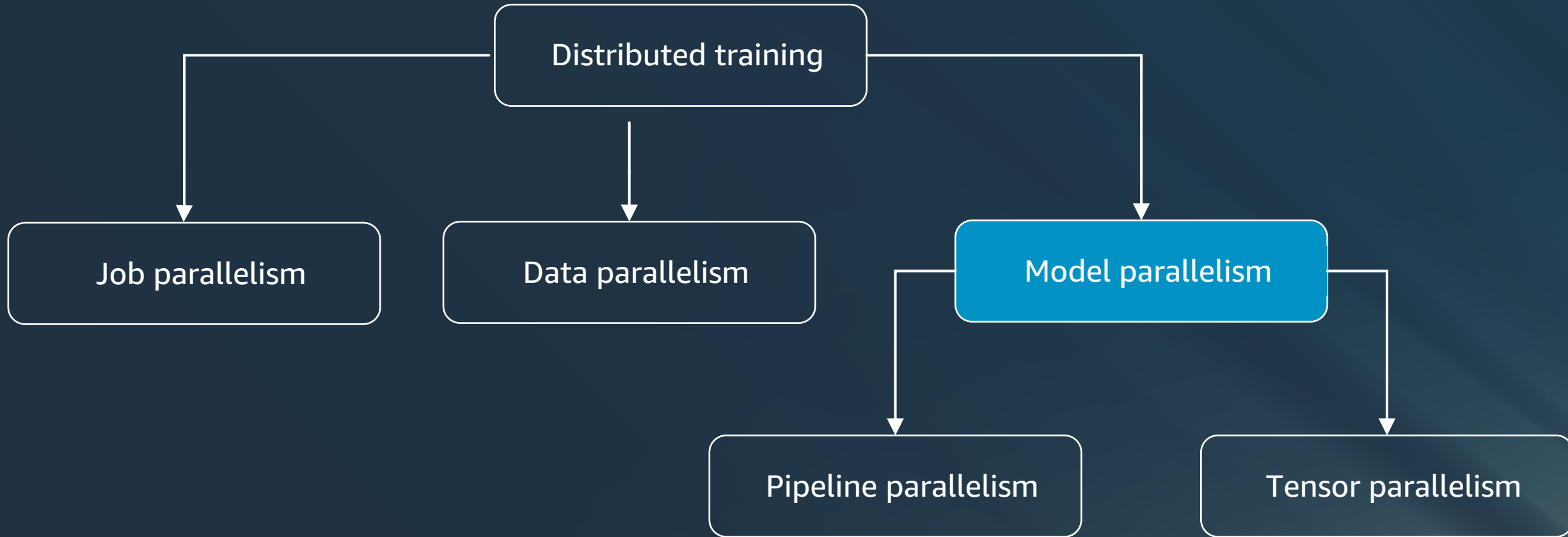
Data parallelism, think "massive data"

Amazon SageMaker Distributed Data Parallel

- **Optimized backend for distributed training of deep learning models in TensorFlow, PyTorch**

- **Accelerates training for network-bound workloads**

- **Built and optimized for AWS network topology and hardware**

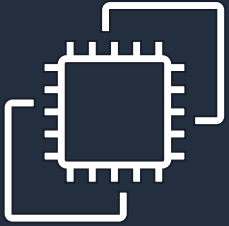- **20%–40% faster and cheaper than NCCL and MPI-based solutions. Best performance on AWS for large clusters.**
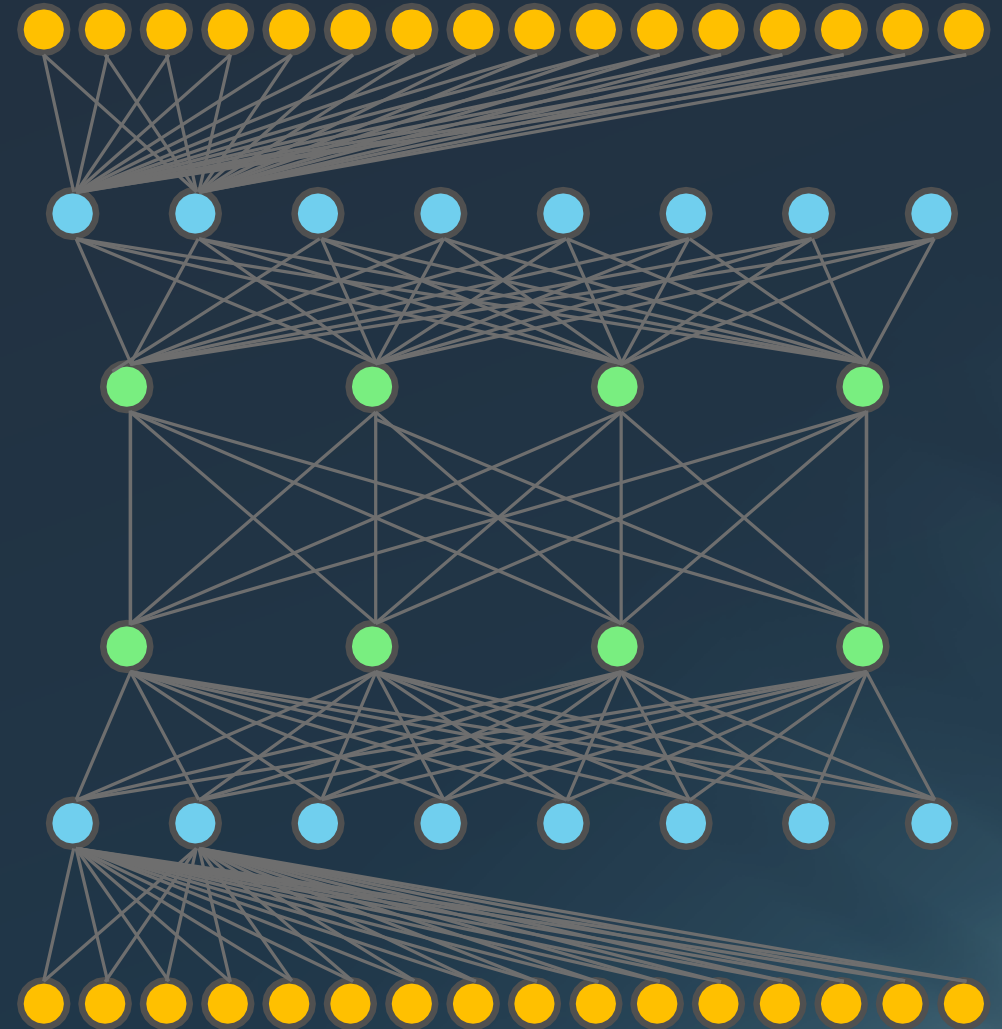
There are many kinds of distributed training

Model parallel, think "massive models"

Model parallelism on Amazon SageMaker (SMP)

**Automated
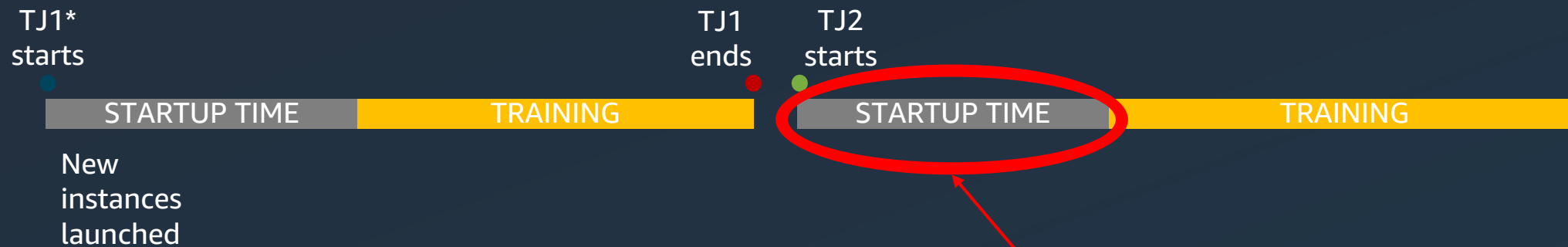model partitioning**

**Interleaved
pipelined training**

**Managed
SageMaker training**
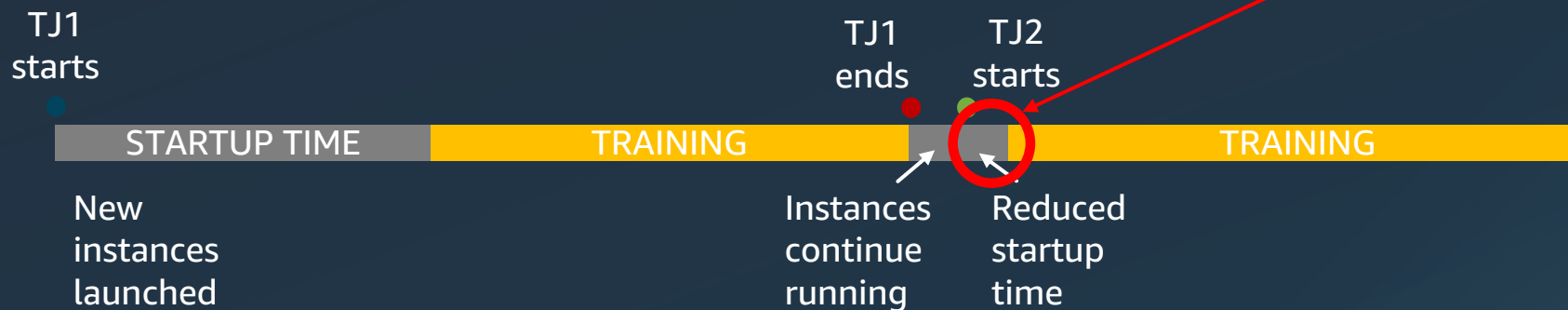
**Clean
framework integration**

Demo

Home Launcher Terminal arn:aws:sagemaker:u SageMaker-train-gpt-simple.ip

Filter files by name

/ sagemaker-distributed-training-
workshop / 2_model_parallel_sdp_Lab

Name

data_pipeline.py
data_prep_512.py
learning_rates.py
memory_tracker.py
requirements.txt
SageMaker-train-gpt-simple.ipynb
sharded_data_parallel_checkpoi...
train_gpt_simple.py

Home

Customize layout

▼ Quick actions

Open Launcher
Create notebooks and other resources

Import & prepare data visually

Open the Getting Started notebook

Read documentation

View guided tutorials

▼ Prebuilt and automated solutions
Deploy built-in algorithms, pre-built solutions, example notebooks, and build models from visual interface.

Quick start solutions
Pretrained models, notebooks, and prebuilt solutions

AutoML
Automatically build, train, and tune the best ML models

▼ Workflows and tasks
Kick off a new step in the machine learning workflow.

Prepare data
- Connect to data sources
- Transform, analyze, and export data

Build, train, tune model
- View all experiments
- Create AutoML experiment

Deploy model
- Get endpoint recommendation
- Manage endpoints and optimize performance

Thank you!

Ayman Salama

aymanq@amazon.com