
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

CMP407

Elastic Fabric Adapter advanced topics for AI/ML and HPC

Matt Koop
Principal Engineer
AWS

Pierre-Yves Aquilanti
Head of Framework ML Solutions
AWS

Maxime Hugues
Solutions Architect
AWS

Timothy Brown
Solutions Architect
AWS

Evan Bollig
Solutions Architect
AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

- Background on Elastic Fabric Adapter (EFA) and the components we will be using during this session
- Hands on with EFA and supporting tools

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is EFA?

Elastic Fabric Adapter (EFA) is an OS-bypass, RDMA-enabled network adapter for high-performance inter-node communication.

This is common for high-performance interconnects

- It uses an AWS-designed transport, Scalable Reliable Datagram (SRD), built on top of the AWS Nitro System hardware.
- The AWS Nitro hardware provides offloading of tasks including I/O and networking from the standard host. Developing our own hardware allows AWS to provide innovations such as EFA.

This provides the elasticity and choice of instance types

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Wide variety of instances

EFA is available on nearly every new instance type that AWS is releasing

2019: 2 instance types → **2022**: 60+ instance types

- **Trn1**: AWS Trainium-based instance: 800Gbps EFA
- **Hpc6a**: AMD Milan, 100Gbps EFA
- **Hpc6id**: Intel Icelake, 200Gbps EFA
- **C6i.32xl**: Intel Ice Lake, 50Gbps EFA, 4GB/core
- **R6i.32xl**: Intel Ice Lake, 50Gbps EFA, 16GB/core
- **P4d:** 8x A100 NVIDIA GPUs, 400Gbps EFA
- …

General purpose:
m5dn.24xlarge | m5dn.metal | m5n.24xlarge | m5n.metal | m5zn.12xlarge | m5zn.metal | m6a.32xlarge | m6a.48xlarge | m6a.metal | m6i.32xlarge | m6i.metal | m6id.32xlarge | m6id.metal

Compute optimized:
c5n.18xlarge | c5n.9xlarge | c5n.metal | c6a.32xlarge | c6a.48xlarge | c6a.metal | c6gn.16xlarge | c6i.32xlarge | c6i.metal | c6id.32xlarge | c6id.metal | c7g.16xlarge | hpc6a.48xlarge | hpc6id.32xlarge

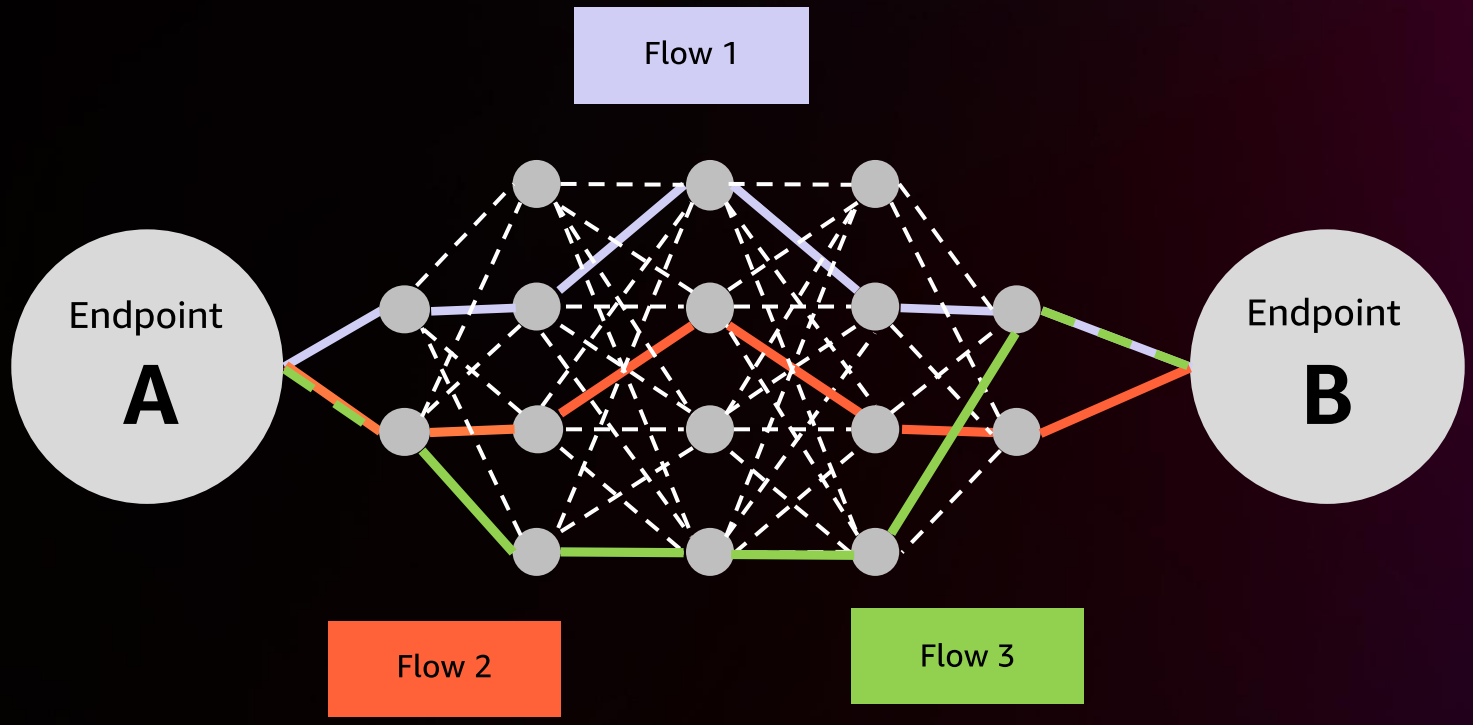
Memory optimized:
r5dn.24xlarge | r5dn.metal | r5n.24xlarge | r5n.metal | r6a.48xlarge | r6a.metal | r6i.32xlarge | r6i.metal | r6id.32xlarge | r6id.metal | x2d.32xlarge | x2d.metal | x2ed.32xlarge | x2ed.metal | x2iezn.12xlarge | x2iezn.metal | x2idn.32xlarge | x2iedn.32xlarge

Storage optimized:
i3en.24xlarge | i3en.12xlarge | i3en.metal | i4i.32xlarge | i4i.metal | im4gn.16xlarge

Accelerated computing:
dl1.24xlarge | g4dn.8xlarge | g4dn.12xlarge | g4dn.metal | g5.48xlarge | inf1.24xlarge | p3dn.24xlarge | p4d.24xlarge | trn1.32xlarge

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How it works:
Scalable Reliable Datagram (SRD)



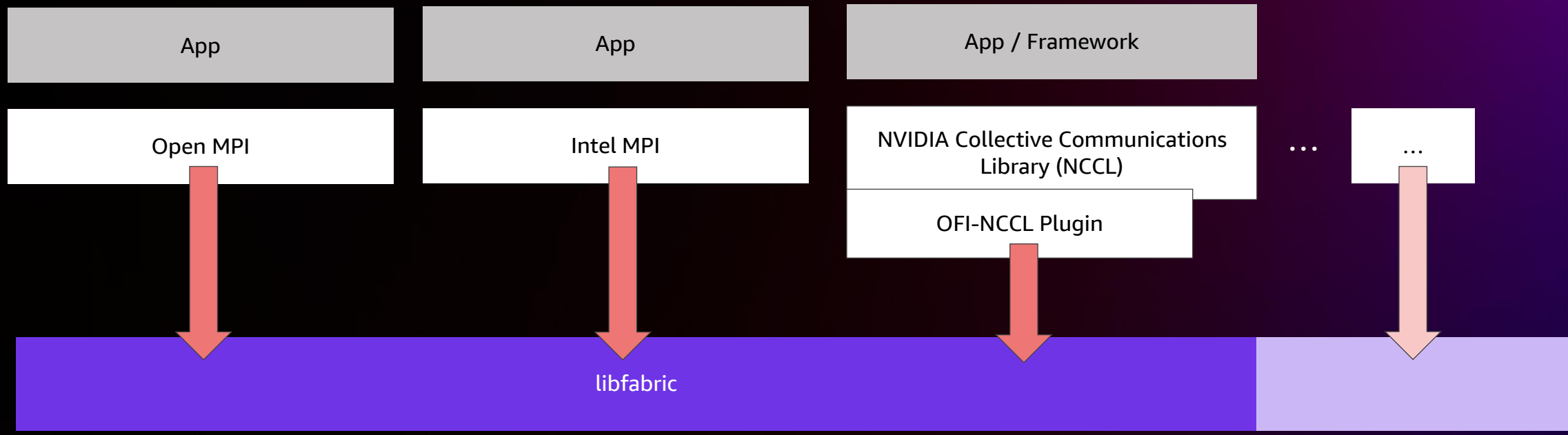
AWS-designed protocol that uses the many paths within the AWS network simultaneously

Designed into the AWS Nitro System Hardware

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How it works:
Software Stack

You can use EFA directly with no code changes, since many libraries support libfabric (which supports EFA)



aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Hands-on portion

1. Learn more about how EFA works
2. Set up the necessary network and instance infrastructure
3. Set up and learn about the software layers
4. Monitor your EFA devices

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Getting started with this workshop

- As a participant, you will have access to an AWS account with pre-provisioned infrastructure and IAM policies needed to complete this workshop.
- The AWS account will only be available for the duration of this workshop. You will lose access to the account thereafter.
- The optional pre-provisioned infrastructure will be deployed to a specific region (eu-north-1).
- Be sure to review the terms and conditions of the event. Do not upload any personal or confidential information in the account.

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

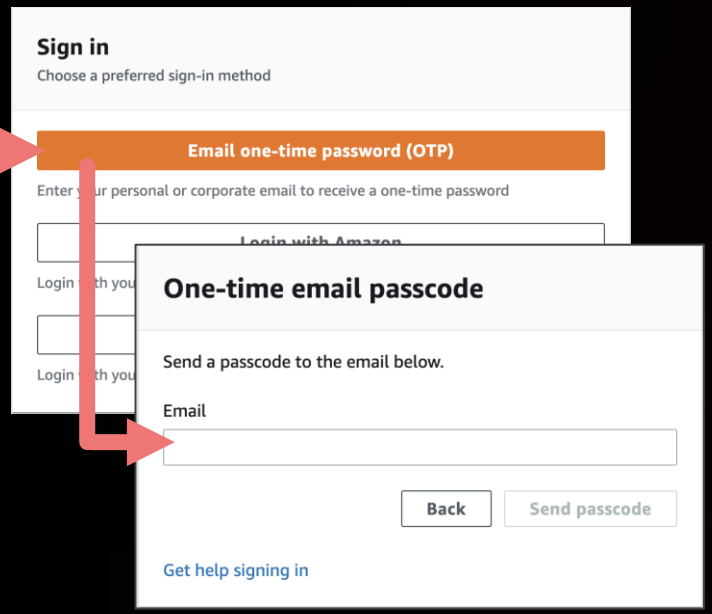
Let’s build!

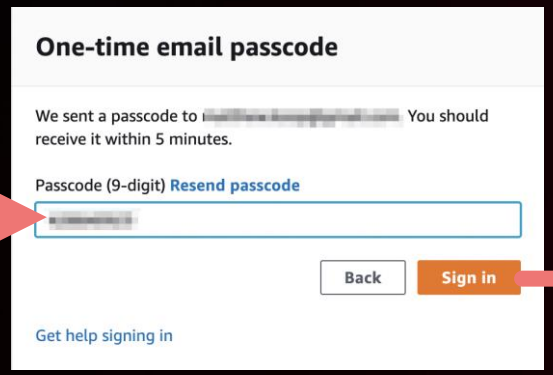
1. http://s12d.com/efa

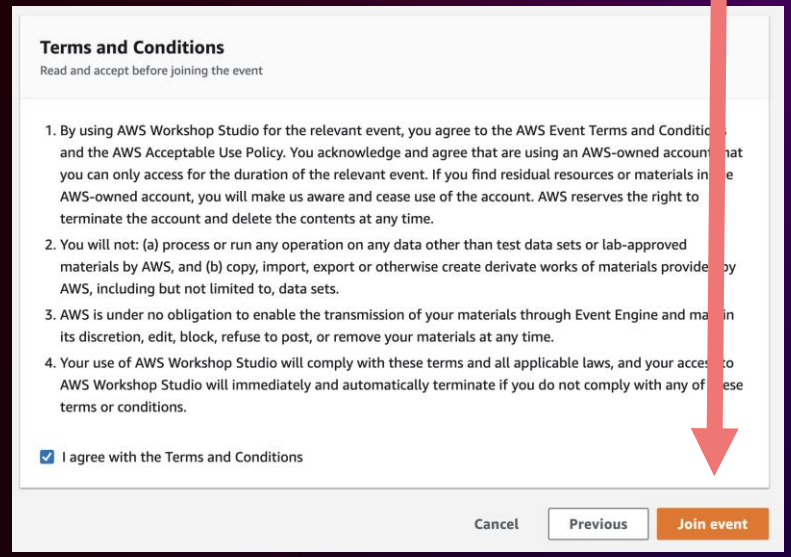
2. Email a one-time passcode

3. Enter passcode after it is emailed

4. Review terms and “Join Event”







© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Matt Koop
mkoop@amazon.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.