

University of Verona

Programming FPGA on cloud - Guide

VR443470 - Valentini Andrea

Last Update: 16th January 2024

Contents

1	Amazon	3
1.1	Introduction to AWS	3
1.1.1	Register a new account	3
1.1.2	Getting Started	5
1.2	Amazon EC2 F1 Instances	6
1.2.1	Description	6
1.2.2	Amazon EC2 Pricing	8
1.2.3	Development Kit	8
1.2.4	FPGA Developer AMI	9
1.2.5	Q&A	10
1.2.6	The reason why you shouldn't choose the EC2 F1 In- stances (for educational purposes)	13
1.2.7	Some interesting blog posts & articles	13
2	Microsoft	14
2.1	Azure Stack Edge Pro FPGA (Expired on February 2024)	14
2.2	Azure Stack Edge Pro 2	16
2.3	Azure Stack Edge Pricing	17

1 Amazon

1.1 Introduction to AWS

This guide is designed to help some developers, and not, to explore the cloud technologies to learn how to program an FPGA (on cloud).

1.1.1 Register a new account

From this site you can register on the AWS platform: <https://aws.amazon.com/>. You can create an account choosing two possibility plans:

- Personal account, with some limitations (see more: [here](#));
- Business account, but attention because you need to register a payment card (see more: [here](#)).

Although, there is an opportunity to students, and anyone else has an institutional e-mail, to register on the AWS Educate platform¹: <https://aws.amazon.com/education/awseducate/>. This plan allows you to access to AWS Educate lab, which provides access to the AWS Console to enable practical application of concepts without the need for credits (in summary, more resources without having to pay).

Unfortunately, Education Plan doesn't give you the "freedom" to do whatever you want because use AWS Console only meanwhile a learn course (sandbox). This choice is still valid because you can use the various tutorials to create guides (like this) or understand how-to-use EC2 F1.

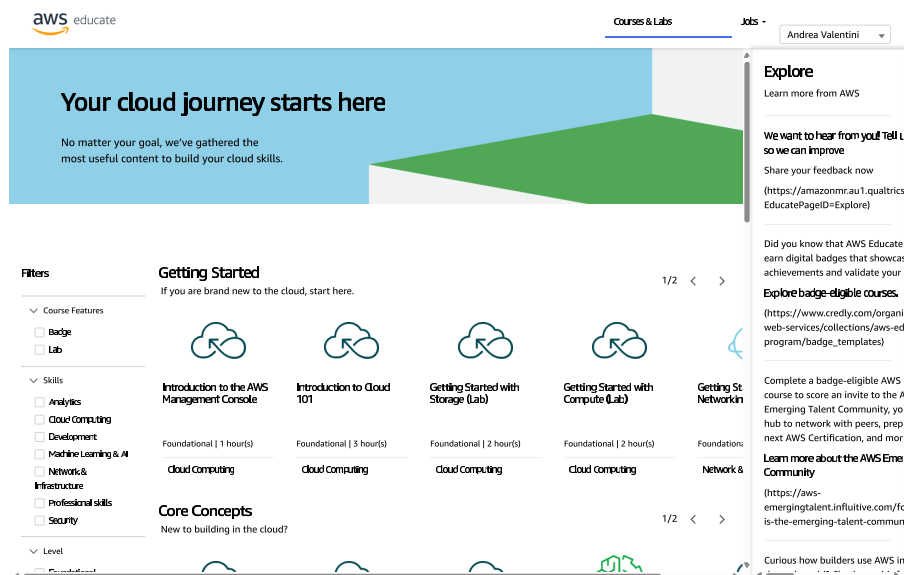



Figure 1: AWS Educate Dashboard.

¹FAQ: https://www.awseducate.com/registration/s/faqs?language=en_US

If you've chosen to use a personal/business plan, once inserted a pay card number, and confirmed it with a 1\$ fee you should have this screen:




Sign up for AWS

Select a support plan

Choose a support plan for your business or personal account. [Compare plans and pricing examples](#)
[You can change your plan anytime in the AWS Management Console.](#)


☒ **Basic support - Free**

- Recommended for new users just getting started with AWS
- 24x7 self-service access to AWS resources
- For account and billing issues only
- Access to Personal Health Dashboard & Trusted Advisor




☐ **Developer support - From \$29/month**


- Recommended for developers experimenting with AWS
- Email access to AWS Support during business hours
- 12 (business)-hour response times



☐ **Business support - From \$100/month**

- Recommended for running production workloads on AWS
- 24x7 tech support via email, phone, and chat
- 1-hour response times
- Full set of Trusted Advisor best-practice recommendations





Need Enterprise level support?

From \$15,000 a month you will receive 15-minute response times and concierge-style experience with an assigned Technical Account Manager. [Learn more](#)

[Complete sign up](#)

On this figure, you can also see the various plans offered. To compare plans and pricing examples, you can visit this page: [link](#).

Anyway, choosing a Free plan will automatic start an AWS Free Tier ([terms](#)). The 12 Month Free Tier is only available to new AWS customers, and is available for 12 months following your AWS sign-up date.

If you have not used the AWS resources provided under an Offer during the previous 3 months, we may reclaim those AWS resources after giving you 30 days' notice. Even if your AWS resources are reclaimed, you may continue to participate in Offers using new AWS resources.

1.1.2 Getting Started

The first thing to do to understand how AWS works is read the Getting Started guide: <https://aws.amazon.com/getting-started/>. In this site, you can find some useful guides that explain cloud functions and others Amazon products.

After that, AWS shows you a sea of tutorials such as “Build a Serverless Web Application”, “Deploy a Web Application on Amazon EC2”, etc. Here’s the link: [Hands-on Tutorials](#).

If you chosen a Business (or free) account plan, this is your dashboard:

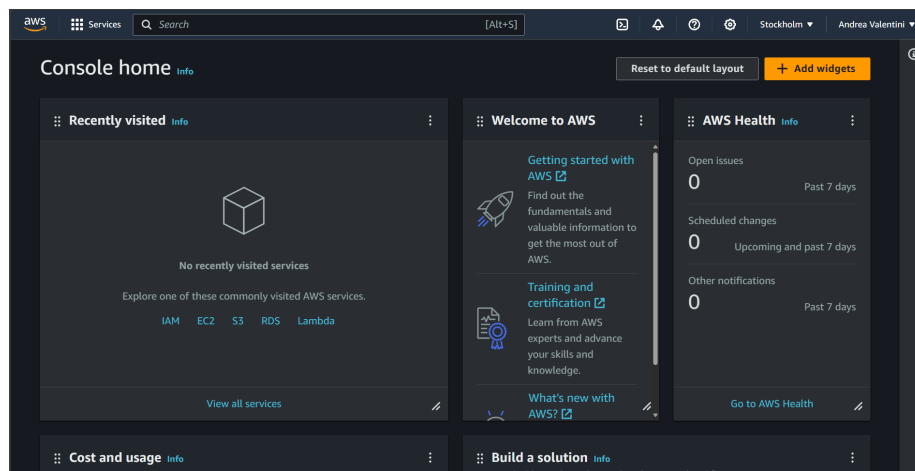


Figure 2: AWS Business/Free Dashboard.

On the left side you can see all resources created and recently visited; at the top you can see all hypertext link that can help you to start.

1.2 Amazon EC2 F1 Instances

1.2.1 Description

Amazon EC2 F1 instances use FPGAs to enable delivery of custom hardware accelerations. With F1 instances you can:

- Develop
- Simulate
- Debug
- Compile

hardware acceleration code, including an FPGA Developer AMI (see 1.2.4 cap.). Furthermore, EC2 supporting hardware level development on the cloud.

F1 instances provide diverse development environments: from low-level hardware developers to software developers who are more comfortable with C/C++ and openCL environments (see more on 1.2.3 par.).

Once your FPGA design is complete, you can register it as an Amazon FPGA Image (AFI), and deploy it to your F1 instance in just a few clicks. You can reuse your AFIs as many times as you like, and across as many F1 instances as you like. There is no software charge for the development tools when using the FPGA developer AMI and you can program the FPGAs on your F1 instance as many times as you like with no additional fees.

Instances Features:

- High frequency Intel Xeon Scalable Processors (Broadwell E5-2686 v4)
- NVMe SSD Storage
- Support for Enhanced Networking

FPGA Features:

- Xilinx Virtex UltraScale+ VU9P FPGAs
- 64 GiB of ECC-protected memory on 4x DDR4
- Dedicated PCI-Express x16 interface
- Approximately 2.5 million logic elements
- Approximately 6,800 Digital Signal Processing (DSP) engines
- FPGA Developer AMI (see 1.2.4)

Instance	FPGAs	vCPU	Memory (GiB)	Instance Storage (GB)	Networking Performance (Gbps)***
f1.2xlarge	1	8	122	1 × 470	Up to 10
f1.4xlarge	2	16	244	1 × 940	Up to 10
f1.16xlarge	8	64	976	4 × 940	25

Table 1: The specifications of each type of *EC2 F1 Instances*.

For **f1.16xlarge** instances, the dedicated PCI-e fabric lets the FPGAs share the same memory space and communicate with each other across the fabric at up to 12 Gbps in each direction.

All instances have the following specs:

- 2.3 GHz (base) and 2.7 GHz (turbo) Intel Xeon E5-2686 v4 Processor
- [Intel AVX*](#), [Intel AVX2*](#), [Intel Turbo](#)
- [EBS Optimized](#)
- [Enhanced Networking*](#)

* AVX and AVX2 are only available on instances launched with HVM AMIs.

*** Instances marked with "Up to" Network Bandwidth have a baseline bandwidth and can use a network I/O credit mechanism to burst beyond their baseline bandwidth on a best effort basis. For more information, see [instance network bandwidth](#).

1.2.2 Amazon EC2 Pricing

The following table is a summary of the prices available. The costs depend on which nation you choose.

Name	On-Demand Price/hr	1-yr Reserved Instance Effective Hourly	3-yr Reserved Instance Effective Hourly
f1.2xlarge	\$1.65	\$1.06	\$0.76
f1.4xlarge	\$3.30	\$2.12	\$1.52
f1.16xlarge	\$13.20	\$8.50	\$6.10

Table 2: Prices shown are for Linux/Unix in the US East (Northern Virginia) AWS Region and rounded to the nearest cent. Prices for 1-year and 3-year reserved instances are for "Partial Upfront" payment options.

You can visit the price page for more details: [click here](#). Although Amazon says you can try EC2 F1 for free, this is untrue. Amazon allows you to use (for 750 hours) a virtual instance (a virtual resource) to give you a taste. Instead, Amazon EC2 F1 are dedicated instances (FPGA).

At least you can estimate the costs with the [AWS Pricing Calculator](#).

1.2.3 Development Kit

Once you created an AWS account, you **must** downloading the AWS FPGA Development Kit by this official repository: <https://github.com/aws/aws-fpga>. It includes all documentation ([shortcut url](#)) on:

- F1
- Internal FPGA interfaces
- Compiler scripts for generating Amazon FPGA Images (AFIs)

AWS FPGAs support multiple development environments to serve both hardware and software developers:

- The FPGA Hardware Development Kit (HDK, [shortcut url](#)) provides fully custom hardware development;
- The FPGA Software Development Kit (SDK, [shortcut url](#)) environment allows developing accelerations using C/C++/OpenCL code with no hardware knowledge needed.

Again, this kit is **essential to develop fast FPGA for the EC2 F1 instances**.

1.2.4 FPGA Developer AMI

The FPGA Developer AMI is an Amazon Linux 2 based AMI provided by Amazon Web Services. The AMI is pre-built with FPGA development tools required to develop and use custom FPGAs for hardware acceleration. The FPGA Developer AMI includes Xilinx tools for simulating your FPGA design, compiling code and building your AFI (Amazon FPGA Image).

The FPGA Developer AMI includes Xilinx Vivado² at no additional software charge as well as a prepackaged tool development environment with scripts and tools for simulating your FPGA design and building and registering your AFI.

As mentioned above, developers can deploy the FPGA Developer AMI on an Amazon EC2 instance and quickly provision the resources they need to write and debug FPGA designs in the cloud. The FPGA Developer AMI is provided at no additional charge to Amazon EC2 users.

Typical Total Price (updated on 17th October 2023): \$0.796/hr.

See more: [AWS marketplace](#).

²Xilinx Vivado is a software suite produced by AMD (previously Xilinx) for synthesis and analysis of hardware description language (HDL) designs, superseding Xilinx ISE with additional features for system on a chip development and high-level synthesis. [Source](#).

1.2.5 Q&A

The following questions and answers can be found on the original site, but they're brought here for convenience ([source](#)).

Questions	Page
What is Amazon EC2 F1?	10
What are FPGAs and why do I need them?	10
How does F1 compare with traditional FPGA solutions?	10
What is an Amazon FPGA Image (AFI)?	11
How do I list my hardware acceleration on the AWS Marketplace?	11
What is available with F1 instances?	11
Do I need to be an FPGA expert to use an F1 instance?	11
I'm an FPGA developer; how do I get started with F1 instances?	12
I'm not an FPGA developer; how do I get started with F1 instances?	12
Does AWS provide a developer kit?	12
Can I use the HDK in my on-premises development environment?	12
Can I add an FPGA to any EC2 instance type?	12

Table of contents of Q&A. Click on the the page number to go to the question/answer.

☛ What is Amazon EC2 F1?

- ☛ Amazon EC2 F1 is a compute instance with programmable hardware you can use for application acceleration. The new F1 instance type provides a high performance, easy to access FPGA for developing and deploying custom hardware accelerations.

☛ What are FPGAs and why do I need them?

- ☛ FPGAs are programmable integrated circuits that you can configure using software. By using FPGAs you can accelerate your applications up to 30x when compared with servers that use CPUs alone. And, FPGAs are reprogrammable, so you get the flexibility to update and optimize your hardware acceleration without having to redesign the hardware.

☛ How does F1 compare with traditional FPGA solutions?

- ☛ F1 is an AWS instance with programmable hardware for application acceleration. With F1, you have access to FPGA hardware in a few simple clicks, reducing the time and cost of full-cycle FPGA development and scale deployment from months or years to days. While FPGA technology has been available for decades, adoption of

application acceleration has struggled to be successful in both the development of accelerators and the business model of selling custom hardware for traditional enterprises, due to time and cost in development infrastructure, hardware design, and at-scale deployment. With this offering, customers avoid the undifferentiated heavy lifting associated with developing FPGAs in on-premises data centers.

☛ **What is an Amazon FPGA Image (AFI)?**

- ☛ The design that you create to program your FPGA is called an Amazon FPGA Image (AFI). AWS provides a service to register, manage, copy, query, and delete AFIs. After an AFI is created, it can be loaded on a running F1 instance. You can load multiple AFIs to the same F1 instance, and can switch between AFIs in runtime without reboot. This lets you quickly test and run multiple hardware accelerations in rapid sequence. You can also offer to other customers on the AWS Marketplace a combination of your FPGA acceleration and an AMI with custom software or AFI drivers.

☛ **How do I list my hardware acceleration on the AWS Marketplace?**

- ☛ You would develop your AFI and the software drivers/tools to use this AFI. You would then package these software tools/drivers into an Amazon Machine Image (AMI) in an encrypted format. AWS manages all AFIs in the encrypted format you provide to maintain the security of your code. To sell a product in the AWS Marketplace, you or your company must sign up to be an AWS Marketplace reseller, you would then submit your AMI ID and the AFI ID(s) intended to be packaged in a single product. AWS Marketplace will take care of cloning the AMI and AFI(s) to create a product, and associate a product code to these artifacts, such that any end-user subscribing to this product code would have access to this AMI and the AFI(s).

☛ **What is available with F1 instances?**

- ☛ For developers, AWS is providing a Hardware Development Kit (HDK) to help accelerate development cycles, a FPGA Developer AMI for development in the cloud, an SDK for AMIs running the F1 instance, and a set of APIs to register, manage, copy, query, and delete AFIs. Both developers and customers have access to the AWS Marketplace where AFIs can be listed and purchased for use in application accelerations.

☛ **Do I need to be an FPGA expert to use an F1 instance?**

- ☛ AWS customers subscribing to an F1-optimized AMI from AWS Marketplace do not need to know anything about FPGAs to take advantage of the accelerations provided by the F1 instance and the AWS Marketplace. Simply subscribe to an F1-optimized AMI from the AWS Marketplace with an acceleration that matches the workload.

The AMI contains all the software necessary for using the FPGA acceleration. Customers need only write software to the specific API for that accelerator and start using the accelerator.

☛ **I'm an FPGA developer; how do I get started with F1 instances?**

- ☛ Developers can get started on the F1 instance by creating an AWS account and downloading the AWS Hardware Development Kit (HDK). The HDK includes documentation on F1, internal FPGA interfaces, and compiler scripts for generating AFI. Developers can start writing their FPGA code to the documented interfaces included in the HDK to create their acceleration function. Developers can launch AWS instances with the FPGA Developer AMI. This AMI includes the development tools needed to compile and simulate the FPGA code. The Developer AMI is best run on the latest C5, M5, or R4 instances. Developers should have experience in the programming languages used for creating FPGA code (i.e. Verilog or VHDL) and an understanding of the operation they wish to accelerate.

☛ **I'm not an FPGA developer; how do I get started with F1 instances?**

- ☛ Customers can get started with F1 instances by selecting an accelerator from the AWS Marketplace, provided by AWS Marketplace sellers, and launching an F1 instance with that AMI. The AMI includes all of the software and APIs for that accelerator. AWS manages programming the FPGA with the AFI for that accelerator. Customers do not need any FPGA experience or knowledge to use these accelerators. They can work completely at the software API level for that accelerator.

☛ **Does AWS provide a developer kit?**

- ☛ Yes. The Hardware Development Kit (HDK) includes simulation tools and simulation models for developers to simulate, debug, build, and register their acceleration code. The HDK includes code samples, compile scripts, debug interfaces, and many other tools you will need to develop the FPGA code for your F1 instances. You can use the HDK either in an AWS provided AMI, or in your on-premises development environment. These models and scripts are available publicly with an AWS account.

☛ **Can I use the HDK in my on-premises development environment?**

- ☛ Yes. You can use the Hardware Development Kit HDK either in an AWS-provided AMI, or in your on-premises development environment.

☛ **Can I add an FPGA to any EC2 instance type?**

- ☛ No. F1 instances comes in two instance sizes: f1.2xlarge, f1.4xlarge, and f1.16 xlarge.

1.2.6 The reason why you shouldn't choose the EC2 F1 Instances (for educational purposes)

The reason why you shouldn't choose the EC2 F1 Instances is quite simple: they're too expensive. If you are a student or a professor, this option is too much for a number of reasons:

1. As mentioned in table 1 on page 7, the computational power of each type of F1 is too much for a beginner that is learning how to program an FPGA. It's like saying: "I don't have a driving licence, so I should drive a Ferrari to train for the exam".
2. Although you can create a free account with 750 hours on EC2, you can't get the access to an EC2 F1 Instance(s) unless you pay a huge amount of money:

f1.2xlarge
Family: f1 8 vCPU 122 GiB Memory Current generation: true
On-Demand RHEL base pricing: 1.78 USD per Hour
On-Demand SUSE base pricing: 1.75 USD per Hour
On-Demand Linux base pricing: 1.65 USD per Hour
f1.4xlarge
Family: f1 16 vCPU 244 GiB Memory Current generation: true
On-Demand Linux base pricing: 3.3 USD per Hour
On-Demand RHEL base pricing: 3.43 USD per Hour
On-Demand SUSE base pricing: 3.4 USD per Hour
f1.16xlarge
Family: f1 64 vCPU 976 GiB Memory Current generation: true
On-Demand RHEL base pricing: 13.33 USD per Hour
On-Demand Linux base pricing: 13.2 USD per Hour
On-Demand SUSE base pricing: 13.3 USD per Hour

1.2.7 Some interesting blog posts & articles

- [University of California, Berkeley uses AWS Educate and Amazon FPGA Instances in Undergraduate Computer Architecture Course](#) by AWS Public Sector Blog Team
- [Time, Randomness, and a \\$100,000 Prize to Forever Change Blockchain](#) by Michael V. Copeland
- [How DNAnexus and Edico Genome are Powering Precision Medicine on Amazon Web Services \(AWS\)](#) by Aaron Friedman and Ujjwal Ratan
- [Bringing Datacenter-Scale Hardware-Software Co-design to the Cloud with FireSim and Amazon EC2 F1 Instances](#) by Mia Champion
- [Accelerating Precision Medicine at Scale](#) by Aaron Friedman and Angel Pizarro

2 Microsoft

2.1 Azure Stack Edge Pro FPGA (Expired on February 2024)

Microsoft provided an interesting service called *Azure Stack Edge Pro FPGA*. In summary, it's an AI-enabled edge computing device with network data transfer capabilities. Unfortunately, these interesting devices will reach end-of-life in February 2024. If you are considering new deployments, Microsoft recommend that you explore Azure Stack Edge Pro 2 (section 2.2 page 16) or Azure Stack Edge Pro GPU ([Azure Stack Edge Pro with GPU](#)) devices for your workloads.

Azure Stack Edge Pro with FPGA is a Hardware-as-a-service solution. Microsoft ships you a cloud-managed device with a built-in Field Programmable Gate Array (FPGA) that enables accelerated AI-inferencing and has all the capabilities of a network storage gateway.

Azure Data Box Edge is rebranded as Azure Stack Edge. You can find it in the marketplace on Azure Cloud Platform. Again, this product will expired, so you can order only Pro 2 or Pro GPU.

In the following pages you can find the researches making to find what Microsoft offers. An interesting official documentation is available online at [this link](#) or you can read the offline documentation findable in `resources` path in this project.

Name	Publisher
 Azure Stack Edge	Microsoft
 AMD-Xilinx Alveo U250 2022.1 Deployment VM - Ubuntu20.04	Xilinx Inc.
 Vitis 2020.2 Alveo U250 Development VM - Ubuntu20.04	Xilinx Inc.
 Vitis 2022.1 Alveo U250 Development VM - Ubuntu18.04	Xilinx Inc.
 Vitis 2022.1 Alveo U250 Development VM - Centos7.8	Xilinx Inc.
 Vitis 2022.1 Alveo U250 Development VM - Ubuntu20.04	Xilinx Inc.
 Vitis 2020.2 Alveo U250 Development VM - Centos7.8	Xilinx Inc.
 Vitis 2021.1 Alveo U250 Development VM - Ubuntu18.04	Xilinx Inc.
 Vitis 2020.2 Alveo U250 Development VM - Ubuntu18.04	Xilinx Inc.
 Vitis 2021.1 Alveo U250 Development VM - Centos7.8	Xilinx Inc.
 Vitis 2021.1 Alveo U250 Development VM - Ubuntu20.04	Xilinx Inc.
 Xilinx Alveo U250 2021.1 Deployment VM-Centos7.8	Xilinx Inc.
 Xilinx Alveo U250 Deployment VM - Ubuntu20.04	Xilinx Inc.
 AMD-Xilinx Alveo U250 2022.1 Deployment VM - Ubuntu18.04	Xilinx Inc.
 Xilinx Alveo U250 Deployment VM - Centos7.8	Xilinx Inc.
 Xilinx Alveo U250 Deployment VM - Ubuntu18.04	Xilinx Inc.
 Xilinx Alveo U250 2021.1 Deployment VM-Ubuntu20.04	Xilinx Inc.
 AMD-Xilinx Alveo U250 2022.1 Deployment VM - Centos7.8	Xilinx Inc.
 Xilinx Alveo U250 2021.1 Deployment VM-Ubuntu18.04	Xilinx Inc.
 Worksafely for Covid	Vulcan AI Pte. Ltd.

Figure 3: Using a student account on Microsoft Azure, you can search what do you want on Marketplace. It's also [available online](#). The services provided by AMD are Virtual Machines. In this devices you can find softwares to develop FPGA. Despite, [Azure Stack Edge](#) is used for over-the-network high performance data transfers. Azure Stack Edge physical devices are available in various form factors and can be tailored for harsh environmental conditions.

2.2 Azure Stack Edge Pro 2

Azure Stack Edge Pro 2 is an evolution of Azure Stack Edge. It uses one or two GPU and is often used for rapid Machine Learning (ML) inferencing.

Azure Stack Edge Pro 2 is a new generation of an **AI-enabled edge computing device** offered as a service from Microsoft. It offers the following benefits over its precursor, the Azure Stack Edge Pro series:

- This series offers multiple models that closely align with your compute, storage, and memory needs. Depending on the model you choose, the compute acceleration could be via one or two Graphical Processing Units (GPU) on the device.
- This series has flexible form factors with multiple mounting options. These devices can be rack mounted, mounted on a wall, or even placed on a shelf in your office.
- These devices have low acoustic emissions and meet the requirements for noise levels in an office environment.

Use cases

The Pro 2 series is designed for deployment in edge locations such as retail, telecommunications, manufacturing, or even healthcare. Here are the various scenarios where Azure Stack Edge Pro 2 can be used for rapid Machine Learning (ML) inferencing at the edge and preprocessing data before sending it to Azure.

- **Inference with Azure Machine Learning** - With this solution, you can run ML models to get quick results that can be acted on before the data is sent to the cloud. The full data set can optionally be transferred to continue to retrain and improve your ML models. For more information, see how to [Deploy Azure ML hardware accelerated models on Azure Stack Edge](#).
- **Preprocess data** - Transform data before sending it to Azure via compute options such as containerized workloads and Virtual Machines to create a more actionable dataset. Preprocessing can be used to:
 - Aggregate data.
 - Modify data, for example, to remove personal data.
 - Subset data to optimize storage and bandwidth, or for further analysis.
 - Analyze and react to IoT Events.
- **Transfer data over network to Azure** - Use this solution to easily and quickly transfer data to Azure to enable further compute and analytics or for archival purposes.

2.3 Azure Stack Edge Pricing

The devices seen in the previous chapters, can be ordered from the Azure Edge Hardware center. These devices are billed as a monthly service through the Azure portal. For more information, see [pricing](#).