



Cloud Computing & Big Data

PARALLEL & SCALABLE MACHINE LEARNING & DEEP LEARNING

Prof. Dr. – Ing. Morris Riedel

Associated Professor

School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland

Research Group Leader, Juelich Supercomputing Centre, Forschungszentrum Juelich, Germany

LECTURE 8

[in](#) @Morris Riedel

[@MorrisRiedel](#)

[@MorrisRiedel](#)

Infrastructure-As-A-Service (IAAS)

October 29, 2020
Online Lecture



EUROPEAN OPEN
SCIENCE CLOUD

EOSC
NORDIC



EuroHPC
Joint Undertaking


ADMIRE


EURO



UNIVERSITY OF ICELAND
SCHOOL OF ENGINEERING AND NATURAL SCIENCES
FACULTY OF INDUSTRIAL ENGINEERING,
MECHANICAL ENGINEERING AND COMPUTER SCIENCE



JÜLICH
Forschungszentrum

JÜLICH
SUPERCOMPUTING
CENTRE

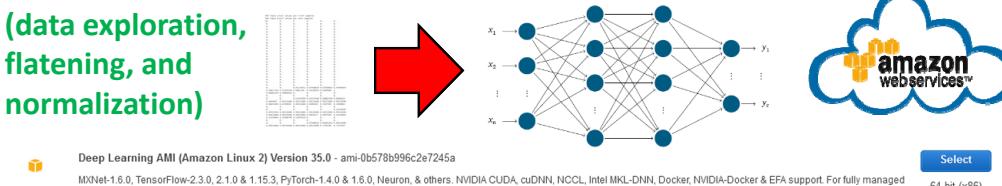
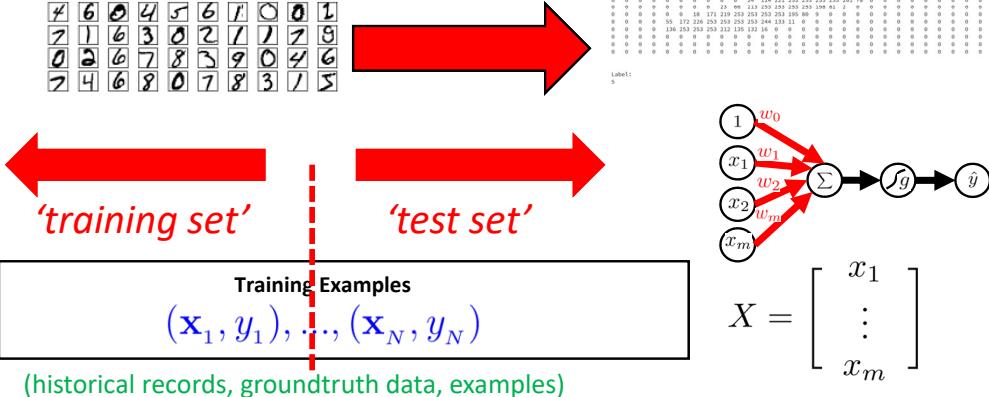

DEEP
Projects

HELMHOLTZAI

ARTIFICIAL INTELLIGENCE
COOPERATION UNIT

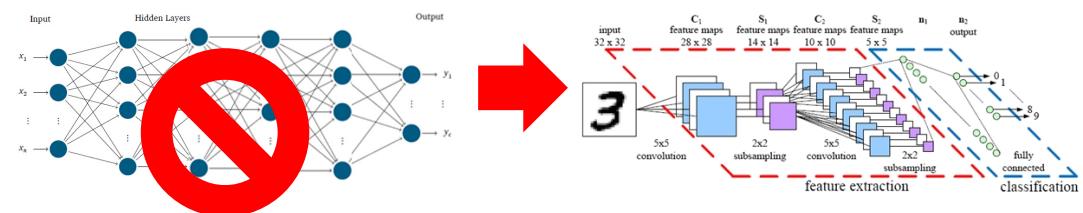
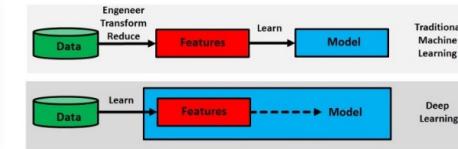
Review of Practical Lecture 7.1 – Using Deep Learning Techniques in Clouds

■ Using Artificial Neural Networks (ANNs)



■ Using Convolutional Neural Networks (CNNs)

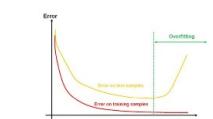
- Deep Learning Technique → Not good with CPUs



- ✓ Multi Output Perceptron: ~91,01% (20 Epochs)
- ✓ ANN 2 Hidden Layers: ~95,14 % (20 Epochs)
- ✓ CNN Deep Learning Model: ~99,36 % (20 Epochs)



(reality: 'noise' in the data)



(fast: careful to not overfit when training long)

[9] www.big-data.tips, 'MNIST Database' [10] www.big-data.tips, 'MNIST Dataset' [3] Amazon Web Services [11] Harley, A.W. et al. [12] A. Rosebrock [13] Google Colaboratory

Outline of the Course

- | | |
|---|--|
| <ol style="list-style-type: none">1. Cloud Computing & Big Data Introduction2. Machine Learning Models in Clouds3. Apache Spark for Cloud Applications4. Virtualization & Data Center Design5. Map-Reduce Computing Paradigm6. Deep Learning driven by Big Data7. Deep Learning Applications in Clouds8. Infrastructure-As-A-Service (IAAS)9. Platform-As-A-Service (PAAS)10. Software-As-A-Service (SAAS) | <ol style="list-style-type: none">11. Big Data Analytics & Cloud Data Mining12. Docker & Container Management13. OpenStack Cloud Operating System14. Online Social Networking & Graph Databases15. Big Data Streaming Tools & Applications16. Epilogue <p>+ additional practical lectures & Webinars for our hands-on assignments in context</p> <ul style="list-style-type: none">▪ Practical Topics▪ Theoretical / Conceptual Topics |
|---|--|

Outline

■ Understanding IAAS Provisioning

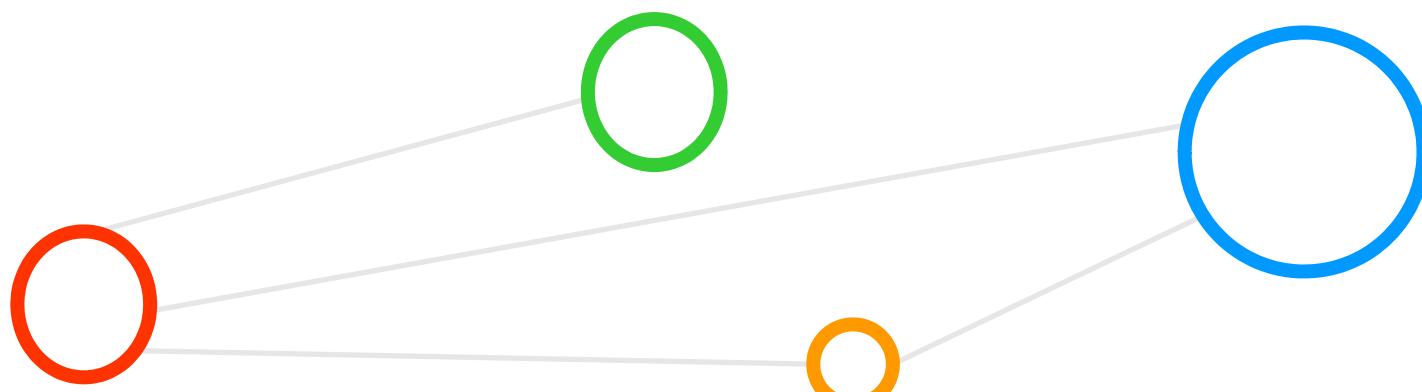
- Different Cloud Service Levels with IAAS, PAAS, and SAAS
- Amazon Web Services (AWS) Examples & AWS Management Console
- Review of Services that Build on top of IAAS such as AWS EC2
- Configuration & Launching EC2 Instances to create Virtual Servers
- Payment Models Reviewed & vCPU/Memory pricing

■ Advanced IAAS Topics & Applications

- Cloud Storage with Amazon S3 / EBS / EFS and Verne Global
- AWS Marketplace & Users for Diverse Applications & Workloads
- Openstack and IAAS Deployment Options with Application Example
- Iceland Advania Data Center Example with HPC IAAS via HPCFlow
- NetApp Application Example as Different Cloud Service Provider

- Promises from previous lecture(s):
- *Lecture 1:* Lecture 8 provides more details about Amazon Web Services and its Infrastructure-as-a-Service (IAAS) models & various cloud services
- *Lecture 4:* Lecture 8 & 9 & 10 offer more insights into concrete cloud systems and their use of virtualization on different levels of cloud services
- *Lecture 4:* Lecture 8 & 9 & 10 will clarify & compare cloud deployment models with the different cloud computing layers IAAS, PAAS, and SAAS
- *Practical Lecture 5.1:* Lecture 8 provides more details about Amazon Web Services and its Infrastructure-as-a-Service (IAAS) models & EC2 Service Elements
- *Practical Lecture 5.1:* Lecture 8 provides more details about Amazon Web Services with different applications using Infrastructure-As-A-Service (IAAS) models

Understanding IAAS Provisioning



Internet Cloud Systems – Examples from Every Day Life – Revisited

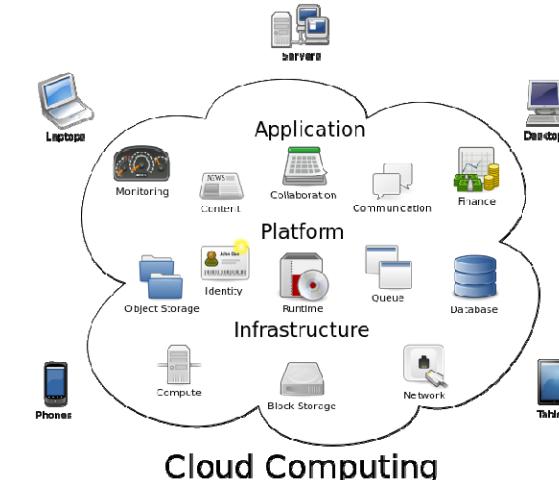
- Selected **Cloud Systems (aka ‘Clouds’)** known today

- Google Cloud → massive computing/storage/applications
- Amazon Web Service → massive computing/storage/services
- Microsoft Azure → massive computing/storage/toolsets
- Facebook → online social networking & advertisement
- SalesForce.com → customer relationship management
- Rackspace → managed cloud provider & hosting
- IBM Bluemix → cloud platform
- Enomaly → elastic computing cloud
- European Open Science Cloud → computing & storage services for research
- Uber Cloud → specialized computing & storage services for engineers



Three Levels of Cloud Service Models: *AAS

- Levels oriented towards different users
 - Full customization to direct usable applications
- Software as a Service (**SAAS**)
 - Provides specific 'ready-to-run applications'
 - Sometimes related to geographical location
- Platform as a Services (**PAAS**)
 - Virtual images ready to deploy your software
 - Includes a 'platform for creation of your services'



- Infrastructure as a Service (**IAAS**)
 - Provides 'bare metal infrastructure' & virtual IT resources (cf. Lecture 4)
 - Use and tune infrastructure as needed (compute, storage, networking, ...)

focus in this lecture

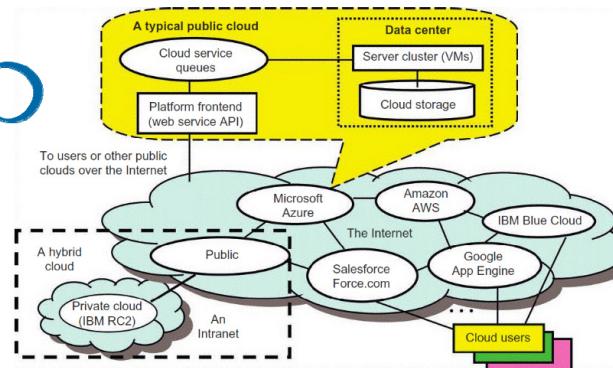
[1] Wikipedia
'Cloud computing'

- Cloud computing infrastructures typically offer services on three different levels: Infrastructure as a Service (IAAS), Platform as a Service (PAAS), and Software as a Service (SAAS) whereby also often some services build on one another (e.g., AWS EMR PAAS builds on top of AWS EC2 IAAS)
- Often Cloud computing service providers (i.e., AWS, MS Azure, Google Cloud) offer service on all these different levels today: IAAS, PAAS, and SAAS

Public/Private/Hybrid Cloud Deployment Models – Different then *AAS Levels



- Public clouds are built over the Internet and can be accessed by any user who has paid for it
- Public clouds are owned by service providers and are accessible through a subscription
- Public clouds deliver a set of business processes or infrastructure resources via price-per-use



[16] *Private Cloud, big-data.tips*

- Hybrid Clouds are a smart combination of public and private clouds supplementing local IT
- Hybrid Clouds typically employing public cloud services for many non-sensitive operations
- Supplements local infrastructure with computing capacity from an external public cloud

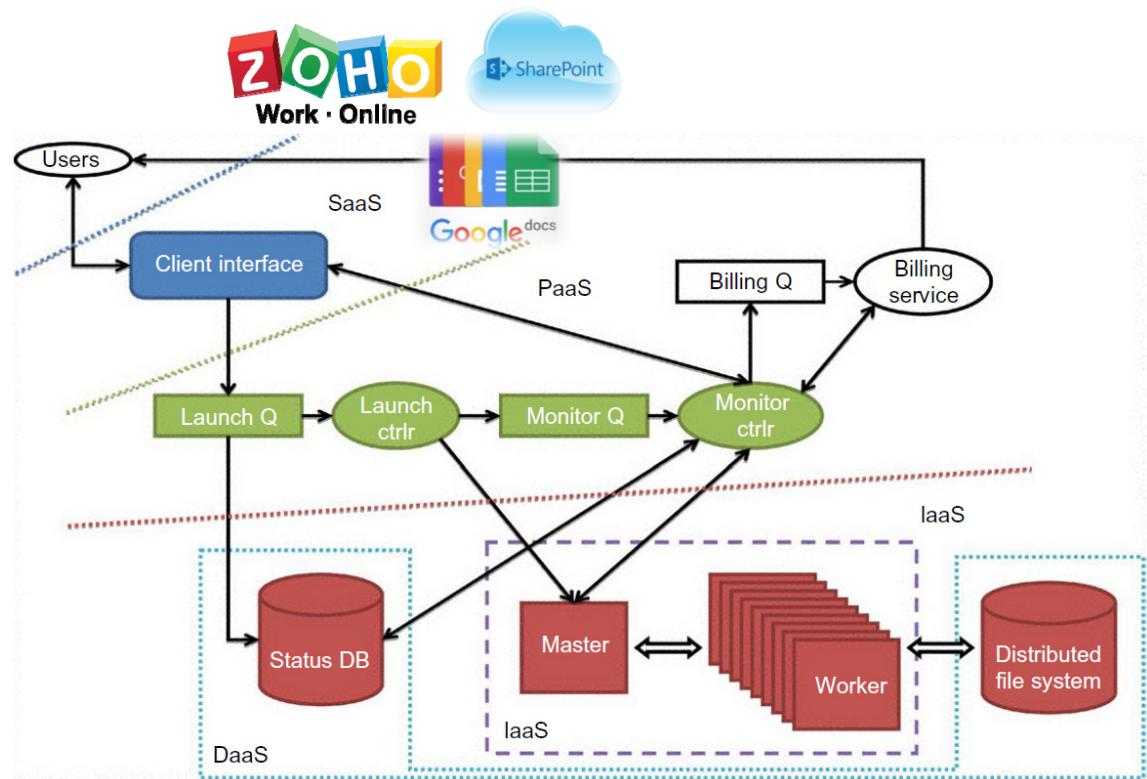
- Private Clouds are built within the domain of an intranet owned by a single organization
- Private Clouds are ‘client owned’ and managed and access is limited to owning clients
- Private Clouds do not sell resources over the Internet but provide them in the organization
- Operations only in one ‘private organization’ - reasons: security w.r.t. the data stored in the cloud; further potentially reasons: Cost of rent is bigger than own long-term hosting

[2] *Distributed & Cloud Computing Book*

Different Cloud Service Models – SAAS

- Software-As-A-Service (SAAS)
 - E.g. services for business processes, consumer applications related to geographical locations, ...
 - E.g. ZOHO, SharePoint, Google Docs
 - E.g. Amazon AWS SageMaker

■ The Conceptual ideas and key usage of the SAAS cloud service model include consume applications, avoid installations, special interfaces & ready-to-run applications
■ SAAS is based on easy accessible software remotely accessed via Web browsers and centrally hosted in data centers



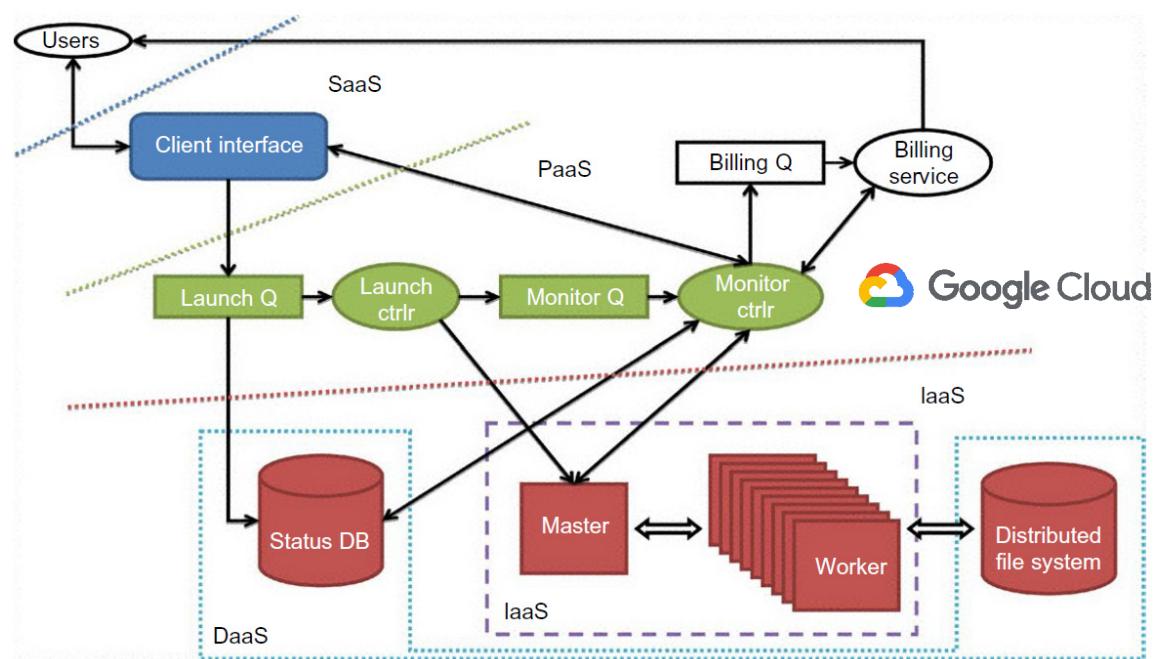
[2] Distributed & Cloud Computing Book

Different Cloud Service Models – PAAS

■ Platform-As-A-Service (PAAS)

- E.g. used to provision billing services, handle compute job queing, launching of images, and monitoring to support application developers
- E.g. Google Cloud

- The Conceptual ideas and key usage of the PAAS cloud service model is building cloud applications with software development kits (SDKs) & application programming Interfaces (APIs) via basic services
- PAAS is based on known application frameworks similiar to ASP, J2EE, JSP and languages like Python, Java, Ruby, etc.



[2] Distributed & Cloud Computing Book

Different Cloud Service Models – IAAS – Focus in this Lecture

■ Infrastructure-As-A-Service (IAAS)

- E.g. used to provision databases (aka Database-As-A-Service), compute instances, distributed file systems, storage **to satisfy 'general' demands**
- E.g. Amazon Web Services (AWS)

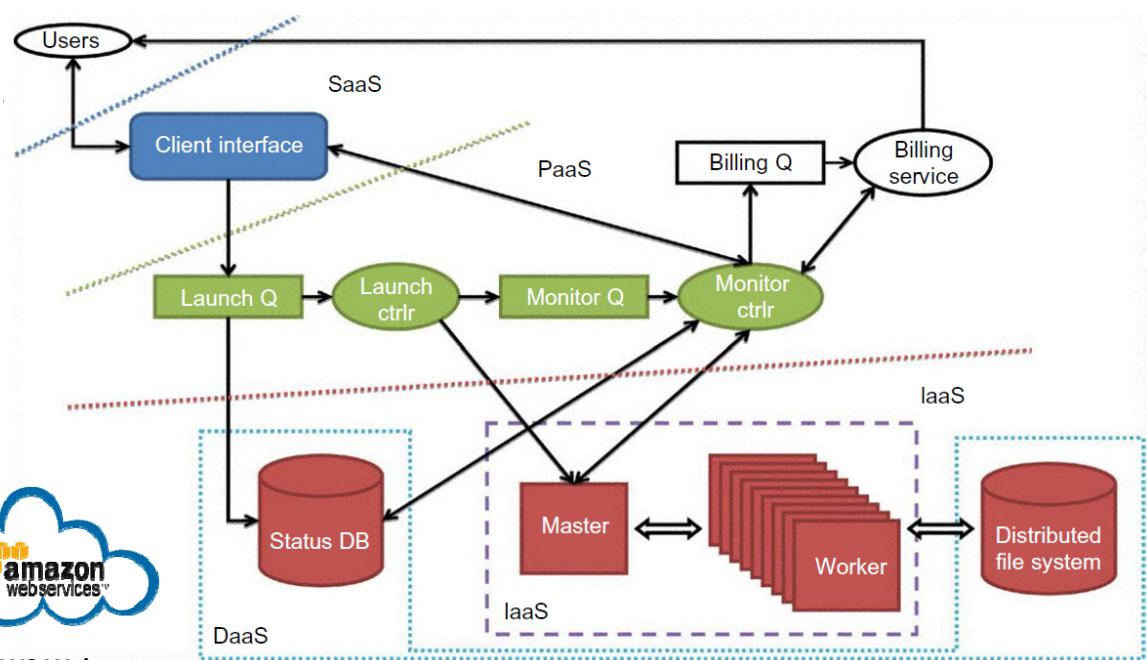
- The Conceptual ideas and key usage of the IAAS cloud service model include hosting, bare metal, and resource provisioning
- IAAS is based on virtual machines (cf. Lecture 4) that are used to flexible share computing and storage resources in a scalable and secure manner

VERNE GLOBAL

[29] Verne Global in Iceland

advania
data centers

[28] Advania HPCFlow IAAS Clusters



[2] Distributed & Cloud Computing Book

IAAS Provider Example – Amazon Web Services (AWS)

- On-Demand delivery of virtualized IT resources

- Compute power, database storage, networking
- Application services, databases, specific software
- Offers '**rapid access**' to flexible and low cost IT resources
- Access **as many resources as you need** (~instantly)
- Using **pay-as-you-go** pricing (pay for what you use)
- Provides tools for **deployment and management of the infrastructure**



[3] AWS Web page

- Selected Business Cases

- Start-ups that need **flexibility**, have **no capital investment**, no experience
- Run applications that share photos to **millions of mobile end users**
- Support **business critical operations** (e.g. understanding customer needs)
- Provide resources for the **educational sector** (e.g., **AWS Educate**) that increases the probability that later students in the job will use it also

- Amazon Web Services (AWS) is a commercial provider using the IAAS cloud service model in order to provide specifically tuned IT resources with roughly 70 different cloud services to choose from
- AWS also provides a wide variety of other services using the PAAS and SAAS cloud service models

AWS Sign-Up Process requires Payment Information – Key Business Idea

■ Account

- Name, Email
- Password
- Address, Phone
- Company or private person

■ Payment(!)

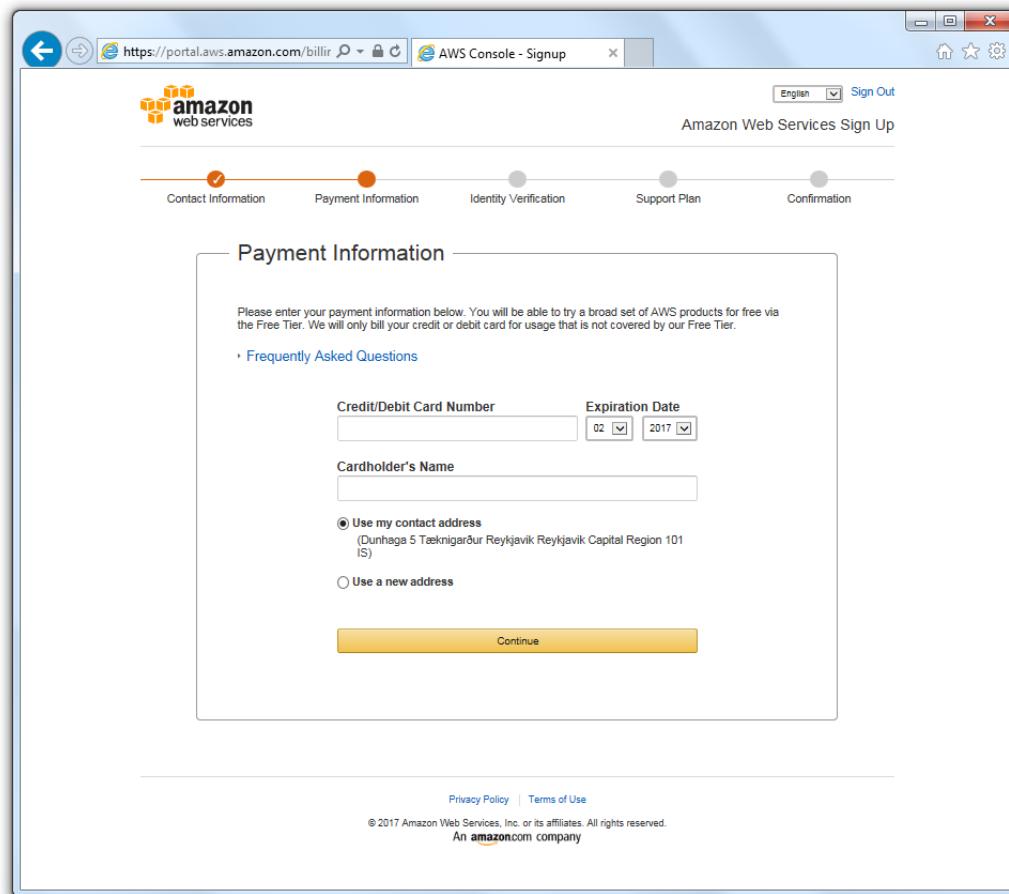
- Credit Card
- Except AWS Educate
(but many limits)

■ Verification

- Auto call & PIN

■ Support plan

- How you plan to use AWS



AWS Management Console – A Plethora of Services – Not only IAAS & Evolutions

Screenshot of the AWS Management Console showing the service catalog. Several services are highlighted with red boxes:

- Compute:** EC2, Lightsail, Lambda, Batch, Elastic Beanstalk, Serverless Application Repository, AWS Outposts, EC2 Image Builder.
- Machine Learning:** Amazon SageMaker, Amazon Augmented AI, Amazon CodeGuru, Amazon Comprehend, Amazon Forecast, Amazon Fraud Detector, Amazon Kendra, Amazon Lex, Amazon Personalize, Amazon Polly, Amazon Rekognition, Amazon Textract, Amazon Transcribe, Amazon Translate, AWS DeepComposer, AWS DeepLens, AWS DeepRacer.
- Storage:** S3, EFS, FSx, S3 Glacier, Storage Gateway, AWS Backup.
- Database:** RDS, DynamoDB, ElastiCache, Neptune, Amazon QLDB, Amazon DocumentDB, Amazon Keyspaces, Amazon TimeStream.
- Analytics:** Athena, Amazon Redshift, EMR, CloudSearch, Elasticsearch Service, Kinesis, QuickSight, Data Pipeline, AWS Data Exchange, AWS Glue, AWS Lake Formation, MSK.
- Networking & Content Delivery:** VPC, CloudFront, Route 53, API Gateway, Direct Connect, AWS App Mesh, AWS Cloud Map, Global Accelerator.
- Developer Tools:** CodeStar, CodeCommit, CodeArtifact, CodeBuild, CodeDeploy, CodePipeline, Cloud9, X-Ray.
- Front-end Web & Mobile:** AWS Amplify, Mobile Hub, AWS AppSync, Device Farm.
- Containers:** ECR, Elastic Container Service, Elastic Kubernetes Service.

(service portfolio 2020)



[3] AWS Web page

The image shows a large red arrow pointing from the AWS Management Console screenshot to a detailed list of AWS services, indicating the evolution of the service portfolio over time. The list is organized into categories:

- Compute:** Amazon EC2, Amazon EC2 Container Registry, Amazon EC2 Container Service, Amazon Lightsail, Amazon VPC, AWS Batch, AWS Elastic Beanstalk, AWS Lambda, Auto Scaling, Elastic Load Balancing.
- Migration**
- Networking & Content Delivery:** Amazon VPC, Amazon CloudFront, Amazon Route 53, AWS Direct Connect, Elastic Load Balancing.
- Storage:** Amazon Simple Storage Service (S3), Amazon Elastic Block Storage (EBS), Amazon Elastic File System (EFS), Amazon Glacier, AWS Storage Gateway, AWS Snowball, AWS Snowball Edge, AWS Snowmobile.
- Analytics:** Amazon Athena, Amazon EMR, Amazon CloudSearch, Amazon Elasticsearch Service, Amazon Kinesis, Amazon Redshift, Amazon QuickSight, AWS Data Pipeline, AWS Glue.
- Developer Tools:** AWS CodeCommit, AWS CodeBuild, AWS CodeDeploy, AWS CodePipeline, AWS X-Ray, AWS Command Line Interface.
- Database:** Amazon Aurora, Amazon RDS, Amazon DynamoDB, Amazon ElastiCache, Amazon Redshift, AWS Database Migration Service.
- Management Tools:** Amazon CloudWatch, Amazon EC2 Systems Manager, AWS CloudFormation, AWS CloudTrail, AWS Config, AWS OpsWorks, AWS Service Catalog, AWS Trusted Advisor, AWS Personal Health Dashboard, AWS Command Line Interface, AWS Management Console, AWS Managed Services.
- Mobile Services:** AWS Mobile Hub, Amazon API Gateway, Amazon Cognito, Amazon Pinpoint, AWS Device Farm, AWS Mobile SDK.
- Security, Identity & Compliance:** AWS Identity and Access Management (IAM), Amazon Inspector, AWS Certificate Manager, AWS CloudHSM, AWS Directory Service, Amazon Cloud Directory, AWS Key Management Service, AWS Organizations, AWS Shield, AWS WAF, AWS Artifact.
- Application Services:** AWS Step Functions, Amazon API Gateway, Amazon Elastic Transcoder, Amazon AppStream.
- Messaging:** Amazon SQS, Amazon Pinpoint, Amazon SES, Amazon SNS.
- Business Productivity:** Amazon WorkDocs, Amazon WorkMail.
- Desktop & App Streaming:** Amazon WorkSpaces, Amazon AppStream 2.0.
- Software:** AWS Marketplace.
- Internet of Things:** AWS IoT Platform, AWS Greengrass, AWS IoT Button.
- Game Development:** Amazon Lumberyard.

(cf. service portfolio 2018)

AWS Management Console – A Plethora of Services – Another Perspective

AWS Management Console

(cf. Lecture 4)

(cf. Lecture 2, 6, 7)

The screenshot shows the AWS Management Console homepage with a grid of service solutions. Solutions are categorized into four columns:

- Launch a virtual machine** (With EC2, 2-3 minutes): Includes a small icon of a computer monitor.
- Build a web app** (With Elastic Beanstalk, 6 minutes): Includes a small icon of a cloud with a person icon.
- Build using virtual servers** (With Lightsail, 1-2 minutes): Includes a small icon of a server tower.
- Register a domain** (With Route 53, 3 minutes): Includes a small icon of a shield with '53' on it.

Connect an IoT device (With AWS IoT, 5 minutes): Includes a small icon of a device with a gear.

Start migrating to AWS (With CloudEndure Migration, 1-2 minutes): Includes a small icon of a cloud with arrows.

Start a development project (With CodeStar, 5 minutes): Includes a small icon of a developer's head with code symbols.

Deploy a serverless microservice (With Lambda, API Gateway, 2 minutes): Includes a small icon of a server with a brain-like interface.

Learn to build: A section for step-by-step guides, labs, and videos.

Websites and Web Apps: Includes a small icon of a website.

Storage: Includes a small icon of a hard drive.

Databases: Includes a small icon of a database.

DevOps: Includes a small icon of a hexagonal network.

Machine Learning: Includes a small icon of two interlocking gears.

Big Data: Includes a small icon of a bar chart with a magnifying glass.

Build with SDKs: Includes a small icon of a gear.

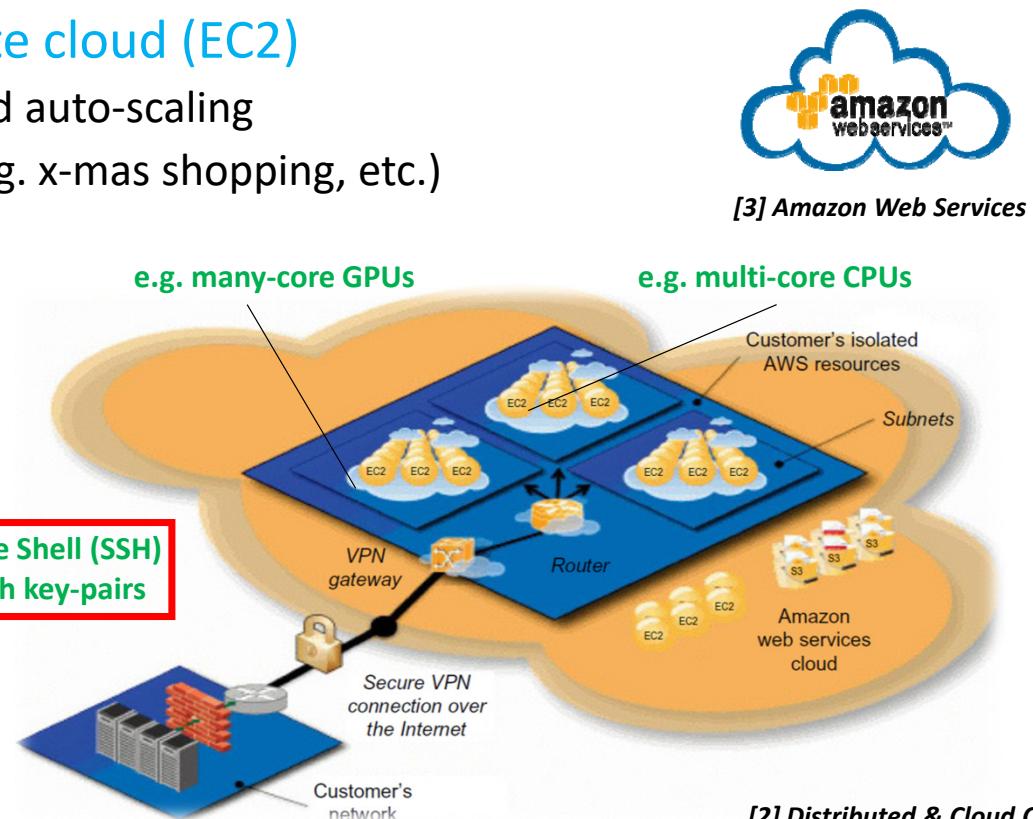
Amazon Web Services – Modern Infrastructure-as-a-Service (IAAS) Example

- Amazon EC2 provides an **elastic compute cloud (EC2)**

- Elastic load balancing services and so-called auto-scaling
- E.g. great **during peak times** in business (e.g. x-mas shopping, etc.)
- Ensures that a **sufficient number of EC2 instances** are provisioned to meet expected performance
- E.g. **New York Times** use it to quickly retrieve pictorial information from millions of articles

- Amazon Web Services (AWS)

- Offers infrastructure used for Amazon shopping also for computing customers
- Ideal situation for Amazon
- Offers **high number of resources & services**
- Central concept is access via key-pairs on IAAS level



[3] Amazon Web Services

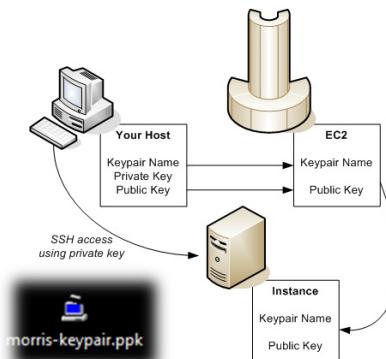
[2] Distributed & Cloud Computing Book

AWS Elastic Compute Cloud (EC2) Virtual Servers & Using Key Pairs – Revisited

■ Secure Shell (SSH)

- Universal technique to securely access remote clusters & HPC machines

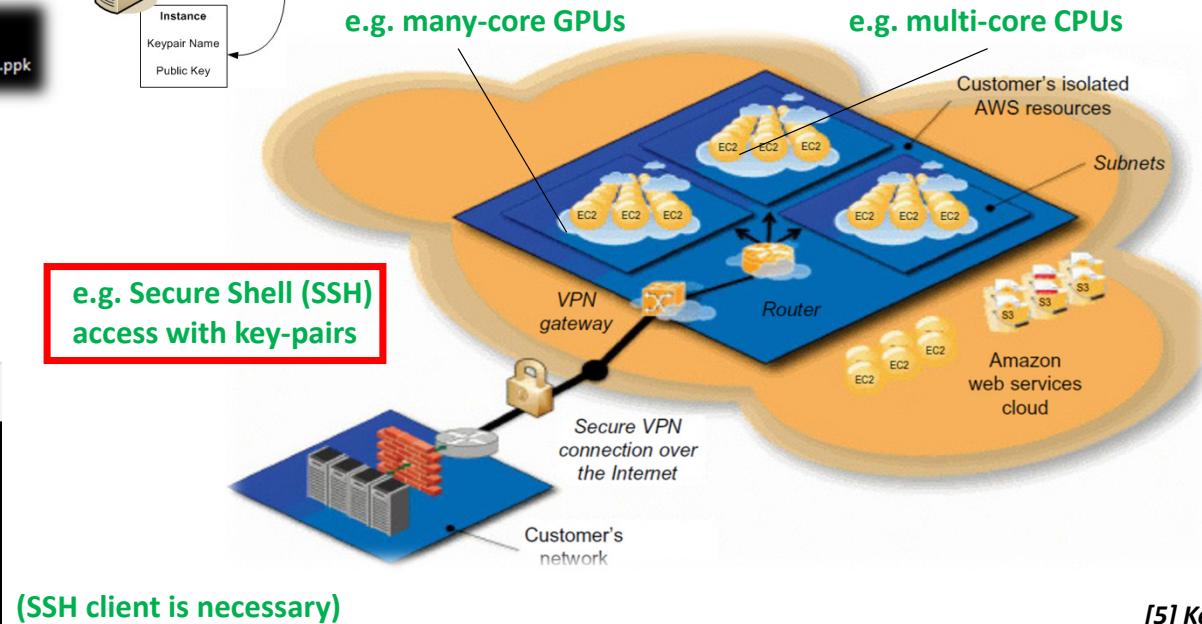
- The Secure Shell (SSH) is a technique to securely access remote AWS computing instances (e.g., AWS EC2) using a named key pair
- An SSH key pair consists of a public key that is known by the Amazon Cloud and a private key that remains only on the laptop of cloud users



- Generated AWS key pairs are created per region (e.g., Virginia) in the AWS Cloud
- Switching regions means new and/or other SSH keys needs to be used as before



[3] Amazon Web Services



[5] Key Concepts from the AWS Cloud

[6] MobaXterm Web page

Understand AWS Cloud Service Portfolio – PAAS Analytics Services on top of IAAS

- Multiple analytics products
 - Extracting insights and actionable information from data requires technologies like analytics & machine learning
- Selected Analytics Services (cf. Lecture 5)
 - Amazon Athena: Serverless Query Service
 - **Amazon ElasticMapReduce (EMR): Hadoop**
 - Amazon ElasticSearch Service: Elasticsearch on AWS
 - Amazon Kinesis: Streaming Data
 - Amazon QuickSight: Business Analytics
 - Amazon Redshift: Data Warehouse

- AWS offers a wide variety of Big Data Analytics Cloud services such as Amazon Athena (interactive analytics), Amazon EMR (big data processing with Apache Hadoop & Spark), Amazon Redshift (data warehousing), Amazon Kinesis (real-time analytics), Amazon Elasticsearch Service (operational analytics), and Amazon Quicksight (dashboards and visualizations)
- Amazon EMR can be seen as a PAAS service that is built on top of IAAS EC2



[3] AWS Web page

AWS Analytics services		
Category	Use cases	AWS service
Analytics	Interactive analytics	Amazon Athena
	Big data processing	Amazon EMR
	Data warehousing	Amazon Redshift
	Real-time analytics	Amazon Kinesis
	Operational analytics	Amazon Elasticsearch Service
	Dashboards and visualizations	Amazon Quicksight

Amazon EMR

Easily run and scale Apache Spark, Hive, Presto, and other big data frameworks

[Get started with Amazon EMR](#)

[Request support for your evaluation](#)

AWS Elastic Compute Cloud (EC2) & Launch Virtual Server Cloud Instances

The screenshot shows the AWS EC2 Dashboard in the US East (N. Virginia) Region. The left sidebar includes links for New EC2 Experience, EC2 Dashboard (highlighted with a red box), Events, Tags, Limits, Instances (with sub-links for Instances, Instance Types, Launch Templates, Spot Requests, Savings Plans, Reserved Instances, Dedicated Hosts, Scheduled Instances, Capacity Reservations), Images (AMIs), and Elastic Block Store (Volumes). The main content area displays 'Resources' with counts for Instances (0), Elastic IPs (0), Dedicated Hosts (0), Snapshots (0), Volumes (0), Key pairs (2), Security groups (5), Placement groups (0), and Load balancers (0). It also shows 'Running instances' (0). A callout box suggests using the AWS Launch Wizard for Microsoft SQL Server. Below this is a 'Launch instance' section with a 'Launch instance' button (highlighted with a red box) and a note stating instances will launch in the US East (N. Virginia) Region. To the right, there's a 'Service health' section showing the service is operating normally in the US East (N. Virginia) region, and a 'Zone status' section. The top right corner shows account attributes like supported platforms (VPC, Default VPC vpc-119c5f6c), settings, EBS encryption, zones, default credit specification, and console experiments. There's also an 'Explore AWS' section with links for Graviton2, Apache Spark on EMR, and more.

AWS Elastic Compute Cloud (EC2) & Amazon Machine Images (AMIs) Example

The screenshot shows the AWS EC2 instance creation process at Step 1: Choose an Amazon Machine Image (AMI). The interface includes a navigation bar with tabs: 1. Choose AMI, 2. Choose Instance Type, 3. Configure Instance, 4. Add Storage, 5. Add Tags, 6. Configure Security Group, and 7. Review. The main area displays a list of AMIs:

- Red Hat Enterprise Linux 8 (HVM), SSD Volume Type** - ami-098f16afa9edf40be (64-bit x86) / ami-029ba835ddd43c34f (64-bit Arm)
Root device type: ebs Virtualization type: hvm ENA Enabled: Yes
Free tier eligible
Select button, 64-bit (x86) radio button selected, 64-bit (Arm) radio button
- SUSE Linux Enterprise Server 15 SP2 (HVM), SSD Volume Type** - ami-0a782e324655d1cc0 (64-bit x86) / ami-06ec4efaf39ca724d4 (64-bit Arm)
SUSE Linux Enterprise Server 15 Service Pack 2 (HVM), EBS General Purpose (SSD) Volume Type. Public Cloud, Advanced Systems Management, Web and Scripting, and Legacy modules enabled.
Root device type: ebs Virtualization type: hvm ENA Enabled: Yes
Free tier eligible
Select button, 64-bit (x86) radio button selected, 64-bit (Arm) radio button
- Ubuntu Server 20.04 LTS (HVM), SSD Volume Type** - ami-0dba2cb6798deb6d8 (64-bit x86) / ami-0ea142bd244023692 (64-bit Arm)
Ubuntu Server 20.04 LTS (HVM), EBS General Purpose (SSD) Volume Type. Support available from Canonical (<http://www.ubuntu.com/cloud/services>).
Root device type: ebs Virtualization type: hvm ENA Enabled: Yes
Free tier eligible
Select button, 64-bit (x86) radio button selected, 64-bit (Arm) radio button
- Ubuntu Server 18.04 LTS (HVM), SSD Volume Type** - ami-0817d428a6fb68645 (64-bit x86) / ami-0f2b11fdc1647918 (64-bit Arm)
Ubuntu Server 18.04 LTS (HVM), EBS General Purpose (SSD) Volume Type. Support available from Canonical (<http://www.ubuntu.com/cloud/services>).
Root device type: ebs Virtualization type: hvm ENA Enabled: Yes
Free tier eligible
Select button, 64-bit (x86) radio button selected, 64-bit (Arm) radio button
- Microsoft Windows Server 2019 Base** - ami-0412e100c0177fb4b
Microsoft Windows 2019 Datacenter edition. [English]
Root device type: ebs Virtualization type: hvm ENA Enabled: Yes
Free tier eligible
Select button, 64-bit (x86)
- Deep Learning AMI (Ubuntu 18.04) Version 35.0** - ami-01aad86525617098d
MXNet-1.6.0, TensorFlow-2.3.0, 2.1.0 & 1.15.3, PyTorch-1.4.0 & 1.6.0, Neuron, & others. NVIDIA CUDA, cuDNN, NCCL, Intel MKL-DNN, Docker, NVIDIA-Docker & EFA support. For fully managed experience, check: <https://aws.amazon.com/sagemaker>
Root device type: ebs Virtualization type: hvm ENA Enabled: Yes
Select button, 64-bit (x86)
- Deep Learning AMI (Ubuntu 16.04) Version 35.0** - ami-0be2e1ead90f81d34
MXNet-1.6.0, TensorFlow-2.3.0, 2.1.0 & 1.15.3, PyTorch-1.4.0 & 1.6.0, Neuron, & others. NVIDIA CUDA, cuDNN, NCCL, Intel MKL-DNN, Docker, NVIDIA-Docker & EFA. For fully managed experience, check: <https://aws.amazon.com/sagemaker>
Root device type: ebs Virtualization type: hvm ENA Enabled: Yes
Select button, 64-bit (x86)
- Deep Learning AMI (Amazon Linux 2) Version 35.0** - ami-0b578b996c2e7245a
MXNet-1.6.0, TensorFlow-2.3.0, 2.1.0 & 1.15.3, PyTorch-1.4.0 & 1.6.0, Neuron, & others. NVIDIA CUDA, cuDNN, NCCL, Intel MKL-DNN, Docker, NVIDIA-Docker & EFA support. For fully managed experience, check: <https://aws.amazon.com/sagemaker>
Root device type: ebs Virtualization type: hvm ENA Enabled: Yes
Select button, 64-bit (x86)

- **AWS Amazon Machine Images (AMIs)** are templates that contain the software configuration (e.g., operating system, libraries, application server, and applications) required to launch a EC2 virtual server instance for a specific purpose
- **AWS EC2 AMI** offers solutions that enormously simplify the deployment of required machine learning and deep learning stacks that can be complicated to make work together due to the many different software versions and fast moving technologies (e.g., different NVIDIA GPUs)
- **AWS EC2 AMI** The AMIs are independent from the underlying hardware infrastructure (i.e. concrete CPUs) and can be easily migrated (cf. Lecture 4) to other hardware – be aware of different hardware costs here
- **Amazon offers pre-configured AMIs for deep learning** consisting of preinstalled deep learning packages such as MXNet, TensorFlow, PyTorch, Keras, etc.
- **Pre-configured AMIs for deep learning feature preinstalled GPU NVIDIA CUDA, cuDNN, and NCCL libraries** that usually requires a lot of efforts in installation and version checks with deep learning packages

Different Payment Models – Reviewed

■ Pay-as-you-go / pay-per-use model

- Cost is often significantly reduced vs. cost of ownership (not always)
- All hardware and software resources are leased from cloud provider
- Without capital investment on the part of the users
- Example: Amazon Web Services (AWS) offers a pay-as-you-go model for pricing ~70 cloud services



[3] AWS Web page

Classical Computing

(Repeat the following cycle every 18 months)

Buy and own

Hardware, system software, applications to meet peak needs

Install, configure, test, verify, evaluate, manage

Use

Pay \$\$\$\$\$ (High cost)

Cloud Computing

(Pay as you go per each service provided)

Subscribe

Use (Save about 80-95% of the total cost)

(Finally)

\$ - Pay for what you use

based on the QoS

- AWS offers also free tier and AWS Educate access with limited resources & certain specific quota on services

- AWS uses four pricing models: 'on-demand' to pay for compute capacity by the hour, EC2 spot instances to bid on spare Amazon EC2 computing 90% off, reserved instances, dedicated hosts

[2] Distributed & Cloud Computing Book

Choose EC2 Instance for AMI & Review Costs Using Free Tier Eligible CPUs

Screenshot of the AWS EC2 instance selection interface:

Step 2: Choose an Instance Type
Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instances are virtual servers that can run applications. They have varying combinations of CPU, memory, storage, and networking capacity, and give you the flexibility to choose the appropriate mix of resources for your applications. [Learn more](#) about instance types and how they can meet your computing needs.

Filter by: All instance families ▾ Current generation ▾ Show/Hide Columns

Currently selected: t2.micro (- ECUs, 1 vCPUs, 2.5 GHz, ~ 1 GiB memory, EBS only)

Family	Type	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance	IPv6 Support
t2	t2.nano	1	0.5	EBS only	-	Low to Moderate	Yes
t2	t2.micro <small>Free tier eligible</small>	1	1	EBS only	-	Low to Moderate	Yes
t2	t2.small	1	2	EBS only	-	Low to Moderate	Yes
t2	t2.medium	2	4	EBS only	-	Low to Moderate	Yes
t2	t2.large	2	8	EBS only	-	Low to Moderate	Yes
t2	t2.xlarge	4	16	EBS only	-	Moderate	Yes
t2	t2.2xlarge	8	32	EBS only	-	Moderate	Yes
t3	t3.nano	2	0.5	EBS only	Yes	Up to 5 Gigabit	Yes
t3	t3.micro	2	1	EBS only	Yes	Up to 5 Gigabit	Yes
t3	t3.small	2	2	EBS only	Yes	Up to 5 Gigabit	Yes
t3	t3.medium	2	4	EBS only	Yes	Up to 5 Gigabit	Yes
t3	t3.large	2	8	EBS only	Yes	Up to 5 Gigabit	Yes
t3	t3.xlarge	4	16	EBS only	Yes	Up to 5 Gigabit	Yes
t3	t3.2xlarge	8	32	EBS only	Yes	Up to 5 Gigabit	Yes
t3a	t3a.nano	2	0.5	EBS only	Yes	Up to 5 Gigabit	Yes
t3a	t3a.micro	2	1	EBS only	Yes	Up to 5 Gigabit	Yes
t3a	t3a.small	2	2	EBS only	Yes	Up to 5 Gigabit	Yes
t3a	t3a.medium	2	4	EBS only	Yes	Up to 5 Gigabit	Yes

Cancel Previous **Review and Launch** Next: Configure Instance Details

Amazon EC2 On-Demand Pricing

PAGE CONTENT

On-Demand Pricing

On-Demand Instances let you pay for compute capacity by the hour or second (minimum of 60 seconds) with no long-term commitments. This frees you from the costs and complexities of planning, purchasing, and maintaining hardware and transforms what are commonly large fixed costs into much smaller variable costs.

The pricing below includes the cost to run private and public AMIs on the specified operating system ("Windows Usage" prices apply to Windows Server 2003 R2, 2008, 2008 R2, 2012, 2012 R2, 2016, and 2019). Amazon also provides you with additional instances for Amazon EC2 running Microsoft Windows with SQL Server, Amazon EC2 running SUSE Linux Enterprise Server, Amazon EC2 running Red Hat Enterprise Linux and Amazon EC2 running IBM that are priced differently.

Linux	RHEL	SLES	Windows	Windows with SQL Standard	Windows with SQL Web
Windows with SQL Enterprise	Linux with SQL Standard	Linux with SQL Web	Linux with SQL Enterprise		
t2.nano	1	Variable	0.5 GiB	EBS Only	\$0.0058 per Hour
t2.micro	1	Variable	1 GiB	EBS Only	\$0.0116 per Hour
t2.small	1	Variable	2 GiB	EBS Only	\$0.023 per Hour
t2.medium	2	Variable	4 GiB	EBS Only	\$0.0464 per Hour
t2.large	2	Variable	8 GiB	EBS Only	\$0.0928 per Hour
t2.xlarge	4	Variable	16 GiB	EBS Only	\$0.1856 per Hour
t2.2xlarge	8	Variable	32 GiB	EBS Only	\$0.3712 per Hour
m6g.medium	1	N/A	4 GiB	EBS Only	\$0.0385 per Hour
m6g.large	2	N/A	8 GiB	EBS Only	\$0.077 per Hour
m6g.xlarge	4	N/A	16 GiB	EBS Only	\$0.154 per Hour
m6g.2xlarge	8	N/A	32 GiB	EBS Only	\$0.308 per Hour
m6g.4xlarge	16	N/A	64 GiB	EBS Only	\$0.616 per Hour
m6g.8xlarge	32	N/A	128 GiB	EBS Only	\$1.232 per Hour
m6g.12xlarge	48	N/A	192 GiB	EBS Only	\$1.848 per Hour
m6g.16xlarge	64	N/A	256 GiB	EBS Only	\$2.464 per Hour
m6g.meta	64	N/A	256 GiB	EBS Only	\$2.464 per Hour

[7] AWS EC2 Pricing

AWS EC2 Compute Payment Models – Elastic Capability – Cost Sensitivity

■ Approach: General Purpose T2

- ‘Pay for what you use’ → t2.SIZES
- Burstable performance Instances
- Provide a baseline level of CPU performance with the ability to burst above the baseline
- T2 instances are for workloads that don’t use the full CPU often, but occasionally need to burst
- Accrue CPU Credits when they are idle
- Use CPU credits when they are active
- A CPU Credit provides the performance of a full CPU core for one minute

Model	vCPU	CPU Credits / hour	Mem (GiB)
t2.nano	1	3	0.5
t2.micro	1	6	1
t2.small	1	12	2
t2.medium	2	24	4
t2.large	2	36	8
t2.xlarge	4	54	16
t2.2xlarge	8	81	32

■ Example: t2.small instance

- Receives credits continuously at a rate of 12 CPU Credits per hour (baseline performance equivalent to 20% of a CPU core)
- Usage below 20% / core stored in credit balance (used when above 20%)

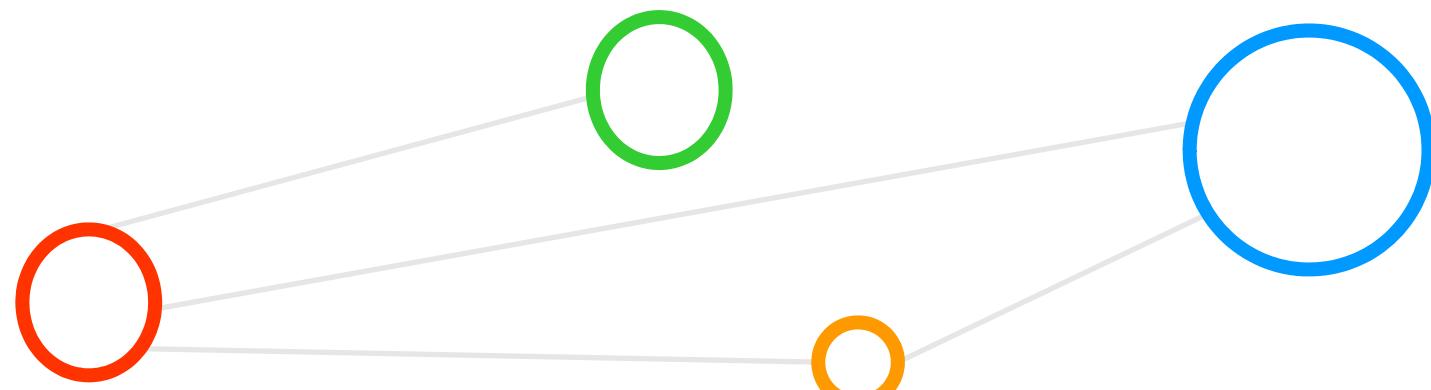
[7] AWS EC2 Pricing

[Video] Infrastructure-As-A-Service Goals & Application Areas



[8] YouTube video, What is IaaS?

Advanced IAAS Topics & Applications



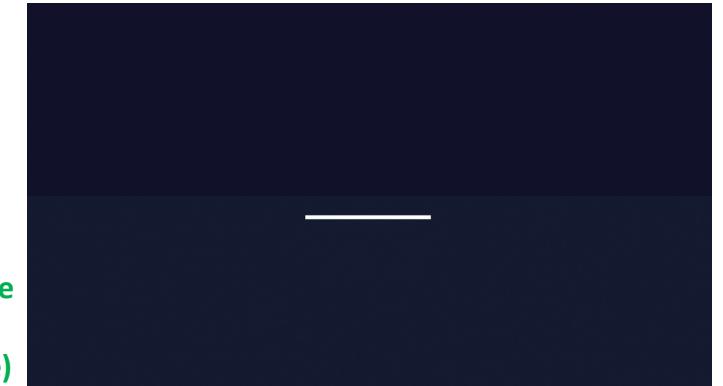
IAAS for different Hardware – Examples from AWS, Verne Global & Advania

The screenshot shows the AWS Management Console with the 'All services' search bar at the top. On the left, there's a sidebar with 'Favorites' and 'Recently visited' sections. The main area lists various AWS services under categories like Compute, Storage, Database, Networking & Content Delivery, Developer Tools, and more. A red box highlights the 'Storage' section under the 'Compute' category, which includes services like S3, EFS, Fsx, S3 Glacier, Storage Gateway, and AWS Backup.



[3] Amazon Web Services

(NVIDIA DGX-Ready as enterprise HPC hardware optimized for machine & deep learning at scale)



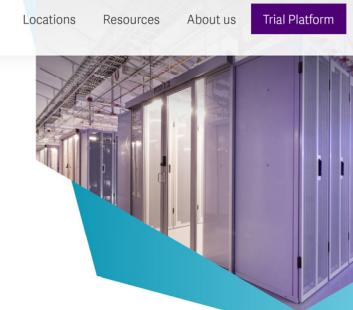
Taking IaaS for AI and HPC to a new level in Iceland

Verne Global takes the complexity and capital costs out of the HPC and AI equation.



[29] Verne Global in Iceland

[31] Verne Global NVIDIA DGX-Ready



Data Center Solutions AI & HPC Solutions Locations Resources About us Trial Platform

HPCFLOW: IaaS Clusters

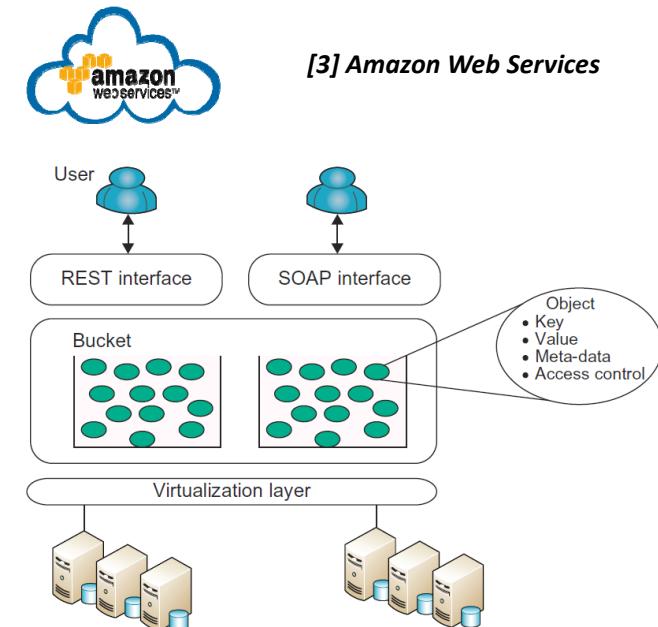
HPCFLOW IaaS clusters enable the missing piece of Hybrid HPC, available as dedicated clusters or pure on demand where you only pay for what you use. HPCFLOW allows you the ultimate AI & HPC cluster elasticity and flexibility.

Contact us →

[28] Advania HPCflow IAAS Clusters

Amazon Web Services – Broadly used Storage-as-a-Service (S3) Example

- S3 is ‘storage as a service’ with a **Web messaging interface**
 - Using API with **Representational State Transfer (REST)**
 - Using API with **Simple Object Access Protocol (SOAP)**
- Remote **object storage**
 - Data considered **objects** to be named by end users
 - Objects alongside metadata are stored in **bucket containers**
 - Buckets enable the organization with **namespace for user identification** & accounting
 - (Automatically) scalable



IAAS Application Example with AWS Amazon S3 Instances – Host a static Website

- Launch a **S3-based service** via console
 - E.g. **Host a static Website** lets one easily create S3 storage for a Website
- Combined with other AWS Services
 - **Amazon Cloud Front**: assigns a URL that is used to access a Website ('buy domain') and managed **content delivery**
 - **Amazon Route 53**: Scalable domain name system
 - **Amazon S3**: data, code, etc.

- **Amazon S3 provides replicated storage at multiple data centers (cf. Lecture 4) in the AWS cloud**



[3] *Amazon Web Services*

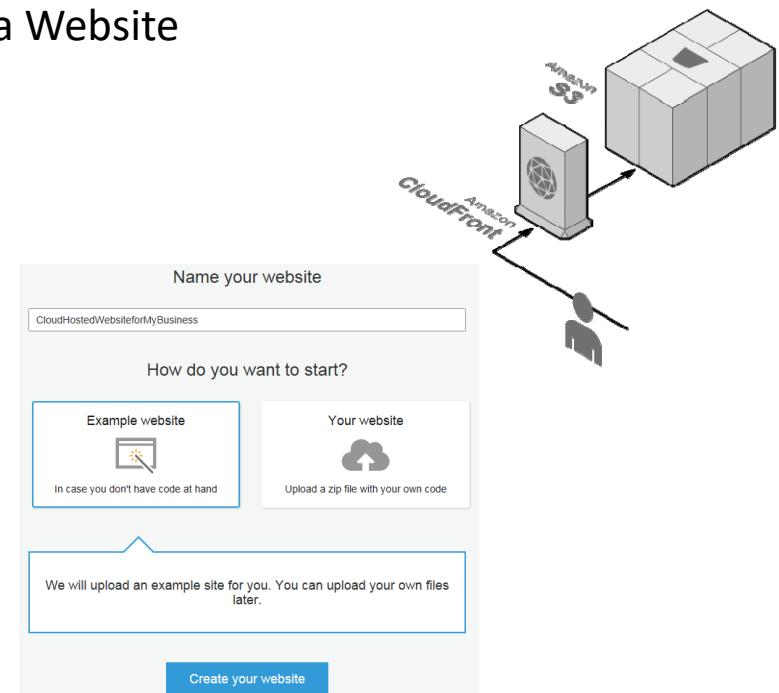
Source code powered by Amazon S3

Your content: 10 files 30 KB

Browse or drag & drop a zip file to upload a new version

[Manage files in Amazon S3](#)

[Create your website](#)



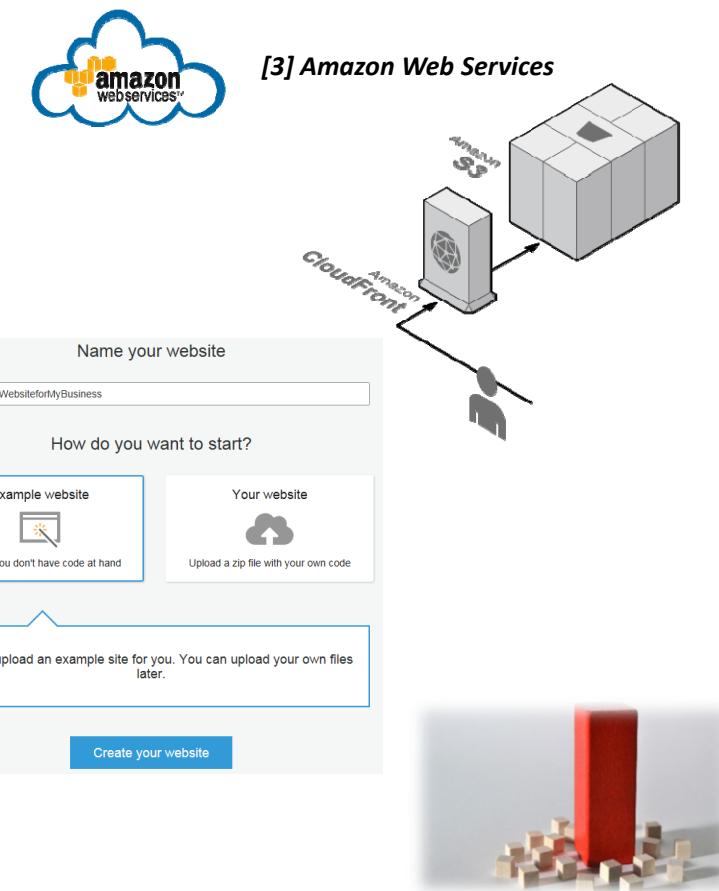
IAAS Application Example with AWS Amazon S3 Instances – Manage Cloud Files

- Using the Web console to [upload/download content](#)
 - Static Website example is mainly based on 'files with content'

The screenshot shows the AWS S3 Management Console interface. At the top, there are tabs for 'Services' and 'Resource Groups'. Below the tabs, there are buttons for 'Upload', 'Create Folder', and 'Actions'. A search bar with the placeholder 'Search by prefix' is followed by three buttons: 'None', 'Properties', and 'Transfers'. The main area displays a table titled 'All Buckets / aws-website-cloudhostedwebsiteformybusiness-j9n2l'. The table has columns for 'Name', 'Storage Class', 'Size', and 'Last Modified'. The data in the table is as follows:

Name	Storage Class	Size	Last Modified
LICENSE.md	Standard	8.9 KB	Sun Feb 05 08:47:08 GMT+000 2017
NOTICE.md	Standard	571 bytes	Sun Feb 05 08:47:08 GMT+000 2017
about.html	Standard	2.4 KB	Sun Feb 05 08:47:08 GMT+000 2017
contact.html	Standard	3.8 KB	Sun Feb 05 08:47:08 GMT+000 2017
css	--	--	--
img	--	--	--
index.html	Standard	3.8 KB	Sun Feb 05 08:47:08 GMT+000 2017
js	--	--	--
services.html	Standard	4.4 KB	Sun Feb 05 08:47:09 GMT+000 2017

At the bottom of the page, there are links for 'Feedback', 'English', and 'aws.amazon.com/GetResc© 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved.' along with 'Privacy Policy' and 'Terms of Use'.



AWS Amazon S3 & Glacier Instances – Cost Models Include Data Transfer & APIs

- S3 ([standard](#)) storage prices and options for long-term archive ([S3 Glacier](#))



Region: US East (Ohio) ▾

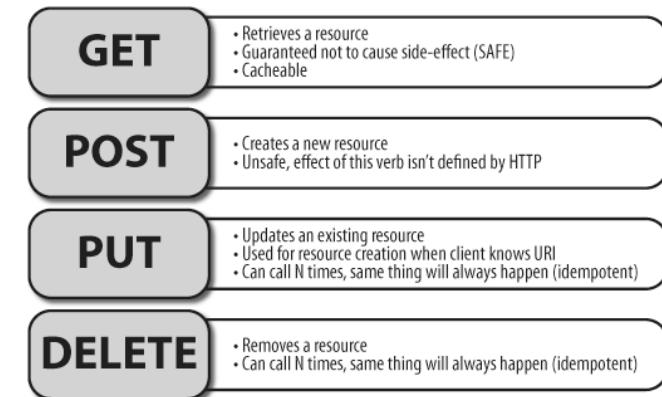
		Storage pricing
S3 Standard - General purpose storage for any type of data, typically used for frequently accessed data		
First 50 TB / Month		\$0.023 per GB
Next 450 TB / Month		\$0.022 per GB
Over 500 TB / Month		\$0.021 per GB
S3 Glacier ** - For long-term backups and archives with retrieval option from 1 minute to 12 hours		
All Storage / Month		\$0.004 per GB
S3 Glacier Deep Archive ** - For long-term data archiving that is accessed once or twice in a year and can be restored within 12 hours		
All Storage / Month		\$0.00099 per GB

- Watch out for transfer in/out prices (!)

	PUT, COPY, POST, LIST requests (per 1,000 requests)	Lifecycle Transition requests (per 1,000 requests)	Data Retrieval requests (per 1,000 requests)	Data retrievals (per GB)
S3 Standard	\$0.005	\$0.0004	n/a	n/a
S3 Intelligent - Tiering	\$0.005	\$0.0004	\$0.01	n/a
S3 Standard - Infrequent Access*	\$0.01	\$0.001	\$0.01	n/a
S3 One Zone - Infrequent Access*	\$0.01	\$0.001	\$0.01	n/a
S3 Glacier **	\$0.05	\$0.0004	\$0.05	See below



- Amazon S3 pricing includes cost models per GB data and for data transfer & Web message exchange requests as well
- Amazon S3 (Deep) Glacier are low-cost Cloud archive storages, but take long to retrieve/restore data
- GET, POST, PUT, DELETE are all Web service message exchanges used by REST/SOAP APIs



[24] [Amazon S3 Pricing Models](#)

[25] [RESTful.NET](#)

IAAS Example – AWS Amazon Elastic Block Storage (EBS)

■ Block storage

- Different from Object storage, but also often used in Cloud services

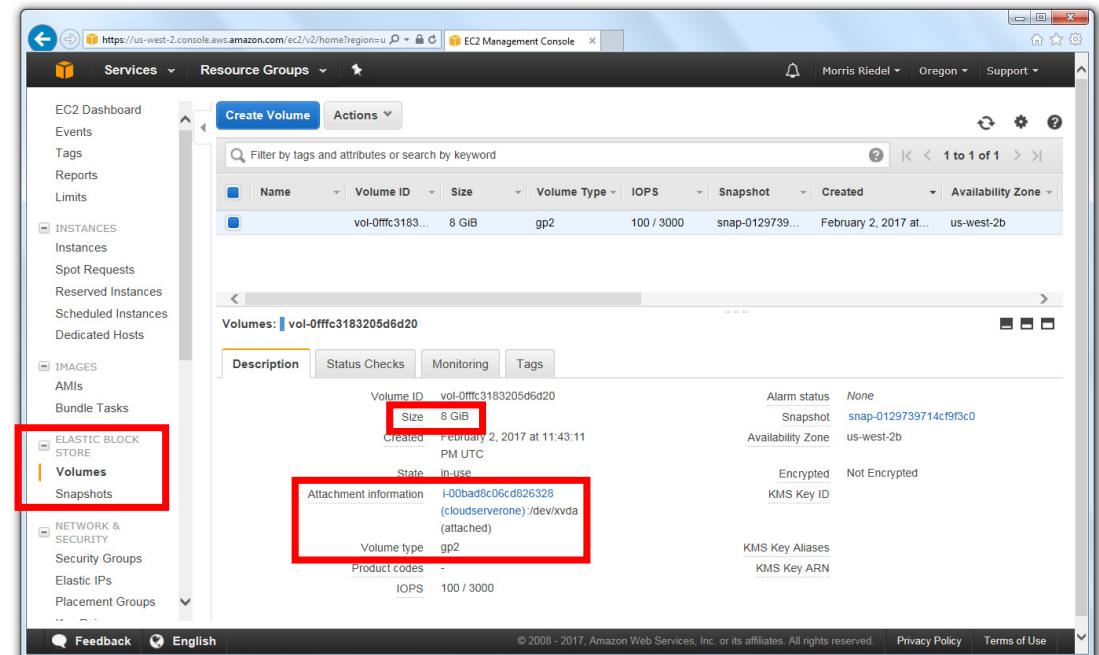
■ EBS Service

- EBS volumes are automatically replicated
- Scales usage up/down (~min)

Amazon EBS Volumes

With Amazon EBS, you pay only for what you use. The pricing for Amazon EBS volumes is listed below

General Purpose SSD (gp2) Volumes	\$0.10 per GB-month of provisioned storage
Provisioned IOPS SSD (io2) Volumes	\$0.125 per GB-month of provisioned storage AND \$0.065 per provisioned IOPS-month
Provisioned IOPS SSD (io1) Volumes	\$0.125 per GB-month of provisioned storage AND \$0.065 per provisioned IOPS-month
Throughput Optimized HDD (st1) Volumes	\$0.045 per GB-month of provisioned storage
Cold HDD (sc1) Volumes	\$0.025 per GB-month of provisioned storage



- Block storage manages data as blocks (compared to object storage with data as objects)
- Block storage typically split files into raw blocks of data whereby each block is controlled like a hard drive
- Amazon Elastic Block Storage (EBS) provides persistent block storage volume used in conjunction with Amazon EC2

[26] Amazon EBS Pricing Models



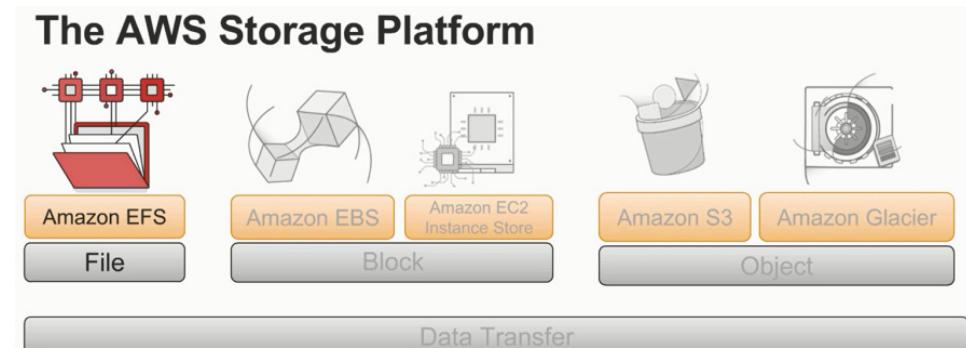
IAAS Example – AWS Amazon Elastic File System (EFS) & File Access Frequency

- EFS for working with **traditional files**
 - Add & remove files, applications have the elastic storage as needed, **only billed whats used**
 - File system object (i.e. directory, file, and link) is **redundantly stored to prevent data loss**

- Typical usage
 - Amazon EFS file system provides a **standard file system interface**
 - **Typical file system access** semantics for existing applications and tools
 - Multiple Amazon EC2 instances **access an Amazon EFS at the same time**
 - Using standard **operating system mount** commands (e.g. UNIX)
(e.g. '**mount**' / connect Amazon EFS file systems via the NFSv4.1 protocol)

Amazon EFS offers two storage classes: the Standard storage class, and the **Infrequent Access storage class (EFS IA)**. EFS IA provides price/performance that's cost-optimized for files not accessed every day. To load data into EFS IA, simply enable Lifecycle Management for your file system and reduce your storage costs by up to 92%.

Industry research and customer analysis shows that on average, 20% of files are actively used and 80% are infrequently accessed. Using that estimate, you can store your files on Amazon EFS at an effective price of \$0.08/GB-month*. See below for detailed pricing examples.



- **Amazon Elastic File System (EFS)** is a simple, scalable and reliable elastic file storage system and can be used with AWS EC2 instances
- **AWS EFS storage capacity is elastic, growing and shrinking automatically depending on the usage**

[27] **Amazon EFS Pricing Models**



AWS Marketplace & Selected Amazon Elastic Map Reduce Applications

■ AWS Marketplace

- E.g. collection of community and Amazon created [pre-installed images](#)
- Software infrastructure, developer tools, business & desktop software
- [User success stories](#) and details of how AWS was adopted in solutions

■ ‘Startup company’ Airbnb (travel)

- [Scales infrastructure automatically](#) using AWS
- Uses [200 Amazon EC2 instances](#) for its application
- Uses [elastic load balancing](#) with Amazon EC2 instances
- Analyzes [50 GB of data daily](#) via [Amazon Elastic MapReduce \(Amazon EMR\)](#)



■ ‘Startup company’ Spotify (music)

- Instant access to [over 16 million licensed songs](#)
- Stores its huge volume of content in [Amazon S3](#)



AWS Marketplace is a digital catalog with thousands of software listings from independent software vendors that make it easy to find, test, buy, and deploy software that runs on AWS.

[View all products](#)

Find AWS Marketplace products that meet your needs.

Categories	Vendors	Pricing Plans	Delivery Methods
All categories	All vendors	All pricing plans	All delivery methods

Total results: 9425

[Clear selection](#) [View results](#)

Popular Categories

Operating Systems	Security	Networking	Storage

Data Analytics	Dev Ops	Machine Learning	Data Products

[View all categories](#)

[23] AWS Marketplace

- [AWS Elastic Map Reduce \(EMR\)](#) enables Hadoop & Spark on AWS cloud & is used in Airbnb applications (50 GB/day) & Spotify applications to access 16 million songs

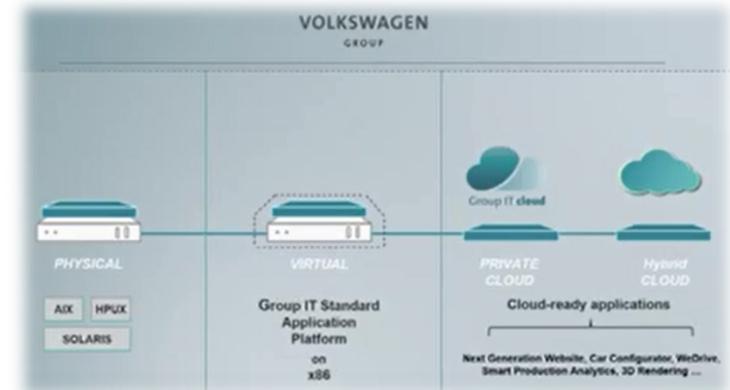
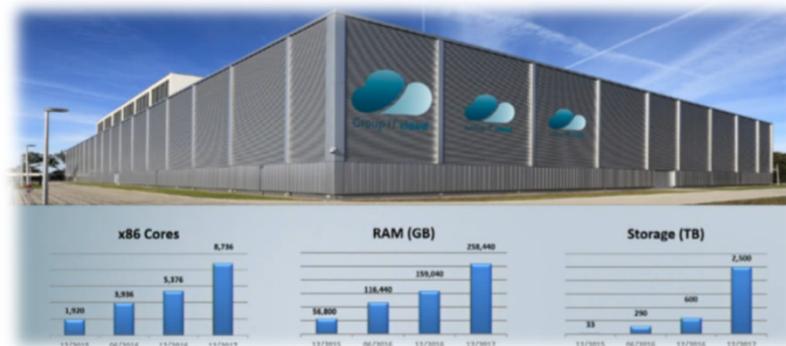
Application Example: OpenStack and IAAS Solutions using Private/Public Clouds

- OpenStack Cloud Operating System used by Volkswagen Group
 - Provides an architecture to **build own IAAS clouds**
 - One can create a massive product **portfolio** (e.g. AWS, but on small level)
 - **Underlying virtual IT resources** like storage, compute, networking needed
 - Openstack software & hardware provide **IAAS basis for many applications**
 - Compatible with public clouds via APIs:
Amazon EC2 and Google Compute Engine



▪ The Volkswagen Group is using OpenStack for their IAAS-based private and public cloud solutions

[21] VW & Openstack



➤ Lecture 13 will provide more details on using the OpenStack Cloud Operating System for creating Cloud Environments for Data Analytics

NetApp IAAS Application Example – A Different Cloud Service Provider

The screenshot shows the NetApp website's IaaS page. At the top, there's a navigation bar with the NetApp logo, a search bar, and links for Solutions & Products, Support & Training, How to Buy, and Community. Below the navigation, a purple background features a grid of white and purple 3D cubes. The page title is "Infrastructure as a Service (IaaS)". Below the title, there's a sub-section titled "Why NetApp for Infrastructure as a Service?" which includes a section about "Deliver differentiated, profitable cloud services". At the bottom, there's a diagram showing two server racks with red cubes on them, connected by a double-headed red arrow labeled "easy migration".

Why NetApp for Infrastructure as a Service?

Deliver differentiated, profitable cloud services

NetApp IaaS solutions help you differentiate your hosted and managed IaaS services to increase service and platform revenue.

By consolidating multitenant mixed workloads and multiple applications on a single cluster, you can reduce management headaches and cut operational costs by up to 67%.

Predictable IaaS performance for every volume in business-critical applications, and the ability to deploy applications in minutes, not months, helps you meet increasing customer demands while winning new ones in the process.

Cloud partners

NetApp cloud data services are available on the biggest clouds and integrated with leading technology partners and orchestration tools.



AWS

NetApp and Amazon Web Services (AWS) maintain a proven partnership to deliver agile, secure, and flexible data services for your most demanding cloud requirements.

[Discover more →](#)



Google Cloud

NetApp extends the reach of its world-class data services to Google Cloud's innovative leadership in application development, analytics, and machine learning.

[Discover more →](#)



Microsoft Azure

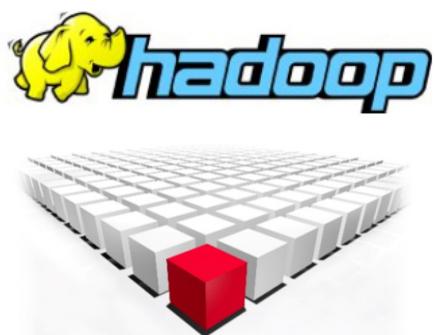
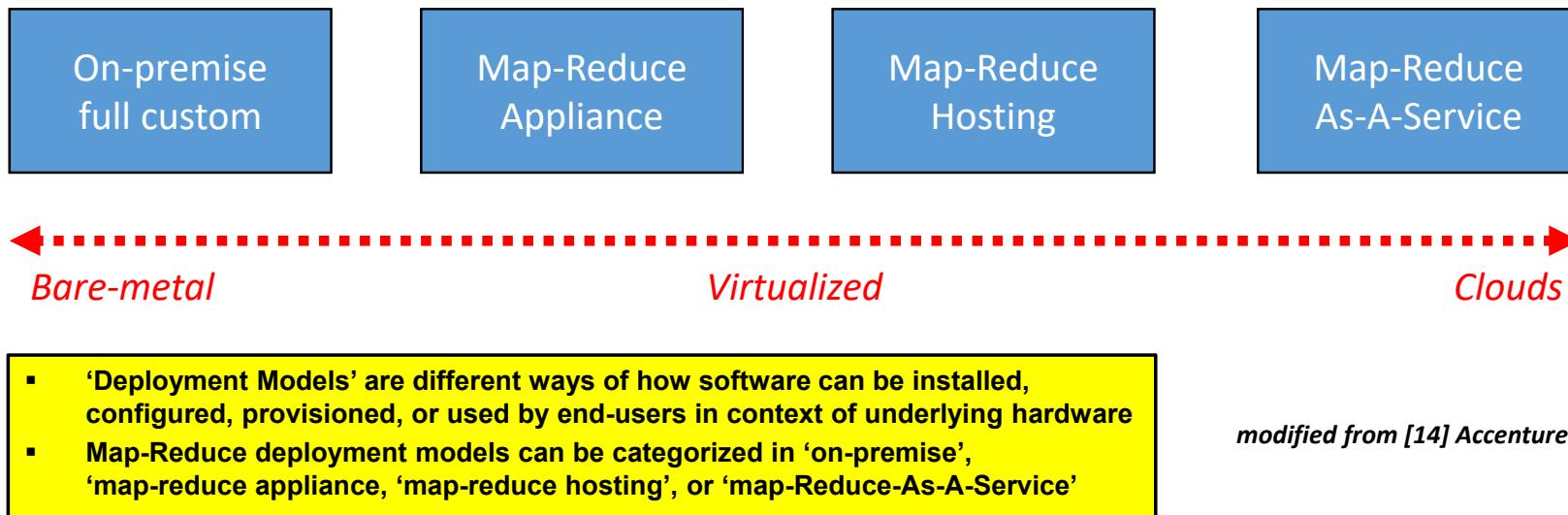
Together, NetApp and Microsoft help you deliver leading NetApp cloud data services as your own for a seamless experience that customers expect from the cloud.

[Discover more →](#)

NetApp Cloud data services build on top of and integrates with other IAAS of Cloud partners

[21] NetApp

IAAS & Map-Reduce Potential Deployment Models Example



- Traditional approaches to store and process data were around the ideas of isolated ‘data warehouses’ with customer data
- Products like Apache Hadoop Map-Reduce and HDFS store and process very large datasets in a scalable and cost effective way

modified from [14] Accenture Deployment Models



[15] Apache Hadoop Web Page

IAAS & Map-Reduce – Choosing Deployment Models & Criteria

- Price-Performance Ratio
 - Virtualized environments safe costs vs. slower with I/O operations
 - Data privacy (cf. Lecture 4 ‘Private Clouds’ vs. ‘Public Clouds’)
 - Concern when storing data outside of a ‘self-owned’ infrastructure
 - Data gravity
 - Data increases tremendously → lock in data infrastructure, migration costs
 - Data enrichment
 - Leveraging multiple datasets to uncover new insights, needs now vs. future
 - (Less) Productivity of data scientists (and developers/admins)
 - Explore the needs, create solution, deploy vs. in clouds already ‘production’

- Infrastructure in Infrastructure-as-a-Service (IAAS) is differently used per users: for some pure hardware for others deploying platforms: Map-Reduce
- The five key-areas that influence a deployment model decision for IAAS solutions are ‘price-performance ratio’, ‘data privacy’, ‘data gravity’, ‘data enrichment’, ‘productivity of developers and data scientists’



modified from [14] Accenture Deployment Models

Deployment Model A: On-premise Full Custom Deployment

On-premise
full custom

- Purchase commodity hardware and employ administrator(s)
- Install and operate the map-reduce software on own costs
- Enabling full control of the map-reduce cluster and configuration



- A bare-metal map-reduce cluster
 - Deployed on top of physical servers without any virtualization layer
 - E.g. plain Hadoop installation on a physical cluster
- ‘Total Cost of Ownership (TCO)’ matters
 - Compare the performance at the matched budget (not hardware specifications)
- Bare-metal with ‘On-premise full custom’ deployment
 - Costs for ‘staff for operation’, ‘technical support’, ‘data center facility’, ‘server hardware’
 - Don’t forget: ‘training the staff to use the deployment’ → ‘Learning curve’

[15] Apache Hadoop Web Page

modified from [14] Accenture Deployment Models

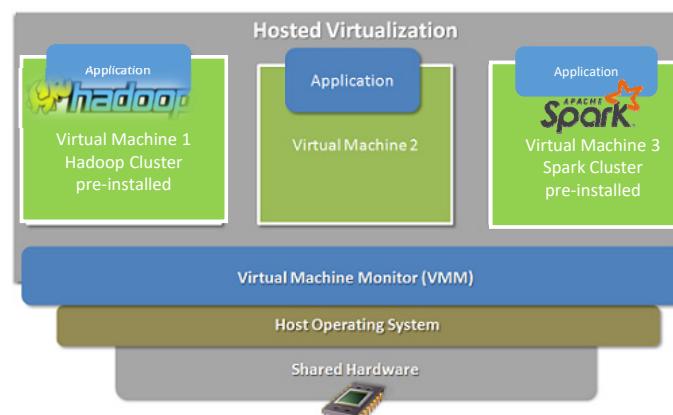
Deployment Model B: Map-Reduce Appliance Deployment

Map-Reduce
Appliance

- ‘Appliance’ means pre-configured map-reduce cluster
- Allows for bypassing detailed technical configuration decisions
- Enables an early start of data analysis based on map-reduce



- Take advantage of advancements in ‘virtualization technology’
 - Leveraging the broad advancements of virtualization technologies (cf. Lecture 4)
 - Not a cloud offering by a service provider, just a pre-installed image and using automated scripts often
 - Can be used to run different ‘map-reduce’ appliances depending on the needs of users
- Example: Pre-installed & configured Hadoop cluster with HDFS
 - Number of workers can be configured with virtual clusters to adjust to workloads of users



(another set of users may like to work with a Spark cluster & map-reduce)

[15] Apache Hadoop Web Page

modified from [16] National Instruments ‘Virtualization’ [17] Apache Spark

modified from [14] Accenture Deployment Models

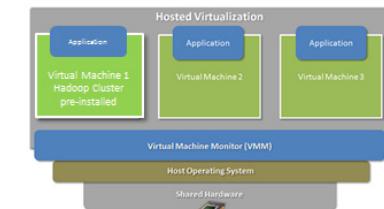
Deployment Model C: Map-Reduce Hosting Deployment

Map-Reduce Hosting

- Take advantage of 'Internet Service Provider (ISP)' model
- Buy-in a service provider to deploy and operate Map-Reduce cluster
- Enabling automating the deployment and operation processes



- Take advantage of '**map-reduce hosting**' by another organization
 - 'Outsourcing' of the deployment and operation of a map-reduce cluster
 - Use '**map-reduce hosting**' based on a '**Service Level Agreement (SLA)**'
- Example: Small-Medium Enterprise (SME)
 - Offers '**map-reduce hosting**'
 - Provides access of fractions of their own Apache Hadoop deployment on a '**pay-per-use**' model
 - Has a business model where map-reduce is not the focus, but is used anyway, so **idle resources** can be 'rented by others'



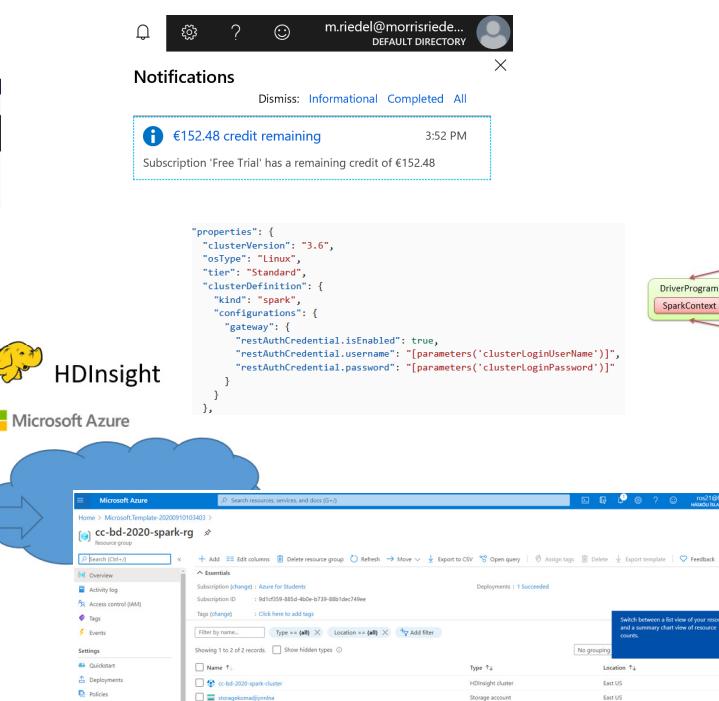
[15] Apache Hadoop Web Page

modified from [14] Accenture Deployment Models

Deployment Example using Template & Possibility to Adapt & Automate

- Microsoft Azure HDInsight Spark Clusters can be deployed via pre-configured resource manager templates that use Azure computing resources & Azure Storage Blobs as cluster storage
- A wide variety of templates are available on Github pages for various general cloud services

The screenshot shows a GitHub repository for an Azure HDInsight Spark cluster template. The repository includes a README file, deployment JSON files, and parameter files. A large blue button labeled "Deploy to Azure" is highlighted with a red box.

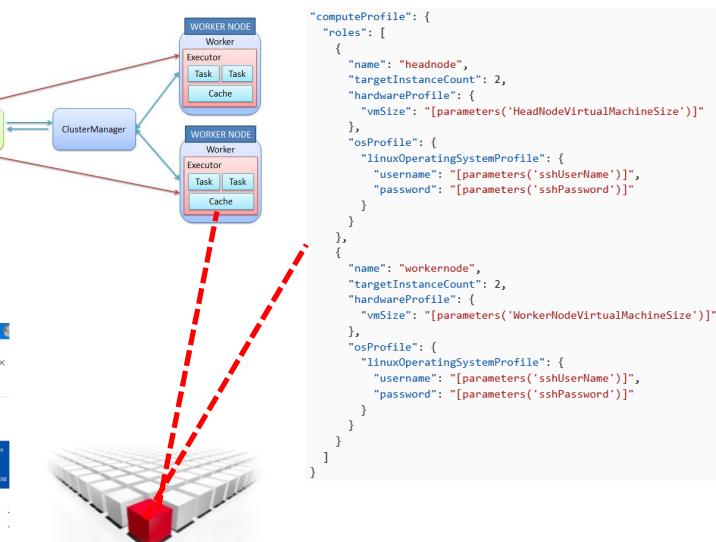


[18] Azure Portal Hub [19] MS Azure Storage

Review the template

The template used in this quickstart is from Azure Quickstart Templates.

```
JSON
{
    "$schema": "https://schema.management.azure.com/schemas/2019-04-01/deploymentTemplate.json#",
    "contentVersion": "1.0.0.0",
    "parameters": {
        "clusterName": {
            "type": "string",
            "metadata": {
                "description": "The name of the HDInsight cluster to create."
            }
        },
        "clusterLoginUserName": {
            "type": "string",
            "metadata": {
                "description": "These credentials can be used to submit jobs to the cluster and to log into cluster dashboards."
            }
        }
    },
    "computeProfile": {
        "roles": [
            {
                "name": "headnode",
                "targetInstanceCount": 2,
                "hardwareProfile": {
                    "vmSize": "[parameters('HeadNodeVirtualMachineSize')]"
                },
                "osProfile": {
                    "linuxOperatingSystemProfile": {
                        "username": "[parameters('sshUserName')]",
                        "password": "[parameters('sshPassword')]"
                    }
                }
            },
            {
                "name": "workernode",
                "targetInstanceCount": 2,
                "hardwareProfile": {
                    "vmSize": "[parameters('WorkerNodeVirtualMachineSize')]"
                },
                "osProfile": {
                    "linuxOperatingSystemProfile": {
                        "username": "[parameters('sshUserName')]",
                        "password": "[parameters('sshPassword')]"
                    }
                }
            }
        ]
    }
}
```



Deployment Model D: Map-Reduce Hosting Deployment

Map-Reduce As-A-Service

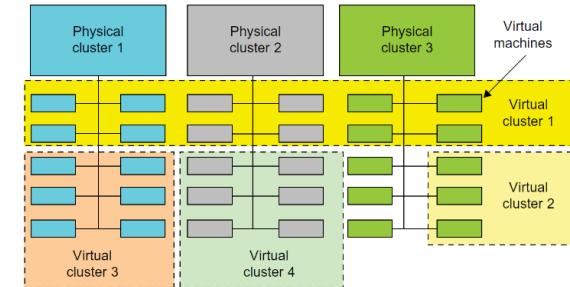
- Provides instant access to a map-reduce cluster as a ‘cloud service’
- Take advantage of the ‘pay-per-use’ model (aka ‘swipe credit card’)
- Offers dynamic usage of the service according to workload peaks

■ A ‘cloud’ that offers a Map-Reduce-As-A-Service capability

- Ideally connected to a ‘storage-as-a-service’ capability (‘data locality’)
- E.g. ‘Hadoop-as-a-Service’ as cloud service directly to use
- No ‘TCO’, but also costs with ‘pay-per-use’ model

■ E.g. Amazon Elastic MapReduce (EMR), cf. Lecture 5

- Enables users to deploy Hadoop clusters on demand
- Fastest way to run MapReduce jobs
- Tear down the clusters when the jobs are completed
- Dynamic usage (perfect for workload peaks)
- Based on Apache Hadoop & Apache Spark



[17] Apache Spark



[15] Apache Hadoop Web Page



[2] Distributed & Cloud Computing Book

modified from [14] Accenture Deployment Models

Deployment Example using AWS Cloud Service – Choose Analytic Service

- Multiple analytics products (cf. Lecture 5)
 - Extracting insights and actionable information from data requires technologies like analytics & machine learning
- Analytics Services (cf. Practical Lecture 5)
 - Amazon Athena: Serverless Query Service
 - **Amazon ElasticMapReduce (EMR): Hadoop**
 - Amazon ElasticSearch Service: Elasticsearch on AWS
 - Amazon Kinesis: Streaming Data
 - Amazon QuickSight: Business Analytics
 - Amazon Redshift: Data Warehouse
- AWS offers a wide variety of Big Data Analytics Cloud services such as Amazon Athena (interactive analytics), Amazon EMR (big data processing with Apache Hadoop & Spark), Amazon Redshift (data warehousing), Amazon Kinesis (real-time analytics), Amazon Elasticsearch Service (operational analytics), and Amazon Quicksight (dashboards and visualizations)



[3] AWS Web page

AWS Analytics services		
Category	Use cases	AWS service
Analytics	Interactive analytics	Amazon Athena
	Big data processing	Amazon EMR
	Data warehousing	Amazon Redshift
	Real-time analytics	Amazon Kinesis
	Operational analytics	Amazon Elasticsearch Service
	Dashboards and visualizations	Amazon Quicksight

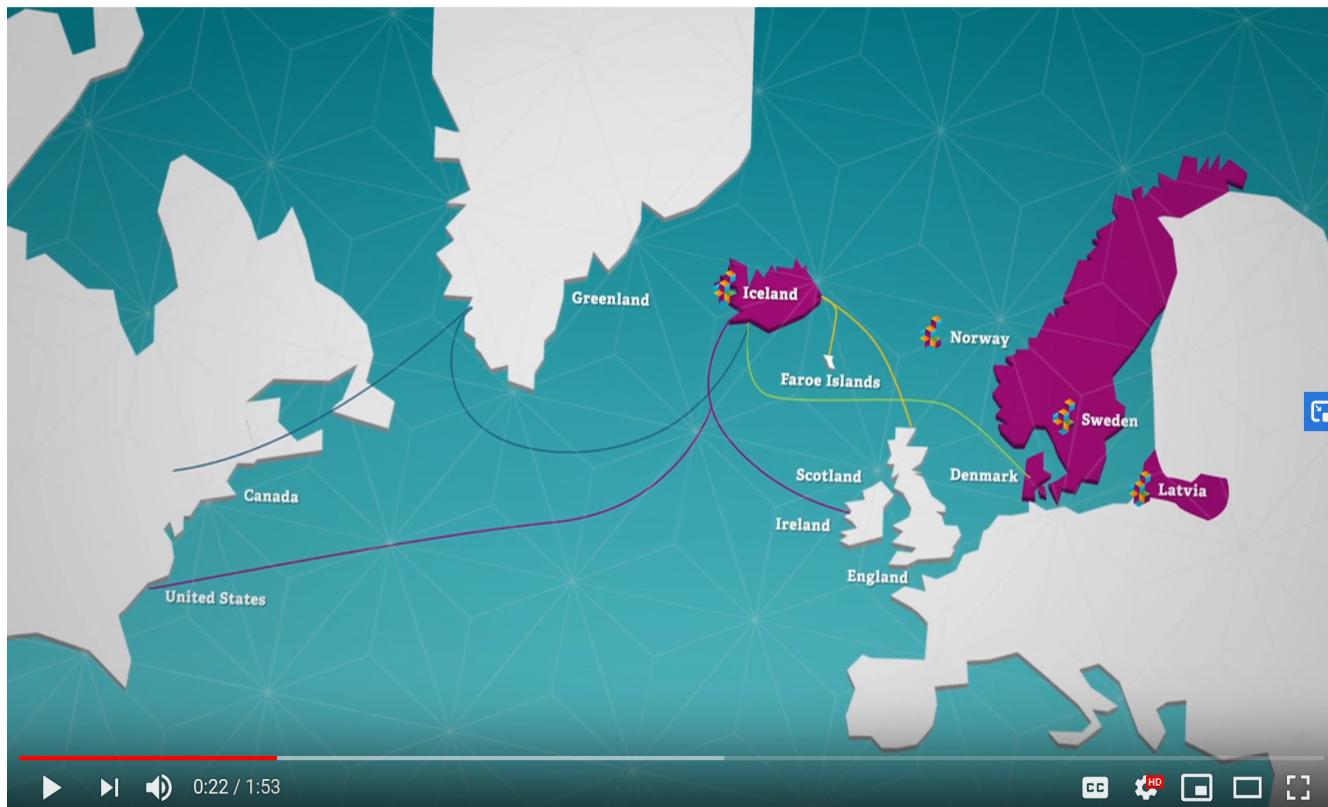
Amazon EMR

Easily run and scale Apache Spark, Hive, Presto, and other big data frameworks

[Get started with Amazon EMR](#)

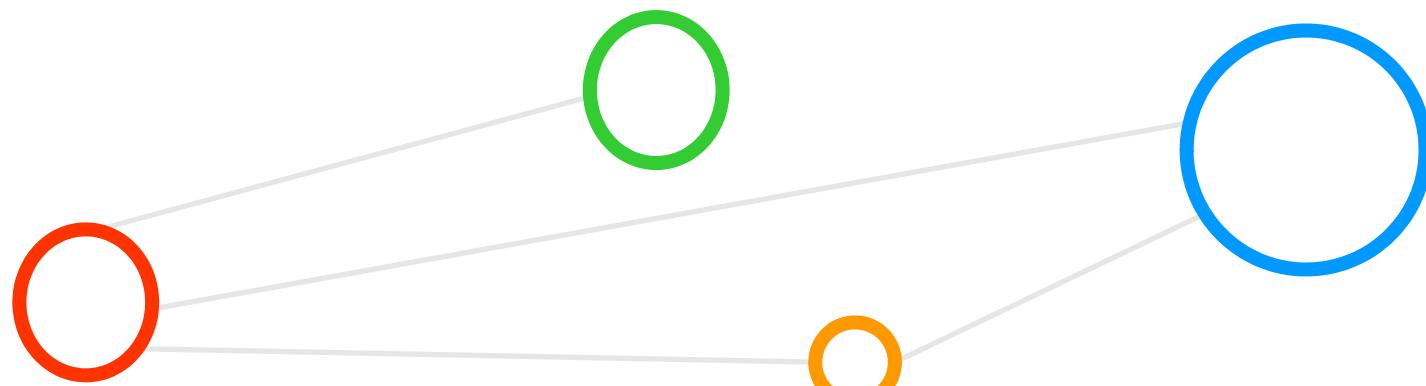
[Request support for your evaluation](#)

[Video] Advania Data Center Example in Iceland



[3o] YouTube video, Advania Data Centers HPC Services

Lecture Bibliography



Lecture Bibliography (1)

- [1] Wikipedia, 'Cloud Computing', Online:
<http://de.wikipedia.org/wiki/Cloud-Computing>
- [2] K. Hwang, G. C. Fox, J. J. Dongarra, 'Distributed and Cloud Computing', Book, Online:
http://store.elsevier.com/product.jsp?locale=en_EU&isbn=9780128002049
- [3] Amazon Web Services Web Page, Online:
<https://aws.amazon.com>
- [4] 'Private Cloud, big-data.tips, Online:
<http://www.big-data.tips/private-cloud>
- [5] Key Concepts from the AWS Cloud, Online:
<https://blogs.sap.com/2010/01/28/key-concepts-from-the-aws-cloud/>
- [6] SSH Client MobaXterm, Online:
<https://mobaxterm.mobatek.net/>
- [7] Amazon Web Services EC2 On-Demand Pricing models, Online:
<https://aws.amazon.com/ec2/pricing/on-demand/>
- [8] YouTube Video, 'What is IaaS?', Online:
https://www.youtube.com/watch?v=tcw_QZ38Mlw
- [9] www.big-data.tips, 'MNIST Database', Online:
<http://www.big-data.tips/mnist-database>
- [10] www.big-data.tips, 'MNIST dataset', Online:
<http://www.big-data.tips/mnist-dataset>
- [11] Harley, A.W., An Interactive Node-Link Visualization of Convolutional Neural Networks, Online:
<http://scs.ryerson.ca/~aharley/vis/conv/flat.html>

Lecture Bibliography (2)

- [12] A. Rosebrock, 'Get off the deep learning bandwagon and get some perspective', Online:
<http://www.pyimagesearch.com/2014/06/09/get-deep-learning-bandwagon-get-perspective/>
- [13] Google Colaboratory, Online:
<https://colab.research.google.com>
- [14] Accenture, Hadoop Deployment Comparison Study, Online:
<http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Hadoop-Deployment-Comparison-Study.pdf>
- [15] Apache Hadoop Web page, Online:
<http://hadoop.apache.org/>
- [16] Virtualization, National Instruments White Paper, Online:
<http://www.ni.com/white-paper/8709/en/>
- [17] Apache Spark Web page, Online:
<http://spark.apache.org/>
- [18] Microsoft Azure Portal Hub, Online:
<https://portal.azure.com/#home>
- [19] MS Azure Storage, Online:
<https://azure.microsoft.com/en-us/pricing/details/storage/>
- [20] OpenStack Cloud Operating System, Online:
<https://www.openstack.org/>
- [21] YouTube Video, 'Driving the Future of IT Infrastructure at Volkswagen Group', Online:
https://www.youtube.com/watch?v=HL_pzkDnai4
- [22] NetApp IAAS, Online:
<https://www.netapp.com/hybrid-cloud/service-provider-infrastructure/infrastructure-as-a-service-iaas/>

Lecture Bibliography (3)

- [23] AWS Marketplace, Online:
<https://aws.amazon.com/marketplace/>
- [24] AWS S3 Pricing Models, Online:
<https://aws.amazon.com/s3/pricing>
- [25] Safari book online, 'RESTful.NET', Online:
<https://www.safaribooksonline.com/library/view/restful-net/9780596155025/ch04.html>
- [26] AWS EBS Pricing Models, Online:
<https://aws.amazon.com/ebs/pricing/>
- [27] AWS EFS Pricing Models, Online:
<https://aws.amazon.com/efs/pricing/>
- [28] Advania Compute IAAS HPCFlow IAAS Clusters, Online:
<https://advaniadc.com/compute/iaas>
- [29] CIO, Taking IaaS for AI and HPC to a new level in Iceland, Online:
<https://www.cio.com/article/3322627/taking-iaas-for-ai-and-hpc-to-a-new-level-in-iceland.html>
- [30] YouTube Video, 'Advania Data Centers HPC Services', Online:
<https://www.youtube.com/watch?v=j-8TZ9rmfSQ>
- [31] YouTube Video, 'Verne Global NVIDIA DGX-Ready', Online:
https://www.youtube.com/watch?v=5EpTVFBHz_o

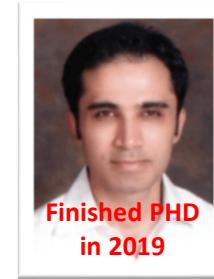
Acknowledgements – High Productivity Data Processing Research Group



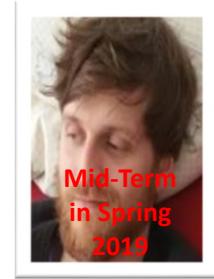
Finished PhD
in 2016



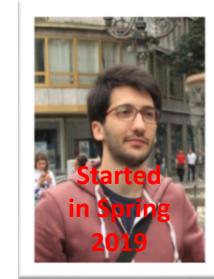
Finishing
in Winter
2019



Finished PhD
in 2019



Mid-Term
in Spring
2019



Started
in Spring
2019



Started
in Spring
2019

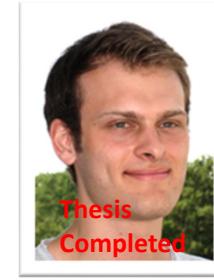
Morris Riedel @MorrisRiedel · Feb 10
Enjoying our yearly research group dinner 'Iceland Section' to celebrate our productive collaboration of @uni_iceland @uisens @Haskell_Islands & @fz_jsc @fz_juelich & E.Erlingsson @emrie passed mid-term in modular supercomputing driven by @DEEPprojects - morrisriedel.de/research

A photograph showing a group of people seated around tables in a restaurant. They are dressed in casual to semi-formal attire. The room has warm lighting and traditional Icelandic decorations on the walls.

Finished PhD
in 2018



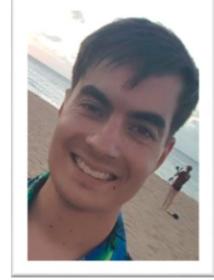
MSc M.
Richerzhagen
(now other division)



MSc
P. Glock
(now INM-1)



MSc
C. Bodenstein
(now
Soccerwatch.tv)



MSc Student
G.S. Guðmundsson
(Landsverkjun)



This research group has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 763558 (DEEP-EST EU Project)

