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

@Morris Riedel @MorrisRiedel @MorrisRiedel

LECTURE 8

Infrastructure-As-A-Service (IAAS)

October 29, 2020
Online Lecture

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

JÜLICH Forschungszentrum
JÜLICH SUPERCOMPUTING CENTRE

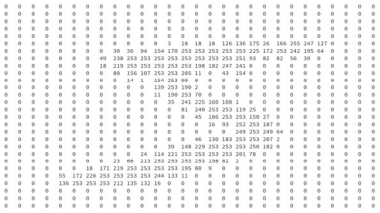
DEEP Projects

HELMHOLTZ AI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

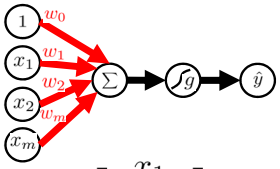
Review of Practical Lecture 7.1 – Using Deep Learning Techniques in Clouds

- Using Artificial Neural Networks (ANNs)

(data inspection & understanding)



'training set' *'test set'*



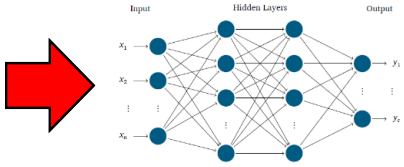
Training Examples

$$(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$$

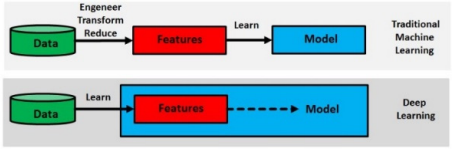
$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$

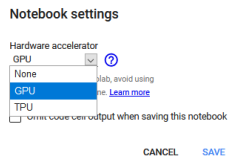
(historical records, groundtruth data, examples)

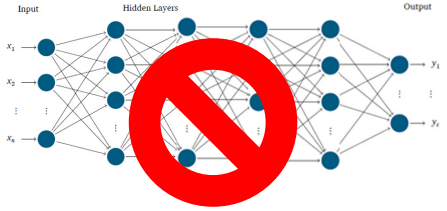
(data exploration, flatening, and normalization)

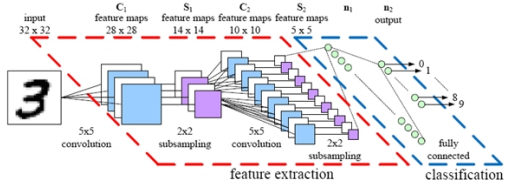


- Using Convolutional Neural Networks (CNNs)
  - Deep Learning Technique → Not good with CPUs



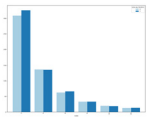


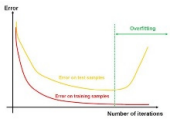




- ✓ **Multi Output Perceptron: ~91,01% (20 Epochs)**
- ✓ **ANN 2 Hidden Layers: ~95,14 % (20 Epochs)**
- ✓ **CNN Deep Learning Model: ~99,36 % (20 Epochs)**

(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

1. Cloud Computing & Big Data Introduction
2. Machine Learning Models in Clouds
3. Apache Spark for Cloud Applications
4. Virtualization & Data Center Design
5. Map-Reduce Computing Paradigm
6. Deep Learning driven by Big Data
7. Deep Learning Applications in Clouds
8. Infrastructure-As-A-Service (IAAS)
9. Platform-As-A-Service (PAAS)
10. Software-As-A-Service (SAAS)
11. Big Data Analytics & Cloud Data Mining
12. Docker & Container Management
13. OpenStack Cloud Operating System
14. Online Social Networking & Graph Databases
15. Big Data Streaming Tools & Applications
16. Epilogue

+ additional practical lectures & Webinars for our hands-on assignments in context

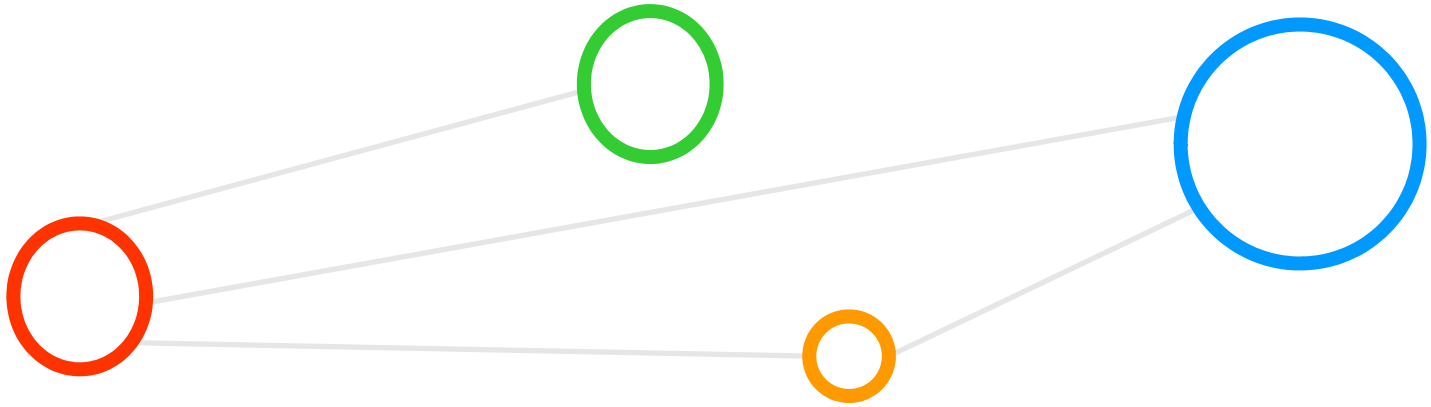
- Practical Topics
- Theoretical / Conceptual Topics

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

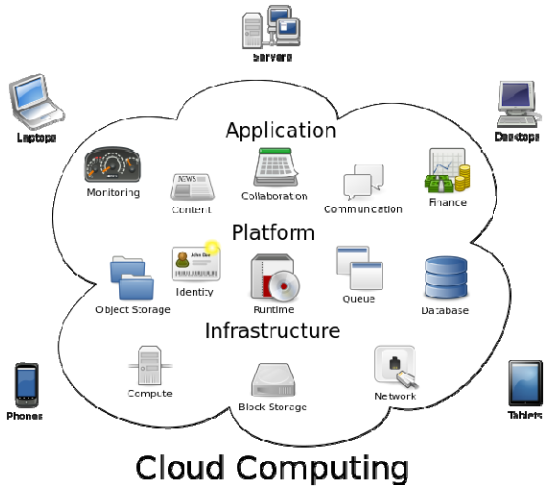
- Selected Cloud Systems (aka 'Clouds') known today
  - Google Cloud → massive computing/storage/applications
  - Amazon Web Service → massive computing/storage/services
  - Microsoft Azure → massive computing/storage/toolsets
  - Facebook → online social networking & advertisement
  - SalesForce.com → 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

- **Cloud systems play an increasingly important role in upgrading traditional Web services and Internet applications to modern scalable online services**
- **Large-scale Internet applications driven by cloud computing have significantly enhanced the quality of life in society today**
- **In one form or another are cloud computing applications requiring large amounts of computing capability or storage capacity**

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

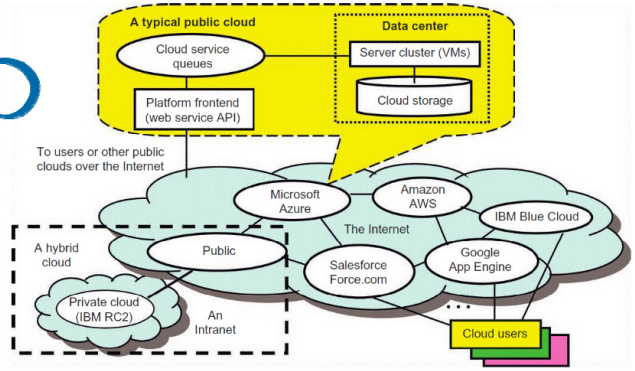


[1] Wikipedia 'Cloud computing'

- **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

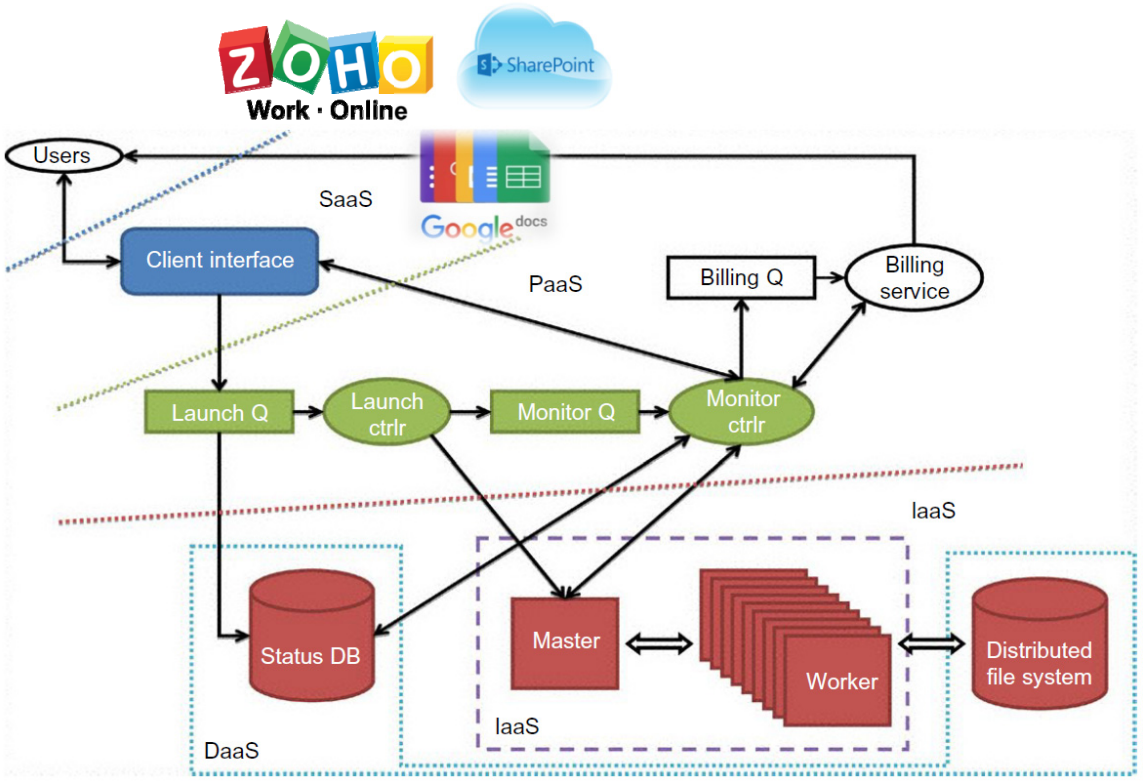
- Private Clouds are built within the domain of an intranet owned by a single organization
- Private Clouds are 'client owned' and managed and access is limited to owning clients
- Private Clouds do not sell resources over the Internet but provide them in the organization
- Operations only in one 'private 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**

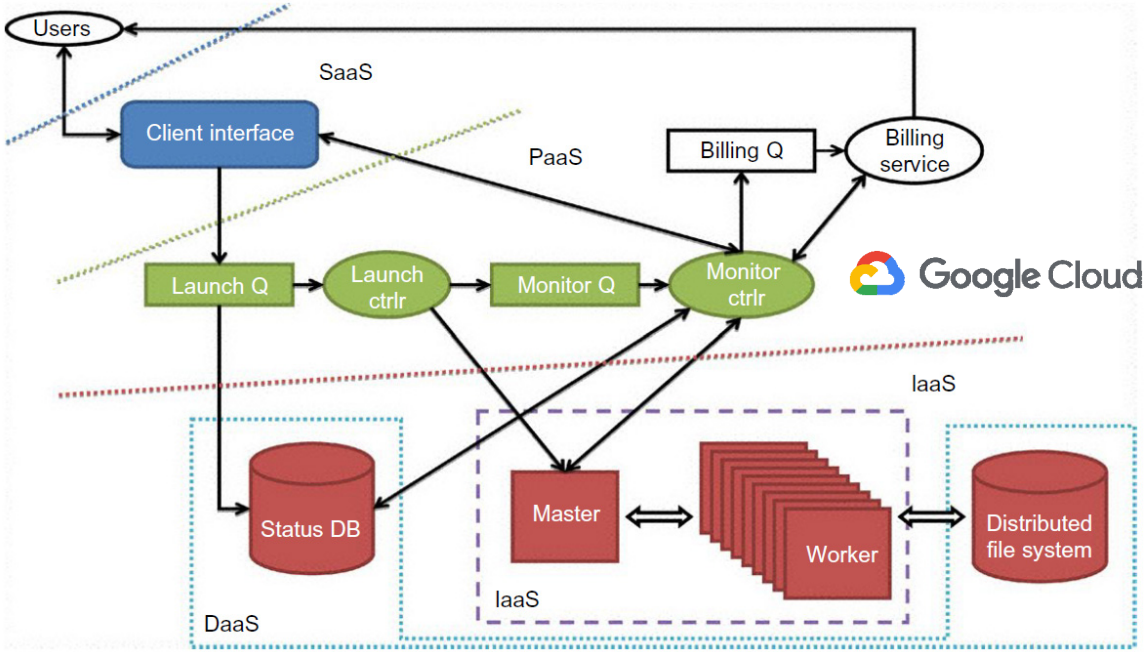


[2] Distributed & Cloud Computing Book

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.**

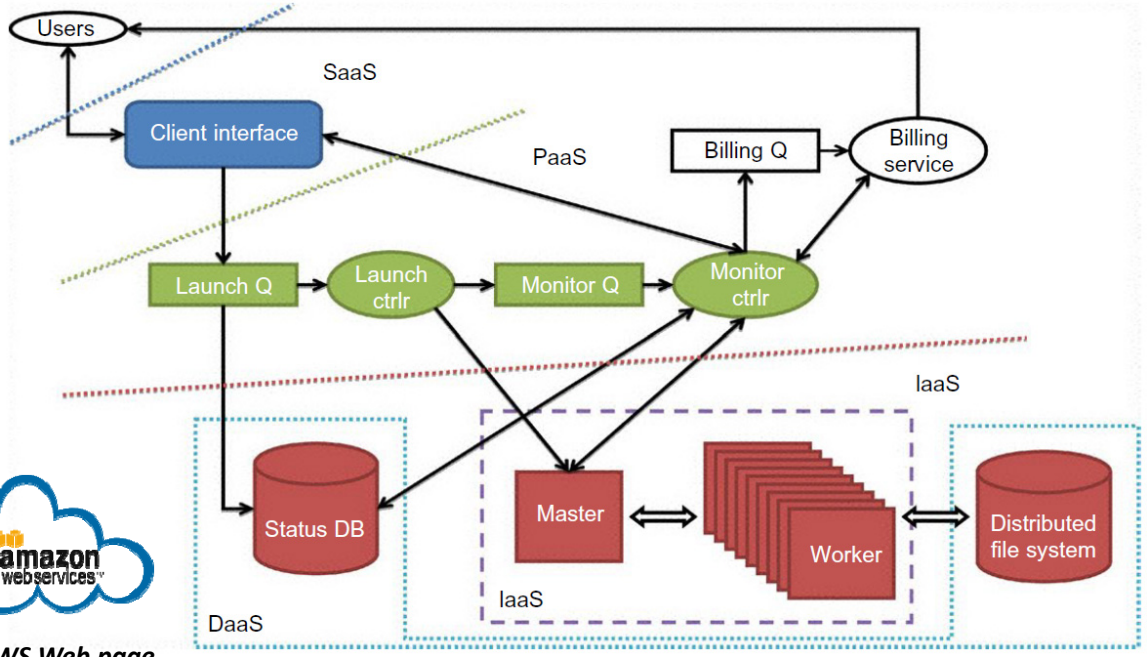


[2] Distributed & Cloud Computing Book

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**



[2] Distributed & Cloud Computing Book

VERNE GLOBAL

[29] Verne Global in Iceland

amazon webservices

[3] AWS Web page

advania data centers

[28] Advania HPCFlow IAAS Clusters

IAAS Provider Example – Amazon Web Services (AWS)

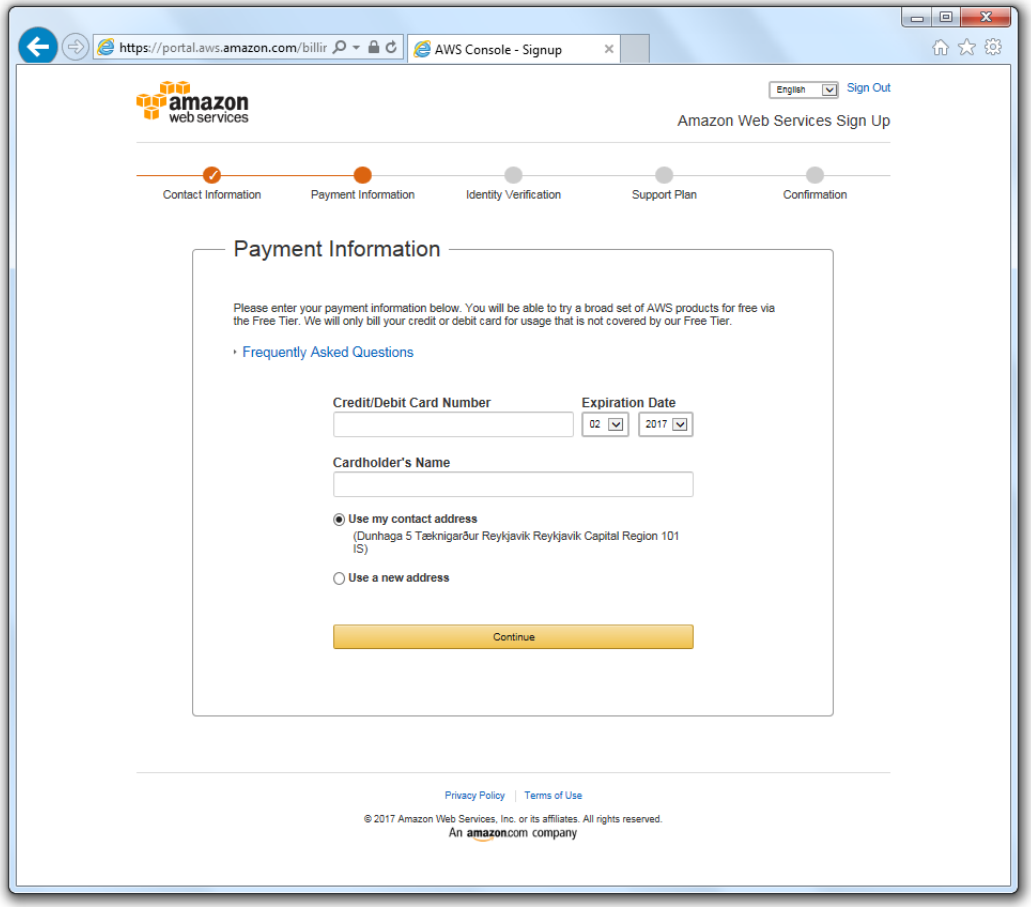
- On-Demand delivery of virtualized IT resources
  - Compute power, database storage, networking
  - Application services, databases, specific software
  - Offers ‘rapid access‘ to flexible and low cost IT resources
  - Access as many resources as you need (~instantly)
  - Using pay-as-you-go pricing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

[3] AWS Web page

- **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

(service portfolio 2020)

[3] AWS Web page

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

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

Database
Amazon Aurora
Amazon RDS
Amazon DynamoDB
Amazon ElastiCache
Amazon Redshift
AWS Database Migration Service

Migration
AWS Discovery Services
AWS Database Migration Service
AWS Server Migration Service
AWS Snowball
AWS Snowball Edge
AWS Snowmobile

Networking & Content Delivery
Amazon VPC
Amazon CloudFront
Amazon Route 53
AWS Direct Connect
Elastic Load Balancing

Developer Tools
AWS CodeCommit
AWS CodeBuild
AWS CodeDeploy
AWS CodePipeline
AWS X-Ray
AWS Command Line Interface

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

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

Analytics
Amazon Athena
Amazon EMR
Amazon CloudSearch
Amazon Elasticsearch Service
Amazon Kinesis
Amazon Redshift
Amazon QuickSight
AWS Data Pipeline
AWS Glue

Artificial Intelligence
Amazon Lex
Amazon Polly
Amazon Rekognition
Amazon Machine Learning

Mobile Services
AWS Mobile Hub
Amazon API Gateway
Amazon Cognito
Amazon Pinpoint
AWS Device Farm
AWS Mobile SDK

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

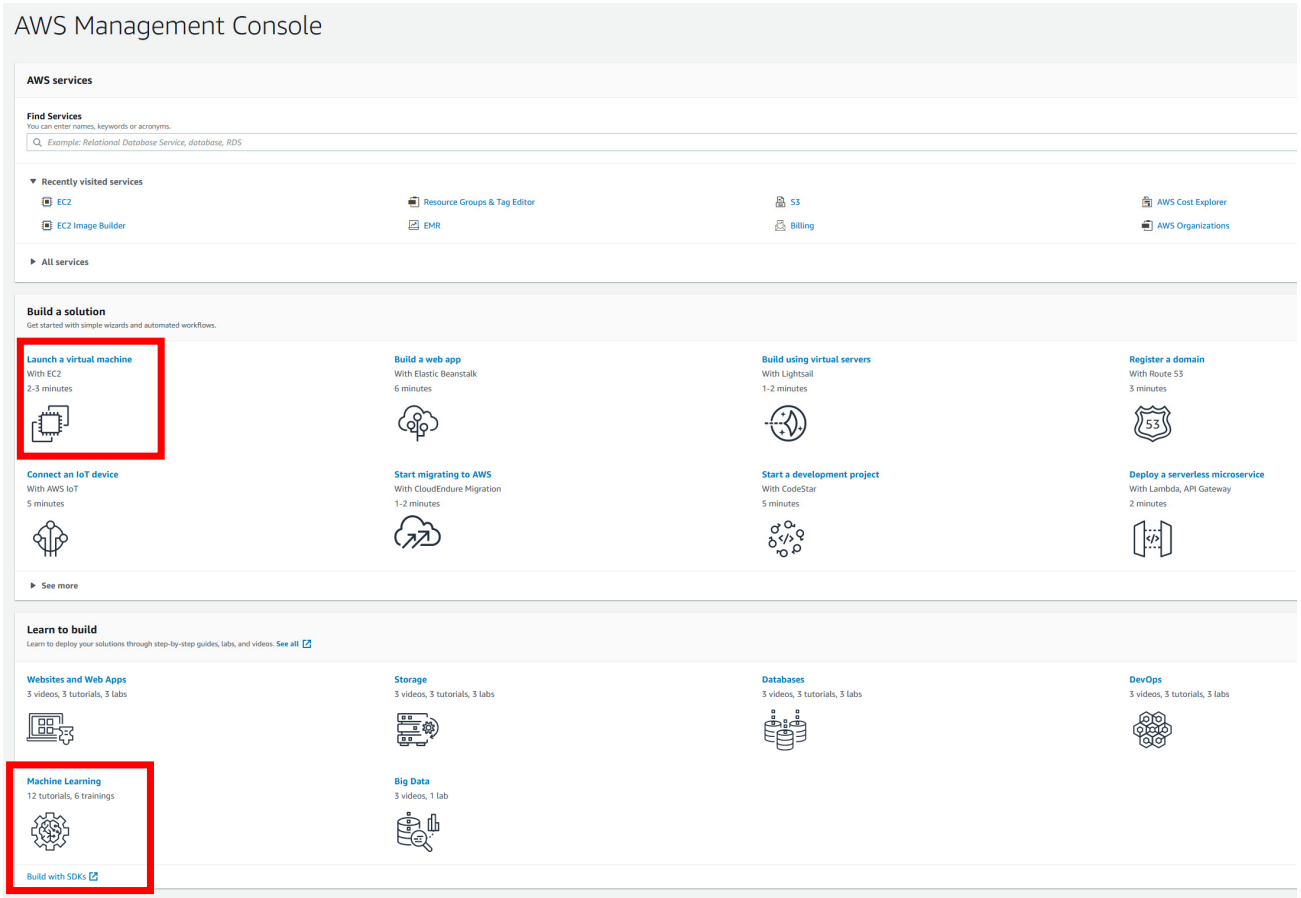
Game Development
Amazon Lumberyard

(cf. service portfolio 2018)

AWS Management Console – A Plethora of Services – Another Perspective

(cf. Lecture 4)

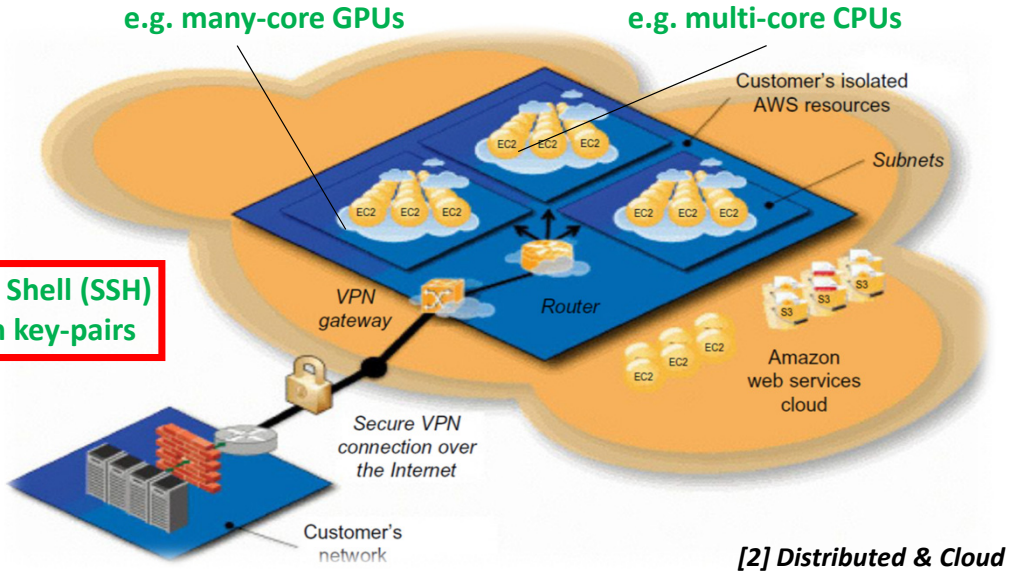
(cf. Lecture 2, 6, 7)



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



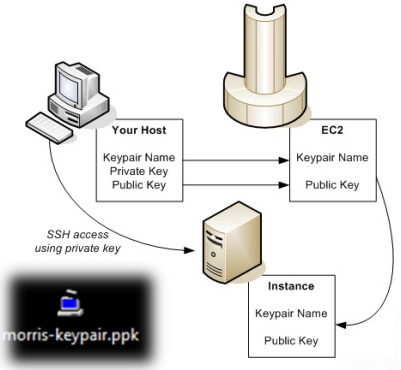
e.g. Secure Shell (SSH) access with key-pairs

[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

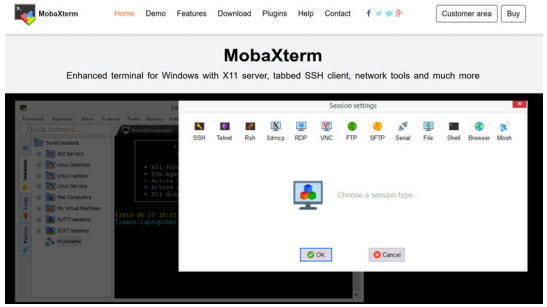
- **The Secure Shell (SSH) is a technique to securely access remote AWS computing instances (e.g., AWS EC2) using a named key pair**
- **An SSH key pair consists of a public key that is known by the Amazon Cloud and a private key that remains only on the laptop of cloud users**



- **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**

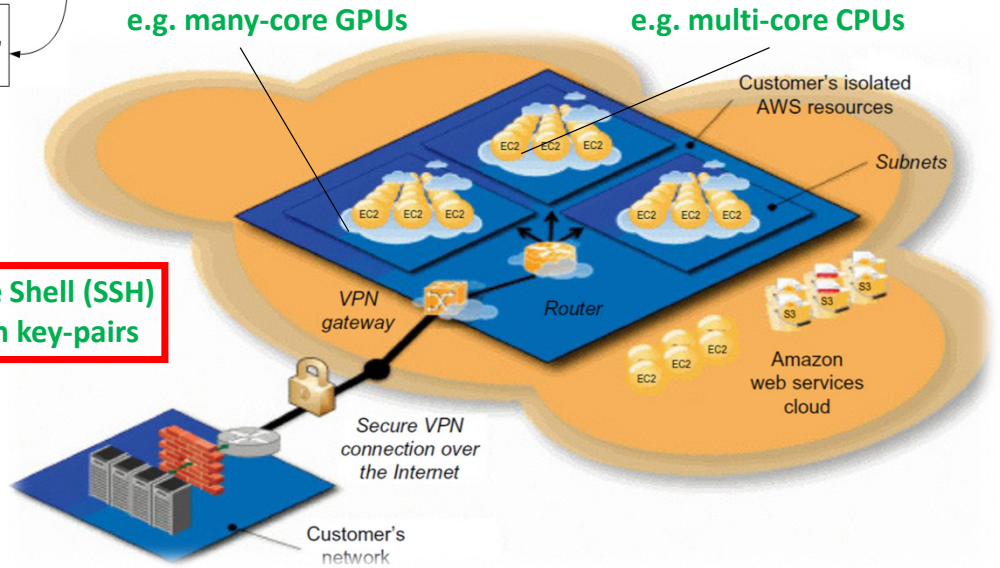
[3] Amazon Web Services

e.g. Secure Shell (SSH) access with key-pairs



(SSH client is necessary)

[6] MobaXterm Web page



[5] Key Concepts from the AWS Cloud

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

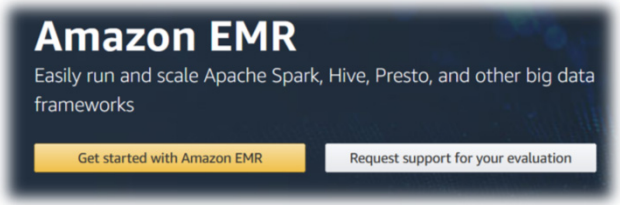
- Multiple analytics products
  - Extracting insights and actionable information from data requires technologies like analytics & machine learning
- Selected Analytics Services (cf. Lecture 5)
  - Amazon Athena: Serverless Query Service
  - Amazon ElasticMapReduce (EMR): Hadoop
  - Amazon ElasticSearch 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**

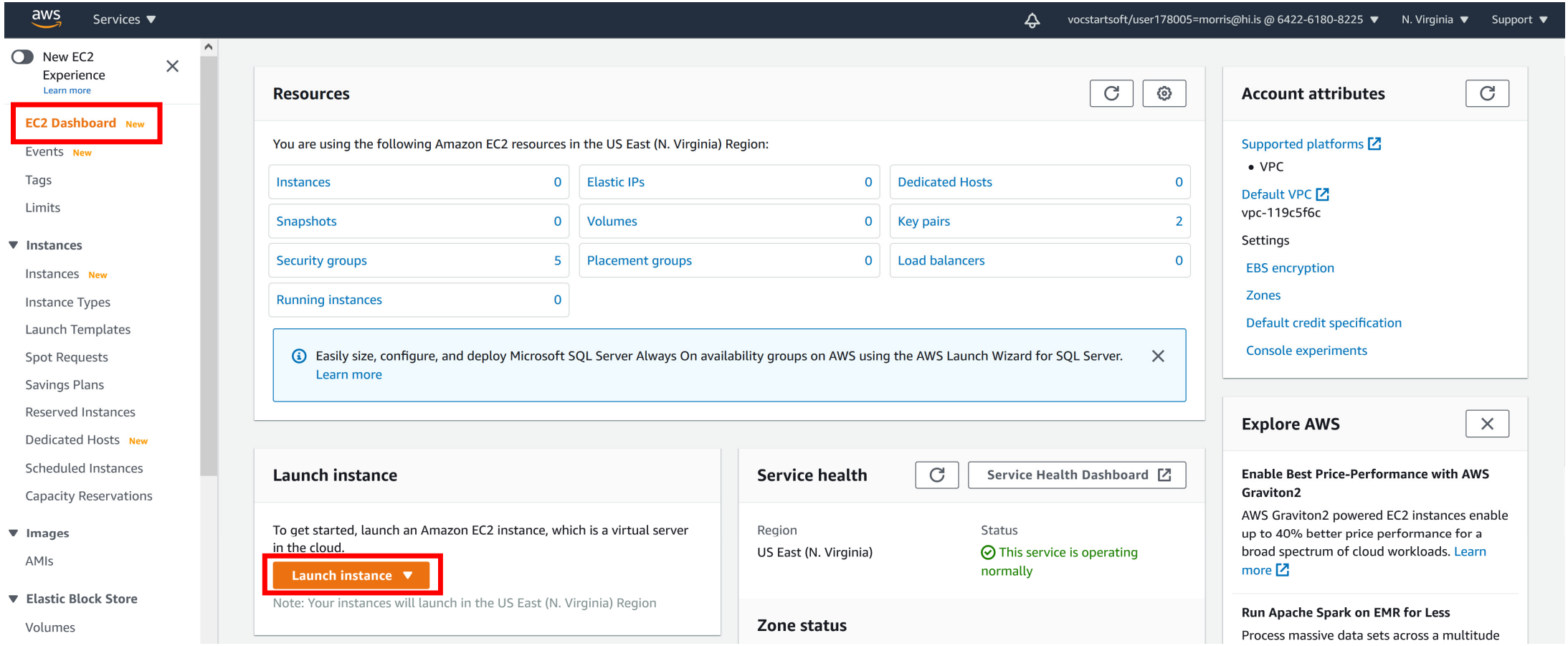
[3] AWS Web page

AWS Analytics services

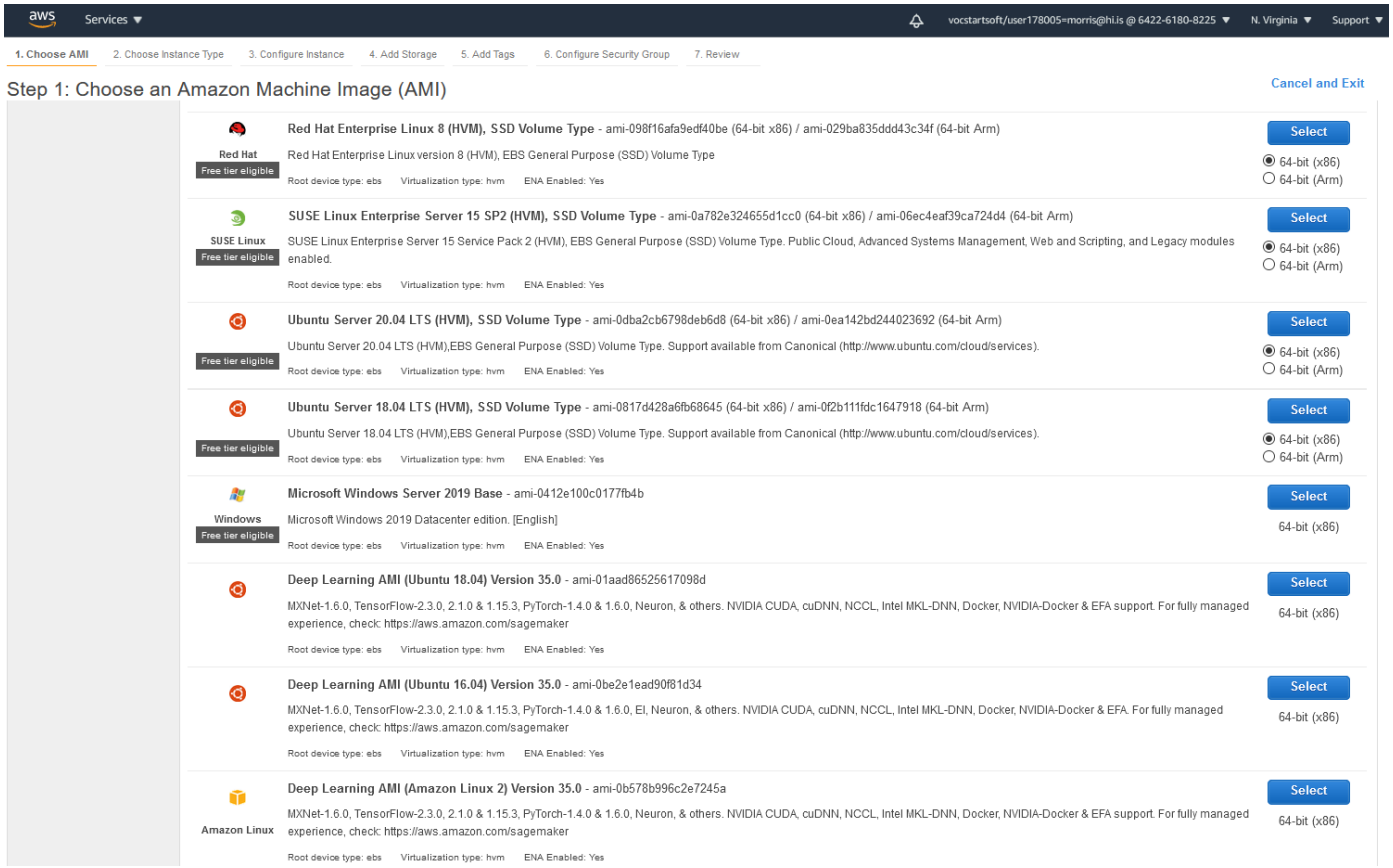
Category	Use cases	AWS service
Analytics	Interactive analytics	Amazon Athena
	Big data processing	Amazon EMR
	Data warehousing	Amazon Redshift
	Real-time analytics	Amazon Kinesis
	Operational analytics	Amazon Elasticsearch Service
	Dashboards and visualizations	Amazon Quicksight



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



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



- **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

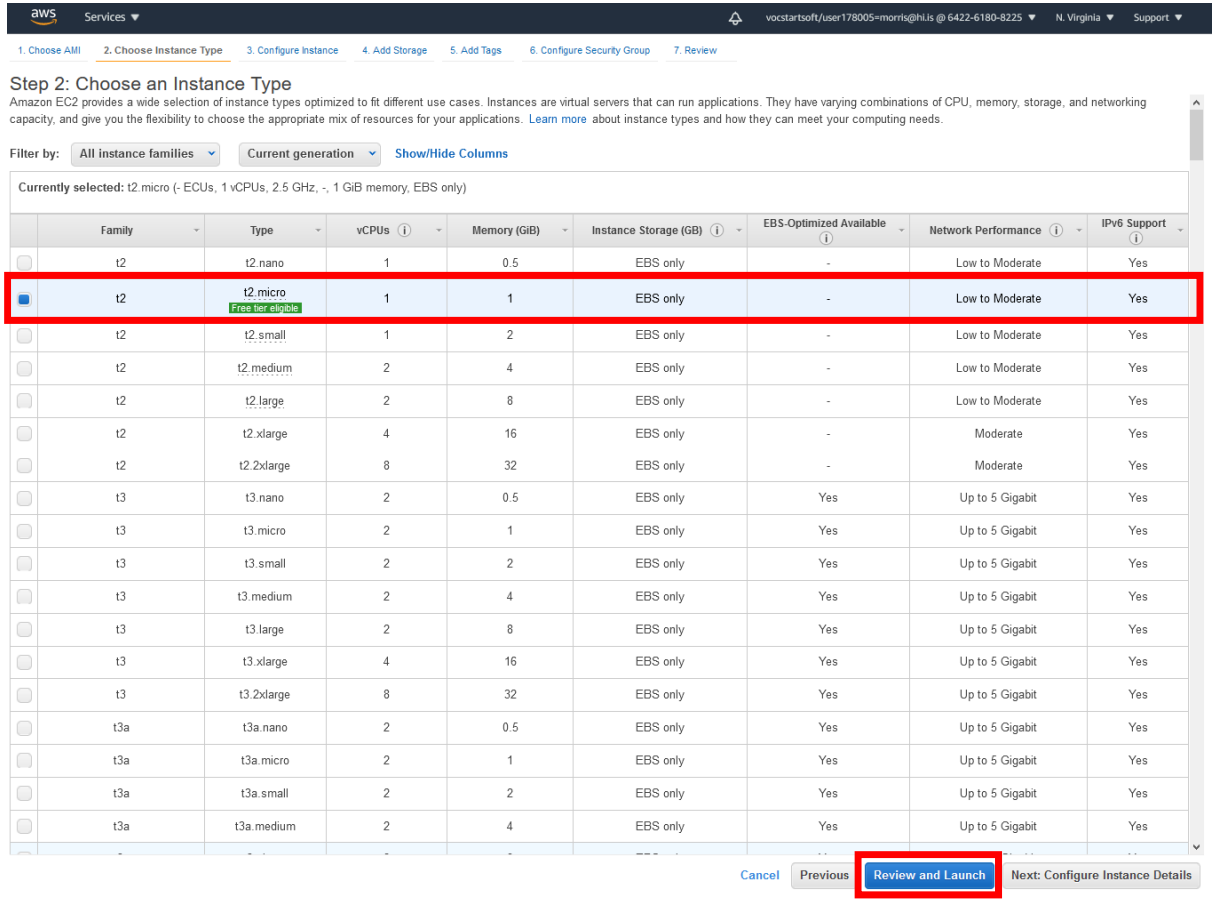
Classical Computing	**Cloud Computing**
(Repeat the following cycle every 18 months)	(Pay as you go per each service provided)
Buy and own	**Subscribe**
Hardware, system software, applications to meet peak needs	- - - -
Install, configure, test, verify, evaluate, manage	**Use** (Save about 80-95% of the total cost)
- - - -	- - - -
Use	(Finally)
- - - -	**$ - Pay for what you use**
Pay $$$$$ (High cost)	based on the QoS

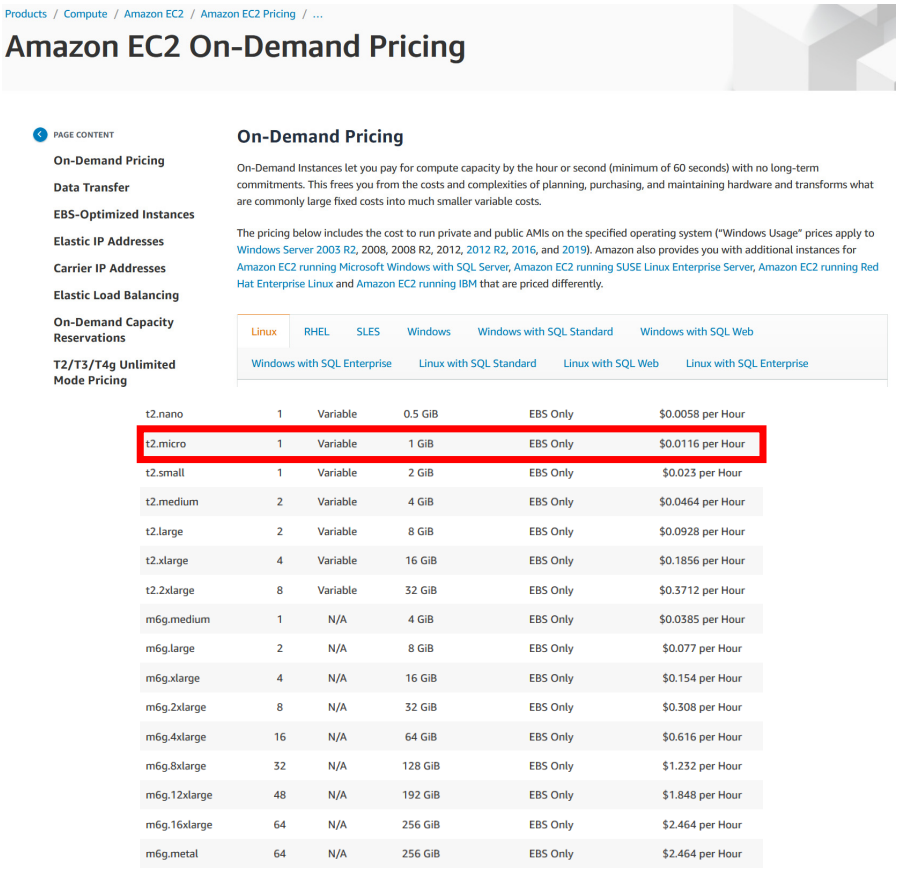
- **AWS offers also free tier and AWS Educate access with limited resources & certain specific quota on services**

- **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





[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
- 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%)

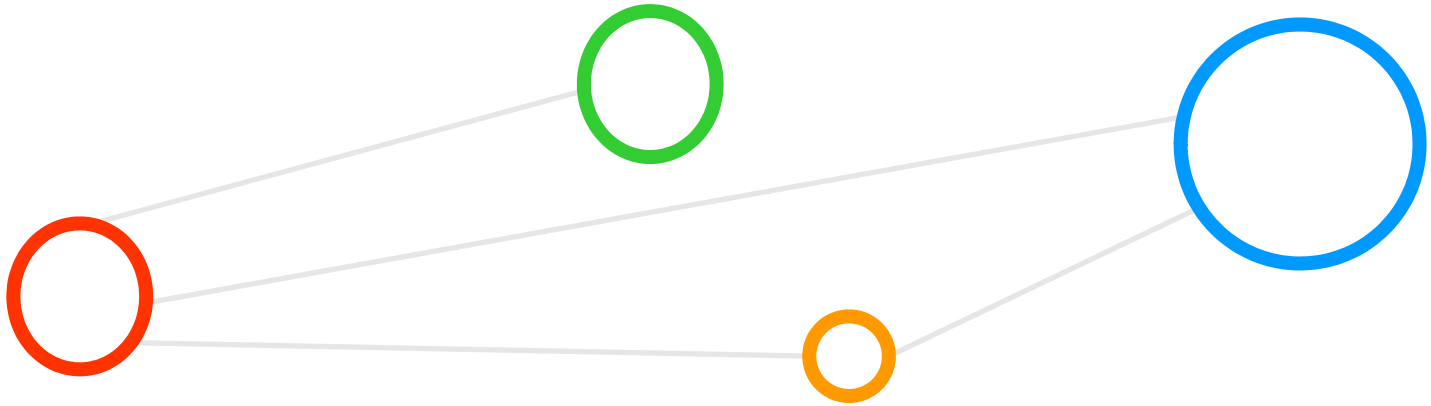
Model	vCPU	CPU Credits / hour	Mem (GiB)
t2.nano	1	3	0.5
t2.micro	1	6	1
t2.small	1	12	2
t2.medium	2	24	4
t2.large	2	36	8
t2.xlarge	4	54	16
t2.2xlarge	8	81	32

[7] AWS EC2 Pricing

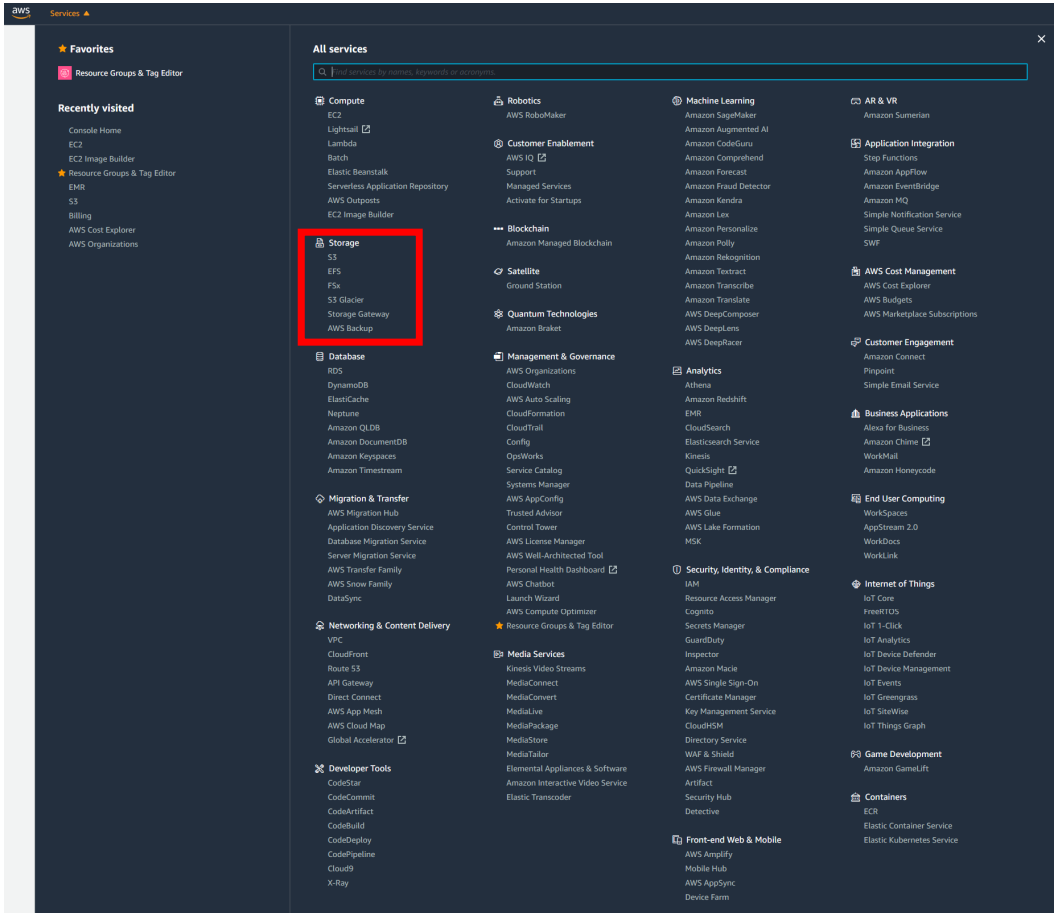
[Video] Infrastructure-As-A-Service Goals & Application Areas

[8] YouTube video, What is IaaS?

Advanced IAAS Topics & Applications

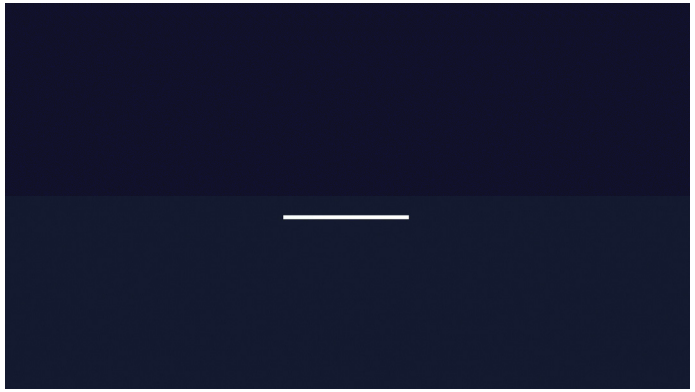


IAAS for different Hardware – Examples from AWS, Verne Global & Advania



[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

advania data centers

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

HPCFLOW: IaaS Clusters

HPC**FLOW** 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. HPC**FLOW** allows you the ultimate AI & HPC cluster elasticity and flexibility.

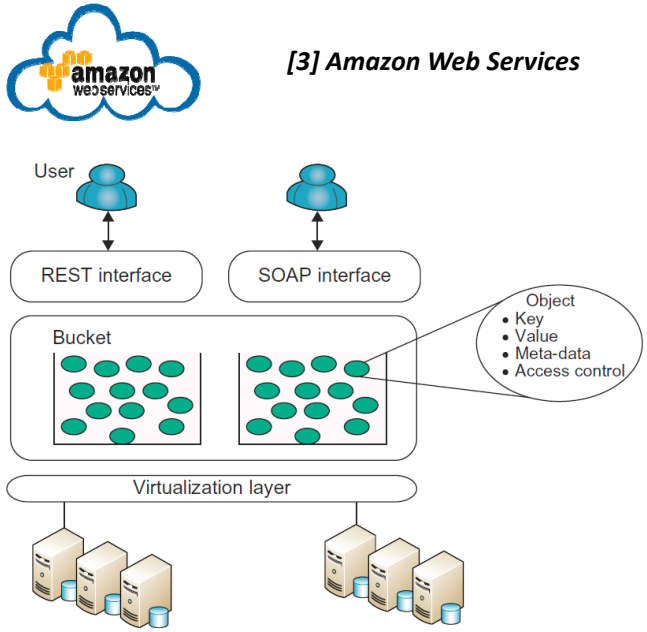
Contact us →

[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

Expedia airbnb NETFLIX Allergan NASA JPL CATHAY PACIFIC

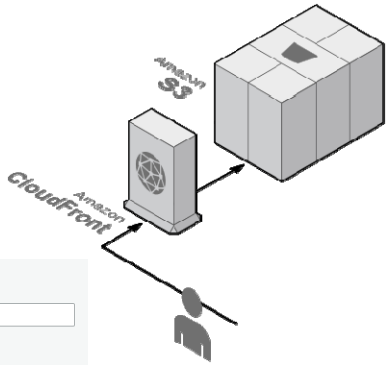


[3] Amazon Web Services

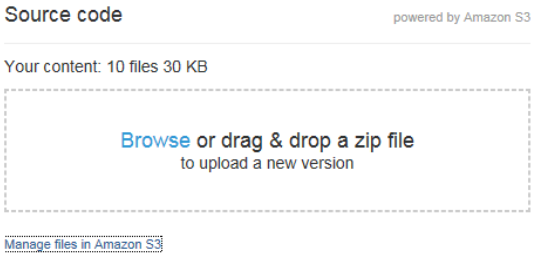
[2] Distributed & Cloud Computing Book

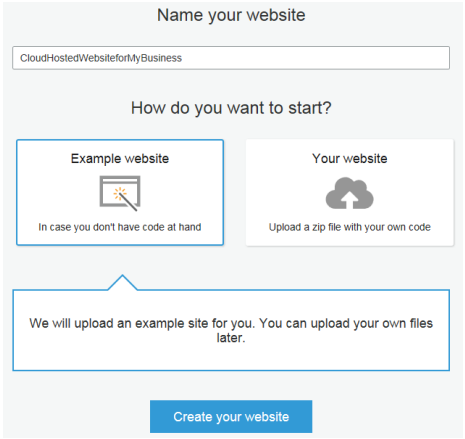
IAAS Application Example with AWS Amazon S3 Instances – Host a static Website

- 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**

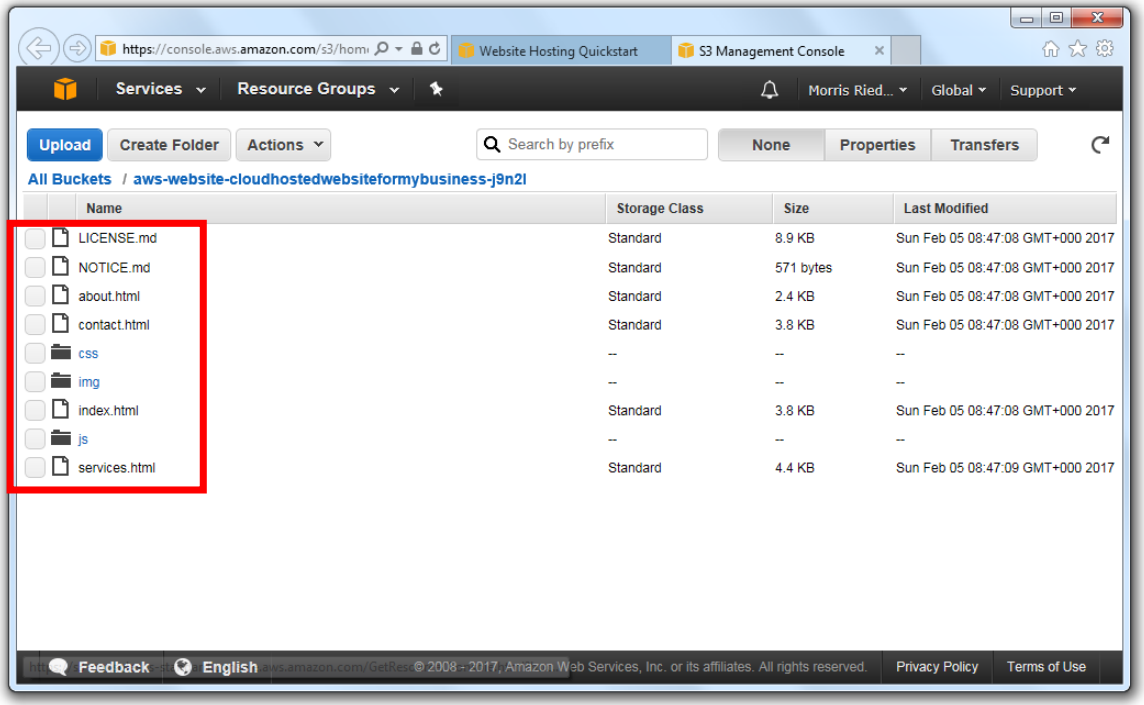




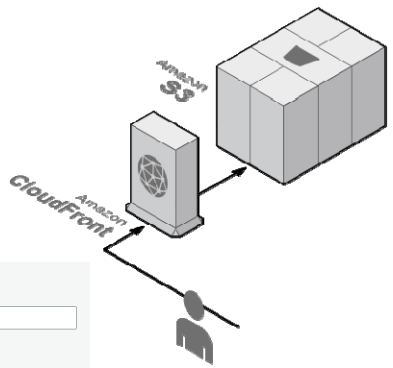
[3] Amazon Web Services

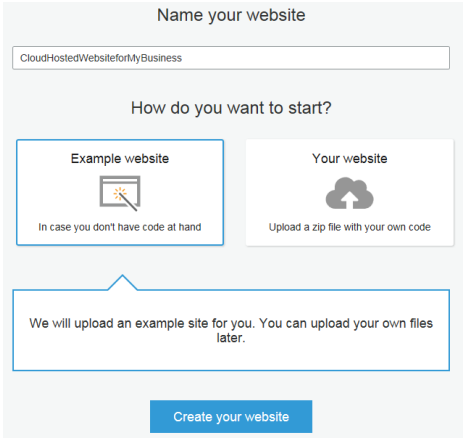
IAAS Application Example with AWS Amazon S3 Instances – Manage Cloud Files

- Using the Web console to upload/download content
  - Static Website example is mainly based on ‘files with content‘



[3] Amazon Web Services





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

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

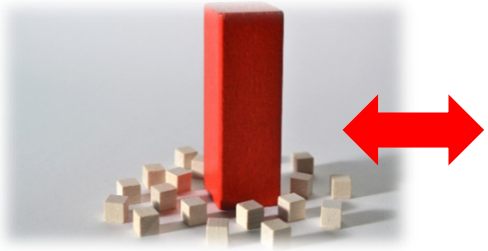
Region: US East (Ohio)

	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)	GET, SELECT, and all other requests (per 1,000 requests)	ifecycle 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	n/a
S3 Intelligent - Tiering	$0.005	$0.0004	$0.01	n/a	n/a
S3 Standard - Infrequent Access*	$0.01	$0.001	$0.01	n/a	$0.01
S3 One Zone - Infrequent Access*	$0.01	$0.001	$0.01	n/a	$0.01
S3 Glacier **	$0.05	$0.0004	$0.05	See below	See below



GET	• Retrieves a resource • Guaranteed not to cause side-effect (SAFE) • Cacheable
POST	• Creates a new resource • Unsafe, effect of this verb isn't defined by HTTP
PUT	• Updates an existing resource • Used for resource creation when client knows URI • Can call N times, same thing will always happen (idempotent)
DELETE	• Removes a resource • Can call N times, same thing will always happen (idempotent)

- **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

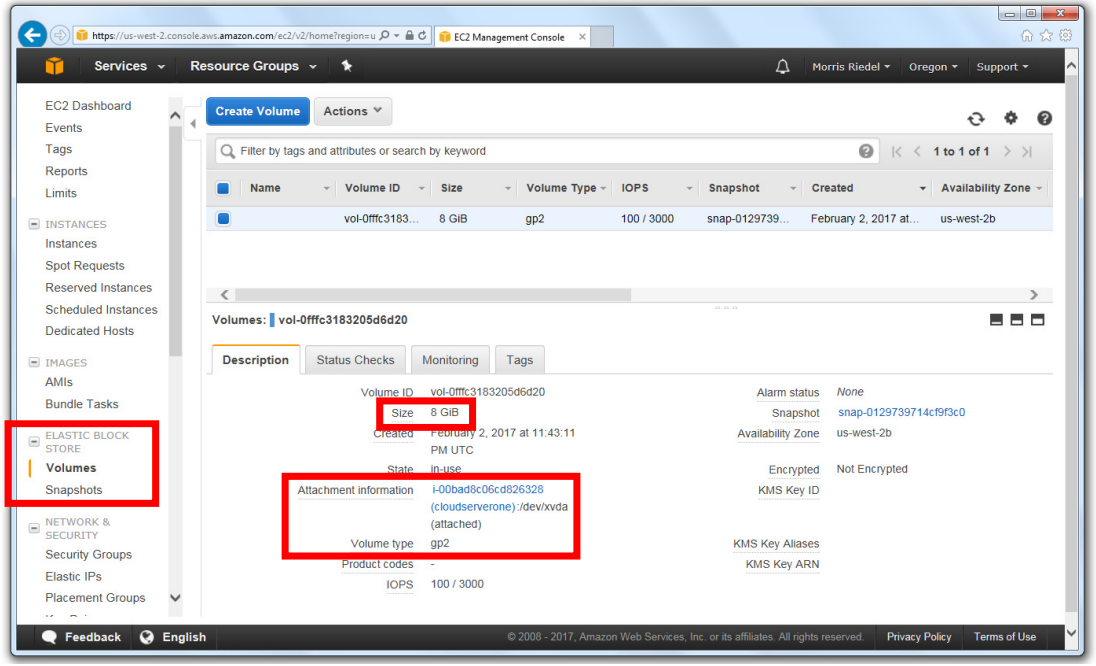
IAAS Example – AWS Amazon Elastic Block Storage (EBS)

- 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

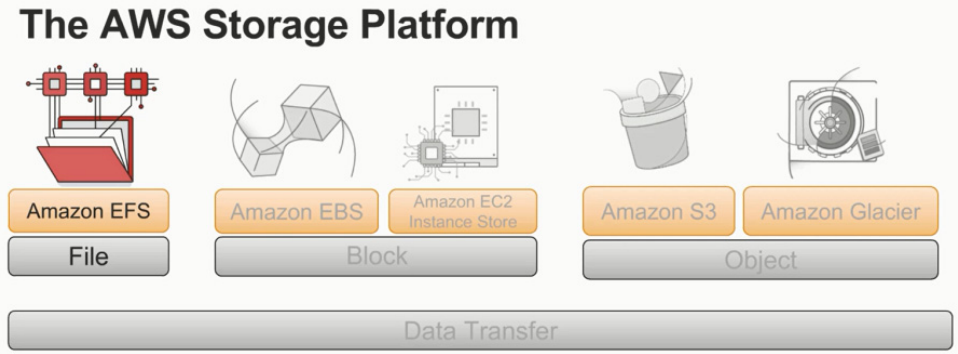


- **Block storage manages data as blocks (compared to object storage with data as objects)**
- **Block storage typically split files into raw blocks of data whereby each block is controlled like a hard drive**
- **Amazon Elastic Block Storage (EBS) provides persistent block storage volume used in conjunction 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 Elastic File System (EFS) is a simple, scalable and reliable elastic file storage sytem and can be used with AWS EC2 instances**
- **AWS EFS storage capacity is elastic, growing and shrinking automatically depending on the usage**

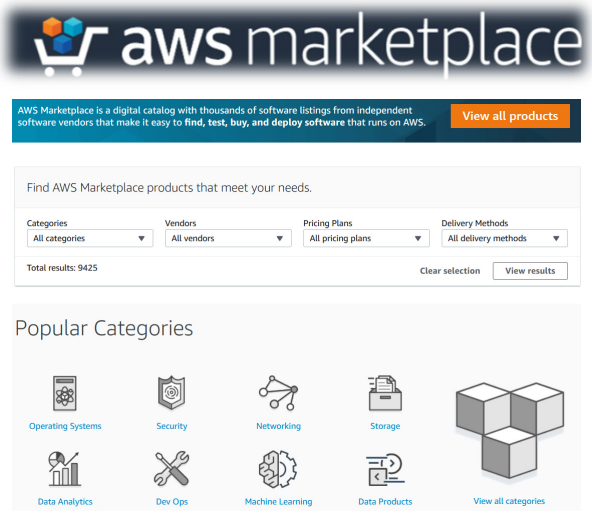
Amazon EFS offers two storage classes: the Standard storage class, and the Infrequent Access storage class (EFS IA). EFS IA provides price/performance that's cost-optimized for files not accessed every day. To load data into EFS IA, simply enable Lifecycle Management for your file system and reduce your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

[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



[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**

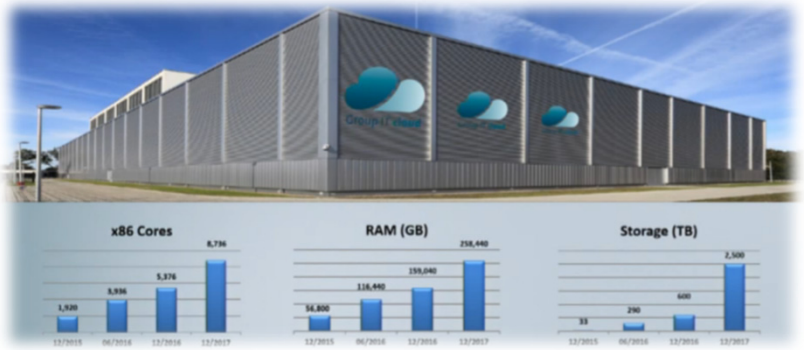
Application Example: OpenStack and IAAS Solutions using Private/Public Clouds

- 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

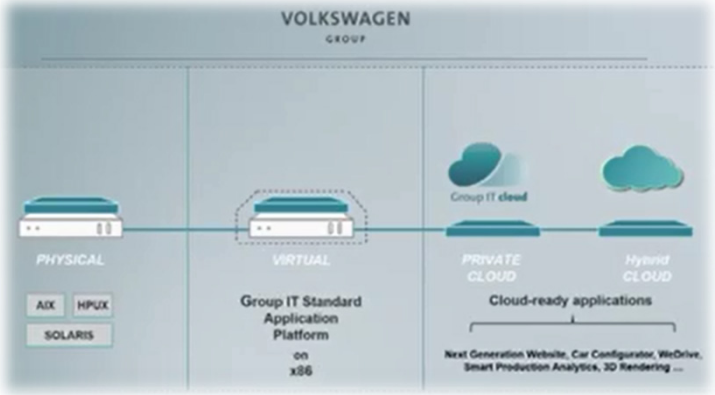
openstack.

[20] OpenStack Web page

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

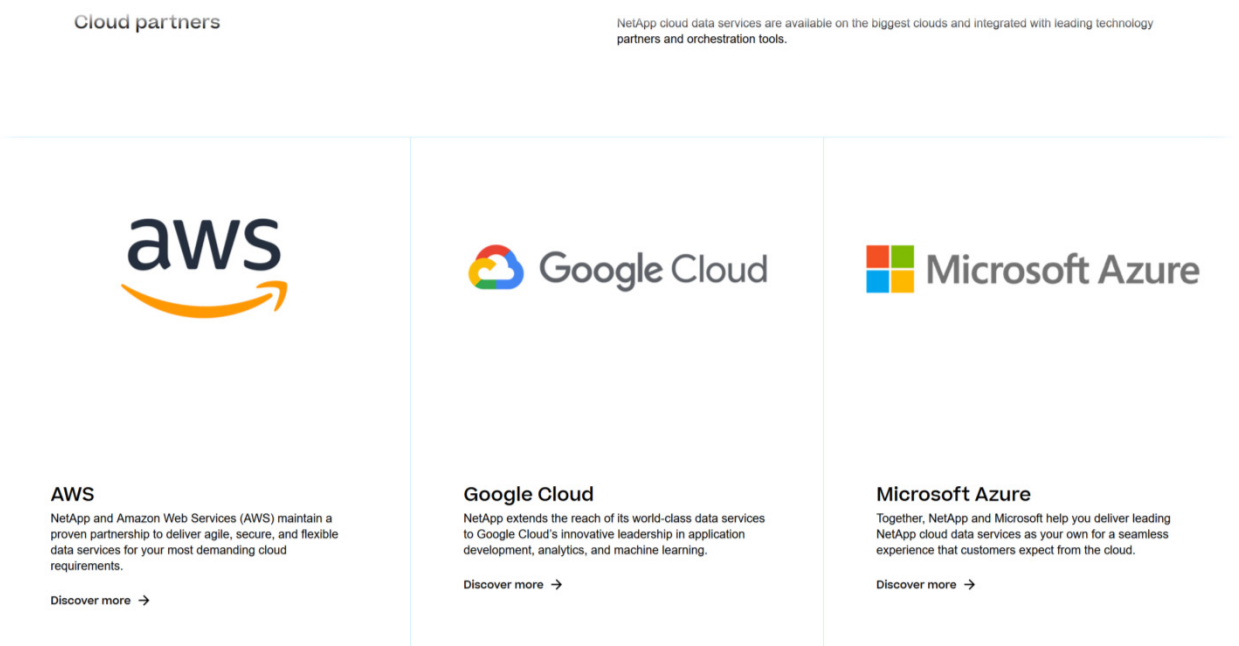


[21] VW & Openstack



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

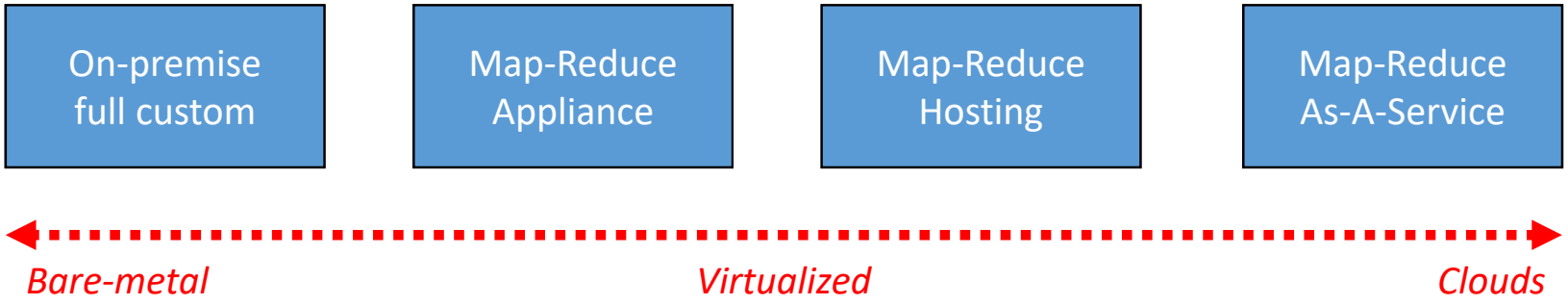
NetApp IAAS Application Example – A Different Cloud Service Provider



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

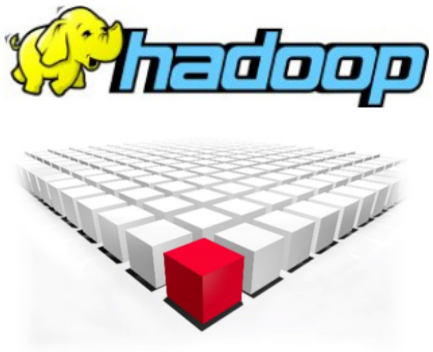
[21] NetApp

IAAS & Map-Reduce Potential Deployment Models Example

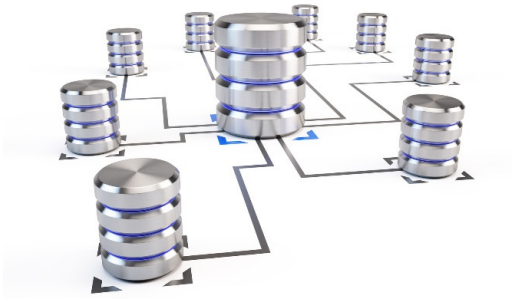


- **'Deployment Models' are different ways of how software can be installed, configured, provisioned, or used by end-users in context of underlying hardware**
- **Map-Reduce deployment models can be categorized in 'on-premise', 'map-reduce appliance, 'map-reduce hosting', or 'map-Reduce-As-A-Service'**

modified from [14] Accenture Deployment Models



- **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**



[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’

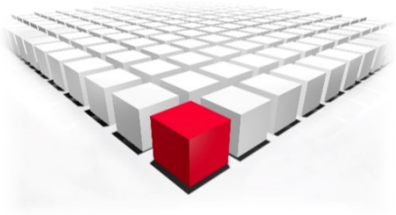
- **Infrastructure in Infrastructure-as-a-Service (IAAS) is differently used per users: for some pure hardware for others deploying platforms: Map-Reduce**
- **The five key-areas that influence a deployment model decision for IAAS solutions are ‘price-performance ratio’, ‘data privacy’,’data gravity’,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

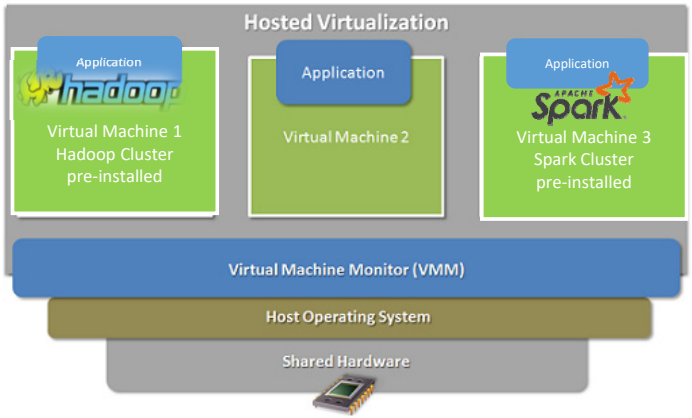
modified from [14] Accenture Deployment Models

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 oepration of a map-reduce cluster
  - Use 'map-reduce hosting' based on a 'Service Level Agreement (SLA)'

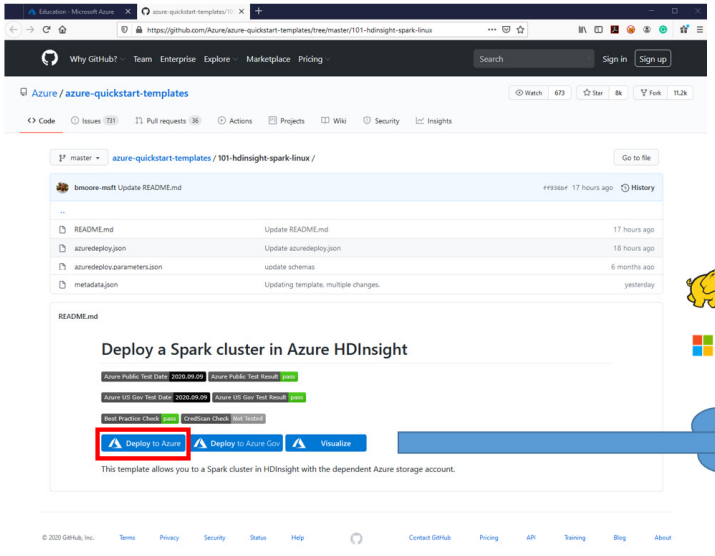
- Example: Small-Medium Enterprise (SME)
  - Offers 'map-reduce hosting'
  - Provides access of fractions of their own Apache Hadoop deployment on a 'pay-per-use' model
  - Has a business model where map-reduce is not the focus, but is used anyway, so idle resources can be 'rented by others'

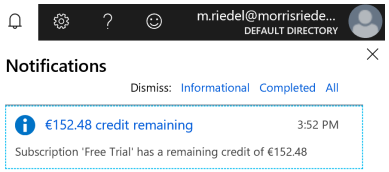
[15] Apache Hadoop Web Page

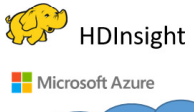
modified from [14] Accenture Deployment Models

Deployment Example using Template & Possibility to Adapt & Automate

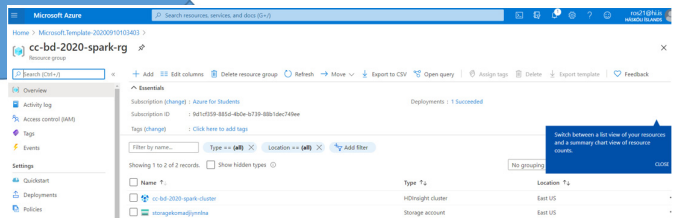
- **Microsoft Azure HDInsight Spark Clusters can be deployed via pre-configured resource manager templates that use Azure computing resources & Azure Storage Blobs as cluster storage**
- **A wide variety of templates are available on Github pages for various general cloud services**



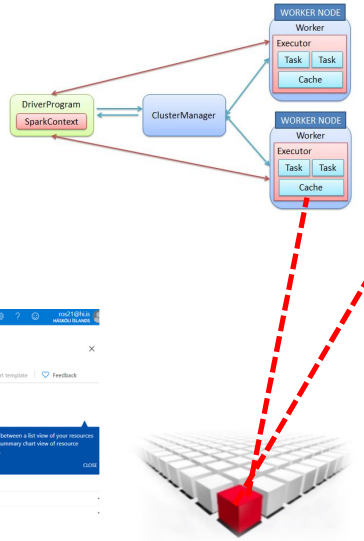




Microsoft Azure



```
"properties": {
  "clusterVersion": "3.6",
  "osType": "Linux",
  "tier": "Standard",
  "clusterDefinition": {
    "kind": "spark",
    "configurations": {
      "gateway": {
        "restAuthCredential.isEnabled": true,
        "restAuthCredential.username": "[parameters('clusterLoginUserName')]",
        "restAuthCredential.password": "[parameters('clusterLoginPassword')]"
      }
    }
  },
```



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."
      }
    },
```

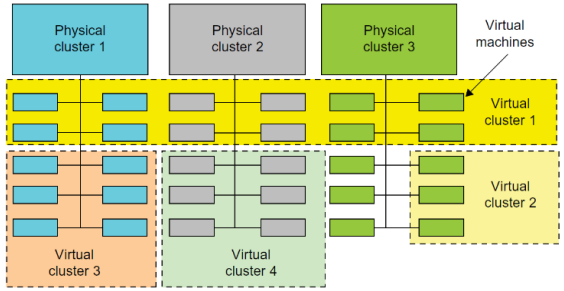
```
"computeProfile": {
  "roles": [
    {
      "name": "headnode",
      "targetInstanceCount": 2,
      "hardwareProfile": {
        "vmSize": "[parameters('HeadNodeVirtualMachineSize')]"
      },
      "osProfile": {
        "linuxOperatingSystemProfile": {
          "username": "[parameters('sshUserName')]",
          "password": "[parameters('sshPassword')]"
        }
      }
    },
    {
      "name": "workernode",
      "targetInstanceCount": 2,
      "hardwareProfile": {
        "vmSize": "[parameters('WorkerNodeVirtualMachineSize')]"
      },
      "osProfile": {
        "linuxOperatingSystemProfile": {
          "username": "[parameters('sshUserName')]",
          "password": "[parameters('sshPassword')]"
        }
      }
    }
  ]
}
```

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

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 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

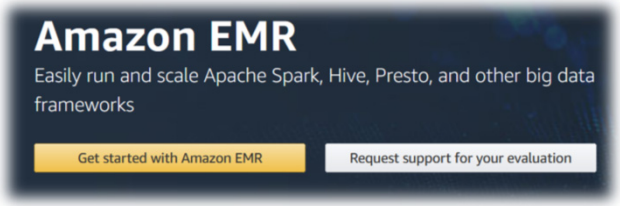
- Multiple analytics products (cf. Lecture 5)
  - Extracting insights and actionable information from data requires technologies like analytics & machine learning
- Analytics Services (cf. Practical Lecture 5)
  - Amazon Athena: Serverless Query Service
  - Amazon ElasticMapReduce (EMR): Hadoop
  - Amazon 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

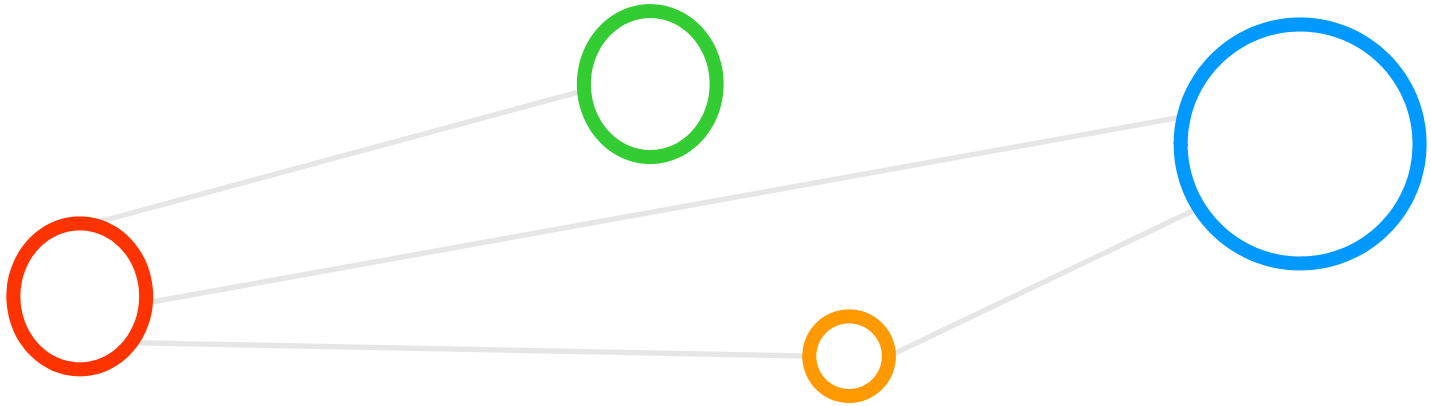
Category	Use cases	AWS service
Analytics	Interactive analytics	Amazon Athena
	Big data processing	Amazon EMR
	Data warehousing	Amazon Redshift
	Real-time analytics	Amazon Kinesis
	Operational analytics	Amazon Elasticsearch Service
	Dashboards and visualizations	Amazon Quicksight



[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

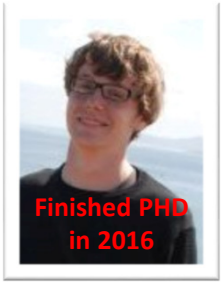
Lecture Bibliography (2)

- [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 Wep 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_pzkDnal4
- [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

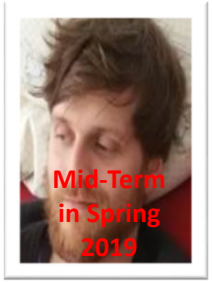
Acknowledgements – High Productivity Data Processing Research Group



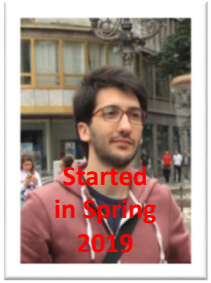
PD Dr.
G. Cavallaro

Senior PhD
Student A.S. Memon

Senior PhD
Student M.S. Memon



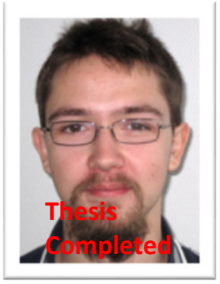
PhD Student
E. Erlingsson



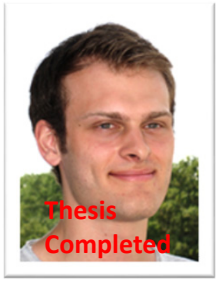
PhD Student
S. Bakarat

PhD Student
R. Sedona

Dr. M. Goetz
(now KIT)



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)

services
measurement
within
fusion
new
funding
Policy-based
concepts
device
Analysis
forms
cross-disciplinary
resources
Centre
climate
Computer
expertise
Approaches
construction
Fusion
year
like
Computational
Test
E.g
Computing
disciplines
results
Enable
different
Cross-Disciplinary
methods
High
big
nodes
PB
community
Environment
Modelling
EU
Science
centers
Key
scientific
important
files
etc
DKRZ
Use
storage
basis
Infrastructure
SMAQ
international
model
simulations
ENES
technologies
engineering
References
formats
brain
increasing
via
EUDAT
physical
IPCC
Big
Project
Data
processing
Common
TB
MapReduce
create
Methods
computing
Using
European
databases
GB
Supercomputing
Work
images
codes
analysis
Large
Summary
data-intensive
Service
Scientific
stored
Resources
analyze
levels
often
DLCL
Level
external
JSC
use
computational
Research
Matters
Concrete
performance
research
e.g
science
HPC
Juelich
datasets
acquisition
large-scale
compute
system
Landscape
management
System
Climate
modelling
AR5
Hardware
distributed
access
hundreds
Nutshell
Services
DLCLs
Understanding
structures
Technologies
Earth
cores
Structure
current
network
Germany
Simulation
Processing
directory
project
General
Provide
NASA
Energy
systems
day
Health