Big Data, Techniques and Platforms
Class 3/8 : AWS

Francesca Bugiotti

CentraleSupélec

October 17, 2023

Objectives

- Describe AWS
- Identify some similarities between cloud platforms

Plan

Who is Amazon?

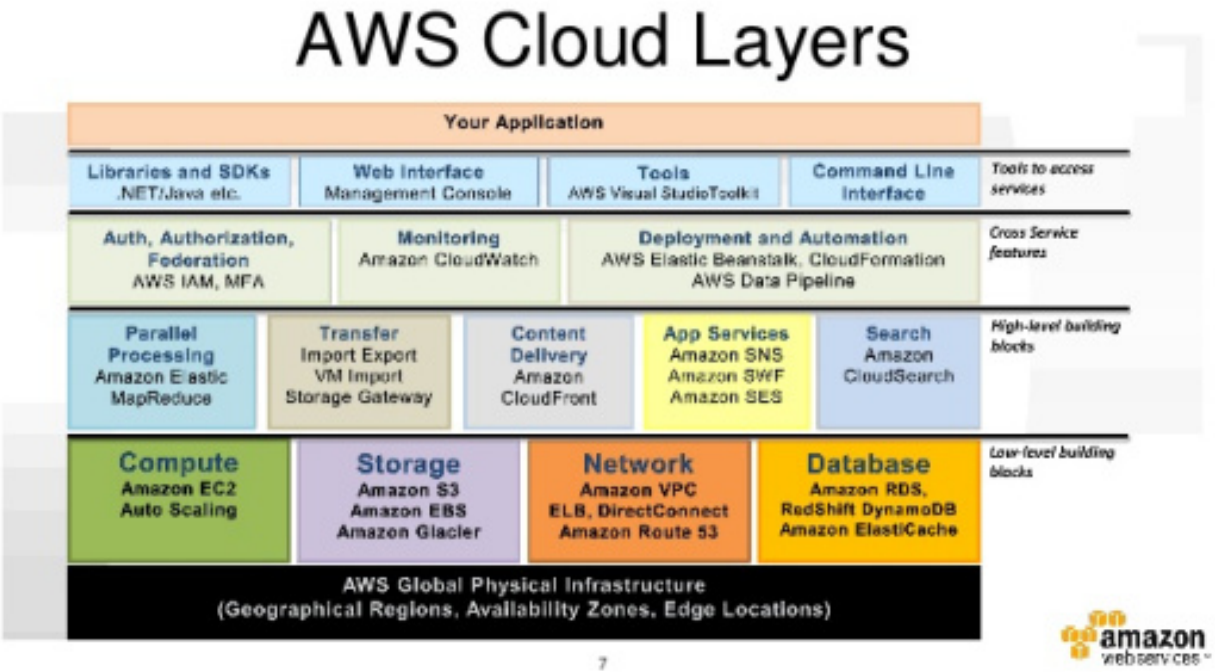
Amazon.com is a multinational company based in Seattle-Washington

- E-commerce
- Cloud computing
- Digital streaming
- Artificial intelligence

History

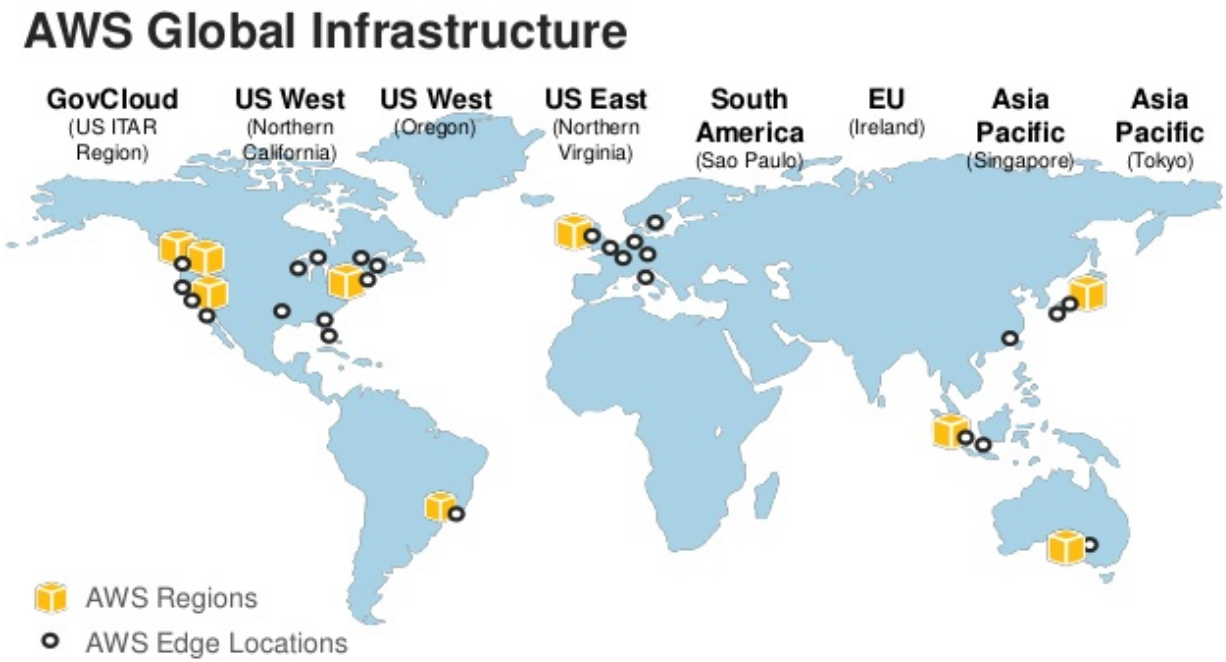
- July 5, 1994: Founded by Jeff Bezos in Bellevue, Washington
- At the beginning "just" an online book-store
- Quickly enriched by new items (electronics, furniture, food, toys, and jewelry)
- 2008: starts producing Kindle Fire tablet computers
- 2006: Amazon launched Amazon Web Services (AWS) for providing cloud computing service

Looking at AWS



Plan

AWS Regions


AWS Global Infrastructure
GovCloud
(US ITAR Region)
US West
(Northern California)
US West
(Oregon)
US East
(Northern Virginia)
South America
(Sao Paulo)
EU
(Ireland)
Asia Pacific
(Singapore)
Asia Pacific
(Tokyo)
AWS Regions
AWS Edge Locations

The cluster

Main components

- Availability Zone (AZ)
- AWS Region
- AWS Edge

Availability Zone

Each group of logical data centers is an Availability Zone

- Independent power
- Cooling
- Physical security
- Redundant and ultra-low-latency network connections

Region

A Region is a physical location around the world where data centers are clustered

- Multiple, isolated, and physically separate AZ's within a geographic area
- North America, South America, Europe, China, Asia Pacific, South Africa, and the Middle East

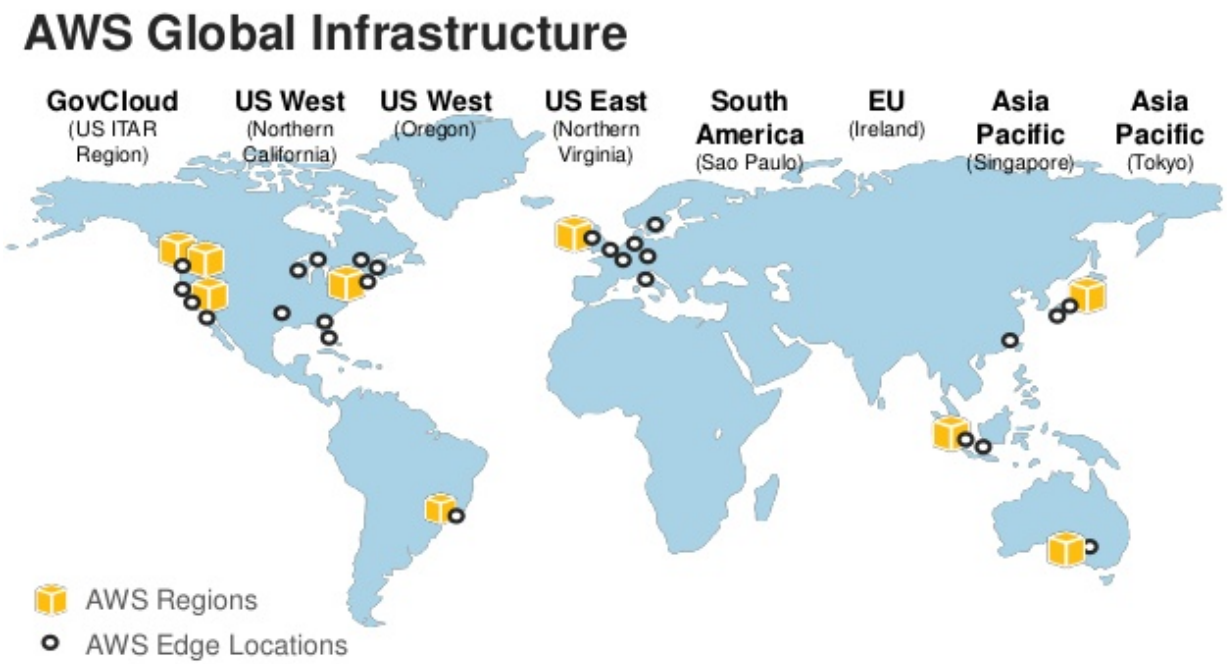
Edge

An edge location is where end users access services located at AWS

- Provides low latency connectivity
- Guarantees that static contents are made available from nearest location of the request
- Automatic routing to the nearest edge location

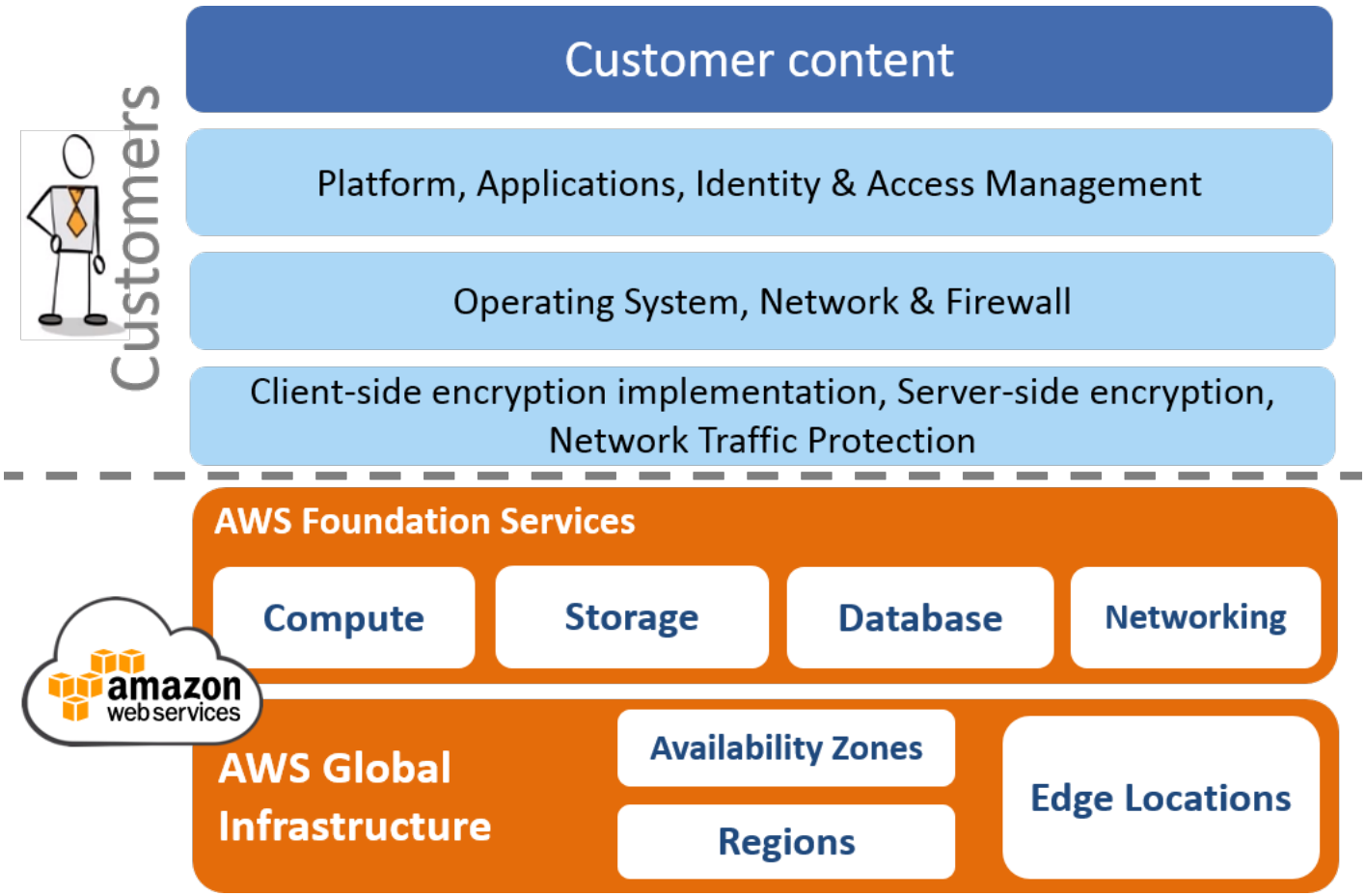
Plan

AWS Services



AWS Services

AWS Services



Elastic Compute Cloud

Elastic Compute Cloud

- Resizable compute capacity in the cloud
- Virtual Machines on-demand
- Auto-scaling the capacity up during demand spikes
- Auto-scaling the capacity down to minimize cost
- Failure resilient application (instances in a separate Availability Zone)
- VM import/export

Elastic Block Store

Elastic Block Store

Provides block level storage volumes (1GB-1TB) for use with Amazon EC2 instances

- multiple volumes on the same instances
- independent-persistence
- network connection
- they behave as hard drives

- they are in a Availability Zone but can be attached to other zones
- automatic replication in the same zone
- snapshots that are persisted
- copy across region

Amazon Simple Storage Service

Web services interface to store and retrieve

Slogan

- any amount of data
- at any time
- from anywhere

- write, read, and delete operations of objects
- storage in buckets
- a bucket can be stored in several regions

Plan

Elastic Map-Reduce

A Web Service that makes it easy to quicly and cost-effectivelu process vast amount of data using Hadoop

How?

- data and processes are distributed across EC2
- usage of a persistent cluster or a temporary cluster
- elastic allocation of EC2 resources

Amazon Relational Database Service

A web service that makes it easy to set up, operate, and scale a relational database in the cloud.

- Access to the capabilities of a familiar MySQL, Oracle or Microsoft SQL Server database engine
- Code, applications, and tools already used with existing databases can be used with RDS.
- Automatic backs up of the database,
- Transparent storage of the backups (for a user-defined retention period)

Amazon Relational Database Service

Amazon RDS provides scaling the compute resources or storage capacity associated with the Database Instance.

Challenge

- Database characteristics
- Cloud capabilities

Amazon DynamoDB

DynamoDB is a fast, fully managed NoSQL database service

Characteristics

- cost-effective storage and retrieving of any amount of data
- capacity to serve any level of request traffic.
- Solid State Drives (SSDs)
- Replication across 3 Availability Zones
- Flexible schema (data organization)

Amazon Cloud Watch

Amazon CloudWatch provides monitoring for AWS cloud resources and the applications customers run on AWS. can also supply their own custom application and system metrics, such as memory usage, transaction volumes, or error rates,.

Characteristics

- Makes easy to retrieve your monitoring data
- Simplifies to take actions based on the state of your cloud environment.
- Provides a monitoring of your AWS resources up-to-the-minute in real-time
  - Amazon EC2 instances
  - Amazon EBS volumes,
  - Elastic Load Balancers,
  - Amazon RDS DB instances.

AWS

Amazon Web Services: Overview

Summary

Launched: 2006

Parent Company: Amazon

Revenue (Q4 2018): US$ 7.3 bil.

Market Share: 32.3%

General assessment

Strengths

- Dominant market position due to first-mover status
- Extensive, mature offerings
- Highest reliability and flexibility of infrastructure services

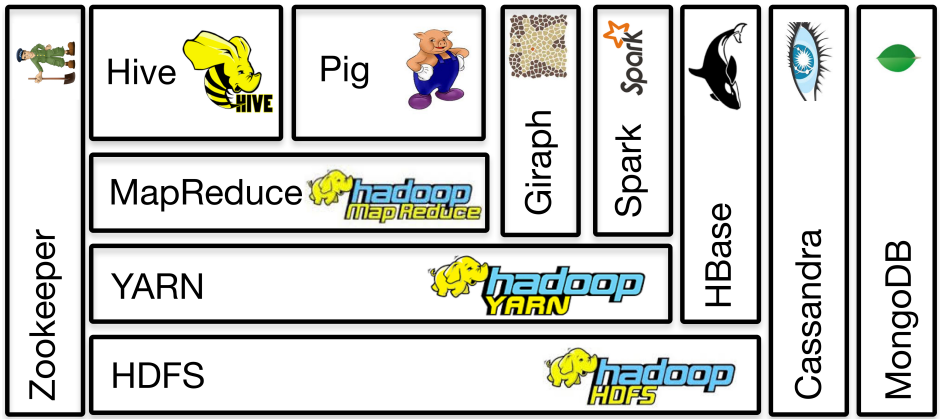
Cautions

- The big breadth and depth of services can sometimes be difficult to navigate
- No strong frontend solution

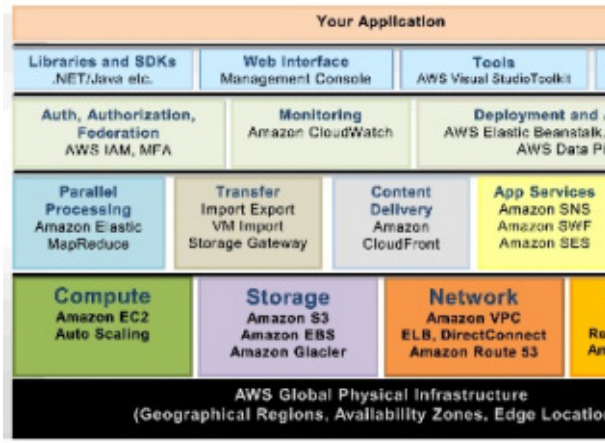
Deployment

AWS has been the **first-mover in the cloud market** and has always been a cloud-first company, not constructed for on-premises implementation. However, *AWS Outposts* (announced in Nov. 2018) is a new development for the company and is meant to **bring any AWS service**, infrastructure, and operating model to data centers, co-location space or **on-premises facility**

Parallel analysis



xmm=-1000pt =



Microsoft Azure:
Overview

Summary

Microsoft Azure

Launched: 2010

Parent Company: Microsoft

Revenue (Q4 2018): US$ 3.7 bil.

Market Share: 16.5%

General assessment

Strengths

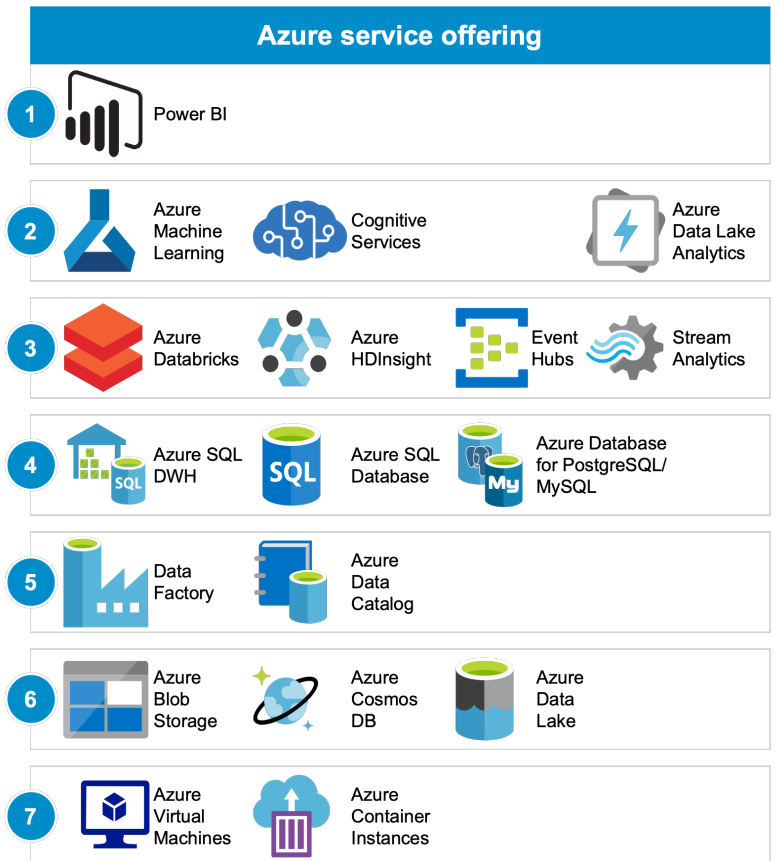
- Second largest provider with Broad feature offering
- Easy integration with Microsoft software
- Strengths in on-premises and hybrid cloud delivery

Cautions

- …

Deployment

Although it has been running longer as a business, under current CEO Nadella the whole company performed a **strong turn towards Cloud offerings**. At the same time, its **core on-premises product Microsoft SQL Server 2019 remains one of the core products** of Microsoft's offering and the query service "Polybase" allows for connectivity to other data sources via virtual queries. With the mix of experience in both fields, the Microsoft **hybrid cloud offering has developed into a strong USP** with extensive experience handling enterprise needs.



IBM Cloud:
Overview

Summary

IBM Cloud

Launched:

Parent Company: IBM

Revenue (Q4 2018): US$ 0.8 bil.

Market Share: 3.6%

General assessment

Strengths

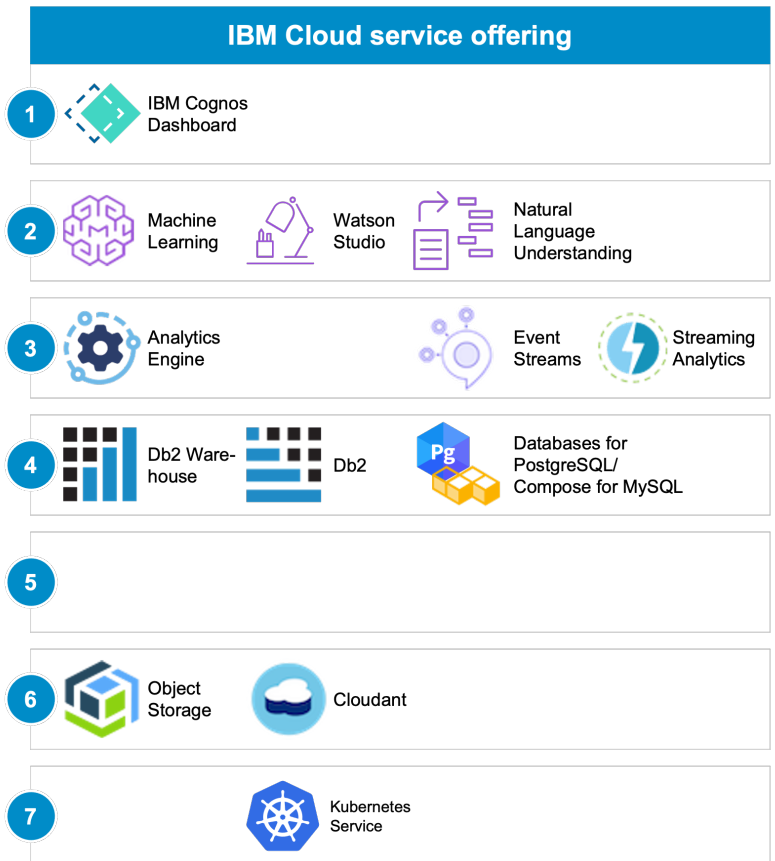
- Integration with IBM companies (“power of incumbency”*)
- Longstanding investments in Watson’s AI
- Brought the mainframe to the cloud

Cautions

- Losing market share to the big three
- Profit from Red Hat deal unclear

Deployment

IBM has been **later to the Cloud game** than the big providers, but is **looking to catch up**. Part of this strategy is the acquisition of Red Hat for a foot in the **container deployment** world, as well as a **strong migration service** from on-premises IBM solutions to cloud solutions. What’s more, IBM also offers services to move and **modernize SAP and Oracle apps and run them in production in the IBM cloud**.



Google Cloud: Overview

Summary

Google Cloud

Launched: 2011

Parent Company: Alphabet

Revenue (Q4 2018): US$ 2.2 bil.

Market Share: 9.5%

General assessment

Strengths

- Designed for cloud-native businesses, but recent move towards hybrid as well
- Commitment to open source (e.g. TensorFlow)
- Strong AI experience and innovativeness

Cautions

- Traditionally less strong positioning in the enterprise market

Deployment

Google has traditionally been recognized as a internet-first actor, **promoting its other cloud services alongside a cheap cloud-storage offering**. In April 2019 they released they **hybrid service "Anthos"** (formerly announced as "Cloud Services Platform"), offering an integrated set of cloud services (e.g. ML applications) that can also be **deployed on-premises** and **integrates with AWS, Azure** and other providers. Other than that, no on-premises strategy is in place.