

Why learn about AWS as a data practitioner?

AWS CLOUD CONCEPTS



Hatim Khouzaimi

Data Scientist

Motivation

- The benefits of learning about cloud computing
- The advantages of choosing AWS

A central technology for the data industry

\$370B

worth industry

A central technology for the data industry

\$370B

worth industry

50%

of all corporate
data

A central technology for the data industry

\$370B **50%** **18.4%**

worth industry of all corporate data more spendings in 2021

Technical advantages



- Technical advantages for advanced machine learning (e.g. deep learning):
 - More data storage
 - More computational power
 - More reliability
- Reduced time to production

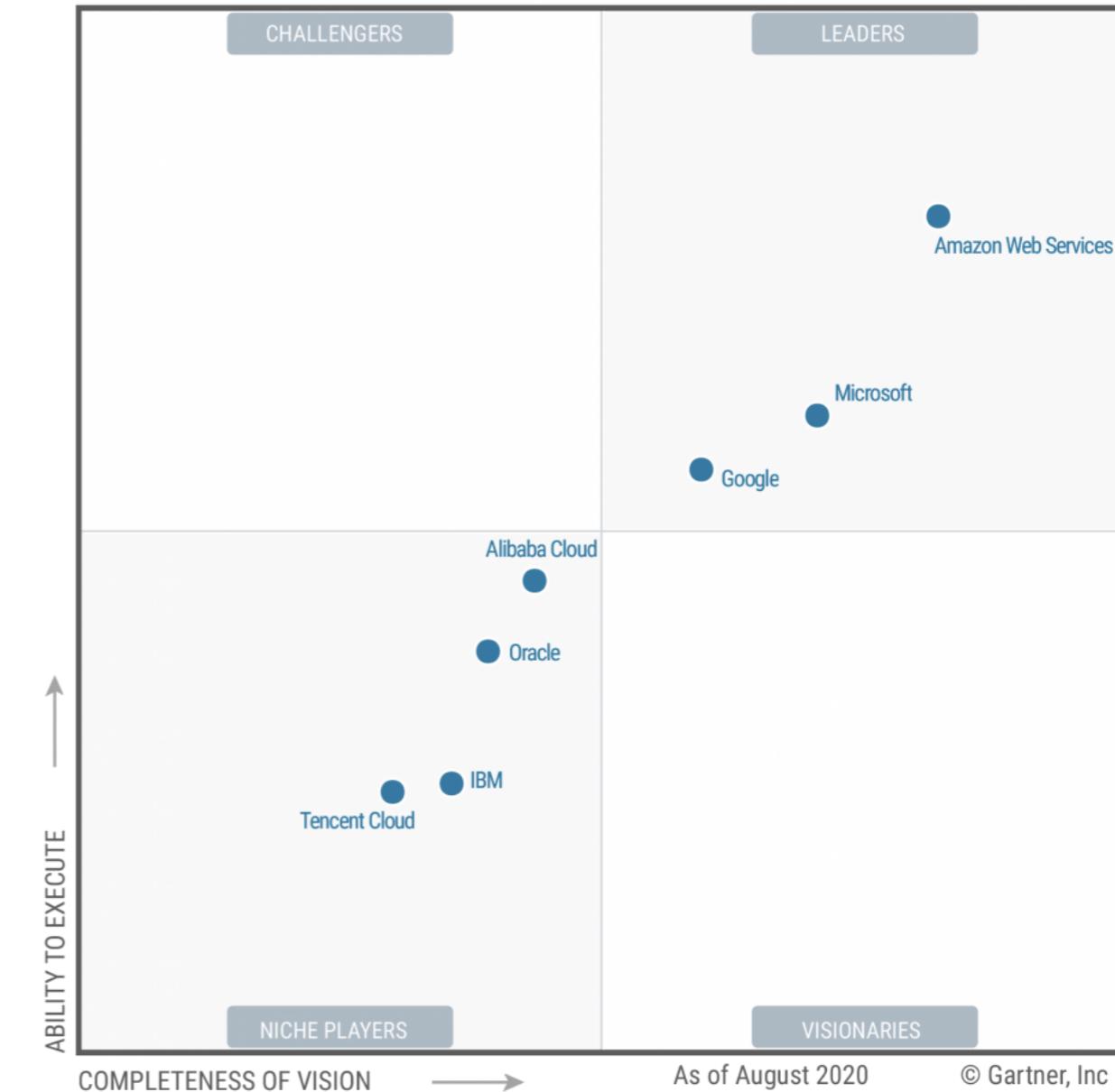
Cloud machine learning services

- Integrated machine learning services, for example:
 - AWS SageMaker
 - Microsoft Azure Machine Learning Studio
 - Google Cloud Datalab
- Native integration to other cloud services, for example:
 - Data storage
 - Security
- Dedicated modules for every step in the Machine Learning process

AWS: Leading cloud provider on the market

- Pioneer: first cloud provider (in the modern sense)
- Recognized as a leader by Gartner in their Cloud Magic Quadrant, for the 10th consecutive year!
- \$13.5B revenue in Q1 2021
- 32% market share, followed by:
 - Microsoft Azure (19%)
 - Google Cloud Platform (7%)

Figure 1. Magic Quadrant for Cloud Infrastructure and Platform Services

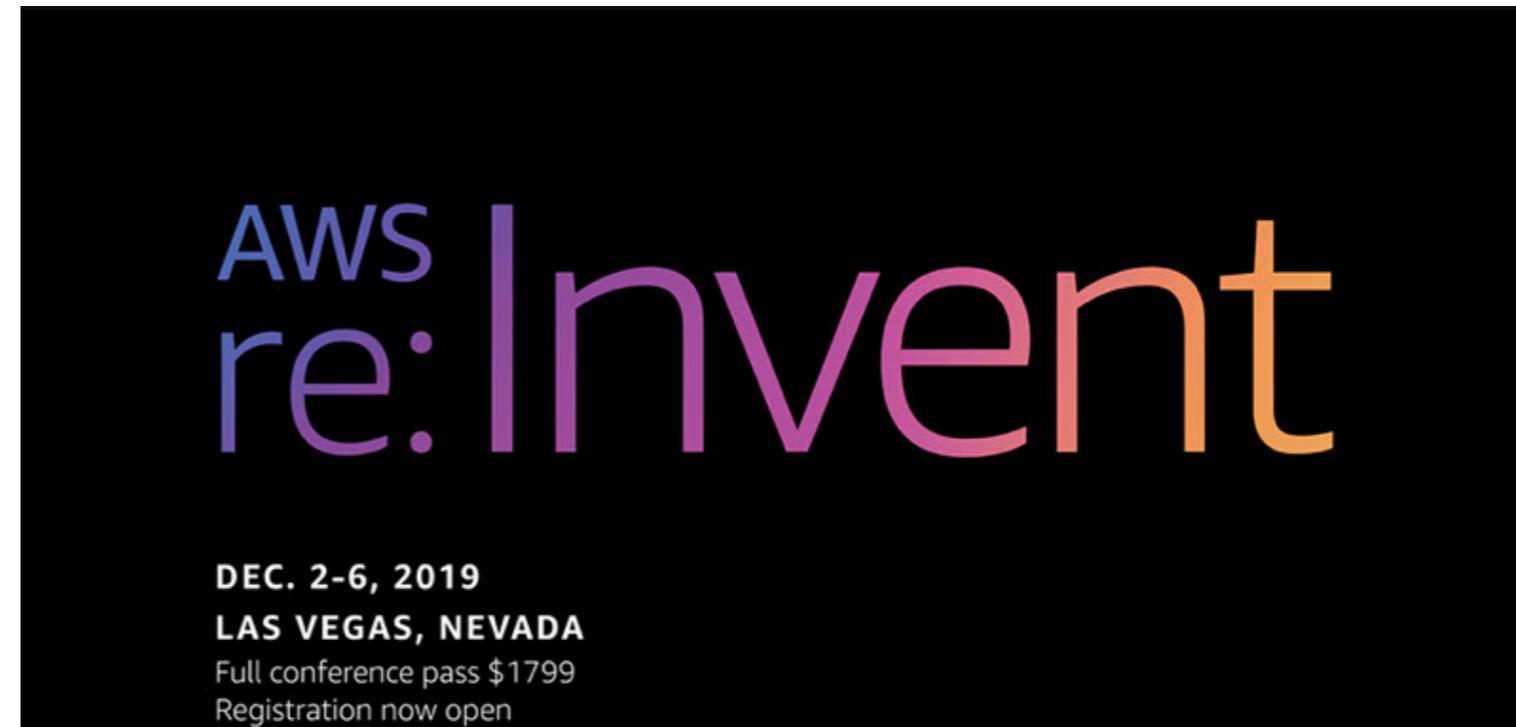


200+ services

- More than 200 AWS services
- Useful offer for data practitioners:
 - **S3:** Ideal for data lakes
 - **SageMaker:** Full Machine Learning toolbox
 - **Kinesis:** Data streams and firehoses
 - etc

AWS community and events

- Large community:
 - Developers
 - Architects
 - System Administrators
 - Data Scientists
 - etc
- Events organized:
 - AWS Summit
 - re:Invent
 - etc



Let's practice!

AWS CLOUD CONCEPTS

AWS cloud offer and history

AWS CLOUD CONCEPTS



Hatim Khouzaimi

Data Scientist

Definition



- Amazon Web Services
- Cloud platform
- Offering more than 200 services:
 - Compute instances
 - Storage
 - Databases
 - Machine Learning
 - AR/VR
 - etc

AWS global infrastructure

- Data centers deployed all over the world
- Region:
 - [...] physical location around the world where we cluster data centers.
aws.amazon.com
- Availability zone (AZ):
 - "one or more discrete data centers with redundant power, networking, and connectivity in an AWS Region."
aws.amazon.com
- **2021:** 25 regions, 81 AZs, 245 countries and territories



The founding of AWS

- **Early 2000:** the need for a common compute, storage and database infrastructure to serve Amazon's internal needs
- **2003:** the idea of a cloud computing platform (Andy Jassy and Jeff Bezos)
- **2006:** the launch of AWS with Andy Jassy at its head
- **2020:** market leader with \$45.3B in revenue



Andy Jassy, founder and former CEO of AWS, replaced Jeff Bezos as CEO of Amazon in July 2021

A growing offer of services

- 2006: S3 (storage), SQS (messaging), EC2 (compute)
- 2009: RDS (relational databases)
- 2010: Route 53 (domain name service)
- 2011: IAM (identity management)
- 2012: DynamoDB (non-relational databases)
- 2013: Kinesis (data streaming)
- 2014: Lambda (serverless compute)
- 2015: Mobile Hub (mobile apps testing and monitoring), IoT (internet of things)
- 2017: SageMaker (machine learning), Sumerian (AR/VR)
- 2020: Braket (quantum computing)

A growing community

- **2007:** Startup challenge
- **2012:** First re:Invent conference
- **2013:** Certification program
- **2014:** AWS Partner Network (APN) and first pop-up loft in San Francisco

Let's practice!

AWS CLOUD CONCEPTS

Getting started with AWS

AWS CLOUD CONCEPTS



Hatim Khouzaimi

Data Scientist

AWS accounts

- **Linked to:**
 - Email address
 - Payment card
- **Account access:**
 - Root access
 - Limited access through AWS IAM
- **Centralized access point:**
 - AWS Organizations
 - Consolidated billing

How to access an AWS account?

- **Console**
- **CLI:** Command Line Interface
- **SDKs:** Software Development Kits
- **API:** Application Programming Interface

The AWS console

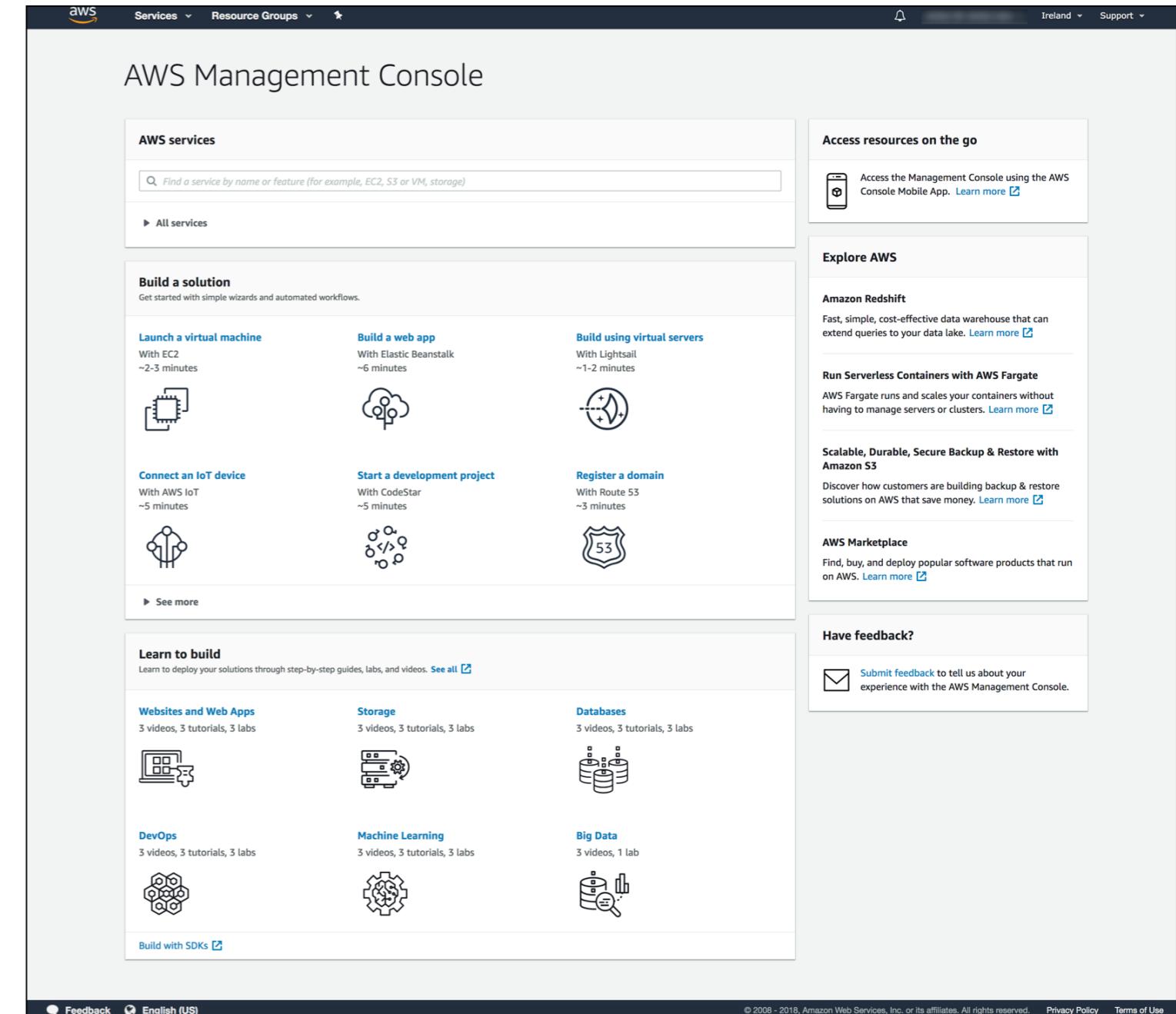
- AWS graphical user interface (GUI)

Advantages

- Simple and intuitive for resource:
 - Creation / Deletion
 - Monitoring (parameters, graphs, etc)
- Direct links to documentation:
 - Recommended when learning about AWS or new services

Disadvantages

- Hard to track resource creation/management actions



The AWS CLI

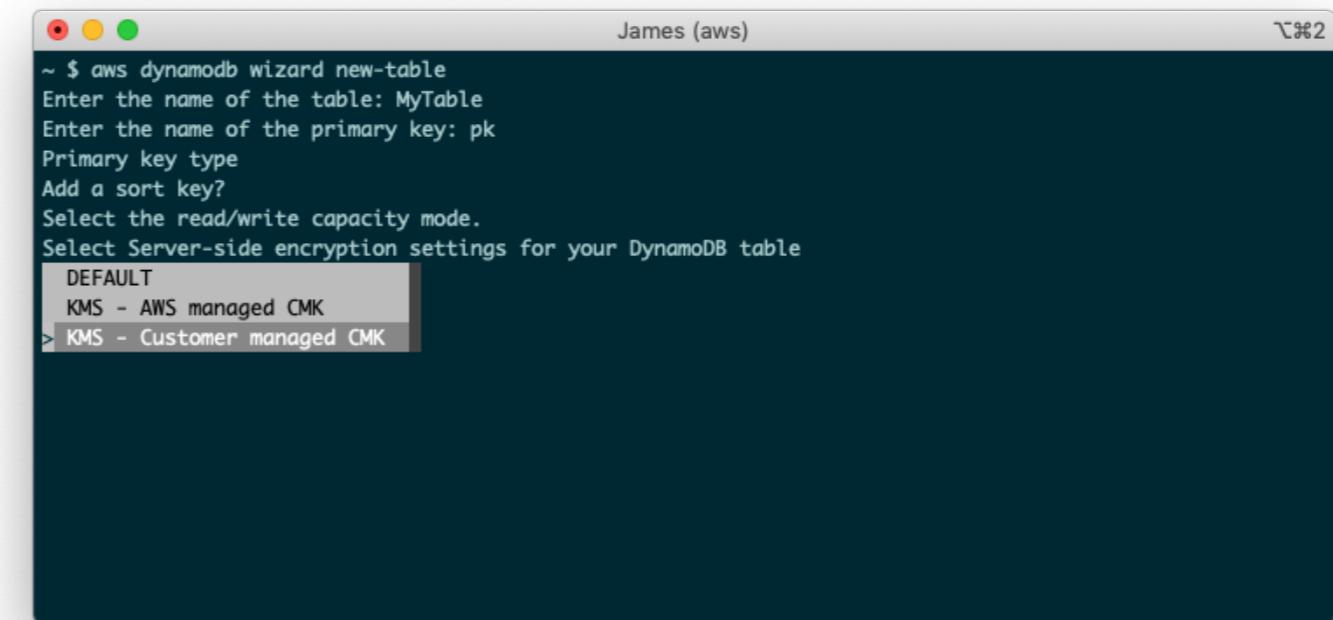
- Interact with AWS accounts using a command line terminal

Advantages

- Handy for administrators to:
 - Automate frequent tasks using scripts
 - Execute large numbers of tasks
- Open source

Disadvantages

- Less visual and intuitive than the console



```
~ $ aws dynamodb wizard new-table
Enter the name of the table: MyTable
Enter the name of the primary key: pk
Primary key type
Add a sort key?
Select the read/write capacity mode.
Select Server-side encryption settings for your DynamoDB table
  DEFAULT
  KMS - AWS managed CMK
> KMS - Customer managed CMK
```

The AWS SDKs

- Specific to a particular programming language:



Advantages

- Simple library to import
- Handy for developers
 - Seamlessly interact with an AWS account inside their code
- Open source

Disadvantages

- Need to know programming

The AWS API

- HTTP request through the Internet
- Most low-level access method
- Building block for AWS SDKs, CLI and console

Advantages

- Maximum flexibility

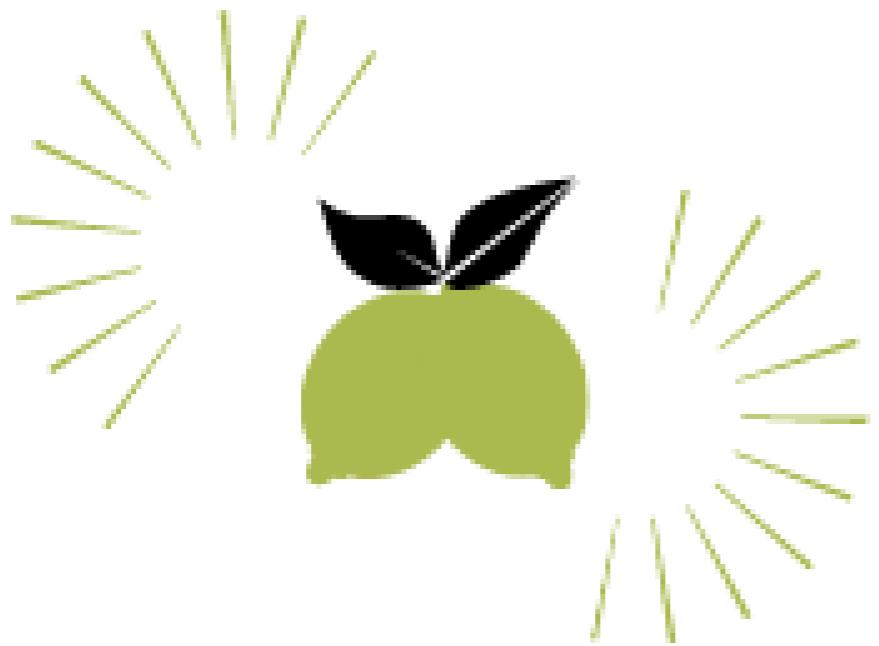
Disadvantages

- Less "human-readable"



¹ Clouds vector created by vectorjuice (<https://www.freepik.com/vectors/clouds>)

MeetFoodies Inc.



MeetFoodies
COMMUNITY

- Mock use case for the course
- Foodies community mobile application
- Recipes sharing
- Restaurant/shops recommendations
- Events
- Contests
- Totally on-premise
- Planning to migrate its infrastructure to AWS

Let's practice!

AWS CLOUD CONCEPTS