NÁSKOK
DÍKY
ZNALOSTEM

PROFINIT

Cloud

Profinit Big Data tým

2020

Osnova

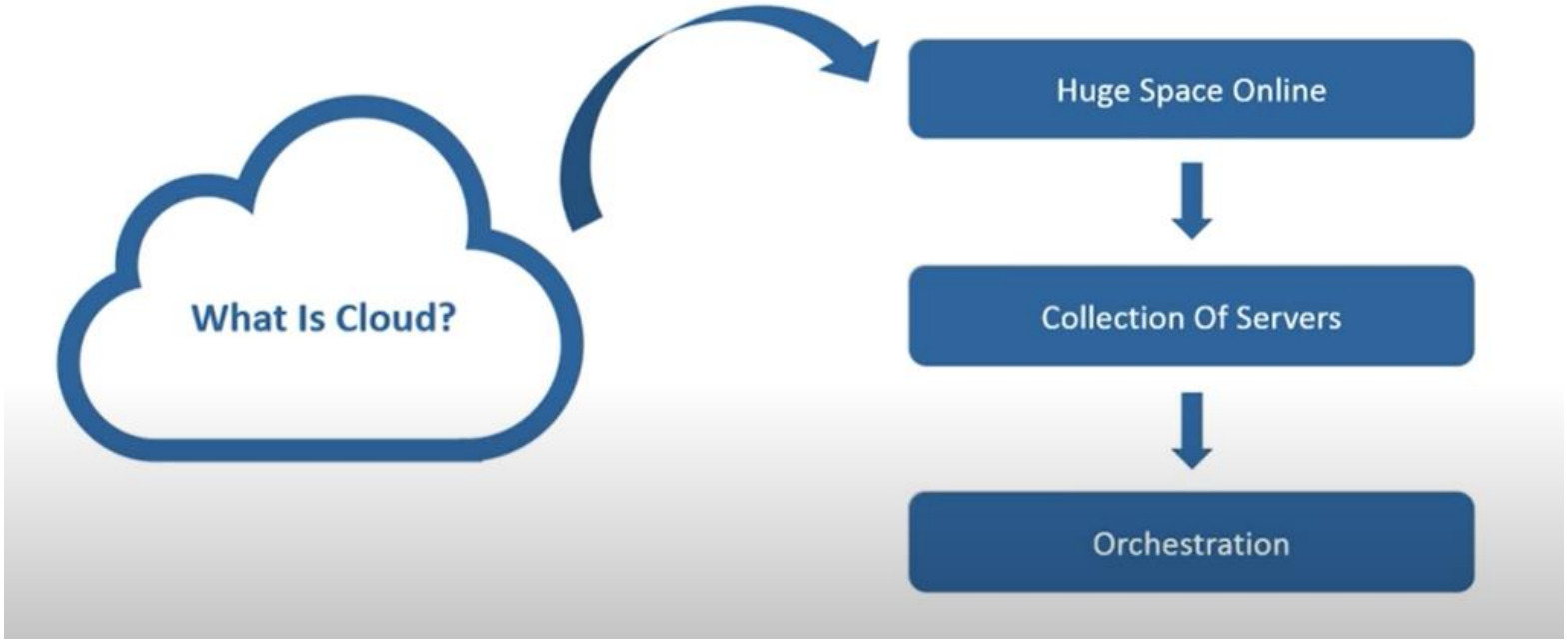
- Co je Cloud?
- Azure
- AWS

PROFINIT

Cloud


ACCORDING TO THIS, THE PLANET EARTH WAS ONCE POPULATED BY HUMANS, THEN IN 2020...
...THEY ALL MOVED TO THE CLOUD.

PROFINIT


What Is Cloud?
Huge Space Online
Collection Of Servers
Orchestration

PROFINIT

Cloud

- Co je vlastně cloud?
  - v podstatě neomezený prostor online,
  - kolekce nejrůznějších serverů,
  - orchestrace prostoru, výkonu.

  To všechno se dá pronajímat a platit pouze za použité prostředky -> často levnější než fyzický hardware.

- Proč cloud použít pro BigData?
  - Škálovatelnost výkonu i úložiště.
  - Predikce nákladů.
  - Možnost volby vhodné architektury a k tomu vhodný HW na pár kliknutí ☺

PROFINIT

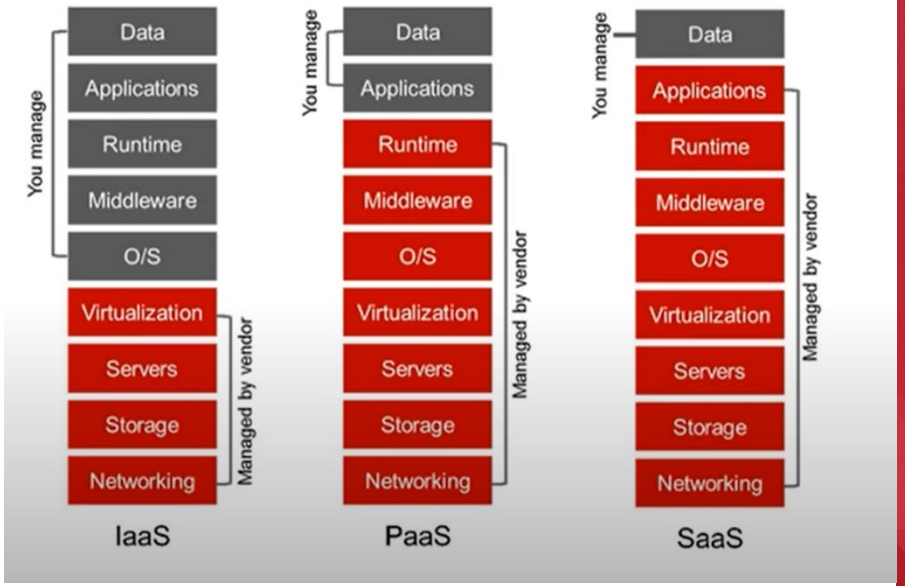
Cloud service poskytovatelé

- Google, IBM, Amazon, Microsoft, …
- Aktuálně AWS patří k největším poskytovatelům.

PROFINIT

Cloud services

- Co to je?
  - Cloudová služba, kterou poskytuje cloudový poskytovatel zákazníkovi přes internet.
- 3 skupiny
  - Různá úroveň řízení uživatelem a poskytovatelem
  - IaaS – Infrastucture as a Service
  - PaaS – Platform as a Service
  - SaaS – Software as a Service



PROFINIT

AWS

Proč je AWS v kurzu?

- Jednoduchá registrace.
- Mikroplatby (platí se za Gigabyte, hodinu užívání).
- Reportování útrat.
- Stabilní – za posledních 7 let pouze několik výpadků a to pouze v několika regionech (pouze několik hodin)
- Důvěryhodná značka.

PROFINIT


US Gov Iowa
Central US
Central US 2
Oregon
N. California
AWS GovCloud
West US
South Central US
Canada Central
Canada East
North Central US
N. Virginia
East US
East US 2
US Gov Virginia
Ireland
UK West
West Europe
North Europe
Frankfurt
UK South
China North
Beijing
Seoul
Japan East
Tokyo
Japan West
China East
East Asia
West India
Central India
Mumbai
South India
Southeast Asia
Singapore
São Paulo
Brasil South
Australia East
Sydney
Australia Southeast
Microsoft Azure
amazon
web services
Atomia

AWS Big Data Portfolio

Collect

Real Time

Amazon Kinesis Firehose

Analyze RT Stream

Amazon Kinesis Analytics

AWS Direct Connect

Amazon Snowball

Store RT in stream

Amazon Kinesis Streams

Store

Object Storage

Amazon S3

Archive

Amazon Glacier

Amazon CloudSearch

RDBMS

Amazon RDS, Amazon Aurora

NoSQL DB

Amazon Dynamo DB

Analyze and store

Amazon Elasticsearch Service

Analyze

Hadoop

Amazon EMR

Amazon EC2

Data Warehouse

Amazon Redshift

Amazon Machine Learning

BI

Amazon QuickSight

Interactive

Athena

AWS Database Migration Service

AWS Data Pipeline

AWS Glue

Přehled služeb

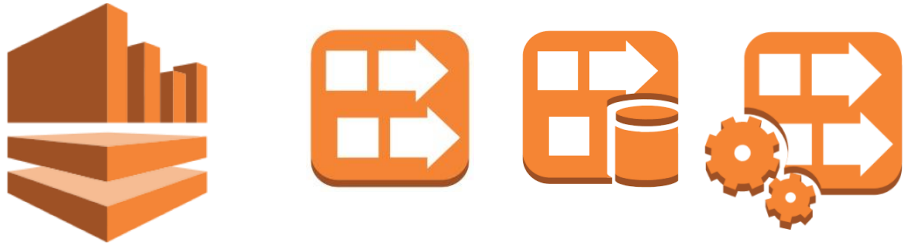
- **EC2 (Elastic Compute Cloud)**
  - Virtuální servery.

- **S3 (Simple Storage Service)**
  - Úložiště, file Storage, Sharing.

- **Relational Database Service**
  - Virtuální realční databáze.
- Route 53
  - Cloudová DNS služba.
- VPC (Virtuál Private Cloud)
  - Virtuální cloud vč. infrastruktury sítě.

PROFINIT

Data Ingestion



- AWS Kinesis
  - Ingest streaming data.
  - Zpracovává data v RT.
  - Ukládá až petabyte dat za hodinu.
  - Kinesis Stream – poskytuje RT metriky, reporty.
  - Kinesis Firehose – masivní load dat do S3 nebo RedShift.
  - Kinesis Analytics – analyzování streamů pomocí SQL.

- AWS Snowball
  - Fyzické zařízení pro „jednorázový“ přenos rychlý dat, 80TB na zařízení.
  - Není vhodné na průběžné přenášení dat na Amazon (AWS Direct Connect).

PROFINIT

Data Storage

- AWS S3
  - Designováno pro online backup a archivaci dat (několik úrovní).
  - Lze uložit a získat jakkoliv velká data pomocí internetu (REST a SOAP).
  - S3 ukládá data v bucketech, do nich ukládá objekty (až 5TB na objekt).
  - Object je uložen jako soubor s metadaty.

- AWS Dynamo DB
  - Fully managed noSQL databáze.
  - Podpora dokumentů a key:value.
  - Velmi nízká latence.
  - Využít pro mobilní hry, živé hlasování, e-commerce.
  - Využivíaji SSD disky, lze nastavit lepší výkon pro konkrétní tabulky.
  - Nevhodné pro velká data s málo I/O operacemi, join a komplexní transakce, BLOB data.

PROFINIT

Data (Pre)Processing and Analyze

- AWS EMR
  - Hadoop ecosystem na Aws – HDFS, Spark, Hive, ...
  - BigData Ucs: log processing, velké ETL, risk modeling, click streaming, reditikvitní analytika, ad-hoc data mining atd.
  - EC2 služba.
  - Lze zálohovat na S3 pro větší odolnost proti výpadkům MN.

- AWS Redshift
  - Fully managed služba.
  - Zvládá petabytové objemy dat, strukturovaná data.
  - Paralelní processing.
  - Využití sloupcového úložiště.
  - Nevhodné pro nestrukturovaná, malá data,OLTP.

PROFINIT

Data (Pre)Processing and Analyze

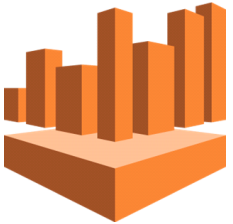
- AWS Machine Learning
  - Managed služba poskytující hotové algoritmy používaní interními DS.
  - RT predikce do 100ms, 200+ transakcí za vteřinu.
  - Modely do 100GB.
  - Interaktivní konzole, SDK, ML API.
  - Nevhodné pro velké modely nad 100GB a unsupervised learning.

- AWS Lambda
  - Event driven služba, fully managed, automaticky škálovaná.
  - RT processing, ETL.
  - Procesing eventu v řádech ms.
  - Podpora Java, Node.js, Python, triggering pomocí timingu nebo eventů.

PROFINIT

Data (Pre)Processing and Analyze

- AWS ElasticSearch
  - Včetně LogStash a Kibany.
  - RT Analýza logů, analýza streamů, monitoring mobilních aplikací…
  - Může fungovat jako storage, EBS storage.
  - Nevhodné pro OLTP, pro loady dat větších než 5TB.



- AWS Athena
  - Serveless řešení pro dotazování nad daty.
  - Popdoruje CSV, JSON, ORC, Avro, Parquet.
  - Využívá S3.

PROFINIT

Data VIsualiaze

- AWS Quicksight
  - Nástroj pro sdílení a spolupráci díky storyboards.
  - Fully managed služba.
  - In-memory processing, paralelní. Rychlé.
  - Levné oproti běžným BI nástrojům.
  - Automaticky najde všechny AWS data zdroje.

- Spousta služeb plní cross funkci – ukládá, zpracovává, vizualizuje.
- Ke službám jsou dostupné často API pro různé jazyky a konzole pro ovládání, či přímo GUI.

PROFINIT

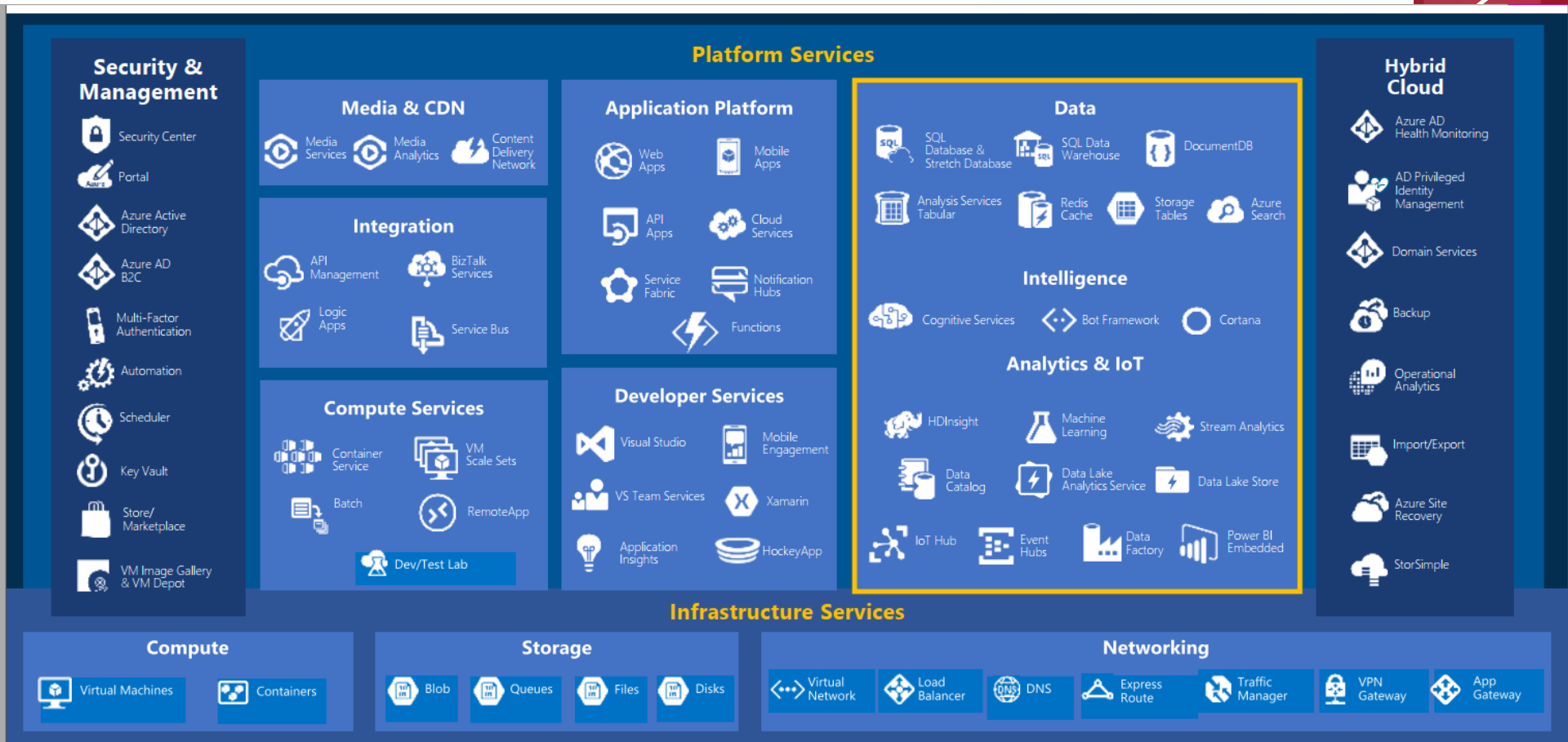
Azure

PROFINIT

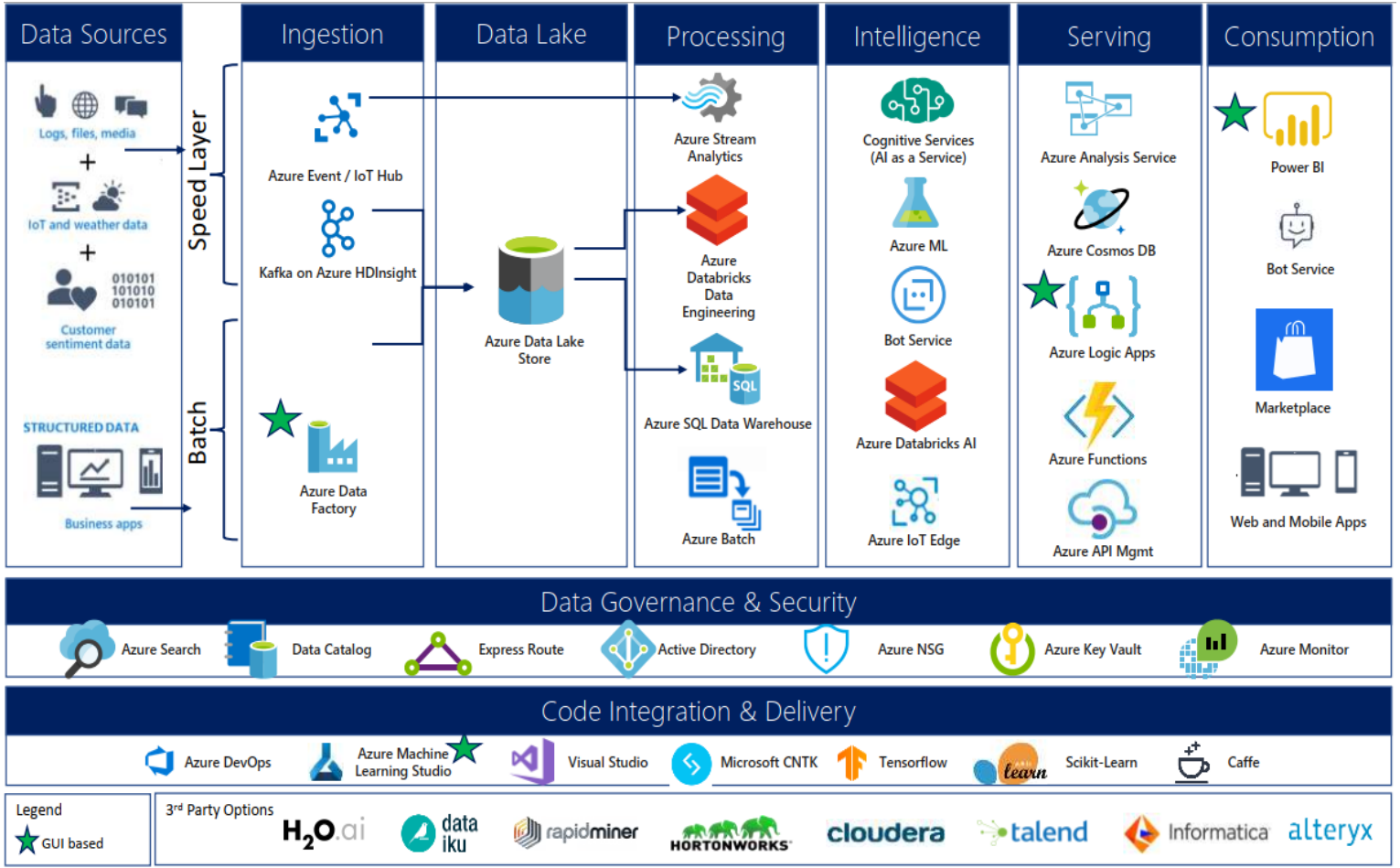
Microsoft Azure

PROFINIT

Azure stack


Security & Management
Security Center
Portal
Azure Active Directory
Azure AD B2C
Multi-Factor Authentication
Automation
Scheduler
Key Vault
Store/ Marketplace
VM Image Gallery & VM Depot
Platform Services
Media & CDN
Media Services
Media Analytics
Content Delivery Network
Integration
API Management
BizTalk Services
Logic Apps
Service Bus
Compute Services
Container Service
VM Scale Sets
Batch
RemoteApp
Dev/Test Lab
Application Platform
Web Apps
Mobile Apps
API Apps
Cloud Services
Service Fabric
Notification Hubs
Functions
Developer Services
Visual Studio
Mobile Engagement
VS Team Services
Xamarin
Application Insights
HockeyApp
Data
SQL Database & Stretch Database
SQL Data Warehouse
DocumentDB
Analysis Services Tabular
Redis Cache
Storage Tables
Azure Search
Intelligence
Cognitive Services
Bot Framework
Cortana
Analytics & IoT
HDInsight
Machine Learning
Stream Analytics
Data Catalog
Data Lake Analytics Service
Data Lake Store
IoT Hub
Event Hubs
Data Factory
Power BI Embedded
Hybrid Cloud
Azure AD Health Monitoring
AD Privileged Identity Management
Domain Services
Backup
Operational Analytics
Import/Export
Azure Site Recovery
StorSimple
Infrastructure Services
Compute
Virtual Machines
Containers
Storage
Blob
Queues
Files
Disks
Networking
Virtual Network
Load Balancer
DNS
Express Route
Traffic Manager
VPN Gateway
App Gateway

BDDS Stack



PROFINIT

Azure Úložiště

- **Blob Storage**
  - Úložiště zejména pro nestrukturovaná data, která nepoužíváme často.
  - Nejlevnější způsob ukládání dat na Azure.
  - Lze škálovat.
  - Přístup pomocí API.

- **Data Lake Storage Gen2**
  - Vylepšený Blob Storage,
  - ACL (posix práva),
  - hierarchiciká struktura,
  - **rychlé úložiště pro velká data**.

PROFINIT

Azure Databáze

- **Azure Cosmos DB**
  - **noSQL databáze, distribuovaná**,
  - Highly responsible, highly available,
  - SQL API, mondoDB API, gremlin API, Cassandra API, Table API.

- **Azure SQL DB** ;
  - Cloudová databáze -> snadné škálování,
  - nativní podpora ML,
  - několik možností škálování podle potřeby (běžná práce, OLTP, autoškálování…),
  - zvládne velký počet uživatelů naráz.

- **Azure SQL DWH**
  - Využívá Massive Parallel Processing,
  - Data jsou uložena odděleně od výpočtu -> šetří se,
  - Vhodné pro menší počet uživatelů.

PROFINIT

Azure ML

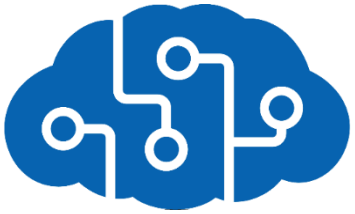
- Proprietární služba Microsoft Azure
  - Dobrá integrace s ostatními službami na Azure.
  - **Možnost MLOps s Azure DevOps**.
  - Platí se náklady pouze za výpočet, nikoliv službu.

- Podporuje PythonSDK a R a spoustu ml frameworků.
  - Integrace spolu s mlflow.
  - **Neumí Spark!**

- Vývoj v populárních IDE a kolaborativní JuPyter Notebooky.

PROFINIT

Cognitive Services

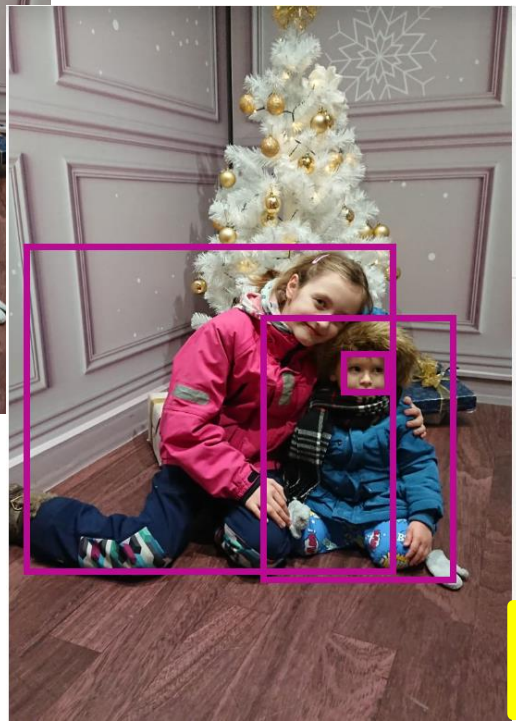


- Již naučené ML algoritmy (text, řeč, obrázky…).
  - Lze je ale naučit na svoje data!
- API mnoha služeb, které je snadné integrovat.
- Licence za použití.

- Ukázky:
  - https://azure.microsoft.com/cs-cz/services/cognitive-services/computer-vision/#features
  - https://azure.microsoft.com/cs-cz/services/cognitive-services/face/#demo
  - https://www.how-old.net/

PROFINIT

Ukázka Cognitive Services ☺



"smile", "confidence": 0.826899648 }, { "name": "baby", "confidence": 0.817665756 }, { "name": "christmas", "confidence": 0.7937247 }, { "name": "child", "confidence": 0.6052515 }]

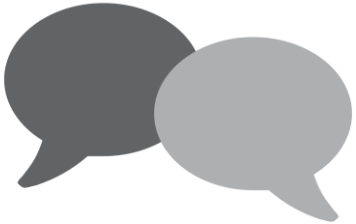
{ "tags": ["person", "indoor", "child", "baby", "sitting", "table", "little", "stuffed", "holding", "toy", "small", "teddy", "girl", "floor", "young", "boy", "bear", "wooden", "playing", "bed", "laying", "blue", "white"], "captions": [{ "text": "a baby holding a stuffed animal", "confidence": 0.6358383 }] }

PROFINIT

HDInsight

- A taky tu máme “Azure Hadoop” ☺
- Něco, co už z BigDat dobře známe ☺
- Široká paleta nástrojů, které známe ☺
  - viz oficiální přehled.

- Seamless integrace s Azure nástroji.
- Dynamická škálovatelnost clusterů.



Dotazy

PROFINIT