**Clustere de Calculatoare**

**Curs 1**

**Cluster de servere** – reuneste o familie de masini(numite *noduri*) care opereaza impreuna si sunt vazute drept o singura entitate de calcul. Au fost folosite initial doar pentru cercetare, nu erau utilizare in business.

* Stocarea datelor era cea mai mare problema

Au ca scop:

* Cresterea puterii de calcul si a disponibilitatii
* Load balancing(<https://www.geeksforgeeks.org/load-balancing-in-cloud-computing/>)
* Simplificarea managementului resurselor

Tipuri de Clustere:

* **Fail-over(des implementat)**
* **Load-balancing(des implementat)**
* High-Performance-Computing

**Grid Computing** – specific conceptului de grid este faptul ca permite o utilizare transparenta a resurselor de calcul in conditiile de distribuire la nivel global **fara** a conta locatia unde sunt plasate resursele de calcul.

1. **HPG(High-Performance-Grid)** – ofera putere de calcul mare prin intermediul unor sisteme de calcul cu mare putere de procesare
2. **Peer-to-peer** – obtinerea de putere de calcul prin disponibilizarea resurselor neutilizate la nivelul nodurilor

**Modelul de programare MapReduce:**

* Procesare masiva a datelor pe un numar mare de calculatoare
* Functiile map si reduce pot fi utilizate in clustere de dimensiuni mari cu sarcini de procesare la nivelul petabaiti-lor(ex: sortare)
* Dependentele intre task-urile elementare sunt reprezentate print-un graf acyclic(DAG)

**Service Computing** – obiectivul sau este de a eficentiza interoperarea dintre serviciile IT si tehnologii de calcul cu serviciile de business. Problematica acoperita de SC porneste din zona consultantei pentru afaceri si “enterprise modelling” pana la servicii Web si cloud computing.

**Cloud computing** este un model care permite accesul permanent la un grup de resurse de calcul configurabile(de exemplu, retele, servere, solutii de stocare, aplicatii si servicii), care pot fi gestionate cu un efort minim.

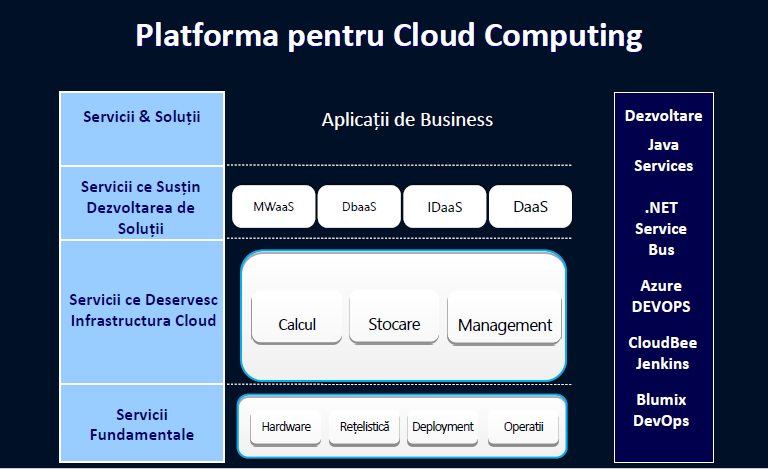
**SaaS + PaaS + IaaS + DaaS**

1. **Software as a Servise (SaaS)** - (din perspectiva utilizatorului) experienta utilizarii functionalitatilor expuse de aplicatie este consumata peste internet (Office 365, GoogleAps)
2. **Platform as a Service (PasS)** – (din perspectiva dezvoltatorului) – furnizorii de cloud ofera dezvoltatorilor care nu doresc utilizarea cloud-ului privat o platforma bazata pe internet cu ajutorul careia sa creeze, instaleze sis a intretina servicii software
3. **Infrastructura as a Service (IaaS)** – furnizorii de cloud construiect retele de centre de date la nivel global. Utilizatorii de cloud inchiriaza servicii de storage, calcul si mentenanta de la furnizorii de cloud (pay-as-you-go).

**Calcul distribuit:** este un model de calcul in care componentele unui system software sunt partajate peste o retea de calculatoare, iar comunicarea intre ele se bazeaza pe protocoale de mesagerie(http, rpc, msmq).

**SOA:** arhitectura de calcul distribuit avand asociata o paradigma de software in care serviciile sunt reprezentate de aplicatii executate la nivelul nodurilor computationale.

|  |  |
| --- | --- |
| **AVANTAJE** | **DEZAVANTAJE** |
| 1. Serviciile web sunt independente de platforma si independente de limbaj deoarece utilizeaza XML. Majoritatea serviciilor Web utilizeaza http pentru transmiterea mesajelor, aceste este un avantaj deoarece proxy-urile de pe internet NU vor afecta traficul http. | 1. Transmiterea tuturor datelor in format XML nu este in mod evident la fel de eficienta ca utilizarea unui cod binar. “Cheltuiala” este de obicei acceptabila pentru majoritatea aplicatiilor. 2. Slaba versatilitate |

****

**Virtualizarea resurselor de calcul**

**Curs 2**

**Open Compute Project –** comunitate globala de lideri in domeniul tehnologiei care coopereaza pentru a elimina blocajele de la nivelul infrastructurii IT proprietare in scopul eficientizarii productiei de hardware prin cresterea flexibilitatii si scalabilitatii acesteia.

Cloud este **un mediu multi-tenant partajat** caracterizat printr-un inalt nivel de automatizare, preferabil bazat pe o infrastructura virtualizata in care resursele IT pot fi provizionate si masurate.

Virtualizarea

* Procesul prin care este create o versiune virtuala(mai degraba posibila decat reala) a unor entitati computationale, incluzand aici platforma hardware virtuale, dispositive de stocare si resurse de retea
* **Virtualizare totala –** proces de natura computationala in care o masina ruleaza pe o alta

**Paravirtualizare**

* Permite lucrul cu sisteme de operare diferite pe acelasi hardware
* Managementul se realizaeaza in conexiune cu un system de operare

**Containere**

* Furnizeaza medii de runtime isolate. Intregul spatiu utilizator este expus ca un container si orice modificare a acestuia **NU** influenteaza restul containerelor

**Cloud privat** – Solutia este operata exclusiv de beneficiar

**Cloud Public –** serviciile de cloud sunt furnizate publicului larg sar organizatiilor si este detinut si administrat de organizatii specilizate in furnizarea de servicii de cloud.

**Cloud al comunitatii** – implica partajarea resurselor de catre una sau mai multe organizatii fiind sustinut de o comunitate specifica caracterizata prin preocupari comune.

Caracteristici:

* Furnizarea de servicii la cerere
* Resurse agregate
* Elsticitatea in alocarea resurselor
* Masurabilitatea resurselor allocate prin servicii
* Partajarea multi-tenant

Servicii: (vizeaza posibilitatea utilizarii de componente disponibile in retea)

* Scalabilitate extinsa
* Capacitate de partajare
* Posibilitatea de a consuma serviciul de pe diferite platforme

**Software as a Service – model de distribuire a componentelor software stocate ca serviciu si accesate de consumatori pe internet**

**Beneficii:**

* Usor de exploatat si intretinut
* Usor configurabil
* Securitate ridicata
* Latimea de banda in permanenta crestere permite o exploatare performanta

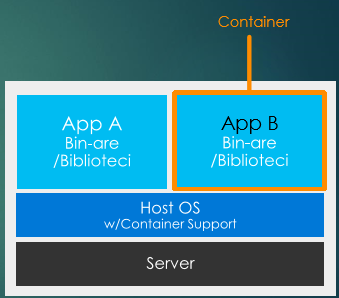
Acesta a fost dezoltat in special pentru a folosi instrumente web. Ofera posibilitatea livrarii software-ului disponibil in comert dintr-o locatie de unde clientii o pot accesa ori de cate ori au acces la internet.

**Containere**

**Curs 3**

**Containerele – cunoscute drept forme de virtualizare la nivelul sistemului de operare si reprezinta o abordare de complexitate redusa**

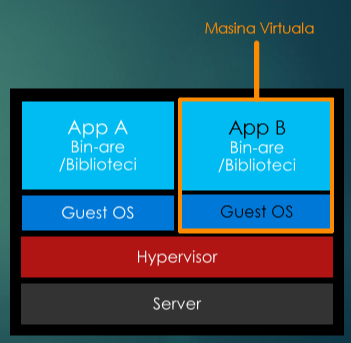
**Pe o singura masina-gazda pot rula mai multe containere**. Fiecare container poate gazdui o aplicatie web sau un serviciu. Containerele IZOLEAZA aplicatiile unele de altele pe un system de operare partajat



**Generalitati**:

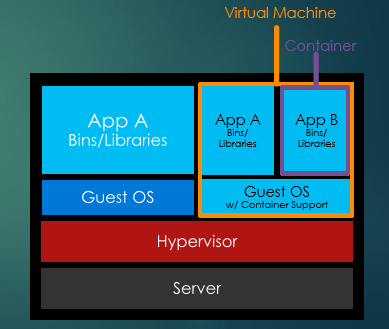
* **Dependente** - Fiecare aplicatie are propriile dependente care include atat elemente software(servicii, biblioteci) cat si hardware(cpu, memorie)
* **Virtualizarea**: Mototrul containerelor este un system de management si orchestrare al acestora, echivalent cu virtualizarea prin care se pot izola dependentele pentru fiecare aplicatie prin “ambalarea” lor in containere virtuale
* **Partajarea SO gazda**: Procesele **partajeaza** kernel-ul cu gazda si cu celelalte containere
* **Flexibilitate**: Diferentele dintre sistemul de operare si infrastructura de baza de abstractioneaza simplificand abordarea prin “deployment oriunde”.
* **Rapiditatea**: Containerele pot fi create aproape instant permitand o scalare elastic asociata cererii

**Diferente fata de masini virtuale(acestea au ca host si un SO spre deosebire de containere)**



* **Dependentele** – fiecare aplicatie virtualizata include aplicatia in sine, avand nevoie de biblioteci, dar si de un host SO
* **SO independent**: Fiecare masina virtuala poate avea un system de operare diferit
* **Flexibilitate**: Masinile virtuale pot fi migrate in alte locatii pentru a balansa resursele utilizate si pentru mentenanta locatiei fara “downtime”.
* **Securitate**: Nivel inalt de izolare securizata

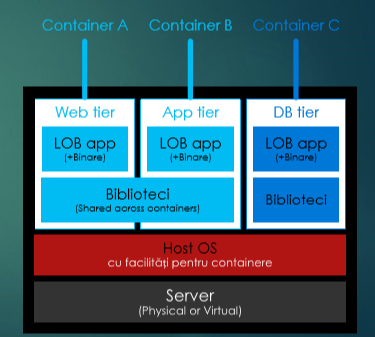
**Scenarii multiple de implementare a aplicatiilor**

****

Containere in MV:

* Utilizatorii pot implementa mai multe sisteme de operare in masini virtuale diferite, iar in interior pot implementa mai multe containere in cadrul acelor SO
* Prin combinarea acestora ar fi necesare mai putine MV pentru a suporta un numar mai mare de aplicatii
* Rezulta reducerea necesarului de stocare

**Containere Windows Server**

****

1. **Build** – se vor folosi instrumente precum Visual Studio/Eclipse pentru a scrie aplicatii si a rula in containere. Prin construirea acestor aplicatii, modulele se pot actualize independent
2. **Run** – functionalitatile containerelor sunt built-in in Windows Server 2016
3. **Management** – deployment-ul si managementul containerelor se realizeaza cu PowerShell sau Docker.
4. **Resurse**

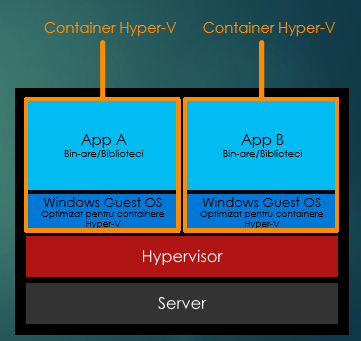
**Docker –** proiect pentru automatizarea implementariiaplicatiilor in containere portabile si care pot functiona in cloud sau local.

**AVANTAJE**: Prin containerizarea aplicatiilor vechi utilizand varianta de containerizare Windows Server, obtinem o consistenta si un management imbunatatit al echipelor de dezvoltatori si testeri

**Terminologie**:

* **Imaginea containerului** – pachet ce contine toate dependintele si informatiile necesare pentru a crea un container. O imagine este imutabile odata ce a fost creata.
* **Container** – instanta a unei imagini Docker. Un container este un mediu de runtime pentru o singura aplicatie, process sau serviciu. Un job batch poate crea multiple containere dintr-o imagine prin transmiterea de parametric diferiti catre fiecare instanta.
* **DockerFile** – fisier text ce contine instructuri pentru generarea unei imagini de docker
* **Build** – Constructia unei imagini bazate pe informatiile si contextul descris in DockerFile

**Containere Hyper-V**

****

* **Consistenta** – asigura coerenta intre seturile de instrumente de gestionare si implementare
* **Compatibilitatea** – utilizeaza aceleasi imagini ca si containerele Windows Server
* **Izolare**(puternica) – fiecare container hyper-x are propria copie dedicate a kernel-ului
* Este foarte sigur
* **Optimizare**: Layer-ul de virtualizare si sistemul de operare au fost special optimizate pentru lucrul cu containere

Containerele reprezinta un excellent mediu pentru: calcul distribuit, scalare, baze de date, task-uri, web app.

**Comparatie intre containere si VM**

Livrare si Integrare Continua

* **Livrare Continua(LC)** – concept ce implica scrierea, testarea si deployment-ul de software la perioade foarte scurte. Este o prelungire a integrarii continue.
* **Integrarea Continua(IC)** – practica care cere dezvoltatorilor sa integreze codul(2-3 ori/zi) intr-un repo comun. Acest lucru permite detectarea intr-o etapa incipienta a eventualelor probleme

**Cloud Storage**

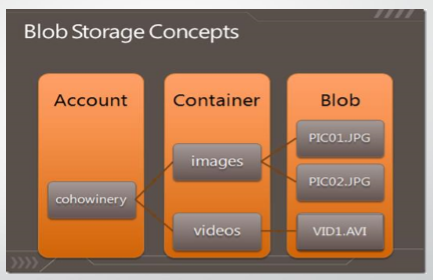
**Curs 4**

Stocarea ca serviciu este un model de business in care o companie furnizore inchiriaza spatiu de stocare in infrastructura proprie, unei alte companii sau persoane fizice.

Abstractiile care fundamenteaza Azure Storage

* **BLOB** – **interfata simpla pentru stocarea fisierelor impreuna cu metadatele associate acestora**
* **Fisiere**
* **Tabele** – ofera stocare structurata de tip noSQL masiv scalabila. O tabela este o familie de entitati care contin o multime de proprietati.
* **Cozi** – ofera o stocare si o livrare sigura a mesajelor pentru o aplicatie pentru a construi fluxul de lucru
* **Drivere**

**BLOB-uri in Windows Azure**



* Politicile de distribuire sunt setate la nivelul containerului, unde un container poate fi setat ca privat sau sa fie accesibil publicului

**BLOB-uri Windows Azure**

**Containere BLOB:**

* Ofera o grupare de seturi de blob-uri
* Numele containereului este cuprins de cont
* Politicile de distribuire sunt setate la nivelul containerului
* Daca un container este “Public”, intregul sau continut poate fi citit de oricine

**Tabele Windows Azure** – system de stocare strutcurat si scalabil non-relational, de tip cheie-valoare, destinat stocarii unor volume mari de date nestructurate. **Extragere simpla si adaugare rapida.** Exista urmatoarele 3 proprietati pentri fiecare entitate**: PartitionKey, RowKey(unic identificator al entitatii in cadrul partitiei), TimeStamp(ora ultimei modificari a entitatii si este gestionata de sistemul storage**). Dimensiune admisa: 200TB.

**Servicii de Stocare in Cloud** – model de cloud computing destinat stocarii datelor in format digital intr-o maniere logica, stocarea fizica fiind realizata pe sisteme distribuite de servere geolocate. Furnizorii de stocare in cloud raspund pentru pastrarea datelor si accesibilitatea acestora, precum si de protejarea si functionarea mediului fizic. Utilizatorii achizitioneaza sau inchiriaza capacitate de stocare de la furnizorii de servicii pentru a stoca date necesare utilizatorilor privati, companiilor sau aplicatiilor detine de acestia.

Arhitectura serviciului Azure Storage

1. Nivelul partitie:
   1. Gestioneaza partajarea tuturor obiectelor de date din system
   2. Obiectele au asociata o cheie de partitionare
   3. Acesta este nivelul care gestioneaza ce partitie este alocata unui server de partitii dat, asigura echilibrarea incarcarii partitiilor de pe servere pentru a raspunde nevoilor de traffic ale blob-urilor.
2. Nivelul de distributie si replicare al fisierelor
   1. Nivelul DFS: cel care gestioneaza stocarea efectiva pe disc si este responsabil de distribuirea si replicarea datelor.

**Windows Azure Storage**

Facebook/Twitter stocheaza date in blob-uri. Utilizeaza cozi pentru managementul workflow-ului. Se poate ajunge la 40000 solicitari/sec si la 2-3 miliarde de cereri/zi.

**Cozi de mesaje: UN MESAJ POATE FI PROCESAT O SINGURA DATA.** Preluarea mesajului il face invizibil in coada pentru un timp specificat. Stergerea mesajului in procesul de prelucrare implica eliminarea lui din coada. Daca procesul cade atunci mesajul devine vizibil in coada pentru alt process worker.

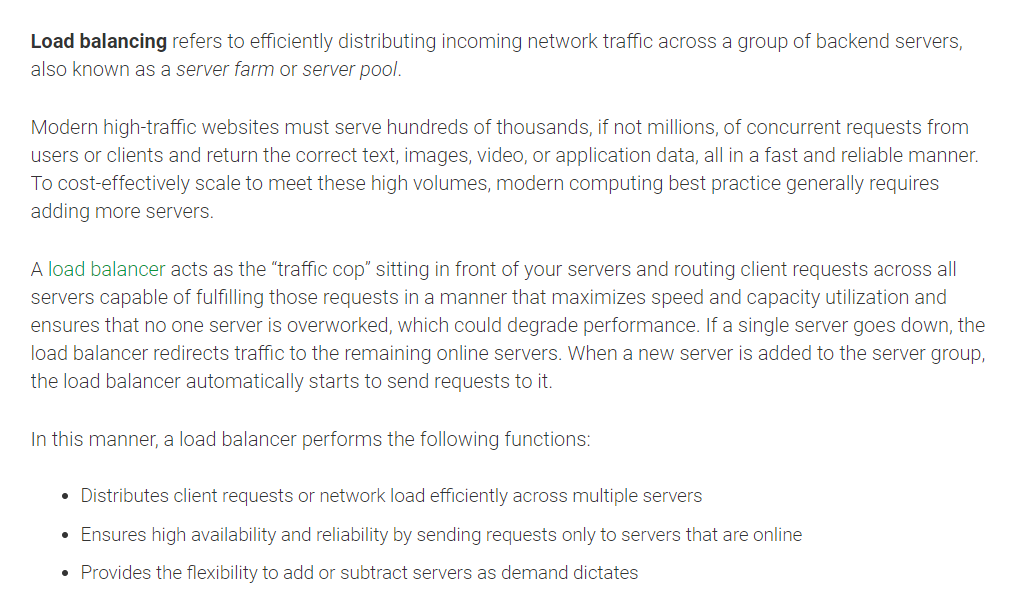
DB as a Service & Data Lake

Curs 5

**Azure Sql Database**

Baza de date livrata ca serviciu, complet gestionata de Microsoft(efort minim de administrare).

**Geo-replicare standard:** generarea unei replici secundare intr-o regiune pereche. Replica este taxata la costuri reduse. Activarea replicii se face de catre Microsoft in cazul unui incident.



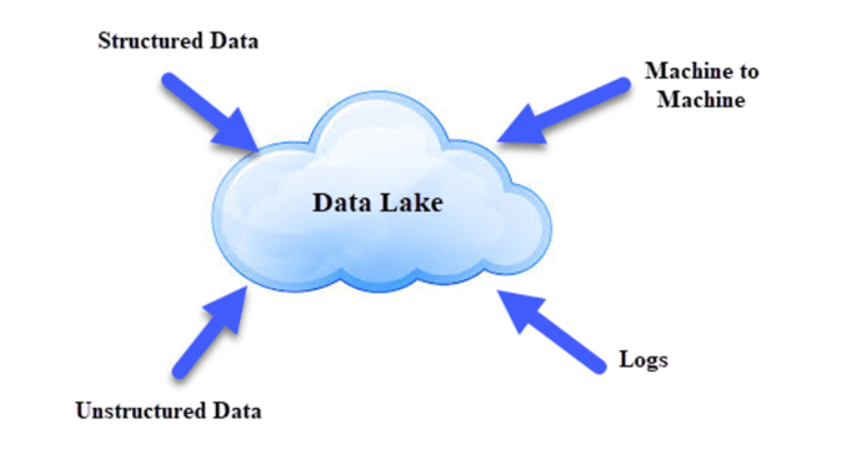
**Zero-administrare –** infrastructura virtualizata elimina aproape tot efortul de intretinere, inclusive aplicarea patch-urilor.

**Data-Warehouse** – cel mai elaborate sisteme de gestiune a datelor din IT.

**Big Data(high volume/velocity/variety)** – is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. (<https://www.guru99.com/what-is-big-data.html>).

* Toate datele au o valoare potential
* Aplicatiile si utilizatorii interpreteaza datele asa cum considera de cuviinta

**Data Lake** – is a storage repository that can store largo amount of structured, semi-structured and unstructured data. It is a place to store every type of data in its native format with no fixed limits on accound size or file.



**Sql Server 2016 R Services**

**Curs 6**

**Bariere in adoptarea analiticelor**

* Deficitul de talente
* Productivitate
* Infrastructura complexa
* Dinamica scazuta a inovatiei
* Costuri mari

**Ce este R –** limbaj de programare pentru statistica, tool pentru vizualizare date**. (Nu este nivel ‘Enterprise’).**

**Introducere in SQL Server R Services**

1. **Simplitate**
   1. Viteza de executie si scalare
   2. Analitice db integrate
   3. Procesare si threading parallel
   4. Reutilizare competente SQL pentru “data engineering”.
2. **Scalabilitate**
   1. Memorie + disk
   2. “write once, deploy anywhere”.
3. **Cost redus**
   1. Included in SQL Server 2016
   2. Reutilizeaza codul R

**Ciclul de viata al analiticelor:**

1. **Pregatirea –** colectarea, “curatarea”, profilarea si transformarea diverselor date relevante pentru subiect.
2. **Model –** utilizarea statisticii si algoritmcii ML pentru a construe clasificatori si modele de regresie.
3. **Operationalizare –** efectuarea de predictii si vizualizari pentru sustinerea app de business.

**Azure Machine Learning**

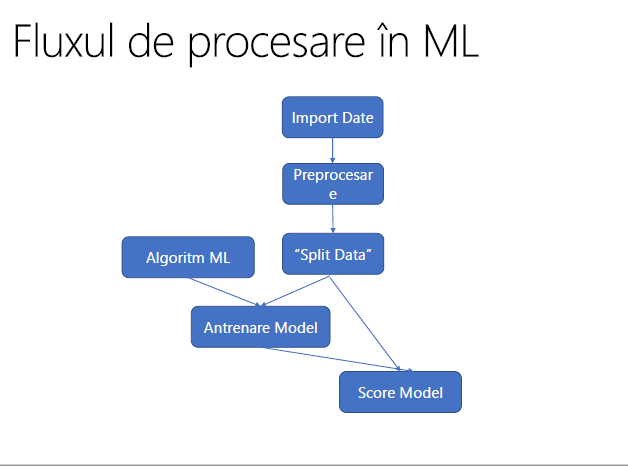
**Curs 7**

**AML** suporta numeroase limbaje de programare, in diverse metodologii de dezvoltate si diverse medii de programare, incluzand atat sisteme de software apartinand Microsoft dar si altor vendori**.**

**Tipuri de invatare:**

1. **Supervizata –** sistemul este instruit folosind datele anterioare(includ si intrarile si iesirile) fiind recomandat a fi aplicat la problemele de decizie sau pentru realizarea de previziuni atunci cand se intalnesc date noi.
2. **Nesupervizata –** sistemul este capabil sa recunoasca regularitatile, luand in considerare numai datele de intrare.
3. **Consolidata –** deciziile sunt luate de system pe baza recompense / “pedepsei” aplicate pentru ultima actiune efectuata.

**Azure Machine Learning Studio –** permite BUILD – TEST – DEPLOY



**Probleme in obtinerea unor predictii de calitate:**

* Cantitati de date adecvate
* Nu trebuie tintita perfectiunea

**Azure SQL Database, Azure Data Warehouse, Azure Data Lake**

**Curs 8**

**Azure SQL Database –** baza de date livrata ca serviciu, complet gestionata de Microsoft

**Arhitecturi Big Data**

**Curs 9**

**Big Data –** caracterizeaza seturi de date de dimensiuni mari si complexitate ridicata pentru care aplicatiile traditionale si analiza a acestora sunt inadecvate.

Solutiile Big Data implica unsa sau mai multe sarcini:

* Procesare serial a volumelor mari de date statice
* Procesara in timp real a volumelor mari de date dinamice
* Explorarea interactive a volumelor mari de date
* Analiza predictive a volumelor learning.

**Trecerea la arhitecturi Bog Data este necesara cand:**

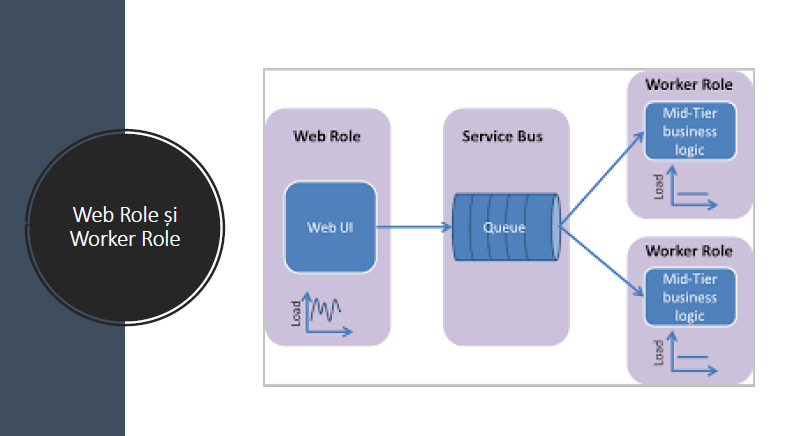
* Stocheaza si proceseaza date in volume prea mari pentru SGBD clasice
* Transforma date nestructurate pentru analiza si rapoarte
* Preia, proceseaza si analizeaza nelimitate stream-uri de date in timp real cu latente neglijabile

**Analiza datelor stocate**

* Multe solutii in Big Data implica pregatirea anterioara a datelor pentru analiza si apoi transferul acestora catre prelucrare intr-un format structurat care poate fi interogat utilizand instrumente analitice.

**Diferenta intre “Web Role” si “Worker Role”**

* Un “Web Role” instaleaza automat si gazduieste aplicatia utilizator in IIS
* Un “Worker Role” nu utilizeaza IID si ruleaza aplicatia in mod autonomy, aceasta fiind instalata si livrata prin intermediul platformei de **Servicii Azure,** ambele pot fi gestionate in acelasi mod si pot fi instalate pe aceeasi instant Azure.

****

**ETL Modernizat cu ADL**

**Curs 10**

**Abordare bazata pe Data Lake**

1. **Preluarea tuturor datelor –** indiferent de cerinte
2. **Stocarea tuturor datelor –** in format nativ, fara schema
3. **Analiza**

**Azure CosmosDB**

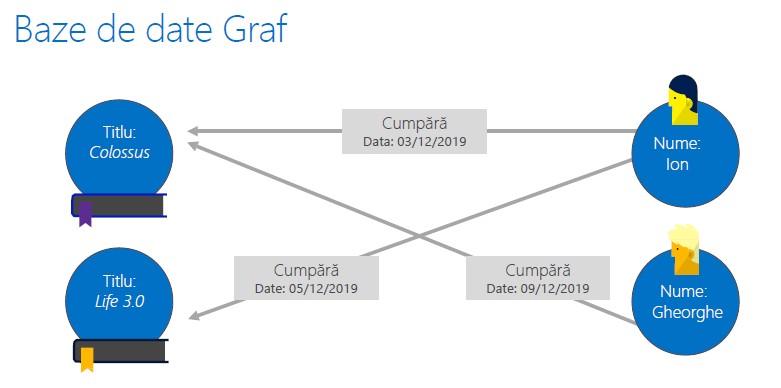
**Curs 11**

O baza de date NoSql utilizeaza patru categorii de stocare: cheie-valoare, columnar, document si graf.

**Stocarea cheie valoare:** - ofera o varianta de stocare rapida prin intermediul unui model de date de complexitate redusa(orice poate fi stocat drept valoare, atat timp cat fiecare valoare este asociata unei chei sau nume).

**Stocare de tip columnar**: datele sunt stocate in cellule grupate pe coloane. Coloanele sunt grupate logic in “column families”. Acestea pot contine un numar theoretic nelimitat de coloane ce pot fi create la runtime sau la definirea schemei.

**Baze de date graf:**

****

**Spatii de stocare pentru documente –** contin obiecte ca date ierarhizate in structure arborescente.

**Proprietati importante**

1. **Indexare automata**
   1. Toate proprietatile documentului sunt disponibile pentru interogari
   2. Nu este necesara schema sau indexul secundar
2. **Managed service**
   1. Spin-up la cerere fara nicio configurare
3. **Auto-scalare**
   1. Scalabilitate si fiabilitate imbunatatita datorita distributiei seturilor de date mari pe mai multe masini
4. **Sql query language**
   1. Poate fi utilizat atat sql cat si .NET LINQ.