



Client Amazon S3



Coordonator științific:

George Iordache

Studenți:

Chiriță Ioan-Alexandru, (SSA2)

Coman Ioan, (AAC)

Cuprins

1. Introducere	3
1.1 Amazon S3 Client	3
1.1.1 Beneficii, cazuri de utilizare și clienți.....	3
1.2 Big Data	4
1.2.1 Caracteristici	4
1.3 Cloud Computing.....	5
2. Definirea problemei și analiza soluției	5
3. Dezvoltarea proiectului	6
3.1 Prima fază a proiectului (Midterm).....	6
3.2 A doua fază a proiectului (Faza finală)	6
4. Amazon S3 vs. Microsoft Azure vs. Google Cloud.....	7
4.1 Amazon Web Services	7
4.2 Microsoft Azure	7
4.3 Google Cloud	7
4.4 Comparăția vitezei de încărcare și descărcare [6]	7
5. Evaluarea rezultatelor proprii obținute	11
6. Concluzie	12
7. Bibliografie	13

1. Introducere

1.1 Amazon S3 Client

S3 Browser este un client Windows gratuit pentru Amazon S3 și Amazon CloudFront. Amazon S3 oferă o interfață simplă de servicii web care poate fi folosită pentru stocarea și preluarea oricărei cantități de date, în orice moment, de oriunde pe web. Amazon CloudFront este o rețea de livrare de conținut (CDN). Acesta poate fi utilizat pentru a livra fișierele folosind o rețea globală de locații de margine.

Serviciul de stocare simplă Amazon (Amazon S3) este un serviciu de stocare a obiectelor care oferă scalabilitate, disponibilitatea datelor, securitate și performanță în domeniul industriei. Acest lucru înseamnă că clienți de toate dimensiunile și industriile îl pot utiliza pentru a stoca și proteja orice cantitate de date pentru o serie de cazuri de utilizare, cum ar fi site-uri web, aplicații mobile, backup și restaurare, arhivă, aplicații de întreprindere, dispozitive IoT și analize de date mari. Amazon S3 oferă funcții de administrare ușor de utilizat, astfel încât să vă puteți organiza datele și să configurați controale de acces ajustate fin pentru a satisface cerințele dumneavoastră specifice de afaceri, organizaționale și de conformitate. Amazon S3 este proiectat pentru o durabilitate de 99,99% și stochează date pentru milioane de aplicații pentru companii din întreaga lume.

1.1.1 Beneficii, cazuri de utilizare și clienți

Beneficii:

- ✓ Performanță, scalabilitate, disponibilitate și durabilitate în industrie
- ✓ O gamă largă de clase de stocare rentabile
- ✓ Gestionarea cu ușurință a datelor și a controalelor de acces
- ✓ Servicii de interogare pentru analiză
- ✓ Cele mai multe servicii de stocare în cloud suportate

Clienți:

- Netflix
- Finra
- Airbnb
- GE Healthcare

Cazuri de utilizare:

- Backup și restaurare
- Recuperare în caz de catastrofe (DR)
- Arhivă
- Baze și analize de date mari
- Spațiu de stocare hibrid
- Date despre aplicația nativă din cloud

1.2 Big Data

BIG DATA



"Datele păstrate și prelucrate în cantități imense, datorită unor medii de stocare mai ieftine, unor metode de procesare mai rapide și unor algoritmi mai performanți"¹

Big Data a devenit o problemă în afaceri, sau cel puțin o problemă pe care oamenii de afaceri încep să o conștientizeze. Presa începe să aloce din ce în ce mai mult spațiu acestui subiect. Pornind cu Wall Street Journal "Companiile sunt inundate cu date"

("Companies are being inundated with data") la Financial Times "Din ce în ce în afaceri sunt aplicate analize din mass-media, cum ar fi Facebook și Twitter" ("Increasingly businesses are applying analytics to social media such as Facebook and Twitter"), Forbes "Big Data a ajuns la Seton Health Care Family" ("Big Data has arrived at Seton Health Care Family"). De ce atâtea articole pe aceasta temă? Deoarece BIG DATA are potențialul de a afecta profund modul în care facem afaceri și chiar modul de a trăi.

1.2.1 Caracteristici

✓ Volumul

Da, volumul de date este în creștere. Experții prezic că volumul de date din lume, va crește la 25 de Zettabytes în 2020. Același fenomen afectează fiecare companie - datele sunt în creștere la aceeași rată exponențială. Dar nu este numai volumul de date care este în creștere, numărul de surse de date este de asemenea în creștere.

✓ Viteza

Datele se creează la viteze din ce în ce mai mari. Companiile își mută aplicațiile de la aplicații de tip "batch" la aplicații în timp real. Și cerințele de afaceri au crescut la fel - de la răspunsuri săptămâna viitoare sau mâine la un răspuns într-un minut sau la secundă. Și lumea este, de asemenea, din ce în ce mai instrumentată și interconectată. Volumul de date de streaming de pe aceste instrumente este exponențial mai mare decât a fost chiar cu 2 ani în urmă.

✓ Varietatea datelor

Varietatea datelor prezintă o provocare la fel de dificilă. Creșterea surselor de date a alimentat și creșterea tipurilor de date. De fapt, 80% din datele generate în lume sunt date nestructurate. Cu toate acestea, metodele tradiționale de analiză se aplică numai la informații structurate.

¹ Viktor Mayer-Schönberger și Kenneth Cukier, "Big Data: A revolution that will transform how we live".

✓ Veridicitatea datelor

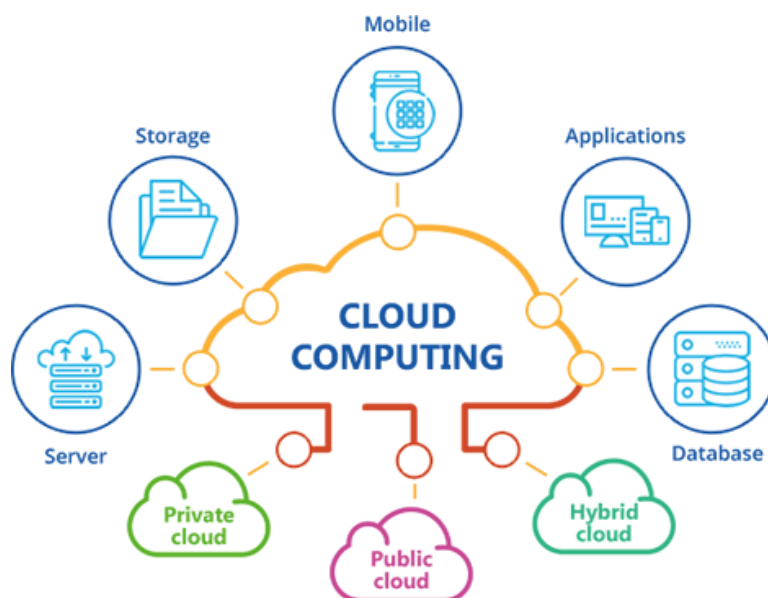
Cum se poate acționa pe baza acestor informații, dacă nu sunt de încredere. Stabilirea încrederii în datele pe care le folosește orice companie reprezintă o provocare uriașă odată cu creșterea surselor și tipurilor de date.

Un alt motiv pentru care Big Data este un subiect fierbinte astăzi este noua tehnologie care permite unei organizații să beneficieze de resursele interne de date. Ceea ce este nou, este tehnologia pentru a procesa și analiza aceste date la volumul și viteza dorită. Scopul tehnologiei Big Data este să analizeze toate datele disponibile, eficient din punct de vedere costuri. Orice date, așa cum sunt. Se pot analiza date structurate, video, audio, date spațiale sau orice tip de date.

1.3 Cloud Computing

Cloud Computing este un concept modern în domeniul computerelor și informaticii, reprezentând un ansamblu distribuit de servicii de calcul, aplicații, acces la informații și stocare de date, fără ca utilizatorul să aibă nevoie să cunoască amplasarea și configurația fizică a sistemelor care furnizează aceste servicii.

Expresia „cloud computing” derivă dintr-o reprezentare grafică simbolică a Internetului des întâlnită în formă de nor („the cloud”), folosită atunci când detaliile tehnice ale Internetului pot fi ignorate, ca în imaginea din dreapta.



2. Definirea problemei și analiza soluției

Proiectul „Client Amazon S3” își propune să testeze eficiența și randamentul serviciului Amazon S3.

Pentru aceasta va fi dezvoltată o aplicație în Python care va primi, ca parametru în linie de comandă, un director care conține fișiere de diferite mărimi. Se vor folosi fișiere de diverse mărimi pentru a vedea timpul necesar rulării operației de upload și, de asemenea, timpul aferent operației de download. Deoarece conținutul fișierelor nu este important, ci ne interesează mărimea acestora în MB, operația de download va fi efectuată pe aceleași fișiere.

Intenționăm ca, odată atins obiectivul principal al proiectului, să realizăm o comparație cu Windows Microsoft Azure și/sau cu Google Cloud Storage pentru a trasa o paralelă între aceste servicii asemănătoare.

3. Dezvoltarea proiectului

3.1 Prima fază a proiectului (Midterm)

Prima faza a proiectului a constat în crearea unui cont pe platforma Amazon Web Services și pregătirea mediului de dezvoltare.

Pași:

1. Am creat un nou utilizator (accesând secțiunea *IAM* din meniul *Services* de pe platforma AWS)
2. Am instalat o mașină virtuală de XUbuntu pe care am instalat python3 și boto3 (SDK Python ce permite dezvoltatorilor să creeze, configureze și organizeze serviciile AWS, precum EC2 sau S3)
3. Am instalat AWS CLI versiunea 2 pe XUbuntu

În ceea ce privește partea de dezvoltare, în această primă etapă am implementat operația de upload a unui fișier de 1MB (am creat un fișier binar de 1MB folosind comanda *fallocate*) pe platforma AWS. Fiind partea de început a dezvoltării este și cea care a avut, după părerea noastră, cel mai slab procent privind implementarea raportată la analiza și investigațiile efectuate.

3.2 A doua fază a proiectului (Faza finală)

În a doua fază a proiectului, cea finală, am implementat și cea de-a doua operație, cea de download.

Programul dezvoltat este format din 4 metode:

- *initialize* – folosită pentru inițializarea clientului de Amazon S3 și a bucketului în care se va face încărcarea, respectiv din care se va face descărcarea fișierelor
- *clean* – folosită în mod opțional pentru a șterge conținutul bucketului și bucketul, dar și folderul local în care s-a realizat descărcarea fișierelor (acest lucru a fost decis deoarece ne-am dat seama că nu putem șterge toate bucket-urile manual direct de pe platforma AWS)
- *upload* – metoda folosită pentru încărcarea unui fișier
- *download* – metoda folosită la descărcarea unui fișier

4. Amazon S3 vs. Microsoft Azure vs. Google Cloud

4.1 Amazon Web Services

Amazon Web Services - Cu un set vast de instrumente care continuă să crească exponențial, capacitățile Amazon nu sunt de neegalat. Cu toate acestea, structura sa de costuri poate fi confuză, iar accentul său special pe cloud public, mai degrabă decât cloud hibrid sau cloud privat înseamnă că interoperarea cu centrul de date nu este prioritatea principală a AWS.

4.2 Microsoft Azure

Microsoft Azure - Un concurent apropiat de AWS cu o infrastructură cloud capabilă excepțională. Dacă sunteți un client de întreprindere, Azure vorbește limba dvs. - puține companii au fondul întreprinderii (și suport pentru Windows) ca Microsoft. Azure știe că rulați încă un centru de date, iar platforma Azure lucrează din greu pentru a interopera cu centrele de date; nor hibrid este o adevărată tărie.

4.3 Google Cloud

Google Cloud - un underdog bine finanțat în competiție, Google a intrat mai târziu pe piața cloud și nu are accentul pe întreprinderi care ajută la atragerea clienților corporativi. Dar expertiza sa tehnică este profundă, iar instrumentele sale lider în industrie în învățarea profundă și inteligența artificială, machine learning și analiza datelor sunt avantaje semnificative.

4.4 Comparația vitezei de încărcare și descărcare [6]

Service name	Transferred	Average speed	Time
Amazon S3	1,91 MB	172,5 kB/s	11,1 s
	650 kB	72,5 kB/s	9,13 s
Google Storage	1,91 MB	168,8 kB/s	11,3 s
	650 kB	69,2 kB/s	9,57 s
Windows Azure	1,91 MB	217,5 kB/s	9,1 s
	650 kB	94,8 kB/s	7,04 s

Tabel 1: Rezultatele testului vitezei de descărcare de către CloudHarmony [6]

La început, viteza de descărcare a acestor trei servicii de stocare a fost testată (în aceeași zi cu următoarea testare). Testul de viteză în cloud conform CloudHarmony [5] a fost câștigat de Windows Azure Storage - vezi Tabelul 1. Aceste rezultate vor fi comparate în secțiunea 5 cu rezultate obținute de noi. [6]

Viteza de încărcare și descărcare a fost măsurată cu diferite fișiere de date de diferite dimensiuni pentru toate cele 3 CSP-uri. Dimensiunile fișierelor: 1) dimensiunea fișierului: 3,57 MB; 2) dimensiunea fișierului: 120 kB; 3) 16 fișiere cu dimensiuni între 100 kB și 500 kB: însumă 3,57 MB și 4) 61 fișiere cu dimensiunea până la 5 kB: însumă 120 kB. Vitezele de transmisie au fost măsurate simultan de la două locații pe oră, de la 10:00 la 16:00 (în timpul zilei de lucru).

Facultatea de Automatică și Calculatoare Universitatea Politehnica din București

Rezultatele primelor două măsurători sunt prezentate în Tabelul 2 și Tabelul 3.

Testing CSP from location / time of test			Test 1 – 10:00	Test 2 – 11:00	Test 3 – 12:00	Test 4 – 13:00	Test 5 – 14:00	Test 6 – 15:00	Test 7 – 16:00
Upload speed	Location 1	Amazon	146,5 s; Ø 26,1 kB/s	145,2 s; Ø 26,3 kB/s	150,6 s; Ø 25,3 kB/s	149,3 s; Ø 25,5 kB/s	148,9 s; Ø 25,6 kB/s	149,1 s; Ø 25,5 kB/s	149,8 s; Ø 25,4 kB/s
		Google	140,5 s; Ø 26,6 kB/s	139,8 s; Ø 26,8 kB/s	140,6 s; Ø 26,7 kB/s	141,3 s; Ø 26,5 kB/s	142 s; Ø 26,3 kB/s	141,8 s; Ø 26,4 kB/s	141,9 s; Ø 26,4 kB/s
		Windows	137,9 s; Ø 27,3 kB/s	138,2 s; Ø 27,1 kB/s	138,1 s; Ø 27,1 kB/s	139 s; Ø 26,9 kB/s	138,1 s; Ø 27,1 kB/s	138,4 s; Ø 27 kB/s	138,5 s; Ø 27 kB/s
	Location 2	Amazon	82 s; Ø 46,6 kB/s	81,6 s; Ø 46,7 kB/s	81,7 s; Ø 46,7 kB/s	82 s; Ø 46,6 kB/s	82,7 s; Ø 46,1 kB/s	82,9 s; Ø 46 kB/s	83,2 s; Ø 45,8 kB/s
		Google	75,1 s; Ø 51,3 kB/s	74,9 s; Ø 51,5 kB/s	75,6 s; Ø 51 kB/s	76 s; Ø 50,7 kB/s	76,3 s; Ø 50,4 kB/s	76,9 s; Ø 50 kB/s	77,3 s; Ø 49,8 kB/s
		Windows	75,6 s; Ø 50,6 kB/s	76,5 s; Ø 49,9 kB/s	75,9 s; Ø 50,7 kB/s	76,1 s; Ø 50,5 kB/s	77,4 s; Ø 49,7 kB/s	77,6 s; Ø 49,6 kB/s	77,8 s; Ø 49,4 kB/s
Download speed	Location 1	Amazon	12,7 s; Ø 302,9 kB/s	12,9 s; Ø 301 kB/s	12,8 s; Ø 301,5 kB/s	12,8 s; Ø 301,4 kB/s	13 s; Ø 295,4 kB/s	13,3 s; Ø 289 kB/s	13,2 s; Ø 291,1 kB/s
		Google	12,8 s; Ø 301,1 kB/s	12,6 s; Ø 306,2 kB/s	12,4 s; Ø 310 kB/s	12,5 s; Ø 308,9 kB/s	12,7 s; Ø 302,8 kB/s	13 s; Ø 295,6 kB/s	13,1 s; Ø 293,5 kB/s
		Windows	12 s; Ø 317,7 kB/s	12,2 s; Ø 313,8 kB/s	12,1 s; Ø 316,9 kB/s	12 s; Ø 317,8 kB/s	12,1 s; Ø 318,6 kB/s	12,5 s; Ø 308,3 kB/s	12,9 s; Ø 300,9 kB/s
	Location 2	Amazon	8,1 s; Ø 471,6 kB/s	8 s; Ø 478,1 kB/s	8,3 s; Ø 460,6 kB/s	8,3 s; Ø 460,7 kB/s	8,5 s; Ø 451,8 kB/s	8,7 s; Ø 442,4 kB/s	8,9 s; Ø 430,5 kB/s
		Google	8,1 s; Ø 471,5 kB/s	8,4 s; Ø 455 kB/s	7,9 s; Ø 483,3 kB/s	8,1 s; Ø 471,3 kB/s	8,2 s; Ø 466,1 kB/s	8,3 s; Ø 460,5 kB/s	8,4 s; Ø 454,7 kB/s
		Windows	7,5 s; Ø 509,4 kB/s	7,8 s; Ø 489,8 kB/s	7,7 s; Ø 496,2 kB/s	7,8 s; Ø 489,4 kB/s	8,2 s; Ø 465,9 kB/s	8,1 s; Ø 471,3 kB/s	8,3 s; Ø 460 kB/s

Tabel 2. Rezultatele vitezei și timpul de încărcare și descărcare - dimensiunea fișierului de 3,57 MB [6]

Testing CSP from location / time of test			Test 1 – 10:00	Test 2 – 11:00	Test 3 – 12:00	Test 4 – 13:00	Test 5 – 14:00	Test 6 – 15:00	Test 7 – 16:00
Upload speed	Location 1	Amazon	5,6 s; Ø 24,6 kB/s	5,4 s; Ø 25,1 kB/s	5,2 s; Ø 26,1 kB/s	5,3 s; Ø 25,6 kB/s	6,1 s; Ø 22,5 kB/s	6,7 s; Ø 20,5 kB/s	6,9 s; Ø 19,8 kB/s
		Google	6,1 s; Ø 22,4 kB/s	5,9 s; Ø 23,2 kB/s	5,6 s; Ø 24,7 kB/s	5,9 s; Ø 23,3 kB/s	6,4 s; Ø 21,5 kB/s	6,6 s; Ø 20,8 kB/s	6,7 s; Ø 20,6 kB/s
		Windows	5,1 s; Ø 26,5 kB/s	5,2 s; Ø 26 kB/s	5 s; Ø 27,1 kB/s	5,1 s; Ø 26,6 kB/s	5,2 s; Ø 26,1 kB/s	5,1 s; Ø 26,5 kB/s	5,3 s; Ø 25,5 kB/s
	Location 2	Amazon	3,8 s; Ø 36,1 kB/s	3,9 s; Ø 35,2 kB/s	4 s; Ø 34,7 kB/s	3,9 s; Ø 35,1 kB/s	4,1 s; Ø 33,9 kB/s	4,2 s; Ø 32,6 kB/s	4,4 s; Ø 31,2 kB/s
		Google	4 s; Ø 34,8 kB/s	4,1 s; Ø 33,9 kB/s	3,7 s; Ø 38,7 kB/s	3,9 s; Ø 35,2 kB/s	4,2 s; Ø 32,6 kB/s	4,3 s; Ø 32 kB/s	4,2 s; Ø 32,7 kB/s
		Windows	3,2 s; Ø 42,7 kB/s	3,3 s; Ø 41,5 kB/s	3,1 s; Ø 44,1 kB/s	3,2 s; Ø 42,7 kB/s	3,2 s; Ø 42,8 kB/s	3,5 s; Ø 39,1 kB/s	3,7 s; Ø 38,8 kB/s
Download speed	Location 1	Amazon	1 s; Ø 138,6 kB/s	1 s; Ø 138,5 kB/s	1 s; Ø 138 kB/s	1,1 s; Ø 134,5 kB/s	1,1 s; Ø 133,9 kB/s	1,2 s; Ø 123,2 kB/s	1,1 s; Ø 132,3 kB/s
		Google	0,9 s; Ø 158,5 kB/s	1 s; Ø 138,8 kB/s	0,9 s; Ø 157,9 kB/s	1 s; Ø 138,5 kB/s	1,1 s; Ø 133,8 kB/s	1 s; Ø 137,9 kB/s	1 s; Ø 137,2 kB/s
		Windows	1 s; Ø 138,7 kB/s	1 s; Ø 138,6 kB/s	0,9 s; Ø 158,5 kB/s	1 s; Ø 139 kB/s	0,9 s; Ø 158,7 kB/s	1 s; Ø 137 kB/s	1 s; Ø 136,9 kB/s
	Location 2	Amazon	0,8 s; Ø 208,7 kB/s	0,8 s; Ø 207,4 kB/s	0,7 s; Ø 235,3 kB/s	0,7 s; Ø 233,4 kB/s	0,8 s; Ø 206,7 kB/s	0,8 s; Ø 205,4 kB/s	0,8 s; Ø 203,9 kB/s
		Google	0,7 s; Ø 233,8 kB/s	0,6 s; Ø 273,9 kB/s	0,6 s; Ø 273 kB/s	0,6 s; Ø 273,1 kB/s	0,7 s; Ø 233,6 kB/s	0,7 s; Ø 232,7 kB/s	0,8 s; Ø 205,3 kB/s
		Windows	0,6 s; Ø 274,5 kB/s	0,7 s; Ø 235,7 kB/s	0,6 s; Ø 272,3 kB/s	0,6 s; Ø 273,9 kB/s	0,7 s; Ø 233,4 kB/s	0,7 s; Ø 232,9 kB/s	0,7 s; Ø 233 kB/s

Tabelul 3. Rezultatele vitezei și timpul de încărcare și descărcare - dimensiunea fișierului de 120 kB [6]

Figurile 1 și 2 arată progresul vitezei de încărcare și descărcare (ambele cu fișierul de date de 3,57 MB) între orele 10:00 - 16:00. În timp ce viteza de încărcare este aproximativ constantă, viteza de descărcare scade lent după-amiaza. Este aparent afectată de sfârșitul orelor de lucru în Europa, când, de exemplu, angajații fac backup de date în stocarea în cloud. Acesta arată cât de scalabile sunt serviciile furnizorului de stocare în cloud. O altă constatare este că viteza de încărcare de la locația 2 la stocarea în cloud Amazon este semnificativ mai mică în comparație cu Google și Windows - a se vedea Fig. 1.

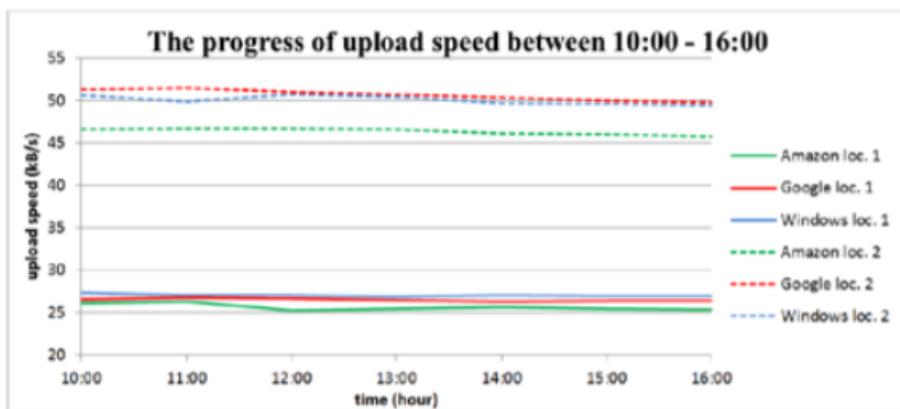


Fig.1. Progresul vitezei de încărcare între 10:00 – 16:00 [6]

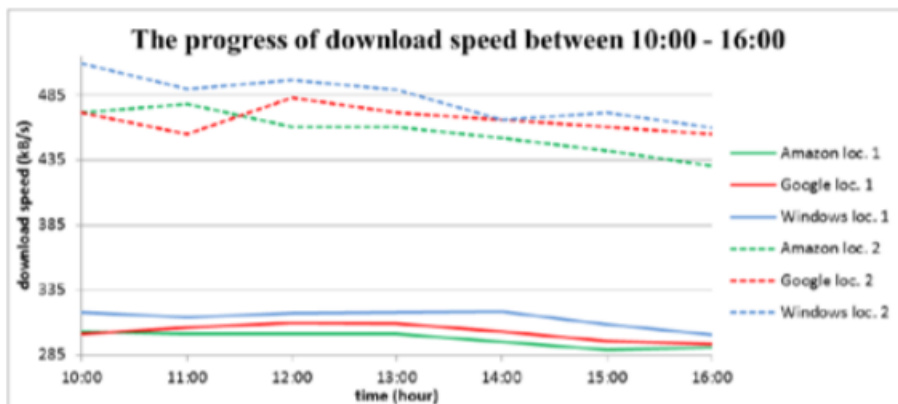


Fig.2. Progresul vitezei de descărcare între 10:00 – 16:00 [6]

Tabelul următor, tabelul 4, prezintă rezultatele timpului mediu de încărcare și descărcare în timp ce lucrați cu date. Durata medie de încărcare a unui singur fișier (dimensiunea de 3,57 MB) este mai scurtă în comparație cu folderul de aceeași dimensiune care conține 16 fișiere. Este și mai evident când se folosește al doilea fișier (dimensiunea de 120 kB). Timpul de încărcare al folderului de aceeași dimensiune de 61 de fișiere este de două ori mai lung decât fișierul unic. Din această comparație, este clar că platforma Cloud a Microsoft - Windows Azure este pentru fișierele de date mai potrivite decât Google și Amazon.

	Amazon loc. 1	Google loc. 1	Windows loc. 1	Amazon loc. 2	Google loc. 2	Windows loc. 2
Average upload time of 3,57 MB GIS file	148,49 s	141,13 s	138,31 s	82,30 s	76,01 s	76,70 s
Average download time of 3,57 MB GIS file	12,96 s	12,73 s	12,26 s	8,40 s	8,20 s	7,91 s
Average upload time of 120 kB GIS file	5,89 s	6,17 s	5,14 s	4,04 s	4,06 s	3,31 s
Average download time of 120 kB GIS file	1,07 s	0,99 s	0,97 s	0,77 s	0,67 s	0,66 s
Average upload time of 16 GIS files (in sum 3,57 MB)	152,20 s	151,93 s	144,93 s	81,74 s	82,04 s	80,57 s
Average download time of 16 GIS files (in sum 3,57 MB)	13,11 s	12,97 s	11,40 s	8,40 s	8,20 s	7,84 s
Average upload time of 61 GIS files (in sum 120 kB)	12,21 s	11,77 s	11,60 s	6,87 s	6,70 s	6,64 s
Average download time of 61 GIS files (in sum 120 kB)	6,47 s	6,26 s	5,79 s	2,31 s	2,24 s	2,21 s

Tabel 4. Comparația timpului mediu de încărcare și descărcare [6]

Concluzia

Această lucrare este doar unul dintre pașii de soluționare a problemei datelor din cloud. Pentru a obține rezultate comparabile, este necesar să se testeze stocarea în cloud din mai multe locații din întreaga lume, de asemenea cu alte condiții inițiale. Returnarea costurilor de investiții și de implementare nu trebuie să fie uitate. A fost făcut un test de încărcare / descărcare pentru a vedea cât de repede este posibil să încărcați și să descărcați date în / din stocarea în cloud. În acest caz, Microsoft cu Windows Azure a fost evaluat ca cea mai bună soluție.

Rezultatele sunt, de asemenea, de acord cu viteza de descărcare a testului de către CloudHarmony [7]. Utilizatorilor nu le place să aștepte, timpul înseamnă bani, astfel încât viteza și timpul de răspuns sunt importante. Cu toate acestea, stocarea în cloud nu este în siguranță pentru datele de afaceri sensibile (strategice) de rezervă. Dar poate fi folosit pentru schimbul de informații cu partenerii de afaceri. Depozitele cloud testate nu sunt încă adecvate pentru aplicațiile de procesare grea a tranzacțiilor. Conform studiului de caz, viteza de încărcare este relativ scăzută pentru toate soluțiile testate. Cu toate acestea, alternativa promițătoare este să descarci date din mai multe locații.

5. Evaluarea rezultatelor proprii obținute

Având soluția implementată am testat timpii necesari operațiilor de upload și download pe mai multe fișiere.

Dacă în prima fază a proiectului am încărcat un simplu fișier de 1MB pentru a testa faptul că fișierul ajunge în bucketul creat de pe AWS, acum am dorit să analizăm timpii de încărcare/descărcare pe diferite fișiere ca mărimi.

Am folosit fișiere de 1MB, 32MB, 64MB, 128MB, 526MB, 1GB, dar și fișiere de 1.91MB, 650kB, 3.57MB și 120kB pentru a face o comparație cu timpii de încărcare/descărcare pe care i-au obținut autorii documentului de la punctul 4 pe ceste fișiere ca dimensiune.

Testele au fost realizate, având o conexiune la internet de 255.9 Mbps pe upload și 150.4 Mbps pe download (DigiNet, pe WiFi).

De asemenea pentru fiecare fișier am efectuat de 5 ori atât operația de upload, cât și pe cea de download, și am făcut o medie.

UPLOAD						Medie
	1	2	3	4	5	
1 MB	1.31	0.53	0.50	0.45	0.38	0.63
32 MB	3.23	1.84	1.87	2.12	2.12	2.23
64 MB	4.81	5.04	5.39	5.89	6.74	5.57
128 MB	8.55	10.39	9.36	20.52	35.14	16.79
256 MB	25.40	92.90	18.70	23.68	20.64	36.26
1.91MB	1.65	0.39	2.36	0.57	2.36	1.46
650KB	0.41	0.36	0.45	0.40	0.40	0.40
3.57MB	0.54	0.52	0.64	0.50	0.60	0.56
120KB	0.41	0.45	0.36	0.45	0.40	0.41

Tabel 5. Timpii de upload folosind Amazon S3

DOWNLOAD						Medie
	1	2	3	4	5	
1 MB	1.11	0.75	0.62	0.84	0.82	0.83
32 MB	3.81	5.15	5.01	4.90	7.70	5.31
64 MB	9.39	11.52	15.29	11.94	10.39	11.71
128 MB	16.17	14.83	12.14	11.76	13.98	13.78
256 MB	25.29	22.55	21.80	19.52	21.96	22.22
1.91MB	1.08	1.85	1.64	1.53	1.63	1.55
650KB	1.09	0.78	0.67	0.62	0.48	0.73
3.57MB	1.71	2.87	1.80	3.28	2.46	2.43
120KB	0.54	0.35	0.39	0.40	0.38	0.41

Tabel 6. Timpii de download folosind Amazon S3

Tabelele 5 și 6 prezintă rezultatele obținute în testarea cu soluția implementată de noi.

Primul lucru observat, evident, este faptul că cu cât fișierele ce sunt încărcate/descărcate sunt mai mari cu atât și timpul necesar operației este mai mare.

Un al doilea lucru ce ne-a surprins au fost câteva spike-uri în ceea ce privește uploadul aceluiași fișier. Posibil acest lucru a fost cauzat și de faptul că în timp ce aplicația se executa, fluxul de trafic de date s-a modificat, fiind situat, eu ca persoana, în momentul testului destul de departe de sursa de WiFi.

Dimensiunile fișierelor au fost împărțite în două (despărțite printr-o linie mai îngroșată). Partea de sus prin care am încercat să observ creșterea timpului de upload/download și partea de jos unde am încercat operațiile de upload/download pe fișiere de dimensiunile folosite de autorii lucrării 6 din bibliografie.

În ceea ce privește fișierele de 1.91MB și 650kB timpii scoși de noi au fost mai buni față de cei scoși în articol [6], însă acest lucru se explică prin faptul că, după cum este menționat și în articol viteza acestora a fost de 172.5 kB/s, în timp ce viteza noastră a fost de 255.9Mbps (~31.97MB/s).

Din ce am observat pentru fișierele de 1MB, 32MB, 64MB, 1.91MB, 650kB, 3.57MB timpii de upload au fost mai mici decât cei de download, însă în rândul fișierelor de dimensiuni mari (128MB și 256MB operația de download se pare că a fost mai rapidă). Surprinzător, după 5 operații repetitive de upload și 5 operații repetitive de download, media obținută pentru 120KB a fost aceeași și anume 0.41 s.

6. Concluzie

În concluzie, proiectul este unul complex și interesant, ce pune în contact două materii care fac subiectul domeniului de masterat (Big Data și Cloud Computing) și care ne prezintă direcția tehnologică de astăzi (procesarea în cloud, performanța serviciilor, volumul mare de date, gestionarea cu ușurință a acestui volum mare de date și a controalelor de acces).

Din câte s-a putut observa și din rezultatele noastre timpii de încărcare/descărcare pentru Amazon S3 sunt destui de buni, însă conform articolului câștigătorul a fost Azure între cele 3 platforme.

7. Bibliografie

- [1] Amazon Web Services, Inc. and/or its affiliates, “*Amazon Simple Storage Service - Developer Guide*”, API Version 2006-03-01, 2019
- [2] A.McAfee, E.Brynjolfsson, “*Big Data: The Management Revolution*”, Harvard Business Review, October 2012
- [3] B.Hayes, “*Cloud Computing*”, News Technology, DOI: 10.1145/1364782.1364786
- [4] Amazon.com, Inc., “*Boto3 Documentation*”, Release 1.1.4, September 28, 2015
- [5] M.Pecherskiy, “*Intro to AWS and Boto3 – Introduction to boto in python*”, Data Camp Tutorials
- [6] M.Lnenicka, J.Komarkova, E.Milkova, “*Performance Testing of Cloud Storage while Using Spatial Data*”
- [7] Cloud Speed Test. CloudHarmony:transparency for the cloud [online]. ©2011 [cit.2012-08-24]. Available from: <http://cloudharmony.com/speedtest>
- [8] C.Harvey, A.Patrizio, “*AWS vs. Azure vs. Google: Cloud Comparison [2019 Update]*”, January 22, 2019, Available from: <https://www.datamation.com/cloud-computing/aws-vs-azure-vs-google-cloud-comparison.html?fbclid=IwAR3gVU8hbzEzkigyp3CFRYlqDHqkCbqhBcUFChRfDfKcYC5pVbwP8rjDHEA>