

Big Data Texnologiyalarında İstifadə Olunan Proqram Təminatı

Aynurə Sadıqova

AMEA, İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
aynura.sadiqova.24@gmail.com

Xülasə— Məqalə böyük verilənlərin emalı problemlərinə həsr olunur və böyük verilənlərdə istifadə olunan bəzi proqram platformaları təhlil olunur. Eyni zamanda böyük həcmli verilənlərin analitikasının mövcud problemləri araşdırılır.

Açar sözlər— Big Data, Hadoop, Apache Pig/Hive, MongoDB, Big Data Analytics

I. GİRİŞ

Bu gün bir çox təşkilatlarda böyük həcmli informasiyanın kütləvi saxlanması, emalı və toplanması əməliyyatları həyata keçirilir. Big Data günümüzdə son dərəcədə məşhur bir anlayış halına gəlmişdir və yeni bir dövrün başlanğıcı kimi şərh olunur. "Big Data" ingilis dilindən tərcümədə “böyük verilənlər” (BV) kimi başa düşülür. Big Data – müxtəlif təbiətli böyük həcmli verilənlərin emalı, analizi və onlardan faydalı biliklərin çıxarılması sahəsində tədqiqatlar üçün istifadə olunan aparıcı istiqamətlərdən biridir [1].

Big Data məlumatların idarə olunması nəticəsində qərarların qəbulunu dəstəkləyən yeni nəsil sistemlərin yaradılmasında istifadə olunur. Verilənlərin analizində istifadə olunan reqressiya təhlili, təsnifatlandırma, maşın təlimi və s. analitik üsullar uzun illərdir ki, mövcuddur. Lakin, kompüter texnologiyalarının inkişafı ilə verilənlərin analizində yeni mənbələrdən və proqram təminatlarından (PT) istifadə etmək tələbləri yarandı. Bu proses Big Data analitikaya olan marağa da təsir etdi. Emal olunmamış verilənlərin həcmnin artması onların real vaxtda analizinin zəruriliyi ilə birlikdə Big Data Analytics məsələsini effektiv həll etməyə imkan verən alətlərin yaradılması və tətbiqini zəruri edir [2]. Verilənlərin analitikası Big Data-nın bir sahəsi olub informasiya texnologiyaları (İT) və proseslərini əhatə edən "Data Science" (verilənlər elmi) anlayışını yaratmışdır. Data science sahəsinin tədqiqat obyektləri böyük verilənlər və verilənlərdən ümumiləşdirilmiş biliyin çıxarılmasına istiqamətlənmiş informatika, riyaziyyat, sosial elmlər, şəbəkə elmi, iqtisadiyyat və s. daxil olmaqla bir çox fənlərin inteqrasiyasından əldə olunan informasiyadır [3].

Big Data analitika təşkilatlarda toplanan verilənlərin strukturlaşdırılması və effektiv analizi nəticəsində əldə olunan bilikdən istifadə etməklə şirkətlərin biznes vəziyyəti haqqında fikir əldə etməyə, riskləri qiymətləndirməyə və gələcəkdə baş verə biləcək hadisələrlə bağlı proqnozlar verməyə kömək edir. Odur ki, Big Data platformalarının düzgün seçilməsi və səmərəli istifadəsi şirkətlər qarşısında duran əsas məsələdir. Məqalədə məqsəd mövcud Big Data platformalarını

araşdırmaq və onların funksionallıqları haqqında məlumat əldə etməkdir.

II. BIG DATA-NIN PROQRAM PLATFORMALARI

Big Data ənənəvi verilənlər bazası (VB) və PT-dan istifadə etməklə emalı çətin olan məlumatların strukturlaşdırılmış və strukturlaşdırılmamış şəkildə istinadı üçün istifadə edilir. Son illər nəhəng şirkətlərdə böyük həcmdə verilənlər toplanmaqdadır. Onların sürətli emalına və etibarlı saxlanmasına ciddi tələblər qoyulur.

Big Data texnologiyalarından düzgün istifadə şirkətlərin işində effektivliyi artırmaq, daha sürətli və asan qərar qəbul etməni mümkün edir. Big Data texnologiyaları verilənlər üzərində manipulyasiya işləri, onların saxlanması, analizi, düzgün təqdimatı üçün şirkətlərə köməklik edir. Bunun nəticəsində şirkətlər yeni müştərilər qazanmaqla öz fəaliyyətlərini yaxşılaşdırmaq, riskləri müəyyənləşdirmək, gəlirlərini artırmaq, rəqabətdə üstün olmaq və s. kimi imkanlar əldə edə bilirlər.

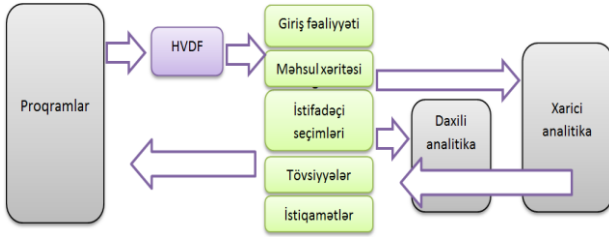
Big Data texnologiyalarından bəziləri nəhəng şirkətlər tərəfindən geniş istifadə edilərək özünü doğrultmuşdur. Böyük həcmli verilənlərin analizi ilə bağlı mövcud problemləri aradan qaldırmaq üçün Big Data texnologiyaların inkişaf etdirilməsi davam edir. Bu texnologiyalardan bəziləri aşağıda qeyd edilmişdir:

- **Apache Pig/Hive** texnologiyaları Map-Reduce işlərinin daha üst səviyyədə yazılmasına imkan verir [4]. Pig-in kodlama dili olan “Pig Latin” özünə xas yazı qaydalarına malikdir. Hive kodlama dili olan HiveQL ilə yazılan sorğular, SQL yazı qaydalarına müəyyən dərəcədə oxşayır. Bu dillərdə yazılan skript böyük və mürəkkəb olan Java dilində (və ya digər dillərdə) yazılan Map-Reduce kodları ilə müqayisədə daha sadə və kiçikdir. Pig və Hive-də yazılan bu skriptlər müəyyən optimallaşdırma alqoritmindən keçdikdən sonra Map-Reduce məlumatlarını tənzimləyir.

- **MongoDB** böyük məlumatların gizli bir şəkildə saxlanıldığı NoSQL VB-dır [5]. Ümumi olaraq Hadoop tərəfindən işlənən məlumatlar MongoDB NoSQL bazasında saxlanılır. VB-ların əlaqəli fəaliyyəti məlumatları saxlamağa və sürətli bir şəkildə daxil edilməsinə şərait yaradır (şəkil 1).

- **Apache Hadoop** Apache Software Foundation-un layihəsi olmaqla daha geniş yayılmış texnologiyadır, BV-nin de-fokto standartı hesab olunur və paylanmış hesablama

mühitində BV-nin emalı və analizi üçün əsas platformadır [2]. Bu texnologiya vasitəsilə 2009-cü ildə 1 terabayt həcmdə məlumat Yahoo tərəfindən Hadoop texnologiyası istifadə olunaraq 62 saniyədə sıralanmışdır.



Şəkil 1. MongoDB texnologiyasının arxitekturası

2013-cü ildə eyni sıralama əməliyyatı 1,42 terabayt həcmdə məlumatlar üzərində ortalama 60 saniyədə reallaşmışdır. Hadoop tərəfindən analiz ediləcək məlumatlar HDFS (Hadoop File System) üzərində saxlanılmaqdadır. Hadoop ümumi olaraq birdən çox kompüterin yaratdığı qruplar (cluster) şəklindədir. Bu həm məlumatın həm də işlərin paylanmasına imkan verir [6-7].

Big Data texnologiyaları müxtəlif proqram təminatları və analitik sistemlər əsasında işləyir.

III. BİG DATA ANALİTİKA VASİTƏLƏRİ

Big Data proqram platformaları və Big Data analitik sistemlərinin əsas işi böyük həcmli verilənlər üçün səmərəli analitikanın təmininə yönəlmişdir. Aşağıda bəzi nəhəng şirkətlərdə Big Data imkanlarından istifadə məsələləri araşdırılmışdır.

A. IBM-in Big Data analitika vasitələri

Texnologiya və konsalting sahəsində ixtisaslaşdırılmış transmilli ABŞ şirkəti IBM (International Business Machines) yeni nəsil analitik həll yolları təklif edir. Məsələn, IBM Big Data analitika sistemi müəssisələrdə toplanan müxtəlif tipli məlumatların analizini həyata keçirir. Big Data daxil olmaqla birdən çox platformada və cihazda hər növ verilənlər haqqında sürətli və asan bir şəkildə məlumat əldə etmək mümkündür. Ayrıca bu sistem vasitəsilə öz-özünə xidmət, ekspertiza və qarşıya qoyulan məsələlərin səmərəli həllinə və tələblərə cavab verən intellektual qərar qəbulu mümkündür. IBM xərclərin azaldılması, risklərin qiymətləndirilməsi məqsədi ilə proqram variantları təklif edir. IBM-in Big Data analitika sistemi bulud xidmətləri (IBM Cloud) də təklif edir.

Bulud əsaslı bələdçi analitika vizuallaşdırma və proqnozlaşdırma vasitəsilə böyük həcmli verilənlərin asan başa düşülməsinə imkan yaradır. IBM-in Big Data analitika sisteminin istifadə etdiyi proqram platformalarından biri Watson Analytics-dır. Watson Analytics bulud üzərində intellektual verilənlərin aşkar edilməsi xidmətini təklif edir. Verilənlərin axtarışı, avtomatik proqnozlaşdırılması və idarəetmə panelində infografiklərin (infographics) yaradılmasına imkan verir. Watson Data proqram platforması mütəxəssisləri tez bir zamanda yeni anlayışları sadə üsullar ilə tapmağa imkan verir [8].

Watson Data proqram platforması koqnitiv dəstəklə maşın təlimi modeli əsasında işləyir. Platforma istifadəçiyə IBM

Cloud-a dair xüsusi alətlər, məlumatlar və analitik xidmətlər göstərir. Watson Data platforması üç əsas yanaşma üsulundan ibarət olmaqla - platforma, metodlar və ekosistem ilə birlikdə müəssisələrdə inkişafı sürətləndirmək mümkündür.

Watson Data platforması bu əsas imkanları təklif edir:

- xüsusi istifadəçi alətləri;
- bütün müəssisələr arasında əməkdaşlığı təmin edən zəngin mühit;
- proqram platforması daxilində məlumatların sürətli axını;
- idrak dəstəklə maşın öyrənmə və Apache Spark ilə avtomatlaşdırılmış yerləşmə;
- açıq ekosistem çərçivəsində müxtəlif texnologiyaların seçimi.

B. Google-nin Big Data analitika vasitələri

7) Google Cloud Google axtarış brauzeri və Youtube kimi saytların da istifadə etdiyi server infrastruktur xidmətlərini istifadəçiyə təqdim edən bulud informasiya platformasıdır. Google Cloud Platforması, server inkişaf vasitələri ilə birlikdə bulud əsaslı xidmətlərlə işləyən və onları təqdim edən Google-un təşkilatı həllər paketinin bir hissəsidir. Müəssisə daxilində müxtəlif mənbələrdən əldə olunan məlumatların böyük həcmi generasiya olunaraq istifadəçilərə ötürülür. Generasiya olunan məlumatların həcmi günü-gündən artmaqdadır. Bu əməliyyatların nəticəsində ətraf mühit və müştərilər haqqında effektiv məlumat əldə etmək mümkündür.

Big Query Google-un təhlili üçün tam idarə olunan, petabayt miqyasda, aşağı qiymətli təşkilatı məlumatlar bazasıdır. BigQuery aparıcı infrastrukturaya malikdir [9]. BigQuery verilənlər bazasında SQL-dən istifadə edərək məlumatlar əldə etmək, analiz etmək mümkündür. BigQuery ilk olaraq Fortune 500 kimi şirkətlərdə istifadə edilən güclü bir Big Data analitik platformadır və aşağıdakı xüsusiyyətlərə malikdir:

- *Sürət və miqyas*, BigQuery platformu məlumatları bir neçə saniyə içində nəzərdən keçirə bilir. Məlumatı real vaxt ərzində təhlil edə bilmək üçün verilənləri Google Cloud Storage-a yüklərək BigQuery'yə köçürmək əlverişlidir.
- *Qiymət*, BigQuery, saxlama və hesablama anlayışlarından asılı olaraq istifadəsi ödənişlidir. Ehtiyacları daha yaxşı qarşılamaq üçün uyğun qiymət variantları təqdim edir.
- *Təhlükəsizlik və etibarlılıq*, BigQuery təhlükəsizlik, praktiklik və dayanıqlılığını təmin etmək üçün məlumatları avtomatik olaraq şifrələyir və çoxaldır. Məlumatları Google Cloud ID idarə sistemindən istifadə edərək qurmaq və giriş nəzarət qurğusu ilə qorumaq mümkündür.

NƏTİCƏ

Hazırda Big Data-nın proqram platformalarından BV-lə işləyən bütün sferalarda istifadə edilir. Big Data texnologiyaları vasitəsilə istifadəçi davranışlarının dərin təhlili,

modelləşdirilməsi dövlət və biznes şirkətlərinin vətəndaşlara daha yaxşı xidmət göstərməsini və marketinqi təmin edə bilər.

BV-nin emalı və analizi üçün tətbiq edilən Map Reduce, Hadoop, HDFS, NoSQL, MongoDB və s. proqram-aparat platformaları inkişaf etdirilməkdədir. Bu texnologiyaların köməyi ilə müəyyən problemlər aradan qaldırılmış olsada hələ də bəzi texniki və elmi-nəzəri problemlər qalmaqdadır. Big Data texnologiyasının böyük potensialını və müxtəlif sahələrə ciddi təsirini nəzərə alaraq düşünülmüş Big Data strategiyasının işlənməsi xüsusi əhəmiyyət daşıyır.

ƏDƏBİYYAT

- [1]. Y.N. İmamverdiyev. “Big Data texnologiyalarının böyük perspektivləri və problemləri”, İnformasiya cəmiyyəti problemləri, №1, 23–3, 2016.
- [2]. R.T. Qasımova “BIG DATA analitikası: mövcud yanaşmalar, problemlər və həllər” İnformasiya texnologiyaları problemləri, №1, 75–93, 2016.
- [3]. R.M. Əliquliyev, M.Ş. Hacırahimova, A.S.Əliyeva. “Big Data-nin aktual elmi-nəzəri problemləri” İnformasiya cəmiyyəti problemləri, №2, 37–49, 2016.
- [4]. S. Dhawan, S. Rathee. “Big Data Analytics using Hadoop Components like Pig and Hive” American International Journal of Research in Science Technology, Engineering & Mathematics, vol. 13, no. 131, p.88–93, 2013.
- [5]. N.Q.Mehmood, R.Culmone, L.Mostarda. “Modeling temporal aspects of sensor data for MongoDB NoSQL database” Journal of Big Data, vol. 4, no. 8, p.13–19, 2017.
- [6]. V. Pellakuri, Dr.D. R. Rao. “Hadoop Mapreduce Framework in Big Data Analytics” International Journal of Computer Trends and Technology (IJCTT), vol. 8, no. 3, p.115–117, 2014.
- [7]. A. Murthy. “Moving Beyond MapReduce and Batch Processing with Apache Hadoop 2”, Addison Wesley Data & Analytics Series, 2014.
- [8]. “A New Platform for the Cognitive Business”, www.ibm.com/analytics/us/en/watson-data-platform/platform.html
- [9]. “Enterprise Cloud Data Warehouse”, www.cloud.google.com/bigquery/