

K.A.Məmmədova, Y.N.Əliyeva (Azərbaycan Dövlət Neft və Sənaye Universiteti)

BÖYÜK VERİLƏNLƏRİN ANALİTİK TƏHLİLİ

Giriş. Son 30 ildə verilənlərin həcmində böyük artım müşahidə olunmuş, 1990-cı illərdə verilənlərin həcmi terabaytlar vasitəsilə ölçülməyə başlanılmışdır. Strukturlaşmış verilənləri sıralar və sütunlar formasında göstərən verilənlər toplusu əlaqələri və verilən anbarlarda müəssisə məlumatlarını yadda saxlamaq və idarə etmək üçün səciyyəvi texnologiyadır. Növbəti onillikdə verilənlər, kontenti idarə olunan depolar və şəbəkə əlaqəli yaddaş sistemləri məhsuldar və nəşriyyat tərəfindən idarə olunan müxtəlif verilən mənbələri ilə məşğul olmağa başladı, nəticədə verilənlərin həcmi petabaytlar vasitəsilə ölçüldü.

2010-cu illər isə rəqəmsallaşdırılmış verilənlərin bir çox mənbəyi və dəyişikliklər əsasında idarə olunan eksponensial verilənlərin həcmi ilə başa çatdı. Böyük miqdarda verilənlər yaradan bəzi tətbiqlər meydana çıxmış, bu qədər verilənlərin yığılması nəticəsində eksabaytlar vasitəsilə ölçmələrə başlanılmışdır [1,2].

Verilənlərin çoxalmasına səbəb olan mənbələri verilənlərin generasiyasının bir çox kateqoriyalarına bölmək olar: mexanizm, insanın təsiri, data işlənilməsi.

Birinci tip generasiya mexanizmin rəqəmsallaşmasının yayılması ilə müəyyən olmuşdür ki, buraya sensorların inteqrasiyası, əlaqələrin çoxalması, səs, şəkil və videoyazan cihazlar, bir-biri ilə əlaqədə olan mexanizmlər daxildir. Xüsusilə, videoyazan kameralar, coğrafi data toplayan cib telefonları, sənaye sisteminin istehsal sahəsindəki mexanizmləri kimi bir sıra cihazlar öz işlərini yerinə yetirərkən vacib informasiyaların mübadiləsini aparırlar.

Sonrakı kateqoriya insanlar arasındakı informasiya mübadiləsi ilə əlaqədardır. Məsələn, Facebook, Twitter, LinkedIn və s. kimi sosial şəbəkələri buna nümunə göstərmək olar. Hər keçən saniyə ərzində bu sistemlər milyonlarla insan tərəfindən paylaşılan nəhəng verilənlər toplusu yaradır. Verilənlərin emalı hər hansı bir nəticə almaq üçün və ya müəyyən layihəyə daxil olan başqa bir proses mərhələsinə çatmaq üçün öncədən emal olunmuş və ya olunmamış verilənlərlə məşğul olmaqda çətinlik yaradır. Hal-hazırda çoxsaylı verilənlər yaradan bir çox mənbələr mövcuddur ki, bu mənbələrdən də gələcəkdə analizlər və araşdırmalar üçün istifadə edillə bilər [3].

Böyük verilənlərin yaranma səbəbləri. Böyük verilənlər müxtəlif sürətlərdə və qeyri-müəyyənliyin müxtəlif dərəcələrində yaradılmış mürəkkəbliyin fərqli dərəcələrində mövcud olan verilənlərin həcmi kimi təsvir oluna bilər. Böyük verilənlər ənənəvi texnologiyalardan, gedişat metodlarından, alqoritmlərdən və ya hər hansı əsas hazır həll yollarından istifadə etməklə emal oluna bilməz. Böyük verilənlərin digər tərifı McKinsey tərəfindən belə verilmişdir: “Böyük verilənlər elə verilənlərdir ki, onların ölçülməsi, paylaşılması, müxtəlifliyi və cariliyi yeni texnoloji arxitekturanın və analitikanın istifadəsini zəruri edir ki, bu da biznes dəyərinin yeni mənbələrini ortaya çıxaran anlayışların formalaşmasına imkan verir” [4,5]. Daha geniş yayılan və istifadə edilən tərif “3V” (volume, velocity, variety) kimi təsvir olunur (şəkl.1):

– volume (həcm) - verilənlər bazasından gələn, terabaytdan tutmuş zettabayta qədər həcmdə olan geniş verilənlər toplusu;

– velocity (sürət) - yüksək yenilənmə sürətilə həyata keçən verilənlərin axını sayəsində baş verən əməliyyatlar nəticəsində yığılan geniş həcmli verilənlər toplusu. Bu cür verilənlər axınının üzərində işləmək üçün vaxt həmişə çox qısa olur. Burada toplu emaldan real zaman axınına keçid etmək mümkündür;

– veracity (doğruluq) - uyğunsuzluq, natamamlıq, qeyri-müəyyənlik, gecikmə, yanılma və model təxminləşdirilməsinə istinadən verilənlərdə şübhə olur. Verilən izahlara əsasən böyük verilənlərin emalındakı mürəkkəbliyin səbəbini görmək olar. Həmçinin verilənlərin qeyri-müəyyənliyi, gecikməsi və virallığı da təsvir olunur.



Şəkil 1. 3V- böyük verilənlərin əsas xarakteristikaları

– variety (müxtəliflik) - verilənlər müxtəlif mənbələrdən gəlir, ilkin olaraq verilən həm daxili, həm də xarici mənbədən gələ bilər. Burada vacib olan verilənlər müxtəlif tətbiqlərdən əməliyyat və log data formasında, verilənlər bazasının cədvəli şəklində, strukturlaşdırılmış data, XML data şəklində, yarı strukturlaşdırılmış data, mətn, şəkil, video, səsli mesaj və s. kimi daxil ola bilər. Burada tək strukturlaşdırılmış data getdikcə daha çox strukturlaşdırılmamış dataya çevrilir, yaxud iki datanın birləşməsi baş verir, bəziləri isə bunu "4V" adlandırır [6].

– ambiguity (qeyri-müəyyənlik) - böyük verilənlərin daxilində az məlumat olduqda və ya heç bir informasiya olmadıqda ortaya çıxır;

– viscosity (gecikmə) - bu termin tez-tez göstərilən hadisə ilə əlaqəli olan verilənlərin vahid zamanda gecikməsini təsvir etmək üçün istifadə olunur;

– virality (sürət) - bir şəbəkəyə qoşulmuş insanlar arasında informasiyanın necə tez yayıldığını təsvir edir. Bu ölçünün nəticəsi informasiyanın müəyyən bir zaman ərzində yayılmasının sürətidir.

Mur qanununda [7,8] təsvir edilən hesablama gücünün eksponensial artımına görə verilənlərin miqdarı onların böyük olub-olmamasının dəqiq meyarı ola bilməz. Məsələn, bu gün böyük verilənlər terabaytlarla ölçülür, sabah isə petabaytlarla ölçülə bilər. Buna görə də Big Datanın əsas xüsusiyyəti onların strukturunun dərəcəsi və təqdimat variantlarıdır. Şəkil 2-də isə böyük



Şəkil 2. 5V - böyük verilənlərin əsas xarakteristikaları

verilənlərin 5V xarakteristikaları təsvir olunmuşdur:

Böyük və kiçik həcmli verilənlərin müqayisəli təhlili. Böyük verilənlər kiçik həcmli verilənlərin sadəcə böyümüş forması sayıla bilməz [1,2,3]. İddia edildiyi kimi gələcək verilənlər artımının təxminən 80-90%-i strukturlaşdırılmamış tip verilənlərdən əldə edilir (şəkil 3).

Strukturlaşdırılmış verilənlər öz formatını təyin edən strukturda saxlanılır. Adətən, strukturlaşdırılmış verilənlər saxlanılmasına (verilənlərin tipi - rəqəm, mətn, və.s.), emal edilməsinə, əldə edilməsinə, məhdudlaşdırılmasına (mümkün olan dəyərlər toplusu - kişi, qadın) görə təsnifatlaşdırılır. Strukturuna görə bu verilənlər asanlıqla saxlanıla, emal edilə və sorğulana bilər.

Yarı strukturlaşdırılmış verilənlər strukturlaşdırılmış verilənlərin bir növü olduğuna baxmayaraq, onun sərt model strukturu yoxdur (məsələn, XML).

Kvazi-strukturlaşdırılmış verilənlər ümumilikdə qeyd edilməməsinə baxmayaraq, bu qrup yarı-strukturlaşdırılmış və strukturlaşdırılmamış kateqoriyaları arasında olan qrup kimi əlavə edilə bilər.

Strukturlaşdırılmamış verilənlər asanlıqla təsnifatlaşdırıla bilmir, bu da onların strukturlaşdırılmış növün tam əksi olduğunu göstərir. Onlara xas olan hər hansı bir struktur yoxdur [8].

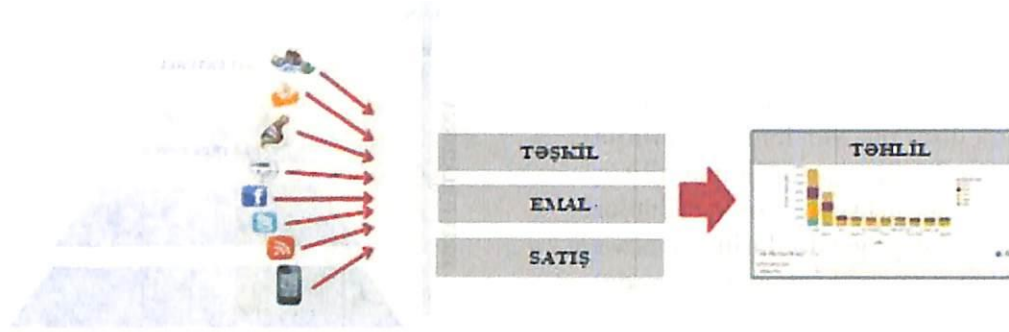
Böyük verilənlərin idarə edilməsi və analizi. Bu zaman ortaya çıxan ən vacib sual bu qədər geniş miqdarda verilənlərin necə saxlanması və emal edilməsidir; onların böyük hissəsi emal olunmamış (raw), yarı-strukturlaşdırılmış və strukturlaşdırılmamışdır. Böyük verilənlər platformaları onların necə miqyaslanmasından, nasazlığa dözümlü olmasından, səmərəli şəkildə saxlanılmasından və emal edilməsindən asılı olaraq təsnifatlaşdırılır [6,7,8,9]. Böyük verilənlərin idarə edilməsi üçün iki mühüm informasiya - idarəetmə üsulu sisteməlik iş yükü üçün təkmilləşdirilmiş, əlaqəli DBMS məhsulları (çox vaxt analitik RDBMS və ya ADBMS kimi tanınır) və raw, yarı-strukturlaşdırılmış və strukturlaşdırılmamış verilənlərin emalı üçün əlaqəsiz üsullardır (bəzən NOSQL sistemləri kimi tanınır). Qeyri-əlaqə üsulları böyük verilənlərdən statistik məlumatlar hazırlamaq və ya məlumat anbarında birləşdirilməmişdən əvvəl böyük məlumatı əvvəlcədən emal etmək üçün istifadə edilə bilər. Əlaqəsiz üsullardan böyük verilənlərdən statistika toplamaq və ya böyük informasiyanı məlumat anbarına birləşdirmədən əvvəl emal etmək üçün istifadə edilə bilər.

Böyük verilənlərin analizi əsasən analitik üsulları, sisteməlik arxitekturanı, intellektual analizi və proqram təminatını əhatə edir. Verilənlərin araşdırılması mənalı dəyərlərin tədqiq edilməsi, təkliflərin və qərarların verilməsi üçün böyük verilənlərdə ən vacib addımdır. Mümkün qiymətlər verilənlərin analizi əsasında tədqiq edilə bilər. Bununla belə, verilənlərin təhlili dinamik və çox mürəkkəb olan geniş bir sahədir.



Şəkil 3. Böyük datanın formatları [7]

Ənənəvi verilənlərin analizi çox böyük verilənlərin analizi üçün statistik metodlardan düzgün istifadə deməkdir. Burada məqsəd verilənlərin qiymətini maksimuma çatdırmaq üçün mürəkkəb verilənlər toplusunun gizli verilənlərinin araşdırılmasından və işlənilib hazırlanmasından ibarətdir. Verilənlərin analizi ölkələr üçün müxtəlif inkişaf planlarına rəhbərlik edir, müştərilərin tələblərini və təşkilatlar üçün bazar tendensiyalarını proqnozlaşdırır. Böyük verilənlərin analizi xüsusi verilənlərin analizi üsulu kimi ifadə edilə bilər. Beləliklə, böyük verilənlərin analizi üçün ənənəvi metodların əksəriyyəti hələ də istifadə olunur (şək.4).



Şək.4. Böyük verilənlərin analizi üçün ənənəvi metodlar [10]

Böyük verilənlərin analitik vasitələri aşağıdakılardan ibarətdir:

- **Hadoop** - server klasterləri arasında çoxsaylı verilənlər şəbəkələrinin bölüşdürülmüş emalını təmin edən açıq mənbə proqram təminatı layihəsidir. O, çox yüksək dərəcədə nasazlığa dözümlülük ilə bir serverdən minlərlə cihaza qədər genişlənmək üçün nəzərdə tutulmuşdur. Hadoop, Google-un MapReduce və Google Fayl Sistemi tərəfindən istehza ilə qarşılanmış, Yahoo onun yaradıcısı olaraq, biznesində Hadoop-dan istifadəyə başlamışdır. Facebook, IBM, Twitter, American Airlines, LinkedIn, The New York Times və digərləri bu proqram təminatının əsas istifadəçiləridir.

Hadoop klasterlərinin dayanıqlılığının əsas xüsusiyyəti proqram təminatının tətbiq səviyyəsindəki nasazlıqları aşkar etmək və idarə etmək qabiliyyətindən irəli gəlir. Hadoopun iki əsas alt layihəsi vardır: bunlardan birincisi MapReduce işi anlayan və klasterdəki qovşaqlara təyin edilən çərçivə, ikincisi isə məlumatların saxlanması üçün Hadoop klasterindəki bütün qovşaqlara spam göndərən paylanmış fayl sistemi HDFS-dir. HDFS bir çox yerli düyünlərdəki fayl sistemlərini birləşdirərək onları bir böyük fayl sistemə çevirir. HDFS düyünlərin uğursuz olacağını güman edərsə, bu zaman verilənləri çoxsaylı düyünlərdə təkrarlamaqla etibarlılığa nail olunur.

Hadoop proqram təminatı Apache layihələrinin ekosistemi ilə tamamlanır, məsələn Pig, Hive və Zookeeper onun dəyərini artırır və onun istifadəsini yaxşılaşdırır. Hadoop böyük miqyaslı hesablamaların iqtisadiyyatını və dinamikasını dəyişməklə, təsirdörd xüsusiyyətə bölünə bilər:

- miqyaslanıla bilən - yeni düyünlər dəyişdirilmədən lazım olduqda əlavə edilə bilər (verilənlərin formatları və ya verilənlərin necə yükləndiyi, işlərin necə yazıldığı və ya yuxarıdakı proqramlar);

- effektiv xərc - Hadoop kütləvi paralel hesablamaları adi serverlərin geniş klasterlərinə gətirir, nəticədə hər terabayt saxlanma xərcləri əhəmiyyətli dərəcədə azalır, bu da öz növbəsində bütün verilənlərin analizini əlverişli edir;

- çevik - Hadoop sxemsizdir və ixtiyari sayda mənbədən strukturlaşdırılmış və ya strukturlaşdırılmamış istənilən növ veriləni qəbul edə bilər. Çoxsaylı mənbələrdən alınan verilənlər ixtiyari üsullarla bilavasitə birləşdirilə bilər ki, bu da digər sistemin təmin edə biləcəyindən daha dərin analizlərə imkan verir.

- nasazlığa dayanıqlıq bir düyün itirilən zaman, sistemin yönəldilməsi klasterdəki başqa nöqtəyə işləyir və bir impulsu qaçırmadan emal etməyə davam edir. Bütün bunlar proqramçılara paralel emal infrastrukturunun mexanikəsindən xəbərdar olmaq məcburiyyətində qalmadan baş verir. Hadoopla verilənlərin idarə edilməsini və analitikasını tamamilə yeni səviyyəyə yüksəltmək mümkündür.

OpenRefine - böyük data üzərində işləmək və analiz aparmaq üçün inkişaf etmiş bir tətbiqdır. OpenRefine verilənlər üzərində bir neçə əməliyyat aparmağa imkan verir. Bunların sırasında datanın təmizlənməsini, verilənlərin bir formatdan digərinə çevrilməsini, web servislərin və datanın genişləndirilməsini və s. göstərmək olar.

OpenRefine tətbiqinin seçilməsinin əsas məqsədi istifadəçilərin pulsuz şəkildə öz dataları üzərində əməliyyatlar aparıb bunları öyrənməkdən ibarətdir. OpenRefine data üzərində rahat şəkildə bir çox əməliyyatlar aparmaq üçün şərait yaradır. İstifadəçilər bu tətbiqi öz datalarını statistik qraflar şəklində görmək üçün istifadə edə bilər. Onlar heç bir risk olmadan öz "kirli" dataları üzərində işləyə, hətta gördükləri işləri geri çəkə (undo) bilərlər. "URL" - lərin təmizlənməsi və çevrilməsi tətbiqin sistemə yüklənməsi vasitəsilə asan bir şəkildə yerinə yetirilə bilər.

Dağınıq, strukturlanmamış, tutarsız data OpenRefine sayəsində araşdırıla bilər. Ümumilikdə, tutarsız və yararsız datanı araşdırmaq çox çətin olardı, lakin OpenRefine bir çox funksiyaları, məsələn, filtrləmə, tutarsızlıqların düzəldilə bilməsi və görüntüləmə kimi bir sıra funksiyaları irəli sürür. Bütün bunlar datanı təmizləmək üçün bir vasitədir.

Elektron cədvəllər də datanı təsnifləşdirmək üçün istifadə oluna bilməsinə baxmayaraq, ən yaxşı vasitə hesab edilmirlər. Belə ki, OpenRefine verilənləri daha sistemli formada tənzimləyir. Tarixi data üzərində işləyərkən boş xanalara, sürəti çıxarılmış yazılara, tutarsız formatlara rast gəlinir. OpenRefine bu cür problemləri aradan qaldırmağa kömək edir.

Hal - hazırda data analizi biznes sahəsində mühüm rol oynayır. Data analistləri qərar qəbul etmənin təkmilləşdirilməsində, büdcənin kiçildilməsində və yeni biznes fərsətlərinin tapılmasında mühüm rol oynayırlar. Datanın analizi dedikdə datanın təmizlənməsi, çevrilməsi, təftiş olunması və modelləşdirilməsi başa düşülür. Verilənlərin analizi faydalı məlumatların tapılması, qərar qəbul etməyə dəstək və məsləhət vermək üçün yerinə yetirilir. Buna görə də datanın analizinin dəqiqliyinə əmin olmaq üçün ilk növbədə datanın təmizlənməsi vacibdir.

Böyük verilənlərin analizinin vacibliyi ondan ibarətdir ki, xüsusi təkmilləşdirilmiş analitik sistemlər və yüksək səviyyəli kompüter sistemləri tərəfindən irəli sürülən böyük data analitikası yeni maliyyə fərsətləri təşkil edən bir çox biznes faydaları, daha effektiv marketing, daha yaxşı müştəri servisi, təkmilləşdirilmiş əməliyyat və rəqiblər üzərində üstünlüklər yaradır.

Big data analitikasının tətbiqləri strukturlanmış əməliyyat datasının həcmlərinin böyüməsinə və ənənəvi biznes zəkası (BI) tərəfindən toxunulmamış datanı data analistləri, data alimləri, proqnoz modelçiləri və statistikaçılar tərəfindən analiz edilməsinə kömək edir. Bunlar strukturlanmamış və yarı-strukturlanmış datanı əhatə edir. Məsələn, internetdə klik axını ilə yaranan data, web şəbəkə girişləri, sosial media

məzmunları, müştəri mailləri və sorğu cavablarından yığılan mətnlər, mobil telefon səs qeydləri və internetlə bağlı vasitələrin avtomatik şəkildə topladığı datanı buna nümunə göstərmək olar.

Sosial media şirkətlərində Böyük Verilənlərdən istifadə. Əgər sosial media şirkətlərinin elmləşdiyi bir sahə varsa, o da data hesab edilir. İstifadəçilərin hər keçən dəqiqə paylaşdığı informasiya sayəsində bu şirkətlərdə həddindən artıq data mövcuddur. Sosial media şirkətlərindəki böyük həcmli data insanların bir-biri ilə necə əlaqədə olduğunu, bu əlaqələrin mərkəzində şəxsi insanların və cəmiyyətin çox dəyərli məlumatlarının cəmləndiyini göstərir. Bu verilənlər sürətli verilən axını ilə birgə böyük verilənlərin özünü təşkil edir.

Sosial mediaya analitikanın tətbiqi insanlar arasındakı əlaqənin məzmununun biznes performansını və insanların brendə baxış bucağının necə dəyişdiyini görmək üçün Böyük data tətbiqləri əlaqələrin də əsas alətinə çevrilir. Məzmun analitikası şirkətlərin istifadəçilər tərəfindən paylaşmalara daha çox diqqət etməsinə xidmət edir. Məsələn olaraq analitika tətbiqlərinin brendin reputasiyasına təsir edən pozitiv və neqativ fikirləri izləmək üçün proqramlaşdırıla bildiyini göstərmək olar [10,11,12]. Böyük verilənlər marketinq qərarlarında əsas amillərdən biridir.

Digər sahələr kimi sosial media da Böyük datanı marketlərin analizi və istehlakçının davranışının proqnozlaşdırılması üçün yararlı hesab edilir. 2012-ci ildə Jay Parikh (Facebook mühəndislik vitse prezidenti) "Facebook"-un hər gün 500 terabayt informasiya, 300 milyon milyard "like" və 2.5 milyard yüklənmiş kontent əldə etdiyini bildirib. Bütün bu verilənlər sadəcə bir dəqiqə ərzində "Facebook"-a istifadəçi reaksiyalarına fikir vermək və ya öz fikirlərini dəyişmək və yaymaq imkanı verir.

Əlavə olaraq, istifadəçilərin yaşı, cinsiyyəti, nikah statusu, coğrafi lokasiyası, gəlir həcmi, təhsil səviyyəsi, müəyyən məhsulu almağa olan meyli şirkətə hansı insanlarla əlaqədə olmalı olduğunu aydınlaşdırır. Belə analizlər həm də reklamların necə müxtəlif müştəri segmentlərinə görə ayrıldığını göstərir. Reklamçılar sürətli reaksiya verib kampaniyalar quraraq gəlir əldə edə bilirlər. Sosial media analizləri göstərir ki, 25-30 yaş arası evlənmiş qadınlar idman zalı endirimi reklamı üçün ideal namizədlərdir. Bu məlumatla əsaslanaraq, satıcı həmin namizədlərə hədəf olaraq Twitter, Facebook və digər sosial şəbəkələrdə endirimli reklamlar yerləşdirə bilər. Əgər analizlərin əksi olarsa və rəylər pis olarsa, təklif performansın təkmilləşdirilməsi üçün yaxşılaşdırılmalıdır.

Bir çox şirkətlər sosial mediaya insanlarla qurduğu qarşılıqlı əlaqə üçün təşəkkür edirlər. Baxmayaraq ki, məhsul təsnifləşdirilməsi sosial mediadan əvvəl qurulmuşdur, biznes tərəfindən edilən sosial media datasına qarşı əhatə dairəsi heyrətləndiricidir. Sosial media analitikasının tətbiqləri vasitəsilə bu şirkətlər bir dəqiqə ərzində data üzərindən qərar verə bilirlər [6,12]. Üstəlik sosial media analitik tətbiq vasitəsilə bizneslər strukturlanmamış data əhatə edən dialoqlara baxıb onların qərarlarına və hərəkətlərinə kömək ola biləcək məlumatlardan yararlanırlar. Sosial medianın statistik analizləri vasitəsilə onlar hər paylaşımдаki reaksiyalara, izləyicilərin paylaşmasına, mobil və masaüstü reaksiyaların qarşılaşdırılmasına, cavablara ("retweet"-lər), "URL"-lərin kliklənməsinə və əməliyyat keçmişinə əsaslanaraq sosial media strategiyalarının effektivliyini brendin reklam tanınmasını və güvənilliyini tənzimləyə bilirlər [8].

Big data həmçinin insanların sosial media qruplarında oynadıqları rola baxmağa imkan yaradır. Məsələn, böyük kütləyə sahib istifadəçilərə təsirçi (influencer) kimi

baxıla bilər. Bu qədər kütləyə aparıcılıq etdikdə, şirkət təhdid daşıyan diskussiyalara baxa və hətta onlara qoşula bilər.

Böyük data sosial media şirkətlərinin insanların şəxsi həyatına müdaxiləsini sürətləndirməyə kömək edəcək Facebook, Twitter, Instagram və Pinterest pul qazanma təkliflərini irəli sürdükcə, görünür ki, Böyük datanın gələcəkdə sosial media üçün əldə edəcəyi faydalar daha da fərdiləşəcək. Cambridge və Stanford Universitetləri tərəfindən aparılan araşdırma göstərir ki, Facebook öz datasından istifadə edərək insanların xasiyyətini onların yaxın dostları və ailəsindən daha yaxşı tanıyır. Hər bir like, paylaşım, rəy sosial media şirkətlərinə insanların nəyi xoşlayıb xoşlamadığını, bir sonrakı hərəkətin nə olacağını, hansı brendi bəyəndiyini və nə almaq istədiyini bildirir. Brauzerlərdəki və axtarış aparatlarındakı əməllər birbaşa sosial media profilinə bağlanır və hər sonrakı gedişi göstərə biləcək uzun bir "ayaq izi" buraxır. Bu vəziyyət sosial mediada paylaşım etməyə güvənən insanlarda daha çox intensivləşir [9].

Nəticə. Big datanın mövcudluğu, yeni informasiyanın idarə edilməsi və analitik proqram təminatı data analizi tarixində unikal moment yaratmışdır. Baxılan tədqiqatlarda verilənlərin sadə konsepti, daha sonra Böyük verilənlər, Big datanın analizi və idarə edilməsi məsələləri şərh olunmuşdur. Böyük və kiçik həcmli verilənlər müqayisəli təhlil edilmiş, onlar mühüm xüsusiyyətlərinə görə iki kateqoriyaya ayrılmış, Big datanın sosial media şirkətlərində istifadəsinin vacibliyi və əhəmiyyəti izah edilmişdir.

Ədəbiyyat siyahısı

1. R. R. Mukkamala, A. Hussain, and R. Vatrappu. "Towards a set theoretical approach to big data analytics", Jun./Jul. 2014.
2. R. R. Mukkamala, A. Hussain, and R. Vatrappu. "Towards a formal model of social data", IT Univ. Copenhagen, Copenhagen, Denmark, Nov. 2013.
3. MathWorks, "MATLAB Getting Started Guide", 2011.
4. https://raw.githubusercontent.com/saodem74/SentimentAnalysis/master/Data/comment_data.csv, April, 2019.
5. [https://github.com/saodem74/Sentiment-Analysis-facebook-comments,\(Sentiment-Analysis-facebook-comments\)](https://github.com/saodem74/Sentiment-Analysis-facebook-comments,(Sentiment-Analysis-facebook-comments)), April, 2019.
6. Puneet Singh Duggal, Sanchita Paul. "Big Data Analysis: Challenges and Solutions", 2013.
7. Jasmine Zakir. "Big Data Analytics", 2015.
8. Miroslav Vozábal. Tools and Methods for Big Data Analysis, Master Thesis, Pilsen, 2016.
9. Zaharia, Matei; Chowdhury, Mosharaf; Das, Tahagata; Dave, Ankur; Ma, Justin; McCauley, Murphy; Franklin, Michael; Shenker, Scott; Stoica, Ion;. Resilient Distributed Datasets: A Fault-Tolerant. Berkeley: University of California at Berkeley, Electrical Engineering and Computer Sciences, 2011.
10. Manyika, James and Chui, Michael. Big data: The next frontier for innovation, competition, and productivity. s.l. : McKinsey Global Institute, 2011. 978-0983179696.
11. Iafate, Fernando and Front, Matter. From Big Data to Smart Data. Chap : John Wiley & Sons, 2015.
12. <http://www.greenplum.com>, January, 2019.

К.А.Мамедова, Е.Н.Алиева

Аналитический анализ больших данных

Резюме

Представлена общая концепция аналитического анализа больших данных, были сопоставлены большие и малые данные, объяснена их структура, анализированы история базы данных, средства ее хранения и управления. Предложены методы анализа больших данных Hadoop и OpenRefine, которые служат для изменения динамики крупномасштабных вычислений. Показана важность и значимость использования больших данных в компаниях, работающих в социальных сетях, проведен их сравнительный анализ и даны их различные характеристики.

К.А.Маммадова, У.Н.Алиева

Analytical analysis of big data

Abstract

The general concept of the analytical analysis of big data is presented, large and small data are compared, their structure is explained, the history of the database, its storage and management tools are analyzed. Hadoop and OpenRefine big data analysis methods are proposed that serve to change the dynamics of large-scale computing. The importance and significance of the use of big data in companies operating in social networks is shown, their comparative analysis is carried out and their various characteristics are given.
