

“Big Data” texnologiyaları: mövcud vəziyyət və perspektivlər

Rasim Əliquliyev¹, Məkrufə Hacırahimova²

^{1,2} AMEA İnformasiya Texnologiyaları İnstitutu, Azərbaycan, Bakı

¹rasim@science.az, ²makrufa@science.az

Xülasə— Məqalədə böyük verilənlərin terminoloji aspektləri, mahiyyəti, mənbələri və imkanları təhlil olunur. Məqalədə həmçinin, böyük verilənlərin problemləri, 3V konsepsiyası və mövcud yanaşmalar tədqiq olunur, böyük verilənlərin analizi məsələləri araşdırılır.

Açar sözlər— Big Data; data science; Big Data analytics; NoSQL, MapReduce; Hadoop; OLAP.

I. GİRİŞ

İnformasiya, bilik və yüksək texnologiyalara əsaslanan informasiya cəmiyyəti (İC) böyük elmi-texniki inqilab çağındadır və bu inqilab, biznesdən başlayaraq elmə qədər həyatımızın bütün sahələrinə nüfuz etməkdədir. Bu eyni zamanda sosial-iqtisadi, təhlükəsizlik, elmi baxımdan mürəkkəb bir problemdir. Problem böyük verilənlər (“Big data”) ideyası ilə daha da dərinləşməkdədir. Bu yeni termin həcm və mürəkkəblik baxımından mövcud idarəetmə metodları və ya intellektual analiz alətləri ilə emal oluna bilməyən verilənlər dəstini identifikasiya etmək üçün istifadə edilir. Son illərdə eksponensial olaraq artan verilənlərin saxlanması, idarə edilməsi, onlardan dəyər yaradılması ciddi problem yaradır. Problem müxtəlif mənbələrdən fasiləsiz olaraq generasiya olunan verilənlərin çoxalması ilə ənənəvi üsul və alətlərin bu verilənləri emal edə bilməməsindədir. Ona görə də bu texnologiyanın elmi-tədqiqat obyekti kimi öyrənilməsi vacibdir. Yüzlərlə terabayt və ekzabayt həcmində strukturlaşdırılmamış informasiyanın toplanması və idarə edilməsi, saxlanması, təhlükəsizliyi, axtarışı, analizi (*analitik hesabatların generasiyası və vizuallaşdırılması, proqnozlaşdırma*) və s. kimi məsələlərin həllində yeni texnologiyalar, yanaşmalar, daha mükəmməl analiz üsulları tələb edir.

Böyük verilənlər (*bundan sonra BV*) artıq ABŞ və bir sıra qərbi dövlətlərində elmi ictimaiyyət, biznes-cəmiyyətləri, hökumət strukturları tərəfindən neft qədər strateji resurs kimi dəyərləndirilir. Məsələn, ABŞ prezident administrasiyası 2012-ci ilin martında BV sahəsində “Big data Research and Development” təşəbbüsünü elan etmişdir [1]. Təşəbbüs çərçivəsində BV texnologiyalarının ABŞ dövlət siyasətinin aparıcı istiqamətlərində istifadəsi üçün kompleks tədbirlərin həyata keçirilməsi üçün 200 milyon dollar həcmində vəsait ayrılmışdır. BV-nin müxtəlif aspektlərinə həsr olunmuş beynəlxalq konfranslar, seminarlar, forumlar keçirilməkdədir, elmi və populyar jurnalların xüsusi nömrələri bu mövzuya həsr olunmuşdur. “Big Data” adı ilə yeni akademik jurnalların nəşrinə başlanmışdır. IDC (International Data Corporation), McKinsey Global İnstitutu, Gartner və s. kimi şirkətlər “big

data”-ni 2013-cü ildə əsas texnoloji istiqamət və 2014-cü ildə informasiya-kommunikasiya texnologiyaları sahəsinin lokomotivi adlandırırlar [2-4].

II. BÖYÜK VERİLƏNLƏRİN TERMİNOLOJİ ASPEKTLƏRİ, MAHIYYƏTİ VƏ MƏNBƏLƏRİ

“Big data” termini ilk dəfə 1998-ci ildə kompüter elmləri üzrə mütəxəssis Con Meşi tərəfindən istifadə edilmişdir [5]. Bir qədər sonra, 2000-ci ildə, F.Dieboldun akademik mühitdə dərc olunan tədqiqatında rast gəlinir [6]. Lakin termin populyarlığını 2008-ci ildə həftəlik “Nature” jurnalının “big data” mövzusunda həsr olunmuş xüsusi nömrəsində Klifford Linçin (Clifford Lunch) “Big Data: How do your data grow?” adlı məqaləsi ilə qazanmışdır. Mütəxəssislər BV-ni bəzən mineral resurslarla – böyük filiz mədəni (*large ore*), yeni neft (*new oil*), gizli biliklərin mənbəyi kimi “*data mining*”, təbiət kataklizmləri (*data tornado, data deluge*), təbii fəlakət (*sunami*) və s. ilə müqayisə edirlər. Onlar qeyd edirlər ki, verilənlərin dəyərini başa düşənlər üçün XXI əsrdə verilənlər XVIII əsrdə istifadə olunmamış neft kimi çox qiymətli aktiv hesab olunur [7]. 2012-ci ildə Davosda Beynəlxalq İqtisadi Forumda BV valyuta və ya qızıl kimi yeni sinif iqtisadi aktiv kimi bəyan edilmişdir. Bəzən bu problemi Mur qanunu kimi də interpretasiya edirlər.

Verilənlərin mənimlənməsi və istifadəsində yeni eranı əks etdirən bu anlayış nisbi anlayışdır və zaman-zaman dəyişə bilər. Yəni bu gün üçün “böyük” hesab olunan informasiya, sabah üçün normal sayıla bilər. Böyük verilənlər həcmi 1 terabaytdan (1Tb=1024 Qb) başlayaraq daha böyük həcmli verilənlər (1Zetabayt=1024Ekzabayt) hesab olunur [2-4].

Vahid unifikasiya olunmuş tərif olmadığından elm, sənaye və kütləvi informasiya vasitələri kimi maraqlı tərəflər müxtəlif təriflər təqdim edirlər [2-4,8]. Bu da “big data” terminin anlanmasında qeyri-müəyyənlik yaradır. Çünki adından da məlum olduğu kimi ilk baxışda BV dedikdə ancaq həcm nəzərdə tutulur. Əslində isə BV-nin müxtəliflik və sürət kimi xarakteristikaları da vardır. Tədqiqatlar göstərir ki, BV-nin ananlanmasında 3V modelini [8] daha çox istifadə olunan və geniş yayılmış tərif kimi qəbul etmək olar. Çünki bu model uyğun texnologiya və məhsullara olan tələbləri əldə etmək üçün BV-ni daha yaxşı xarakterizə edir.

Son onillikdə kommunikasiya vasitələrinin artması və informasiyanın əlyətərliliyinin asanlaşması nəticəsində hər il rəqəmsal verilənlərin həcmi həndəsi silsilə ilə artır. Bunu statistik rəqəmlərdə də görmək olar. Bəşəriyyətin mövcudluğundan 2003-cü ilə qədərki dövrdə dünyada cəmi 5

ekzabayt məlumat generasiya olunduğu halda, 2012-ci ildə rəqəmsal informasiyanın həcmnin 500 dəfə artaraq 2.7 zetabayt olduğu və 2015-ci ildə üç dəfə artması, növbəti hər il 40% artaraq, 2020-ci ildə dünyada informasiyanın həcmnin 44 zetabayta çatacağı proqnozlaşdırılır [2]. Bu da Yer kürəsinin hər sakininə düşən 5200 qiçabayt informasiyaya ekvivalentdir. IBM-in məlumatına görə, dünyada hər gün 2.5 trilyon bayt məlumat hazırlanır, hər dəqiqədə 100 milyon email göndərilir, Google axtarış sistemində 2 milyon axtarış sorğusu, Facebook sosial şəbəkəsində 350 Qb məlumat emal olunur və 570-dən çox veb-sayt yaradılır, hər dəqiqə 72 saatlıq yeni video YouTuba yüklənir və s. İnformasiyanın 90%-nin isə son iki ildə yaradıldığı bildirilir [9].

BV-nin əsas mənbələrini elmi eksperimentlər, sensor və sosial şəbəkələr, dövlət qurumlarının portalları, iqlim haqqında məlumatlar, GPS siqnalları, GIS sistemlər, böyük şirkətlərin bazaları, elektron poçt, mobil qurğular, böyük satış mərkəzləri, bank əməliyyatları və s. təşkil edir. BV-nin mənbələri əsasən beş kateqoriyaya bölünür: veb və sosial media; maşın-maşın (machine-to-machine – M2M); tranzaksiyalar; elmi; insanların yaratdığı verilənlər [10,11]. “Əşyaların interneti” ideyası M2M kommunikasiyasına ən yaxşı nümunədir.

III. BÖYÜK VERİLƏNLƏRİN ƏSAS PROBLEMLƏRİ VƏ YA 3V KONSEPSİYASI

Tətbiq olunduğu sahələrdən asılı olmayaraq, BV-ni təsvir edən, eyni zamanda onların əsas problemlərini özündə əks etdirən ümumi xarakteristikalar mövcuddur: həcm (*volume*), sürət (*velocity*) və müxtəliflik (*variety*). İngilis dilli mənbələrdə bunu «3V»lər də adlandırırlar. Bu parametrlərin konvergensiyası BV-ni təyin etməyə və digər verilənlərdən fərqləndirməyə kömək edir. Bu model ilk dəfə 2001-ci ildə D.Laney tərəfindən verilmişdir [8] və “big data” texnologiyalarının əsas konsepsiyasını təşkil edir. Bu konsepsiya çox böyük sürətlə və müxtəlif mənbələrdən toplanan çox böyük həcmdə verilənləri daha səmərəli istifadə etmək, saxlamaq, analiz edərək ondan daha qiymətli informasiyanı əldə etmək ideyasını özündə əks etdirir. Analitiklər bəzən “5V”lər kimi təsvir edilən dördüncü – həqiqilik (*veracity*) [9] və beşinci – dəyər (*value*) [2, 3] xarakteristikalarını da qeyd edirlər.

Həcm - BV-nin əsas xarakteristikasıdır. Həcm problemi ilk növbədə saxlama problemi yaradır ki, genişmiqyaslı saxlama və paylanmış emal tələb edir. Hazırda saxlama məsələsinin həllində informasiyanın qurğular arasında miqrasiyasını həyata keçirən bir sıra texnologiyalar: DAS (*Direct-Attached Storage*), NAS (*Network Attached Storage*), SAN (*Storage Area Networks*), HSM (*Hierarchical Storage Management*), ILM (*Information Life-cycle Management*) mövcuddur. Son zamanlar isə hesablamalar və yaddaş resurslarının klasterləşməsi və virtuallaşdırılmasını həyata keçirməklə, verilənlərin emalı və yadda saxlanılmasına xidmət edən “grid” və “cloud computing” texnologiyalarının tətbiqi saxlama sahəsindəki problemləri demək olar ki, aradan qaldıra bilmişdir [5,11,12]. Həcmindən asılı olaraq BV üç qrupa bölünür [13-15]: Sürətli verilənlər (Fast Data) – onların həcmi terabaytlarla ölçülür; böyük analitika (Big Analytics) - onların həcmi petabaytlarla

ölçülür; dərinə nüfuzetmə (Deep Insight) - onların həcmi ekzabaytlarla və zetabaytlarla ölçülür.

Sürət. Həcm artıqca emal üçün də çox böyük sürət tələb olunur. Burada iki hal nəzərdə tutulur. Birinci, yeni verilənlər böyük sürətlə generasiya olunur, mövcudlar yenilənir və toplanır. İkincisi, sürət zaman problemi kimi dəyərləndirilir və mövcud emal texnologiyalarının verilənləri real-vaxtda analiz etmək imkanına malik olması ilə izah olunur [22].

Müxtəliflik BV-nin təbii özəlliklərindəndir. BV ənənəvi relyasiya verilənlər bazasının sətir və sütunlarında ifadə olunmuş strukturlaşdırılmış verilənlərlə bərabər, strukturlaşdırılmamış – *mətn, video-audio fayllar, təsvirlər* və s. tiplərdə olur. Bu tip verilənlər isə dünyada, bütün informasiyanın 80-90%-ni təşkil edir. Bunları bir araya yığmaq və birgə emal etmək və analiz üçün onları uyğun şəkə salmaq çox çətindir [13].

IV. “BIG DATA” ALƏTLƏRİ VƏ TEXNOLOJİ HƏLLƏR

«Big Data» texnologiyaları İnternetin nailiyyəti ilə “kamil” kontent səviyyəyə çatmış və eksponensial inkişafda ifadə olunan bir faza – “*informasiya partlayışı*”nın yaranması ilə əlaqədardır. İnformasiyanın həddindən artıq çoxalması ilə onların idarə olunması daha da çətinləşir, informasiya yükü problemi yaradır. Problemin həlli məqsədi ilə İT sahəsinin nəhənglərindən olan Google şirkəti tərəfindən Google File System [16] və MapReduce [17] proqram-aparat platforması yaradılmışdır. Bunun əsasında açıq kodlu Apache Hadoop və Hadoop File System [18] proqram təminatları işlənmiş və bununla da BV texnologiyalarının əsası qoyulmuşdur.

MapReduce paylanmış hesablamalar modeli Google şirkəti tərəfindən 2004-cü ildə təqdim olunmuşdur və BV üzərində paralel proqramlaşdırmanın əsasıdır [16,17]. Onun əsas ideyası, böyük və mürəkkəb verilənləri kiçik hissələrə bölməklə emal etməkdir. MapReduce-un işi iki mərhələdən (Map və Reduce) ibarətdir. “Map” mərhələsində giriş verilənləri ilkin emal üçün əsas qovşağa (*master node*) göndərilir və orada digər kompüterlər (*worker node*) arasında paylanılır. “Reduce” mərhələsində əsas qovşaq emal olunmuş verilənləri işçi qovşaqlardan toplayır və onun əsasında məsələnin həllinin nəticəsini formalaşdırılır.

BV-nin de-fakto standartı hesab olunan Apache Software Foundation-un layihəsi **Hadoop** daha geniş yayılmış texnologiyadır, paylanmış hesablamalar mühitində böyük verilənlərin emalı və analizi üçün əsas platformadır, MapReduce modelinin açıq kodlu (*open access*) sistemidir və 1000 qovşaqdan çox miqyaslaşmanı təmin edir. Hadoop iki əsas komponentdən ibarətdir: Hadoop MapReduce və Hadoop Distributed File System (HDFS). MapReduce paralel hesablamalara, HDFS paylanmış fayl sistemi isə verilənlərin idarə edilməsinə cavab verir [18].

NoSQL (*Not Only SQL*) bu gün BV aləminin əsası hesab olunur və verilənlərin idarə edilməsinin miqyaslılıq (*scalability*), əlverişlilik (*availability*) və verilənlərin uyğunlaşdırılması (*consistency*) kimi problemlərin həllində tətbiq olunur [53]. Ədəbiyyatda paylanmış sistemlərin bu üç xüsusiyyəti Berkli universitetinin professoru Eric Brewer

tərəfindən təklif olunmuş CAP (*Consistency, Availability and Partition Tolerance*) teoremi kimi də tanınır [19].

Hazırda böyük həcmli verilənlərin saxlanması, idarə olunması, analizi və vizuallaşdırılması üçün IBM, Microsoft, SAS, HP, EMC kimi informasiya texnologiyaları nəhəngləri tərəfindən paralel emalı təmin edən müxtəlif proqram-aparat həlləri mövcuddur. Ənənəvi informasiya xəzinələri çoxölcümlü analiz (OLAP), klassifikasiya, klasterizasiya və s. alətlər dəstini təqdim edir, bu gün isə operativ yaddaşda terabaytlarla informasiyanın analitik emalı üçün SAP şirkətinin– HaNa (*High-performance Analytic Appliance*), Oracle şirkətinin Oracle Exalytics, Oracle Exadata məhsulları mövcuddur. Bundan başqa Netezza, Teradata, Greenplum və s. şirkətlərinin ənənəvi relyasiya verilənlərinin idarə edilməsi sistemi əsasında terabaytlar və ekzabaytlarla verilənləri səmərəli emal edən proqram-aparat alətləri vardır.

Bu gün müasir İT faktorları: böyük verilənlər, analitika və bulud texnologiyalarını bir-birindən ayrı təsəvvür etmək mümkün deyildir. **Bulud texnologiyaları** saxlanma, böyük hesablamaların aparılmasında son dərəcə müvəffəqiyyətli yanaşmalardandır. Burada böyük həcmli rəqəmsal informasiya **IaaS** – Infrastructure as a service, **PaaS** – Platform as a service, **SaaS** – software as a service “bulud” xidmətləri vasitəsi ilə mərkəzləşdirilmiş qaydada idarə olunur və saxlanılır [12].

V. BÖYÜK VERİLƏNLƏRİN ANALİTIKASI

Biz hazırda ekzabayt və zetabaytlarla böyük verilənlər axınının istehsalını təmin etmiş elm, texnika və texnologiyaların geniş yayıldığı erada yaşayırıq. Elm sahəsində BV artır, çünki indi elmi tədqiqatlar nəzəri düşüncələrdən daha çox elmi eksperimentlərə (petabaytlarla informasiya generasiya edən LHS–Large Hydron Collider, LSST–Large Synoptic Survey Telescope, Hubble teleskopu və s.) köklənmişdir [20]. Kommersiyada BV massivi yaranır, ona görə ki, hazırda insan fəaliyyətinin böyük hissəsi İnternetdədir, onlayndır.

Real-vaxta maksimum yaxın rejimdə verilənlərin analizinə olan tələbat müxtəlif parametrlər, xarakteristikalar, hadisələr və s. arasındakı korrelyasiyanı tapmağa, klassifikasiya və analitik hesabatlar və bunun əsasında proqnozların verilməsinə imkan verən BV analitikasının (**Big Data Analytics**) yaranmasına gətirib çıxardı [5,11,14,15]. Mütəxəssislər BV-nin analizində iki yanaşmanı qeyd edirlər: saxlamaq və analiz etmək (**store and analyze**); analiz etmək və saxlamaq (**analyze and store**) [14]. Birinci halda verilənlərə analitik alətləri tətbiq etməzdən əvvəl tranzaksiya sistemlərindən (**OLTP- Online Transaction Processing**) alınan xam verilənlər emal olunur və xəzinəyə (**data warehouse**) yüklənir. Məhz bu prinsip əsasında yaradılmış ənənəvi analitik həllər (**OLAP - Online Analytical Processing**) BV-nin analizində uyğunsuzluq meydana çıxır. Məlumdur ki, **superkompüterlər** saniyədə böyük sayda (**flops**) hesab əməliyyatlarını yerinə yetirmək imkanına malikdirlər və ancaq strukturlaşdırılmış verilənlərlə işləyərkən faydalıdır. İndi verilənlər axınının analizi üçün hipotetik bir qurğu lazımdır ki, o, verilənləri xüsusi saxlanma yerlərində toplanmadan, yığım

yerinə daha yaxın yerdə dərhal emal etsin. **“DataScope”** adlandırılan bu qurğu superkompüterlərdən fərqli olaraq eksperimental verilənlərlə işləmək məqsədi güdür. Ümumiyyətlə petabaytlarla verilənlər massivlərinin emalının zəruriliyi **“Data – Intensive Computing”** [11,21] yanaşmasını ortaya çıxarmışdır. Bu halda verilənlər hesablamalara deyil, hesablamalar verilənlərə istiqamətləndirilir.

BV-nin analizi sahəsində ROLAP (*Relational On-Line Analytical Processing*), MOLAP (*Multi-Dimensional On-Line Analytical Processing*), HOLAP (*Hybrid Online Analytical Processing*) kimi ən müasir texnologiyalar mövcuddur [40]. Bunlardan hansısa birinin seçilməsi verilənlərin yenilənməsindən asılıdır.

Strukturlaşdırılmamış verilənlərin daha dərin intellektual analizi (**mining**) [15] və nəticələrinin vizuallaşdırılması [14,15,22,23] BV analitikanın əsas məsələlərindəndir. Problemin həllində data mining texnologiyaları sinfindən olan klassifikasiya, klasterləşdirmə, neyron şəbəkələr və s. kimi üsullar tətbiq olunur [22].

BV-nin problemləri ilə məşğul olmaq üçün ənənəvi informatikadan başlayaraq riyaziyyata qədər müxtəlif sahələr üzrə bacarıq və vərdislərə malik **“data scientist”** adlanan daha ixtisaslaşmış mütəxəssislərə ehtiyac vardır [14]. 2013-cü ildən başlayaraq Dandi Universitetində (Şotlandiya), Okled Universitetində (Yeni Zelandiya), London İmperial Kollecinə, Cənubi Kaliforniya Universitetində, Vaşinqton, Berkli, Nyu-York universitetlərində **“data science”** ixtisası üzrə magistr proqramları tədris olunur. Bu problem təhsilin yeni məsələsi kimi qiymətləndirilir.

VI. BÖYÜK VERİLƏNLƏRİN TƏHLÜKƏLƏRİ VƏ FAYDALARI

BV-nin də digər texnologiyalar kimi iki tərəfi: zərərləri və faydaları vardır. Biri ilə mübarizə edərkən, digərini yaddan çıxarmaq olmaz. Heterogen xam verilənlər cəmiyyətin bütün sahələrini kökündən dəyişə biləcək təsirə malik bilik mənbəyidir. Yeni-yeni elmi kəşflərə imza atmaq, iqtisadi inkişafa nail olmaq məqsədi ilə bu verilənlər biliyə çevrilməlidir. BV texnologiyalarının faydasını göstərmək üçün 2009-cu ildə BMT-nin “Global Puls” təşəbbüsünü qeyd etmək kifayətdir [24]. **“İnsan-maşın”** və **“maşın-maşın”** kimi ikitərəfli qarşılıqlı əlaqə nəticəsində müxtəlif mənbələrdən və müxtəlif formatlarda toplanan verilənlərin birgə analizi və onlardan yeni biliklərin və faydalı məlumatların əldə olunması yeni elmi kəşflərin edilməsində, dövlət və özəl təşkilatlarda düzgün qərarların qəbul edilməsində, hüquq qaydalarının qorunmasında, sosial təminat, milli təhlükəsizlik, terrorizm, iş yerlərinin yaradılması, xəstəlik epidemiyalarını əvvəlcədən söyləməyə, marketing işlərinin yaxşılaşdırılmasına, insanların gizli davranışlarını üzə çıxarmaq, məqsəd və niyyətlərini anlamaq, onların digər insanlarla, ətraf mühitlə qarşılıqlı əlaqəsini başa düşməkdə, maliyyə sektorunda milli səviyyədə iqtisadi riskləri daha yaxşı anlamaq siyasətçiləri və tənzimləyici orqanları istiqamətləndirmək və risk sistemlərini daha yaxşı idarə etməkdə bu texnologiyadan istifadə önəmlidir.

Qeyd olunan üstünlüklərə baxmayaraq, verilənlər çoxaldıqca informasiyalara əlçatan və ondan istifadə edən

subyektlərin sayı da çoxalır. Verilənlərin əksəriyyəti fərdi məlumat olduğundan, xüsusi mühafizə olunmalıdır. İnsanların xəbəri olmadan onların haqqındakı verilənlərin analiz olunması etik və hüquqi cəhətdən yolverilməzdir. Bu vəziyyət insanlarla təşkilatlar arasında yeni rəqəmsal partlayışa gətirib çıxara bilər.

VII. PROQNOZLAR VƏ ƏKS ƏLAQƏ İQTİSADİYYATI

Qeyd olunduğu kimi verilənlər hər il həndəsi silsilə ilə artmaqdadır. Proqnozlar isə bu prosesin hələ davam edəcəyindən xəbər verir [2-4]. Bu baxımdan BV-nin saxlanması, idarə edilməsi və analitikası sahəsində tədqiqatçıların və praktiklərin yaxın illərdə məşğul olacağı problemlər *arxitektura, analitika və mining, vizuallaşdırma, faydalı verilənlərin aşkarlanması, müxtəlif tip verilənlərin inteqrasiyası* və s.-dir.

Maddi (*xammal*) və qeyri-maddi (*verilənlər və ya informasiya*) resursların emalının inteqrasiya olunduğu “informasiya iqtisadiyyatı” və ya “rəqəmsal iqtisadiyyat” dövründə iqtisadiyyat bazar vasitəsilə reallaşdırılan əks-əlaqə olmadan mövcud ola bilməz. IDC və IIA (*International Institute of Analytics*), Wikibon cəmiyyətinin tədqiqatlarında BV texnologiyaları bazarının həcmi 2014-cü ildə proqram-aparat təminatı, xidmət üzrə 16,1 milyard dollar olacağı göstərilir. 2012-2017-ci illər ərzində 50 milyard dollardan bir qədər çox olacağı proqnozlaşdırılır. (*Hazırda bu hədd 7 milyard dollar təşkil edir*) [2-4]. Bunlardan proqram-təminatı 24%, xidmətlər 29% və 45% iş saxlama sistemləri təşkil edir. Wikibonun 2013-ildəki tədqiqatında [25] (*70 şirkət üzrə*) dünyada “Big data” texnologiyaları bazarına IBM (*\$1,368 milyon*) və HP (*\$869 milyon*) şirkətləri liderlik edir. Həmin tədqiqatda həmçinin göstərilir ki, 2015-ci ildə bu sahədə 4,4 milyon IT-iş yerlərinin yaradılması gözlənilir, onlardan 1,9 milyonu ABŞ-da olacaqdır.

VIII. NƏTİCƏ

BV böyük informasiya massivlərini istifadə etməyə imkan verən yeni nəsil texnologiyadır. Böyük həcm, sürət və müxtəliflik kimi xüsusiyyətlərlə xarakterizə olunan verilənlərin emalı və analizi məqsədlə yaradılmış *MapReduce, Hadoop, HDFS, NoSQL* və s. proqram-aparat platformalar mövcuddur. Böyük həcmli verilənlərlə işləməkdə superkompüterlər, qrid və bulud texnologiyaların köməyi ilə müəyyən problemlər aradan qaldırılmış olsa da, bu sahədə hələ də kədr, praktiki və elmi-nəzəri problemlər mövcuddur. Problem ancaq böyük həcmdə verilənlərin saxlanması və idarə olunmasında deyil, həm də strukturlaşdırılmamış verilənlərin analizi və nəticələrin interpretasiyasındadır. BV texnologiyalarının tətbiq etməklə bir çox elm sahəsində tez bir zamanda yüksək nailiyyətlər, idarəetmədə və biznes fəaliyyətində gəlir, rəqabətdə müəyyən üstünlüklər əldə etmək

mümkündür. Ona görə də bu yeni elmi istiqamət daha da dərindən tədqiq olunmalı və inkişaf etdirilməlidir.

ƏDƏBİYYAT

1. Big Data Research and Development Initiative. www.whitehouse.gov/
2. Worldwide Big Data Technology and Services 2013–2017 Forecast. <http://www.idc.com>
3. Big data: The next frontier for innovation, competition, and productivity. Analyst report. McKinsey Global Institute, May 2011. <http://www.mckinsey.com/>.
4. Beyer M. A., Laney D. “The importance of big data: A definition.” Stamford, CT: Gartner, 2012.
5. Chen J., Chen Y., Xiaoyong D., et.al., “Big data challenge: a data management perspective,” // *Frontiers of Computer Science in China*, 2013, vol.7, №2, pp. 157–164.
6. Diebold F. “Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting.” Discussion Read to the Eighth World Congress of the Econometric Society, 2000.
7. Data Is the New Oil of the Digital Economy. 2014, <http://www.wired.com>
8. Laney D. “3D Data Management: Controlling Data Volume, Velocity and Variety.” Technical report, META Group, Inc (now Gartner, Inc.), February 2001. <http://blogs.gartner.com/>
9. What is big data? - Bringing big data to the enterprise, 2013. <http://www-01.ibm.com/>
10. Soares S. “Big Data Governance - An Emerging Imperative.” MC Press Online, LLC, 1st edition, 2012.
11. Philip C.L., Zhang C-Y. “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data,” // *Information Sciences*, 2014, vol. 275, № xx, pp. 314–347.
12. Agrawal D., Das S., Amr El Abbadi “Big Data and Cloud Computing: Current State and Future Opportunities”, / *Proceedings of the 14th International Conference on Extending Database Technology*, 2011, pp.530-533.
13. McAfee A., Brynjolfsson E. “Big Data: The Management Revolution. *Harvard Business Review*” 2012, vol.90, №10, pp. 60–68.
14. Селезнев К. Проблемы анализа больших данных // *Открытые системы*, 2012, №7, с.25-29.
15. Fan W., Bifet A. “Mining Big Data: Current Status, and Forecast to the Future,” / *Proceedings of the SIGKDD*, vol.14, issue 2, pp.1-5.
16. Ghemawat S., Gobioff H., Leung S.T. “The Google File System”, / *Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles (SOSP’03)*, 2003, pp. 29–43.
17. Dean J., Ghemawat S. “MapReduce: Simplified Data Processing on Large Clusters,” / *Proceedings of the Sixth Symposium on Operating System Design and Implementation*, vol.6, 2004, pp.137–150.
18. Hadoop. <http://hadoop.apache.org/>.
19. Stonebraker M. “Errors in Database Systems, Eventual Consistency, and the CAP Theorem”, // *Communications of the ACM*, vol.53, № 4, pp.10-11.
20. Anderson C. “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete”, *Wired Magazine*, July 2008. <http://www.wired.com/science/>
21. Черняк Л. Вычисления с акцентом на данные // *Открытые системы*, 2008, №8, с.36-39.
22. Wu X., Zhu X., Wu G.Q., Ding W. “Data mining with big data // *IEEE Transactions on Knowledge and Data Engineering*.” 2014, vol.26, no. 1, pp.97-107.
23. Zhang J., Huang M. L. “5Ws Model for Big Data Analysis and Visualization”, / *Proceedings of the IEEE 16th International Conference on Computational Science and Engineering (CSE)*, 2013, pp.1021–1028.
24. UN Global Pulse. <http://www.unglobalpulse.org>.
25. “Big Data Market Size and Vendor Revenues”. <http://wikibon.org>