

Böyük Verilənlər Analizinin Problemləri

Rəna Qasımova

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
depart1@iit.ab.az

Xülasə — *Big Data* mənbələrinin əsas hissəsini təşkil edən strukturlaşdırılmamış və qismən strukturlaşdırılmış böyük verilənlərin anazi ciddi problemlərdəndir. Məqalədə böyük verilənlər problemləri araşdırılır və onların analizi üçün bəzi metodlar təhlil olunur.

Açar sözlər — *verilənlər xəzinəsi, bulud, VBİS, big data, big data analytics, OLAP, Fast Data, Deep Insight.*

I. GİRİŞ

Zaman keçdikcə kompüter proqramları özünün çoxcəhətliyi ilə real dünyaya daha da yaxınlaşır. Emal olunmamış verilənlərin həcmində artması onların real vaxtda analizinin zəruriliyi ilə birlikdə *Big Data Analytics* məsələsini effektiv həll etməyə imkan verən alətlərin yaradılması və tətbiqini zəruri edir. Bunun nəticəsidir ki, son zamanlar böyük həcmdə informasiya mənbələrinin toplanması, eyni zamanda onların çox böyük sürətlə artması, həm akademik mühitdə, həm də informasiya texnologiyaları (İT) sənayesində daha çox diqqət cəlb etməyə başlamışdır. Miniwatts Marketing Group analitik şirkətinin apardığı statistik hesabatda əsasən 2015-ci ilin birinci rübündə 3 milyardan çox insan, yəni planetin əhalisinin 42,4%-i İnternetə qoşulmuş və mobil rabitə abunəçilərinin sayı 7,1 milyarda çatmışdır. 2020-ci ilə İnternetə qoşulan qurğuların sayının 50 milyard olacağı gözlənilir. 2012-ci ildə dünyada rəqəmsal informasiyanın həcmi 2,7 zətabayt olmuşdur. 2015-ci ildə bu həcm üç dəfə və növbəti hər il üçün 40% artması proqnozlaşdırılır [1-5].

Rəqəmsal informasiyanın belə sürətlə artımı, verilənlərin müxtəlifliyi, onların ötürülmə sürətinin yüksək artımı çoxsaylı problemlərin yaranmasına səbəb olur. Qeyd edildiyi kimi, artıq böyük verilənlərin (BV) saxlanması, real-vaxt rejimində emalı, analizi və idarə edilməsi problemlər yaratmışdır. Bununla belə, böyük verilənlər (BV) problemi hələ ilkin araşdırmalar səviyyəsindədir, yəni bu sahə hələ də tam olaraq təhlil olunmayıb. Aparılan tədqiqatlar “Big Data” anlayışını, onun mahiyyətini, müxtəlif xarakteristikalarını təsnifləndirməyi, BV-nin mənbələrini, bu texnologiyanın imkanlarını, problemlərini, təhlükəsizlik məsələlərini hərtəfli tədqiq etməyə imkan verir. Tədqiqatlar göstərir ki, BV-in emalı və analizi mükəmməl analitik texnologiyalar və alətlər tələb edir [6-8].

II. BİG DATA ANALİTİKA

Yüzlərlə terabayt və ekzabayt həcmində böyük verilənlərin mövcud metodologiyalarla və ya alətlərlə toplanması, idarə edilməsi, saxlanması və onlardan faydalı informasiyanın əldə edilməsi ciddi problemdir. Həm strukturlaşdırılmamış, həm də

strukturlaşdırılmamış informasiya ilə işləmək, daha dərin intellektual analiz aparmaq və analiz nəticələrini vizuallaşdırılmaq BV-nin analitikasının əsas məsələlərindəndir.

Verilənlərin həcmində artması və real zamanda onların analizinə olan tələbat BV-nin ən əsas problemlərindən sayılan böyük verilənlərin analitikası (Big Data Analytics) istiqamətinin yaranmasına gətirib çıxarmışdır [9-11]. Big Data Analytics müxtəlif tipli verilənlərdən ibarət olan böyük verilənlər yığımının öyrənilməsi prosesidir. Yəni, böyük verilənlərdən gizli qanunauyğunluqları, naməlum korrelyasiyaları və digər faydalı işgüzar informasiyanı aşkarlamaq üçündür. Analitik verilənlər daha səmərəli marketingə, yeni gəlir almaq imkanlarına, müştərilərə xidmət keyfiyyətinin yaxşılaşmasına, işin effektivliyinin artmasına, təşkilatların rəqabət və digər biznes üstünlüklərinə gətirib çıxara bilər.

Bu istiqaməti digər tətbiqlərdən fərqləndirən böyük həcm, sürət və mürəkkəbliyi kimi xarakteristikalar uyğun texnologiyalar tələb edir. Buna görə də, bu gün Big Data Analytics sahəsində əsas istehsalçılar xüsusi proqram-aparat sistemlərini təklif edirlər: SAP HANA, Oracle Big Data Appliance, Oracle Exadata Database Machine, Oracle Exalytics Business Intelligence Machine, Teradata Extreme Performance Appliance, NetApp E-Series Storage Technology, IBM Netezza Data Appliance, EMC Greenplum, HP Converged Infrastructure əsasında Vertica Analytics Platform. Bununla yanaşı kiçik və yeni başlayan şirkətlərin də böyük həcmli verilənləri səmərəli emal edən proqram-aparat alətləri vardır. Onlara Cloudera, DataStax, Northscale, Splunk, Palantir, Factual, Kognitio, Datameer, TellApart, Paracel, Hortonworks aiddir [12-14].

Verilənlər informasiya almaq üçün emal edilir, bu informasiyanın həcmi o qədər olmalıdır ki, insan onu biliyə çevirə bilsin. Həcm böyük verilənlərin ən əsas xarakteristikasıdır. Həcmindən asılı olaraq BV üç qrupa bölünür [15-18]:

- Tez (sürətli) verilənlər (Fast Data) – onların həcmi terabaytlarla ölçülür;
- Böyük analitika (Big Analytics) – onların həcmi petabaytlarla ölçülür;
- Dərinə nüfuz etmə (Deep Insight) – onların həcmi ekzabaytlarla və zətabaytlarla ölçülür.

Qruplar bir-birindən yalnız verilənlərin həcmində görə deyil, həm də onların keyfiyyətli emalına görə fərqlənilir. Statistik rəqəmlərdə əks olunan verilənlərin həcmi bir daha

mütəxəssisləri bu sahədə yeni metod və alətlərin işlənməsinə sövq edir.

Fast Data üçün emal yeni biliklərin alınmasını nəzərdə tutmur, onun nəticələri aprior biliklərlə əlaqələndirilir, bu və ya digər prosesləri necə keçməsinə nəzarət edir, eyni zamanda baş verənləri daha yaxşı və ətraflı şəkildə görməyə, hansısa hipotezləri təsdiq və ya inkar etməyə imkan verir. Yalnız mövcud texnologiyaların kiçik bir hissəsi *Fast Data* məsələlərinin həllinə yararır. Bunlara xəzinələrlə işləyən bəzi texnologiyaları göstərmək olar, məsələn, Greenplum, Netezza, Oracle Exadata, Teradata, Verica tipli verilənlər bazasının idarəetmə sistemləri (VBİS) və s. Bu texnologiyaların iş sürəti verilənlər həcmının artımı ilə sinxron artır.

Big Analytics vasitələri ilə həll olunan məsələlər kəmiyyət və keyfiyyətə çox fərqlənirlər. Uyğun texnologiyalar isə yeni biliklərin alınmasına və faydalı məlumatların əldə olunmasına kömək etməklə, verilənlərdə olan informasiyanı yeni biliyə çevirirlər. Başqa sözlə qərarın seçilməsində süni intellekt texnologiyaları nəzərdə tutulmur, analitik sistem “müəllimlə təlim” prinsipi üzrə qurulur və onun bütün analitik potensialı təlim prosesində ona tətbiq edilir. Belə analitikanın klassik nümunələri MATLAB, SAS, Revolution R, Apache Hive, SciPy Apache və Mahout məhsullarıdır [19-21].

Yüksək səviyyə, *Deep Insight* müəllimsiz təlimi (unsupervised learning) və analitikanın müasir metodlarının istifadəsini, həmçinin müxtəlif vizuallaşma üsullarını nəzərdə tutur. Bu səviyyədə aprior bilik və qanunauyğunluqların aşkarlanması mümkündür.

Keyfiyyət baxımından *Big Data Analytics* proqramları nəinki yeni texnologiyalar, həm də yeni düşüncə tərzini tələb edir. Analitikaya ilkin verilənlərin hazırlanması vasitələrindən, vizuallaşmadan və nəticələri insana təqdim edən digər texnologiyalarından ayrıca baxılır. Hətta The Data Warehousing Institute kimi təşkilatın analitikaya baxışları başqadır. Təşkilatın məlumatına görə hazırda müəssisələrin 38%-i idarəetmə praktikasında *Advanced Analytics* vasitələrindən istifadə imkanlarını tədqiq edirlər, 50%-i isə yaxın üç il ərzində bunu etməyi nəzərdə tuturlar. Qeyd etmək lazımdır ki, müəssisələrdə bu sahəyə belə maraq biznes sahəsindən çoxlu arqument gətirməklə əsaslandırılır. Belə ki, müəssisələrə yeni şəraitdə daha təkmil idarəetmə sistemi tələb olunur. Onun yaradılmasına əks əlaqə qurulmasından, yəni qərar qəbulunu dəstəkləyən sistemlərdən başlamaq tələb olunur ki, bunun da nəticəsində gələcəkdə qərar qəbulunu avtomatlaşdırmaq mümkün olacaqdır.

Texnoloji obyektlərin avtomatlaşdırılmış idarəetmə sistemlərinin yaradılması problemi heç də yeni deyildir. Bu verilənlər bazası sahəsindəki çoxdan mövcud olan mövzudur ki, paylanmış verilənlər bazası, resursların birgə istifadəsi arxitekturasının yaranması da bu məsələnin həllinə xidmət etmişdir [22].

Analiz üçün yeni vasitələr verilənlər mənbələrinin çox olmasına, həmçinin müxtəlif formatlarda (strukturlaşdırılmış, strukturlaşdırılmamış, qismən strukturlaşdırılmış) olmasına, həm də müxtəlif indeksləşmə sxemlərindən (relyasion, çoxölçülülük, noSQL) istifadə edilməsinə görə tələb olunur. Belə

ki, böyük həcmli verilənləri toplamaq, birgə emal etmək və analiz üçün uyğun şəkllə salmaq çox çətin olur.

Ənənəvi üsullarla verilənlərlə işləmək artıq mümkün olmur. *Big Data Analytics* daha böyük və mürəkkəb massivlərə tətbiq edildiyindən *Discovery Analytics* və *Exploratory Analytics* terminlərindən də istifadə edilir. Necə adlandırılmasından asılı olmayaraq, mahiyyət eynidir – qərar qəbul edən şəxsləri müxtəlif proseslər haqqında məlumatlarla təmin edən əks əlaqəni yaratmaq tələb olunur [23-26].

Müasir İT faktorları: böyük verilənlər, analitika və bulud texnologiyalarını bu gün bir-birindən ayrı təsəvvür etmək mümkün deyildir. Çoxsəviyyəli saxlama sistemlərinə artan tələbat, bulud texnologiyalarının real olaraq mövcudluğu həm də BV-nin analitikasına marağı artırmışdır. Bulud texnologiyaları böyük hesablamaların aparılmasında son dərəcə müvəffəqiyyətli yanaşmalardandır, buludlarda saxlamaq və buludlarda hesablamasız BV-lə işləmək mümkün deyildir. 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. İT sahəsindəki nəhəng şirkətlər yeni nəsil saxlama sistemlərində məhsul miqyaslaşma aspektlərinə və verilənlərin çoxsəviyyəli saxlanmasına böyük diqqət ayırmışdır [27, 28].

Praktika göstərir ki, bu gün analitik məsələlərin yerinə yetirilməsi üçün sistemləri çox yükləmək tələb olunur. Lakin biznes tələb edir ki, bütün servis, əlavələr və verilənlər həmişə əlçatan olmalıdır. Bundan başqa, hazırda analitik tədqiqatların nəticələrinə tələbat çox yüksəkdir. Çünki, savadlı, düzgün və vaxtında keçirilən analitik proseslər bütövlükdə biznes işinin nəticələrini əhəmiyyətli artırmağa imkan verir.

III. BÖYÜK VERİLƏNLƏRİN ANALİZİ PROBLEMLƏRİ

Bu gün BV-nin analizi üçün VX (Data Warehouse) texnologiyalarının yanaşma və metodlarından istifadə etməyə cəhd göstərilir. Bununla yanaşı ənənəvi metodların bəzi xüsusiyyətləri BV-nin emalı spesifikasiyasına zidd ola bilər. Verilənlərin operativ və analitik emalı məsələlərinin əhəmiyyətli fərqi verilənlər bazaları texnologiyalarının inkişafının ilk vaxtlarında yaranmışdı. VX termini 70-ci illərdə Bill İnmon tərəfindən təklif edilmiş, lakin bu texnologiyalara maraq 20 il keçəndən sonra, belə sistemlərə real tələbat yaranıqda və lazımı hesablama gücləri əlçatan olduqda baş vermişdi [29].

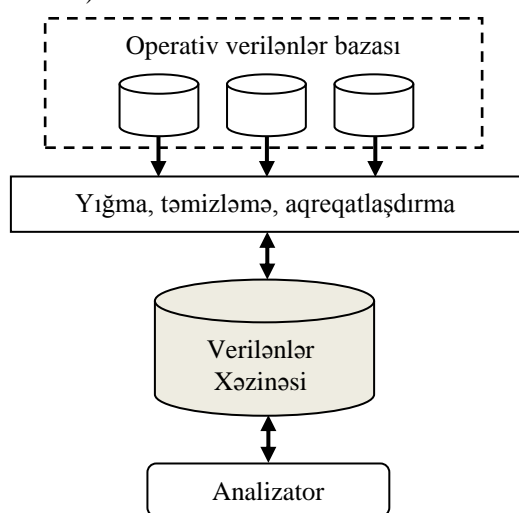
VX-də verilənlərin emalı mərhələləri verilənlərin *toplanmasından, təmizləmədən, yükləmədən, analizdən* və nəhayət analizin nəticələrinin *təqdimatından* ibarətdir. Bu mərhələlərin hər birində verilənlər üzərində xüsusi əməliyyatlar yerinə yetirilir. Qeyd etmək lazımdır ki, əgər BV-nin analizi üçün VX texnologiyalarının tətbiqinə cəhd olsa, onda yalnız alqoritmlərin analizinə deyil, həm də verilənlərlə işin bütün mərhələlərinə diqqət yetirmək lazımdır.

Verilənlərin toplanması. VX-nin yaradılmasında məqsəd ayrı-ayrı verilənlərin emal sistemlərində yığılmış verilənlərin təmizlənməsi, razılaşdırılması, inteqrasiya edilməsi və analizdə istifadə üçün rahat formaya gətirilməsidir (şəkil 1).

VX-də nəzərdə tutulur ki, informasiya operativ informasiya bazasından çıxarılır, lazımı şəkllə çevrilir, yoxlanılır və yalnız bundan sonra sistemə yüklənir. Yəni, VX-də verilənlərin hazırlanması texnologiyası bir-biri ilə əlaqəli üç mərhələdən ibarətdir [30]:

1. Verilənlərin toplanması (Data Acquisition).
2. Verilənlərin təmizlənməsi (Data Cleaning).
3. Verilənlərin aqreqatlaşdırılması (Data Consolidation).

Consolidation).



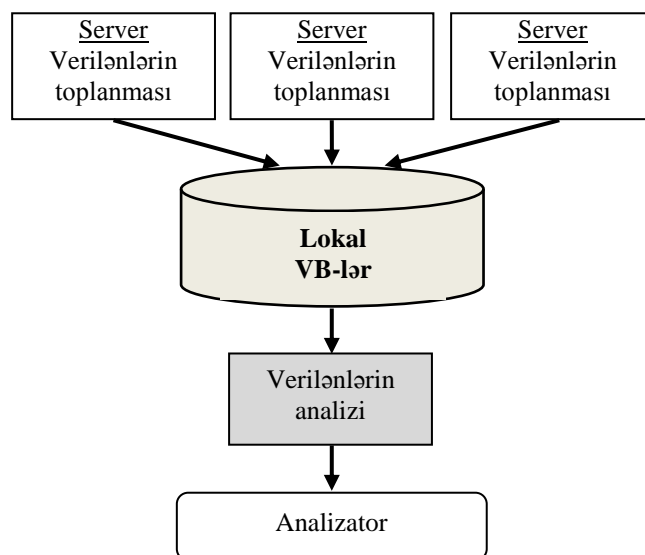
Şəkil 1. Ənənəvi analizdə verilənlərin xəzinəyə toplanması

Sadalanan əməliyyatlar dövriliklə yerinə yetirilir. Lakin BV-lərlə iş zamanı müxtəlif mənbələrdən girişə daxil olan verilənlərin analizini təmin etmək üçün belə dövrilik həmişə mümkün deyil. İnformasiyanın toplanması ilə onun analiz üçün ölçətanlığı arasındakı müddət VX-nin qurulması zamanı əməliyyatların yerinə yetirilməsinə lazım olan vaxtdan kiçik ola bilər. Belə məsələyə nümunə kimi sosial şəbəkələrdə tez yayılan informasiyanın və onun mənbələrinin müəyyən edilməsi, fəal istifadəçilərin təyin edilməsi, neqativ ifadələrin aşkarlanması və ya konfidensial informasiyanın sızması faktlarının aşkarlanması məqsədlə monitorinqi göstərilə bilər [31].

Məlumdur ki, bütün bu hadisələr daha tez aşkarlanmalı və neytrallaşdırılmalıdır. Lakin burada verilənlərin qeyri-formal təsviri mövcuddur ki, onların da emalı üçün mətnlərin intellektual analizi alqoritmləri tələb olunur. Məsələn, sosial şəbəkələrin monitorinqi məsələsində girişə istifadəçilərin ifadələri, şərhlər, qoyulmuş qiymətlər, fotoşəkillər və s. verilir. Aydın ki, məşhur sosial şəbəkələrin monitorinqini çoxlu sayda istifadəçilərlə keçirmək olar. Həm də aydındır ki, istifadəçilərin çox olması və onların yüksək fəallığı səbəbindən bu məsələnin həlli qeyri-formal şəkildə təsvir edilən böyük həcmli verilənlərin yığılımı və emalını nəzərdə tutur [32]. VX-də verilənlərin ilkin emalı zamanı (məsələn, uyğunsuzluqların axtarışında) xəzinənin əvvəl yığılımı tərkibinin (verilənlərinin) istifadə edilməsi nəzərdə tutulur ki, bu işə BV-lərlə iş zamanı çətin yerinə yetirilir. Yəni, problem ondadır ki, bu verilənlər həmişə paylaşılır, həm də tək analiz üçün deyil, yığılım üçün rahat olur. Məsələn, əgər söhbət

telekommunikasiya sistemlərindən gedirsə, onda verilənlər regional serverlərdə toplanır (şəkil 2).

Beləliklə, aparılan təhlillər göstərir ki, ənənəvi VX-lərdə bütün verilənlər həmişə vahid məntiqi blokdan keçir, burada onlar konvertasiya olunur, yoxlanılır, təmizlənir, yüklənir və bu əməliyyatların yerinə yetirilməsi vaxtı nadir hallarda bütün qalan sistemlər üçün vacib olur. Lakin BV-nin emalı zamanı belə vahid blok ola bilməz. Qeyd etmək lazımdır ki, hələ ki, intensiv girişli verilənlər axınına malik məsələlər çox deyil və paylaşmış, lakin məntiqi vahid sistem şəklinə yığılım, təmizləmə, çevirmə və yükləmə bloku reallaşdırmaq mümkündür.



Şəkil 2. Böyük verilənlərin analizi

Böyük verilənlərin analizi. Böyük həcmli verilənləri müxtəlif məqsədlər üçün operativ analiz etmək həmişə idarə edənlər üçün ciddi problem yaradır. Elmi-texniki ədəbiyyatın analizi göstərir ki, tətbiq olunan məhsullar, eləcə də digər analitik sistemlər hələ də tam olaraq bütün tələblərə cavab vermir. Gartner, McKinsey Global İnstitutu, IDC (International Data Corporation) və s. kimi kompaniyaların son tədqiqatlarında BV böyüyən, dinamik inkişaf edən sahə kimi təqdim edilir. McKinsey Global İnstitutunun analitik hesabatında BV-lərə tətbiq edilən analiz metodlarının adları verilmişdir [5]:

- **Data Mining sinfinin metodları** - assosiativ qaydaların öyrənilməsi (*association rule learning*), təsnifat, klaster analizi, rəqressiya analizi və s.;
- **onlayn analitik emal (OLAP)**;
- **kraudsorsinq (crowdsourcing)** - əmək münasibətləri olmadan, ictimai müqavilə əsasında cəlb edilən bir qrup şəxsin gücü ilə verilənlərin emalı;
- **verilənlərin birləşməsi və inteqrasiyası (data fusion and integration)** - dərindən analizi təmin etmək üçün müxtəlif mənbələrdən müxtəlif verilənləri inteqrasiya etməyə imkan verən texniki vasitələrdir. Buna nümunə olaraq siqnalların rəqəmsal emalı və təbii dilin emalını (tonal analiz daxil olmaqla) göstərir;

– **maşın təlimi** - bura müəllimlə və müəllimsiz təlim, həmçinin kollektiv təlim (Ensemble learning), baza modelləri əsasında kompleks proqnozlar almaq üçün statistik analiz və ya maşın təlimi əsasında qurulan modellərdən istifadə daxildir;

– **süni neyron şəbəkələri, şəbəkə analizi, optimallaşdırma**, o cümlədən **genetik alqoritmlər**;

– **obrazların tanınması**;

– **prediktiv analitika**;

– **imitasiya modelləşdirilməsi**;

– **fəza verilənlərinin analizi (Spatial analysis)** - verilənlərdə topoloji, həndəsi və coğrafi informasiyadan istifadə edən metodlar;

– **statistik analiz**, buna nümunə A/B testləşməsi və zaman sıralarının analizi göstərilir;

– **analitik verilənlərin vizuallaşdırılması** - nəticələrin alınması, sonrakı analiz, ilkin verilənlər kimi istifadə üçün informasiyanın şəkillər, diaqramlar, interaktiv imkanlardan və animasiyadan istifadə etməklə təqdimatı.

BV üzərində çoxölçülü əməliyyat aparmaq üçün klassik B-ağaclardan başlamış bütün mövcud metodların tətbiqi yararlıdır. Bu gün ənənəvi VX-lər verilənlərin analizi alətlərinin təxminən eyni dəstini təqdim edirlər [33-35]. BV-lərdə sadalanan metodları işləməyə imkan verən, operativ yaddaşda terabaytlarla informasiyanı analitik emal edən məhsullar da yaranmışdır, məsələn, SAP HANA, Greenplum Chorus, Oracle Exalytics, Oracle Exadata, Aster Data nCluster. Bundan başqa, Netezza, Teradata, Greenplum və s. şirkətlərinin ənənəvi relyasion verilənlərinin idarə edilməsi sistemi əsasında terabaytlar və ekzabaytlarla verilənləri səmərəli emal edən proqram-aparat alətləri vardır. Belə həllərin potensial imkanlarını başa düşmək üçün həmin metodların əsasında olan alqoritmlərə baxmaq, həmçinin onların BV-nin emalına açar olan mümkün paralel hesablamalar üçün yolları analiz etmək lazımdır. Bu zaman verilənlərin konkret paylanmış emalı texnologiyalarına bağlanmamalı, yalnız BV üçün xarakterik olan əsas parametrləri (sürət, müxtəliflik, həcm və s.) nəzərə almaq lazımdır. İstənilən VBIS-lərin iş sürətinə təsir edən mühüm faktor giriş/çıxış əməliyyatlarının sayı və yaradılmış indekslərin effektivliyidir.

Beləliklə, BV-lərlə işləyərkən məsələnin həllində ilk addım kimi emal olunacaq verilənlərin formatına, analiz tipinə, tətbiq olunan emal metodlarına, eyni zamanda məqsədli sistemin alacağı, yükləyəcəyi, emal və analiz edəcəyi, saxlayacağı verilənlərin mənbələrinə görə təsnifatlandırmaq tövsiyə olunur.

NƏTİCƏ

Verilənlərin həcmünün artması, onların real vaxtda analizinin zəruriliyi Big Data Analytics məsələsini effektiv həll etməyə imkan verən alətlərin yaradılması və tətbiqini tələb edir. Big Data Analytics vasitələrini və biznes-analitika texnologiyalarını əsasən İT nəhəngləri istifadə edirlər. Bu da biznes-analitikanın müəssisələrdə istifadəsinin aşağı mədəniyyətindən və biznes-istifadəçi tərəfindən mövcud analiz metodlarının qavrama çətinliyindən irəli gəlir. Bunu nəzərə alaraq bəzi şirkətlərin (məsələn, Information Builders)

analitikləri tərəfindən real vaxt rejimində istənilən mənbədən daxil olan verilənlərlə işləməyə imkan verən, istifadədə ən sadə qiymətləndirilən məhsul təklif edirlər.

Big Data sadə düşüncə deyil, gələcək texniki inqilabın simvoludur. BV-lərlə analitik iş zəruriliyi İT-sənayesini tamam dəyişəcək yeni proqram və aparat platformalarının yaranmasını stimullaşdırır. Artıq bu gün böyük həcmli verilənlərin analizi üçün ən qabaqcıl metodlar təmin edilir: süni neyron şəbəkələri – bioloji neyron şəbəkələrinin təşkili və funksiyası prinsipi üzrə qurulmuş modellər, prediktiv analitika, statistika, Natural Language Processing metodları və s. Həmçinin kraudsorsinq, A/B testləşmə, sentiment analiz və ekspertləri cəlb edən metodlardan istifadə edilir.

Nəticələrin vizuallaşdırılması üçün məlum metodlar, məsələn, bulud teqləri və ən yeni Clustergram, History Flow və Spatial Information Flow istifadə edilir. BV texnologiyası tərəfindən Google File System, Cassandra, HBase, Lustre və ZFS paylanmış fayl sistemləri, MapReduce və Hadoop proqram konstruksiyaları və digər qərarlar dəstəklənir. Ekspertlərin qiymətlərinə görə, BV-nin təsiri altında ən çox transformasiyaya istehsalat, səhiyyə, ticarət, inzibati idarəetmə uğrayacaqdır.

ƏDƏBİYYAT

- [1] Miniwatts Marketing Group, Worldwide Internet Market Research, www.miniwatts.com
- [2] The digital universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. Study report, IDC, December 2012. www.emc.com/leadership/digital-universe
- [3] Worldwide Big Data Technology and Services 2013-2017 Forecast, <http://www.idc.com>
- [4] Data Science Central, The online resource for Big Data practitioners, www.datasciencecentral.com
- [5] Big data: The next frontier for innovation, competition, and productivity. Analyst report, McKinsey Global Institute, May 2011. <http://www.mckinsey.com>
- [6] Madden S. From Databases to Big Data, IEEE Internet Computing, 2012, vol.16, No. 3, p. 4-6.
- [7] What is big data? - Bringing big data to the enterprise, 2013. <http://www-01.ibm.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] Черняк Леонид. Вычисления с акцентом на данные, Открытые системы, 2008, № 8, с. 36-39.
- [10] InfoSphere Platform: Big Data Analytics, 2013, <http://www-01.ibm.com/software>
- [11] Jacobs A. The pathologies of big data, Communications of the ACM, 2009, vol. 52, No. 8, p. 36-44.
- [12] Вахрамеев Кирилл. СУБД для анализа Больших Данных, Открытые системы, 2011, № 10, с. 26-29.
- [13] Babu S., Herodotou H. Massively Parallel Databases and MapReduce Systems3, Foundations and Trends in Databases, 2013, vol.5, No.1, p.1-104.
- [14] Vignesh Prajapati, Big Data Analytics with R and Hadoop, Publisher: Packt Publishing Ltd, 2013, p. 238.
- [15] Черняк Леонид. Большие Данные – новая теория и практика, Открытые системы, 2011, № 10, с. 18-25.
- [16] Черняк Леонид. Свежий взгляд на Большие Данные, Открытые системы, 2013, № 7, с. 48-51.
- [17] Krish Krishnan. Data Warehousing in the Age of Big Data. 1st Edition, Morgan Kaufmann Publishers Inc. San Francisco, USA, 2013, p. 370.

- [18] Билл Фрэнкс. Укрощение больших данных. Как извлекать знания из массивов информации с помощью глубокой аналитики, пер. с англ. Андрея Баранова, М.: Манн, Иванов и Фербер, 2014, 352 с.
- [19] Big Data - What Is It? 2013, <http://www.sas.com/big-data>
- [20] MathWorks, <http://www.mathworks.com/discovery/big-data-matlab.html>
- [21] Hadoop Distributed File System. <http://hadoop.apache.org/docs>
- [22] Witt D., Gray J. Parallel Database Systems: The Future of High Performance Database Systems, Communications of the ACM, 1992, vol. 35, No. 6, p. 85-98.
- [23] Билл Фрэнкс. Укрощение больших данных. Как извлекать знания из массивов информации с помощью глубокой аналитики, пер. с англ. Андрея Баранова, М.: Манн, Иванов и Фербер, 2014, 352 с.
- [24] Селезнев К. Проблемы анализа больших данных, Открытые системы, 2012, №7, с.25-29.
- [25] Gudivada V.N., Rao D., Raghavan V.V. NoSQL Systems for Big Data Management, Proceedings of the 2014 IEEE World Congress on Services (SERVICES '14), USA, 2014, p. 190-197.
- [26] Майер-Шенбергер Виктор, Кукьер Кеннет. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим, пер. с англ. Инны Гайдюк, М.: Манн, Иванов и Фербер, 2013, 240 с.
- [27] Əliquliyev R.M., Nəcirəhimova M.Ş. Big data fenomeni: problemlər və imkanlar, İnformasiya Texnologiyaları Problemləri, 2014, №2, səh. 3-16.
- [28] Marcos D. Assunção, Rodrigo N., Silvia Bianchi, Marco A.S. Netto, Rajkumar Buyya. Big Data computing and clouds, Journal of Parallel and Distributed Computing, 2015, vol. 79, No. C, p. 3-15.
- [29] Inmon W. H. “Building the Data Warehouse,” 3rd Edition, John Wiley & Sons, Inc., New York, 2002, p. 412.
- [30] Əliquliyev R.M., Qasimova R.T., Ələkbərova İ.Y. Qərarların qəbul edilməsini dəstəkləyən müasir konsepsiyalar haqqında, AMEA-nın Xəbərləri, fizika-riyaziyyat və texnika elmləri seriyası, 2005, №2, s.70-75.
- [31] Tonkin E.L., Pfeiffer H.D. Zombies Walk Among Us: Cross-Platform Data Mining for Event Monitoring, Proceedings of the 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW '13), USA, 2013, p. 452-459.
- [32] Krishna Kumar K.P., Geethakumari G. A taxonomy for modelling and analysis of diffusion of (mis)information in social networks, International Journal of Communication Networks and Distributed Systems, Switzerland, 2014, vol. 13, No.2, p. 119-143.
- [33] Alguliev R.M., Gasimova R.T. Identification of Categorical Registration Data of Domain Names in Data Warehouse Construction Task, Intelligent Control and Automation, 2013, vol. 4, No. 2, p. 227-234.
- [34] Alguliev R.M., Gasimova R.T. On a approach for intellectual analysis of registration data of domain names, International Journal of Ubiquitous Computing and Internationalization, 2011, vol. 3, No. 1, p. 27-30.
- [35] Ordonez C. Can we analyze big data inside a DBMS?, Proceedings of the sixteenth international workshop on Data warehousing and OLAP (DOLAP '13), USA, 2013, p. 85-92.