

UOT 004.62

Hacırahimova M.Ş.¹, İsmaylova M.İ.²

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

¹makrufa@science.az, ²imarziya@gmail.com

BÖYÜK VERİLƏNLƏRİN SAXLANILMA PROBLEMLƏRİ VƏ MÖVCUD HƏLLƏR

Məqalə böyük verilənlərin saxlanılma problemlərinin və mövcud həllərin analizinə həsr olunmuşdur. Verilənlərin saxlanması və emalı problemi hər zaman mövcud olmuşdur. İnformasiya-kommunikasiya texnologiyaları vasitəsilə yaradılan böyük həcmli verilənlər bu gün saxlama problemini bir daha gündəmə gətirmişdir. Məqalədə verilənlərin mövcud saxlanma sistemləri müqayisəli şəkildə araşdırılmış, böyük verilənlərin saxlanılma problemləri və mövcud yanaşmaları analiz edilmişdir.

Açar sözlər: böyük verilənlər, böyük verilənlərin saxlanması, NoSQL verilənlər bazaları, verilənlər mərkəzi, bulud hesablamaları, bulud saxlanıcı.

Giriş

Son onillikdə informasiyaya əlyətərliyin asanlaşması və kommunikasiya vasitələrinin çoxalması nəticəsində hər il rəqəmsal informasiyanın həcmi həndəsi silsilə şəklində artmağa başlamışdır. IDC (*ing. International Data Corporation*) analitik şirkətinin məlumatına görə, dünyada yaradılmış və replikasiya olunmuş informasiyanın həcmnin təxminən 7,5 zətabayta bərabər olması və informasiyanın 90%-dən çox hissəsini strukturlaşdırılmamış verilənlərin (foto, video, mp3 faylları, sosial media faylları və s.) təşkil etdiyi bildirilir, 2020-ci ildə dünyada rəqəmsal informasiyanın həcmnin 44 zətabayta çatması haqqında proqnoz verilmişdir [1, 2]. Bu rəqəmlər həqiqətən də verilənlərin həcmnin çox böyük olduğunu bir daha sübut edir.

Böyük verilənlər (BV) həcm, sürət və müxtəliflik ilə xarakterizə olunur. Həcm BV-nin ən əsas xarakteristikasıdır. Həcm ilk növbədə saxlama problemi yaradır ki, bu da genişmiqyaslı saxlanma və paylanmış emal tələb edir. İnformasiyanın həcmi həndəsi silsilə şəklində artdığı üçün onun etibarlı saxlanması, təhlükəsizliyinin təmini, idarə olunması və ona operativ müraciətin təşkili hər bir təşkilatın qarşılaşdığı problemdir [1].

BV elmi-tədqiqat, idarəetmə və biznes proseslərə böyük təsir göstərmiş və onun saxlanması üçün qənaətbəxş texnologiyaların tətbiqi problemlərinin həlli mütəxəssislərin diqqətini cəlb etmişdir. Relyasion verilənlər bazaları (RVB) uzun illər ərzində strukturlaşdırılmış verilənlərin saxlanması və əldə edilməsi nöqtəyi-nəzərindən çox səmərəli idi. Ancaq strukturlaşdırılmış verilənlərdən strukturlaşdırılmamış verilənlərə keçid tendensiyası ilə əlaqədar olaraq ənənəvi RVB BV-nin saxlanması üçün əlverişli hesab olunmur. RVB-nin bu uyğunsuzluğu effektiv paylanmış saxlanma mexanizmlərinin yaradılmasını zəruri etmişdir.

Yeni modellərin, optimallaşma üsullarının, təhlükəsizlik modellərinin işlənməsi BV-nin saxlanması və əlyətərliyinin təmin edilməsinin problemlərindəndir. Belə ki, həm saxlanma sistemləri arxitekturu səviyyəsində (bulud saxlamaları və ehtiyat surətlərin alınması, paylanmış saxlanclar), həm də fiziki səviyyədə yeni həllər tələb olunur.

Böyük verilənlərin saxlanma qurğuları

BV əsri üçün analitik alətlərlə yanaşı böyük həcmli və sürətli saxlama sistemləri tələb olunur. Ümumiyyətlə, verilənlərin saxlanma qurğuları 4 tipə bölünür: sərt disklər (*ing. Hard Disk Drive, HDD*), optik disklər (*ing. Optical Disk Drive, ODD*), sərtgövdəli qurğular (*ing. Solid State Drive, SSD*) və maqnit lentlər [3].

Sərt disklər. HGST (*ing. Hitachi Global Storage Technologies*) disklərin satış şirkətinin tədqiqatlar üzrə vitse-prezidenti K.Muncenin (*ing. Currie Munce*) proqnozuna görə, sərt disklər ("onlar hər yerdə var, praktikdir və ucuzdur") 2020-ci ilə qədər əsas saxlama sistemi olaraq qalacaqdır. Sərt disklərin tutumunun artırılması istiqamətində daima işlər davam etdirilir. Hal-

hazırda istifadə olunan iri sərt diskin tutumu 4 terabaytdır. Sərt diskilər üçün nanolitoqrafiya, helium diskilər və HAMR (ing. Heat-assisted magnetic recording) kimi üç yeni yanaşma daha çox imkanlar vəd edir.

Nanolithography – iki innovativ “*nanoimprinting*” və “*molecularself-assembly*” kimi texnologiya vasitəsilə yaradılan sərt disk yazı sıxlığını artırmaqla disk tutumunu ikiqat böyüdür.

Helium drives. 2013-cü ilin sonunda HGST şirkəti helium elementi ilə hazırlanmış ilk sərt disk platformasını təqdim etmişdir. İstehsalçının qeyd etdiyinə görə, bu texnologiya verilənlərin emal edilmə mərkəzi və bulud infrastrukturunu üçün sərt diskin tutumunu və onun ümumi dəyərini əhəmiyyətli dərəcədə artırmağa imkan verir. Bu innovasiya sayəsində 2017-ci ildə təqdim edilən HDD-də 1 kvadrat düym-ə düşən yazı sıxlığının 5 terabayta bərabər olduğu bildirilmişdir. Hava ilə hazırlanmış diskilərdən fərqli olaraq, helium elementli diskilər daha tutumludur. K.Munceyə görə, bu disk yeddi lövhə dəstəkləməyə imkan verir, həm də enerjiden istifadəni 23 %-ə qədər azaldır. Ancaq hava ilə hazırlanmış disk ən çoxu beş lövhəyə malikdir.

HAMR texnologiyası sərt diskin tutumunu artırmağa imkan verir. Bu texnologiya ilə hazırlanmış diskilər, hazırda istifadə olunan diskilərdən fərqli olaraq, verilənləri böyük sıxlıqla saxlamağa imkan verir.

Verilənlərin saxlanması üçün sərt diskilərin ən böyük istehsalçılarından olan “Seagate Technology” şirkəti laboratoriyaya sınaqlarında disk sahəsinin hər kvadrat düym-ə 1 terabayt olmaqla saxlama sıxlığını artırmaq üçün HAMR-dan istifadə edir. Bugünkü diskilər hər kvadrat düym-ə maksimum 620 qıqabayt sıxlıq təklif edir. Şirkət, həm də HAMR texnologiyalı 6 terabayt informasiya saxlayan sərt diskilərin ilk məhsulunun 2017-ci ildə satışa çıxarıldığını bəyan etmişdir [3, 4].

Sərt gövdəli qurğular. Təşkilat üçün ciddi problemlərdən biri böyük həcmli verilənlərin dinamik olması və onların sürətli əldə edilməsidir. Bu problem SSD diskilərindən istifadə edilməklə həll edilir, belə ki, onlar sərt diskilərdən fərqli olaraq hərəkət edən hissələrə malik olmadığından verilənlərə sürətli müraciət əldə edilir və analiz üçün işçi yaddaş qismində istifadə olunur. Hazırda tədqiqatçılar SSD üçün yeni yanaşmalar üzərində işləyirlər. Məsələn, 3D NAND qurğusu əhəmiyyətli dərəcədə saxlama sıxlığını artırır və ilk qurğunun 2016-cı ildə istehsal olunduğu bildirilir [3–5].

Maqnit lentlər. Qeyd etmək lazımdır ki, saxlama texnologiyalarının inkişafına baxmayaraq, maqnit daşıyıcıları əvvəlki kimi qiymət (həcm) cəhətdən ən yaxşı həll kimi təklif edilir, bundan başqa, onlar yüksək etibarlılığa malikdirlər və xidmətdə çox yararlı hesab edirlər. Bu səbəbdən, onlardan hələ də, xüsusilə nadir hallarda istifadə edilən verilənlərin arxivləşdirilərək saxlanması üçün geniş istifadə edilir [5].

Optik diskilər. Optik diskilər BV-nin saxlanması üçün çox da effektiv deyildir. Yaddaşın həcmi artırmağa imkan verən daha yüksək sıxlığa nail olmaq üçün holoqrafik, yaxud çoxlaylı həllərdən istifadə edilməlidir. Bu da mürəkkəb texniki problemdir [5].

Bulud saxlamaları

Bu gün saxlama məsələsinin həllində informasiyanın qurğular arasında miqyasını həyata keçirən bir sıra texnologiyalar mövcuddur [1, 6, 7]: DAS (ing. *Direct-Attach-Storage*), NAS (ing. *Network Attached Storage*), SAN (ing. *Storage Area Networks*), HSM (ing. *Hierarchical Storage Management*), ILM (ing. *Information Life-cycle Management*). Lakin bu texnologiyalar BV əsri üçün o qədər də yararlı deyildir. Son zamanlar isə saxlama qurğularının yaddaş tutumunun artması, çoxsaylı kompüterlərin (server, kompüter və s.) hesablama 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, nisbətən aradan qaldıra bilmişdir [1, 8].

Müasir informasiya texnologiyalar (İT) faktorları: BV, analitika və bulud texnologiyalarını bu gün bir-birindən ayrı təsəvvür etmək mümkün deyildir. Buludlarda saxlama, buludlarda hesablamaşız BV ilə işləmək mümkün deyildir. Qeyd etmək lazımdır ki, bulud texnologiyaları

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 [1, 9].

Buludda saxlama əksər təşkilatlara böyük həcmli informasiyaların saxlanmasına imkan verir. Belə ki, təşkilatların çoxu öz informasiyalarını müstəqil saxlamaqdan yayınaraq buludda, BV-nin emalı mərkəzində saxlayırlar. Bulud saxlama xidməti tələb üzrə resursların ödəniş əsasında əldə edilməsinə xidmət edən ehtiyat və saxlama modelidir. Bu, əsasən böyük sayda verilənlər toplayan və onların saxlanması üçün resursları çatışmayan təşkilatlar üçün sərfəlidir [10, 11].

Bulud saxlama növləri. İstifadəçilərə təqdim olunan bulud saxlamanın müxtəlif formaları mövcuddur. Onlardan aşağıdakıları fərqləndirmək olar:

Şəxsi bulud saxlama (ing. *Personal Cloud Storage*) – və ya mobil bulud saxlama kimi tanınmış saxlama növü buludda informasiyanın fərdi qaydada saxlanması və ona istifadəçilərin dünyanın istənilən nöqtəsindən daxil olmasını nəzərdə tutur. Bu da verilənlərin çoxsaylı qurğular vasitəsilə sinxronlaşdırılması və mübadilə imkanlarını təmin edir. iCloud şəxsi bulud saxlamanın bariz nümunəsidir.

Ümumi bulud saxlama (ing. *Public Cloud Storage*) – müəssisənin və saxlamanın servis provayderləri ayrılır və müəssisənin verilənlərin emalı mərkəzində saxlanan heç bir bulud resursu yoxdur. Bulud saxlama provayderi tamamilə müəssisənin şəxsi bulud saxlamalarını idarə edir.

Özəl bulud saxlama (ing. *Private Cloud Storage*) – vahid təşkilat daxilində fəaliyyət göstərən bulud infrastrukturudur. Burada bütün resurslar toplusu yalnız bir təşkilata məxsus olur.

Hibrid bulud saxlama (ing. *Hybrid Cloud Storage*) – dövlət və özəl bulud saxlamalarının birləşməsidir. Bir sıra qiymətli məlumatlar müəssisənin özəl buludunda saxlanılır. Eyni zamanda digər məlumatlar saxlama provayderi ilə ictimai buludda saxlanılır [11–13].

Bulud saxlamanın üstünlükləri və çatışmazlıqları. Verilənlərin buludda saxlanması başlıca üstünlükləri ondan ibarətdir ki, aparat və proqram təminatına çəkilən xərclər azalır. Elektrik enerjisi üçün xərclər, serverlər üçün sahə, sistem administratorlarının işi və s. bulud saxlama provayderi tərəfindən aparılır. Verilənlərin bulud saxlama üstünlüklərinə aşağıdakılar aiddir:

- Fərdi kompüterlərin hesablama gücünə olan tələbat aşağı düşür (vacib şərt yalnız İnternetə girişin olmasıdır);
- İmtinalara davamlılıq;
- Təhlükəsizlik;
- Verilənlərin qorunması və əlyətərlik saxlama arxitekturu məxsusdur;
- Verilənlərin emalında yüksək sürət;
- Disk fəzasına qənaət (verilənlər və proqramlar İnternetdə saxlanılır).

Bulud saxlama çatışmazlıqlarına isə aşağıdakıları aid etmək olar:

- Saxlanan verilənlərin bulud hesablamaları xidməti təqdim edən şirkətlərdən asılılığı;
- Yeni bulud inhisarçıların meydana gəlməsi;
- Müxtəlif bulud saxlama provayderləri arasında qarşılıqlı əlaqəni təmin edən ümumi qəbul edilmiş standartların olmaması [11].

Böyük verilənlərin müasir saxlama sistemləri

Relyasion verilənlər bazaları. Son bir neçə onillik ərzində strukturlaşdırılmış verilənlərin idarəetmə texnologiyaları kimi RVB-dən istifadə edilmişdir. Lakin verilənlərin həcmnin artması ilə meydana çıxan problemlərin öhdəsindən klassik arxitektura malik RVB gələ bilmədiyi üçün bu sahənin mütəxəssisləri yeni həllər düşünüb tapmaqdadırlar. Belə ki, RVB yerinə yetirilən əməliyyatların həcmnin öhdəsindən gəlmədikdə aşağıdakı addımların yerinə yetirilməsinə cəhd etmək lazımdır [14]:

1. İlk addım ən az xərc tələb olunan üsulun yoxlanılmasıdır. Bu, maliyyə imkanları olduqda, daha güclü avadanlıq – server (şaquli miqyaslaşdırma) almaq üsuludur. Lakin bu üsul sonsuzluğa qədər davam edə bilməz.

2. Daha çox xərc sərf edilən üsul – sorğuların icra edilmə planını təhlil etmək və əlavə indekslər yaratmaqla sorğuları optimallaşdırmaqdan ibarətdir. Bu üsul müvəqqəti olaraq asanlıq yaradır, ancaq əlavə indekslər əlavə əməliyyatlar yaradır və emal edilən verilənlərin həcmnin artması ilə bu əlavə əməliyyatlar tənəzzülə səbəb olur.

3. Növbəti addım oxuma əməliyyatlarında keşin tətbiqidir. Belə bir həllin düzgün təşkili verilənlərin uyğunluğunu (*ing. consistency*) pozmaqla RVB-ni oxuma əməliyyatlarının əhəmiyyətli bir hissəsindən azad edir. Bu yanaşma isə kliyent proqram təminatının mürəkkəbləşməsinə səbəb olur.

4. Yerləşdirmə/yenilənmə əməliyyatlarının növbəyə düzülməsi – pis qərar hesab edilmir, ancaq növbənin ölçüsü məhduddur. Bundan başqa, uyğunluğu təmin etmək üçün növbənin özünün daimiliyini (*ing. persistence*) təşkil etmək lazımdır, ancaq bu asan bir məsələ deyil.

5. Nəhayət, bütün digər üsullar işləməyincə, verilənlərin özünün necə təşkil edildiyini yenidən nəzərdən keçirmək lazımdır. İlk növbədə – lazımsız müraciətlərin sayını azaltmaq üçün normallaşma sxemindən imtina etmək lazımdır.

6. Lakin, bu üsul da işləməsə, onda hesablamaları müxtəlif qovşaqlara paylamaq, yəni üfüqi miqyaslama aparmaq lazımdır. Bu halda normallaşma və xarici açarlardan tamamilə imtina etmək lazımdır. Bu zaman “yeni kortejləri hansı əlamətə görə qovşaqlar üzrə paylamaq” və “mövcud sxemi necə miqrasiya etmək” kimi suallar cavablandırılmalıdır. Nəticədə RVB-nin BV ilə işləməyə uyğunlaşdırılması cəhdləri aşağıdakılara səbəb olmuşdur:

- Verilənlərin uyğunluğundan imtina etmək;
- Normallaşma sxemindən və informasiya artıqlığından qaçmaq;
- SQL dilinin ifadəliliyinin itməsi və onun funksiyalarının bir hissəsinin proqramla modelləşdirilməsi;
- Kliyent proqram təminatının əhəmiyyətli dərəcədə mürəkkəbləşməsi;
- Alınmış qərarın işləmək qabiliyyətinin və imtinaya davamlılığının dəstəklənməsinin çətinliyi.

NoSQL verilənlər bazaları. Qeyd etmək lazımdır ki, RVB istehsalçıları yuxarıda sadalanan bütün bu problemləri dərk edərək artıq miqyaslanı bilən klaster həllər təklif etməyə başlamışdılar. Verilənlərin artan həcmi və onların effektiv idarə edilməsinə uyğunlaşdırılmış arxitektura layihənin hazırlanması NoSQL (*Not Only SQL*) saxlama sistemlərinin meydana çıxmasına səbəb olmuşdur. NoSQL terminindən ilk dəfə 1998-ci ildə SQL-dən istifadə etməyən RVB-ni təsvir etmək üçün istifadə olunmuşdur. 2009-cu ildə böyük miqdarda veb startapların meydana çıxması NoSQL saxlama sistemlərini populyarlaşdırdı. NoSQL sistemlərin əsas xüsusiyyətləri aşağıda sadalanmışdır [14, 15]:

1. *Lazımsız mürəkkəbləşdirmənin kənar edilməsi.* RVB çox böyük sayda müxtəlif funksiyaları yerinə yetirir və verilənlərin uyğunluğunu təmin edir. Lakin bir çox əlavələr üçün belə funksiyalar dəstinin və ACID (*ing. atomicity, consistency, isolation, durability*) tələblərinin ödənilməsi gərəksizdir.

2. *Yüksək ötürmə qabiliyyəti.* Bir çox NoSQL həlləri ənənəvi RVB-dən fərqli olaraq verilənlərin daha yüksək ötürmə qabiliyyətini təmin edir. Misal üçün, Google Bigtable yanaşmasını reallaşdıran sütun şəkilli Hypertable saxlancısı Zvent axtarış sürgüsünə gün ərzində milyarda yaxın yazını saxlamağa imkan verir. Digər bir misal kimi gün ərzində 20 petabayt informasiya emal etmək qabiliyyətinə malik Bigtable VB-ni göstərmək olar.

3. *Qeyri-məhdud üfüqi miqyaslama.* RVB-dən fərqli olaraq, NoSQL həllər qeyri-məhdud üfüqi miqyaslama üçün nəzərdə tutulmuşdur. Bu halda qovşaqların klasterə əlavə edilməsi və kənarlaşdırılması sistemin iş qabiliyyətinə təsir etmir. Belə arxitekturanın digər bir üstünlüyü NoSQL klasteri adı aparat təminatına yerləşdirməklə bütövlükdə sistemin qiymətini əhəmiyyətli dərəcədə aşağı salmaqdır.

4. *Uyğunluqdan imtina etməklə yüksək məhsuldarlığın əldə edilməsi.* NoSQL bu gün BV aləminin əsası hesab olunur və verilənlərin idarə edilməsinin miqyaslılıq (*ing. scalability*),

əlyetərlik (*ing. availability*) və verilənlərin uyğunlaşdırılması (*ing. consistency*) kimi problemlərinin həllində tətbiq olunur [15]. Ədəbiyyatlarda paylanmış sistemlərin bu üç xüsusiyyəti Berkli Universitetinin professoru Eric Brewer tərəfindən təklif olunmuş CAP (*Consistency, Availability u Partition Tolerance*) teoremi kimi də tanınır. Bu teoremə görə, NoSQL verilənlər bazaları (VB) verilənlərin uyğunluğundan imtina etməklə yüksək məhsuldarlığın əldə edilməsi (*ing. availability*) və bölünməyə dayanaqlılığı (*ing. partition tolerance*) həyata keçirir.

NoSQL verilənlər bazalarının klassifikasiyası. Bu gün çoxlu sayda NoSQL həllər yaradılmışdır. Onların hamısı yuxarıda sadalanan dörd əsas prinsipə əsaslanır, lakin bir-birindən əsaslı surətdə fərqlənirlər. Bir çox nəzəriyyəçilər NoSql VB-nin verilənlər modellərinə görə klassifikasiyalarını yaratmışlar. Lakin verilənlər modelindən istifadəyə əsaslanan ən sadə və hamının qəbul etdiyi sistem VB-nin müstəqil məsləhətçisi Rick Kettel (*ing. Rick Cattell*) tərəfindən verilmişdir. Aşağıda verilənlər modellərinin tipləri şərh edilmişdir [16–20]:

1. *Açar-qiyət saxlama (ing. key-value stores).* “Açar-qiyət” VB BV-nin saxlanması üçün işlənilib hazırlanmışdır və yalnız strukturlaşdırılmış verilənləri deyil, xüsusilə də strukturlaşdırılmamış verilənləri açar və hər bir yazının veriləninə uyğun olan qiymət şəklində saxlayır. Bu VB sorğuların daha sürətli cavablandırılması tələb olunan BV-nin daim artan, ardıcıl olmayan qiymətləri üçün uyğun saxlama strukturuna malikdir. Bloklarla saxlama sistemlərindən fərqli olaraq, “açar-qiyət” verilənlər strukturunda verilənlər kiçik obyektlərdə saxlanılır və onlar asanlıqla konfigurasiya edilə bilirlər. “Açar-qiyət” VB-də yazılardakı qiymətlər fərqli ola bilər və ya müxtəlif cür təqdim edilə bilər. Bu struktur böyük sayda yazıların əlavə edilməsi zamanı daha az yaddaş sərfini və çevikliyi təmin edir. “Açar-qiyət” saxlamanın ən çox yayılmış əlavələrinə misal olaraq onlayn oyunları, onlayn alış-verişləri və s. göstərmək olar.

2. *Sənəd yönümlü saxlama (ing. document-oriented stores).* Sənəd yönümlü verilənlər modeli “açar-qiyət” strukturuna uyğundur və verilənlər sənədə istinad kimi “açar-qiyət” şəklində saxlanılır. Lakin sənəd yönümlü VB daha mürəkkəb sorğuları və iyerarxik əlaqələri dəstəkləyir. Bu verilənlər modeli adətən JSON (JavaScript Object Notation) formatını həyata keçirir və çox çevik sxem təklif edir. Sənəd yönümlü VB sonrakı optimallaşdırma üçün meta-verilənləri çıxarır və onları sənəd kimi saxlayır. Sənəd yönümlü VB-yə sosial şəbəkələrdə istifadəçi hesabları, veb saytların analizi və mürəkkəb tranzaksiya verilənləri aiddir. Məsələn, Facebook səhifəsində istifadəçilər öz hesabları haqqında qısa və ya geniş məlumat təqdim edə bilirlər. Bu zaman hər bir istifadəçi hesabı sənəd kimi saxlanılır.

3. *Sütun yönümlü saxlama (ing. column-oriented stores).* NoSQL VB-nin bu kateqoriyası şaquli bölünmüş, bir-biri ilə əlaqəli və sıxılmış saxlama sistemləri üçün uyğundur. Verilənləri yazılar şəklində saxlanan ənənəvi saxlama sistemlərindən fərqli olaraq, sütun yönümlü VB verilənlərin sütunlarını ayrı-ayrılıqda saxlayır. Bu sistemlərdə atributların oxunması və əldə edilməsi olduqca sürətli və nisbətən daha ucuzdur, çünki yalnız müvafiq sütun istifadə olunur və hər bir sütun üçün proses birbaşa icra edilir. Sütun yönümlü VB yüksək dərəcədə miqyaslanabilir və dayanıqlıdır. O, etibarlı və yüksək əlçatanlı saxlama əlavələrinin dəstəklənməsini təmin edir. Sütun yönümlü VB müştəri yazılarının analizi, verilənlər anbarı, pasiyent verilənlərinin idarə olunması, kitabxana sistemləri və verilənlərin oxşar elementlərinin toplanması üçün analiz tələb olunan hər bir sahəyə tətbiq oluna bilər. Bundan başqa, sütun yönümlü strukturda bütün sətirlərə yeni xüsusiyyətlər əlavə etmək əlverişlidir. Məsələn, bir onlayn alış-veriş saytı müəyyən bir vaxtda ən çox baxılan və sifariş edilən mallar, onlayn alış-veriş üçün dəbdə olan bölgələr və bir il ərzində onlayn ticarətin uğur qazanma dərəcəsinə görə aqreqasiyanı tətbiq edə bilər.

4. *Qraf saxlama (ing. Graph stores).* Qraf VB verilənləri əlaqələrlə birlikdə saxlamaq üçün ən yaxşı seçimdir. Qraf VB obyektlərin və əlaqələrin daimi saxlanmasını təmin edir və onların sintaksisləri ilə bir yerdə sadə və anlaşılabilir sorğuları dəstəkləyir. Müasir müəssisələr, öz mürəkkəb biznes prosesləri və qarşılıqlı əlaqəli verilənləri üçün qraf VB-ni tətbiq edirlər, çünki bu əlaqəli verilənlərin strukturu asan keçidi təmin edir. Çox sıx ticarət sistemləri və təkliflər sistemləri

gecikmənin azalmasına nail olmaq üçün qraf VB-ni üstün tuturlar. Məsələn, ənənəvi VB üçün bir alış-veriş veb-saytında müştəri ilə əks əlaqənin məlumatına görə təkliflərin alınması müstəqil çoxsəviyyəli sorgular tələb edir ki, bu da çox mürəkkəb bir prosesdir. Əksinə, qraf VB üçün bu verilənlərin manipulyasiyası verilənlərin strukturuna təsir etmədən iki sətirlik kod sətiri olaraq olduqca bəsitdir. Verilənlər modellərinə görə, NoSQL VB-nin klassifikasiyası aşağıdakı cədvəldə təqdim edilmişdir.

Cədvəl 1

NoSQL VB-nin verilənlər modelləri üzrə klassifikasiyası

Verilənlər modelinin tipi	NoSQL VB
Açar-qiymət	Redis, DynamoDB, Scalaris, Riak, Tokyo Tyrant
Sənəd yönümlü	SimpleDB, CouchDB, MongoDB, Terrastore
Sütun yönümlü	BigTable, HBase, HyperTable, Cassandra
Qraf	HyperGraphDB, Neo4j, OrientDB

Onlardan bəziləri aşağıda şərh edilmişdir [17, 19]:

DynamoDB – 2012-ci ildə Amazon şirkəti tərəfindən təqdim olunan bulud xidmətidir. DynamoDB verilənlər modeli “açar-qiymət” tipli sistemlər üçün kifayət qədər çevik və zəngindir. DynamoDB istifadəçilərə replikasiyaları və bölünmələri avtomatik olaraq həyata keçirən sürətli və miqyaslanan verilənlərin idarəetmə sistemlərini təqdim edir. Sistemin yüksək məhsuldarlığı SSD qurğularının hesabına əldə edilir. Bulud xidmətlərinin istifadəsi ilə istifadəçilər serverlərin quraşdırılması və idarə edilməsi yükündən azad olur, yalnız trafik və saxlanılan informasiyaların həcmi ödənilir. DynamoDB ilə iş üçün müştəri kitabxanaları Java, .NET, PHP, Perl, Python, Ruby və s. daxil olmaqla çoxlu sayda proqramlaşdırma dilləri üçün əlverişlidir.

Redis – C proqramlaşdırma dilində yazılmış və eyni zamanda bu sistemlər üçün kifayət qədər zəngin bir verilənlər modelini dəstəkləyən açıq mənbəli “açar-qiymət” tipli verilənlərin idarə edilmə sistemidir. Redis-də verilənlərin əməli yaddaşda (RAM) saxlanması hesabına yüksək məhsuldarlıq əldə edilir. Uzunmüddətli saxlamanın təmin olunması üçün zamanın müəyyən anlarında verilənlərin sürətlərinin alınması və ya diskə verilənlərin modifikasiya əməliyyatlarının daimi yazılması tətbiq edilə bilər; həmçinin disklərdən istifadə etmədən də işləmək mümkündür. Üfüqi miqyaslama verilənlərin bölünməsi yolu ilə əldə edilə bilər (məntiq müştəri tərəfdə reallaşır). Redis müştəri kitabxanaları proqramlaşdırma dillərinin əksəriyyəti üçün əlverişlidir, sistemin özü isə Twitter, Instagram, StackOverflow, Flickr və s. kimi iri layihələrdə istifadə olunur.

MongoDB – C++ proqramlaşdırma dilində yazılmış və 10gen şirkəti tərəfindən hazırlanmış açıq mənbəli sənəd verilənlər bazalarının idarəetmə sistemidir. MongoDB kifayət qədər zəngin funksiyalara malikdir və hal-hazırda ən populyar NoSql sistemlərdən biridir. MongoDB-də miqyaslama seçilmiş açar əsasında qovşaqlar üzrə kolleksiyalara görə sənədlərin bölünməsi hesabına əldə olunur. “Baş-asılı” rejimdə asinxron replikasiyaları dəstəkləyir: yazı əməliyyatları yalnız baş qovşaqlarda emal edilir, oxuma əməliyyatları həm baş qovşaqlarda, həm də asılı qovşaqlardan birində həyata keçirilə bilər. Müştəri müxtəlif rejimlərdə: bloklamaqla (bəlli sayda qovşaqların təsdiqlənməsini gözləməklə) və ya bloklamadan (təsdiqlənməni gözləmədən) işləyə bilər. MongoDB, həmçinin GridFS funksionalı sayəsində paylanmış fayl sistemi kimi də istifadə edilə bilər. MongoDB ilə iş üçün müştəri kitabxanalar çoxlu sayda proqramlaşdırma dilləri üçün əlverişlidir, bundan başqa REST (*ing. Representational State Transfer*) interfeysi dəstəklənir. Bu sistem SourceForge, Foursquare, The Guardian, Forbes, The New York Times və s. daxil olmaqla çoxlu sayda böyük şirkət və layihələrdə istifadə olunur.

BigTable – 2004-cü ildə Google Earth, Google Maps, Google Analytic və s. kimi Google xidmətlərin dəstəklənməsi üçün işlənilib hazırlanmışdır. BigTable Google fayl sistemə (*ing. Google File System, GFS*) əsaslanır. GFS verilənlərin və jurnalın saxlanması üçün istifadə olunur. BigTable yüzlərlə və minlərlə qovşaqlarda etibarlı miqyaslanıla bilən və petabayt həcmində verilənlərlə işləmək üçün nəzərdə tutulmuşdur. BigTable-də cədvəl və sütunlar ailəsinin

yaradılması və silinməsi, meta-verilənlərin dəyişdirilməsi (məsələn, giriş hüququ), qiymətlərin yazılması və silinməsi, müəyyən sətirlərin oxunması, verilənlərin alt çoxluqlarının gözdən keçirilməsi (məsələn, müəyyən ailədən sütunlar) üçün əməliyyatlar nəzərdə tutulmuşdur. Müştəri kitabxanaları seqmentlərin yerləşməsi haqqında meta-verilənləri keş edir və vaxtın çox hissəsi bilavasitə verilənləri saxlayan qovşaqlara müraciət edirlər.

Mövcud yanaşmalar

Hal-hazırda BV saxlanması sahəsindəki problemlər (*ing. Big Data Storage*) nüfuzlu beynəlxalq təşkilatlar (*ing. IEEE, Storage Networking Industry Association, SNIA v.s.*) tərəfindən keçirilən konfransların, simpoziumların, seminarların və forumların əsas məzahirə mövzusu sırasında [21, 22]. BV-nin saxlanması problemlərinin həllinə həsr olunmuş çox sayda yanaşmalar irəli sürülmüşdür [23–36]. [23]-də paylanmış saxlama sistemlərində replikasiyanın alternativini kimi təmizləmə kodlarından (*ing. erasure coded*) verilənlərin itki və izafiliyinin qorunması üçün yanaşma təklif edilmişdir. Belə ki, bu kodlar cüzi xərclərlə yüksək elastikliyi təmin edir. Lakin bu təmizləmə kodunda verilənlərin effektiv bərpası və yenilənməsi imkanlarının olmaması onun ən əsas çatışmayan cəhətidir. Reed-Solomon kodlaşmasından istifadə etməklə verilənlərin bərpasına çəkilən xərcləri kifayət qədər azaltmaq olar.

Bulud saxlamalara tələbat get-gedə sürətlə artır. Belə ki, böyük sayda müəssisələrin və müştərilərin verilənləri buludda saxlanılır və emal edilir. Bu səbəbdən paylanmış obyektlərin verilənlərinin saxlanması (məsələn, *ing. QFS (Quantcast File System), OpenStack Swift, HDFS (ing. Hadoop Distributed File System)*) istənilən bulud infrastrukturunun çox mühüm tərkib hissəsini təşkil edir. Bu sistemlər çoxlu sayda serverlər arasında informasiyaları paylaşmaqla yaxşı etibarlılıq təklif edir. Ancaq bu sistemlərin çatışmayan cəhəti onların vahid verilənlərin emalı mərkəzi daxilində yerləşdirilmək üçün nəzərdə tutulmasıdır. Açıq kodlu əksər saxlama sistemləri üçün geo-replikasiya (verilənlərin emalı mərkəzində lazımsız informasiyaların paylaşılması) qeyri-əsinxron olaraq şəffaf həyata keçirilir. Geo-replikasiyanın çox yüksək dərəcədə etibarlılığı təmin etməsini nəzərə alaraq [24]-də bilavasitə qlobal şəbəkə şəraitində yerləşdirilmiş bu saxlama sistemlərinin həyata keçirilməsinə baxılmışdır. Bu məqsədlə Quantcast-QFS, Swift və Tahoe-LAFS kimi üç populyar paylanmış saxlama obyektini götürülmüş və üç böyük verilənlərin emalı mərkəzində testdən keçirilmiş və qiymətləndirilmişdir.

Sənaye və elmi dairələr üçün [25]-də çoxlu passiv verilənlər saxlayan BV sistemində kodlaşdırma texnologiyası əsasında həll təklif edilmişdir. Kodlaşdırma texnologiyalarının tədqiqi sistemin etibarlılığını qoruyur, həm də təhlükəsizliyi və verilənlərin saxlama sistemlərindən istifadəni yaxşılaşdırır, boş yerlərə qənaət edir. Tədqiqatçılar yükləmə tarazlığını saxlamaq üçün saxlama qovşağını bir neçə virtual qovşağa bölmüşdür. Müxtəlif “server codec” üçün müxtəlif virtual qovşaqların saxlama qruplarını yaratmaqla sistemin əlyətərliyi təmin edilmişdir. Qovşağın və verilənlər blokunun şifrinin açılmasında paralel hesablama istifadə etməklə sistemin bərpası üçün vaxt azaldılmışdır. Bundan başqa, çoxlu sayda istifadəçilər müxtəlif kodlaşdırma parametrlərindən istifadə etməklə BV-nin saxlama sisteminin etibarlılığını yüksəldir. Müəlliflər müxtəlif sayda verilənlər bloku və kalibrəmə bloku yaratmaqla külli miqdarda eksperimentlərdə istifadə əmsalını yaxşılaşdırmışdır. İstifadə edilən ICRS (*ing. International Standard Recording Code*) şifrləmə ilə kodlaşdırmanın effektivliyi CRS-dən (*ing. Cyclic redundancy check*) 34,2% və RS (*ing. Reed-Solomon*) şifrlənmədən isə 56,5% yüksəkdir. ICRS-lə deşifrlənmənin sürəti CRS-dən 18,1%, RS-dən 31,1 % yüksəkdir. [26]-də təqdim edilmiş klaster saxlama sistemi adlanan intellektual şəbəkə saxlama diskləri BV-nin saxlanması, əlyətərliyi, mühafizəsi və idarə edilməsi kimi problemləri həll edir. Yaddaş tıxacının (*ing. bottleneck*) aradan qaldırılması üçün ucuz (*ing. cheap*) və dəqiq (*ing. accurate*) kanal işlənmişdir.

Hal-hazırda verilənlərin buludda təhlükəsiz və effektiv saxlanması bulud hesablamaları sahəsində ən böyük çağırışlardan biridir. Buludda saxlanan verilənlərin etibarlı şəkildə qorunmasına zəmanət yoxdur. [27]-də bulud hesablama sistemində təhlükəsizliyi təmin edən

etibarlı hesablama mühiti yaratmaq üçün metod təklif edilmişdir. Bu metod istifadəçilərə verilənlərini buludda təhlükəsiz və effektiv saxlamağa imkan verir. BV-nin idarə edilməsi və təhlükəsizlik problemləri verilənlərin bulud saxlamaya yüklənməsi zamanı şifrlənmə və sıxma (*ing. compression*) metodu ilə həll edilir.

BV-nin həcmi və müxtəlifliyi kimi özəllikləri saxlanma və analitikada problemlər yaradır. [28]-də bulud mühitində BV-nin saxlanması üçün universal arxitektura təqdim edilmişdir. Burada klaster analizindən istifadə edilməklə müxtəlif qovşaqlar arasındakı kommunikasiya xətlərinin qiymətinə görə bulud qovşaqları bir neçə klasterə bölünmüşdür. Bu məqsədlə istifadəçilər üçün universal saxlama və sorğu interfeysi təmin edən, ən böyük hesablama gücünə malik klaster seçilmişdir. Digər klasterlərin hər biri relyasion verilənləri, açar-qiymət (*ing. key-value*) verilənləri (DynamoDB, Redis, Riak), sənəd (*ing. document*) verilənləri (MongoDB, CouchDB) və s. kimi konkret modellə verilənlərin saxlanmasına cavab verir. BV-nin saxlanması ilə əlaqəli problemlər elmi sahələr kimi sənayenin də diqqət mərkəzinə çevrilmişdir. Mövcud işlərin analizi əsasında BV-nin saxlanmasının qarşılaşdığı problemlərin həllində paylanmış fayl sistemləri, o cümlədən dörd əsas: kiçik faylların saxlanması, yüklənmənin balanslaşdırılması, sürət çıxarmanın daimiliyi (*ing. consistency*) və təkrarlanmaların aradan qaldırılması (*ing. deduplication*) üsulları ilə tədqiqat inkişaf etdirilmişdir [29].

Yeni hesablama paradigması kimi formalaşan bulud hesablamaları, xüsusilə BV-nin emalı üçün resurslara çevik və geniş miqyasda girişi – əlyətərliyi təmin edir. Burada həll edilməmiş əsas problem verilənlərin buludda səmərəli emalı üçün uzun müddət ərzində müxtəlif coğrafi məkanlarda yerləşmiş verilənlərin effektiv miqrasiyasıdır. Sərt disklə daşınma yanaşması qıvraq və etibarlı deyildir. Pərakəndə verilənlərin emaletmə mərkəzini əhatə edən buluda yönəldilmiş bu iş verilənlərin miqrasiyasına çəkilən xərclərin minimallaşdırılması problemini modelləşdirir və iki onlayn alqoritm təklif edir: verilənlərin emaletmə mərkəzində istənilən vaxt verilənlərin emal və aqreqasiyası üçün seçilməsinin minimallaşdırılması; verilənlərin ötürülməsində marşrutun seçilməsi [30]. Verilənlərin tamlığı məzmunla əlaqəli istənilən verilən və hesablamaları praktik olaraq təmin etmək üçün mühüm amildir. O, yalnız xidmətin keyfiyyətindən biri kimi xidmət etmir, həm də verilənlərin təhlükəsizliyinin və məxfiliyinin mühüm hissəsidir. Bulud hesablamaların yayılması və BV-nin analizinə artan tələbat verilənlərin, xüsusilə kənar (*ing. outsourced*) verilənlərin tamlığının yoxlanmasını daha da zəruri edir. Buna görə də verilənlərin tamlığının yoxlanması ilə əlaqədar tədqiqat mövzuları tədqiqatçıların böyük marağına səbəb olmuşdur. Bütün göstəricilər arasında iki ən maraqlı meyar effektivlik və təhlükəsizlikdir. [31]-də autentikliyin təhlili əsasında verilənlərin tamlığının effektiv yoxlanması analiz edilir və bu problemin əsas aspektləri üzrə icmal verilir, tədqiqat metodologiyaları, həmçinin bir neçə xarakterik yanaşmaların əsas nailiyyətləri ümumiləşdirilir.

Yer kürəsi müşahidələri repozitorisinin həcmi hər gün bir neçə terabayt artır. Qiymətli informasiyaları və bilikləri təqdim etmək üçün bu böyük həcmli verilənlərin emalı və saxlanması üçün effektiv metodun yaradılmasına ehtiyac vardır. [32]-də bulud virtuallaşması əsasında arxitektura həll təqdim edilir və Yer kürəsinin bu təsvirlərinin böyük həcmli verilənlərindən biliklərin əldə edilməsi üçün çevik (*ing. flexible*) və adaptiv (*ing. adaptive*) metod təklif edilir.

[33]-də BV-nin saxlanması və idarə edilməsi problemləri analiz edilir və həllər təklif edilir.

BV ilə əlaqəli böyük potensiala görə bir çox sektorlar BV-yə maraq göstərirlər. BV verilənlərlə bağlı problemləri həll etmək üçün böyük hesablama gücü və paylanmış saxlanma tələb etdiyinə görə bulud infrastrukturunu BV üçün tələb olunan çevik hesablama gücünü və saxlamaları təmin edir. BV-nin həcmi yüksək sürətlə artıqca, bulud verilənlər mərkəzlərinin həcmi də genişlənir və onlar BV üçün bulud infrastrukturunun fasiləsiz olaraq təmin edilməsinə görə böyük miqdarda enerji sərf edirlər. Məqsəd BV üçün virtuallaşmanın köməyiylə bulud verilənlər mərkəzindən səmərəli istifadə etməklə enerjiyə qənaət etməkdir [34].

Elmi, peşəkar, sosial və s. kimi sahələrdə emal edilən verilənlərin həcmi sürətlə artır, onların idarə edilməsi və saxlanması getdikcə çətinləşir. Yüksək miqyaslı infrastrukturlarının meydana çıxması, saxlanmanın idarəetmə texnologiyalarının inkişafına səbəb olmuşdur. Lakin, bununla bərabər

verilənlərin uyğunluğu və əldə olunması, sahələrin miqyaslanması və ya verilənlərə rəqabətli giriş kimi çoxsaylı problemlər ortaya çıxmışdır. [35]-də Google Fayl Sistemi (*ing. Google File System, GFS*) və IBM Ümumi Paralel Fayl Sistemi (*ing. IBM General Parallel File System, GPFS*), yaxud Hadoop Paylanmış Fayl Sistemi (*ing. Hadoop Distributed File System, HDFS*), Blobseer və Andrew Fayl Sistemi (*ing. Blobseer and Andrew File System, AFS*) kimi açıq kodlu sistemlərdə mövcud texnologiyaların bir sıra başlıca xüsusiyyətləri nəzərdən keçirilmiş, müzakirə və müqayisə edilmişdir. Məqalədə, həmçinin hər bir hal üzrə ən yaxşı xidmətlərin təqdim edilməsi üçün BV-nin saxlanması idarə edilməsi və onların həlli üsullarının bir sıra əsas problemləri müzakirə edilmişdir.

[36]-da BV-nin saxlama problemləri üçün ən yaxşı həlli təmin edən daxili server əsasında böyük tutumlu SSD və sərt disklərin konfigurasiyası ilə paylanmış saxlama üçün heterogen saxlama arxitekturu təklif edilmişdir. Bu, dinamik verilənlərin qiymətini hesablama və verilənlərin müxtəlif qiymətinə uyğun olaraq saxlamanın müvafiq strategiyasını seçmə və sistemin ümumi məhsuldarlığını yaxşılaşdırma bilər. Sonda məqalədə Hadoop paylanmış fayl sistemi əsasında müxtəlif strategiyaların səmərəlilik göstəriciləri və dəyərləri müqayisə edilmiş və analiz edilmişdir.

Nəticə

Verilənlərin həcmi çox böyük sürətlə artması onların saxlanması sahəsində də bir sıra problemlər yaratmışdır. Bu isə mövcud texnologiyaların təkmilləşdirilməsinə və ya yeni arxitekturaların yaradılmasına səbəb olmuşdur. Aparılmış tədqiqatlardan da görünür ki, ümumiyyətlə, verilənlərin ənənəvi saxlama sistemləri (HDD, SSD və s.) hələ geniş istifadə olunmaqdadır. Bu sahənin istehsalçıları və mütəxəssisləri mövcud saxlama sistemlərinin həcm və sürət kimi xarakteristikalarının yaxşılaşdırılması istiqamətində işlər aparmışlar. Lakin sərt disklə daşınma elastikliyə və etibarlılığa zəmanət vermir. Bulud infrastrukturunu BV üçün tələb olunan çevik hesablama gücünü və saxlamayı təmin edir. Bulud saxlama xidməti tələb üzrə resursların ödəniş əsasında əldə edilməsinə xidmət etdiyindən bu model saxlanan verilənlərin bulud hesablamaları xidməti təqdim edən şirkətlərdən asılılığının yaranmasına, yeni bulud inhisarçıların meydana gəlməsinə səbəb olur. Verilənlərin buludda səmərəli emalı üçün uzun müddət ərzində müxtəlif coğrafi məkanlarda yerləşmiş verilənlərin effektiv miqyası da bulud saxlamanın həll edilməmiş əsas problemlərindəndir. BV-nin saxlanması əsas məqsəd dinamik artan verilənlər üçün yüksək miqyaslı, etibarlı və effektiv saxlamayı təmin edən vasitələrdən istifadə etməkdir. Bu da təkmilləşdirilmiş səmərəli giriş və imtinaya davamlı NoSQL saxlama sistemlərinin innovativ inkişafını gerçəkləşdirmişdir. Nəticə olaraq qeyd etmək lazımdır ki, BV-nin saxlanma problemi həmişə olduğu kimi, yenə də alimlərin maraq göstərdiyi mövzu olaraq qalmaqdadır. Belə ki, rəqəmsal verilənlərin həcmi hər gün bir neçə terabayt artır. Dəyərli informasiyaların əldə edilməsi məqsədi ilə belə böyük həcmli verilənlərin emalı və saxlanması üçün effektiv metodların yaradılmasına, yeni modellərin, optimallaşma üsullarının, təhlükəsizlik modellərinin işlənməsinə və təkmilləşdirilməsinə ehtiyac vardır.

Ədəbiyyat

1. Əliquliyev R.M., Hacırahimova M.Ş. “Big Data” fenomeni: problemlər və imkanlar // İnformasiya Texnologiyaları Problemləri, 2014, №2, s.3–16.
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. Черняк Л. Большие Данные — новая теория и практика, М.: Открытые системы, 2011, №10, с.18–25.
4. Leavitt N. Storage Challenge: Where Will All That Big Data Go? // Computer, 2013, vol.48, no.9, pp.22–25.
5. Тютляева Е.О., Тютляев М.М. Системы хранения данных лидирующих суперкомпьютеров / Третий национальный суперкомпьютерный форум (НСКФ-2014), 25-27 ноября 2014 г. <http://2014.nscf.ru/video/nscf-plenar>.
6. Fətəliyev T.X., İsmayılova M.İ. SAN texnologiyasının korporativ şəbəkədə tətbiqi haqqında /

- PCI 2006 "Kibernetika və informatika problemləri" Beynəlxalq konfransı, cild II, Bakı, 2006, s.87–89.
7. Fətəliyev T.X, İsmayılova M.İ. Korporativ verilənlərin saxlanılmasının bəzi məsələləri / 3-rd International Conference on Information Technologies and Telecommunication ITTC-2007, Ganja, 2007, pp.258–260.
 8. Agrawal D., Das S., Amr El A. Big Data and Cloud Computing: Current State and Future Opportunities / Proceedings of the International Conference on Extending Database Technology (EDBT), 2011, pp.530–533.
 9. Foster Y., Kesselman C., Tuecke S. The Anatomy of the Grid: Enabling Scalable Virtual Organizations // Intern. J. of High Performance Computing Applications, 2001, vol.15, no.3, pp.200–222, www.globus.org.
 10. Big Data Processing using Apache Hadoop in Cloud System, www.researchgate.net.
 11. Коваленко О.С. Обзор состояний, проблем и перспектив хранения и анализа данных в облаке // Информатика, вычислительная техника и инженерное образование, 2011, т. 7, №5, с.39–49.
 12. Elmustafa A. Sayed A, Saeed R..Ali. A Survey of Big Data Cloud Computing Security // International Journal of Computer Science and Software Engineering (IJCSSE), 2014, vol.3, no.1, pp.78–85.
 13. Implementing, Serving, and Using Cloud Storage. www.snia.org.
 14. Клеменков П.А., Кузнецов С.Д. Большие данные: современные подходы к хранению и обработке // Труды Института системного программирования РАН, 2012, т. 23, с.143–158.
 15. Pokorny. J. NoSQL databases: a step to database scalability in web environment / Proceedings of the 13th International Conference on Information Integration and Webbased Applications and Services, 2011, pp.278–283.
 16. Stonebraker M. “Errors in Database Systems, Eventual Consistency, and the CAP Theorem” blog, Communications of the ACM, April, 2010. <http://cacm.acm.org>.
 17. Siddiqa A, Karim A. Gani A. Big data storage technologies: a survey // Frontiers of Information Technology & Electronic Engineering, 2017, vol.18, no.8, pp.1040–1070.
 18. Cattel R. Scalable SQL and NoSQL data stores. ACM SIGMOD Record, 2010, vol.39, no.4, pp.12–27.
 19. Strauch C. NoSQL Databases. <http://christof-strauch.de/nosql dbs.pdf>.
 20. Mattu B.M. NOSQL Databases: The Need of Enterprises // International Journal of Allied Practice, Research and Review, 2017, vol.4, no.5, pp.15–19.
 21. www.ieeebigdata.org.
 22. www.snia.org.
 23. Esmaili K.S., Chiniah A., Datta A. Efficient updates in cross-object erasure-coded storage systems / Proceedings of the IEEE International Conference on Big Data, 2013, pp.28–35.
 24. Yih-Farn C., Daniels S., Hadjieleftheriou M., Liu P., Tian C., Vaishampayan V. Distributed storage evaluation on a three-wide inter-data center deployment / Proceedings of the IEEE International Conference on Big Data, 2013, pp.17–22.
 25. Yin C., Wang J., Xie C., Wan J., Long C., Bi W. Robot: An efficient model for big data storage systems based on erasure coding / Proceedings of the IEEE International Conference on Big Data, 2013, pp.163–168.
 26. Haixia L., Chuiwei L., Sheng S. Intelligent network storage on the big data / Proceedings of the International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI), 2014, pp.1–6.
 27. Kumar A., Hoonjae L., Singh R.P. Efficient and secure Cloud storage for handling big data / Proceedings of the 6th International Conference on Information Science and Service Science and Data Mining (ISSDM), New Trends, 2012, pp.162–166.

28. Zhang Q., Chen Z., Lv A., Zhao L., Liu F., Zou J. A Universal Storage Architecture for Big Data in Cloud / Proceedings of the IEEE International Conference on Internet of Things (iThings/CPSCoM), 2013, pp.476–480.
29. Zhang X., Xu F. Survey of Research on Big Data / Proceedings of the 12th International Symposium on Storage Distributed Computing and Applications to Business, Engineering & Science (DCABES), 2013, pp.76–80.
30. Lofstead J., Jimenez I., Maltzahn C., Koziol Q., Bent J., Barton E. Moving big data to the cloud / Proceedings of the IEEE International Conference on Cluster Computing (CLUSTER), 2014, pp.280–281.
31. Liu C., Ranjan R., Zhang X., Yang C., Georgakopoulos D.J. Chen Public Auditing for Big Data Storage in Cloud Computing – A Survey / Proceedings of the IEEE 16th International Conference on Computational Science and Engineering (CSE), 2013, pp.1128–1135.
32. Bica M., Bacu V., Mihon D., Gorgan D. Architectural solution for virtualized processing of big earth data / Proceedings of the IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), 2014, pp.399–404.
33. Li J., Xu Z., Jiang Y., Zhang R. The overview of big data storage and management / Proceedings of the IEEE 13th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), 2014, pp.510–513.
34. Solanki N., Kachhwaha R. Energy efficient strategy for cloud based big data / Proceedings of the 6th International Conference on Reliability, Infocom Technologies and Optimization, 2017, pp.519–522.
35. Elomari A., Maizate A., Hassouni L. Data storage in big data context: A survey / Proceedings of the IEEE Third International Conference on Systems of Collaboration (SysCo), 2017, pp.1–4.
36. Li H., Wen Z., Mo J., Wu J. Distributed heterogeneous storage based on data value / Proceedings of the IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2017, pp.264–271.

УДК 004.62

Гаджирагимова Макруфа Ш.¹, Исмайылова Марзия И.²

Институт Информационных Технологий НАНА, Баку, Азербайджан

¹makrufa@science.az, ²imarziya@gmail.com

Проблемы хранения больших данных и существующие решения

Статья посвящена анализу проблем хранения больших данных и существующих решений. Проблемы хранения и обработки данных все время существовал. Большие объемы данных, генерируемые с помощью информационно-коммуникационных технологий, сегодня вновь внесены в повестку дня проблемы хранения. В статье проверен сравнительный анализ систем хранения данных, а также исследованы современные подходы и проблемы хранения больших данных.

Ключевые слова: большие данные, хранение больших данных, NoSQL база данных, дата-центр, облака вычислений, облака хранения.

Makrufa Sh. Hajirahimova¹, Marziya I. Ismayilova²

Institute of Information Technology of ANAS, Baku, Azerbaijan

¹makrufa@science.az, ²imarziya@gmail.com

Problems of storing big data and existing solutions

The article reviews big data and existing solutions. Data storage and processing issues have always existed. Today, large volumes of data generated through information and communication technologies has raised the problem of once again. This article comparatively analyzes the storage systems and highlights the problems of big data storage and modern approaches to them.

Keywords: big data, big data storage, data centre, NoSQL database, cloud computing, cloud storage.