



**KATTA HAJMLI MA'LUMOTLARNI SAQLASH TIZIMINING
ISHLASH ALGORITMI VA QURULISH ARXITEKTURASINI
TAHLILI**

**АНАЛИЗ АЛГОРИТМА РАБОТЫ И АРХИТЕКТУРЫ
ПОСТРОЕНИЯ БОЛЬШОЙ СИСТЕМЫ ХРАНЕНИЯ ДАННЫХ**

**ANALYSIS OF OPERATION ALGORITHM AND ARCHITECTURE
OF BUILDING A LARGE DATA STORAGE SYSTEM**

Норматова Дилбар Тургуновна

Муҳаммад ал-Хоразмий номидаги ТАТУ,

Телекоммуникация инжиниринги кафедраси катта ўқитувчиси,

Email: normatova_1972@mail.ru, +998901684889

Annotatsiya. Maqolada katta hajmli ma'lumotlarni saqlash tizimlarining xususiyatlari, turlari, ishlash algoritmlari va qurilish arxitekturalarini tahlili, ma'lumotlarni saqlash tizimining tarkibi va funktsional xususiyatlarining muhim tomonlari, tizimlar uchun alohida ishlab chiqishda foydalaniladigan arxitekturaviy yondashuvlar ko'rsatiladi. Shu bilan birga, paralellik, taqsimlash, tartiblash va uzatish kabi katta hajmli ma'lumotlarni saqlash tizimlarining yuqori darajadagi ish rejalarini bajarish uchun qo'llaniladigan xususiyatlari ham keng yoritilgan.

Kalit so'zlar: *NoSQL (Not Only SQL), Distributed File Systems, ClickHouse, Postgre SQL DBMS, DAS (Directed Attach Storage), SAN (Storage Area Network), SAS (Advanced SCSI), Secure Hash Algorithm (SHA).*

Annotation. The article shows the characteristics, types, performance algorithms and construction architectures of large-scale data storage systems, the important aspects of the structure and functional characteristics of the data storage system, and the architectural approaches used in individual development for the systems. At the same time, the features of large-scale data storage systems, such as parallelism, distribution, ordering, and transfer, which are used to execute high-level work plans, are also covered.





Аннотация. В статье показаны характеристики, типы, алгоритмы работы и архитектуры построения крупномасштабных систем хранения данных, важные аспекты структуры и функциональных характеристик системы хранения данных, а также архитектурные подходы, используемые при индивидуальной разработке систем. При этом также рассматриваются особенности крупномасштабных систем хранения данных, такие как распределение, упорядочивание и передача, которые используются для выполнения планов работы высокого уровня.

Kirish. Dunyo bo'ylab ma'lumotlar hajmi tez sur'atlar bilan o'sib bormoqda. Ko'pgina jarayonlarni raqamlashtirish, turli xil ijtimoiy media platformalarining paydo bo'lishi, bloglar, kundalik hayotda ishlatiladigan kichik hajmli raqamli (smart soatlar, GPS qurilmalari va hokazolar) qurilmalarning paydo bo'lishi, qurilmalar va internetdan foydalanishning tez o'sishi katta miqdordagi ma'lumotlarning doimiy ravishda ishlab chiqarilishiga sabab bo'lmoqda. Katta ma'lumotlar tobora deyarli barcha sohalarning ajralmas qismiga aylanib bormoqda. «Big Data» – bu juda tez sur'atlarda o'sib borayotgan katta hajmdagi tizimli va tizimsiz raqamli ma'lumotlar to'plami. «Big Data» biznes jarayonlarini optimallashtirish va avtomatlashtirish, to'plangan ma'lumotlarga asoslangan eng samarali qarorlarni qabul qilinishini ta'minlash maqsadida axborotni saqlash va qayta ishlashning innovatsion usullarini o'z ichiga oladi. Katta ma'lumotlar hajmi o'sishda davom etar ekan, unga sho'ng'ish va amaliy tushunchalarni olish uchun o'qitilgan ma'lumotlar tahlilchilariga ehtiyoj ortib bormoqda. «Katta ma'lumotlar» tahlili moliya, hukumat va sog'liqni saqlash kabi sohalarda o'zgarishlar yaratish uchun ajoyib imkoniyatlarni taqdim etadi, shuningdek, firibgarlikning oldini olish, tabiiy ofat yuz berganda resurslarni taqsimlash yoki sog'liqni saqlashni yaxshilash orqali odamlar hayotini o'zgartirishga yordam beradi.

Tadqiqot ob'ekti va usullari. Katta hajmli ma'lumotlarni saqlash tizimlari ma'lumotlarni yig'ish, saqlash, qayta ishlash va izlashni osonlashtiruvchi vositalardir. Katta hajmli ma'lumotlarning o'sishiga bir qancha omillar turtki bo'ldi. Endi odamlar korxonalar orasida qog'oz yozuvlarini raqamlashtirishning keng tarqalishi tufayli har qachongidan ham ko'proq ma'lumotni saqlaydilar [1].





Tizim turli manbalardan olingan ma'lumotlarni tahlil qilish uchun "Massively Parallel Processing" ma'lumotlar bazalariga tayanadi. Katta hajmli ma'lumotlar ko'pincha tuzilmaga ega emas va turli manbalardan kelib chiqadi, shuning uchun ular relyatsion ma'lumotlar bazasi bilan ishlash uchun yomon mos keladi. Apache Hadoop Distributed File System (HDFS) katta hajmli ma'lumotlar uchun eng keng tarqalgan tahliliy vosita bo'lib, odatda NoSQL ma'lumotlar bazasining ba'zi afzalliklari bilan birlashtiriladi. Hadoop - Java dasturlash tilida yozilgan ochiq kodli dasturiy ta'minot. HDFS ma'lumotlar tahlilini yuzlab yoki hatto minglab server tugunlari bo'ylab ishlashga ta'sir qilmasdan tarqatadi. MapReduce komponenti orqali Hadoop halokatli nosozlikdan himoya qilish uchun qayta ishlashni shu tarzda tarqatadi. Bir nechta tugunlar tarmoqning chekkasida ma'lumotlarni tahlil qilish uchun platforma bo'lib xizmat qiladi. So'rov kelganda, MapReduce to'g'ridan-to'g'ri ma'lumotlar joylashgan saqlash tugunida qayta ishlashni amalga oshiradi. Tahlil tugallangandan so'ng, MapReduce har bir serverdan umumiy natijalarni to'playdi va yagona yaxlit javobni taqdim etadi.

Tadqiqot natijalari va ularning muhokamasi. Katta hajmli ma'lumotlar bazalari bilan ishlash uchun bir nechta usullar mavjud. SQL: SQL (Structured Query Language) katta hajmli ma'lumotlar bazalari bilan ishlash uchun keng qo'llaniladigan usul. SQL, ma'lumotlarni so'rov qilish, qo'shish, o'chirish, yangilash va tahlil qilishga imkon beradi [2].

NoSQL: NoSQL (Not Only SQL) katta hajmli ma'lumotlar bazalari uchun boshqa bir usuldir. NoSQL tizimlar relatsionallikning cheklanganligini olib tashlab, ma'lumotlarni dokumentlar, kalitli juftliklar, grafiklar yoki boshqa formatlarda saqlashga imkon beradi. Ular ma'lumotlarni tashkil etishga moslashtirilgan tashqi interfeyslarni taqdim etishadi.

Ma'lumotni saqlash har qanday biznes yoki tashkilotning asosiy tarkibiy qismidir, lekin u o'ziga xos qiyinchiliklar bilan birga keladi. Masalan katta hajmdagi ma'lumotlar bilan ishlashda miqyoslash imkoniyatlarining yetishmasligi. Bunga qarshi kurashish uchun tashkilotlar kerak bo'lganda o'zlarining mavjud infratuzilmasini osongina kengaytirishlari uchun bulutli hisoblash yechimlariga sarmoya kiritishlari kerak. Bundan tashqari, ushbu yechimlar korxonalariga qimmatli ma'lumotlarini bulutda ishonchli tarzda arzon narxlarda saqlashga imkon beradi.





Bugungi kunda ma'lumotlarni saqlashda eng katta muammolar nima va ularni qanday yengish mumkin? Keling, bir ko'rib chiqaylik.

1. Infratuzilma. Agar siz katta hajmdagi ma'lumotlarni saqlashni rejalashtirmoqchi bo'lsangiz, uni saqlash uchun zarur bo'lgan infratuzilma kerak bo'ladi, bu ko'pincha ofisingiz yoki binongizda katta joy egallaydigan yuqori texnologiyali serverlarga sarmoya kiritishni anglatadi. Eng oson yechimlardan biri bu bulutli hosting va bulutli saqlashdan foydalanish bo'lib, ular boshqa kompaniyaning infratuzilmasidan foydalanib, o'sha joyni va sozlamalarni o'zingizga moslashtirib olishingiz mumkin bo'ladi.

2. Narx. Shaxsiy ma'lumotlar markazini ishga tushirish qimmat operatsiya hisoblanadi. Siz dastlabki sozlash, doimiy texnik xizmat ko'rsatish va uni saqlash uchun mas'ul bo'lgan odamlar bilan bog'liq xarajatlarga pul sarflashingiz kerak bo'ladi. Shunga qaramay, bu yerda eng yaxshi yechim ishni ijaraga olishdir, ehtimol siz oylik to'lovni to'lashingiz kerak bo'ladi, lekin bu uzoq muddatda pulingizni tejaydi.

3. Xavfsizlik. Xavfsizlik - bu yengib o'tish kerak bo'lgan asosiy muammo. Taxminlarga ko'ra, agar sizning ma'lumotlaringiz biror joyda saqlangan bo'lsa, uchinchi tomon ularni olishi mumkin. Ushbu ruxsatsiz kirishning oldini olishga yordam beradigan ko'plab xavfsizlik qatlamlari mavjud, shu jumladan shifrlash va uchinchi tomon provayderlariga ishonish, lekin ular sizni qanchalik yaxshi himoya qilishi mumkinligining chegarasi bor - hatto FBI o'z ma'lumotlarining xavfsizligini ta'minlashda muammolarga duch keladi. Siz eng yaxshi sheriklarni va ishonchli provayderlarni tanlashingiz kerak bo'ladi.

4. Masshtab. Siz hozirgi ehtiyojlaringizga mos keladigan saqlash yechimini topishingiz mumkin, ammo bu ehtiyojlar to'satdan o'zgarsa nima bo'ladi? Sizning ehtiyojlaringizni 5 yil ichida qanday hisoblaysiz? Sizning ma'lumotlarni saqlash yechimingiz masshtablash uchun biroz sig'imga muhtoj. Bu yerda o'zingizga iloji boricha ko'proq imkoniyatlarni taqdim etish foydalidir, chunki kelajakda ehtiyojlaringiz qanday o'zgarishiga ishonchingiz komil bo'lmaydi.

5. UI va foydalanish imkoniyati. Agar kirish qiyin bo'lsa, ma'lumotlaringiz siz uchun unchalik yaxshi bo'lmaydi. Axir, ma'lumotlarni saqlash vaqtinchalik chora bo'lib, keyinchalik ma'lumotlarni tahlil qilish va





undan yaxshi foydalanish mumkin. Shunga ko'ra, sizga intuitiv, foydalanish mumkin bo'lgan foydalanuvchi interfeysi (UI) va istalgan funksionallik uchun oson kirish imkoniyati bo'lgan qandaydir tizim kerak bo'ladi [3].

6. Muvofiqlik. Agar siz ma'lumotlaringiz bilan bir nechta tizim yoki ilovalardan foydalanishni rejalashtirmoqchi bo'lsangiz, ularning mos kelishiga ishonch hosil qilishingiz kerak bo'ladi. Buning uchun siz ochiq API va toza o'tish("clean system") tizimiga ega ma'lumotlarni saqlash tizimini topishingiz kerak bo'ladi.

Bizneslarni, internet xizmatlarini va ijtimoiy media hayotimizni qo'llab-quvvatlovchi ma'lumotlar markazlari ko'p energiya sarflaydi. Ularning soni ortib borishi bilan ularni energiyani tejamkor qilish va atrof-muhitga ta'sirini kamaytirish zarur bo'ladi.

Big Data ishlash texnologiyalari: Katta hajmli ma'lumotlar bazalari uchun Big Data ishlash texnologiyalari keng qo'llaniladi. Ular ma'lumotlarni tezkor va parallel tarzda ishlashga imkon beradi. Hadoop va Spark, misol uchun, ma'lumotlarni bo'shatish, tarqatib bo'lish, tezkor ishlash va tahlil qilishda foydalaniladigan Big Data ishlash platformalari hisoblanadi [4,5].

Ilmiy tadqiqot natijalari va xulosasi. Katta ma'lumotlarning aylanishini har bir bosqichidagi maxfiylik muammolari katta ma'lumotlar ilovalari kontekstida mavjud maxfiylikni saqlaydigan texnologiyalarning afzalliklari va kamchiliklari bilan birga taqdim etiladi. Katta ma'lumotlarda maxfiylikni saqlashning an'anaviy va so'nggi usullari bilan birga, identifikatsiyaga asoslangan anonimlashtirish va differentsial maxfiylik tushunchalari hamda katta ma'lumotlar maxfiyligining turli so'nggi texnikalari o'rtasidagi qiyosiy o'rganish chiqildi.

Yakuniy xulosa. Katta ma'lumotlar tobora deyarli barcha sohalarning ajralmas qismiga aylanib bormoqda. Ushbu maqolada "Big Data" haqida qiziqarli ma'lumotlar bilan tanishib, katta hajmli ma'lumotlarni saqlash tizimlari ma'lumotlarni yig'ish, saqlash, qayta ishlash va izlashni osonlashtiruvchi vositalari, korxonalar orasida qog'oz yozuvlarini raqamlashtirishning keng tarqalishi tufayli har qachongidan ham ko'proq ma'lumotni saqlaydilar. O'z navbatida katta hajmli ma'lumotlarni qayta ishlash uchun bizga katta hajmli ma'lumotlar saqlash tizimlari zarur bo'ladi. Katta hajmli ma'lumotlarni saqlash tizimlari ma'lumotlar sonining katta hajmini, turli manbalardan olingan





ma'lumotlarni yig'ish va yuqori darajada ishlov berish talablari bilan bog'liq bo'lishi o'rganildi. Bunga ko'ra ma'lumotlar bazasini boshqarish tizimi (DBMS) - bu ma'lumotlar bazasini yaratish va ma'lumotlarni boshqarish imkonini beruvchi dasturiy ta'minot va til vositalari to'plami bo'lishi bilan birga, ma'lumotlar bazasi ma'lumotlar bazasini tashkil qilish, boshqarish va boshqarish imkonini beruvchi dasturlar to'plami ekanligi va ko'pgina saytlar ma'lumotlar bazasiz ishlay olmasligi sababli, DBMS deyarli hamma joyda qo'llaniladi.

Ma'lumotlar maxfiyligi bu raqamli dunyoda katta muammodir. U shaxsiy yoki maxfiy ma'lumotlarni kiberhujumlar, buzilishlar va qasddan yo'qotishdan himoya qilishga qaratilgan. Korxonalar ma'lumotlarni himoya qilishni kuchaytirish uchun bulutdagi kirishni boshqarish xizmatlari yordamida, jumladan, maxfiylikka qat'iy rioya qilish orqali qat'iyroq ma'lumotlar maxfiyligi tamoyillariga amal qilishlari kerak. Bir yoki bir nechta ma'lumotlar xavfsizligi texnologiyalarini joriy qilish bilan bir qatorda bir nechta qoidalarga amal qilish yaxshidir.

Foydalanilgan adabiyotlarlar ro'yxati:

1. A. B. Веретенников, BigData: анализ больших данных сегодня. Молодой ученый. 2017. № 32 (166).
2. S. A. Weil, S. A. Brandt, E. L. Miller, D. D. Long, and C. Maltzahn, "\Ceph: A scalable, high-performance distributed file system," in *Proceedings of the 7th symposium on Operating systems design and implementation*. USENIX Association, 2006, pp. 307-320.
3. K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "\The hadoop distributed file system," in *2010 IEEE 26th symposium on mass storage systems and technologies (MSST)*. IEEE, 2010, pp. 1-10.
4. I. Gorton, J.Klein, Distribution, data, deployment: software architecture convergence in big data systems. *IEEE Softw.* 2014;32(3):78–85.
5. P.Gölzer, P.Cato, M.Amberg, Data Processing Requirements of Industry 4.0 - Use Cases for Big Data Applications. In: *Proceedings of the 23th European Conference on Information Systems (ECIS)*, paper 61; 2015.

