

**KATTA MA'LUMOT BAZALARINI PARALLEL ISHLOV BERISH
USULLARINI VA MODELLARINI O'RGANISH**

Kodirov Akbar Shuxratovich

Shahrisabz Davlat Pedagogika institutining AKT o'qituvchisi.

Homidova Mahliyo Husniddin qizi

Sharof Rashidov Nomidagi Samarqand Davlat Universiteti,

1-basqich magistratura talabasi

Annotatsiya: Hozirgi vaqtda axborot tizimlari orqali ma'lumotlarga ishlov berish texnologiyalari sohasiga sezilarli ta'sir ko'rsatadigan hodisalardan biri bu juda katta ma'lumotlar bazasini yaratilishi, ularni saqlash va foydalanishdir. Real vaqtda ma'lumotlar bazasiga ishlov berish bugungi kundagi dolzarb masalalardan biri bo'lib bormoqda. Ko'p yadroli prosessorlar asosida jihozlangan kompyuterlarda relyatsion ma'lumotlar bazasi bilan integratsiyalashuvga imkon beruvchi klasterli hisoblash tizimlariga yo'naltirilgan ma'lumotlarning ustunli tasviridan foydalangan holda katta ma'lumotlar bazalarini parallel ishlov berishning samarali usullarini tadqiq va etish va foydalanish bo'yicha tavsiyalar berishdan iborat.

Kalit so'zlar: Ma'lumotlar bazasi, katta hajmdagi ma'lumotlar, parallel hisoblash kompleksi, dasturiy ta'minot, ma'lumotlar bazasini qayta ishlash, Big Data, faylning aktivlik koeffisientni, RAM, parallellashtirish.

Bugungi kunga qadar inson faoliyatining turli sohalarida, tarmoqlarda katta hajmdagi ma'lumotlarni saqlash, qayta ishlash va uzatish kompyuterlar va kompyuter tarmoqlari orqali qo'llashning muhim yo'nalishlaridan: iqtisodiyot, bank, savdo, transport, tibbiyot, fan va hokazolardan biridir. Mavjud zamonaviy axborot tizimlari orqali juda katta hajmdagi saqlanadigan va qayta ishlanadigan ma'lumotlar, turi va murakkab tashkil etilishidan qat'iy nazar ko'plab foydalanuvchilarning turli talablarini qondirish zarurati axborot tizimlaridan keng foydalanishiga olib kelmoqda.

Axborot tizimi – bu ma'lumotlarni yig'ishni, qayta ishlashni va manipulyatsiyasini avtomatlashtirilgan holda amalga oshiradigan va ma'lumotlarni qayta ishlashning texnik vositalarini, dasturiy ta'minot va xizmat ko'rsatuvchi xodimlarni o'z ichiga olgan tizimdir. Har qanday axborot tizimining maqsadi dunyodagi real obyektlar haqidagi ma'lumotlarni qayta ishlash hisoblanadi. Axborot tizimining asosini ma'lumotlar bazasi tashkil etadi. Ma'lumotlar bazasi keng ma'noda – bu har qanday predmet sohasidagi dunyoning aniq obyektlar haqidagi ma'lumotlar to'plamidir. Ma'lumotlar bazasini yaratish orqali foydalanuvchi turli xil xususiyatlarga ko'ra ma'lumotni tartibga solishga intiladi va tezlik bilan kerakli xususiyatlarning kombinatsiyasi tanlovini amalga oshiradi. O'z o'zidan ma'lumki ma'lumotlar bazasini

kombinatsiya qilish va ularni boshqarishda yetarli darajada texnik vositalar mavjud bo'lishi, bazani ko'yiigan sari ularni saqlash kerak bo'ladi.

Ma'lumotlar bazasi bilan ishlashda kompyuter texnologiyalarining hozirgi holati va dasturlash tizimlarining imkoniyatlari turli ko'rinishlarda, darajalarda bo'lishi va turli usullardan foydalangan holda baza ma'lumotlarini optimallashtirishni amalga oshirish imkonini beradi.

Katta hajimdagi ma'lumotlarni qayta ishlash jarayonini ishlab chiqish texnologiyasi quyidagi bosqichlardan iborat:

- ma'lumotlar modelini tanlash va ma'lumotlar bazasi sxemasini qurish;
- so'rovlarni ishlab chiqish va ularni algebraik ifodalar ko'rinishida taqdim etish;
- parallel hisoblash kompleksi arxitekturasini va shu arxitekturaga mos keladigan hisoblash modelini tanlash;

–dastlabki ma'lumotlar modeli va hisoblash modelini bog'lovchi oraliq ma'lumotlar modelini tanlash;

–so'rovlar rejasini bir vaqtda parallellashtirish va optimallashtirish bilan so'rov ifodalarini ma'lumotlar modeli tilidan oraliq model tiliga tarjima qilish;

–so'rovni parallel amalga oshirish talablariga muvofiq ma'lumotlarni tashqi xotiraga joylashtirish;

–so'rovga mos keladigan meta ma'lumotlar asosida so'rovni tashkil etuvchi operatsiyalarni bajarish tartiblarini sozlash;

–apparatni sozlash va so'rovni bajarish.

Hozirgi vaqtda an'anaviy dasturiy ta'minot va apparat vositalaridan foydalangan holda katta hajimdagi ma'lumotlarni qayta ishlash mutlaqo mantiqsiz va ko'pincha bunday imkonsizdir, chunki mavjud ma'lumotlarning katta miqdori bunga imkon bermaydi. Biroq, bunday operatsiyalarni amalga oshirishning bir qancha samarali usullari hali ham mavjud. Garchi buning ham o'z qiyinchiliklari bo'lsada.[2]

Texnologik va iqtisodiy nuqtai nazardan olib qaraganda har bir aniq holat uchun bazada mavjud ma'lumotlarni qayta ishlashning eng mos usulini tanlash kerak, shundan keyingina biz natijalarga erisha olamiz. An'anaga ko'ra, katta hajimdagi ma'lumotlar bazasini qayta ishlash parallel hisoblash bilan bog'liq va buni quyidagicha tushuntirishimiz mumkin: yuqori hajimdagi parallel ishlov berish - bu katta hajimdagi ma'lumotlarni ko'p sonli protsessorlarning parallel qayta ishlash usuli hisoblanadi.

Bugunga kelib ma'lumotlarni ommaviy qayta ishlash Big Data deb nomlangan yo'nalish bilan bog'liq. Katta hajimdagi ma'lumotlar umumiy atama bo'lib, muammolarni hal qilish jarayonida yangi yaratilgan juda katta va doimiy ravishda o'sib borayotgan hajimdagi tuzilgan, tuzilmagan va yarim tizimli ma'lumotlarni anglatadi. Bunday ma'lumotlarni an'anaviy (masalan, relyatsion) ma'lumotlar bazasiga yuklash va keyinchalik qayta ishlash juda ko'p hisoblash resurslarini talab qiladi.[5]

Ma'lumotlar bazasida saqlanadigan ko'pgina yozuvlarning qayta ishlashdagi

ishtirok etishini fayl faolligi koeffitsienti deb ataladigan qiymat yordamida aniqlash mumkin. L - bu fayldagi yozuvlarning umumiy soni bo'lsin, L_e - bu faylda operatsiya jarayoni davomida qayta ishlanadigan bir xil faylning yozuvlari soni deb oladigan bo'lsak, K_a - faylning aktivlik koeffitsientini quyidagicha hisoblash mumkin.

$$K_a = \frac{L_e}{L}$$

Agar faylning aktivlik koeffitsientni yuqori bo'lsa, ya'ni birga yaqin bo'lsa, faylga ketma-ket kirish samaraliroq bo'ladi, aks holda u nolga yaqin bo'lganda, tasodifiy kirishdan foydalanish tavsiya etiladi. Shunday qilib, katta hajimdagi ma'lumotlarni qayta ishlashning asosiy xususiyati shundaki, barcha qayta ishlangan fayllarning faollik koeffitsientlari birga yaqin bo'lishi samarali natija ko'rstkichini beradi.

Ma'lumotlar bazasidagi fayllarni operatsiyalarni bajarish jarayoni ikki yoki undan ortiq fayllarni birgalikda qayta ishlash operatsiyalarini parallellashtirish imkoniyatini ta'minlashi kerak. Algebraik nazar bilan qaraydigan bo'lsak nazariy-ko'payish usulidagi operatsiyalarning rasmiy ta'rifi ma'lumotlar oqimi tizimlariga, ko'p o'lchovli matritsali modelga - vektor (matritsa) protsessorlariga asoslangan tizimlarga xos bo'lgan parallellashtirish usullaridan foydalanishga imkon beradi. Bundan tashqari, model RAM va tashqi xotira vositalari (an'anaviy tashqi xotira yoki ma'lumotlarni saqlash) o'rtasidagi almashinuvni parallellashtirish qobiliyatini ta'minlashi kerak.

Katta ma'lumotlar hajmi yuzlab terabayt va petabaytdan ortiq bo'lishi mumkin bo'lgan ma'lumotlarni o'z ichiga oladi. Bundan tashqari, bunday ma'lumotlar muntazam ravishda yangilanadi. Masalan, aloqa markazlaridan, ijtimoiy tarmoqlardan, birja savdolari ma'lumotlaridan va hokazolardan keladigan ma'lumotlarni o'z ichiga oladi. Katta ma'lumotlar bazasiga parallel ishlov berish usullaridan foydalanish orqali klasterga yo'naltirilgan hisoblash tizimlari bilan jihozlangan ko'p yadroli tezlatgichlar yordamida relyatsion MBBT bilan integratsiya qilinadi va samaradorlik ortadi.

Foydalanilgan adabiyotlar:

1. А. А. Прихожий. Распределенная и параллельная обработка данных. Учебно-методическое пособие Минск БНТУ, 2016
2. Методы параллельной обработки сверхбольших баз данных с использованием распределенных колоночных индексов. Журнал Программирование, 2017. №3, 3-21.
3. <https://gb.ru/blog/obrabotka-bolshikh-dannykh/>, 06.12.2021
4. Захаров В.Н., Мунерман В.И. Модели и методы параллельной обработки структурированных больших данных.
5. Ступников С.А., Скворцов Н.А., Будзко В.И., Захаров В.Н., Калиниченко Л.А. Методы унификации нетрадиционных моделей данных. – Системы высокой доступности, № 1, 2014, т. 10, с. 18-40.