

8-§. Katta hajmdagi ma'lumotlar bilan ishlash texnologiyalari.

Reja:

- 1. Katta ma'lumotlarni qayta ishlash texnologiyalari (Big Data)*
- 2. Katta ma'lumotlarni qayta ishlashda taqsimlangan hisoblash tizimlari*

Katta ma'lumotlarni o'rganish va tahlil qilish rivojlanish va kengayishda davom etadigan murakkab va juda muhim yo'nalishdir. Har yili inson tobora ko'proq ma'lumot ishlab chiqaradi va uning asosiy qismi tartibsiz shaklda bo'ladi. Shu turdagi ma'lumotlarni tahlil qilishni o'rganish, alohida ma'lumotlar to'plamlari orasidagi aloqalarni aniqlash davrimizning eng muhim vazifasidir. IBM korporatsiyasi ushbu vazifaning muhimligini tushunib, katta ma'lumotlar bilan ishlash vositalarini ishlab chiqishda BIG DATA sohasida katta ilmiy ish olib bormoqda.

Katta ma'lumotlar bilan ishlash deyarli barcha sohalarda zarur: fan, tibbiyot, biznes. Big Data ni qayta ishlash ayniqsa biznes yechimlari uchun foydalidir. Ushbu sohada saralanmagan ma'

"BIG DATA" iborasi 2008-yilda Klifford Linchning yengil qo'li bilan paydo bo'ldi. Nature jurnalining maxsus sonida mutaxassis axborot oqimlarining portlovchi o'sishini – BIG DATA deb atadi. Unda u kuniga 150 Gb dan ortiq bo'lgan har qanday ma'lumotlarning massivlarini kiritdi. 2011-yilgacha katta ma'lumot texnologiyalari faqat ilmiy tahlil sifatida ko'rib chiqilgan va amaliy yechim topmagan. Biroq, ma'lumotlar miqdori keskin o'sib bordi va juda ko'p miqdordagi tartiblanmagan ma'lumotlar muammosi 2012-yil boshida dolzarb bo'lib qoldi.

Katta ma'lumotlar texnologiyasining ishlash tamoyili foydalanuvchini har qanday ob'ekt yoki hodisa to'g'risida maksimal darajada xabardor qilishga asoslanadi. Ma'lumotlar bilan tanishishning maqsadi to'g'ri qaror qabul qilish uchun ijobiy va salbiy tomonlarni ko'rib chiqishga yordam berishdir. Aqlli

mashinalarda bir qator ma'lumotlar asosida kelajak modeli quriladi, shuning uchun turli xil variantlar simulyatsiya qilinadi va natijalar kuzatiladi.

Zamonaviy analitik agentliklar g'oyani, taxminni yoki muammoni hal etishda millionlab shunga o'xshash simulyatsiyalarni qo'llaydilar. Jarayon avtomatlashtirilgan.

Katta ma'lumot manbalariga quyidagilar kiradi:

- Internet bloglar, ijtimoiy tarmoqlar, saytlar, OAV va turli forumlar;
- Arxiv, tranzaksiyalar, ma'lumotlar bazasi;
- o'qish moslamalari - meteorologik qurilmalar, uyali aloqa dachchiklari va boshqalar.

Hozirgi vaqtda Big Data axborot texnologiyalarining eng tez rivojlanayotgan sohalaridan biri hisoblanadi, statistika ma'lumotlariga ko'ra, olingan va saqlanadigan ma'lumotlarning umumiy miqdori har 1 yildan oshiqroq vaqtda ikki baravar ko'paymoqda. Big Data biznesning ko'plab sohalarida keng tarqalgan va kompaniyalar rivojlanishida muhim rol o'ynaydigan texnologiya sohasi bo'lib, 2018-yilda Big Data va biznesni tahlil qilish global bozori 169 milliard AQSh dollarini tashkil etgani va 2022-yilga kelib 274 milliard AQSh dollarigacha o'sishi kuzatildi. Bundan tashqari, PwC hisobotida 2020-yil oxiriga kelib faqat AQShda Data Science va Analyticsda 2,7 millionga yaqin ish bajarildi. Global Hadoop hisobotida katta analitik bozor hajmi 2020-yilda 12,8 milliard AQSh dollaridan 2025 yilgacha 23,5 milliard AQSh dollarigacha o'sadi, prognoz davrida yillik o'sish sur'ati (CAGR) 13,0 foizni tashkil etadi, deyiladi. Katta hajmdagi Hadoop katta analitik bozori ushbu vaqt oralig'ida (2018-2023) 36,37% CAGR darajasida rivojlanishi kutilmoqda. Bozor, ehtimol 2023-yil oxiriga qadar 52 milliard AQSh dollaridan ko'proq qiymatda qolishi mumkin. Big Data sohasi-saqlash texnologiyalari, hisoblash va xizmatlarni o'z ichiga oladi. Shuni ta'kidlash kerakki, ushbu sohaga an'anaviy usullardan foydalangan holda qayta ishlash qiyin bo'lgan katta miqdordagi

ma'lumotlarni qayta ishlash kiradi. An'anaviy va Big Data ma'lumotlar bazasining qiyosiy jadvali.

Xarakteristikasi	An'anaviy ma'lumotlar bazasi	Big Data
Axborot hajmi	Gigabaytdan (10^9 bayt)terabaytgacha (10^{12} bayt)	Petabaytdan (10^{15} bayt) eksabaytgacha (10^{18} bayt)
Saqlash usuli	markazlashgan	Markazlashmagan
Ma'lumotning strukturasi	Strukturalashgan	Yarim strukturalashgan va strukturalashmagan
Ma'lumotlarni saqlash va qayta ishlash modeli	Vertikal model	Gorizantal model
Ma'lumotlarning bog'lanishi	Kuchli	Kuchsiz

Katta ma'lumotlarni qayta ishlash texnologiyalari biz uchun ishlab chiqarishni rejalashtirish, ta'lim, sog'liqni saqlash va boshqa sohalarda yangi ufqlarni ochadi. Agar ularning rivojlanishi davom etsa, BigData texnologiyalari ishlab chiqarish omili sifatida axborotni mutlaqo yangi sifat darajasiga ko'tarishi mumkin. Axborot nafaqat mehnat va kapitalga tenglashadi, balki zamonaviy iqtisodiyotning eng muhim resursiga aylanadi.

Inson faoliyatining turli sohalarida katta hajmdagi ma'lumotlarni qayta ishlash va saqlash kompyuterlarni qo'llashning muhim yo'nalishlaridan iqtisodiyot, bank, savdo, transport, tibbiyot, fan va hokazolardan biridir. Mavjud zamonaviy axborot tizimlari juda katta hajmdagi saqlanadigan va qayta ishlanadigan ma'lumotlar, murakkab tashkil etilishi, ko'plab foydalanuvchilarning turli talablarini qondirish zarurati bilan ajralib turadi.

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 ob’ektlar 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 ob’ektlar 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. Bunda ma’lumotlar modelini to’g’ri tanlash juda muhimdir.

Ma’lumotlar modeli – bu dunyoni idrok etishning asosiy toifalarini, uning ob’ektlari, bog’lanishlari, xususiyatlari, shuningdek, ularning o’zaro ta’sirlari bilan ifodalanishining tasviridir.

Ma’lumotlar bazasi (MB) - ko’rib chiqilayotgan sohadagi ob’ektlar holati va ularning o’zaro munosabatlarini aks ettiruvchi ma’lumotlar to’plamidir¹. Ma’lumotlar bazasidagi ma’lumotlar tartibli ravishda saqlanadi. Shunday qilib, daftarchada barcha yozuvlar alifbo tartibida, kutubxona katalogida esa alifbo bo’yicha (alifbo katalogi) yoki sohaga (mavzu katalogi) muvofiq tartiblangan. Ma’lumotlar bazasini yaratishga, unda saqlanadigan ma’lumotlarni yangilashga, ko’rish va izlashga va unga qulay kirish imkoniyatini yaratishga imkon beradigan dasturlar tizimi ma’lumotlar bazasini boshqarish tizimi deb ataladi (MBBT). Zamonaviy ma’lumotlar bazalariga va binobarin, (MBBT) ga quyidagi asosiy talablar qo’yiladi:

➤ Yuqori harakatchanlik (so’rovga qisqa javob vaqti).

Javob vaqti - ma’lumotlar bazasiga so’rov yuborilgan paytdan boshlab ma’lumotlarni qabul qilinishigacha bo’lgan vaqt oralig’i. Shunga o’xshash yana bir atama bu - “Kirish Vaqti” hisoblanadi.

Kirish vaqti - buyrug’ berilishi va ma’lumotlarni qabul qilish o’rtasidagi vaqt oralig’i. Kirish deganda ma’lumotlarni qidirish, o’qish yoki yozish operatsiyasi

tushuniladi. Ma'lumotlarni yozish, o'chirish va o'zgartirish operatsiyalari ko'pincha yangilanish operatsiyasi deb nomlanadi.

- Ma'lumotlar yangilanishining osonligi.
- Ma'lumotlarning mustaqil bo'lishi.
- Ko'p foydalanuvchilar o'rtasidagi ma'lumot almashinuvi.
- Ma'lumotlar bazalarini qurish va ishlashini standartlashtirish (MBBT).
- Tegishli mavzu sohasidagi ma'lumotlarni ko'rsatishning yetarliligi.
- Foydalanuvchi uchun do'stona interfeys.

Bir-biriga qarama-qarshi bo'lgan dastlabki ikki talab eng muhimlardan hisoblanadi:

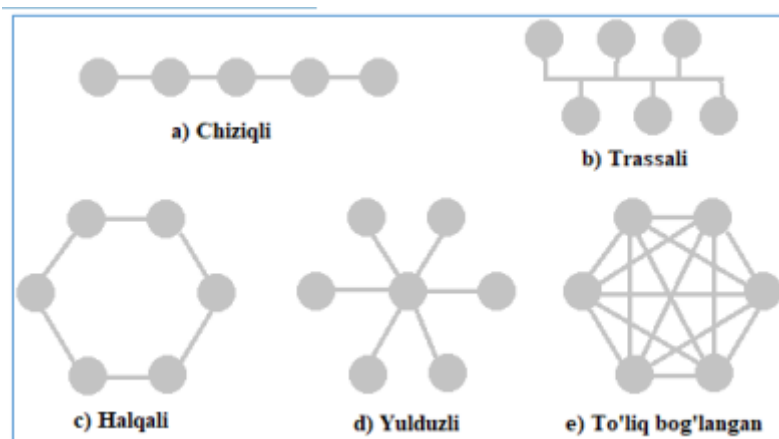
Yuqori harakatchanlik ma'lumotlar bazasi tuzilishini soddalashtirishni talab qiladi. Bu esa o'z navbatida ma'lumotlarni yangilash tartibini murakkablashtiradi va ularning ortiqcha hajmini oshiradi.

Ma'lumotlarning mustaqil bo'lishi - foydalanuvchi qarashlarini o'zgartirmasdan ma'lumotlar bazasining mantiqiy va fizik tuzilishini o'zgartirish qobiliyati. Ma'lumotlarning mustaqilligi ma'lumotlarni saqlash, dasturiy ta'minot va apparat vositalarining xususiyatlariga qarab o'zgarmaslikni anglatadi. Ma'lumotlarga kirish strategiyasini va asl ma'lumotlarning tuzilishini o'zgartirganda, MB tarkibida minimal o'zgarishlarni ta'minlaydi. Bunga loyihalashtirish davrida barcha o'zgarishlarni konseptual va mantiqiy bosqichlariga minimal o'zgartirishlar bilan "siljitish" orqali erishiladi.

Bugungi kunda zamonaviy ilm fan taraqqiy etishi, dunyo bo'ylab ma'lumotlarni raqamlashtirish jarayonining tezlashishi katta va kichik serverlarda hamda shaxsiy qurilmalarda hozirgi zamonaviy texnologiyalar yordamida to'plash va qayta ishlash murakkab bo'lgan katta hajimli turli xil ma'lumotlar oqimi (Big Data) ning hosil bo'lishiga sabab bo'lmoqda. Bu katta hajimli turli xil ma'lumotlar oqimi ularni saqlash va qayta ishlash bilan bog'liq bir qancha muammolarni hosil qilishi bilan birgalikda, ulardan samarali foydalanish ya'ni uni to'liq va to'g'ri tahlil qilish ma'lumotlar ishonchligini

oshirib, insonlarga barcha sohalarda to'g'ri qarorlar qabul qilishda katta imkoniyatlarni ochib berish imkoniyatiga ega. Shu maqsadda hozirgi kunda katta hajimli ma'lumotlar (Big Data)ga oid ko'plab ilmoy tadqiqot ishlari olib borilmoqada. Bu ishlarning aksariyati Big Dataning asosiy xususiyati ya'ni katta hajimli ma'lumotlarni tezkorlik bilan qayta ishlash va tahlil qilish samaradorligini oshirishga qaratilgan bo'lib, tadqiqot natijalari odatiy yakka hisoblash tizimlaridan foydalanish kutilgan natijalarni bermasligini ko'rsatmoqda.

Ma'lumotlarning katta hajimda ekanligini va uni yagona kompyuter yoki serverda qayta ishlash imkoniyati pastligi hisobga olinsa, eng yuqori natija beruvchi yondashuv taqsimlangan hisoblash mexanizmlari orqali ma'lumotlarni yig'ish va qayta ishlashdir. Ma'lumotlarning katta hajimda ekanligi sababli taqsimlangan hisoblash tizimlarida jarayonlarni to'g'ri tashkil etish tizim samaradorligiga ijobiy tasir etadi. Ya'ni Big Datani qayta ishlashda taqsimlangan hisoblash tizimining samaradorligi tizimning bir nechta xususiyat (omil)lariga bog'liq bo'ladi. Shunday taqsimlangan hisoblash tizimi samaradorligiga chambarchas bog'liq bo'lgan muhim omillardan biri - bu tizim abonentlari o'rtasidagi jarayonlararo aloqa va sinxronizatsiya. Ayniqsa, katta hajimdagi turli xillik xususiyatiga ega ma'lumotlarni bir serverdan ikkinchisiga to'qnashuvlarsiz va yo'qotishsiz uzatish tizim samaradorligining qiyin bo'g'inidir. Bu, shuningdek, taqsimlangan tizimlarda parallel ishlash jarayoniga ham tasir qiladi. Shuning uchun Big Datani qayta ishlash uchun taqsimlangan hisoblash tizimining o'zaro bog'lanish sxemasini tanlash ish samaradorligiga ta'sir qiladi. Taqsimlangan hisoblash tizimlarining bir qancha o'zaro bog'lanish sxemalari mavjud bo'lib, ulardan asosiylari 2.5-rasmda ko'rsatilgan.



2.5-rasm. Taqsimlangan hisoblash tizimlarining oddiy o'zaro bog'lanish sxemalari.

Ushbu rasmda aylana hisoblash tugunini, chiziqli esa tugunlar orasidagi to'g'ridan - to'g'ri aloqa kanalini ifodalaydi Chiziqli o'zaro bog'lanish sxemasi - hisoblash tugunlari bitta chiziqqa joylashtirilgan va ulangan. Marshrutlash oddiy va topologiyani rekursiv sifatida ko'rish mumkin. Biroq, qo'shni bo'lmagan har qanday ikkita tugun o'rtasidagi aloqa boshqa tugunlarning yordamiga muhtoj, har qanday oraliq tugundagi nosozlik butun tizimni buzadi.

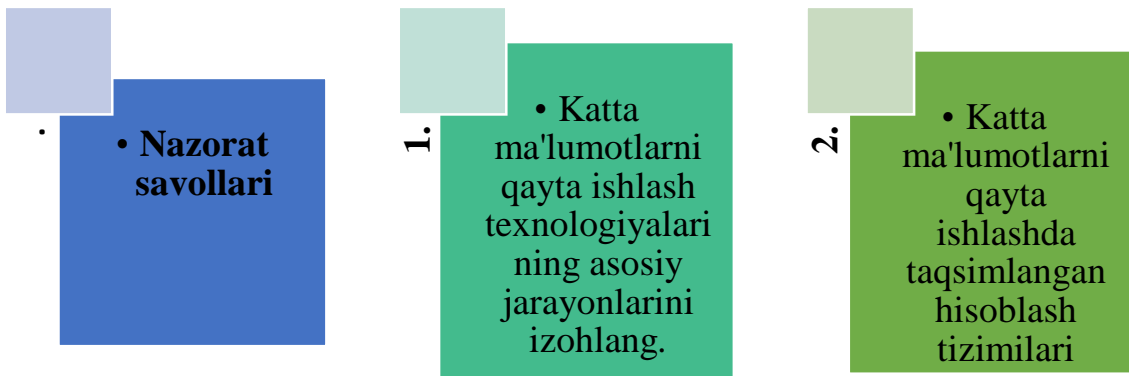
Trassali o'zaro bog'lanish sxemasida har qanday ikkita tugun o'rtasida to'g'ridan-to'g'ri ulanish mavjud. O'zaro ulanish umumiy trassa orqali amalga oshiriladi. Bu bog'lanishlardagi murakkablikni sezilarli darajada kamaytiradi. Biroq, trassadan katta ma'lumotlarning oqib o'tishini hisobga olsak har bir abonent ma'lumotlarni uzatishda aloqa kanalini uzoq vaqt band qiladi. Natijada bir vaqtda uzatilgan xabarlar to'qnashuvi sababli o'z manziliga yetib bormaydi.

Halqali o'zaro bog'lanish sxemasi - bu chiziqli o'zaro bog'lanish sxemasining ikki uchi o'rtasida qo'shimcha ulanish bilan takomillashtirilgan sxema. Bu o'zaro ulanish masofasini 2 martaga kamaytiradi. Biroq, asosiy xususiyatlar hali ham bir xil, Big Datani qayta ishlash uchun yuqori ishlash yoki ishonchlilikni yarata olmaydi hamda tizim miqyosi kattalashgani sari unumdorlik tez pasayadi.

Yulduzli o'zaro bog'lanish sxemasi, barcha abonentlarni birlashtiruvchi markaziy abonentga ega. Har bir aloqa kanali faqatgina ikkita abonentga xizmat

qilganligi uchun ma'lumotlar to'qnashuvi kuzatilmaydi. Bundan tashqari o'zaro ulanish masofasi 2 ga teng bo'lib, markaziy kommutator tugunining yordami bilan kollektiv aloqani oson amalga oshirishni qo'llab - quvvatlaydi va rekursiv kengayish imkonini beradi. Lekin ma'lumotlar hajmining kattaligi, ularni qayta ishlash va abonentlar o'rtasidagi sinxronizatsiyani amalga oshirish vazifalarining ko'pligi sababli markaziy kompyuter doimo yuklanish bilan ishlaydi va markaziy kompyuterning har qanday nosozlikka uchrashi o'zaro aloqani yo'qotadi. Bu Big Datani qayta ishlash ishonchliligiga zarar yetkazadi.

To'liq bog'langan o'zaro bog'lanish sxemasi. Unda har qanday ikkita hisoblash tugunlari o'rtasida to'g'ridan-to'g'ri aloqa mavjud. O'zaro bog'lanish masofasi 1 ga teng va har bir aloqa kanal faqatgina ikkita hisoblash tugunini bog'laydi. Natijada bular Big Datani qayta ishlash ishonchliligini oshiradi. Katta hajimdagi ma'lumotlarni saqlash va qayta ishlashda to'liq bog'langan o'zaro bog'lanish sxemasi mumkin bo'lgan eng yuqori samaradorlikni va ishonchlilikni keltirib chiqaradi.



Mavzuni mustahkamlash uchun savollar.

1. Quyidagi ta'riflardan qaysi biri BigData-ga taalluqli emas?

- a) Yuqoridagilarning barchasi taalluqli
- b) Katta hajmdagi ma'lumotlarni saqlash va qayta ishlash muammosi
- c) Katta hajmdagi ma'lumotlarni qayta ishlash usullarining kompleksto'plami
- d) Ma'lumotlar manbalarining yuqori o'zgaruvchanligi va munosabatlarning murakkabligi bilan bog'liq ma'lumotlar;

2. Muayyan hodisaning yuzaga kelish ehtimolini bashorat qilish uchun modelni yaratish va tanlash jarayoni nima deb nomlanadi?

- a) Bashoratli modellashtirish
- b) OLAP
- c) Ma'lumotlarni qazib olish
- d) Ma'lumotlar ilmi

3. Biznes intellektida qanday analitik vositalar mavjud?

- a) Bashoratli modellashtirish va ma'lumotlarni qazib olish
- b) MS Office bilan integratsiya
- c) BI infratuzilmasi
- d) Hisobot vositalari