

9-§. Katta hajmdagi ma'lumotlarga bog'liq muammolar va kamchiliklar.

REJA:

1.Katta hajmdagi ma'lumotlarga bog'liq muammolar va kamchiliklar

2.Katta ma'lumotlarni qayta ishlash texnologiyalarining asoslari

Katta hajmli ma'lumotlar platformalari tizimda yuqori tezlik va aniq hamda katta miqdordagi ma'lumotlarni qayta ishlash uchun maxsus ishlab chiqilgan. Ushbu katta hajmli ma'lumotlar platformalari odatda turli xil serverlar, ma'lumotlar bazalaridan iborat bo'lib, ular ma'lumotlar izlash qayta izlash uchun ma'lumotlarni manipulyatsiya qilish imkonini beradi.

Katta ma'lumotlar mohiyatiga ko'ra tushunchaga ega bo'lish uchun har bir atributni batafsil ko'rib chiqish foydali bo'ladi.

Katta hajmli ma'lumotlar (Big Data) o'lchami juda ulkandir. An'anaviy ma'lumotlar megabayt, gigabayt va terabayt kabi tanish o'lchamlarda o'lchansada, katta hajmli ma'lumotlar petabayt va zettabaytlarda saqlanadi.

An'anaviy yondashuv	Big Data yondashuvi
Katta bo'lmagan ma'lumotlar paketlarini bosqichma-bosqich tahlil qilish	Bir vaqtning o'zida mavjud ma'lumotlarning barchasini tahlil qilish
Qayta ishlashdan oldin tahrirlash va saralash	Ma'lumotlar asl shaklida qayta ishlanadi
Gipotezani ma'lumotlarga nisbatan solishtirish	Ma'lumot olish uchun barcha ma'lumotlar bo'yicha korrelyatsiyalarni qidirish
Ma'lumotlar yig'iladi, qayta ishlanadi, saqlanadi va keyinchalik tahlil qilinadi	Katta ma'lumotlarni real vaqt rejimida tahlil qilish va qayta ishlash

Masshtabdagi farqning ulkanligini tushunish uchun Berkeley Axborot maktabidagi ushbu taqqoslashni ko'rib chiqing: bir gigabayt HD formatidagi yetti daqiqali videoga teng, bir zettabayt esa 250 milliard DVDga teng. Bu aysbergning faqat uchi. EMC hisobotiga ko'ra, raqamli olam har ikki yilda ikki baravar o'sib boradi va 2020 yilda 44 trillion zettabaytga yetgan.

Katta hajmli ma'lumotlar ushbu turdagi ma'lumotlarni qayta ishlash arxitekturasini ta'minlaydi. Saqlash va qayta ishlash bo'yicha tegishli yechimlarsiz, tushunchalar tagiga yetish qiyinchilik tug'dirardi.

Uni yaratish tezligidan tortib, tahlil qilish uchun zarur bo'lgan vaqtgacha, katta hajmli ma'lumotlardan real vaqt rejimida foydalanish va undan tushunchalar yaratish qobiliyatiga ega bo'lishi kerak, aks holda bu unchalik foydali emas. Haqiqiy vaqtda ishlov berish qaror qabul qiluvchilarga tezkorlik bilan harakat qilish imkonini beradi, bu esa ularga raqobatda ustunlik beradi.

Ma'lumotlarning ba'zi shakllari ommaviy qayta ishlanishi va vaqt o'tishi bilan dolzarb bo'lib qolishi bilan ahamiyatlidir.

Masalan, sog'liqni saqlash tizimida katta hajmli ma'lumotlar bilan ishlash jarayoni ancha murakkab bo'lib, qurilmalaridan olingan sensor ma'lumotlarini bazada yig'ish hamda qayta ishlash va foydalanuvchiga taqdim etish ancha murakkabliklarni keltirib chiqaradi.

Katta hajmli ma'lumotlarda bulutli saqlashdan foydalanish, onlayn ma'lumotlarni saqlash va ular bilan ishlash texnologiyalari kichik va o'rta biznes uchun ko'plab muammolarni hal qilishi mumkin: ma'lumotlar markaziga xizmat ko'rsatishdan ko'ra bulutni sotib olish arzonroq, xodimlar masofadan ishlashi mumkin va joy(ofis) kabi maommalarga ehtiyoj sezilmaydi.

Katta hajmli ma'lumotlarda chuqur o'rganish, sun'iy intellect alohida ahamiyatga egadir. Analitik mashinalar inson miyasini taqlid qiladi, ya'ni sun'iy neyron tarmoqlardan foydalaniladi. O'rganish katta hajmdagi ma'lumotlar asosida mustaqil ravishda amalga oshiriladi.

Dark Data - kompaniya to'g'risidagi raqamlashtirilmagan ma'lumotlarni to'plash va saqlash, ular biznesni rivojlantirishda muhim rol o'ynamaydi, lekin ular texnik va qonunchilik rejalarida zarur.

Blokcheyn. Internet tranzaksiyalarini soddalashtirish, bu operatsiyalarning narxini pasaytirish. O'z-o'ziga xizmat ko'rsatish tizimlari - 2016 yildan boshlab kichik va o'rta biznes uchun maxsus platformalar joriy etildi,

ularda siz mustaqil ravishda ma'lumotlarni saqlashingiz va tartibga solishingiz mumkin.

Katta ma'lumotlar bilan bog'liq eng katta muammo - uni qayta ishlash narxi. Bu qimmat uskunalar va katta hajmdagi ma'lumotlarga xizmat ko'rsatishga qodir bo'lgan malakali mutaxassislarga ish haqi xarajatlarini o'z ichiga olishi mumkin. Shubhasiz, ma'lumotlar miqdori oshgani sayin minimal ishlashni yo'qotmaslik uchun uskunani muntazam yangilab turish kerak bo'ladi. Ikkinchi muammo yana qayta ishlanishi kerak bo'lgan katta hajmdagi ma'lumotlar bilan bog'liq. Agar, masalan, tadqiqot 2-3 emas, balki ko'p sonli natijalarni bersa, ob'ektiv bo'lib qolish va umumiy ma'lumotlar oqimidan faqat har qanday hodisaning holatiga haqiqiy ta'sir ko'rsatadiganlarini tanlash juda qiyin. Katta ma'lumotlarning maxfiyligi muammosi. Aksariyat mijozlarga xizmat ko'rsatuvchi provayderlar ma'lumotlardan onlayn foydalanishga o'tishlari bilan kiberjinoyatchilarning navbatdagi nishoniga aylanish juda oson. Hech qanday onlayn tranzaksiyalarsiz shaxsiy ma'lumotlarni oddiy saqlash ham bulutli saqlash mijozlari uchun nomaqbul oqibatlarga olib kelishi mumkin.

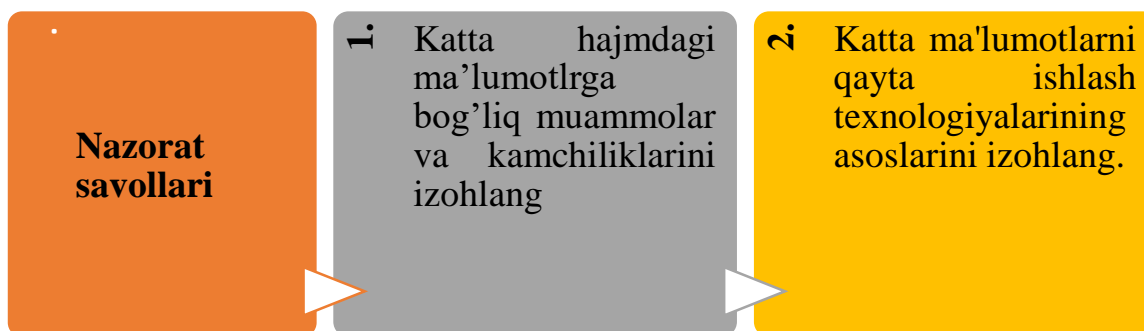
Axborotni yo'qotish muammosi. Ehtiyot choralari oddiy bir martalik ma'lumotlarni zaxiralash bilan cheklanib qolmaslikni, balki saqlashning kamida 2-3 zaxira nusxasini yaratishni talab qiladi. Biroq, hajm oshgani sayin, ortiqcha ish bilan bog'liq qiyinchiliklar o'sib bormoqda - va IT mutaxassislari ushbu muammoning maqbul echimini topishga harakat qilmoqdalar. Katta ma'lumotlardan unchalik kam foydalanilmayotganining sabablaridan biri katta ma'lumotlar va katta ma'lumotlar texnologiyalari ham ko'p muammolarni keltirib chiqaradi. Bir tadqiqot shuni ko'rsatdiki, katta ma'lumotlar loyihalarining 55 foizi hech qachon tugallanmagan. Ushbu topilma ikkinchi so'rovda takrorlandi, natijada mahalliy yirik ma'lumotlar loyihalarining aksariyati muvaffaqiyatli emas.

Masshtablilik. Katta ma'lumotlar bilan talab bo'yicha kattalashtirish va pasaytirish juda muhim. Ko'pgina tashkilotlar katta ma'lumotlar loyihasi

qanchalik tez o'sishi va rivojlanishini hisobga olmaydilar. Qo'shimcha resurslarni qo'shish uchun loyihani doimiy ravishda to'xtatib turish ma'lumotlarni tahlil qilish vaqtini qisqartiradi. Katta ma'lumotlarning ish yuklari ham tez sur'atda bo'ladi, bu esa resurslarni qaerga taqsimlash kerakligini oldindan aytishni qiyinlashtiradi. Ushbu katta ma'lumotlar muammosining darajasi yechimga qarab farq qiladi. Bulutdagi yechim mahalliy yechimga qaraganda ancha oson va tezroq masshtablanadi.

Iste'dodning etishmasligi: korxonalar ma'lumotlarning etishmasligini his qilmoqdalar. Nafaqat ma'lumotlar bo'yicha olimlarning etishmasligi, balki katta ma'lumotlar loyihasini muvaffaqiyatli amalga oshirish uchun qimmatli tushunchalarni aniqlash uchun etarli miqdordagi domen bilimiga ega bo'lgan ishlab chiquvchilar, ma'lumotlar bo'yicha olimlar va tahlilchilarning murakkab jamoasi talab qilinadi. Ko'pgina yirik ma'lumotlar sotuvchilari o'zlarining ta'lim resurslarini taqdim etish yoki boshqaruvning asosiy qismini ta'minlash orqali bu katta ma'lumotlar muammosini engishga intilishadi.

Hadoop qiyin. Hadoop va uning atrofidagi vositalar ekotizimining katta hajmdagi tuzilgan va tuzilmagan ma'lumotlar bilan ishlash qobiliyati maqtovg'a sazovor bo'lsa-da, dasturiy ta'minotni boshqarish yoki ishlatish oson emas. Texnologiya nisbatan yangi bo'lganligi sababli, ko'plab ma'lumotlar mutaxassislari Hadoopni qanday boshqarishni bilishmaydi. Bundan tashqari, Hadoop ko'pincha qo'llab-quvvatlash uchun keng ichki resurslarni talab qiladi va ko'plab kompaniyalar o'z resurslarining katta qismini hal qilmoqchi bo'lgan haqiqiy katta ma'lumotlar muammosiga emas, balki texnologiyaga bag'ishlaydilar. Yuqorida aytib o'tilgan so'rovda respondentlarning 73% katta ma'lumotlar platformasini tushunish katta ma'lumotlar loyihasining eng muhim muammosi ekanligini ta'kidladi.



Mavzuni mustahkamlash uchun savollar.

1. Umumiy samaradorlikni oshirish uchun tizimning har bir komponentini ishlashini oshirish- bu...?

- a) Vertikal masshtablash
- b) Master- slave replikatsiyasi
- c) Gorizontal masshtablash
- d) Peer-to-pee replikatsiya

2.«Business Intelligence» integratsiyasi vositasi nima?

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

3. Anʼanaviy AT tizimlari Hadoop kabi katta hajmdagi maʼlumotlar texnologiyalari bilan birlashganda nimaga asos boʻlishi mumkin?

- a) Katta hajmdagi maʼlumotlarni boshqarish va maʼlumotlarni qidirish
- b) Katta hajmdagi maʼlumotlarni boshqarish
- c) Katta hajmdagi maʼlumotlarni qidirish
- d) Strukturalanmagan maʼlumotlarni toʻplash va saqlash