

17-§. Katta hajmdagi ma'lumotlarni yaratish va boshqarish

Reja:

- 1. Katta hajmdagi ma'lumotlarni yaratish va boshqarish usullari.*
- 2. Katta ma'lumotlarni boshqarish arxitekturasini yaratish.*

Katta hajmdagi ma'lumotlar – qimmat resurs hisoblanadi. Ma'lumot dolzarbligi, foydaliligi va ichidagi kontentiga qarab uning qanchalik qimmat tranzaksiya ma'lumotlari ham onlayn, ham oflayn rejimda amalga oshiriladigan barcha kundalik tranzaksiyalardan hosil bo'ladi. Hisob-fakturalar, to'lov topshiriqlari, saqlash yozuvlari, yetkazib berish kvitansiyalari - barchasi tranzaksiya ma'lumotlari sifatida tavsiflanadi, biroq ma'lumotlarning o'zi deyarli ma'nosizdir va aksariyat tashkilotlar o'zlari yaratayotgan ma'lumotlar va undan qanday foydalanish mumkinligini tushunish uchun kurashadilar.

Katta ma'lumotlar raqobatbardosh ustunlikni ta'minlashi mumkinligi haqida umumiy e'tirof etilgan tushuncha mavjud bo'lsa-da, murakkab uchinchi tomon provayderlari bilan hamkorlik qilayotganlar yuqori sifatli va hamyonbop ma'lumotlardan foydalanishda ancha yaxshi imkoniyatga ega. Katta ma'lumotlar davri biz uchun yaxshi va haqiqatan ham keldi va endi korxonalar katta ma'lumotlar bilan shug'ullanishi kerakmi yoki yo'qmi degan savol emas, balki qanday qilib. Texnologik gigant Cisco 2020-yilda ishlab chiqarilgan ma'lumotlar miqdori bugungi kunga nisbatan 50 baravar ko'p bo'lishini bashorat qilmoqda. Ajablanarli emaski, kompaniyalar o'zlarini haddan tashqari ko'tarib, o'z bizneslarini tushunadigan va natijalarni berish uchun uni texnologiya bilan birlashtira oladigan.

Faol bo'lish asosiy hisoblanadi. An'anaviy hisobot va BI o'z o'rnini Advanced Analyticsga bo'shatib bermoqda. Tashkilotlar ma'lumotlarni yetkazib berish va boshqarishdan tortib, ma'lumotlardan foydalanishgacha bo'lgan bulutli katta ma'lumotlar xizmati sifatida analitik talablarning to'liq spektri bo'ylab

biznes ehtiyojlarini qondirishi mumkin. Bulutga asoslangan keng qamrovli katta ma'lumotlar ma'lumotlar strategiyasini ishlab chiqish orqali ular tushuncha doirasini belgilashi va korxona ma'lumotlarining umumiy qiymatini optimallashtirishi mumkin. Biroq, bulutga asoslangan katta ma'lumotlar tahlili yagona o'lchamdagi yechim emas va CloudMoyo kabi AT bo'yicha mutaxassis hamkor bu yo'lda sizga yordam berishi mumkin.

Katta ma'lumotlar manbalari ko'p yillar davomida mavjud bo'lgan. Marketingda ularni mijozlar xaridlari, kredit tarixlari, turmush tarzi va hokazolarning ma'lumotlar bazasi deb atash mumkin. Yillar davomida tahlilchilar ushbu ma'lumotlardan kompaniyalarga mijozlarning kelajakdagi ehtiyojlarini bashorat qilish, xatarlarni baholash, iste'molchilarning xohish-istaklarini shakllantirish va h.k.

Hozirgi vaqtda vaziyat ikki jihatdan o'zgargan:

- turli xil ma'lumotlar to'plamlarini tahlil qilish va taqqoslash uchun yanada murakkab vositalar va usullar paydo bo'ldi;

Katta ma'lumotlar haqiqatan ham yangi yoki bu ma'lumotlarni boshqarishda boshqa to'lqinlar singari, katta ma'lumotlar ham so'nggi besh o'n yillikda ma'lumotlarni boshqarish amaliyoti evolyutsiyasi asosida qurilgan. Yangilik shundaki, birinchi marta hisoblash tsikllari va saqlash narxi eng yuqori darajaga yetdi. Faqat bir necha yil oldin, tashkilotlar odatda tasvirlarni yoki muhim ma'lumotlarning pastki qismlarini saqlash bilan murosaga kelishadi, chunki saqlash va qayta ishlashga sarflanadigan xarajatlar tahlil qilishni istagan hamma narsani saqlashga taqiq qo'ydi.

Katta ma'lumotlarga ega bo'lgan holda, endi ma'lumotlarni virtualizatsiya qilish mumkin, shunda ular samarali tarzda saqlanishi mumkin va bulutga asoslangan xotiradan foydalanib, iqtisodiy jihatdan ham samaraliroq bo'ladi. Bundan tashqari, tarmoq tezligi va ishonchliligining yaxshilanishi katta miqdordagi ma'lumotlarni maqbul darajada boshqarish imkoniyatiga ega bo'lishning boshqa jismoniy cheklovlarini

olib tashladi. Virtualizatsiya, parallel ishlov berish, tarqatilgan fayl tizimlari va xotiradagi ma'lumotlar bazalari kabi katta ma'lumotlarning asosida joylashgan ko'plab texnologiyalar o'nlab yillar davomida mavjud edi. Ilg'or tahlillar ham o'nlab yillar davomida mavjud bo'lib, ular har doim ham amaliy bo'lmagan. Hadoop va MapReduce kabi boshqa texnologiyalar sahnada bir necha yil davomida mavjud. Texnologiya yutuqlarining ushbu kombinatsiyasi endi muhim biznes muammolarini hal qilishi mumkin. Korxonalar har xil ma'lumotlardan kerakli tezlikda tushunchalar va amaliy natijalarga erishishni xohlashadi - qancha ma'lumot jalb qilinmasin. Agar kompaniyalar petabayt ma'lumotlarini (matnli fayllar bilan to'ldirilgan to'rt millionli to'rtta tortma fayllar shkafiga yoki 13,3 yillik HDTV tarkibiga teng) maqbul ishlash ko'rsatkichlari bilan naqsh va anomalialarni tahlil qila olsalar, korxonalar ma'lumotlarni yangi usullar bilan tushunishni boshlashlari mumkin.

Katta ma'lumotlarga o'tish nafaqat biznes bilan bog'liq. Ilm-fan, tadqiqotlar va hukumat faoliyati ham uni oldinga siljishga yordam berdi. Ma'lumotlar bilan ishlashga turli xil yondashuvlar harakatdagi ma'lumotlar yoki o'zgarmas holatidagi ma'lumotlarga asoslangan holda mavjud. Harakatdagi ma'lumotlar, asosan kompaniya faoliyatida xatolarga yo'l qo'ymaslik uchun ishlab chiqarish jarayonida mahsulot sifatini tahlil qila olishda foydalaniladi. O'zgarmas holatidagi ma'lumotlar biznes-tahlilchi tomonidan xaridorlar munosabatlarining barcha jabhalariga, shu jumladan sotuvlar, ijtimoiy tarmoqlar ma'lumotlari va mijozlarga xizmat ko'rsatish jarayonlariga asoslangan mijozlarning amaldagi sotib olish tartibini yaxshiroq tushunish uchun foydalaniladi.

Katta ma'lumotlar - bu bitta texnologiya emas, balki kompaniyalarga amaliy tushuncha olishga yordam beradigan eski va yangi texnologiyalarning kombinatsiyasi. Shuning uchun katta ma'lumotlar - bu turli xil ma'lumotlarning kerakli hajmini, kerakli tezlikda va real vaqt rejimida tahlil qilish uchun kerakli

vaqt ichida boshqarish imkoniyatidir. Katta ma'lumotlar odatda uchta xususiyat bo'yicha taqsimlanadi:

- Hajmi: qancha ma'lumot;
- Tezlik: bu ma'lumotlar qanchalik tez qayta ishlanadi;
- Xilma-xillik: har xil turdagi ma'lumotlar.

Ma'lumotlar aniqlik va kontekst asosida tekshirilishi kerak. Innovatsion biznes ushbu mijozning qiymatini va ushbu mijozga qo'shimcha takliflarni taqdim etish imkoniyatini tezda baholash uchun real vaqtda katta hajmdagi ma'lumotlarni tahlil qilishni xohlashi mumkin. Biznes natijalariga ta'sir qilish uchun tahlil qilinadigan ma'lumotlarning to'g'ri miqdori va turlarini aniqlash kerak. Katta ma'lumotlar barcha ma'lumotlarni, shu jumladan tuzilgan ma'lumotlar va electron pochtdan, ijtimoiy tarmoqlardan, matnli oqimlardan va boshqa narsalardan tuzilmaydigan ma'lumotlarni o'z ichiga oladi. Ushbu turdagi ma'lumotlarni boshqarish kompaniyalardan o'zlarining tuzilgan va tuzilmagan ma'lumotlaridan foydalanishni talab qiladi.

Katta ma'lumotlarni boshqarishda CMSning roli. Tashkilot ma'lumotlar bazalarida ba'zi strukturalanmagan ma'lumotlarni saqlaydi. Shu bilan birga, ular kontentning to'liq hayotiy davrini boshqarishi mumkin bo'lgan korporativ kontentni boshqarish tizimlaridan (CMS) foydalanadilar. Bunga veb-kontent, hujjat kontenti va boshqa ommaviy axborot vositalari kirishi mumkin.

Axborot va tasvirlarni boshqarish assotsiatsiyasi ma'lumotlariga ko'ra, ta'lim, tadqiqotlar va ilg'or tajribalarni ta'minlaydigan notijorat tashkilot, korporativ kontentni boshqarish (ECM) tarkibiga "strategiyalar, usullar va vositalar tashkiliy jarayonlar bilan bog'liq kontent va hujjatlarni olish, boshqarish, yig'ish, saqlash va etkazib berish" kiradi. ECM-ga kiritilgan texnologiyalar hujjatlarni boshqarish, yozuvlarni boshqarish, tasvirlash, ish oqimlarini boshqarish, veb-kontentni boshqarish va hamkorlik.

Butun sanoat kontentni boshqarish asosida o'sdi va ko'plab kontentni savdosi bilan shug'ullanuvchilar bu katta hajmdagi strukturalanmagan

ma'lumotlarga ishlov berish o'zlarining yechimlari hajmini kengaytirmoqda. Shu bilan birga, strukturalanmagan ma'lumotlarni qo'llab-quvvatlashga va strukturalanmagan ma'lumotlarni tahlil qilishga yordam beradigan yangi texnologiyalar ham rivojlanib bormoqda. Ulardan ba'zilar ham strukturalangan, ham strukturalanmagan ma'lumotlarni qo'llab-quvvatlaydi. Ba'zilar real vaqt potoklarini qo'llab-quvvatlaydi. Ular orasida Hadoop, MapReduce va potok kabi texnologiyalar mavjud.

Tarkibni kontentni boshqarish tizimlari shaklida saqlash uchun mo'ljallangan tizimlar endi mustaqil yechimlar emas. Aksincha, ular ma'lumotni boshqarish bo'yicha umumiy yechimning bir qismi bo'lishi mumkin. Masalan, sizning tashkilotingiz keyinchalik CMS qidiruvi dasturiy triggeri bo'ladigan Twitter-kanallarini kuzatishi mumkin. Endi tvitni qo'ygan shaxs u odam o'zi qidirayotgan mahsulotni topishi mumkin bo'lgan joyni taklif qiladigan javobni oladi. Bunda ushbu turdagi o'zaro ta'sirni real vaqtda sodir bo'lishi muhim hisoblanadi. Bundan tashqari, u real vaqt rejimida strukturalanmagan, strukturalangan (mijozning tvitterda yozgan shaxs haqidagi ma'lumotlari) va yarim strukturalangan (CMSdagi asl tarkib) ma'lumotlaridan foydalanish qiymatini ko'rsatadi.

Nazorat savollari

1.

Katta hajmdagi ma'lumotlarni yaratish boshqarish usullarini izohlang.

2.

Katta ma'lumotlarni boshqarish arxitekturasini yaratishni izohlang.



Mavzuni mustahkamlash uchun savollar.

1.Ko'pchilik foydalanuvchilar tomonidan ma'lumotlar bazasini yaratish, to'ldirish va birgalikda foydalanish uchun mo'ljallangan dasturiy vositalar majmuasi nima deyiladi?

- a) Ma'lumotlar bazasini boshqarish tizimi
- b) Sonli o'zgarma
- c) Ma'lumotlar lug'ati
- d) Hisoblash tizimi

2.Quyidagilardan qaysi biri Big Data vizualizatsiya turi emas?

- a) Arxitektura vizualizatsiyasi

- b) Ilmiy vizualizatsiya
- c) Dasturiy ta'minotni vizualizatsiya qilish
- d) Axborotni vizualizatsiya qilish

3. Quyidagi faoliyat tamoyillaridan qaysi biri Big Dataga taalluqli emas?

- a) Vertikal masshtablilik
- b) Gorizontal masshtablilik
- c) Xatolarga chidamlilik
- d) Ma'lumotlarning joylashuvi