

KATTA MA'LUMOTLAR TEXNOLOGIYALARI, ULARNI TAHLIL QILISH, SAQLASH VA QAYTA ISHLASH USULLARI

Yarasheva Gulnoza Baxtiyor qizi

Toshkent davlat iqtisodiyot universiteti Samarqand filiali. Magistrant.

E-mail: gulnozayarasheva0@gmail.com

ANNOTATSIYA

Ushbu ishda Big Data (Katta ma'lumotlar) tushunchasi nazariy jihatdan ko'rib chiqilgan. Katta hajmdagi ma'lumotlarning zamonaviy hayotdagi o'rni o'rganilgan, shuningdek katta ma'lumotlar bilan ishlashda yuzaga keladigan muammolar ko'rib chiqilgan. Katta ma'lumotlar qanday saqlanadi va ularni kim saqlaydi kabi savvollarga javob berilgan.

Kalit so'zlar: Big Data, an'anaviy yondashuv, AI va nutqni aniqlash, DG amaliyoti, axborotni qayta ishlash, DevOps.

BIG DATA TECHNOLOGIES, THEIR ANALYSIS, STORAGE AND PROCESSING METHODS

ABSTRACT

In this work, the concept of Big Data is considered theoretically. The role of large amounts of data in modern life has been studied, as well as problems arising in working with large data have been considered. Questions such as how big data is stored and who stores it are answered.

Key words: Big Data, traditional approach, AI and speech recognition, DG practice, information processing, DevOps.

Katta ma'lumotlar (Big Data) - hajm, tezlik va xilma-xillik. Katta ma'lumotlar - bu hajmi doimiy ravishda o'sib borayotgan va ortib borayotgan tezlikda keladigan turli xil ma'lumotlar. U 2001 yilda paydo bo'lgan va hozirgacha qo'llanilmoqda. Big Data biznes muammolarini hal qiladi: teledasturni tomosha qilishdan tortib tashqi reklama samaradorligigacha har bir sanoatning katta ma'lumotlar mahsulotlariga ehtiyoji va qo'llanilishi mavjud. Qoida tariqasida, bunday mahsulotlar ilgari imkonsiz bo'lgan vazifani to'liqroq tasavvur qilish imkonini beradi.

An'anaviy yondashuv - talab qilinadigan ko'rsatkichlar vakillik, ammo cheklangan tanlov yordamida baholanganda panel tadqiqotlarini o'tkazishdir. Masalan, Moskvadan kelganlar televizorda nima tomosha qilayotganini bilish uchun kompaniya bor-yo'g'i 10 000 ta panel varaqlarini tahlil qiladi va keyin butun shahar aholisiga ekstrapolyatsiya qiladi. Agar dasturni 10 000 kishidan 1000 nafari tomosha qilgan bo'lsa, unda barcha moskvaliklarning taxminan 10 foizi buni qilgan. Katta ma'lumotlar tahlil qilish uchun sezilarli darajada kattaroq namunalardan foydalanishga imkon beradi, chunki u anonim bo'lsa-da, lekin foydalanuvchi xatti-harakatlari haqidagi haqiqiy ma'lumotlarni tahlil qiladi. Bu nimaga olib kelishi mumkinligiga yaxshi misol tashqi tahlillar tarixidir. An'anaga ko'ra, bu sohada o'lchovlarning cheklangan to'plamlari qo'llanilgan: masalan, shaharning ma'lum nuqtalarida ma'lum soatlarda o'tgan avtomobillar soni hisoblangan. Keyin tahlilchilar butun oy uchun murakkab matematik model yordamida ma'lumotlarni ekstrapolyatsiya qilishdi. Tadqiqot natijasida mijoz har bir reklama taxtasi uchun kuniga o'rtacha auditoriya bahosini oldi.

Navigatsiya tizimlari va mobil ilovalardan olingan ma'lumotlardan foydalanish avvalgi modellarni sezilarli darajada aniqroq qildi. Tahlilchilar haftaning kuniga / soatiga qarab har bir dizayn uchun auditoriya profillarini ajratishi mumkin. Tuzilmalarda Wi-Fi datchiklaridan foydalanish esa tomoshabinlarning xatti-harakatlarini o'rganish: ularning yashash joylari va qiziqishlarini ko'rish imkonini berdi. Shunday qilib, datchiklar o'tgan avtomobillarning hammasi moskvaliklarga tegishli emasligini aniqladi. Magistral yo'llarda ko'p odamlar mintaqadan sayohat

qiladilar, boshqa shaharlardan tranzit transporti mavjud. Big Data eng to‘liq tahlilni ta‘minlaydi va undan tashqariga chiqadi. Shuning uchun, katta ma‘lumotlarga asoslanib, siz to‘g‘ri bashorat qilishingiz va mashinani o‘rganish modellarini o‘rgatishingiz mumkin. Misol uchun, qarz oluvchi, ba‘zi xususiyatlarga asoslanib, ishonchsiz bo‘lishi mumkin, deb taxmin qilish. Yoki, aniq dalillarga asoslanib, zavoddagi birlik tez orada texnik xizmat ko‘rsatishni talab qiladi, deb taxmin qiling.

Big Datada kadrlar etishmasligi

Haqiqatan ham mutaxassislar etarli emas: birinchi navbatda, biz DeVOps haqida gapiramiz. Endi ularda haqiqiy bum bor va etarli malakaga ega odamlar etarli emas. Siz katta ma‘lumotlar, klasterlar, Kubernetes bilan ishlaydigan oddiy muhandis yoki tizim ma‘murini topishingiz mumkin, ya‘ni DevOps - bunday mutaxassislarning etishmasligi. MLOplarga kelsak, ba‘zida bunday odamlar odatda faqat nazariy jihatdan mavjud bo‘lib tuyuladi. Bu sohada tizimli ravishda o‘qitilgan odamlar umuman yo‘q.

Bozorda ma‘lumotlar bo‘yicha olimlarning taklifi ko‘proq yoki kamroq talabni qo‘lga kiritdi: endi siz duch kelgan birinchi mutaxassisni qo‘lga kiritishingiz mumkin emas, balki nomzodlarni kuzatib, suhbatlashishingiz mumkin. Xuddi shu narsa ma‘lumotlar muhandislari uchun ham amal qiladi.

AI va nutqni aniqlash: Katta ma‘lumotlarning kelajagi o‘z-o‘zidan uzoq vaqtdan beri ko‘plab sohalarga bo‘lingan: bu shunday atama bo‘lib, bugungi kunda turli xil texnologiyalarni bildiradi. Katta ma‘lumotlarning asosiy o‘sish potentsiali uning sun‘iy intellektda qo‘llanilishi va amaliy Big Data yechimlarini izlash, ular haqiqatan ham foydalidir. Bozor ishtirokchilari allaqachon shov-shuv va umidsizliklar vodiysidan o‘tib ketishdi, endi texnologiyalar mahsuldorlik platosiga kirmoqda. Big Data ilovalari diapazoni monitoring tizimiga "aqli" narsalarni qo‘shish uchun avtomatik tizimlar, hisobotlar, NLP - nutqni tanib olish va uni tushunarli sintez qilish qobiliyati, video tahlilidan foydalangan holda oxirgi foydalanuvchi uchun yanada "aqli" va to‘liq mahsulotlarni taqdim qiladi. Katta ma‘lumotlar atamasi turli yo‘llar bilan talqin qilinishi mumkin. Ba‘zi manbalar biz juda ko‘p tuzilmagan ma‘lumotlar haqida

gapirayotganini aytsa, boshqalari bu ma'lumotlarga vositalar va yondashuvlarni nazarda tutadi.

Katta ma'lumotlar qayerda saqlanadi?

Savollar, albatta, bir-biriga bog'liq - shuning uchun men xarajatlardan boshlashni taklif qilaman. Ko'p saqlash variantlari mavjud, ammo optimallashtirish va oxirgi variantni tanlash ko'pincha qimmatga tushadi. Umuman olganda, ma'lumotlar bilan ishlash xarajatlari quyidagilarga bog'liq:

- ma'lumotlar hajmi va turlari;
- ma'lumotlar yordamida hal qilinadigan vazifalar;
- kerakli ma'lumotlarni qayta ishlash tezligi;
- kiritilgan ma'lumotlarning xilma-xilligi;
- bulutlardan foydalanish imkoniyati;
- ma'lumotlar bilan ishlash nuqtai nazaridan kompaniyaning yetuklik darajasi.

Shu bilan birga, bu erda biz texnologiyalar va ular bilan bog'liq bo'lgan barcha narsalar haqida gapirmayapmiz, chunki. texnologiyalar yuqoridagilarning natijasidir. Vazifalar, menimcha, qiymatning asosiy omili bo'lib, ular biznesning ma'lumotlardan foydalanishdan oladigan qadriyatlarini bilan qoplanishi kerak. Shubhasiz, agar ma'lumotlar iste'molchisi yuqori sifatli vitrinalar yoki murakkab agregatlarni talab qiladigan ML modellari bilan ta'minlanishi kerak bo'lgan tahliliy bo'lim bo'lsa, unda ma'lumotlarni qayta ishlashning murakkabligi sovuq arxivni taqdim etish kerak bo'lgan holatga qaraganda bir necha baravar yuqori bo'ladi. saqlash. Va ishlov berishning murakkabligi protsessor kuchi, xotira. Agar iste'molchiga “besh to‘qqiz” ning ishonchliligi ham kerak bo'lsa, bu ham xarajat multiplikatoridir. Xo'sh, unutmang - ETL qanchalik murakkab bo'lsa, uni yozish va keyingi qo'llab-quvvatlash shunchalik qimmatroq bo'ladi. Kerakli ishlov berish tezligi. Agar siz daqiqada bir marta yangilanishni ko'rsatishi kerak bo'lgan asboblarning paneliga ma'lumotlar oqimini yoki firibgarlarni real vaqtda ushlashi kerak bo'lgan firibgarlikka qarshi modelni ta'minlashingiz kerak bo'lsa, bu haftada bir marta yangilanishi mumkin bo'lgan bir xil

vitriinalarga qaraganda ancha qimmatroq bo‘ladi. dam olish kunlari yoki har oyda bir marta hamma narsa birlashtiriladigan sovuq arxiv. Shunday qilib, tezlik bizga protsessor kuchiga qimmat (ayniqsa yaqinda) talablarni qo‘shadi va undan ham ko‘proq - xotira uchun, shuningdek, bu yechimning ishlab chiqarilishiga ta‘sir qiladi, bu esa ba‘zan umuman “tinga” qo‘shmaydi.

Saqlash uchun bulutdan foydalanish juda jozibador bo‘lishi mumkin: o‘z vaqtida to‘lang, kapital xarajatlarni operatsion xarajatlarga o‘tkazing, tizimni, joyni, elektr energiyasini va boshqa ko‘plab nuanslarni sotib olayotganda aktsiyalar haqida o‘ylashning hojati yo‘q. Boshqa tomondan, bulutli saqlash barcha turdagi ma‘lumotlar uchun mavjud emas (me‘yoriy talablarga rioya qilish kerak) va u barcha vazifalar uchun qo‘llanilmaydi. Shuning uchun, bulutlar yaxshi, lekin bulut va o‘zingizning saqlashingiz o‘rtasidagi chiziqni qaerga chizish kerakligini tushunish uchun juda malakali mutaxassis kerak.

Nima uchun big data kompaniyasining yetuklik darajasi juda muhim?

Ko‘p odamlar ma‘lum jarayonlarning etuklik darajalari qanday ekanligini bilishadi yoki eshitgan. Misol uchun, agar biz IT haqida gapiradigan bo‘lsak, unda siz ITIL-ni ochishingiz mumkin - u erda hamma narsa yaxshi tasvirlangan. Va ma‘lumotlar bilan ishlashda taxminan bir xil, kamroq rasmiylashtirilgan bo‘lsa-da, etuklik darajalari mavjud. Men bu darajalarni bu erda bermayman, chunki ularga hali yagona yondashuv yo‘q va hech bo‘lmaganda de-fakto standartga aylanmagan narsalarni tasvirlash alohida muhokama yoki maqola mavzusidir. Shuning uchun men kompaniyaning ma‘lumotlar bo‘yicha etuklik darajasini korxonada Ma‘lumotlarni boshqarish amaliyotini joriy etish kengligi bilan tenglashtirishni taklif qilaman. Agar siz bu nima ekanligini bilmasangiz va ma‘lumotlarni qayta ishlash amaliyotiga ega bo‘lmasangiz, biz sizni 0 yoki 1-darajada ekanligingizni taxmin qilamiz. Ammo keyin bu “etuklik darajasi” saqlash va qayta ishlash xarajatlariga hech qanday ta‘sir ko‘rsatmaydi. Ammo siz DG amaliyotlarini amalga oshirishni boshlaganingizdan so‘ng, bu sizning etuklik darajangizni oshiradi, lekin ma‘lumotlar bilan ishlash uchun

qo‘shimcha xarajatlarni ham oshiradi: birinchi navbatda ushbu amaliyotlarni ishlab chiqish va amalga oshirish, keyin esa qo‘llab-quvvatlash va qo‘llab-quvvatlash. Shu bilan birga, qaerdan boshlashingiz muhim emas - ma‘lumotlar egalarining namunasi, ma‘lumotlar boshqaruvchilari instituti, biznes lug‘atini yaratish va yuritish, ma‘lumotlar sifatini ta‘minlash yoki boshqa tadbirlar bilan. Va odatdagidek, qaramlik hech qachon chiziqli emas, balki juda eksponensial emas - ya‘ni har bir keyingi daraja avvalgisidan qimmatroq. DG amaliyotlari yomon va keraksiz yuk bo‘lib tuyulishi mumkin. Lekin, albatta, bunday emas: ular ma‘lumotlar sifatini, mavjudligini va boshqarilishini yaxshilaydi. Natijada, ma‘lumotlardan biznesga qiymat keltiradigan ma‘lumotlar mahsuloti yaxshilanadi va shuning uchun qimmatroq bo‘ladi. Yoki yangi ma‘lumot mahsulotini yaratish va yangi biznes qiymatini olish tezligi oshadi. Shuning uchun, etuklik darajasida har ikki tomon uchun mos keladigan xarajatlari va qiymat o‘rtasidagi muvozanatni topish muhimdir. Axir, agar biznes oxirgi vergulgacha tasvirlangan ushbu ajoyib, tezkor ma‘lumotlarni qanday qilib pul ishlashni tushunmasa, etuklik darajasini 5-darajaga tezlashtirishning ma‘nosi yo‘qligi aniq. Mening hamkasbim aytganidek, “do‘koningizdagi har bir atribut aniq belgilangan pulni talab qiladi - menga ulardan qanday ta‘sir ko‘rsatishingizni ko‘rsating”: va bu to‘g‘ri savol, garchi unga javob topish ba‘zan juda qiyin bo‘lishi mumkin.

Yuqoridagilardan kelib chiqqan holda, yechimni tanlashda asosiy uchta jihat o‘rtasidagi muvozanatni topishdir: ma‘lumotlarga biznes qiymatini ta‘minlaydigan joriy vazifalarni belgilash, kelajakdagi vazifalarni aniqlab olish, saqlash va qayta ishlash xarajatlari. Va bu jihatlar o‘rtasidagi muvozanatni saqlash uchun balki bir qator mutaxassislar jalb qilinishi kerak.

FOYDALANILGAN ADABIYOTLAR RO‘YHATI: (REFERENCES)

1. **Абдыкаримова А. Т.**, «Технология больших данных», Наука и жизнь Казахстана, №2(78), 2019 г., с.223-226.
2. **[Elektron havola]**. – Режим доступа:<https://freshprogs.ru/programmy/7335->

statsoft-statistica-v-100.html

3. [Elektron havola]. – Режим доступа: <http://www.jetinfo.ru/stati/bolshiedannyebolshaya-bolshaya-problema>

4. **Taylakov, Norbek Isakulovich, and Fotima Urazalievna Anarbaeva.** "Pedagogical Features Of Moodle Platform." *The American Journal of Applied sciences* 2.07 (2020): 104-107.

5. **Anarbaeva F.** Methodological properties of the moodle platform in the organization of e-learning // *IEJRD-International Multidisciplinary Journal*. – 2020. – Т. 5. – №. 4. – С. 7.

6. **А.Т. Абдыкаримова.** «Big Data: проблемы и технологии» *Технические науки* № .1. – С. 55-57.

7. **Ubaydullayev, U. S.** (2020). The inverse problem for a mixed loaded equation with the riemann-liouville operator in a rectangular domain. *Vestnik KRAUNC. Fiziko-Matematicheskie Nauki*, 31(2), 18-31.