

BIG DATA TEXNALOGIYALARI

Xurshid Ko'charov Xurram o'g'li

Toshkent davlat iqtisodiyot universiteti

Samarqand filiali magistranti

Tel: 919002369

Annotatsiya: Big data(katta ma'lumotlar) - juda katta hajmdagi bir jinsli bo'lmagan va tez tushadigan raqamli ma'lumotlar bo'lib, ularni odatiy usullar bilan qayta ishlab bo'lmaydi. Ba'zi hollarda, katta ma'lumotlar tushunchasi bilan birga shu ma'lumotlarni qayta ishlash ham tushuniladi. Asosan, analiz obyekti katta ma'lumotlar deb ataladi.

Kalit so'zlar: Big data texnologiyalari, sun'iy intellekt, axborot texnologiyalari, Blockchain texnologiyasi, dastur, loyiha, algoritmlar.

Big data atamasi 2008-yilda dunyoga kelgan. Nature jurnali muharriri Klifford Linch dunyo ma'lumotlar hajmining juda tez sur'atda o'sishiga bag'ishlangan maxsus sonida big data atamasini qo'llagan. Biroq, katta ma'lumotlar avval ham bo'lgan. Mutaxassislar fikricha, kuniga 100 gb dan ko'p ma'lumot tushadigan oqimlarga big data deb aytilar ekan. Katta ma'lumotlarni analiz qilish, inson his etish imkoniyatidan tashqarida bo'lgan qonuniylatlarni aniqlashda yordam beradi. Bu esa kundalik hayotimizdagi barcha sohalar, hukumatni boshqarish, tibbiyot, telekommunikatsiya, moliya, transport, ishlab chiqarish va boshqa sohalarni yanada yaxshilash, ularning imkoniyatlarini oshirish, muommolarga muqobil yechimlar izlab topish imkonini yaratadi. Data lake(ma'lumotlar ko'li) - qayta ishlanmagan katta ma'lumotlar ombori. "Ko'l" har xil manbalardan kelgan, har xil formatda bo'lgan ma'lumotlarni saqlaydi. Bu esa odatiy relatsion ma'lumotlar omborida ma'lumotlarni aniq struktura asosida saqlashdan ko'ra arzonroqqa tushadi. Ma'lumotlar ko'li, ma'lumotlarni boshlang'ich holatida analiz qilish imkonini beradi. Bundan tashqari, "ko'l"lardan bir vaqtni o'zida bir nechta ishchilar foydalanishlari mumkin. Data science(ma'lumotlar haqidagi fan) - analiz muommolarini, ma'lumotlarni qayta ishlash va ularni raqamli ko'rinishda taqdim etishni o'rganadigan fan. Bu atama dunyoga kelgan vaqt 1974-yil hisoblanadi. O'sha yili Daniyalik informatik, Peter Naur "A Basic Principle of Data Science" nomli kitobini chop ettirgan. 2010-yillar boshida katta ma'lumotlarni tarqalishi natijasida bu yo'nalish juda foydali va kelajagi bor biznesga aylandi. Va o'shandi katta ma'lumotlar bilan ishlaydigan mutaxassislar talab juda oshib ketdi. Data science tushunchasiga ma'lumotlar omborini loyihalash va raqamlangan ma'lumotlarni qayta ishlashning barcha metodlari kiradi. Ko'plab mutaxassislar fikricha, aynan data science big dataning biznes nuqtai nazaridan hozirgi zamonoviy

o'rindoshi hisoblanadi. Data mining (ma'lumotlarni topish) - biron qonuniyatni topish maqsadida ma'lumotlarni intellektual analiz qilishga aytiladi. Isroillik matematik Grigoriy Pyatetskiy-Shapiro 1989-yilda bu atamani fanga kiritgan. Texnologiyalar, avvalari noma'lum va foydali bo'lgan qayta ishlanmagan (hom) ma'lumotlarni topish jarayoniga data mining (ma'lumotlarni topish) deyiladi. Data mining metodlari ma'lumotlar ombori, statistika va sun'iy intellekt tutashgan nuqtada joylashadi. Machine learning (mashinali o'qitish) - o'zi o'rganadigan dasturlar yaratish amaliyoti va nazariyasi, sun'iy intellektning katta qismi. Dasturchilar o'z algoritmlariga xususiy hollarda umumiy qonuniyatlarni aniqlashni o'rgatishadi. Natijada, kompyuter, inson avvaldan ko'rsatib o'tgan buyruqlaridan emas, balki, o'z shaxsiy malakasidan kelib chiqib qaror qabul qiladi. Bunday o'qitishning juda ko'p metodlari data mining'ga oid bo'lishi mumkin. Mashinali o'qitishga birinchi tarifni 1959-yilda amerikalik informatik Artur Samuel bergan. U sun'iy intellekt elementlariga ega bo'lgan shashka o'yini, dunyoda birinchi o'zi o'rganadigan dasturni yaratgan. Deep learning (chuqur o'qitish) - yanada murakkab va yanada mustaqil bo'lgan o'zi o'qidigan dasturlar yaratadigan mashinali o'qitish turi. Oddiy mashinali o'qitish hollarida boshqariladigan malaka yordamida kompyuter bilimlarni aniqlab oladi: dasturchi algoritmgaga ma'lum misollarni ko'rsatadi, xatolarni qo'lda to'g'rilaydi. Deep learningda esa, tizim o'zi o'z funksiyalarini loyihalaydi, ko'p darajali hisob-kitoblar amalga oshiradi va atrof-muhit haqida xulosalar qiladi. Odatda chuqur o'qitishni neyron tarmoqlarga tadbiiq qilishadi. Bu texnologiya asosan rasmlarni qayta ishlashda, nutqni tanishda, neyromashina tarjima, farmatsevtikadagi hisoblashlarda va boshqa zamonaviy texnologiyalarda qo'llaniladi. Asosan Google, facebook va Baidu tomonidan loyihalarga tadbiiq qilinadi. Sun'iy neyron tarmog'i - oddiy protsessorlar (sun'iy neyronlar) birlashtirilgan tizimi bo'lib, insonning nerv tizimini imitatsiya qiladi. Bunday struktura evaziga, neyron tarmoqlari dasturlanmaydi, ular o'qitishadi. Huddi haqiqiy neyronlar kabi, protsessorlar signallarni oddiygina qabul qilishadi va boshqa protsessorlarga o'zatishadi. Shu bilan birga, boshqa butun tizim algoritmlar bajara olmaydigan murakkab topshiriqlarni bajaradi. 1943-yilda amerikalik olimlar Uorren Makkalok va Uolter Pittslar sun'iy neyron tarmog'i tushunchasini fanga kiritishgan. Business intelligence (biznes-analitika) - aniq strukturaga ega bo'lmagan juda katta hajmdagi ma'lumotlarni qayta ishlash natijasida muqobil biznes yechimlar izlashga aytiladi. Effektiv biznes-analitika ichki va tashqi ma'lumotlarni analiz qiladi - ham bozor axborotlarini, ham mijoz-kompaniyaning hisobotlarini hisobga oladi. Bu biznesni butunlay tushunishga yordam beradi, shu bilan birga, strategik va operatsion qarorlar qabul qilishga zamin yaratadi (mahsulot narxini aniqlashda, kompaniya rivojlanishining asosiy yo'nalishlarini belgilab olishda).

Bu atama 1958-yilda IBM taqdiqotchisi Xans Piter Lun maqolasida birinchi bo'lib ko'rsatilgan. 1996-yilda axborot texnologiyalari bozorini o'rganishga asoslangan

Gartner analitik agentligi business intelligence tarkibiga data mining metodikasini ham qo'shgan. Katta ma'lumotlarga misollar. Big Data texnologiyalarini bozorga va zamonaviy hayotga faol joriy etish, dunyoning taniqli kompaniyalari dunyoning deyarli har bir burchagida o'z mijozlariga ega bo'lganlaridan foydalanishni boshlaganidan keyin boshlandi.

Big Data nima?

Aslida, atamaning ta'rifi yuzaga chiqadi: "katta ma'lumotlar" juda katta hajmdagi ma'lumotlarni boshqarish va ularni tahlil qilishni anglatadi. Agar siz kengroq qarasangiz, bu juda katta hajm tufayli klassik usullar bilan qayta ishlanmaydigan ma'lumotlar. Big Data atamasi o'zi nisbatan yaqinda paydo bo'lgan. Google Trends ma'lumotlariga ko'ra, ushbu mashhurlikning faol o'sishi 2011 yil oxiriga to'g'ri keladi: 2010 yilda katta ma'lumotlarni qayta ishlash bilan bevosita bog'liq bo'lgan birinchi mahsulotlar va yechimlar paydo bo'ldi. 2011 yilga kelib, eng yirik IT-kompaniyalar, shu jumladan IBM, Oracle, Microsoft va Hewlett-Packard, o'z biznes strategiyalarida Big Data atamasidan faol foydalanmoqdalar. Asta-sekin, axborot texnologiyalari bozori tahlilchilari ushbu kontseptsiya bo'yicha faol izlanishlarni boshlaydilar.[1]

Hozirgi vaqtda ushbu atama katta shuhrat qozondi va turli sohalarda faol qo'llanilmoqda. Ammo Katta Ma'lumotlar bu mutlaqo yangi hodisa, deb aniq aytish mumkin emas, aksincha, 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;

- Tahlil vositalari turli xil yangi ma'lumotlar manbalari bilan to'ldirildi, bu raqamli texnologiyalarga keng o'tish bilan bog'liq, shuningdek ma'lumotlarni yig'ish va o'lchashning yangi usullari.

Tadqiqotchilar Big Data texnologiyalaridan ishlab chiqarish, sog'liqni saqlash, savdo, davlat boshqaruvi va boshqa turli sohalar va sohalarda eng faol foydalanilishini taxmin qilishmoqda.

Big Data bu ma'lum bir ma'lumotlar to'plami emas, balki ularni qayta ishlash usullari to'plami. Katta ma'lumotlar uchun aniqlovchi xususiyat nafaqat uning hajmi, balki ma'lumotlarni qayta ishlash va tahlil qilishning og'ir jarayonlarini tavsiflovchi boshqa toifalardir.

Qayta ishlash uchun dastlabki ma'lumotlar bo'lishi mumkin, masalan:

- Internet foydalanuvchilari xulq-atvori jurnallari;
- narsalar Interneti;
- ijtimoiy media;
- meteorologik ma'lumotlar;
- yirik kutubxonalarning raqamli kitoblari;
- transport vositalarining GPS signallari;
- bank mijozlarining operatsiyalari to'g'risidagi ma'lumotlar;[2]
- uyali aloqa abonentlarining joylashuvi to'g'risidagi ma'lumotlar;
- yirik chakana savdo tarmoqlarida xaridlar to'g'risida ma'lumot va boshqalar.

Big Data asosiy tamoyillari:

- Landshaft kengaytirilishi - ma'lumotlar massivlari juda katta bo'lishi mumkin va bu katta ma'lumotlarni qayta ishlash tizimining dinamik ravishda kengayib borishi kerakligini anglatadi.

- Xatolarga bardoshlik - uskunaning ba'zi elementlari ishdan chiqsa ham, butun tizim ishlashi kerak.

- ma'lumotlarning joylashuvi. Katta taqsimlangan tizimlarda ma'lumotlar odatda juda ko'p sonli mashinalarda tarqatiladi. Ammo, iloji boricha va resurslarni tejash maqsadida ma'lumotlar ko'pincha o'sha serverda saqlanadi.

Uchala tamoyilning barqaror ishlashi va shunga mos ravishda katta ma'lumotlarni saqlash va qayta ishlashning yuqori samaradorligi uchun, masalan, blockchain kabi yangi katta texnologiyalar kerak. Shunday qilib, Big Data texnologiyalarini amaliy qo'llash marketing sohasida yotadi. Internetning rivojlanishi va barcha turdagi aloqa vositalarining tarqalishi tufayli xatti-harakatlar ma'lumotlari (masalan, qo'ng'iroqlar soni, xarid qilish odatlari va xaridlar) real vaqtda mavjud bo'ladi. Katta ma'lumot texnologiyalaridan moliya, sotsiologik tadqiqotlar va boshqa ko'plab sohalarda samarali foydalanish mumkin. Mutaxassislarining ta'kidlashicha, katta ma'lumotlardan foydalanishning barcha imkoniyatlari bu aysbergning ko'rinadigan qismidir, chunki bu texnologiyalar katta hajmlarda razvedka va kontr-razvedkada, harbiy ishlarda, shuningdek, axborot urushi deb ataladigan barcha narsalarda qo'llaniladi. Umuman olganda, Big Data bilan ishlashning ketma-ketligi ma'lumotlarni to'plash, hisobotlar va boshqaruv paneli yordamida olingan ma'lumotlarni tuzish, shuningdek harakatlar bo'yicha tavsiyalarni shakllantirishdan iborat.

Marketingda Big Data texnologiyalaridan foydalanish imkoniyatlarini qisqacha ko'rib chiqing. Ma'lumki, marketolog uchun ma'lumot bashorat qilish va strategiyani tuzishning asosiy vositasidir. Ma'lumotlarning katta tahlili uzoq vaqtdan beri maqsadli auditoriya, qiziqish, talab va iste'molchilarning faolligini aniqlash uchun muvaffaqiyatli ishlatilgan. Katta ma'lumotlarni tahlil qilish, xususan, reklama (RTB

kim oshdi savdosi modeli asosida - Real vaqt savdosi) faqat mahsulot yoki xizmatga qiziqqan iste'molchilarga namoyish qilish imkonini beradi.

Marketingda Big Data-dan foydalanish ishbilarmonlarga:

- iste'molchilar bilan tanishish, Internetda shunga o'xshash auditoriyani jalb qilish yaxshiroqdir;

- mijozlarning qoniqishini baholash;

- taklif etilayotgan xizmat istiqbol va ehtiyojlarga javob berishini tushunish;

- mijozlar ishonchini oshiradigan yangi usullarni topish va amalga oshirish;

- talabga javob beradigan loyihalarni yaratish va boshqalar.

Blokchain texnologiyasi va Big Data o'zaro ta'sirining istiqbollari. Big Data bilan integratsiya sinergistik samara beradi va biznes uchun ko'plab yangi imkoniyatlarni ochadi, shu jumladan:

- iste'molchilar imtiyozlari haqida batafsil ma'lumotga ega bo'lish, ular asosida aniq etkazib beruvchilar, mahsulotlar va mahsulot tarkibiy qismlari uchun batafsil tahliliy profillarni yaratish mumkin;

- foydalanuvchilarning turli toifalari bo'yicha ayrim tovarlar guruhlari bo'yicha tranzaktsiyalar va iste'mol statistikasi to'g'risidagi batafsil ma'lumotlarni birlashtirish;

- yetkazib berish va iste'mol zanjirlari, tashish paytida mahsulotni yo'qotish bo'yicha batafsil tahliliy ma'lumotlarni olish (masalan, ayrim tovarlarning qurishi va bug'lanishi natijasida vazn yo'qotish);

- mahsulot firibgarligiga qarshi kurashish, pul yuvish va firibgarlikka qarshi kurash samaradorligini oshirish va boshqalar.[3]

Tovarlarni iste'mol qilish va iste'mol qilish bo'yicha batafsil ma'lumotlarga kirish Big Business texnologiyasining asosiy biznes jarayonlarini optimallashtirish imkoniyatlarini sezilarli darajada ochib beradi, tartibga soluvchi xatarlarni kamaytiradi va mavjud iste'molchilar talablariga eng mos keladigan mahsulotlarni yaratish va yaratish uchun yangi imkoniyatlarni ochib beradi. Ma'lumki, yirik moliya institutlari vakillari, shu jumladan va boshqalar, blockchain texnologiyasiga katta qiziqish bildirmoqdalar. Shveysariyaning UBS moliyaviy xoldingining IT-menejeri Oliver Bussmannning so'zlariga ko'ra blockchain texnologiyasi "tranzaktsiyalarni qayta ishlash vaqtini bir necha kundan bir necha daqiqagacha qisqartirishga" qodir. Big Data texnologiyasidan foydalangan holda blockchain tahlilining imkoniyatlari juda katta. Taqsimlangan ro'yxatga olish texnologiyasi ma'lumotlarning yaxlitligini, shuningdek tranzaktsiyalar tarixini ishonchli va oshkora saqlashni ta'minlaydi. Big Data, o'z navbatida, samarali tahlil qilish, prognozlash, iqtisodiy modellashtirish uchun yangi vositalarni taqdim etadi va shunga muvofiq boshqaruvni yanada xabardor qilish uchun yangi imkoniyatlarni ochib beradi. Sog'liqni saqlashda blockchain va Big Data tandemidan muvaffaqiyatli foydalanish mumkin. Ma'lumki, bemorning sog'lig'i to'g'risida to'liq va to'liq bo'lmagan ma'lumotlar ba'zida noto'g'ri tashxis qo'yish va

noto'g'ri davolanish xavfini oshiradi. Tibbiyot muassasalari mijozlarining sog'lig'i to'g'risida tanqidiy ma'lumotlar iloji boricha xavfsiz bo'lishi kerak, o'zgaras xususiyatlarga ega, tekshirilishi kerak va hech qanday manipulyatsiya qilinmasligi kerak. Blokchindagi ma'lumotlar yuqoridagi barcha talablarga javob beradi va Big Data yangi texnologiyalaridan foydalangan holda chuqur tahlil qilish uchun yuqori sifatli va ishonchli manba sifatida xizmat qilishi mumkin. Bundan tashqari, blockchain yordamida tibbiy muassasalar tibbiy ma'lumotlarga muhtoj bo'lgan sug'urta kompaniyalari, adliya organlari, ish beruvchilar, ilmiy muassasalar va boshqa tashkilotlar bilan ishonchli ma'lumotlar almashishlari mumkin.

Katta ma'lumotlar va axborot xavfsizligi.[4]

Keng ma'noda, axborot xavfsizligi bu tabiiy yoki sun'iy tabiatning tasodifiy yoki qasddan salbiy ta'siridan axborot xavfsizligi va yordamchi infratuzilma.

Axborot xavfsizligi sohasida Big Data quyidagi muammolarga duch keladi:

- ma'lumotlarni himoya qilish va ularning yaxlitligini ta'minlash muammolari;
- maxfiy ma'lumotlarni buzish va yashirinish xavfi;
- maxfiy ma'lumotlarni noto'g'ri saqlash;
- ma'lumotni yo'qotish xavfi, masalan, boshqa bironing zararli harakatlari tufayli;
- shaxsiy ma'lumotlardan uchinchi shaxslar tomonidan noto'g'ri foydalanish xavfi va boshqalar.

Blokchain yechishga mo'ljallangan katta ma'lumotlarning asosiy muammolaridan biri, axborot xavfsizligi sohasida. Uning barcha asosiy tamoyillariga rioya qilinishini ta'minlash, taqsimlangan ro'yxatga olish texnologiyasi ma'lumotlarning yaxlitligi va ishonchliligini kafolatlashi mumkin va bitta nosozlik nuqtasi yo'qligi sababli blockchain axborot tizimlarini barqaror qiladi. Tarqalgan ro'yxatga olish texnologiyasi ma'lumotlarga bo'lgan ishonch muammosini hal qilishga yordam beradi, shuningdek, ularni umumiy ravishda baham ko'rish imkoniyatini beradi.[5]

Xulosa:

Axborot muhim ahamiyatga ega, demak, axborot xavfsizligining asosiy jihatlarini ta'minlash masalasi birinchi o'rinda turadi. Raqobatda omon qolish uchun kompaniyalar vaqtni bosib o'tishlari kerak, ya'ni ularni blockchain texnologiyasi va Big Data vositalari mavjud bo'lgan imkoniyatlar va afzalliklarga e'tibor bermaslik kerak. Faqat dangasa Big ma'lumotlar haqida gapirmaydi, lekin u nima ekanligini va qanday ishlashini deyarli tushunmaydi. Eng sodda - terminologiyadan boshlaylik. Rus tilida gapirsangiz, Katta ma'lumotlar bu har xil vositalar, yondashuvlar va tarkibiy tuzilmalar va ishlov berilmagan ma'lumotlardan ularni aniq vazifa va maqsadlarda foydalanish uchun ishlatishdir. Tuzilmaydigan ma'lumotlar - bu oldindan belgilangan tuzilishga ega bo'lmagan yoki ma'lum tartibda tashkil etilmagan ma'lumotlar. "Katta ma'lumotlar" atamasi 2008 yilda Tabiat muharriri Clifford Lynch tomonidan dunyo

miqyosidagi ma'lumotlar hajmining keskin o'sishiga bag'ishlangan maxsus sonida kiritilgan. Albatta, katta ma'lumotlarning o'zi oldin ham mavjud edi. Mutaxassislarning fikriga ko'ra, kuniga 100 Gb dan ortiq ma'lumotlar oqimining katta qismi Katta ma'lumotlar toifasiga kiradi.

Foydalanilgan adabiyotlar:

1. Kovalev, I., Nezhmetdinov, R., & Kvashnin, D. (2019). Big data analytics of the technological equipment based on Data Lake architecture. MATEC Web of Conferences. <https://doi.org/10.1051/matecconf/201929800079>[1]
2. <https://www.guru99.com/big-data-tools.html>[2]
3. Hussain, T., Sanga, A., & Mongia, S. (2019). Big Data Hadoop Tools and Technologies: A Review. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.3462554>[3]
4. Alshammari, H., Lee, J., & Bajwa, H. (2018). H2hadoop: Improving hadoop performance using the metadata of related jobs. IEEE Transactions on Cloud Computing. <https://doi.org/10.1109/TCC.2016.2535261>[4]
5. Gupta, S., & Narsimha, G. (2017). Efficient query analysis and performance evaluation of the Nosql data store for bigdata. Advances in Intelligent Systems and Computing. https://doi.org/10.1007/978-981-10-2471-9_53[5]