

9-mavzu. Qishloq xo'jaligida katta ma'lumotlar (Big Data)

Reja:

- 1. Big data haqida ma'lumotlar;**
- 2. Big data qanchalik katta;**
- 3. Big data texnologiyasi va undagi ma'lumotlarni tahlil etish usullari;**
- 4. Big dataning asosiy sakkizta atamasi;**

Kalit so'zlar: big data- katta hajmli ma'lumot, descriptive analytics- tavsiflovchi tahlil, predictive analytics-bashoratli tahlil, prescriptive analytics- retseptiv tahlil, diagnostic analytics- tashxisli tahlil, data lake-ma'lumotlar ko'li, data science-ma'lumotlar haqidagi fan, deep learning-chuqur o'qitish.

1. Big data haqida ma'lumotlar

Katta hajmli ma'lumot (inglizcha: Big data) — katta hajmdagi ma'lumotlarga nisbatan qo'llanadigan termin hisoblanadi(odatda terabayt, ekzabayt va petabaytlar darajasida). Ushbu ma'lumotlarni amaldagi va yangi yaratilayotgan texnologiyalar yordamida toplash, taqsimlash, o'zgartirishlar kiritish, tahlil qilish, saqlash va vizual shaklda tasvirlash imkonи bo'lishi kerak. Ma'lumotlar miqdorining tez o'sishiga so'nggi yillarda kuzatilayotgan texnologik taraqqiyot va buning natijasi hisoblanmish mashina tomonidan yaratilayotgan ma'lumotlarning ko'payib borishi sababchidir. RFID-texnologiyasining keng tarqalishi, elektron transaksiyalarning moliya sohasida keng qo'llanishi, ilmiy-tadqiqot muassasalarining toplagan natijalari, web-texnologiyalarining keng tarqalishi, telekommunikasion tizimlardagi protokollar shular jumlasidandir. Big Data aslida ilmiy-tadqiqotga aloqador sohalarda avvaldan ham mavjud bo'lgan bo'lsa-da, lekin so'nggi yillardagina ushbu fenomen ko'proq tilga olinadigan bo'ldi. Buning sababi shundaki, katta miqdordagi ma'lumotlarni tahlil qilish endi iqtisodiy tashkilotlar tomonidan ham keng qo'llanila boshlanishi va buning oqibatida ma'lumotlar tahlilining raqobatbardoshlikni oshirish va samaradorlikni kuchaytirish kabi muhim masalasarda asosiy o'rinni egallaganidir. Big Data odatiy Bosh Informasion Mutaxassis (Chief Information Officer - CIO) uchun qo'shimcha vazifalar va talablar yaratishdan tashqari, yangi Ma'lumot Boshqaruvchisi (Data Steward) va Ma'lumot bo'yicha mutaxassis (Data Scientist) kabi kasblarning paydo bo'lishiga olib keldi.

"Katta ma'lumotlar" bizning davrimizning eng ko'p ishlataladigan shov-shuvli so'zlaridan biridir, ammo bu aslida nimani anglatadi?

Bu erda katta ma'lumotlarning tez va oddiy ta'rifi. Katta ma'lumot an'anaviy ma'lumotlarni qayta ishlash va saqlash usullari bilan ishlov berish uchun juda katta va murakkab ma'lumotlar. Bu siz evristik sifatida foydalanishingiz mumkin bo'lgan tezkor ta'rif bo'lsa-da, katta ma'lumotlarni chuqurroq va to'liqroq tushunish foydali

bo'ladi. Keling, saqlash, tuzilish va qayta ishlash kabi katta ma'lumotlar asosidagi ba'zi tushunchalarni ko'rib chiqaylik.

2. Big data qanchalik katta

Bu "X" o'lchamidan ortiq har qanday ma'lumot katta ma'lumotdir" degandek oddiy emas, ma'lumotlar bilan ishlanayotgan muhit juda muhim omil hisoblanadi. katta ma'lumotlarga nimalar tegishli ekanligini aniqlash. Katta ma'lumotlar deb hisoblanishi uchun ma'lumotlarning hajmi kontekstga yoki ma'lumotlardan foydalanilayotgan vazifaga bog'liq. Turli xil o'lchamdagি ikkita ma'lumotlar to'plamini turli kontekstlarda "katta ma'lumotlar" deb hisoblash mumkin.

Aniqroq qilib aytadigan bo'lsak, agar siz 200 megabaytli faylni elektron pochta ilovasi sifatida yuborishga harakat qilsangiz, buni amalga oshira olmaysiz. Shu nuqtai nazardan, 200 megabaytlik faylni katta ma'lumotlar deb hisoblash mumkin. Bundan farqli o'laroq, 200 megabaytlik faylni bitta LAN ichidagi boshqa qurilmaga nusxalash umuman vaqt talab qilmasligi mumkin va bu kontekstda u katta ma'lumot sifatida qaralmaydi.

Biroq, faraz qilaylik, 15 terabaytlik videoni kompyuterda ko'rish dasturlarini o'qitishda foydalanish uchun oldindan ishlov berish kerak. Bunday holda, video fayllar shunchalik ko'p joy egallaydiki, hatto kuchli kompyuter ham ularning barchasini qayta ishlash uchun uzoq vaqt talab etadi va shuning uchun ishlov berish vaqtini qisqartirish uchun odatda bir-biriga ulangan bir nechta kompyuterlar bo'ylab taqsimlanadi. Ushbu 15 terabaytlik video ma'lumotlari, albatta, katta ma'lumotlar sifatida tasniflanadi.

Katta ma'lumotlar tuzilmalarining turlari

Katta ma'lumotlar strukturaning uch xil toifasida bo'ladi: tuzilmagan ma'lumotlar, yarim tizimli va tuzilgan ma'lumotlar.

Strukturalanmagan ma'lumotlar - bu aniq tuzilmaga ega bo'limgan ma'lumotlar, ya'ni ma'lumotlar asosan bitta katta hovuzda joylashgan. Tarkibi bo'limgan ma'lumotlarga misol sifatida etiketlanmagan tasvirlar bilan to'la ma'lumotlar bazasi bo'lishi mumkin.

Yarim tuzilgan ma'lumotlar - bu rasmiy tuzilishga ega bo'limgan, ammo bo'sh tuzilmada mavjud bo'lgan ma'lumotlar. Masalan, elektron pochta ma'lumotlari yarim tizimli ma'lumotlar sifatida hisoblanishi mumkin, chunki siz alohida elektron pochta xabarlarida mavjud bo'lgan ma'lumotlarga murojaat qilishingiz mumkin, ammo rasmiy ma'lumotlar namunalari o'rnatilmagan.

Strukturaviy ma'lumotlar - bu rasmiy tuzilishga ega bo'lgan, ma'lumotlar nuqtalari turli xususiyatlar bo'yicha tasniflangan ma'lumotlar. Strukturaviy

ma'lumotlarga misollardan biri ismlar, elektron pochtalar, telefon raqamlari va veb-saytlar kabi aloqa ma'lumotlarini o'z ichiga olgan Excel elektron jadvalidir.

Agar siz ushbu ma'lumotlar turlarining farqlari haqida ko'proq o'qishni istasangiz, bu yerdagi havolani tekshiring.

Katta ma'lumotlarni uch xil ko'rsatkich bo'yicha tahlil qilish mumkin: hajm, tezlik va xilma-xillik.

Hajmi ma'lumotlar hajmini bildiradi. Ma'lumotlar to'plamlarining o'rtacha hajmi ko'pincha ortib bormoqda. Misol uchun, 2006 yilda eng katta qattiq disk 750 Gb qattiq disk edi. Bundan farqli o'laroq, Facebook bir kunda 500 terabaytdan ortiq ma'lumot ishlab chiqaradi va bugungi kunda mavjud bo'lgan eng katta iste'molchi qattiq disk 16 terabayt qattiq disk hisoblanadi. Bir davrda katta ma'lumot sifatida ko'rsatilgan narsa boshqa davrda katta ma'lumot bo'lmasligi mumkin. Bugungi kunda ko'proq ma'lumotlar yaratilmoqda, chunki bizni o'rab turgan ko'proq ob'ektlar sensorlar, kameralar, mikrofonlar va boshqa ma'lumotlarni yig'ish moslamalari bilan jihozlangan.

Tezlik ma'lumotlarning qanchalik tez harakatlanishini yoki boshqacha qilib aytganda, ma'lum vaqt oralig'ida qancha ma'lumot ishlab chiqarilishini anglatadi. Ijtimoiy media oqimlari har daqiqada yuz minglab xabarlar va sharhlarni yaratadi, shu bilan birga shaxsiy elektron pochta qutingizda faollik kamroq bo'lishi mumkin. Katta ma'lumotlar oqimlari ko'pincha real vaqt rejimida yuz minglab yoki millionlab voqealarni boshqaradigan oqimlardir. Ushbu ma'lumotlar oqimlariga misollar - onlayn o'yin platformalari va yuqori chastotali birja savdosi algoritmlari.

Varete ma'lumotlar to'plamida mavjud bo'lgan har xil turdag'i ma'lumotlarni anglatadi. Ma'lumotlar audio, video, matn, fotosuratlar yoki seriya raqamlari kabi juda ko'p turli formatlardan iborat bo'lishi mumkin. Umuman olganda, an'anaviy ma'lumotlar bazalari bir yoki bir nechta ma'lumotlar turini qayta ishlash uchun formatlangan. Boshqacha qilib aytadigan bo'lsak, an'anaviy ma'lumotlar bazalari juda bir xil va izchil, bashorat qilinadigan tuzilishga ega bo'lgan ma'lumotlarni saqlash uchun tuzilgan. Ilovalar xilma-xil bo'lib, turli funktsiyalarga to'la va ko'proq odamlar tomonidan foydalanilar ekan, ma'lumotlar bazalari ko'proq turdag'i ma'lumotlarni saqlash uchun rivojlanishiga to'g'ri keldi. Strukturalanmagan ma'lumotlar bazalari katta ma'lumotlarni saqlash uchun ideadir, chunki ular bir-biriga bog'liq bo'limgan bir nechta ma'lumotlar turlarini saqlashi mumkin.

3. Big data texnologiyasi va undagi ma'lumotlarni tahlil etish usullari;

Katta ma'lumotlarni tahlil qilishni osonlashtirish uchun mo'ljallangan turli xil platformalar va vositalar mavjud. Ma'lumotlardan mazmunli naqshlarni olish uchun katta ma'lumotlar hovuzlarini tahlil qilish kerak, bu an'anaviy ma'lumotlarni tahlil qilish vositalari bilan juda qiyin bo'lishi mumkin. Katta hajmdagi ma'lumotlarni

tahlil qilish vositalariga bo'lgan ehtiyojga javoban, turli kompaniyalar katta ma'lumotlarni tahlil qilish vositalarini yaratdilar. Katta ma'lumotlarni tahlil qilish vositalari ZOHO Analytics, Cloudera va Microsoft BI kabi tizimlarni o'z ichiga oladi.

Big Data texnologiyasi va undagi ma'lumotlarni tahlil etish usullari

«Big Data» atamasi «katta ma'lumotlar» degan ma'noni bildirib, uni birinchi marta 2008-yilda «Nature» jurnalining muharriri Klifford Linch ishlatgan. U dunyoda axborot hajmlarining jadal sur'atlarda o'sishi haqida gapirib, faqatgina yangi vositalar va yanada ilg'or texnologiyalar ularni o'zlashtirishga yordam berishini ta'kidlagan.

Linch «katta ma'lumotlar»ga sutkasiga 150 Gb hajmda bo'lgan har qanday ko'rinishdagi ma'lumotlar tegishli deb hisoblaydi, ammo hanuzgacha ma'lumotlarni yagona o'lchov birligi mavjud emas. Dunyoda raqamlangan ma'lumotlar hajmi tinimsiz o'sib bormoqda.

IBC kompaniyasining ma'lumotlariga qaraganda, 2003 yilda 5 eksabayt (1 eksabayt = 1 milliard gigabayt) ma'lumot yig'ilgan. 2015-yil may oyiga qadar dunyoda yig'ilgan ma'lumotlar hajmi 6,5 zettabaytdan oshgan bo'lsa, 2020-yilda insoniyat 40-44 zettabayt raqamli ma'lumot hosil qilgan.

«Big Data» – bu juda tez sur'atlarda o'sib borayotgan katta hajmdagi tizimli va tizimsiz raqamli ma'lumotlar to'plami. «Big Data» biznes jarayonlarini optimallashtirish va avtomatlashtirish, to'plangan ma'lumotlarga asoslangan eng samarali qarorlarni qabul qilinishini ta'minlash maqsadida axborotni saqlash va qayta ishslashning innovatsion usullarini o'z ichiga oladi.

«Katta ma'lumotlar» tushunchasi o'zi nisbatan yangi bo'lsa-da, aslida uning kelib chiqishi birinchi ma'lumotlar markazlari va relyasion ma'lumotlar bazalarini ishlab chiqish boshlangan 1960-1970-yillarga to'g'ri keladi.

2005-yilda odamlar Facebook, YouTube va boshqa onlayn xizmatlar foydalanuvchilari qancha ma'lumot ishlab chiqarayotganini tushuna boshladilar. Shu yili Hadoop (katta ma'lumotlarni saqlash va tahlil qilish uchun maxsus yaratilgan ochiq kodli platforma) ishlab chiqildi.

2011-yilga qadar katta ma'lumotlar tahlili faqat ilmiy-statistik tadqiqotlar doirasida amalga oshirilgan, ammo 2012-yilning boshiga kelib ma'lumotlar hajmining katta ko'lamda o'sishi sababli ularni tizimlashtirish va amalda qo'llash ehtiyoji tug'ildi.

2014-yildan boshlab, katta ma'lumotlarga amaliy muhandislik va IT mutaxassislar tayyorlaydigan dunyoning yetakchi oliy o'quv yurtlari ye'tibor qaratishdi. Keyin Microsoft, IBM, Oracle, YEMC, undan keyin esa Google, Apple, Facebook va Amazon kabi yirik korporatsiyalar qo'shildi. Bugungi kunda katta ma'lumotlar yirik kompaniyalar va davlat idoralari tomonidan turli sohalarda ishlatilmoqda.

Barcha yig‘ilgan Big Data ma’lumotlari olingan manbalarga qarab tasniflanishi mumkin. AQShning Gartner kompaniyasi «katta ma’lumotlar»ga kiritilgan ma’lumotlarning quyidagi tasnifini taklif qiladi:

1) Operatsion ma’lumotlar. Bu onlayn tranzaksiyalarni qayta ishlash jarayonida mavjud bo‘lgan yoki onlayn tahliliy ma’lumotlar bazasidan olingan mijozlar, yetkazib beruvchilar, hamkorlar va xodimlar to‘g‘risidagi ma’lumotlar. Odatda tranzaksiya ma’lumotlari, aloqa ma’lumotlari va jismoniy shaxslar haqidagi umumiyligi ma’lumotlarni o‘z ichiga oladi. Ko‘pincha bunday ma’lumotlar qo‘sishimcha sensorlar va korxona jarayonlarini monitoring qilish yordamida muvaffaqiyatli to‘planadi. Masalan, bank tizimiga ulangan kassa apparatlari, aqli hisoblagichlar, ovozli aloqa, radiochastotani identifikatsiyalash va boshqalar;

2) «Qorong‘i» ma’lumotlar. Tashkilotlar tomonidan ataylab saqlanmaydigan yoki to‘planmagan, lekin biznesni yuritish yoki tarmoq xizmatlari bilan ishlash jarayonida tasodifiy (yo‘lda) yaratilgan va Internet arxivlarida qoladigan ma’lumotlar. Bunday ma’lumotlar ommaga ochiq va qisman tahlil qilish uchun tuzilgan, jumladan, elektron pochta xabarlari, elektron shartnomalar, hujjatlar, multimedia, tizim jurnallari va boshqalar;

3) Tijorat ma’lumotlari. Katta ma’lumotlar texnologiyasi imkoniyatlari paydo bo‘lishidan oldin, turli sohalarda qimmatli tijorat ma’lumotlarining aggregatorlari mavjud edi. Ushbu aggregatorlar obuna ma’lumotlarining o‘z kataloglariga to‘liq kirishni ta’minlaydi. Biroq zamonaviy bozor munosabatlarining potensial investorlar va mijozlar uchun ma’lumotni ochishga yo‘naltirilganligini hisobga olgan holda, tijorat manfaatlariga oid ko‘plab ma’lumotlar raqamli muhitda ochiq joylashtiriladi;

4) Rasmiy ma’lumotlar. Davlat organlari tomonidan tarqatiladigan ma’lumotlar (bayonetlar, press-relizlar, ob-havo prognozlari, munitsipal rivojlanish rejalarini to‘g‘risidagi ma’lumotlar), ochiq davlat registrlari, e’lon qilingan normativ-huquqiy hujjatlar (shu jumladan, ularning loyihalari) eng ishonchli va eng ko‘p tuzilgan hisoblanadi;

5) Ijtimoiy tarmoqlar va xizmatlardan olingan ma’lumotlar. Yirik ijtimoiy tarmoqlarning (Facebook, VKontakte, LinkedIn, Twitter, Instagram va boshqalar) funksiyalari, biznes, bozor munosabatlarida jismoniy shaxslarning ishtiroki, yangi va istiqbolli mahsulotlar, xizmatlar, kompaniyalar haqida yana bir ma’lumot manbasini yaratdi. Xabarlar, sharhlar, takliflar («repostlar») maqsadli mijozlarni, tijorat imkoniyatlarini, raqobatbardosh munosabatlarni, biznes va potensial hamkorlarni aniqlash hamda prognoz qilish uchun faol foydalilanildi.

Big Data tahlil qilish uchun zarur bo‘lgan ma’lumotlarni tanlaydi va tizimlaydi, katta ma’lumotlardan ilg‘or tahliliy vazifalarni bajarish uchun sun’iy intellektdan

foydalananadi. Ko‘proq katta ma’lumotlar kengaytirilgan tahliliy vazifalar uchun qo‘llaniladi, shu jumladan, sun’iy aql ham.

Tahlil qilishning to‘rtta asosiy usuli mavjud:

Tavsiflovchi tahlil (descriptive analytics) – eng keng tarqalgan usul bo‘lib, u «nima bo‘ldi?» degan savolga javob beradi hamda real vaqtida kelayotgan va tarixiy ma’lumotlarni tahlil qiladi. Asosiy maqsad ma’lum bir sohadagi muvaffaqiyat yoki muvaffaqiyatsizlik sabablari va uning muntazamligini aniqlash hamda ushbu ma’lumotlardan samarali foydalanish.

Bashoratli tahlil (predictive analytics) – mavjud ma’lumotlarga asoslanib, voqealarning eng ehtimoliy rivojlanishini bashorat qilishga yordam beradi. Buning uchun shunga o‘xhash xususiyatlar to‘plamiga ega bo‘lgan har qanday ob’yekt yoki hodisalarga asoslangan tayyor shablonlardan foydalanadi.

Retseptiv tahlil (prescriptive analytics) – bashoratli tahlil bilan solishtirganda undan keyingi daraja. Big Data va zamonaviy texnologiyalar yordamida biznesdagi yoki boshqa faoliyatdagi muammoli nuqtalarni aniqlash va kelajakda qaysi yo‘l bilan ulardan qochish mumkinligini hisoblash mumkin.

Tashxisli tahlil (diagnostic analytics) – sodir bo‘lgan voqealarning sabablarini tahlil qilish uchun ma’lumotlardan foydalanadi. Bu hodisalar va harakatlar o‘rtasidagi anomaliyalar va tasodifiy aloqalarni aniqlashga yordam beradi.

Ma’lumotlar yangi boylik manbasiga aylangani sayin, katta ma’lumotlar yechimlari kompaniyalarning o‘sishida tobora muhim rol o‘ynashi kutilmoqda. Kompaniyalar har doim ma’lumotlarga kirish huquqiga ega bo‘lgan, ammo mazmunli ma’lumotlarni olish uchun ma’lumotlarga kirish va qayta ishslash imkoniyati cheklangan. «Katta ma’lumotlar» yechimlari kompaniyalarga ushbu qiyinchiliklarni yengish imkonini beradi.

«Katta ma’lumotlar» tahlili kompaniyalarning asosini tashkil qilib, real vaqtida katta ma’lumotlar to‘plamini boshqarish, qayta ishslash, optimallashtirishga yordam beradi va qaror qabul qilish qobiliyatini yaxshilaydi. Bundan tashqari, katta ma’lumotlar va biznes-tahlilning asosiy maqsadi kompaniyalarga o‘z mijozlarini yaxshiroq tushunishga yordam berish va marketing kampaniyalarini yaxshilashdir.

Ba’zi tarmoqlar katta ma’lumotlardan foydalanishda katta yutuqlarga erishgan bo‘lsa-da, boshqalari hali ham katta ma’lumotlarga yetarli e’tibor qaratishgani yo‘q. 2025 yilga kelib bank sohasidagi katta ma’lumotlar tahlil bozori 62,10 mldr dollargacha o‘sishi mumkin. Bank sohasidagi katta ma’lumotlarning statistikasiga ko‘ra, jahon bank sektori infratuzilmasi allaqachon katta ma’lumotlarni tahlil qilishni o‘z ichiga oladi.

Masalan, 2013-yil holatiga ko‘ra, jahon moliyaviy sektorning 64 foizi allaqachon katta ma’lumotlarni o‘z infratuzilmasiga kiritgan bo‘lib, 2015-yilda bozor

hajmi 12 mld dollarga yetgan. 2019-yilga kelib katta ma'lumotlarning bank operatsiyalari tahlili bozori 29,87 mld dollarni tashkil etdi. 2020-2025-yillar orasida bank sohasidagi katta ma'lumotlar tahlili bozori o'rtacha yillik o'sish sur'ati 12,97 foizni tashkil etishi kutilmoqda.

Katta ma'lumotlar bozori 2020-yilda 198,08 mld dollarga baholangan bo'lsa, 2022-yilga kelib 274,3 mld dollargacha o'sadi va yaqin 5 yillikda yillik o'sish sur'ati 13,2 foizni tashkil qiladi. 2027-yilga kelib 103 mld dollargacha o'sishi va dasturiy ta'minot segmenti 45 foizni tashkil qilishi kutilmoqda.

BARC ma'lumotlariga ko'ra, tashkilotlar «katta ma'lumotlar»dan foyda ko'rmoqda. Xususan, strategik qarorlar qabul qilish imkoniyati 69 foizga, operatsion jarayonlar ustidan ko'proq nazoratni qo'lga kiritish imkoniyati 54 foizga, iste'molchilarni yaxshiroq tushunish imkoniyati 52 foizga va xarajatlarni kamaytirish imkoniyati 47 foizga o'sgan. Katta ma'lumotlar imkoniyatlaridan foydalananayotgan kompaniyalar daromadlarining o'rtacha 8 foizga oshgani va xarajatlari 10 foizga kamayganini e'tirof etishgan.

Forbes 2025-yilga borib real vaqt rejimida 150 zettabayt yoki 150 trln gigabaytdan ortiq ma'lumotlar tahlil qilinishini bashorat qilgan. Forbes ma'lumotlariga ko'ra, kompaniyalarning 95 foizdan ortig'i tuzilmagan ma'lumotlarning bir nechta to'plamini boshqarishda yordamga muhtoj, kompaniyalarning 40 foizi esa katta ma'lumotlar bilan shug'ullanish kerakligini aytadi.

StrategyMRC ma'lumotlariga ko'ra, Hadoop va katta ma'lumotlar bozori hozirda 138,9 mld dollarga baholanmoqda va o'rtacha yillik o'sish sur'ati 28,5 foizni tashkil qilgan holda 2022-yilda 30 mld dollarga o'sishi kutilmoqda.

Wikibon ma'lumotlariga ko'ra, katta ma'lumotlar, tahliliy va amaliy ma'lumotlar bazasi yechimlari hajmi o'n yil ichida 6 foiz o'rtacha yillik o'sish sur'atida 2017-yildagi 6,4 mld dollardan 2027-yilga kelib 12 mld dollargacha o'sishi kutilmoqda. Hadoop, Kafka, Spark va TensorFlow kabi katta ma'lumotlar ekotizimidagi ochiq kodli platformalarga bo'lgan talab uning sun'iy intellekt, mashinani o'rganish, chuqur o'rganish yoki ma'lumotlar faniga bevosita murojaat qilishi tufayli kamayishi mumkin. Ammo Hadoop, NoSQL, xotira, oqim va boshqa ko'plab ma'lumotlar bazalari kabi ma'lumotlarni tahlil qilish platformalarining gibridd o'rnatilishi ma'lumotlar ko'lami va ma'lumotlar ishlab chiqarish yechimlari bozor ulushini oshiradi.

Bugungi kunda «Big Data» texnologiyalari bozorida AQShning Amazon.com, Inc., IBM, Microsoft, Oracle, Dell Technologies, Cisco Systems, Inc., Cloudera, Inc., Salesforce.com, Inc., Teradata, Tableau Software, Hewlett-Packard, Prolifics, Inc., Xplenty, Clairvoyant, Teradata, EquBot Inc., Dell Technologies, Vmware, Inc.

Databrix korporatsiyalari va SAS instituti, Germaniyaning SAP SE va Software AG, Yaponiyaning Fujitsu Limited va CMIC Co., Ltd., Hindistonning Infosys Limited, Xitoyning Huawei Technologies Co., Ltd. va Buyuk Britaniyaning Deloitte Touche Tohmatsu Limited kompaniyalari yetakchilik qilmoqda.

2021-yilda mamlakatlarning global yirik ma'lumotlar va biznes-tahlil bozoridagi ulushi: AQSh – 51%; Yaponiya – 5,7%; Xitoy – 5,5% Buyuk Britaniya – 5,1%; Germaniya – 4,4%; boshqalar – 28,3%.

«Katta ma'lumotlar»ning kelajagi tahlilchilarni talab qiladi. Katta ma'lumotlar hajmi o'sishda davom etar ekan, unga sho'ng'ish va amaliy tushunchalarni olish uchun o'qitilgan ma'lumotlar tahlilchilariga ehtiyoj ortib bormoqda. «Katta ma'lumotlar» tahlili moliya, hukumat va sog'liqni saqlash kabi sohalarda o'zgarishlar yaratish uchun ajoyib imkoniyatlarni taqdim etadi, shuningdek, firibgarlikning oldini olish, tabiiy ofat yuz berganda resurslarni taqsimlash yoki sog'liqni saqlashni yaxshilash orqali odamlar hayotini o'zgartirishga yordam beradi.

Manba: Tuyg'un Asatov, «Axborot texnologiyalari va axborot resursslari rivojlantirish markazi» davlat unitar korxonasi Normativ-texnik hujjatlar ishlab chiqish bo'limi bosh mutaxassisi.

4. Big dataning asosiy sakkizta atamasi;

Dunyoda raqamlangan ma'lumotlar hajmi ekponent bo'yicha o'sib bormoqda. IBS kompaniyasining ma'lumotlariga qaraganda, 2003-yilda 5 eksabayt(1 eksabayt - 1 milliard gigabayt) ma'lumot yig'ilgan ekan. 2008-yilda u 0.18 zettabayt(1 zettabayt = 1024 eksabayt) gacha, 2011-yilga kelib 1.76 zettabayt, 2013-yilda 4.4 zettabaytgacha yetibdi. 2015-yilning mayida dunyoda yig'ilgan raqamlanga ma'lumotlar hajmi 6.5 zettabaytdan oshib ketibdi. 2020-yilga kelib insoniyat 40-44 zettabayt raqamli ma'lumot hosil qilar ekan.

IBS mutaxassislarining fikriga ko'ra, 2013-yilda yig'ilgan ma'lumotlar massivining atiga 1.5% i qandaydiy axborot qiymatiga ega bo'lgan ekan. Baxtga qarshi, hozir dunyoda katta ma'lumotlarni qayta ishslash texnologiyalari bo'lib, ular yordamida juda katta ma'lumotlar massividan insonlarga kerak, qiziq bo'lgan, foydali ma'lumotlarni ajratib olish mumkin bo'ladi.

Ushbu maqolada Big Data(katta ma'lumotlar)ning 8 eng asosiy atamalari va ular haqida qisqacha tushunchalar beriladi.

Big data(katta ma'lumotlar) - juda katta hajmdagi bir jinsli bo'lмаган 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 ishslash ham tushuniladi. Asosan, analiz obyekti katta ma'lumotlar deb ataladi.

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. Mutaxassislarning 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 hayotimizdag'i 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 vaqt ni 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 aytildi. Isroiilik 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 bergen. 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'lida to'g'rileydi. 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 tadbiq 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 tadbir 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'limgan juda katta hajmdagi ma'lumotlarni qayta ishlash natijasida muqobil biznes yechimlar izlashga aytildi.

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.