

## **1- 1 - §. Katta hajmli ma'lumotlar fanining maqsadi, vazifalari va asosiy tushunchalari.**

### ***Reja:***

- 1. Katta hajmli ma'lumotlar fanining maqsad va vazifalari*
- 2. Katta hajmli ma'lumotlarning tarkibiy tuzilmasi*

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 (1zettabayt = 1024 eksabayt) gacha, 2011-yilga kelib 1.76 zettabayt, 2013-yilda 4.4 zettabaytgacha yetibdi. 2015-yilning mayida dunyoda yig'ilgan raqamlarga 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 qandaydir axborot qiymatiga ega bo'lgan ekan. Baxtga qarshi, hozir dunyoda katta ma'lumotlarni qayta ishlash texnologiyalari bo'lib, ular yordamida juda katta ma'lumotlar massividan insonlarga kerak, qiziq bo'lgan, foydali ma'lumotlarni ajratib olish mumkin bo'ladi.

**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.

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.

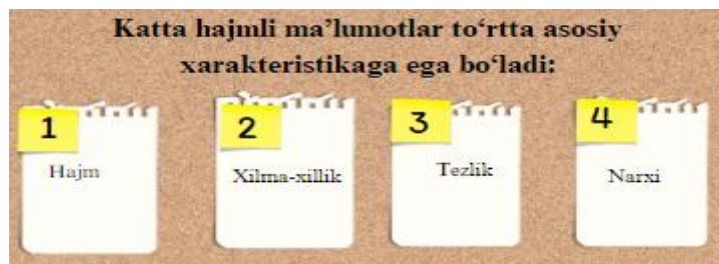
**Katta hajmli ma'lumot** (inglizcha: *Big data*) — katta hajmdagi ma'lumotlarga nisbatan qo'llanadigan termin hisoblanadi (odatda terabayt, ekzabayt va petabaytlar darajasida). Ushbu ma'lumotlarni amaldagi va yangi yaratilayotgan texnologiyalar yordamida to'plash, taqsimlash, o'zgartirishlar kiritish, tahlil qilish, saqlash va vizual shaklda tasvirlash imkoni 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.

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.

RFID-texnologiyasining keng tarqalishi, elektron transaksiyalarning moliya sohasida keng qo'llanilishi, ilmiy-tadqiqot muassasalarining to'plagan 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 hajmli multimedia ma'lumotlar to'rtta asosiy xarakteristikaga ega bo'ladi: hajm, xilma-xillik, tezlik, narhi.



**1. Hajm.** Insonlar va mashinalar tomonidan xosil qilinayotgan ma'lumotlarning o'sib borayotgan soni, axborot texnologiyalari infrastrukturalariga ma'lumotlarni saqlash, ishlov berish va taqdim etishida yangi talablarni qo'yadi.

**2. Xilma-xillik.** Turli strukturalarda taqdim etilgan ma'lumotlar xilma-xil bo'ladi. Bular kredit kartalar bo'yicha operatsiyalar bo'ladimi, ilmiy tadqiqot natijalari bo'ladimi, fotografik suratlar bo'ladimi, video va audio ma'lumotlar bo'ladimi barchasi o'ziga xos ishlov va saqlash shart sharoitlarini talab etadi.

**3. Tezlik.** Tezlik deganda nafaqat ma'lumotlarning ma'lumotlar bazasiga kelib tushish tezligi, balki bu ma'lumotlar bazasidan olinadigan ma'lumotlarni chiqarib olish tezligi ham anglanadi.

**4. Narhi.** Katta hajmdagi ma'lumotlar – qimmat resurs hisoblanadi. Ma'lumotning dolzarbligi, foydaliligi va ichidagi kontentiga qarab uning qanchalik qimmat.

**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 mutaxassislarga 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 zamonaviy 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.

**Deep learning**(*chuqur o'qitish*) - yanada murakkab va yanada mustaqil bo'lgan o'zi o'qidigan dasturlar yaratadigan mashinali o'qitish turi. Deep learningda 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.

### Nazorat savollari

1. Katta hajmli ma'lumotlar fanining maqsad va vazifalarini izohlang.

2. Katta hajmli ma'lumotlarning tarkibiy tuzilmasini izohlang.



**Mavzuni mustahkamlash uchun savollar.**

1. Katta hajmdagi ma'lumotlarning minimal o'lchovi qancha bo'lishi kerak?
  - a) Minimal cheklovlari yo'q
  - b) 1TB dan ortiq ma'lumotlar hajmi
  - c) 10TB dan ortiq ma'lumotlar hajmi
  - d) 100TB dan ortiq ma'lumotlar hajmi
2. Aksariyat qaysi sanoat tarmoqlari " Big Data"dan foydalanadi?
  - a) Ob-havo ma'lumotlari, Marketing, Sog'liqni saqlash, Walmartdo'koni
  - b) Walmart do'koni, internet do'konlar
  - c) Internet marketing, intellektual tahlil
  - d) Yetkazib berish tizimi, xususiy tadbirkorlik
3. Katta hajmdagi ma'lumotlarning muhim xususiyatlarini tanlang.
  - a) Hajm, tezlik, turli xillik
  - b) Tartib, tezlik, turli xillik
  - c) Hajm, harakat, xilma –xillik
  - d) Tartib, tezlanish, turli xillik

