

## 10-dars. BIG DATA VA DATA MINING

### Big Data

Taraqqiyot yangi texnologiyalarning yaratilishiga, yangi texnologiyalar esa katta ma'lumotlar to'plamlarining paydo bo'lishiga olib kelmoqda. Bugungi kunda raqamli ma'lumotlarning to'xtovsiz o'sishi insonlar hayoti va jamiyatning barcha jabhalarini qamrab oldi. Quyidagi raqamlar orqali bu jarayonni aniqroq tasavvur qilish mumkin. Yer yuzida har kuni insonlar



tomonidan 500 million tvitt , 294 milliard elektron xatt , 4 million gigabayt Facebook ma'lumotlari, 65 milliard WhatsApp xabarlari va YouTube'da 720 ming soatlik yangi tarkib yaratilmoqda. Bu birgina ijtimoiy platformalarda juda katta axborotlar saqlanayotganligini va ular kun sayin o'sib borayotganligini anglatadi. Hisob-kitoblarga ko'ra, 2020-yilda dunyodagi raqamli ma'lumotlarning umumiy hajmi 59 ZB ga yetgan. Bu 65 milliard gigabaytga tengdir. Ma'lumotlarning shu zaylda o'sib borishi natijasida 2025-yilga kelib raqamli ma'lumotlar 175 ZB ga yetishi taxmin qilinmoqda. Katta hajmdagi raqamli ma'lumotlarning asosiy qismi Data Center (ma'lumot markazi)da saqlanadi. Bugungi kunda dunyoda 600 ga yaqin katta Data Centerlar mavjud. Eng ko'p Data Centerlar AQSHda joylashgan bo'lsa, eng katta Data Center Xitoyda joylashgan. Dunyoda har yili yangi raqamli ma'lumotlarni saqlash uchun 50 ga yaqin katta Data Centerlar qurilmoqda. Data Centerlarning asosiy vazifasi Big Data, ya'ni katta hajmdagi raqamli axborotlarni saqlashdan iboratdir.

### BILASIZMI?

Axborot o'lchov birliklari:

1 bit = Binary Digit (eng kichik) 8 bit = 1 bayt

1 KB (kilobayt) = 1024 Bayt

1 GB (gigabayt) = 1024 MB

1 PB (petabayt) = 1024 TB

1 ZB (zetabayt) = 1024 EB

1 Bronto Bayt = 1024 YB

1 MB (megabayt) = 1024 KB

1 TB (terabayt) = 1024 GB

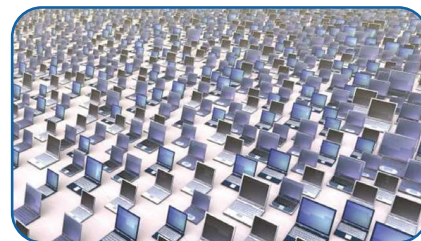
1 EB (ekzabayt) = 1024 PB

1 YB (yotabayt) = 1024 ZB

1 Geop Bayt = 1024 Bronto Bayt

Geop Bayt – eng yuqori birlik

**Big Data** – vaqt o'tgan sayin o'sib boruvchi katta ma'lumotlar to'plami. U o'zida ma'lumotlar ustida amallar bajarishga yo'naltirilgan texnik va dasturiy vositalarni jamlaydi. An'anaviy ma'lumotlar bazasidan farqli ravishda, Big Data katta hajmdagi ma'lumotlarni yig'ish, qayta ishlash va uzatish imkoniyatlarini taqdim etadi. Bu esa ma'lumotlarni samarali tahlil qilish orqali kerakli axborotni qo'lga kiritishga yordam beradi. Big Datani yaxshiroq tushunish uchun uning tur va xususiyatlarini yaxshi bilishimiz lozim.



Big Data 3 ta asosiy turga bo'linadi:

- 1) tuzilmalashtirilgan;
- 2) tuzilmalashtirilmagan;
- 3) yarim tuzilmalashtirilgan.



#### DIQQAT!

- **Tuzilmalashtirilgan Big Data** – aniq formatda qayta ishlanishi, saqlanishi va olinishi mumkin bo'lgan katta ma'lumotlar to'plami. U oddiy jadval ko'rinishidagi ma'lumotlar bo'lishi mumkin. Ya'ni bunda ma'lumotlar oldindan tartibli ravishda tuzilgan bo'ladi. Bunga misol tariqasida maktab o'quvchilari haqidagi elektron jadvalni keltirish mumkin.
- **Tuzilmalashtirilmagan Big Data** – aniq shakl yoki tuzilishga ega bo'lmagan katta hajmdagi ma'lumotlar to'plami. Misol tariqasida elektron pochta manziliga kelayotgan xabarlarini keltirish mumkin. Pochtaga kelgan ma'lumotlar tartibsiz, turli fayl formatlarida (matn, audio, video, tasvir, hujjat) bo'ladi. Tuzilmalashtirilmagan ma'lumotlarni qayta ishlash yoki tahlil qilish qiyin va ko'p vaqtni talab etadi.
- **Yarim tuzilmalashtirilgan Big Data** – bir vaqtning o'zida tuzilmalashtirilgan va tuzilmalashtirilmagan ma'lumotlarni o'z ichiga olgan katta ma'lumotlar to'plami.

### Big Dataning xususiyatlari

Big Data 5 ta asosiy xususiyati orqali oddiy ma'lumotlar bazasidan ajralib turadi.

#### 1. Hajm

Big Data o'zida katta hajmdagi ma'lumotlarni saqlashi va ularni boshqarishi mumkin. An'anaviy MBBT vositalari esa katta hajmdagi ma'lumotlar bilan ishlay olmaydi.

## 2. Tezlik

An'anaviy ma'lumotlar bazasidan farqli ravishda Big Datada ma'lumotlar hajmi juda tez sur'atlarda o'sadi.

## 3. Xilma-xillik

Big Datada ma'lumotlar to'plami turli xil ma'lumotlardan tashkil topadi. Masalan, matn, rasm, video, veb-sahifalar va boshqalar. Shu bilan birga, ma'lumotlarni olish manbalari ham turlichadir.

## 4. Aniqlik

Big Datada ma'lumotlarni yig'ish va saqlashda e'tibor aniqlikka qaratiladi. Ma'lumotlarning aniqligi uning to'g'ri va haqqoniyligini belgilovchi asosiy omildir. Agar ma'lumotlar aniq bo'lmasa, keyinchalik katta muammolar kelib chiqishi mumkin.

## 5. Qiymat

Big Datadagi ma'lumotlar qiymatli bo'lishi lozim. Bu keyinchalik ma'lumotlarni qayta ishlash natijasida kerakli axborotlarning taqdim etilishiga olib keladi. Qiymatli ma'lumotlar jarayonlarni oldindan bashorat qilish imkoniyatini taqdim etadi. To'plangan ma'lumotlar qiymatli bo'lishi uchun so'rovlar berish va tahlil qilish jarayonlari oldindan loyihalashtirilishi lozim.

Big Data bilan ishlashga yordam beruvchi ko'plab dasturiy vositalar yaratilgan. Bularga Xplenty, Adverity, Apache Hadoop, Cassandra kabilarni misol keltirish mumkin. Bu dasturlarning umumiy vazifasi katta hajmdagi axborotlarni to'plash, ularni saralash va foydalanuvchilarga taqdim etishdan iborat. Big Data dasturiy vositalari, asosan, bulutli saqlash tizimlaridan foydalanadi.

Big Data ta'lim, iqtisod, qishloq xo'jaligi va boshqa sohalarning rivojlanishini belgilovchi asosiy omillardan biri hisoblanadi.

## Data Mining

Raqamli texnologiyalarning rivojlanishi va katta hajmdagi axborotlarning paydo bo'lishi ulardan samarali foydalanish zaruratini yuzaga keltirdi. Masalan, biznes sohasining rivojlanishi mavjud ma'lumotlar tahlili va ehtimoliy bashorat



bilan uzviy bog'liqlikni kashf etdi. Ya'ni ma'lumotlarning noto'g'ri tahlil va bashorat qilinishi bugungi axborotlashgan davrda juda katta talafotga olib kelishi mumkin. Bu jarayonlarni to'g'ri tashkil etish uchun Data Miningdan foydalaniladi.

**Data Mining** – katta ma’lumotlar bazasidagi ma’lumotlarni tahlil qilish, ularning o’xshash va umumiy xususiyatlarini topish orqali yangi axborotlarni yaratish jarayoni. U katta ma’lumotlar to’plamidan kerakli ma’lumotlarni saralashga, shu tariqa to’g’ri qaror qabul qilishga yordam beradi. Data Mining sun’iy intellekt va mashinani o’qitish imkoniyatlaridan foydalangan holda yanada rivojlanmoqda.

### **Data Miningni amalga oshirish jarayoni**

Data Mining quyidagi ketma-ketlikda amalga oshiriladi:

- ma’lumotlar to’plamidan kerakli ma’lumotlarni ajratish;
- ma’lumotlardan o’xshash va umumiy xususiyatlilarni saralash;
- saralangan ma’lumotlar asosida foydali axborotlarni yaratish.

### **Ma’lumotlar to’plamidan kerakli ma’lumotlarni ajratish**

Odatda, ma’lumotlar bazasi o’zida katta miqdordagi ma’lumotlarni saqlaydi (masalan, dunyo mamlakatlari haqidagi ma’lumotlar saqlanadigan BMT ma’lumotlar bazasi). Foydalanuvchiga esa kerakli so’rovlarni amalga oshirish uchun ma’lumotlarning ma’lum bir qismi yetarli. Katta ma’lumotlar bazasidan kerakli ma’lumotlarni ajratish esa ish jarayonini osonlashtiradi.

Kerakli ma’lumotlarni ajratish uchun quyidagi savollarga javob topish lozim:

- Qaysi jadvallardan foydalanish kerak?
- Qaysi maydonlarni ajratib olish kerak?
- Qanday so’rovlarni yaratish kerak?

Ma’lumotlar to’plamidan kerakli ma’lumotlar ajratilganidan so’ng keyingi bosqichga o’tiladi.

### **Ma’lumotlardan o’xshash va umumiy xususiyatlilarni saralash**

Katta hajmdagi ma’lumotlar bilan ishlashda oddiy so’rovlar samarali natijalarga olib kelmasligi mumkin. Shu sababli qo’shimcha vositalar asosida murakkab so’rovlar yaratish va ma’lumotlarni tahlil qilish lozim. Tahlil esa ma’lumotlar orasidan o’xshash va umumiy xususiyatga egalarini topib beradi. Bu, o’z navbatida, maqbulroq qaror qabul qilish va samarali strategiyalarni ishlab chiqish uchun yangi ma’lumotlarni taqdim etadi.

### Saralangan ma'lumotlar asosida foydali axborotlarni yaratish

Bu bosqich ko'proq inson intellekti bilan bog'liq. Ma'lumotlarni saralash foydalanuvchiga bir qancha muhim ma'lumotlarni taqdim etadi. Foydalanuvchi esa shu ma'lumotlar asosida o'z faoliyatini amalga oshiradi va samaradorlikka erishadi.

#### Misollar

1. Facebook (va boshqa Internet gigantlari) foydalanuvchilar shaxsiy ma'lumotlarini to'playdi. To'planayotgan ma'lumotlar hajmi vaqt o'tgan sayin ko'payib boradi. Ma'lumot qancha ko'p bo'lsa, shuncha ko'p foydali ma'lumotni saralash mumkin. Ushbu ma'lumotlardan foydalanish huquqiga ega bo'lgan kompaniyalar, jismoniy shaxslar va Facebook ma'lumotlar to'plamidan foydalilarini ajratadi. Ajratilgan ma'lumotlarni kichikroq ma'lumotlar bazasida saqlaydi. Saqlangan ma'lumotlarni tahlil qilish orqali foydalanuvchilarning do'stlari va do'stlarining do'stlari, yashash manzili, yozishmalar tahlili, tarmoqdagi harakatlari kabilarni tahlil qilish imkoniyatiga ega bo'ladi. Bu ma'lumotlar keyinchalik profil egasining hozirgi emotsional holatini aniqlashi, uning holatidan kelib chiqib, xabarlarini lentaga saralashi, do'stlarni taklif qilishi, reklama yoki tijoriy takliflarni berishi mumkin.



2. Data Mining orqali telekanal faoliyatini tahlil qilish quyidagi ma'lumotlarni taqdim etadi:

- tomoshabinlar soni;
- kanalni kunning qaysi vaqtida ko'proq foydalanuvchi ko'rishi;
- tomoshabinlarning demografik joylashuvi.

Bu ma'lumotlar televideniye rahbariyatiga kelajakda qaysi film yoki ko'rsatuvni qaysi vaqtda qo'yishi, reklama narxlarini vaqtga qarab belgilashida yordam beradi. Bu orqali kelajakda televideniye kanali muxlislarining soni oshishiga erishish mumkin.

Bugungi kunda Data Mining texnologiyasidan Google, Instagram, Facebook, Uber, Netflix kabi gigant kompaniyalar faol foydalanishmoqda.

#### TAKRORLASH UCHUN SAVOL VA TOPSHIRIQLAR

1. Big Data ma'lumotlarni qanday saqlaydi?
2. Big Dataning qanday xususiyatlari mavjud?
3. Big Data va kichik ma'lumotlar bazasi o'rtasida qanday farq bor?

4. Data Mining jarayonini amalga oshirish ketma-ketligini tushuntiring.
5. Data Miningni qo'llash bo'yicha misollar keltiring.

## UYGA VAZIFA

1. Internet manbalaridan Xplenty, Adverity, Apache Hadoop, Cassandra dasturiy vositalari haqida ma'lumot to'plang. To'plangan ma'lumot asosida jadvalni to'ldiring:

Dastur nomi	Afzalliklari	Kamchiliklari	Narxi
Xplenty			
Adverity			
Apache Hadoop			
Cassandra			

2. Data Mining uchun sizda Amazon sotuvlari haqida quyidagi ma'lumotlar mavjud:

- ko'p sotilayotgan mahsulotlar;
- sotuvlarning demografik ko'rsatkichlari.

Bu ma'lumotlar mahsulot ishlab chiqaruvchilarga qanday yangi axborotlarni taqdim etadi? Buning mahsulot sotuvlariga ta'sirini tushuntiring.