

## KATTA HAJMLI MA'LUMOTLARNI QAYTA ISHLASH

Axmedova Nilufar Farxodovna

*Qarshi davlat texnika universiteti Kompyuter tizimlarining texnik va dasturiy ta'minoti  
kafedrası dotsenti*

Abdurazzoqov Abdunazar Abdug'aniyevich

*Qarshi davlat texnika universiteti Telekommunikatsiya texnologiyalari yo'nalishi 1-  
bosqich talabasi*

**Annotatsiya:** *Mazkur maqolada katta hajmli ma'lumotlarni qayta ishlash texnologiyalari, ularning asosiy tushunchalari va zamonaviy axborot tizimlaridagi o'rni yoritilgan. Katta hajmli ma'lumotlar (Big Data) tushunchasi, uning asosiy xususiyatlari hamda 3V modeli (Volume, Velocity, Variety) haqida ma'lumot berilgan. Shuningdek, katta hajmli ma'lumotlarni qayta ishlashda qo'llaniladigan texnologiyalar, jumladan Hadoop, Apache Spark va tarqatilgan hisoblash tizimlari haqida tushuntirishlar keltirilgan. Maqolada Big Data texnologiyalarining biznes, sog'liqni saqlash, moliya, transport va sun'iy intellekt sohalarida qo'llanilishi ham ko'rib chiqilgan. Bundan tashqari, katta hajmli ma'lumotlar bilan ishlashda yuzaga keladigan muammolar va ularni hal etish yo'llari haqida ham ma'lumot berilgan. Ushbu maqola katta hajmli ma'lumotlar bilan ishlashning ahamiyatini tushunishga yordam beradi.*

**Kalit so'zlar:** *katta hajmli ma'lumotlar, Big Data, ma'lumotlarni qayta ishlash, Hadoop, Apache Spark, tarqatilgan tizimlar, ma'lumotlar tahlili, axborot texnologiyalari, bulutli texnologiyalar, NoSQL ma'lumotlar bazasi.*

**Аннотация:** *В данной статье рассматриваются технологии обработки больших данных, их основные концепции и роль в современных информационных системах. Представлены концепция больших данных (Big Data), их основные характеристики и модель 3V (объем, скорость, разнообразие). Также приводятся пояснения к технологиям, используемым в обработке больших данных, включая Hadoop, Apache Spark и распределенные вычислительные системы. В статье также рассматривается применение технологий больших данных в сферах бизнеса, здравоохранения, финансов, транспорта и искусственного интеллекта. Кроме того, представлена информация о проблемах, возникающих при работе с большими данными, и способах их решения. Данная статья помогает понять важность работы с большими данными.*

**Ключевые слова:** *большие данные, Big Data, обработка данных, Hadoop, Apache Spark, распределенные системы, анализ данных, информационные технологии, облачные технологии, база данных NoSQL*

Katta hajmli ma'lumotlarni qayta ishlash

Hozirgi raqamli davrda ma'lumotlar hajmi juda tez sur'atlarda ortib bormoqda. Internet, ijtimoiy tarmoqlar, mobil qurilmalar, sensorlar va turli axborot tizimlari har kuni ulkan miqdorda ma'lumotlar ishlab chiqaradi.

Ushbu ma'lumotlarni samarali saqlash, qayta ishlash va tahlil qilish katta hajmli ma'lumotlar (Big Data) tushunchasini yuzaga keltirdi. Katta hajmli ma'lumotlarni qayta ishlash zamonaviy IT sohasining muhim yo'nalishlaridan biri hisoblanadi.

Shu sababli so'nggi yillarda katta hajmli ma'lumotlar (Big Data) tushunchasi keng qo'llanila boshladi. Big Data – bu an'anaviy ma'lumotlar bazalari yordamida samarali qayta ishlash qiyin bo'lgan katta hajmdagi ma'lumotlar to'plamidir. Bunday ma'lumotlar turli formatlarda bo'lishi mumkin: matn, rasm, video, audio yoki sensor ma'lumotlari.

Katta hajmli ma'lumotlarni qayta ishlash zamonaviy biznes, fan va texnologiya sohalarida muhim ahamiyatga ega. Ushbu texnologiyalar yordamida tashkilotlar katta hajmdagi ma'lumotlarni tahlil qilib, muhim qarorlar qabul qilishlari mumkin.

Katta hajmli ma'lumotlar tushunchasi

Katta hajmli ma'lumotlar – bu hajmi juda katta, tez yaratiladigan va turli xil formatlarda bo'lgan ma'lumotlar majmuasidir. Ularni qayta ishlash uchun maxsus texnologiyalar, algoritmlar va dasturiy vositalar talab etiladi.

Big Data odatda 3V modeli orqali tushuntiriladi:

1 Volume (Hajm)

Bu katta hajmdagi ma'lumotlarni anglatadi. Hozirgi kunda kompaniyalar petabayt va hatto eksabayt miqdoridagi ma'lumotlar bilan ishlashmoqda. Masalan, ijtimoiy tarmoqlar har kuni milliardlab foydalanuvchi ma'lumotlarini saqlaydi.

2 Velocity (Tezlik)

Ma'lumotlar juda tez yaratiladi va real vaqt rejimida qayta ishlanishi kerak bo'ladi. Masalan, bank operatsiyalari yoki internetdagi tranzaksiyalar soniyasiga minglab bo'lishi mumkin.

3 Variety (Turli xillik)

Katta hajmli ma'lumotlar turli formatlarda bo'ladi. Masalan:

- matnli ma'lumotlar
- audio va video fayllar
- rasm va grafiklar
- sensorlardan keladigan ma'lumotlar

Ba'zi mutaxassislar bu modelga yana ikkita element qo'shishadi:

- Veracity – ma'lumotlarning ishonchliligi
- Value – ma'lumotlardan olinadigan foyda

Katta hajmli ma'lumotlarni qayta ishlash texnologiyalari

Katta hajmdagi ma'lumotlarni qayta ishlash uchun maxsus texnologiyalar qo'llaniladi. Ular ma'lumotlarni tez va samarali qayta ishlash imkonini beradi.

1 Tarqatilgan hisoblash tizimlari

Tarqatilgan hisoblash tizimlari ma'lumotlarni bir nechta kompyuterlar yoki serverlar o'rtasida taqsimlab ishlaydi. Bu tizimlar yordamida katta hajmdagi ma'lumotlarni tez qayta ishlash mumkin.

Masalan, ma'lumotlar bir nechta serverlarda saqlanadi va har bir server ma'lumotning ma'lum qismini qayta ishlaydi. Natijada hisoblash jarayoni tezlashadi.

2 Hadoop texnologiyasi

Hadoop – katta hajmli ma’lumotlarni saqlash va qayta ishlash uchun mo’ljallangan ochiq kodli platforma hisoblanadi. U quyidagi asosiy qismlardan iborat:

- HDFS (Hadoop Distributed File System) – tarqatilgan fayl tizimi
- MapReduce – ma’lumotlarni qayta ishlash modeli
- YARN – resurslarni boshqarish tizimi

Hadoop yordamida katta hajmdagi ma’lumotlar turli serverlarda saqlanadi va parallel tarzda qayta ishlanadi.

### 3 Apache Spark

Apache Spark – katta hajmli ma’lumotlarni tezkor qayta ishlash uchun mo’ljallangan platforma. U Hadoop bilan solishtirganda tezroq ishlaydi va real vaqt tahlilini amalga oshirish imkonini beradi.

Spark quyidagi sohalarda qo’llaniladi:

- ma’lumotlarni tahlil qilish
- mashinaviy o’qitish
- real vaqt ma’lumotlarini qayta ishlash

Katta hajmli ma’lumotlarni saqlash tizimlari

Katta hajmdagi ma’lumotlarni saqlash uchun maxsus ma’lumotlar bazalari va tizimlar qo’llaniladi.

### 1 NoSQL ma’lumotlar bazalari

An’anaviy SQL bazalar katta hajmdagi ma’lumotlar bilan ishlashda ba’zan yetarli bo’lmaydi. Shu sababli NoSQL bazalari paydo bo’lgan.

NoSQL bazalarining asosiy turlari:

- hujjatga asoslangan bazalar
- kalit-qiymat bazalari
- grafik bazalar
- ustunli bazalar

Bu tizimlar katta hajmdagi ma’lumotlarni tez saqlash va qayta ishlash imkonini beradi.

### 2 Bulutli saqlash tizimlari

Bulut texnologiyalari ham katta hajmli ma’lumotlar bilan ishlashda muhim rol o’ynaydi. Bulutli saqlash tizimlari orqali ma’lumotlar internet orqali saqlanadi va ularga istalgan joydan kirish mumkin.

Bulut texnologiyalarining afzalliklari:

- yuqori hajmda saqlash imkoniyati
- moslashuvchanlik
- xarajatlarni kamaytirish

Katta hajmli ma’lumotlarning qo’llanilish sohalari

Big Data texnologiyalari turli sohalarda keng qo’llanilmoqda.

### 1 Biznes va marketing

Kompaniyalar mijozlar xatti-harakatlarini tahlil qilish uchun katta hajmli ma’lumotlardan foydalanadi. Bu orqali ular mijozlar ehtiyojlarini aniqlab, samarali marketing strategiyalarini ishlab chiqishlari mumkin.

### 2 Sog’liqni saqlash

Tibbiyot sohasida katta hajmli ma'lumotlar kasalliklarni aniqlash, davolash usullarini takomillashtirish va ilmiy tadqiqotlar olib borishda qo'llaniladi.

### 3 Bank va moliya

Bank tizimlarida Big Data texnologiyalari moliyaviy operatsiyalarni tahlil qilish va firibgarlik holatlarini aniqlashda ishlatiladi.

### 4 Transport va logistika

Transport sohasida katta hajmli ma'lumotlar yo'l harakati tahlili, logistika jarayonlarini optimallashtirish va transport oqimini boshqarishda qo'llaniladi.

### 5 Sun'iy intellekt

Sun'iy intellekt tizimlarini o'qitish uchun katta hajmdagi ma'lumotlar talab qilinadi. Shu sababli Big Data texnologiyalari AI rivojlanishida muhim rol o'ynaydi.

Katta hajmli ma'lumotlar bilan ishlash muammolari

Katta hajmdagi ma'lumotlarni qayta ishlashda bir qator muammolar mavjud.

### 1 Ma'lumot xavfsizligi

Katta hajmdagi ma'lumotlar maxfiy bo'lishi mumkin. Shu sababli ularni himoya qilish juda muhim.

### 2 Saqlash xarajatlari

Katta hajmdagi ma'lumotlarni saqlash uchun ko'p resurs va xarajat talab qilinadi.

### 3 Mutaxassislar yetishmasligi

Big Data texnologiyalari bilan ishlay oladigan mutaxassislar yetishmasligi ham muhim muammolardan biridir.

## Foydalanilgan adabiyotlar:

1. Han J., Kamber M. Data Mining: Concepts and Techniques
2. Tom White. Hadoop: The Definitive Guide
3. Viktor Mayer-Schönberger, Kenneth Cukier. Big Data: A Revolution
4. www.ibm.com – Big Data texnologiyalari haqida ma'lumotlar
5. www.oracle.com – ma'lumotlar bazasi tizimlari haqida