

KATTA MA'LUMOTLAR VA ULARNI BOSHQARISH TIZIMLARI

Borotov Jo'raqo'zi Shukurjon o'g'li
Muhiddinova Dilmuraxon Saminjon qizi
 Qo'qon universitet, Maxsus pedagogika
 1-kurs MP 4-25 guruh talabasi
<https://doi.org/10.5281/zenodo.18081071>

Annotatsiya: Ushbu maqolada katta ma'lumotlar (Big Data) nazariyasi, uning asosiy xususiyatlari, arxitektura va qayta ishlash jarayonlari, zamonaviy boshqaruv tizimlari (Hadoop, Spark, NoSQL va boshqalar), xavfsizlik va maxfiylik muammolari hamda amaliy qo'llanish sohalari tahlil qilinadi. Maqola oxirida kalit so'zlar va ilmiy-adabiyotlar (manbalar) ro'yxati berilgan.

Kalit so'zlar: Big Data, katta ma'lumotlar, data management, ma'lumotlarni boshqarish, Hadoop, HDFS, YARN, Apache Spark, Spark Streaming, MLib, MapReduce.

Kirish. So'nggi o'n yilliklarda raqamli qurilmalar, IoT, ijtimoiy tarmoqlar va korporativ tizimlar orqali yaratilayotgan ma'lumotlar hajmi va murakkabligi keskin oshdi. An'anaviy ma'lumotlar bazalari va tizimlari bu yangi o'lchovdagi ma'lumotlar bilan samarali ishlay olmay qoladi — shuning uchun Big Data texnologiyalari rivojlandi. Big Data nafaqat hajm (volume) balki tezlik (velocity), xilma-xillik (variety), ishonchlilik (veracity) va qiymat (value) jihatlardan ham o'lchanadi — bu 5V modeli amaliy tahlillarda keng qo'llaniladi. (Laney tomonidan ilgari surilgan 3V modeli ham keng tarqalgan; tarixiy ma'noda 3V kontseptsiyasi va undan keyingi kengaytmalar Big Data ga nazariy asos bo'ladi).

Asosiy qism. Katta ma'lumotlar — an'anaviy ma'lumotlar boshqaruv vositalarining chegaralaridan oshib ketadigan hajm, tezlik va xilma-xillikka ega bo'lgan ma'lumotlar majmui. Ular biznes-yagona qarorlar, ilmiy tadqiqotlar va davlat boshqaruvi uchun yangi imkoniyatlar yaratadi, biroq maxsus arxitektura va algoritmlar talab qiladi.

5V modeli. Volume (Hajm): petabayt va undan yuqori hajmlar.

- ✓ Velocity (Tezlik): real-vaqt yoki near-real-vaqt oqimlari.
- ✓ Variety (Xilma-xillik): strukturalangan (jadval), yarim strukturalangan (JSON, XML) va strukturasiz (matn, tasvir, video) ma'lumotlar.
- ✓ Veracity (Ishonchlilik): ma'lumotlardagi shovqin va noaniqlik.
- ✓ Value (Qiymat): tahlil orqali olinadigan biznes va ilmiy qiymat.

Big Data tizimlari strukturaviy xilma-xillikka javob berishi kerak: ma'lumot manbalari (sensör, tranzaksiya, ijtimoiy media), tez-tez o'zgaruvchi sxemalar, metama'lumotlar (metadata) va lineage (ma'lumot kelib chiqishi) muhim rol o'ynaydi. Ma'lumotlarning semantik boyligi tahlil qilishda kontekstual transformatsiyalarni talab qiladi.

Arxitektura va qayta ishlash modellari

- ✓ Ma'lumot yig'ish va ingestiya. Ma'lumotni real vaqt oqimida yoki to'plam (batch) shaklida qabul qilish uchun brokerlar va ingestiya tizimlari kerak. Apache Kafka, Flume, va boshqa messaging/streaming vositalari ma'lumot oqimlarini ishonchli va o'lchovli tutish uchun keng qo'llanadi.
- ✓ Saqlash: taqsimlangan omborlar

Katta hajmli fayllarni saqlash uchun taqsimlangan fayl tizimlari va ob'ekt omborlari ishlatiladi. Apache Hadoop HDFS – commodity-serverlarda katta fayllarni saqlash uchun

mo'ljallangan keng tarqalgan yechimdir; HDFS arxitekturasi bloklar, NameNode va DataNode prinsiplari asosida ishlaydi. hadoop.apache.org+1

- ✓ Ma'lumotlar bazalari: NoSQL va boshqa yondashuvlar. Relatsion DBMS cheklovlari bor joylarda NoSQL tizimlari (document — MongoDB; wide-column — Cassandra; key-value — Redis; graph — Neo4j) qo'llanadi. Har bir model o'zining konsistensiya, mavjudlik va kechikish (latency) xususiyatlariga ega. NoSQL asoslari va amaliy yondashuvlar haqida keng tavsiyalar mavjud. [ACM Digital Library](#)
- ✓ Qayta ishlash: batch vs stream. Batch (yig'ma): MapReduce modeli katta to'plamlarni parallel qayta ishlash uchun asos solgan; Jeff Dean va Sanjay Ghemawatning MapReduce ishlanmasi bu paradigmaning nazariy va amaliy poydevorini tashkil qiladi. [Google Research](#)

Stream (oqim): real-vaqt tahlil uchun Spark Structured Streaming, Kafka Streams, Flink va boshqalar ishlatiladi. Apache Spark xotira asosida ishlash tufayli ko'plab vazifalar uchun yuqori tezlik beradi. [Apache Spark+1](#)

- ✓ Orkestratsiya va resurs boshqaruv. YARN kabi resurs boshqaruv tizimlari hisoblash va xotira resurslarini taqsimlash, ishlarni nazorat qilish va monitoring qilishni osonlashtiradi. hadoop.apache.org

Katta ma'lumotlarni boshqarish tizimlari (BDMS) va ekotizim.

- ✓ Hadoop ekotizimi. Hadoop ekotizimi HDFS, MapReduce/YARN, Hive (SQL-like tili), Pig va boshqa asboblar bilan katta ma'lumotlarni saqlash va analiz qilish uchun keng qo'llanadi. HDFS ombor sifatida, Hive esa analitik so'rovlarni yozishda yordam beradi. hadoop.apache.org+1
- ✓ Apache Spark. Spark — umumiy maqsadli, tezkor, xotira asosidagi analiz dvigateli. U Spark SQL, MLlib (mashina o'qitish), GraphX (graf tahlil) va Structured Streaming kabi modullar orqali keng imkoniyatlar beradi. Spark yirik klasterlarda oson parallelizatsiya va iterativ hisoblashlarni amalga oshiradi. [Apache Spark+1](#)
- ✓ Data Lake va Data Warehouse. Data Warehouse: strukturaviy, optimallashtirilgan so'rovlar uchun (OLAP), an'anaviy ETL jarayonlari bilan.

Data Lake: xom va yarim strukturalangan ma'lumotlarni saqlash, keyinchalik kerakli transformatsiyalar bilan ishlov berish. Ko'p tashkilotlar hybrid yondashuvni (lake + warehouse) qo'llaydi.

- ✓ Integratsiya va metadata boshqaruvi. Ma'lumot kataloglari, schema registry va lineage tizimlari (masalan, Apache Atlas, AWS Glue Catalog) data governance uchun zarur. Bu tizimlar audit, qidiruv va moslik (compliance) jarayonlarini soddalashtiradi.

Ma'lumotlarni tahlil qilish va AI integratsiyasi

- ✓ Statistik tahlil va BI. ETL/ELT jarayonlaridan so'ng ma'lumotlar BI (Business Intelligence) vositalari yordamida ko'rsatkichlar (KPIs) va hisobotlar uchun tayyorlanadi. Tableau, Power BI, Kibana kabi vizualizatsiya vositalari keng qo'llanadi.
- ✓ Mashinaviy o'qitish va sun'iy intellect. Big Data infrastrukturasi mashina o'qitish bilan sinergiyada bo'lib, katta hajmdagi ma'lumotlar asosida model qurish (ML ops), modelni joylashtirish va monitoring qilish amaliyotlari rivojlandi. Spark MLlib va boshqa kutubxonalar bu jarayonni osonlashtiradi. [Apache Spark+1](#)
- ✓ Graf va tavsiya tizimlari. Graf bazalari (Neo4j va boshqalar) ijtimoiy tarmoqlar, tavsiya tizimlari, yo'l va tarmoq optimizatsiyasida samarali ishlaydi.

Xavfsizlik, maxfiylik va huquqiy me'yorlar

- ✓ Xavfsizlik choralari. Big Data kodlash (encryption), autentifikatsiya (Kerberos, OAuth), ruxsat nazorati (RBAC), auditi va tarmoq segmentatsiyasi kabi qat'iy choralarga muhtoj. Yana, tizimlar uchun DoS/DDoS, ruxsatsiz kirish va ma'lumot yo'qolishi tahdidlari doimiy xavotir manbai hisoblanadi.
- ✓ Maxfiylik va me'yoriy talablar. Shaxsiy ma'lumotlarni qayta ishlashda GDPR (Yevropa Ittifoqi) kabi qoidalar amal qilishi mumkin — ular ma'lumot sub'ektining huquqlari, shaffoflik va ma'lumotni minimal darajada saqlash prinsplarini belgilaydi. AQSh kontekstida sog'liqni saqlash ma'lumotlari HIPAA bilan himoyalangan — bu Big Data loyihalarida maxsus shifrlash, kirish nazorati va audit talablarini keltirib chiqaradi. EUR-Lex+2gdpr-info.eu+2
- ✓ Etika va bias. Katta ma'lumotlardan kelib chiqib yaratilgan modellarda adolatsizlik (bias), diskriminatsiya xavfi mavjud. Data scientist va muhandislar modellarni adolatli, shaffof va tushunarli qilish uchun metodologiyalar joriy etishi kerak.

Amaliy qo'llanmalar va sanoat misollari

- ✓ Tibbiyot va genomika. Genomik ma'lumotlar, elektron sog'liq kartalari va real-vaqt monitoring ma'lumotlarining birikmasi klinik qarorlarni qo'llab-quvvatlashga yordam beradi.
- ✓ Moliyaviy texnologiyalar. Firibgarlikni aniqlash, risk-baholash va real-vaqt tranzaksiya monitoringi Big Data yondashuvlari yordamida amalga oshiriladi.
- ✓ Transport va logistika. Yo'l-harakatini optimallashtirish, yuk manzillarini rejalashtirish va tarmoq samaradorligini oshirish uchun katta hajmdagi telemetriya va sensor ma'lumotlari tahlil qilinadi.
- ✓ Davlat va ijtimoiy soha. Epidemiologiya monitoringi, fuqarolar so'rovlarini tahlil qilish, shuningdek shahar infratuzilmasini optimallashtirish kabi sohalarda Big Data muhim rol o'ynaydi.

Rivojlanish istiqbollari va tavsiyalar

Edge computing va hybrid arxitekturalar — ma'lumotlarni chekkada (edge) oldindan qayta ishlash tarmoqqa bo'lgan yukni kamaytiradi va kechikishni qisqartiradi.

AI va avtomatizatsiya integratsiyasi - ML modelari real-vaqt oqimlarda integratsiyalashuvi va avtomatizatsiyalashuvi kuchayadi. Kvantsar hisoblash (quantum) — uzoq muddatda kvant algoritmlari ba'zi ma'lumot tahlili vazifalarini sezilarli tezlashtirishi mumkin.

Ma'lumot guvohnomalari (data provenance), etika va regulyatorlik - tashkilotlar data governance va etik qoidalarni kuchaytirishi zarur.

Xulosa shuki, katta ma'lumotlar va ularni boshqarish tizimlari zamonaviy iqtisodiyotning, ilm-fanning va davlat boshqaruvining ajralmas qismiga aylangan. Ularning muvaffaqiyati texnologik yechimlar, kuchli governance, xavfsizlik va me'yoriy muvofiqlikni birlashtirishga bog'liq. Kelajakda edge computing, AI integratsiyasi va yangi hisoblash paradigmatlari Big Data sferasini yanada boyitadi.

Adabiyotlar, References, Литературы:

1. Quyida maqolada ishlatilgan asosiy manbalar keltirilgan. (Internet asosidagi manbalarga havolalar va rasmiy hujjatlar qo'shildi.)

2. Dean, J., & Ghemawat, S. (2004). *MapReduce: Simplified Data Processing on Large Clusters*. Google Research. PDF. [Google Research](#)
3. Laney, D. (2001). *3Vs — Volume, Velocity, Variety* (Gartner / blog overview). Ta'rif va tarixiy kontekst. [community.aiim.org+1](#)
4. Apache Hadoop Project — HDFS, YARN va ekotizim hujjatlari. Apache Hadoop rasmiy sayti. [hadoop.apache.org+1](#)
5. Apache Spark Documentation — Overview, Structured Streaming, MLlib. Apache Spark rasmiy hujjati. [Apache Spark+1](#)
6. Kleppmann, M. (2017). *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. O'Reilly Media. (Kitob tavsifi va qo'llanma). [Амазон+1](#)
7. Sadalage, P. J., & Fowler, M. *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*. (NoSQL asoslari). [ACM Digital Library+1](#)
8. GDPR — Regulation (EU) 2016/679 (General Data Protection Regulation). Rasmiy matn va sharh. [EUR-Lex+1](#)
9. U.S. Department of Health & Human Services — HIPAA overview va Security Rule summary. (HIPAA hujjatlari va tavsiflari). [hhs.gov+1](#)
10. Nurmatov, S. (2025). SIMMETRIYALAR VA GRUH HARAKATLARI: ALGEBRAIK GEOMETRIYA NUQTAI NAZARIDAN. Наука и инновация, 3(23), 39-40.