KATTA MA'LUMOTLAR BILAN ISHLASH TEXNOLOGIYALARI

Qurbonova Diyoraxon Dilmurodjon qizi
Farg'ona davlat universiteti, Amaliy matematika mutaxassisligi, 1-kurs magistranti
E-mail: diyoraxon@example.com

Annotatsiya. Ushbu maqolada katta ma'lumotlar (Big Data) bilan ishlashning zamonaviy texnologiyalari va usullari tahlil qilinadi. Raqamli iqtisodiyot va sun'iy intellekt rivojlanishi fonida har kuni ishlab chiqariladigan ma'lumotlar hajmining keskin o'sishi, ularni samarali saqlash, qayta ishlash va tahlil qilish muammosini dolzarb masalaga aylantirmoqda. Tadqiqotda Hadoop, Apache Spark, NoSQL ma'lumotlar bazalari va bulutli hisoblash kabi asosiy Big Data texnologiyalari ko'rib chiqiladi. Shuningdek, ma'lumotlarni qayta ishlashda qo'llaniladigan algoritmlar, parallel va taqsimlangan hisoblash modellari hamda tizim samaradorligini baholash mezonlari tahlil etiladi. Natijada Big Data texnologiyalari zamonaviy ilm-fan, sanoat va davlet boshqaruvida muhim rol o'ynashi asoslab beriladi.

Kalit so'zlar. Big Data, Hadoop, Apache Spark, NoSQL, bulutli hisoblash, taqsimlangan tizimlar, ma'lumotlar tahlili, parallel hisoblash, MapReduce, ma'lumotlarni qayta ishlash.

Bugungi kunda Internet, ijtimoiy tarmoqlar, IoT qurilmalari va biznes tizimlari har soniyada ulkan hajmdagi ma'lumotlarni ishlab chiqarmoqda. IDC kompaniyasining ma'lumotlariga ko'ra, 2025-yilga borib global ma'lumotlar hajmi 175 zettabaytga yetishi bashorat qilinmoqda. Bu ko'rsatkich katta ma'lumotlar (Big Data) texnologiyalariga bo'lgan ehtiyojni keskin oshirmoqda. An'anaviy ma'lumotlar bazalari va qayta ishlash tizimlari

bunday hajmdagi ma'lumotlarni boshqarishga qodir emas. Shu sababli maxsus Big Data texnologiyalari va arxitekturalari ishlab chiqilgan.

Big Data tushunchasi odatda uchta asosiy xususiyat — hajm (Volume), tezlik (Velocity) va xilma-xillik (Variety) bilan tavsiflanadi. Keyinchalik bu model ikkita qo'shimcha xususiyat — haqiqiylik (Veracity) va qadriyat (Value) bilan kengaytirildi, natijada "5V" modeli hosil bo'ldi. Ushbu xususiyatlar Big Data tizimlariga qo'yiladigan talablarni belgilab beradi va texnologiyalar tanloviga bevosita ta'sir qiladi.

Big Data arxitekturasi va texnologik yondashuv

Katta ma'lumotlar bilan ishlashda eng keng tarqalgan texnologik yondashuv Lambda va Kappa arxitekturalaridir. Lambda arxitekturasida ma'lumotlar ikki yo'nalishda — "batch" (to'plam) va "stream" (oqim) rejimlarida parallel qayta ishlanadi. Batch qatlam tarixiy ma'lumotlarni to'liq qayta ishlaydi, stream qatlam esa real vaqt ma'lumotlarini tezkor tahlil qiladi. Kappa arxitekturasi esa soddalashtirilgan variant bo'lib, faqat stream qayta ishlashga asoslanadi.

Big Data tizimlarining samaradorligi quyidagi asosiy ko'rsatkichlar orqali baholanadi:

Tizim o'tkazuvchanligini hisoblash formulasi:

$$\mathbf{T} = \mathbf{N} / \mathbf{t} \quad (1)$$

bu yerda: **T** — o'tkazuvchanlik (record/s); **N** — qayta ishlangan yozuvlar soni; **t** — vaqt oralig'i (soniya).

MapReduce modelida umumiy ishlash vaqtini baholash:

$$\mathbf{T}total = \mathbf{T}map + \mathbf{T}shuffle + \mathbf{T}reduce \quad (2)$$

Tmap — xaritalash bosqichi vaqti; **T**shuffle — ma'lumotlarni tarqatish vaqti; **T**reduce — qisqartirish bosqichi vaqti.

Amdahl qonuni asosida parallel hisoblash tezlashuvini baholash:

$$S(n) = 1 / (1 - p + p/n) \quad (3)$$

bu yerda: **S(n)** — n ta protsessor bilan erishiladigan tezlashuv; **p** — parallel bajarilishi mumkin bo'lgan vazifalar ulushi; **n** — protsessorlar soni.

Apache Hadoop ekotizimi

Apache Hadoop — katta ma'lumotlarni taqsimlangan muhitda saqlash va qayta ishlash uchun mo'ljallangan ochiq kodli platforma. Uning ikkita asosiy komponenti mavjud: HDFS (Hadoop Distributed File System) va MapReduce hisoblash modeli. HDFS ma'lumotlarni bloklarga bo'lib, klasterdagi turli nodalarga tarqatib saqlaydi. Odatda har bir blok uchun 3 ta nusxa saqlanadi, bu esa tizim ishonchliligini ta'minlaydi.

HDFS saqlash samaradorligini hisoblash:

$$Seff = (S \times r) / n \quad (4)$$

bu yerda: **S** — umumiy ma'lumotlar hajmi; **r** — replikatsiya koeffitsienti (odatda 3); **n** — nodallar soni.

Apache Spark va in-memory hisoblash

Apache Spark — Hadoopdan farqli ravishda ma'lumotlarni asosan operativ xotirada (in-memory) qayta ishlaydi. Bu esa Sparkni Hadoopga nisbatan batch ishlov berishda 10–100 barobar, real vaqt tahlilida esa yanada tezroq ishlashini ta'minlaydi. Spark RDD (Resilient Distributed Dataset) abstraktsiyasiga asoslanadi, bu esa nosozliklarga chidamli taqsimlangan ma'lumotlar to'plamlarini boshqarish imkonini beradi.

Spark klasterida ma'lumot hajmiga qarab optimal partitsiyalar sonini hisoblash:

$$P = ceil(DataSize / BlockSize) \quad (5)$$

bu yerda: **P** — partitsiyalar soni; **DataSize** — ma'lumotlar hajmi (MB); **BlockSize** — bir partitsiya hajmi (odatda 128 MB yoki 256 MB).

Asosiy Big Data texnologiyalari taqqoslash

Xususiyat	Hadoop	Apache Spark	Apache Flink	Kafka
Qayta ishlash turi	Batch	Batch + Stream	Stream	Message broker
Tezlik	Past	Yuqori	Juda yuqori	Juda yuqori
Xotira	Disk	RAM (asosan)	RAM	Disk + RAM
Murakkablik	O'rtacha	O'rtacha	Yuqori	Past
Qo'llanish sohasi	Arxiv tahlil	ML, SQL, Batch	Real vaqt	Ma'lumot oqimi

NoSQL ma'lumotlar bazalari

Katta hajmdagi tuzilmagan yoki yarim tuzilmagan ma'lumotlarni saqlash uchun NoSQL ma'lumotlar bazalari keng qo'llaniladi. Ular to'rtta asosiy turga bo'linadi: kalit-qiymat (Redis, DynamoDB), ustunli (Cassandra, HBase), hujjatli (MongoDB, CouchDB) va grafli (Neo4j) bazalar. Har bir tur o'ziga xos ma'lumot modeli va so'rovlar mexanizmiga ega.

Big Data tizimlarining asosiy afzalliklari quyidagilardan iborat:

– ma'lumotlarni gorizontal kengaytirish (horizontal scaling) imkoniyati;
– real vaqt rejimida ulkan hajmdagi ma'lumotlarni tahlil qilish;
– turli formatlardagi (tuzilmagan, yarim tuzilmagan) ma'lumotlar bilan ishlash;
– bulutli infratuzilma bilan integratsiya;
– mashinali o'qitish modellari uchun ma'lumot tayyorlash.

Big Data texnologiyalarining amaliy qo'llanish sohalari

Moliya va bank sektori Big Data texnologiyalaridan firibgarlikni aniqlash (fraud detection) va kredit risklarini baholashda foydalanadi. Real vaqtda millionlab

tranzaksiyalarni tahlil qilib, g'ayritabiiy moliyaviy harakatlarni soniyalar ichida aniqlash mumkin. Streaming ma'lumotlarni qayta ishlash tizimlari, jumladan Apache Kafka va Spark Streaming, bu jarayonning asosini tashkil etadi.

Sog'liqni saqlash sohasida genomika ma'lumotlarini tahlil qilish, kasalliklarni bashorat qilish va tibbiy tasvirlarni qayta ishlash uchun Big Data keng qo'llaniladi. Bitta inson genomi taxminan 200 GB ma'lumotdan iborat bo'lib, bu hajmdagi ma'lumotlarni qayta ishlash faqat taqsimlangan hisoblash tizimlarida mumkin.

Aqlli shahar tizimlarida transport oqimlarini boshqarish, energiya sarfini optimallashtirish va fuqarolar xavfsizligini ta'minlash maqsadida kamera va sensorlardan kelayotgan ma'lumotlar real vaqtda tahlil qilinadi. O'zbekistonda ham Toshkent shahri miqyosida aqlli shahar loyihasi doirasida bunday texnologiyalar joriy etilmoqda.

Xulosa

Tahlillar shuni ko'rsatdiki, katta ma'lumotlar bilan ishlash texnologiyalari zamonaviy raqamli iqtisodiyotning asosiy poydevorlaridan biriga aylanib bormoqda. Apache Hadoop, Spark va Flink kabi platformalar taqsimlangan hisoblash orqali petabayt miqyosidagi ma'lumotlarni samarali qayta ishlash imkonini beradi. NoSQL ma'lumotlar bazalari esa gorizontal kengayish va yuqori yuklamada ishlash imkoniyatlari bilan an'anaviy relatsion tizimlarga munosib alternativa bo'lmoqda. Kelajakda sun'iy intellekt va avtonom hisoblash tizimlarining integratsiyasi Big Data texnologiyalarining imkoniyatlarini yanada kengaytiradi. O'zbekistonda raqamli transformatsiya strategiyasi doirasida ushbu texnologiyalarni davlat boshqaruvi, sanoat va ilm-fanga joriy etish muhim ustuvor yo'nalishlardan biri bo'lib qolmoqda.

Foydalanilgan adabiyotlar ro'yxati

1. White T. Hadoop: The Definitive Guide. O'Reilly Media, 2015.

2. Karau H., Warren R. High Performance Spark. O'Reilly Media, 2017.

3. Marz N., Warren J. Big Data: Principles and Best Practices. Manning Publications, 2015.

4. Kleppmann M. Designing Data-Intensive Applications. O'Reilly Media, 2017.

5. Hasanov A.B. Katta ma'lumotlar texnologiyalari va ularning qo'llanilishi. Toshkent, 2024.

6. O'zbekiston Respublikasi Prezidentining «Raqamli O'zbekiston — 2030» strategiyasi. Toshkent, 2020.

7. IDC. Data Age 2025: The Evolution of Data to Life-Critical. Framingham, 2018.