ISNN 3030-3907

2026
MAY
ISSUE - 5
VOLUME - 8

DEVELOPMENT OF SCIENCE

scientific journal



Development of science

Ilmiy jurnal

MAY 2026
Issue - 5, VOLUME – 8

ISSN 3030 -3907

Ilmiy jurnal OAK rayosatining 2024 yil 27 sentabrdagi 361-son qaroriga asosan 02.00.00 kimyo fanlari hamda 2025 yil 12 fevral 367- son qaroriga asosan 05.00.00 texnika fanlari bo‘yicha falsafa doktori (PhD) va fan doktori (DSc) ilmiy darajasiga talabgorlarning dissertatsiya ishlari yuzasidan dissertatsiyalari asosiy ilmiy natijalarini chop etish tavsiya etilgan milliy ilmiy nashrlar ro‘yxatiga kiritilgan.

BOSH MUHARRIR:
Temirov Alisher Hoshim o‘g‘li – texnika fanlari falsafa doktori (PhD),Dotsent.
TAHRIRIYAT KENGASH RAISI:
Fozilov Sadriddin Fayzullayevich – texnika fanlari doktori, professor
TAHRIRIYAT KENGASH A’ZOLARI

TEXNIKA FANLARI/ TECHNICAL SCIENCES/ ТЕХНИЧЕСКИЕ НАУКИ

Bozorov G‘ayrat Rashidovich
Buxoro Davlat Texnika Universiteti, professor
Abdurahmonov Olim Rustamovich
Buxoro Davlat Texnika Universiteti, professor
Adizov Bobur Zamirovich
O‘zbekiston Respublikasi Fanlar akademiyasi Umumiy va noorganik kimyo instituti, professor
Hayitov Ruslan Rustamjonovich
Buxoro Davlat Texnika Universiteti, professor
Jumayev Qayum Karimovich
Buxoro Davlat Texnika Universiteti, professor.
Berdiyev Dorob Murodovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor
Maxmudov Rafiq Amonovich
Buxoro Davlat Texnika Universiteti, professor.
Tilloyev Lochin Ismatilloyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Raxmonov Qahramon Sanakulovich
Buxoro Davlat Texnika Universiteti, dotsent.
Barakayev Nusratullo Rajabovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor.
Jo‘rayev Xayrullo Fayziyevich
Buxoro Davlat Texnika Universiteti, professor.
Panoyev Erali Rajabboyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Ataullayev Sherzod Nabiullyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Malik Negmurodovich
Buxoro Davlat Texnika Universiteti, dotsent.
Axmedova Ozoda Bahronovna
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti.
Safarov Bahri Jumayevich
Buxoro Davlat Texnika Universiteti, dotsent.
Xaydarov Axtam Amonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Safarov Jasur Alijon o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.

KIMYO FANLARI/ CHEMICAL SCIENCES/ ХИМИЧЕСКИЕ НАУКИ

Axmedov Voxid Nizomovich
Buxoro Davlat Texnika Universiteti, professor
Umarov Baqo Bafoyevich
Buxoro Davlat Universiteti, professor
Kodirov Abduaxad Abduraximovich
Qarshi davlat Universiteti, professor
Qurbonov Mingniqul Jumagulovich
Qarshi davlat Universiteti, dotsent
Turemuratov Sharibay Naurizbaevich
O‘zbekiston Respublikasi Fanlar akademiyasi Qoraqalpog‘iston bo‘limi Bosh ilmiy kotibi, professor
Tursunov Kaxor Shonazarovich
Qarshi davlat Universiteti, professor
Do‘stov Hamro
Buxoro Davlat Texnika Universiteti, professor
Panoyev Nodir Shavkatovich
Buxoro Davlat Texnika Universiteti, dotsent.
Olimov Bobur Bahodir o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.
Kasimov Sherzod Abduazimovich
Termiz Davlat Universiteti, professor.
Aliqulov Rustam Valiyevich
Termiz Davlat Universiteti, professor
Tropov Xamza Tursunovich
Samarqand Davlat Universiteti, professor

IQTISODIYOT FANLARI/ ECONOMICS/ ЭКОНОМИЧЕСКИЕ НАУКИ

Nizamov Aslitdin Badreddinovich
Buxoro Davlat Texnika Universiteti, professor.
Boboyev Akmal Choriyevich
Buxoro Davlat Texnika Universiteti, professor.
Ergashxojayeva Shaxnoza Djasurovna
Toshkent davlat iqtisodiyot universiteti, professor
Saidaxmedova Nodira Ilxomovna
Toshkent davlat iqtisodiyot universiteti, professor
Muxsinov Bekzod Toxirovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shakarov Qulmat Ashirovich
Renessans ta'lim universiteti,dotsent.
Jumaeva Zulfiya Qayumovna
Buxoro Davlat Texnika Universiteti, dotsent.
Norova Salomat Yusupovna
Buxoro Davlat Texnika Universiteti, dotsent.
Avezova Shaxnoza Maxmudjonovna
Buxoro Davlat Texnika Universiteti, dotsent.
Jalilov Jamshid G‘anijonovich
Toshkent davlat iqtisodiyot universiteti, dotsent
Xasanova Gulruh Djumanazarovna
Buxoro Davlat Texnika Universiteti, dotsent.

PEDAGOGIKA FANLARI/ PEDAGOGICAL SCIENCES / ПЕДАГОГИЧЕСКИЕ НАУКИ

Siddiqova Sadoqat G‘afforovna
Buxoro Davlat Texnika Universiteti, dotsent.
Zoyirov Erkin Xalilovich
Buxoro Davlat Texnika Universiteti, DSc, professor
Gaffarov Feruz Hasanovich
Puchon Universiteti, *professor*
Ruziyev Davron Yuldashevich
Buxoro Davlat Universiteti, professor
Muradova Firuza Rashidovna
Buxoro Davlat Texnika Universiteti, professor.
Yunusova Gulandon Samiyevna
Buxoro Davlat Texnika Universiteti, professor.
To‘xtayeva Zebo Sharifovna
Buxoro Davlat Texnika Universiteti, professor.
Navro‘zova Gulchixra Nigmatovna
Buxoro Davlat Texnika Universiteti, dotsent.
Uzoqova Laylo Polvonovna
Buxoro muhandislik – texnologiya instituti, dotsent.
To‘raqulova Marjona Qiyom qizi
Buxoro Davlat Texnika Universiteti, dotsent.
Muradova Zarina Rashidovna
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Sanjar Aslonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shodiyev Jahongir Jo‘raqulovich
Buxoro Davlat Texnika Universiteti, dotsent
Gaffarov Laziz Hasanovich
Buxoro muhandislik – texnologiya instituti, dotsent.

Tahririyat a’zolarning to‘liq ro‘yxati jurnalning www.devos.uz saytida keltirilgan.

Tahririyat hay’ati a’zolarining tarkibini o‘zgartirish hamda qo‘shimcha kiritish yoki tahririyat a’zoligidan chiqarish bosh muharrir tomonidan amalga oshiriladi.

ISSN 3030-3907

Development of science ilmiy jurnali 2024 yil 15 aprel kuni №256445 sonli guvohnoma bilan Davlat ro‘yxatidan o‘tgan.

To‘plam elektron ko‘rinishda (pdf formatida) mualliflarga taqdim etiladi. To‘plamga kiritiladigan maqolalarning mazmuni, ulardagi statistik ma’lumotlar va me’yoriy hujjatlarning to‘g‘riligiga hamda har qanday fikr - mulohazalarga mualliflarning o‘zlari mas’ul hisoblanadilar. Talablarga javob bermaydigan maqolalar to‘plamga kiritilmaydi. Tashkiliy qo‘mita maqola matnini qisqartirish, qisman tuzatish kiritish, seksiyalarga joylashtirish huquqiga ega.

ISSN 3030-3907

KATTA HAJMLI TA'LIM MA'LUMOTLARINI (BIG DATA) INTEGRATSIYALASH VA TAQSIMLANGAN MA'LUMOTLAR BAZALARINI OPTIMALLASHTIRISH

Xusenov Murodjon Zoxirovich
Buxoro davlat universiteti Axborot tizimlari va raqamli texnologiyalar kafedrasi dotsenti,
E-mail: m.z.xusenov@buxdu.uz,
ORCID: 0000-0002-1533-3102
Istamov Bexzod Bafoqulovich
Buxoro davlat universitetining "Kompyuter tizimlari va ularning dasturiy ta'minoti" ta'lim yo'nalishi magistranti,
E-mail: b.b.istamov@buxdu.uz
ORCID: 0009-0009-7499-9906)

Annotatsiya. ushbu maqolada ta'lim tizimida shakllanayotgan katta hajmli ma'lumotlarni yig'ish, integratsiyalash, saqlash, qayta ishlash va tahlil qilish masalalari yoritiladi. Elektron ta'lim platformalari, o'quv jarayonini boshqarish tizimlari, onlayn testlar, raqamli kutubxonalar, talabalarning faollik jurnallari, mobil ilovalar va ma'muriy axborot tizimlari ta'lim muassasalarida turli formatdagi ma'lumotlar oqimini hosil qilmoqda. Bunday ma'lumotlardan samarali foydalanish uchun integratsiya arxitekturasi, metama'lumotlar boshqaruvi, taqsimlangan ma'lumotlar bazalari, sharding, replikatsiya, indekslash, kesh, so'rovlarni optimallashtirish, xavfsizlik va boshqaruv tamoyillari kompleks yondashuvda ko'rib chiqilishi zarur. Maqolada Big Data tushunchasi, ta'lim ma'lumotlarining xususiyatlari, ETL/ELT jarayonlari, data lake va lakehouse yondashuvlari, Hadoop, Spark, Cassandra, MongoDB, PostgreSQL kabi texnologik asoslar, shuningdek taqsimlangan tizimlarda CAP, konsistensiya, mavjudlik va kechikish o'rtasidagi muvozanat masalalari tahlil qilinadi. Ish yakunida ta'lim muassasalari uchun amaliy arxitektura modeli va optimallashtirish bo'yicha tavsiyalar beriladi.

Kalit so'zlar: Big Data, ta'lim analitikasi, ma'lumotlar integratsiyasi, taqsimlangan ma'lumotlar bazasi, sharding, replikatsiya, Hadoop, Spark, Cassandra, MongoDB, PostgreSQL, data lake, ETL, ELT, optimallashtirish.

Abstract. This article examines the collection, integration, storage, processing, and analysis of large volumes of data emerging in the education system. E-learning platforms, learning management systems, online tests, digital libraries, student activity logs, mobile applications, and administrative information systems are generating flows of data in various formats within educational institutions. To use such data effectively, integration architecture, metadata management, distributed databases, sharding, replication, indexing, caching, query

optimization, security, and governance principles must be considered through a comprehensive approach. The article analyzes the concept of Big Data, the characteristics of educational data, ETL/ELT processes, data lake and lakehouse approaches, technological foundations such as Hadoop, Spark, Cassandra, MongoDB, and PostgreSQL, as well as issues related to CAP, consistency, availability, and the balance between latency in distributed systems. At the end of the study, a practical architecture model and optimization recommendations are provided for educational institutions.

Keywords: Big Data, educational analytics, data integration, distributed database, sharding, replication, Hadoop, Spark, Cassandra, MongoDB, PostgreSQL, data lake, ETL, ELT, optimization.

Аннотация. В данной статье рассматриваются вопросы сбора, интеграции, хранения, обработки и анализа больших объемов данных, формирующихся в системе образования. Электронные образовательные платформы, системы управления учебным процессом, онлайн-тесты, цифровые библиотеки, журналы активности студентов, мобильные приложения и административные информационные системы создают в образовательных учреждениях потоки данных различных форматов. Для эффективного использования таких данных необходимо комплексно рассматривать архитектуру интеграции, управление метаданными, распределенные базы данных, шардинг, репликацию, индексирование, кэширование, оптимизацию запросов, безопасность и принципы управления данными. В статье анализируются понятие Big Data, особенности образовательных данных, процессы ETL/ELT, подходы data lake и lakehouse, технологические основы Hadoop, Spark, Cassandra, MongoDB, PostgreSQL, а также вопросы CAP, согласованности, доступности и баланса между задержкой в распределенных системах. В заключение предлагается практическая архитектурная модель и рекомендации по оптимизации для образовательных учреждений.

Ключевые слова: Big Data, образовательная аналитика, интеграция данных, распределенная база данных, шардинг, репликация, Hadoop, Spark, Cassandra, MongoDB, PostgreSQL, data lake, ETL, ELT, оптимизация.

KIRISH.

Raqamli transformatsiya ta’lim sohasida ma’lumotlar hajmi va xilma-xilligini keskin oshirdi. Ilgari ta’lim jarayoni asosan qog‘oz hujjatlar, jurnal va imtihon natijalari bilan cheklangan bo‘lsa, bugungi kunda o‘quvchi yoki talabaning har bir raqamli harakati ma’lumotga aylanmoqda: platformaga kirish vaqti, topshiriqni bajarish tezligi, test savollariga javob berish ketma-ketligi,

ISSN 3030-3907

video darsni ko‘rish davomiyligi, forumdagi faollik, baholash natijalari, davomad, kutubxona resurslaridan foydalanish va boshqa ko‘rsatkichlar.

Katta hajmli ta’lim ma’lumotlari strategik qiymatga ega. Ular yordamida individual ta’lim trayektoriyalarini shakllantirish, talabalar xavf guruhini erta aniqlash, o‘qitish metodikasini takomillashtirish, o‘quv dasturlarini baholash, resurslardan foydalanishni rejalashtirish, ta’lim sifatini monitoring qilish va boshqaruv qarorlarini dalillarga asoslash mumkin. Biroq bu imkoniyatlar ma’lumotlar to‘g‘ri integratsiyalangan, sifatli, xavfsiz va tezkor ishlov berishga tayyor bo‘lgandagina yuzaga chiqadi.

Muammo shundaki, ta’lim muassasalarida ma’lumotlar ko‘pincha alohida tizimlarda saqlanadi: LMS, SIS, elektron jurnal, test tizimi, HR, moliya, kutubxona, ilmiy faoliyat bazasi, masofaviy ta’lim platformasi va tashqi sertifikatlash tizimlari. Bu tizimlarning ma’lumot modeli, identifikatori, formatlari va yangilanish davri turlicha bo‘ladi. Natijada ma’lumotlarni yagona tahliliy muhitga keltirish texnik, metodologik va tashkiliy masalaga aylanadi.

Ushbu monografiyaning maqsadi katta hajmli ta’lim ma’lumotlarini integratsiyalash va taqsimlangan ma’lumotlar bazalarini optimallashtirish bo‘yicha nazariy hamda amaliy yondashuvlarni tizimli ko‘rib chiqishdan iborat.

Katta hajmli ta’lim ma’lumotlarining mohiyati

Big Data odatda hajm, tezlik, xilma-xillik, ishonchlilik va qiymat kabi mezonlar orqali tavsiflanadi. Ta’lim muhitida bu mezonlar quyidagicha namoyon bo‘ladi.

Hajm: yirik universitet yoki milliy ta’lim platformasi millionlab foydalanuvchi harakatlari, test javoblari, videodars ko‘rish loglari va baholash natijalarini saqlaydi.

Tezlik: real vaqtga yaqin monitoring, adaptiv test, onlayn proktoring va o‘quvchining faoliyatiga darhol javob beruvchi tizimlar tezkor qayta ishlashni talab qiladi.

Xilma-xillik: ta’lim ma’lumotlari jadval, matn, video, audio, JSON, log, graf, rasm va hujjat ko‘rinishida bo‘lishi mumkin.

Ishonchlilik: baholar, davomad, sertifikatlar va shaxsiy ma’lumotlar xatolarsiz bo‘lishi zarur, chunki ular huquqiy va akademik oqibatlarga ega.

Qiymat: ma’lumotlar faqat saqlansa emas, balki ta’lim sifatini yaxshilash, qaror qabul qilish va individual yordam berish uchun ishlatilsa qiymat yaratadi.

Ta’lim ma’lumotlari tabiatan ko‘p manbali va kontekstga bog‘liq. Masalan, test natijasi faqat ball emas; u topshiriq murakkabligi, urinishlar soni, vaqt, mavzu, o‘qituvchi, guruh, oldingi bilim darajasi bilan birga talqin qilinishi kerak. Shuning uchun integratsiya jarayoni oddiy fayl birlashtirish emas, balki semantik moslashtirish, sifat nazorati va ma’lumotlar boshqaruvini ham o‘z ichiga oladi.

ISSN 3030-3907

Ta’lim ma’lumotlarini integratsiyalash tamoyillari

Ma’lumotlar integratsiyasi - turli manbalardan kelgan ma’lumotlarni yagona, izchil va qayta ishlashga tayyor shaklga keltirish jarayonidir. Ta’lim muhitida integratsiya quyidagi bosqichlarni qamrab oladi:

1. Manbalarni aniqlash: LMS, elektron jurnal, test platformasi, kutubxona, HR, moliya, kontingent va tashqi APIlar.
2. Identifikatorlarni moslashtirish: talaba ID, login, pasport yoki ichki kodlar o‘rtasida yagona kalit yaratish.
3. Sxemani moslashtirish: turli tizimlardagi maydon nomlari, tiplari va kodlarini yagona modelga keltirish.
4. Sifatni nazorat qilish: bo‘sh qiymatlar, dublikatlar, noto‘g‘ri sanalar, format xatolari va nomuvofiq kodlarni aniqlash.
5. Transformatsiya: normalizatsiya, agregatsiya, kategoriyalash, anonimlashtirish va hisoblangan ko‘rsatkichlar yaratish.
6. Saqlash: ma’lumotlar ombori, data lake, lakehouse yoki operatsion taqsimlangan baza.
7. Kuzatuvchanlik: ma’lumot qayerdan kelgani, qachon o‘zgargani va qanday transformatsiyadan o‘tganini qayd etish.

An’anaviy ETL yondashuvida ma’lumot avval olinadi, transformatsiya qilinadi va keyin omborga yuklanadi. ELT yondashuvida esa xom ma’lumot avval saqlash qatlamiga yuklanadi, transformatsiya keyin hisoblash infratuzilmasida bajariladi. Ta’lim Big Data muhitida ELT ko‘pincha qulay, chunki xom loglar, matnlar va yarim tuzilgan JSON ma’lumotlarini yo‘qotmasdan saqlab, keyinchalik turli tahliliy ehtiyojlar uchun qayta ishlash mumkin.

Integratsiyada eng muhim masalalardan biri metama’lumotlar boshqaruvidir. Har bir jadval, ustun, fayl, ko‘rsatkich va transformatsiya qoidasi izohlanmasa, tizim vaqt o‘tishi bilan tushunarsiz bo‘lib qoladi. Shu sababli data catalog, data lineage va data dictionary kabi vositalar ta’lim analitikasi infratuzilmasining ajralmas qismiga aylanishi kerak.

Ta’lim Big Data arxitekturasi

Katta hajmli ta’lim ma’lumotlari uchun tavsiya etiladigan arxitektura bir necha qatlamdan iborat:

Ma’lumot manbalari qatlami. Bu yerda LMS, SIS, MOOC platformalari, test tizimlari, elektron kutubxonalar, mobil ilovalar, video platformalar va IoT qurilmalar mavjud bo‘ladi.

Yig‘ish qatlami. Batch yuklash, API orqali olish, xabarlar navbati, streaming va fayl almashinuvi mexanizmlari ishlatiladi.

ISSN 3030-3907

Saqlash qatlami. Xom ma’lumotlar data lake muhitida, tozalangan va modellashtirilgan ma’lumotlar data warehouse yoki lakehouse muhitida, operatsion tezkor ma’lumotlar esa taqsimlangan bazalarda saqlanadi.

Qayta ishlash qatlami. Apache Spark, Hadoop MapReduce, SQL dvigatellari, oqimli qayta ishlash platformalari va mashinaviy o‘qitish kutubxonalari qo‘llanadi.

Xizmat ko‘rsatish qatlami. Dashboard, API, hisobot, tavsiya tizimi, ogohlantirish moduli va ilmiy-tadqiqot muhitlari shu qatlamga kiradi.

Boshqaruv va xavfsizlik qatlami. Autentifikatsiya, avtorizatsiya, audit, shifrlash, anonimlashtirish, zaxira nusxa, monitoring va siyosatlar bilan ishlaydi.

NIST Big Data Reference Architecture yondashuvi bunday tizimlarda rollar, komponentlar va o‘zaro bog‘lanishlarni standartlash uchun foydali nazariy asos beradi. Ta’lim muassasasi ushbu modeldan foydalanib, ma’lumot provayderi, ma’lumot istemolchisi, tizim orchestratori, xavfsizlik va boshqaruv funksiyalarini alohida aniqlashi mumkin.

Taqsimlangan ma’lumotlar bazalarining nazariy asoslari

Taqsimlangan ma’lumotlar bazasi - ma’lumotlarni bir nechta server yoki tugunlarda saqlaydigan, lekin foydalanuvchiga mantiqan yagona tizim sifatida ko‘rinadigan ma’lumotlar boshqaruvi tizimidir. Katta hajmli ta’lim ma’lumotlari uchun bunday yondashuv zarur, chunki bitta server hajm, tezlik va mavjudlik bo‘yicha yetarli bo‘lmasligi mumkin.

Taqsimlangan bazalarning asosiy tushunchalari quyidagilar:

Fragmentatsiya: ma’lumotlarni bo‘laklarga ajratish. Gorizontal fragmentatsiyada satrlar, vertikal fragmentatsiyada ustunlar bo‘linadi.

Sharding: ma’lumotlarni shard deb ataluvchi mustaqil bo‘laklarga taqsimlash. Masalan, talabalar ma’lumotini hudud, fakultet yoki hash kalit bo‘yicha bo‘lish mumkin.

Replikatsiya: bir xil ma’lumot nusxalarini bir nechta tugunda saqlash. Bu mavjudlik va o‘qish tezligini oshiradi.

Konsistensiya: turli nusxalarda ma’lumotning mosligi. Kuchli konsistensiya aniqlikni oshiradi, ammo kechikishni ko‘paytirishi mumkin.

Partitsiya bardoshliligi: tarmoq uzilishi yoki tugunlar o‘rtasidagi aloqa muammosida tizim ishlashda davom etishi.

CAP teoremasi taqsimlangan tizimlarda konsistensiya, mavjudlik va partitsiyaga bardoshlilik o‘rtasidagi muvozanatni tushuntiradi. Ta’lim tizimida bu tanlov amaliy ahamiyatga ega. Masalan, baholar va sertifikatlar uchun kuchli konsistensiya muhim, lekin video ko‘rish loglari yoki forum faolligi uchun vaqtinchalik nomuvofiqlik qabul qilinishi mumkin.

Ma’lumotlarni taqsimlash strategiyalari

ISSN 3030-3907

Taqsimlangan bazani optimallashtirishning birinchi bosqichi – to‘g‘ri taqsimlash strategiyasini tanlash. Noto‘g‘ri shard kaliti tizimda “issiq nuqta” hosil qiladi: ayrim tugunlarga haddan tashqari yuk tushadi, boshqalari esa bo‘sh qoladi. Ta’lim ma’lumotlari uchun quyidagi strategiyalar qo‘llanilishi mumkin:

Hash asosida taqsimlash. Talaba ID yoki foydalanuvchi ID hash qilinadi va ma’lumot shardlarga teng taqsimlanadi. Bu yukni balanslash uchun yaxshi, lekin hudud yoki vaqt bo‘yicha tahlilda qo‘shimcha xarajat tug‘dirishi mumkin.

Diapazon asosida taqsimlash. Ma’lumotlar sana, kurs yili yoki hudud bo‘yicha bo‘linadi. Vaqtli tahlillar uchun qulay, ammo ayrim davrlar juda faol bo‘lsa, balans buzilishi mumkin.

Ro‘yxat asosida taqsimlash. Fakultet, ta’lim turi yoki hudud bo‘yicha bo‘lish mumkin. Boshqaruv hisobotlari uchun tushunarli, lekin yirik fakultetlar shardni ortiqcha yuklashi ehtimoli bor.

Kompozit kalit. Masalan, region_id + student_hash yoki year + course_id + hash kabi kombinatsiyalar balans va tahliliy qulaylikni birlashtiradi.

Ta’lim tizimida odatda vaqt va foydalanuvchi kalitlarini birlashtirish maqsadga muvofiq. Masalan, log ma’lumotlari oy bo‘yicha partitsiyalanadi, har bir oy ichida foydalanuvchi hash bo‘yicha bo‘linadi. Bu eski ma’lumotlarni arxivlash, yangi ma’lumotlarni tez yozish va vaqtli so‘rovlarni tezlashtirishga yordam beradi.

So‘rovlarni optimallashtirish

Taqsimlangan ma’lumotlar bazalarida so‘rov samaradorligi faqat indeksga bog‘liq emas; u ma’lumot qayerda joylashgani, so‘rov necha shardga borishi, tarmoq orqali qancha ma’lumot qaytishi va agregatsiya qayerda bajarilishiga ham bog‘liq.

Asosiy optimallashtirish usullari:

Indekslash. Ko‘p ishlatiladigan filtr, join va tartiblash ustunlari indekslanadi. Lekin har bir indeks yozish xarajatini oshiradi, shuning uchun indekslar real so‘rov profiliga asoslanishi kerak.

Partition pruning. Agar jadval vaqt, hudud yoki boshqa kalit bo‘yicha partitsiyalangan bo‘lsa, so‘rov faqat kerakli partitsiyalarni o‘qishi kerak. PostgreSQL hujjatlarida bunday pruning rejalashtirish va bajarish vaqtida amalga oshishi mumkinligi ko‘rsatiladi.

Denormalizatsiya. O‘qish tezligi muhim bo‘lgan analitik tizimlarda ayrim ma’lumotlarni oldindan birlashtirib saqlash foydali. Masalan, talabaning kurs, guruh va fakultet nomini har safar join qilmasdan analitik jadvalda saqlash mumkin.

ISSN 3030-3907

Materialized view. Ko‘p takrorlanadigan hisobotlar uchun oldindan hisoblangan ko‘rinishlar yaratiladi: kunlik faollik, kurs bo‘yicha o‘rtacha ball, xavf guruhidagi talabalar soni.

Predicate pushdown. Filtrlash imkon qadar ma’lumot joylashgan qatlamda bajariladi. Bu tarmoq orqali ortiqcha ma’lumot uzatilishini kamaytiradi.

Kesh. Tez-tez so‘raladigan dashboard natijalari, statistik ko‘rsatkichlar yoki tavsiya natijalari keshda saqlanishi mumkin.

So‘rov rejalarini tahlil qilish. SQL tizimlarda EXPLAIN va ANALYZE yordamida qaysi indeks ishlatilgani, qaysi jadval to‘liq o‘qilgani, join tartibi va xarajatlar tekshiriladi.

Ta’lim tizimida optimallashtirish amaliy maqsadga bog‘liq bo‘lishi kerak. Masalan, real vaqt ogohlantirish tizimida kechikish eng muhim; semestr yakuni hisobotida esa aniqlik va to‘liq qamrov muhimroq.

Replikatsiya va ishonchlilik

Replikatsiya taqsimlangan bazaning muhim mexanizmidir. U ikki asosiy vazifani bajaradi: tizim ishdan chiqishlarga bardoshli bo‘ladi va o‘qish so‘rovlari bir nechta nusxalar orqali tezroq bajariladi.

Ta’lim ma’lumotlari turiga qarab replikatsiya siyosati farqlanishi mumkin: Baholar, diplom va sertifikat ma’lumotlari: kuchliroq konsistensiya, qat’iy audit va zaxira nusxa talab etiladi.

Faollik loglari: yuqori yozish tezligi va hajm muhim, vaqtinchalik kechikkan konsistensiya qabul qilinishi mumkin.

Dashboard agregatlari: qayta hisoblash mumkin bo‘lganligi sababli kesh va materialized view yetarli bo‘lishi mumkin.

Shaxsiy ma’lumotlar: shifrlash, kirish nazorati, minimal saqlash va anonimlashtirish muhim.

Replikatsiya faktorini tanlashda saqlash xarajati, tarmoq trafigi, tiklanish vaqti va ma’lumot yo‘qotish xavfi hisobga olinadi. Masalan, uch nusxali replikatsiya keng tarqalgan, chunki bitta tugun ishdan chiqsa ham tizim ishlashda davom etadi.

Hadoop, Spark va data lake yondashuvi

Hadoop ekotizimi katta hajmli ma’lumotlarni arzon serverlar klasterida saqlash va qayta ishlash g‘oyasini ommalashtirdi. HDFS fayllarni bloklarga bo‘lib, DataNode tugunlarida saqlaydi, NameNode esa metama’lumotlarni boshqaradi. Bu yondashuv yirik fayllar va batch ishlov berish uchun qulay.

Apache Spark esa xotira asosida tezkor qayta ishlash, SQL, streaming, graph va machine learning kutubxonalari bilan keng qo‘llanadi. Ta’lim analitikasida Spark quyidagi vazifalarda foydali:

- Katta loglarni tozalash va agregatsiya qilish.

ISSN 3030-3907

- Talabalar faolligi bo‘yicha xususiyatlar yaratish.
- Xavf guruhini bashoratlash uchun mashinaviy o‘qitish.

Reyting va tavsiya tizimlari.

Data lake yondashuvi turli formatdagi xom ma’lumotlarni yagona saqlash muhitida jamlaydi. Uning afzalligi - ma’lumot avvaldan qat’iy sxemaga solinmasdan saqlanadi. Kamchiligi - metama’lumotlar va sifat boshqaruvi yetarli bo‘lmasa, data lake “data swamp”, ya’ni tartibsiz ma’lumot omboriga aylanib qolishi mumkin.

Lakehouse yondashuvi data lake moslashuvchanligi va data warehouse boshqariluvchanligini birlashtirishga intiladi. Ta’lim muassasalari uchun bu qulay model: xom loglar saqlanadi, tozalangan ma’lumotlar jadval sifatida boshqariladi, analitik SQL va ML jarayonlari bir infratuzilmada bajariladi.

NoSQL va NewSQL yondashuvlari

Ta’lim Big Data tizimlarida barcha ma’lumotlarni bitta klassik relatsion bazada saqlash har doim ham samarali emas. Ma’lumot turiga qarab turli bazalar tanlanadi.

MongoDB hujjatga yo‘naltirilgan baza bo‘lib, yarim tuzilgan ma’lumotlar uchun qulay. Masalan, test savollari, kurs kontenti, foydalanuvchi sozlamalari yoki turli strukturadagi anketa javoblarini hujjat sifatida saqlash mumkin. MongoDB sharding orqali juda katta dataset va yuqori throughput talablarini qo‘llab-quvvatlaydi.

Cassandra yuqori yozish tezligi, gorizontal kengayish va ko‘p tugunli replikatsiya talab qilinganda foydali. Talaba faoliyati loglari, vaqt qatorlari, klikstream va sensor ma’lumotlari uchun mos keladi. Cassandra ma’lumot modeli so‘rovlar asosida oldindan loyihalanadi: avval qanday o‘qish kerakligi aniqlanadi, keyin jadval tuziladi.

PostgreSQL kuchli SQL, tranzaksiya, indeks, partitsiyalash va kengaytmalar ekotizimi bilan ta’limning asosiy ma’muriy ma’lumotlari uchun qulay. Katta jadvallarni partitsiyalash, indekslarni to‘g‘ri tanlash va materialized view orqali analitik yuklarni yengillashtirish mumkin.

NewSQL tizimlar relatsion model va SQL qulayligini taqsimlangan kengayuvchanlik bilan birlashtirishga harakat qiladi. Ular kuchli konsistensiya va gorizontal kengayish talab qilingan holatlarda ko‘rib chiqilishi mumkin.

Xavfsizlik, maxfiylik va etik masalalar

Ta’lim ma’lumotlari ko‘pincha shaxsiy va sezgir hisoblanadi. Shuning uchun Big Data tizimining muvaffaqiyati faqat texnik tezlik bilan emas, balki ishonch, qonuniylik va etik boshqaruv bilan ham belgilanadi.

Asosiy talablar:

ISSN 3030-3907

Minimal yig‘ish: faqat zarur ma’lumotlar yig‘iladi.

Maqsadga muvofiqlik: ma’lumotlar oldindan belgilangan ta’limiy yoki boshqaruv maqsadlarida ishlatiladi.

Ruxsat va rollar: o‘qituvchi, dekanat, administrator, tadqiqotchi va talaba uchun kirish darajalari farqlanadi.

Shifrlash: ma’lumotlar uzatishda va saqlashda himoyalanadi.

Anonimlashtirish va psevdonimlashtirish: ilmiy tahlil uchun shaxsni bevosita aniqlovchi ma’lumotlar ajratiladi yoki yashiriladi.

Audit: kim, qachon, qaysi ma’lumotni ko‘rgani va o‘zgartirgani qayd qilinadi.

Model adolatliligi: bashoratlovchi algoritmlar talabalarni noto‘g‘ri tamg‘alamasligi, ijtimoiy yoki demografik noxolislikni kuchaytirmasligi kerak.

Ta’limda analitika yordam beruvchi vosita bo‘lishi lozim; u o‘qituvchi qarorini butunlay almashtirmasligi kerak. Ayniqsa xavf guruhini aniqlash tizimlari insoniy maslahat, qo‘llab-quvvatlash va shaffof mezonlar bilan birga ishlatilishi zarur.

Optimallashtirish metodikasi

Taqsimlangan ma’lumotlar bazasini optimallashtirish bir martalik sozlama emas, doimiy jarayondir. Quyidagi metodika tavsiya etiladi:

1. Ish yukini o‘lchash. Eng ko‘p ishlatiladigan so‘rovlar, yozish tezligi, o‘qish tezligi, kechikish, xotira va disk sarfi aniqlanadi.

2. Ma’lumot modelini baholash. Jadval, hujjat, partitsiya va shard kalitlari real so‘rovlar bilan mosligi tekshiriladi.

3. Indeks siyosatini qayta ko‘rish. Keraksiz indekslar olib tashlanadi, zarur indekslar qo‘shiladi.

4. Taqsimot balansini tekshirish. Shardlar o‘rtasida ma’lumot hajmi va so‘rov yuki tengmi yoki yo‘qmi aniqlanadi.

5. Agregatsiyalarni oldindan hisoblash. Dashboard va hisobotlar uchun materialized view yoki agregat jadvallar yaratiladi.

6. Monitoring va alert. CPU, RAM, disk I/O, tarmoq, query latency, replication lag va error rate kuzatiladi.

7. Sinov. Katta yuk ostida stress-test va benchmark o‘tkaziladi.

8. Zaxira va tiklash mashqi. Backup borligi yetarli emas; restore jarayoni ham muntazam tekshiriladi.

Ta’lim tizimlarida semestr boshida, oraliq nazorat davrida va yakuniy imtihon paytida yuk keskin oshishi mumkin. Shuning uchun capacity planning akademik kalendar bilan bog‘langan bo‘lishi kerak.

Amaliy arxitektura modeli

Quyidagi model o‘rta yoki yirik ta’lim muassasasi uchun tavsiya etiladi:

ISSN 3030-3907

Operatsion qatlam:

PostgreSQL - talaba, guruh, fan, baho, davomad, o‘quv reja kabi tranzaksion ma’lumotlar.

MongoDB - kurs kontenti, savol banki, anketa, yarim tuzilgan hujjatlar.

Cassandra yoki shunga o‘xshash kengayuvchan NoSQL - faollik loglari, vaqt qatorlari, klikstream.

Integratsiya qatlami:

- API gateway va message queue orqali tizimlar ulanishi.
- Batch import va streaming import jarayonlari.
- Yagona identifikator xizmati.
- Saqlash va tahlil qatlami:
- Data lake - xom fayllar, loglar, eksportlar.

Lakehouse yoki warehouse - tozalangan, sifat nazoratidan o‘tgan jadval ma’lumotlari.

Spark - transformatsiya, agregatsiya va ML vazifalari.

Xizmat qatlami:

- Dashboard: davomad, o‘zlashtirish, xavf ko‘rsatkichlari.
- Tavsiya tizimi: individual resurs va topshiriqlar.
- Ogohlantirish: past faollik, muddatdan kechikish, bahoning keskin pasayishi.

Boshqaruv qatlami:

Data catalog, lineage, audit, rolga asoslangan kirish, shifrlash, zaxira nusxa.

Bu model modulli bo‘lishi kerak. Barcha tizimni birdan joriy qilish shart emas; avval eng muhim manbalar va aniq biznes savollar tanlanadi. Masalan, birinchi bosqichda LMS + elektron jurnal + baholar integratsiya qilinadi, keyin kutubxona va video platforma qo‘shiladi.

XULOSA.

Katta hajmli ta’lim ma’lumotlarini integratsiyalash va taqsimlangan ma’lumotlar bazalarini optimallashtirish zamonaviy ta’lim boshqaruvi uchun strategik ahamiyatga ega. Bunday tizimlar o‘quv jarayonini chuqur tahlil qilish, individual yondashuvni kuchaytirish, xavf guruhidagi talabalarni erta aniqlash, resurslarni oqilona taqsimlash va ta’lim sifatini oshirish imkonini beradi.

Maqolada ko‘rsatilganidek, muvaffaqiyatli yechim faqat texnologiya tanlash bilan cheklanmaydi. Ma’lumotlar sifati, yagona identifikator, metama’lumotlar, xavfsizlik, etik tamoyillar, monitoring va doimiy optimallashtirish birgalikda ishlashi zarur. Taqsimlangan bazalarda shard kaliti, partitsiya, indeks, replikatsiya, konsistensiya va kesh mexanizmlari tizim samaradorligini belgilaydi.

ISSN 3030-3907

Ta’lim muassasalari uchun eng maqbul yondashuv bosqichma-bosqich integratsiya, aniq maqsadli analitika va moslashuvchan arxitekturadir. Bunda operatsion bazalar, data lake/lakehouse, Spark kabi qayta ishlash platformalari va kuchli boshqaruv siyosatlari yagona ekotizim sifatida loyihalanishi kerak. Kelgusida sun’iy intellekt, real vaqt analitikasi va semantik ma’lumot boshqaruvi ta’lim Big Data tizimlarining asosiy rivojlanish yo‘nalishlariga aylanishi kutiladi.

FOYDALANILGAN ADABIYOTLAR

1. NIST. Big Data Interoperability Framework: Volume 6, Reference Architecture.https://www.nist.gov/publications/nist-big-data-interoperability-framework-volume-6-reference-architecture
2. NIST. NBDIF Version 2.0 Final. https://www.nist.gov/itl/big-data-nist/big-data-nist-documents/nbdif-version-20-final
3. Apache Hadoop. HDFS Architecture. https://hadoop.apache.org/docs/r2.10.2/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html
4. Apache Hadoop. HDFS Users Guide. https://hadoop.apache.org/docs/r3.3.5/hadoop-project-dist/hadoop-hdfs/HdfsUserGuide.html
5. Apache Spark. Documentation. https://spark.apache.org/documentation
6. Apache Spark. Cluster Mode Overview. https://spark.apache.org/docs/latest/cluster-overview.html
7. Apache Cassandra. Architecture Overview. https://cassandra.apache.org/doc/latest/cassandra/architecture/overview.html
8. MongoDB. Sharding. https://www.mongodb.com/docs/current/sharding/
9. PostgreSQL. Table Partitioning. https://www.postgresql.org/docs/current/ddl-partitioning.html
10. Chang, F., Dean, J., Ghemawat, S. et al. Bigtable: A Distributed Storage System for Structured Data. Google Research. https://research.google/pubs/pub27898
11. Bigtable PDF. https://research.google.com/archive/bigtable-osdi06.pdf
12. DeCandia, G. et al. Dynamo: Amazon’s Highly Available Key-value Store. https://web.stanford.edu/class/cs244/papers/amazon-dynamo-sosp2007.pdf
13. Srivastava, D. Big Data Integration. Proceedings of the VLDB Endowment. https://www.vldb.org/pvldb/vol6/p1188-srivastava.pdf
14. Nargesian, F. et al. Data Lake Management: Challenges and Opportunities. PVLDB. https://www.vldb.org/pvldb/vol12/p1986-nargesian.pdf
15. Giebler, C. et al. Data Lakes: A Survey of Concepts and Architectures. Computers, MDPI. https://www.mdpi.com/2073-431X/13/7/183
16. Hai, R. et al. Data Lakes: A Survey of Functions and Systems. arXiv. https://arxiv.org/abs/2106.09592

ISSN 3030-3907

17. Brewer, E. Towards Robust Distributed Systems. CAP keynote reference. https://people.eecs.berkeley.edu/~brewer/cs262b-2004/PODC-keynote.pdf
18. Kleppmann, M. A Critique of the CAP Theorem. University of Cambridge. https://www.cl.cam.ac.uk/research/dtg/archived/files/publications/public/mk428/cap-critique.pdf
19. Abadi, D. PACELC design principle reference. https://www.cs.umd.edu/~abadi/papers/abadi-pacelc.pdf
20. Zaharia, M. et al. Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster
Computing. https://www.usenix.org/system/files/conference/nsdi12/nsdi12-final138.pdf

ISSN 3030-3907