



BIR KOMPYUTERDA KATTA MA'LUMOTLAR BILAN ISHLASH

Qodirjonova Nilufar Abduvohidjon qizi
Tursunova Nafosatxon G'ayratjon qizi
Parpiboyev Nasimjon Asqarali o'g'li

Farg'ona davlat universiteti talabalari

Israil Tojimamatov

Farg'ona davlat universiteti o'qituvchisi

<https://doi.org/10.5281/zenodo.7830510>

ARTICLE INFO

Qabul qilindi: 05-april 2023 yil

Ma'qullandi: 10-april 2023 yil

Nashr qilindi: 15-april 2023 yil

KEY WORDS

Big data, inson, analiz, hajm, axborot, tizim, operatsiya.

ABSTRACT

Big data (katta ma'lumotlar) – juda katta hajmdagi bir jinsli bo'lmagan va tez tushadigan raqamli ma'lumotlar bo'lib, ularni odatiy usullar bilan qayta ishlab bo'lmaydi. Ba'zi hollarda, katta ma'lumotlar tushunchasi bilan birga shu ma'lumotlarni qayta ishlash ham tushuniladi. Asosan, analiz obyekti katta ma'lumotlar deb ataladi. Big data atamasi 2008-yilda dunyoga kelgan. Nature jurnali muharriri Klifford Linch dunyo ma'lumotlar hajmining juda tez sur'atda o'sishiga bag'ishlangan maxsus sonida big data atamasini qo'llagan. Biroq, katta ma'lumotlar avval ham bo'lgan. Mutaxassislarning fikricha, kuniga 100 gb dan ko'p ma'lumot tushadigan oqimlarga big data deb aytilar ekan. Ushbu maqolada kompyuterda Katta ma'lumotlar bilan ishlash haqida ma'lumotlar, Katta ma'lumotlar bilan ishlash tahlili hamda usullari haqida so'z yuritiladi.

Katta ma'lumotlarni analiz qilish, inson his etish imkoniyatidan tashqarida bo'lgan qonuniylatlarni aniqlashda yordam beradi. Bu esa kundalik hayotimizdagi barcha sohalarda, hukumatni boshqarish, tibbiyot, telekommunikatsiya, moliya, transport, ishlab chiqarish va boshqa sohalarni yanada yaxshilash, ularning imkoniyatlarini oshirish, muommolarga muqobil yechimlar izlab topish imkonini yaratadi.

Zamonaviy axborot tizimlarida foydalaniladigan axborot (ayniqsa video, audiomultimediali ma'lumotlar) hajmini jadal o'sishi oldimizga ushbu axborotni ma'lumotlar bazasida saqlash va ularni boshqarishning yangi murakkab masalalarni qo'yadi. Quyida katta hajmdagi multimedia ma'lumotlarni tashkil qilishga, shuningdek, ular bilan ishlash vositalari tahlil qilib chiqamiz. Katta hajmli multimedia ma'lumotlar to'rtta asosiy xarakteristikaga ega bo'ladi: hajm, xilma-xillik, tezlik, narxi.

1. Hajm. Insonlar va mashinalar tomonidan xosil qilinayotgan ma'lumotlarning o'sib borayotgan soni, axborot texnologiyalari infrastrukturalariga ma'lumotlarni saqlash, ishlov berish va taqdim etishida yangi talablarni qo'yadi.

2. Xilma-xillik. Turli strukturalarda taqdim etilgan ma'lumotlar xilma-xil bo'ladi. Bular kredit kartalar bo'yicha operatsiyalar bo'ladimi, ilmiy tadqiqot natijalari bo'ladimi, fotografik

suratlar bo'ladimi, video va audio ma'lumotlarbo'ladimi barchasi o'ziga xos ishlov va saqlash shart sharoitlarini talab etadi.

3. Tezlik. Tezlik deganda nafaqat ma'lumotlarning ma'lumotlar bazasiga kelibtushish tezligi, balki bu ma'lumotlar bazasidan olinadigan ma'lumotlarni chiqariblish tezligi ham anglanadi.

4. Narxi. Katta hajmdagi ma'lumotlar – qimmat resurs hisoblanadi. Ma'lumotdolzarliligi, foydaliligi va ichidagi kontentiga qarab uning qanchalik qimmat baholigi aniqlanadi.

Hozirgi vaqtda ma'lumotlarning hajmi keskin o'sib bormoqda. Bozor o'zgarishlariga tezda javob berish, raqobatdosh ustunliklarga ega bo'lish va ishlab chiqarish samaradorligini oshirish uchun juda katta hajmdagi ma'lumotlarni olish, qayta ishlash va tahlil qilish kerak. Bunday hajmdagi ma'lumotlar bilan ishlash uchun muhandislar barcha ma'lumotlarni tahlil qilish ustida ishlash vositalarini modernizatsiya qilishga majbur bo'lishdi. Shunday qilib, 2000-yillarda BigData kontseptsiyasi shakllandi, bu faqat tor doiradagi mutaxassislar uchun qiziqarli edi. Endi bu so'zni axborot texnologiyalari sohasiga qiziqqan har bir kishi eshitadi. BigData texnologiyalari sizga katta hajmdagi ma'lumotlarni qayta ishlash, ularni tizimlashtirish, tahlil qilish va inson miyasi ularni hech qachon sezmaydigan naqshlarni aniqlash imkonini beradi. Bu ma'lumotlardan foydalanish uchun mutlaqo yangi imkoniyatlarni ochadi. BigData kontseptsiyasi nafaqat ma'lumotlarning katta qatlamlarini anglatadi. Bular yuzlab gigabaytlik va hatto petabaytlik ma'lumotlardan iborat ulkan saqlangan va qayta ishlangan massivlardir. Ulardan foydali ma'lumotlarni qayta ishlash va olish mumkin bo'lgan ma'lumotlar. Xulosa qilib aytganda, BigData ma'lumot olish uchun axborotni qayta ishlash texnologiyalari to'plami sifatida belgilanishi mumkin. Shuni ta'kidlash kerakki, BigData orqali qayta ishlangan ma'lumotlar hajmi, shuningdek, uni qayta ishlash tezligi doimiy ravishda o'sib bormoqda. Ushbu yo'nalishning rivojlanishi zamonaviy dunyoga juda mos keladi, tezkor va innovatsiondir. BigData ning rivojlanishi bilan texnologiya rivojlandi va aksincha. BigData-dan foydalanish bugungi kunda yirik IT-kompaniyalar rivojlanishining zaruriy shartiga aylanib bormoqda. Foydalanuvchilarning xatti-harakatlarini tahlil qilmasdan, bashorat qilish imkoniyatisiz, faqat tajriba va sezgi asosida, raqobatbardosh bo'lib qolish juda qiyin. Sozlangan va ishlaydigan BigData tizimi bir necha soniya ichida kompaniya mijozlarining milliardlab harakatlari tahlili natijasida olingan qimmatli ma'lumotlarni taqdim etishga qodir. Hozirgi biznesda ma'lumotlarga asoslangan boshqaruv kontseptsiyasi allaqachon paydo bo'lgan, bu kompaniyani qat'iy ma'lumotlar tahliliga asoslangan boshqarishni nazarda tutadi. Va bunday boshqaruv usullari ajoyib natijalarni ko'rsatadi. Facebook, Google, Mail.ru, Yandex uzoq vaqtdan beri qaror qabul qilish uchun tahliliy vositalardan foydalanib kelmoqda. Bugungi kunda an'anaviy biznes ham BigData ga qiziqish bildirmoqda, ularning vakillari samaradorlikni oshirish uchun yangi vositalarga muhtoj.

Katta ma'lumotlarga misollar

Big Data texnologiyalarini bozorga va zamonaviy hayotga faol joriy etish, dunyoning taniqli kompaniyalari dunyoning deyarli har bir burchagida o'z mijozlariga ega bo'lganlaridan foydalanishni boshlaganidan keyin boshlandi.

Bu Facebook va Google, IBM. Kabi ijtimoiy gigantlar, shuningdek Master Card, VISA va Bank of America singari moliyaviy tuzilmalardir.

Masalan, IBM davom etayotgan kassa operatsiyalariga katta ma'lumotlar usullarini qo'llaydi.

Ularning yordami bilan 15 foizga ko'proq soxta bitimlar aniqlandi, bu esa himoyalangan mablag'lar miqdorini 60 foizga oshirish imkonini berdi. Shuningdek, tizimning noto'g'ri signallari bilan bog'liq muammolar hal qilindi - ularning soni yarmidan ko'piga kamaytirildi. VISA shu kabi yoki boshqa operatsiyani amalga oshirish uchun qilingan firibgarliklarni kuzatib, Big Data-dan foydalangan. Buning yordamida ular har yili 2 milliard dollardan ko'proq mablag'ni tejashga imkon beradi.

Germaniya Mehnat vazirligi ishsizlik bo'yicha nafaqa berish bo'yicha katta ma'lumot tizimini joriy etish orqali xarajatlarni 10 milliard evroga kamaytirishga muvaffaq bo'ldi. Shu bilan birga, fuqarolarning beshdan biri bu imtiyozlarni hech qanday sababsiz olayotganligi aniqlandi.

Big Data o'yin sanoatini qo'ldan boy bermadi. Shunday qilib, World of Tanks-ning ishlab chiquvchilari barcha o'yinchilar haqida ma'lumotni o'rganishdi va ularning faoliyatining mavjud ko'rsatkichlarini solishtirishdi. Bu kelajakda o'yinchilarning ketishini taxmin qilishda yordam berdi - qilingan taxminlarga asoslanib, tashkilot vakillari foydalanuvchilar bilan yanada samarali aloqada bo'lishdi.

Taniqli yirik ma'lumotlar tashkilotlariga HSBC, Nasdaq, Coca-Cola, Starbucks va AT&T kiradi. Katta ma'lumotlarning eng katta muammosi - uni qayta ishlashning narxi. Bunga qimmatbaho uskunalar va katta miqdordagi ma'lumotlarga xizmat ko'rsatadigan malakali mutaxassislarining ish haqi xarajatlari ham kirishi mumkin. Shubhasiz, uskunalar doimiy ravishda yangilanib turilishi kerak, shunda u ma'lumotlarning ko'payishi bilan minimal ish qobiliyatini yo'qotmaydi.

Ikkinchi muammo yana qayta ishlanishi kerak bo'lgan juda ko'p ma'lumotlar bilan bog'liq. Agar, masalan, tadqiqot 2-3 emas, balki ko'p sonli natijalarni beradigan bo'lsa, ob'ektiv bo'lib qolish va umumiy ma'lumot oqimidan faqat hodisaning holatiga real ta'sir ko'rsatadigan narsalarni tanlash juda qiyin.

xizmatlari onlayn ma'lumotlardan foydalanishga o'tishlari sababli, kiber jinoyatchilar uchun boshqa maqsadga aylanish juda oson. Hatto biron bir onlayn tranzaksiya qilmasdan shaxsiy ma'lumotlarni oddiy saqlash ham bulutli saqlash mijozlari uchun nomaqbul oqibatlariga olib kelishi mumkin.

Axborotni yo'qotish muammosi. Ehtiyot choralar oddiy bitta ma'lumotni zaxiralash bilan cheklanmaydi, lekin kamida 2-3 ta zaxira nusxasini talab qiladi. Biroq, hajmning o'sishi bilan, zaxira bilan bog'liq qiyinchiliklar kuchaymoqda - va IT-mutaxassislari ushbu muammoning maqbul echimini topishga harakat qilmoqdalar.

BigData bilan ishlashning asosiy tamoyillari

1. Gorizontallik miqyoslilik: ma'lumotlar juda ko'p bo'lishi mumkinligi sababli, ular saqlanadigan tizim kengaytirilishi kerak. Agar ma'lumotlar miqdori ikki baravar ko'paygan bo'lsa, u holda klasterlar soni ikki baravar ko'payadi.

2. Xatolarga chidamlilik: gorizontallik miqyoslilik klasterda ko'p sonli mashinalar mavjudligini anglatadi. Va, albatta, bu mashinalar u yoki bu sabablarga ko'ra muvaffaqiyatsiz bo'ladi. Masalan, Yahoo ning Hadoop klasterida 42 000 dan ortiq mashinalar mavjud. BigData bilan ishlash usullari ushbu omilni hisobga olishi va ko'rinadigan yo'qotishlarsiz ishlashni davom ettirishi kerak.

3. Ma'lumotlarning joylashishi: katta tizimlarda ma'lumotlar ko'p sonli mashinalar bo'ylab taqsimlanadi. Agar ma'lumotlar bir mashinada joylashgan bo'lsa va boshqa mashinada qayta

ishlansa, bu ma'lumotlarni uzatish narxi hatto qayta ishlash narxidan oshib ketishi mumkin. Shuning uchun BigData dizaynidagi muhim masala - ma'lumotlarning joylashuvi, ma'lumotlarni u saqlanadigan joyda qayta ishlash printsipli. Big Data axborot texnologiyalarining eng tez rivojlanayotgan yo'nalishlaridan biri bo'lib, statistik ma'lumotlarga ko'ra, qabul qilingan va saqlanadigan ma'lumotlarning umumiy hajmi har 1-2 yilda ikki barobar ortadi. 2012-2014-yillarda mobil tarmoqlar orqali har oyda uzatiladigan ma'lumotlar

Data lake(ma'lumotlar ko'li) - qayta ishlanmagan katta ma'lumotlar ombori.

"Ko'l" har xil manbalardan kelgan, har xil formatda bo'lgan ma'lumotlarni saqlaydi. Bu esa odatiy relatsion ma'lumotlar omborida ma'lumotlarni aniq struktura asosida saqlashdan ko'ra arzonroqqa tushadi. Ma'lumotlar ko'li, ma'lumotlarni boshlang'ich holatida analiz qilish imkonini beradi. Bundan tashqari, "ko'l"lardan bir vaqtni o'zida bir nechta ishchilar foydalanishlari mumkin.

Data science(ma'lumotlar haqidagi fan) - analiz muommolarini, ma'lumotlarni qayta ishlash va ularni raqamli ko'rinishda taqdim etishni o'rganadigan fan.

Bu atama dunyoga kelgan vaqt 1974-yil hisoblanadi. O'sha yili Daniyalik informatik, Peter Naur "A Basic Principle of Data Science" nomli kitobini chop ettirgan.

2010-yillar boshida katta ma'lumotlarni tarqalishi natijasida bu yo'nalish juda foydali va kelajagi bor biznesga aylandi. Va o'shandi katta ma'lumotlar bilan ishlaydigan mutaxassislariga talab juda oshib ketdi.

Data science tushunchasiga ma'lumotlar omborini loyihalash va raqamlangan ma'lumotlarni qayta ishlashning barcha metodlari kiradi. Ko'plab mutaxassislar fikricha, aynan data science big dataning biznes nuqtai nazaridan hozirgi zamonoviy o'rindoshi hisoblanadi.

Data mining(ma'lumotlarni topish) - biron qonuniyatni topish maqsadida ma'lumotlarni intellektual analiz qilishga aytiladi. Isroillik matematik Grigoriy Pyatetskiy-Shapiro 1989-yilda bu atamani fanga kiritgan.

Texnologiyalar, avvalari noma'lum va foydali bo'lgan qayta ishlanmagan(hom) ma'lumotlarni topish jarayoniga data mining(ma'lumotlarni topish) deyiladi. Data mining metodlari ma'lumotlar ombori, statistika va sun'iy intellekt tutashgan nuqtada joylashadi.

Machine learning(mashinali o'qitish) - o'zi o'rganadigan dasturlar yaratish amaliyoti va nazariyasi, sun'iy intellektning katta qismi.

Dasturchilar o'z algoritmlariga xususiy hollarda umumiy qonuniyatlarni aniqlashni o'rgatishadi. Natijada, kompyuter, inson avvaldan ko'rsatib o'tgan buyruqlaridan emas, balki, o'z shaxsiy malakasidan kelib chiqib qaror qabul qiladi. Bunday o'qitishning juda ko'p metodlari data mining'ga oid bo'lishi mumkin.

Mashinali o'qitishga birinchi tarifni 1959-yilda amerikalik informatik Artur Samuel bergan. U sun'iy intellekt elementlariga ega bo'lgan shashka o'yini, dunyoda birinchi o'zi o'rganadigan dasturni yaratgan.

Deep learning(chuqur o'qitish) - yanada murakkab va yanada mustaqil bo'lgan o'zi o'qidigan dasturlar yaratadigan mashinali o'qitish turi. Oddiy mashinali o'qitish hollarida boshqariladigan malaka yordamida kompyuter bilimlarni aniqlab oladi: dasturchi algoritmgaga ma'lum misollarni ko'rsatadi, xatolarni qo'lda to'g'rilaydi. Deep learningda esa, tizim o'zi o'z funksiyalarini loyihalaydi, ko'p darajali hisob-kitoblar amalga oshiradi va atrof-muhit haqida xulosalar qiladi.

Odatda chuqur o'qitishni neyron tarmoqlarga tadbiiq qilishadi. Bu texnologiya asosan rasmlarni qayta ishlashda, nutqni tanishda, neyromashina tarjima, farmatsevtikadagi hisoblashlarda va boshqa zamonaviy texnologiyalarda qo'llaniladi. Asosan Google, facebook va Baidu tomonidan loyihalarga tadbiiq qilinadi.

Sun'iy neyron tarmog'i - oddiy protsessorlar(sun'iy neyronlar) birlashtirilgan tizimi bo'lib, insonning nerv tizimini imitatsiya qiladi. Bunday struktura evaziga, neyron tarmoqlari dasturlanmaydi, ular o'qitishadi. Huddi haqiqiy neyronlar kabi, protsessorlar signallarni oddiygina qabul qilishadi va boshqa protsessorlarga o'zlashadi. Shu bilan birga, boshqa butun tizim algoritmlar bajara olmaydigan murakkab topshiriqlarni bajaradi.

1943-yilda amerikalik olimlar Uorren Makkalok va Uolter Pittslar sun'iy neyron tarmog'i tushunchasini fanga kiritishgan.

Business intelligence(biznes-analitika) - aniq strukturaga ega bo'lmagan juda katta hajmdagi ma'lumotlarni qayta ishlash natijasida muqobil biznes yechimlar izlashga aytiladi.

Effektiv biznes-analitika ichki va tashqi ma'lumotlarni analiz qiladi - ham bozor axborotlarini, ham mijoz-kompaniyaning hisobotlarini hisobga oladi. Bu biznesni butunlay tushunishga yordam beradi, shu bilan birga, strategik va operatsion qarorlar qabul qilishga zamin yaratadi(mahsulot narxini aniqlashda, kompaniya rivojlanishining asosiy yo'nalishlarini belgilab olishda).

Bu atama 1958-yilda IBM taqdiqotchisi Xans Piter Lun maqolasida birinchi bo'lib ko'rsatilgan. 1996-yilda axborot texnologiyalari bozorini o'rganishga asoslangan Gartner analitik agentligi business intelligence tarkibiga data mining metodikasini ham qo'shgan.

Katta ma'lumotlar bilan ishlash

Ma'lumotlar- kompyuter ishlaydigan va magnit, optik yoki mexanik tashuvchilarda qayd etilgan elektr signallari shaklida saqlanishi va uzatilishi mumkin bo'lgan miqdorlar, belgilar yoki belgilar.

Muddati Katta ma'lumotlar vaqt o'tishi bilan eksponent ravishda o'sib boruvchi katta ma'lumotlar to'plamini tasvirlash uchun ishlatiladi. Ushbu hajmdagi ma'lumotlarni qayta ishlash uchun ajralmas hisoblanadi.

Big Data taqdim etadigan imtiyozlar:

- ✓ Turli manbalardan ma'lumotlarni yig'ish.
- ✓ Real vaqt tahlili orqali biznes jarayonlarini takomillashtirish.
- ✓ Katta hajmdagi ma'lumotlarni saqlash.

Insights. Katta ma'lumotlar tuzilgan va yarim tizimli ma'lumotlardan foydalangan holda yashirin ma'lumotlarga nisbatan aniqroqdir.

Katta ma'lumotlar xavflarni kamaytirishga yordam beradi va tegishli xavf tahlillari yordamida aqlli qarorlar qabul qiladi

Katta ma'lumotlarga misollar

Nyu-York fond birjasi har kuni hosil qiladi 1 terabayt oxirgi sessiyadagi savdolar to'g'risidagi ma'lumotlar.

Ijtimoiy tarmoqlar: statistika shuni ko'rsatadiki, Facebook ma'lumotlar bazalari har kuni yuklanadi 500 terabayt yangi ma'lumotlar asosan ijtimoiy tarmoq serverlariga fotosuratlar va videolarni yuklash, xabar almashish, postlar ostidagi sharhlar va boshqalar tufayli hosil bo'ladi.

Reaktiv dvigatel hosil qiladi 10 terabayt parvoz paytida har 30 daqiqada ma'lumotlar. Har

kuni minglab parvozlar bo'lganligi sababli, ma'lumotlar miqdori petabaytlarga etadi.

Katta ma'lumotlarning tasnifi

Katta ma'lumotlar shakllari:

- Strukturaviy
- Tarkibi bo'lmagan
- Yarim tuzilgan
- Strukturaviy shakl

Ruxsat etilgan formatda saqlanishi, foydalanilishi va qayta ishlanishi mumkin bo'lgan ma'lumotlar tuzilgan deb ataladi. Yillar davomida kompyuter fanlari ushbu turdagi ma'lumotlar bilan ishlash texnikasini takomillashtirishda katta muvaffaqiyatlarga erishdi (bu erda format oldindan ma'lum) va qanday foyda olishni o'rgandi. Biroq, bugungi kunda hajmlarning bir necha zettabayt oralig'ida o'lchanadigan o'lchamlarga o'sishi bilan bog'liq muammolar mavjud.

1 zettabayt milliard terabaytga teng.

Ushbu raqamlarga qarab, Big Data atamasining to'g'riligiga va bunday ma'lumotlarni qayta ishlash va saqlash bilan bog'liq qiyinchiliklarga ishonch hosil qilish oson.

Relyatsion ma'lumotlar bazasida saqlanadigan ma'lumotlar tuzilgan bo'lib, masalan, kompaniya xodimlarining jadvalariga o'xshaydi

Strukturasisiz shakl

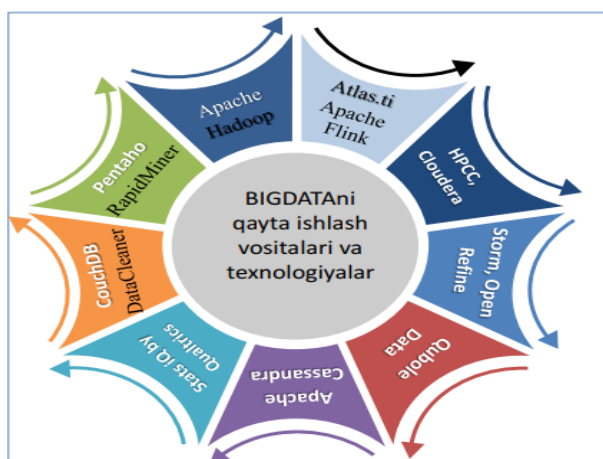
Noma'lum tuzilishdagi ma'lumotlar tuzilmagan deb tasniflanadi. Katta bo'lishidan tashqari, bu shakl foydali ma'lumotlarni qayta ishlash va olishda bir qator qiyinchiliklar bilan tavsiflanadi. Tarkibi bo'lmagan ma'lumotlarning odatiy misoli oddiy matnli fayllar, rasmlar va videolarning kombinatsiyasini o'z ichiga olgan heterojen manbadir. Tashkilotlar bugungi kunda katta hajmdagi xom yoki tuzilmagan ma'lumotlardan foydalanish imkoniyatiga ega, ammo undan qanday foyda olishni bilishmaydi.

Katta ma'lumotlar texnologiyalari asosan mijoz muhitini tahlil qilish uchun ishlatiladi. SAP CIS (SAP CIS) bosh direktorining o'rinbosari Dmitriy Shepelyavvi bir nechta misollar keltiradi: "Amerika PNC banki o'z mijozlarining saytlardagi xatti-harakatlari, xaridlar va turmush tarzi haqidagi ma'lumotlarni moslashuvchan foiz stavkalari siyosatiga aylantiradi, bu esa pirovardida amalga oshiriladi. o'sish ko'rsatkichlari kapitallashuvida ifodalangan. Avstraliya Hamdo'stlik Banki (CBA) o'z omonatchilarining barcha operatsiyalarini tahlil qiladi va bu tahlilni ular haqida ijtimoiy tarmoqlarda ma'lumotlarni to'plash orqali to'ldiradi. Ushbu ma'lumotlar oqimini bog'lash orqali bank kreditlar to'lanmaslik foizini sezilarli darajada kamaytirishga erishdi. Rossiyada esa Ural tiklanish va taraqqiyot bankining tajribasi qiziq - ular ma'lum bir mijoz uchun maksimal qiziqish uyg'otishi mumkin bo'lgan kredit takliflari, depozitlar va boshqa xizmatlarni yaratish uchun mijozlar bazasi to'g'risidagi ma'lumotlar bilan ishlay boshladilar. Taxminan bir yil davomida IT-yechimlarni qo'llash davomida UBRDning chakana kredit portfeli qariyb 55 foizga o'sdi".

BigDatani qayta ishlash vositalari va texnologiyalari

BigData - 2000 yil oxirida paydo bo'lgan va an'anaviy ma'lumotlar bazalarini boshqarish tizimlari va Business Intelligence sinfining yechimlariga muqobil bo'lgan, katta hajmdagi va sezilarli xilma-xillikdagi strukturlangan va strukturlanmagan ma'lumotlarni qayta ishlashga mo'ljallangan dasturlash texnologiyasi hisonlanadi. Hozirda Big Data texnologiyalari hayotning turli sohalariga tobora chuqurroq kirib bormoqda, savdo,

reklama, dam olish va hatto shaxsiy hayot tuzilishi haqidagi odatiy g'oyalarni o'zgartirmoqda. Mamlakarimizda elektron hukumat doirasida turli axbrot tizimlarining joriy etilishi natijasida ma'lumotlar bazasida katta hajmdagi ma'lumotlarning yuzaga kelishi kuzatilmoqda. Ushbu ma'lumotlarni qayta ishlash uchun BigData texnologiyalaridan foydalanishga to'g'ri keladi. Hozirgi kunda Bigdata bilan ishlash uchun ko'plab vositalar va texnologiyalar mavjud. Ushbu vositalar Bigdata ma'lumotlarini tahlil qilinish samaradorligini oshirishga va vaqtni tejashga imkon beradi. 2022 yil holatiga ko'ra Bigdata ma'lumotlarini qayta ishlash uchun qulay bo'lgan vositalar va texnologiyalar rasmda keltirilgan keltirilgan



BIGDATA ni qayta ishlash vositalari va texnologiyalari.

Apache Hadoop [<https://hadoop.apache.org/releases.html>] – Bigdata platformasi bo'lib, katta hajmdagi ma'lumotlarni kompyuter klasterlari yordamida taqsimlangan tarzda qayta ishlash imkonini beradi. Ushbu amallar serverlar yoki minglab mashinalardan foydalanilgan holda bajariladi.

Atlas.ti [<https://atlasti.com>] – keng qamrovli tadqiqot dasturi hisoblanib, katta hajmdagi ma'lumotlarni tahlil qilishda turli platformalardan universal tarzda foydalanish imkoni tadqim etadi. Ushbu vositadan ilmiy tadqiqotlarni amalga oshirishda foydalanishingiz mumkin.

HPCC [<https://hpccsystems.com/try-now>] – LexisNexis Risk Solution tomonidan ishlab chiqilgan bo'lib, katta hajmdagi ma'lumotlarni tahlil qilishda yagona platforma, arxitektura va dasturlash tilidan iborat yechimni taklif qiladi.

Storm [<http://storm.apache.org/downloads.html>] – katta hajmdagi ma'lumotlarni tahlil qilish uchun ochiq kodli hisoblash tizimi. Katta hajmdagi ma'lumotlarni ayni vaqtda tizimda qayta ishlanganda, taqsimlanganda yuzaga keladigan nozozliklarga chidamli bo'lgan yagona tizimdir.

Qubole Data [<https://www.qubole.com/>] – katta hajmdagi ma'lumotlarni tahlil qiladigan avtomon platforma. Unda o'z-o'zini boshqarish, optimalashtirish va qayta ishlangan ma'lumotlarni guruhlashga ruhsat beradigan, ochiq manbali katta hajmdagi ma'lumotlar vositasidir.

Apache Cassandra [<http://cassandra.apache.org>] – ma'lumotlar bazasi bugungi kunda katta hajmdagi ma'lumotlarni samarali boshqarishni ta'minlash uchun keng qo'llaniladi.

Stats iQ by Qualtrics [<https://www.qualtrics.com/au/iq/stats-iq/>] – Katta hajmdagi ma'lumotlarni tahlil qilish va foydalanish uchun qulay bo'lgan statistic vositadir. Uning zamonaviy interfeysi statistik testlarni avtomatik ravishda tanlaydi.

CouchDB [<http://couchdb.apache.org/>] – Internet orqali kirish yoki JavaScript ni ishlatib, so'rovni bajarish uchun JSON hujjatlarida ma'lumotlarni saqlaydi. Unda noto'g'ri saqlash bilan tarqalgan masshtabni taklif etadi. Bu replikasiya protokolini belgilash orqali ma'lumotlarga kirish imkonini beradi.

Pentaho [<https://www.hitachivantara.com/>] – Katta hajmdagi ma'lumotlarni olish, tayyorlash va birlashtirish vositasi. U ma'lumotlarni vizualizatsiya va tahlil qilishni taklif qiladi. Katta hajmdagi ma'lumotlar bilan ishlash uchun ushbu vosita katta ma'lumotlarni katta g'oyalarga aylantirish imkonini beradi.

Apache Flink [<https://flink.apache.org/>] – Katta hajmdagi ma'lumotlar oqimini tahlil qiladigan eng yaxshi ochiq kodli vositadir. U ma'lumotlarni taqsimlangan, yuqori tezlikda ishlash, har doim mavjud va ma'lumotlar oqimini yuboradigan dasturiy ilova.

Cloudera [<https://www.cloudera.com/>] – katta hajmdagi ma'lumotlar bilan ishlashda eng tez, oddiy va xavfsiz zamonaviy platforma. Bunda foydalanuvchiga bitta kengaytiriladigan platformada istalgan muhitda istalgan ma'lumotlarni olish imkonini beradi.

Open Refine [<https://openrefine.org/>] – katta hajmdagi ma'lumotlar bilan ishlash uchun qulay bo'lgan vosita. Unda katta hajmdagi ma'lumotlarni tozalash va bir formatdan boshqa formatga aylantirib, tartibsiz ma'lumotlarni tahlil qilib beradi. Undagi ma'lumotlar veb-xizmat va tashqi ma'lumotlar bilan kengayadi.

RapidMiner [<https://my.rapidminer.com/>] – katta hajmdagi ma'lumotlarni tahlil qilish uchun eng yaxshi ochiq kodli vosita. Unda ma'lumotlarni tayyorlash, mashinani o'rganish va modelni joylashtirish uchun ishlatiladi. Katta hajmdagi ma'lumotlarni qidirish va bashoratli ma'lumotlar tahlil to'plamni taklif qiladi.

DataCleaner [<https://github.com/datacleaner>] – bu ilova ma'lumotlar sifatini tahlil qilish va yechimlar platformasi hisoblanadi. Unda ma'lumotlarni tezkor profillash mexanizmi mavjud bo'lib, ma'lumotlar ko'paygan sari ma'lumotlarni o'chirish, o'zgartirish va birlashtiradi.

Kaggle [<https://www.kaggle.com/>] – dunyodagi eng yirik katta hajmdagi ma'lumotlar bilan ishlaydigan tashkilotdir. U tadqiqodchilarga o'z ma'lumotlari va statistik ma'lumotlarini nashr etishga yordam beradi. Katta hajmdagi ma'lumotlarni tahlil qilish uchun eng yaxshi joy hisoblanadi.

Hive [<https://hive.apache.org/downloads.html>] – katta hajmdagi ma'lumotlar bilan ishlaydigan ochiq kodli dasturiy vosita hisoblanadi. Unda dasturchilarga Hadoop da katta hajmdagi ma'lumotlar to'plamini tahlil qilish, katta hajmdagi ma'lumotlar to'plamini boshqarish va so'rovlarni bajarish imkonini beradi.

Foydalanilgan adabiyotlar:

1. Силен Д. Основы Data Science, Big Data. Python и наука о данных / Д. Силен. М.: Питер, 2017.
2. Михнев И.П. Информационная безопасность на просторах мобильного интернета // Образовательные ресурсы и технологии. – 2015.
3. Bernard Marr. "Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance". John Wiley& Sons Ltd, 2015.