

BIG DATANI TAHLIL QILISH MUAMMOLARI

Kenjaev Xamdam Bazarbaevich

*(Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti
Nukus filiali "Kompyuter tizimlari" kafedrasida assistenti)*

Xojanova Alfiya Mayrambay Qizi

*(Muhammad al-Xorazmiy nomidagi TATU Nukus filiali Kompyuter injiniringi
fakul'teti 2-kurs talabasi)*

Abilxanova Balauza Marat Qizi

*(Muhammad al-Xorazmiy nomidagi TATU Nukus filiali Kompyuter injiniringi
fakul'teti 2-kurs talabasi)*

Bugungi kunda Big Datani tahlil qilish uchun axborot saqlash texnologiyalarini yaratishda ishlab chiqilgan yondashuv va usullardan foydalanilmoqda va ularga miqdoriy ko'rsatkichlar asosida tuzatishlar kiritishda ma'lumotlar yig'ish, qayta ishlash va tahlil qilish tamoyillarini hisobga olinmoqda. Shu bilan birga, an'anaviy operatsiyalarning ayrim xususiyatlari Big Datani qayta ishlash xususiyatlariga zid bo'lishi mumkin.

Operatsion va analitik ma'lumotlarni qayta ishlash vazifalaridagi sezilarli farq ma'lumotlar bazasi texnologiyasi rivojlanishining boshida o'zini namoyon qila boshladi. Ma'lumotlar bazasi atamasi Bill Inmon tomonidan 1970 yillardayoq taklif qilingan edi, lekin bu texnologiyalarga qiziqish ortishi faqatgina 20 yil o'tib sodir bo'ldi, birinchi navbatda, bunday tizimlar uchun real ehtiyoj bor edi, ikkinchidan, ular mavjud bo'lgan hisoblash ishlarini oshirish uchun zarur edi.

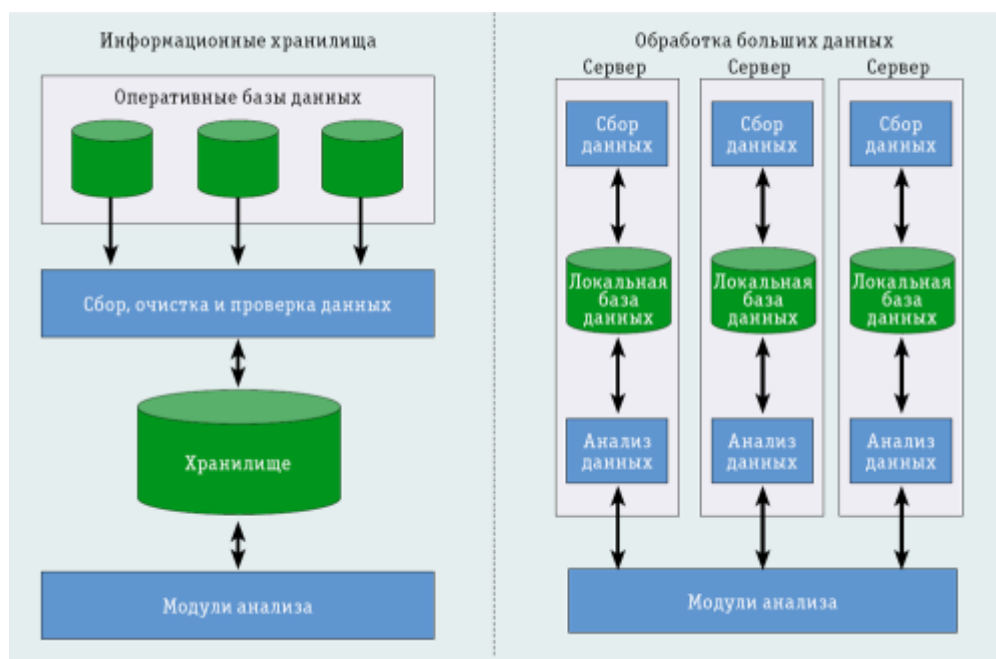
Ma'lumotlar bazasidagi ma'lumotlarni qayta ishlash tsikli o'z ichiga ma'lumotlarni yig'ish, tozalash, yuklash, tahlil qilish va nihoyat tahlil natijalarini taqdim etishdan iborat. Ushbu bosqichlar haqida batafsil ma'lumot berish mantiqsiz bo'ladi, lekin asosiy tezisni aniq belgilash kerak - agar siz Big Datani tahlil qilish uchun axborot saqlash texnologiyalarini qo'llashga harakat qilsak, unda siz nafaqat tahlil algoritmlariga, balki ma'lumotlar bilan ishlashning barcha bosqichlariga ham e'tibor qaratishimiz kerak.

Axborotni yig'ish.

Axborot tizimiga mos keladigan ma'lumot, operatsion ma'lumotlar bazasidan olinadi, kerakli shaklga o'tkaziladi, tekshiriladi va keyin tizimga yuklanadi. Ro'yxatga olingan operatsiyalar muayyan davriylik bilan amalga oshiriladi va bu yerda savol tug'iladi: "Big Data" bilan ishlashda doimo bunday "davriylik" har doim mumkin va tahlil qilish uchun imkon qadar tezroq bo'lishi kerakmi? Axborotning ko'rinishi bilan tahlil qilish uchun mavjudligi o'rtasidagi vaqt oralig'i axborot saqlash amaliyotlarini bajarish uchun zarur bo'lgan vaqtdan kamroq bo'lishi mumkin. Bunday vazifaning namunasi ijtimoiy tarmoqlarni salbiy bayonotlarni aniqlash yoki maxfiy axborotning yo'qolishi faktlarini aniqlash uchun kuzatib borishdir. Bu voqealar imkon qadar qisqa

vaqt ichida aniqlash va zararsizlantirish kerak. Ammo, biz bu yerda ma'lumotni rasmiy bo'lmagan taqdimot bilan shug'ullanmoqdamiz, dastlab birinchi bosqichda yuqori tezlik bilan ta'riflanmagan matnni ishlab chiqarish algoritmlarini qayta ishlash zarur.

Ma'lumotlar bazasining dastlabki ishlashida (misol uchun, ma'lumotni izlashda), ilgari to'plangan tarkibni ishlatish mumkin, bu Big Data bilan ishlashda qiyin bo'lgan degan ma'noni anglatadi. Muammo birinchi navbatda ularning har doim taqsimlanishi bilan bog'liq bo'lib, tahlil qilish uchun qulay bo'lishi bilan emas, balki uni yig'ib olish uchun qulay bo'lgan - masalan, agar biz telekommunikatsiya tizimlari haqida gapiradigan bo'lsak, ma'lumotlar mintaqaviy serverlarda "qo'shilgan" (1-Rasm). Tahlil nuqtai nazaridan ma'lumotni hududiy emas, balki vaqti bilan tarqatish yanada qulayroq (har bir server ma'lum bir vaqt uchun javobgar bo'ladi) va hokazo. Ammo yana ma'lumot tahlil uchungina kerakli serverlarga o'tkazilgandan so'nggina mavjud bo'ladi.



Rasm. 1. Ma'lumot yig'ish

Shunday qilib, an'anaviy ma'lumot saqlash tizimlarida barcha ma'lumotlar har doim ham konvertatsiya qilish, nazorat qilish, tozalash, yuklash va ushbu operatsiyalarni amalga oshirish uchun javobgar bo'lgan yagona mantiqiy blok orqali o'tadi. Shu bilan birga, katta hajmdagi ma'lumotlarni qayta ishlash bunday birlikni tashkil qilolmaydi. Shuni ta'kidlash kerakki, hozirgi kunga qadar jadal kirish ma'lumotlar oqimiga ega bo'lgan juda ko'p ishlar yo'q, ammo mantiqiy birlashgan tizim sifatida yig'ish, tozalash, konvertatsiya qilish va yuklash uchun blokni amalga oshirish mumkin.

Tahlil

An'anaviy ravishda ma'lumotlar bazalari taxminan bir xil ma'lumotlar tahlil vositalari to'plamini beradi: ko'p o'lchovli tahlil (OLAP), regressiya, tasniflash,

klasterlashtirish va qidirish. Bugungi kunda bu yo'nalishda dasturiy mahsulotlar ham paydo bo'lgan - masalan, SAP Xon, Greenplum Chorus, Aster Data nCluster, bu usullarni Big Data-da ishlashga imkon beradi.

Bunday yechimlarning potentsialini tushunish uchun, asosiy algoritmlarni ko'rib chiqish, shuningdek, ularning katta parallelizatsiyalash yo'llarini tahlil qilish katta hajmdagi ma'lumotlarni qayta ishlash kalitidir. Shu bilan birga, taqsimlangan ma'lumotlarni qayta ishlashning muayyan texnologiyalariga (masalan, MapReduce) nisbatan emas, balki faqat Big Data (tarmoq ta'sirining intensivligi va hajmlari) xarakteristikasining asosiy parametrlarini hisobga olish muhim. Big Data bilan ishlash uchun barcha mavjud usullar klassik B-daraxtlaridan murakkab tuzilmalarga ko'p o'lchovli ma'lumotlarni boshqarish uchun qo'llaniladi.

FOYDALANGAN ADABIYOTLAR:

1. Tomas Konnolli Karolin Begg, *Вазы данных, проектирование, реализация и сопровождение, теория и практика*, Университет Пеysли, Шотландия, изд. М.- СПб.- Киев, 2003.
2. Deyt K.Dj., *Vvedenie v sistemi baz dannix*.
3. Ulman Djeffri D., Djennifer Uidom, *Vvedenie sistemi baz dannix*. Per.s angl. M.: «Lori», 2000.
4. Meyer D., *Teoriya relyatsionnix baz dannix*. Per.s angl. M.: Mir, 1987.