

O‘ZBEKISTON RESPUBLIKASI
OLIY TA’LIM, FAN VA INNOVATSIYALAR VAZIRLIGI

BUXORO DAVLAT UNIVERSITETI HUZURIDAGI PEDAGOG
KADRLARNI QAYTA TAYYORLASH VA ULARNING MALAKASINI
OSHIRISH MINTAQAVIY MARKAZI

“BIG DATA VA MA’LUMOTLAR TAHLILI”
MODULI BO‘YICHA

O‘QUV – USLUBIY MAJMUA

Buxoro 2025

Modulning ishchi dasturi Oliy ta'lim, fan va innovatsiyalar vazirligining 2024-yil 27-dekabrda 485-sonli buyrug'i bilan tasdiqlangan oliy ta'lim muassasalari pedagog kadrlarini qayta tayyorlash va malaka oshirish yo'nalishlari o'quv reja va dasturlariga muvofiq ishlab chiqilgan.

Tuzuvchi: N.S. Sayidova - BuxDU Axborot tizimlari va raqamli texnologiyalar kafedrasida dotsenti, f.-m.f.n.

Taqrizchi: J. Jumayev - BuxDU Amaliy matematika va dasturlash texnologiyalari kafedrasida professori, f.-m.f.n.

**O'quv-uslubiy majmua Buxoro davlat universiteti Ilmiy
Kengashining qarori bilan nashrga tavsiya qilingan
(2024 yil "27" dekabrda 5-sonli bayonnoma)**

MUNDARIJA

I.	ISHCHI DASTUR	4
II.	MODULNI O‘QITISHDA FOYDALANILADIGAN INTERFAOL TA’LIM METODLARI	9
III.	NAZARIY MASHG‘ULOT MATERIALLARI	15
IV.	AMALIY MASHG‘ULOT MATERIALLARI	41
V.	KEYSLAR	64
VI.	GLOSSARIY	73
VII.	FOYDALANILGAN ADABIYOTLAR RO‘YXATI	79

I. ISHCHI DASTURI

Kirish

Ushbu ishchi dastur O'zbekiston Respublikasining 2020 yil 23 sentabrda tasdiqlangan "Ta'lim to'g'risida"gi Qonuni, O'zbekiston Respublikasi Prezidentining 2020 yil 3 dekabrda "Iqtidorli yoshlarni saralab olish tizimi va akademik litseylar faoliyatini takomillashtirish chora-tadbirlari to'g'risida"gi PQ-4910-son hamda Vazirlar Mahkamasining 2022 yil 1 iyundagi "Akademik litseylar rahbar va pedagog xodimlarining uzduksiz malakasini oshirish tizimini joriy etish to'g'risida"gi 296-son Qarorlarida belgilangan ustuvor yo'nalishlar hamda namunaviy dastur mazmunidan kelib chiqqan holda tuzilgan bo'lib, u zamonaviy talablar asosida malaka oshirish jarayonlarining mazmunini takomillashtirish hamda akademik litseylar pedagog xodimlarining kasbiy kompetentligini muntazam oshirib borishni maqsad qiladi.

Ishchi dastur doirasida berilayotgan mavzular orqali akademik litseylarda faoliyat olib borayotgan pedagog xodimlarning big data va ma'lumotlar tahliliga oid bilimlar darajasini kengaytirish va chuqurlashtirish hisobiga ularning pedagogik mahorat va kasbiy kompetentligini muntazam takomillashtirish bilan bog'liq bilim, ko'nikma va malakalarga ega bo'lishlari ta'minlanadi.

Modulning maqsad va vazifalari

Modulning maqsadi tinglovchilarni Data Mining bilan OLAP-tizimlari o'rtasidagi farqni, Data Mining usullari hisoblangan assotsiatsiya, klassifikatsiya, ketma-ketlik, klasterizatsiya, prognozlash asoslarini o'rgatish va ulardan foydalanish ko'nikmalarini bo'yicha bilim, ko'nikma va malakalarini oshirishdan iborat.

Modulning vazifalariga quyidagilar kiradi:

- Data Mining metodlari: neyron to'rlari;
- qarorlar (yechimlar) daraxti;
- pereborni cheklash usullari;
- genetik algoritmlar;
- evolyutsion dasturlash; klasterli modellar va aralash metodlar orqali ma'lumotlarni taxlil qilish;
- OLTP, OLAP, ROLAP, MOLAP kabi metodlarni o'rganish.

Modul yakunida tinglovchilarning bilim, ko'nikma va malakalari hamda kompetentligiga qo'yiladigan talablar:

"Big data va ma'lumotlar tahlili" modulini o'zlashtirish jarayonida amalga oshiriladigan masalalar doirasida:

Tinglovchi:

- katta hajmdagi ma'lumotlarni aniqlash va ularni loyihalash haqida tasavvurga ega bo'lishni;
- katta hajmdagi ma'lumotlar modelini tavsiflash usullarini;
- katta hajmdagi ma'lumotlar saqlashda ma'lumotlar bazasini boshqarish tizimlarini tanlash va ma'lumotlar bazalarini shaxsiy hisoblash mashinalarida yaratishni bilishi va ulardan foydalana olishni;
- ma'lumot saqlagichlar va ulardagi Data Mining texnologiyalari va usullari haqida tasavvurga ega bo'lishni;
- OLTP, OLAP, ROLAP, MOLAP texnologiyalarini bilishi va ulardan foydalana olishni ***bilishi*** kerak.

Tinglovchi:

- ma'lumot saqlagichlar va ulardagi Data Mining texnologiyalari va usullari haqida tasavvurga ega bo'lish;
- Katta ma'lumotlarni va Data scienceni qo'llash sohalari hamda afzalliklari bilish;
- Data science jarayonlar kirish, topshiriq loyihalarini yaratish va tadqiq qilish;
- Data Mining texnologiyalari orqali ma'lumotlarni taxlil qilish ko'nikmalariga ega bo'lish;
- katta hajmdagi ma'lumotlarni aniqlash va ularni loyihalash haqida tasavvurga ega bo'lish;
- masalani yechish uchun ma'lumotlar bazasini boshqarish tizimlarini tanlash;
- ishlab chiqilgan model asosida katta hajmdagi ma'lumotlar bazasini loyihalash ***ko'nikma va malakalariga*** ega bo'lishi lozim.

Tinglovchi:

- taqsimlangan tizimlar ma'lumot bazalarini loyihalash malakalariga ega bo'lish;
- ma'lumot saqlagichlar va ulardagi Data Mining texnologiyalari va usullari haqida tasavvurga ega bo'lish;
- OLTP, OLAP, ROLAP, MOLAP texnologiyalarini bilishi va ulardan foydalana olish;
- katta ma'lumotlarni va Data scienceni qo'llash sohalari hamda afzalliklari bilish;
- Data science jarayonlar kirish, topshiriq loyihalarini yaratish va tadqiq qilish;

- Data Mining texnologiyalari orqali ma'lumotlarni taxlil qilish ko'nikmalariga ega bo'lishi;
- Data Mining texnologiyalari orqali jahonning yetuk kompaniyalarida ishlab chiqilgan dasturiy ta'minotlardan foydalanish *kompetensiyalariga* ega bo'lishi lozim.

MODUL BO'YICHA SOATLAR TAQSIMOTI

№	Modul mavzulari	Auditoriya o'quv yuklamasi		
		Jami	Nazariy	Amaliy mashg'ulot
1.	Katta hajmli ma'lumotlarni tashkillash vazifalari va metodlari.	2	2	
2.	Qarorni qo'llab quvvatlash tizimlari. OLAP- tizimlari.	4	2	2
3.	Katta hajimli ma'lumotlarni qayta ishlashda qo'llanlayotgan texnologiyalar.	2		2
	Jami:	8	4	4

NAZARIY MASHG'ULOTLAR MAVZULARI

1-mavzu: Katta hajmli ma'lumotlarni tashkillash vazifalari va metodlari. (2 soat ma'ruza)

Katta hajmli ma'lumotlarni tashkillash vazifalari va metodlari. Data science jarayonlar. Data Mining texnologiyalari, metodlari va bosqichlari. Katta hajmli ma'lumotlar tushunchasi va ahamiyati. Katta hajmli ma'lumotlarning o'sishi va unga bo'lgan talab. Katta hajmli ma'lumotlarni tashkillashtirishning zaruriyati. Strukturaviy ma'lumotlar: ma'lumotlar bazalari, jadvallar, yozuvlar. Semistrukturaviy ma'lumotlar: XML, JSON formatlaridagi ma'lumotlar. Katta hajmli ma'lumotlarni saqlash metodlari. Katta hajmli ma'lumotlarni qayta ishlash metodlari. Ma'lumotlarni tahlil qilish va vizualizatsiya. Ma'lumotlar xavfsizligi va maxfiyligi.

2-mavzu: Qarorni qo'llab quvvatlash tizimlari. OLAP- tizimlari.

OLAP (On-Line Analytical Processing) tizimlari, ma'lumotlar bazalarida analiz qilish va qarorlar qabul qilish jarayonlarini qo'llab-quvvatlashga mo'ljallangan texnologiyalardir. Qarorlarni qo'llab-quvvatlash tizimlari (DSS) va OLAP

tizimlarining ahamiyati. DSS va OLAP o'rtasidagi farqlar. Qarorlarni qo'llab-quvvatlash tizimlarining umumiy tushunchasi. OLAP tizimlarining asosiy tushunchasi. OLAP tizimlarining turlari. OLAP tizimlarining ishlash prinsiplari. OLAP tizimlarida ma'lumotlarni tahlil qilish. OLAP tizimlarining afzalliklari va cheklavlari. OLAP tizimlarining qo'llanilish sohalari.

AMALIY MASHG'ULOTLARI MAVZULARI

1-mavzu: Qarorni qo'llab quvvatlash tizimlari. OLAP- tizimlari. (2 soat amaliy)

Qarorni qo'llab quvvatlash tizimlari. OLAP- tizimlari. Qaror daraxтини qurish usullari. Ma'lumotlar bazasining ko'p o'lchovli, kontseptual, mantiqiy va fizik modeli. OLTP, OLAP, ROLAP, MOLAP texnologiyalari.

2-mavzu: Katta hajimli ma'lumotlarni qayta ishlashda qo'llanlayotgan texnologiyalar. (2 soat amaliy)

Katta hajimli ma'lumotlarni qayta ishlashda qo'llanlayotgan texnologiyalar. Katta hajimli ma'lumotlarni qayta ishlashda qo'llanlayotgan texnologiyalar: NoSQL, MapReduce, Hadoop, R, ERP, SAP. Grafik ma'lumotlar bazasi. Matnlarni chuqur va ma'lumotlarni vizuval tahlil qilish.

Modulni tashkil etish va o'tkazish bo'yicha tavsiyalar

“Big data va ma'lumotlar tahlili” moduli ma'ruza va amaliy mashg'ulotlar shaklida olib boriladi.

Kursni o'qitish jarayonida ta'limning zamonaviy metodlari, pedagogik texnologiyalar va masofaviy ta'limga asoslangan raqamli texnologiyalari qo'llanilishi nazarda tutilgan:

- video ma'ruzada zamonaviy interaktiv texnologiyalar yordamida prezentatsiya va elektron-didaktik texnologiyalar;

- o'tkaziladigan amaliy mashg'ulotlarda bulutli texnologiyaga asoslangan dasturiy vositalar, ekspress-so'rovlar, test so'rovlari va boshqa interaktiv ta'lim usullarini qo'llash nazarda tutiladi.

II. MODULNI O‘QITISHDA FOYDALANILADIGAN INTERFAOL TA’LIM METODLARI

«BLUM KUBIGI» metodi

Metodning maqsadi: Mazkur metod tinglovchilarda yangi axborotlar tizimini qabul qilish va bilimlarni o‘zlashtirilishini engillashtirish maqsadida qo‘llaniladi, shuningdek, bu metod tinglovchilar uchun “Ochiq” savollar tuzish va ularga javob topish mashqi vazifasini belgilaydi.

Metodni amalga oshirish tartibi:

1. Ushbu metodni qo‘llash uchun, oddiy kub kerak bo‘ladi.
Kubning har bir tomonida quyidagi so‘zlar yoziladi:
 - Sanab bering, ta’rif bering(oddiy savol)
 - Nima uchun (sabab-oqibatni aniqlashtiruvchi savol)
 - Tushuntirib bering (muammoga har tomonlama qarash savoli)
 - Taklif bering (amaliyot bilan bog‘liq savol)
 - Misol keltiring (ijodkorlikni rivojlantiruvchi savol)
 - Fikr bering(tahlil qilish va baholash savoli)
2. O‘qituvchi mavzuni belgilab beradi.
3. O‘qituvchi kubikni stolga tashlaydi. Qaysi so‘z chiqsa, unga tegishli savolni beradi.

“KWHL” metodi

Metodning maqsadi: Mazkur metod tinglovchilarni yangi axborotlar tizimini qabul qilishi va bilimlarni tizimlashtirishi uchun qo‘llaniladi, shuningdek, bu metod tinglovchilar uchun mavzu bo‘yicha quyidagi jadvalda berilgan savollarga javob topish mashqi vazifasini belgilaydi.

Izoh. KWHL:

Know – nimalarni bilaman?

Want – nimani bilishni xohlayman?

How - qanday bilib olsam bo‘ladi?

Learn - nimani o‘rganib oldim?

“KWL” metodi	
1. <i>Nimalarni bilaman:</i> -	2. <i>Nimalarni bilishni xohlayman, nimalarni bilishim kerak:</i> -
3. <i>Qanday qilib bilib va topib olaman:</i> -	4. <i>Nimalarni bilib oldim:</i> -

“W1H”metodi

Metodning maqsadi: Mazkur metod tinglovchilarni yangi axborotlar tizimini qabul qilishi va bilimlarni tizimlashtirishi uchun qo‘llaniladi, shuningdek, bu metod tinglovchilar uchun mavzu bo‘yicha quyidagi jadvalda berilgan oltita savollarga javob topish mashqi vazifasini belgilaydi.

What?	Nima? (ta’rifi, mazmuni, nima uchun ishlatiladi)	
Where?	Qaerda? (joylashgan, qaerdan olish mumkin)	
What kind?	Qanday? (parametrlari, turlari mavjud)	
When?	Qachon? (ishlatiladi)	
Why?	Nima uchun? (ishlatiladi)	
How?	Qanday qilib? (yaratiladi, saqlanadi, to‘ldiriladi, tahrirlash mumkin)	

“SWOT-tahlil” metodi

Metodning maqsadi: mavjud nazariy bilimlar va amaliy tajribalarni tahlil qilish, taqqoslash orqali muammoni hal etish yo‘llarini topishga, bilimlarni mustahkamlash, takrorlash, baholashga, mustaqil, tanqidiy fikrlashni, nostandart tafakkurni shakllantirishga xizmat qiladi.

S – (strength)	• kuchli tomonlari
W – (weakness)	• zaif, kuchsiz tomonlari
O – (opportunity)	• imkoniyatlari
T – (threat)	• xavflar

2.1-rasm.

“VEER” metodi

Metodning maqsadi: Bu metod murakkab, ko‘p tarmoqli, mumkin qadar, muammoli xarakterdagi mavzularni o‘rganishga qaratilgan. Metodning mohiyati shundan iboratki, bunda mavzuning turli tarmoqlari bo‘yicha bir xil axborot beriladi va ayni paytda, ularning har biri alohida aspektlarda muhokama etiladi. Masalan, muammo ijobiy va salbiy tomonlari, afzallik, fazilat va kamchiliklari, foyda va zararlari bo‘yicha o‘rganiladi. Bu interfaol metod tanqidiy, tahliliy, aniq mantiqiy fikrlashni muvaffaqiyatli rivojlantirishga hamda o‘quvchilarning mustaqil g‘oyalari, fikrlarini yozma va og‘zaki shaklda tizimli bayon etish, himoya qilishga imkoniyat yaratadi. “Veer” metodidan ma’ruza mashg‘ulotlarida individual va juftliklardagi ish shaklida, amaliy vaseminar mashg‘ulotlarida kichik guruhlardagi ish shaklida mavzu yuzasidan bilimlarni mustahkamlash, tahlil qilish va taqqoslash maqsadida foydalanish mumkin.

Metodni amalga oshirish tartibi:



trener-o'qituvchi ishtirokchilarni 5-6 kishidan iborat kichik guruhlariga ajratadi;



trening maqsadi, shartlari va tartibi bilan ishtirokchilarni tanishtirgach, har bir guruhga umumiy muammonitahlil qilinishi zarur bo'lgan qismlari tushirilgan tarqatma materiallarni tarqatadi;



har bir guruh o'ziga berilgan muammoni atroflicha tahlil qilib, o'z mulohazalarini tavsiya etilayotgan sxema bo'yicha tarqatmaga yozma bayon qiladi;



navbatdagi bosqichda barcha guruhlar o'z taqdimotlarini o'tkazadilar. Shundan so'ng, trener tomonidan tahlillar umumlashtiriladi, zaruriy axborotlr bilan to'ldiriladi va mavzu yakunlanadi.

4.2. Tahlil.

Muammoli savol					
1-usul		2-usul		3-usul	
afzalligi	kamchiligi	afzalligi	kamchiligi	afzalligi	kamchiligi
Xulosa:					

“Keys-stadi” metodi

«**Keys-stadi**» - inglizcha soʻz boʻlib, («case» – aniq vaziyat, hodisa, «stadi» – oʻrganmoq, tahlil qilmoq) aniq vaziyatlarni oʻrganish, tahlil qilish asosida oʻqitishni amalga oshirishga qaratilgan metod hisoblanadi. Mazkur metod dastlab 1921 yil Garvard universitetida amaliy vaziyatlardan iqtisodiy boshqaruv fanlarini oʻrganishda foydalanish tartibida qoʻllanilgan. Keysdaochoq axborotlardan yoki aniq voqea-hodisadan vaziyat sifatida tahlil uchun foydalanish mumkin.

“Keys metodi” niamalga oshirish bosqichlari

Ishbosqichlari	Faoliyat shakliva mazmuni
1-bosqich: Keys va uning axborot ta'minoti bilan tanishtirish	✓ yakka tartibdagi audio-vizual ish; ✓ keys bilan tanishish(matnli, audio yoki media shaklda); ✓ axborotni umumlashtirish; ✓ axborot tahlili; ✓ muammolarni aniqlash
2-bosqich: Keysni aniqlashtirish va o'quvtopshiriqni belgilash	✓ individual va guruhda ishlash; ✓ muammolarni dolzarblik ierarxiyasini aniqlash; ✓ asosiy muammoli vaziyatni belgilash
3-bosqich: Keysdagi asosiy muammoni tahlil etish orqali o'quv topshirig'iningechimini izlash, hal etish yo'llarini ishlab chiqish	✓ individual va guruhda ishlash; ✓ muqobil echim yo'llarini ishlab chiqish; ✓ har bir echimning imkoniyatlari va to'siqlarini tahlil qilish; ✓ muqobil echimlarni tanlash
4-bosqich: Keys echiminishakllantirish va asoslash, taqdimot.	✓ yakka va guruhda ishlash; ✓ muqobil variantlarni amalda qo'llash imkoniyatlarini asoslash; ✓ ijodiy-loyiha taqdimotini tayyorlash; ✓ yakuniy xulosa va vaziyat echimining amaliy aspektlarini yoritish

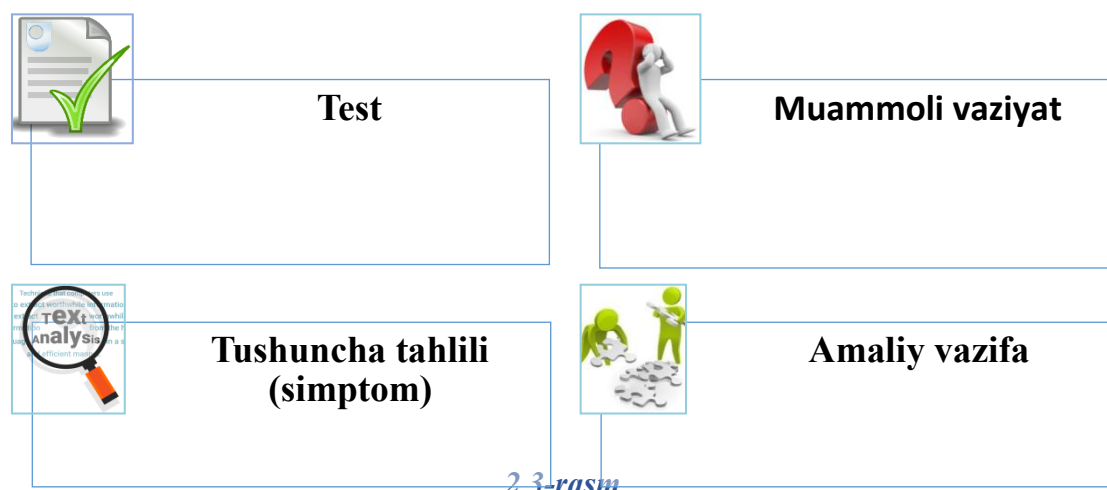
“Assessment”metodi

Metodning maqsadi:mazkur metod ta'lim oluvchilarning bilim darajasinibaholash, nazorat qilish, o'zlashtirish ko'rsatkichi va amaliy ko'nikmalarini tekshirishga yo'naltirilgan.Mazkur texnika orqali ta'lim oluvchilarningbilishfaoliyati turliyo'nalishlar (test, amaliy ko'nikmalar, muammoli vaziyatlar mashqi, qiyosiy tahlil, simptomlarni aniqlash) bo'yicha tashhis qilinadi va baholanadi.

Metodni amalga oshirish tartibi:

“Assessment”lardanma'ruza mashg'ulotlarida talabalarning yoki qatnashchilarning mavjud bilim darajasini o'rganishda, yangi ma'lumotlarni bayon qilishda, seminar,amaliy mashg'ulotlarda esa mavzu yoki ma'lumotlarni o'zlashtirish darajasini baholash, shuningdek, o'z-o'zini baholash maqsadida individual shakldafoydalanish tavsiya etiladi. Shuningdek, o'qituvchining ijodiy yondashuvi hamda o'quv maqsadlaridan kelib chiqib,assessmentga qo'shimcha topshiriqlarni kiritish mumkin.

Har bir katakdagi to'g'ri javob 5 ball yoki 1-5 balgacha baholanishi mumkin.



“Insert” metodi

Metodni amalga oshirish tartibi:

- o‘qituvchi mashg‘ulotga qadar mavzuning asosiy tushunchalari mazmuni yoritilgan matnni tarqatma yoki taqdimot ko‘rinishida tayyorlaydi;
- yangi mavzu mohiyatini yorituvchi matn ta‘lim oluvchilargatarqatiladi yoki taqdimot ko‘rinishida namoyishetiladi;
- ta‘lim oluvchilarindividual tarzda matn bilan tanishib chiqib, o‘z shaxsiy qarashlarini maxsusbelgilar orqali ifodalaydilar. Matn bilan ishlashda talabalar yoki qatnashchilarga quyidagi maxsus belgilardan foydalanish tavsiya etiladi:

Belgilar	Matn
“V” – tanish ma’lumot.	
“?” – mazkur ma’lumotni tushunmadim, izoh kerak.	
“Q”bu ma’lumot men uchun yangilik.	
“– ” bu fikr yoki mazkur ma’lumotga qarshiman?	

Belgilangan vaqt yakunlangach,ta‘lim oluvchilar uchun notanish va tushunarsiz bo‘lgan ma’lumotlar o‘qituvchi tomonidan tahlil qilinib, izohlanadi,ularning mohiyati to‘liq yoritiladi.Savollarga javob beriladi va mashg‘ulot yakunlanadi.

III. NAZARIY MATERIALLAR

1-mavzu: Katta hajmli ma'lumotlarni tashkillash vazifalari va metodlari. (2 soat)

1. Big Data va ma'lumotlar tahlilining asosiy tushunchalari
2. Ma'lumotlar, ma'lumot tiplari. axborot va bilim.
3. Katta ma'lumotlarda data science.
4. Data mining texnologiyalari, metodlari va bosqichlari.
5. Data mining qo'llash soxalari. prognozlash va vizualizatsiya.

1.1. Big data haqida tushuncha.

Big data ya'ni – “Katta ma'lumotlar” umumiy atamasi ma'lumotlarni qayta ishlashning an'anaviy vositalari (masalan, RDBMS - relyatsion ma'lumotlar bazasini boshqarish tizimlari) tomonidan qayta ishlanishi uchun etarlicha katta va murakkab bo'lgan har qanday ma'lumotlar to'plami hisoblanadi. Bugunda keng tarqalgan relyatsion ma'lumotlar bazasini boshqarish tizimlari uzoq vaqt mobaynida universal vosita hisoblangan. Ammo katta ma'lumotlarni qayta ishlashga bo'lgan talab esa buni aksini ko'rsatmoqda, ya'ni relyatsion ma'lumotlar bazasini boshqarish tizimlari ayni vaqtda katta hajmli ma'lumotlarni boshqarishda yetarli imkoniyatlarni bermaydi.

Katta ma'lumotlar (inglizcha big data, ['big 'deɪtə]) - bu 2000-yillarning oxirida paydo bo'lgan va an'anaviy ma'lumotlar bazasini boshqarish tizimlari hamda biznesga muqobil bo'lgan gorizontol ravishda kengaytiriladigan dasturiy vositalar tomonidan katta hajmli va xilma-xillikdagi tuzilgan va tuzilmagan ma'lumotlarning samarali qayta ishlanadigan va Business Intelligence sinfiga kiruvchi ma'lumotlar to'plamidir.

Katta ma'lumotlarni xususiyatlarini 3 ta V xarifi bilan ifodalash mumkin. Ya'ni Volume, Velocity, Variety. Birinchi V-hajm, ikkinchi V-tezlik, uchinchi V-xilma-xillik. Big Data birinchi navbatda hajmga bog'liq. Xozirda ma'lumotlar hajmi eksponent ravishda o'sib bormoqda: masalan, samolyotlar dvigatellariga o'rnatilgan sensor datchiklari orqali har yili 2,5 milliard TB ma'lumot generatsiya qilinadi. Shu bilan birga, ma'lumotlar doimiy ravishda yangilanadi, yangilari hosil bo'ladi va ularni "katta" deb hisoblash uchun yangilanish tezligi ham muhimdir. Misol uchun, dunyo bo'ylab har daqiqada Google qidiruv tizimiga deyarli 2,5 million so'rov yuboriladi. Katta ma'lumotlar loyihalari uchun qiyinchilik - bu ma'lumotlarning yaratilish tezligini yengish va ularni real vaqtda tahlil qilishdir.

Xilma-xillik shuni anglatadiki, Big Data loyihalari turli formatlardagi ma'lumotlarni o'z ichiga olishi kerak: tuzilgan va tuzilmagan ma'lumotlar, matn,

grafikalar, korporativ pochta yoki ijtimoiy media ma'lumotlari, video va boshqa ko'rinishdagi ma'lumotlar bo'lishi mumkin. Ushbu turdagi ma'lumotlarning har biri har xil turdagi tahlillarni va tegishli vositalarni talab qiladi. Ijtimoiy media brend egalariga mijozlarning kayfiyatini tahlil qilishda yordam berishi mumkin va sensorli ma'lumotlar mahsulotning uni yaxshilash uchun ushbu bilimlarni qo'llash uchun eng ko'p ishlatilishi haqida ma'lumot berishi mumkin. Big Data ni aynan mazkur 3 ta asosiy xususiyati bilan ifodalash lozim:

- ❖ Hajm – Volume – to'plamda qancha ma'lumot mavjudligi;
- ❖ Tezlik –Velocity – yangi ma'lumotlar qanday tezlikda generatsiya bo'lishi;
- ❖ Xilma-xillik –Variety – ma'lumotlarni tipi bir-biridan qanchaga farq qilishi.

Demak bundan shuni tushunish mumkinki Big Data texnologiyalarining mohiyati ulkan ma'lumotlar massivlari bilan ishlashini bildiradi (bu Big Data atamasining o'zidan kelib chiqadi). Big Data da ma'lumotlar miqdori asosiy tushuncha, biroq tahlilchilar Big Data ni to'liq formulasini ishlab chiqdilar. Unga ko'ra ettita muhim xususiyat, "7 V" Big Data loyihasining ta'rifida paydo bo'lishi kerak, deb hisoblashadi: hajm, tezlik, xilma-xillik, haqiqat, o'zgaruvchanlik, vizualizatsiya, qiymat. Har bir "V" Big Data ni umumiy tushunish uchun muhimdir.

Ma'lumotlar fanining kontseptsiyasi katta hajmdagi ma'lumotlarni tahlil qilish va ulardagi ma'lumotlarni olish usullaridan foydalanishni o'z ichiga oladi. Katta ma'lumotlar va ma'lumotlar fanlari o'rtasidagi munosabatlar xom neft va neftni qayta ishlash zavodi o'rtasidagi kabi. Ma'lumotlar fani va katta ma'lumotlar statistika va an'anaviy ma'lumotlarni boshqarishdan rivojlangan, ammo hozir alohida fanlar hisoblanadi.

1.2.Katta hajmdagi ma'lumotlarning asosiy terminlari.

Nomidan ko'rinib turibdiki, "katta ma'lumotlar" atamasi shunchaki katta hajmdagi ma'lumotlarni boshqarish va tahlil qilishni anglatadi. McKinsey Instituti hisobotiga ko'ra katta ma'lumotlar: innovatsiyalar, raqobat va mahsuldorlikning navbatdagi chegarasi bo'lib, katta ma'lumotlar atamasi ma'lumotlarni kiritish, saqlash, boshqarish va tahlil qilish uchun odatiy ma'lumotlar bazalaridan (MB) kattaroq bo'lgan ma'lumotlar to'plamini anglatadi.

Biroq, "katta ma'lumotlar" nafaqat katta hajmdagi ma'lumotlarni tahlil qilishni o'z ichiga oladi balki, tashkilotlar katta hajmdagi ma'lumotlarni yaratadilar, lekin ularning aksariyati an'anaviy tuzilgan ma'lumotlar bazasi formatiga mos kelmaydigan formatda taqdim etadilar - bu veb-jurnallar, videolar, matnli hujjatlar, mashina kodi yoki geofazoviy ma'lumotlar bo'lishi mumkin. Bularning barchasi

juda ko'p turli xil ma'lumotlar bazalarilarda, ba'zan tashkilot axborot saqlash qurilmalaridan tashqarida saqlanadi. Natijada, korporatsiyalar o'zlarining katta miqdordagi ma'lumotlariga kirishlari mumkin va bu ma'lumotlar o'rtasida aloqalarni o'rnatish hamda ulardan mazmunli xulosalar chiqarish uchun zarur vositalarga ega bo'lmaydi. Bunga ma'lumotlar tez-tez yangilanib borayotganini qo'shing va axborotni tahlil qilishning an'anaviy usullarini doimiy yangilab turadigan katta hajmdagi ma'lumotlarga mos kela olmaydigan vaziyatga duch kelinadi, bu esa pirovardida katta ma'lumotlar texnologiyalaridan foydalanishga yo'l ochadi.

Ma'lumotlar massivlarini tahlil qilishning turli usullari mavjud bo'lib, ular statistika va informatikadan olingan vositalarga asoslangan bo'lishi mumkin (masalan, mashinali o'qitish). Mashinali o'qitishda ro'yxat to'liq deb da'vo qilmaydi, lekin u turli sohalaridagi eng mashhur yondashuvlarni aks ettiradi. Shu bilan birga, shuni tushunish kerakki, tadqiqotchilar yangi usullarni yaratish va mavjudlarini takomillashtirish ustida ishlashni davom ettirmoqdalar. Bundan tashqari, sanab o'tilgan usullarning ba'zilari faqat katta ma'lumotlar uchun qo'llanilishi shart emas va undan kichikroq massivlar uchun muvaffaqiyatli ishlatilishi mumkin (masalan, A/B testi, regressiya tahlili). Albatta, massiv qanchalik hajmi katta bo'lsa va diversifikatsiyali tahlil qilinsa, natijada shunchalik aniq va sohaga tegishli ma'lumotlarni olish mumkin bo'ladi.

Shuning uchun quyda keltirilgan katta ma'lumotlarga doir atamalarni faqat shu sohaga qo'llaniladi deb cheklash mumkin emas:

A/B testi (A/B testing) . Nazorat namunasi navbat bilan boshqalar bilan solishtiriladigan texnika. Shunday qilib, masalan, marketing taklifiga eng yaxshi iste'molchi javobiga erishish uchun ko'rsatkichlarning optimal kombinatsiyasini aniqlash mumkin. Katta ma'lumotlar sizga juda ko'p iteratsiyalarni o'tkazishga va shu bilan statistik jihatdan muhim natijaga erishishga imkon beradi.

Assotsiatsiya qoidalarini o'rganish (Association rule learning). O'zaro munosabatlarni aniqlash uchun texnikalar to'plami, ya'ni. katta ma'lumotlar massivlaridagi o'zgaruvchilar orasidagi assotsiatsiya qoidalari. Ma'lumotlarni qazib olishda qo'llaniladi.

Tasniflash (Classification). Muayyan bozor segmentida iste'molchilarning xatti-harakatlarini bashorat qilish imkonini beruvchi texnikalar to'plami (sotib olish qarorlari, ishdan chiqish, iste'mol hajmi va boshqalar). Ma'lumotlarni qazib olishda qo'llaniladi.

Klasterli tahlil (Cluster analysis). Oldindan ma'lum bo'lmagan umumiy xususiyatlarni aniqlash orqali ob'ektlarni guruhlariga ajratishning statistik usuli. Ma'lumotlarni qazib olishda qo'llaniladi.

Kraudsorsing (Crowdsourcing). Ko'p sonli manbalardan ma'lumotlarni yig'ish texnikasi.

Ma'lumotlarni birlashtirish va ma'lumotlar integratsiyasi (Data fusion and data integration). Ijtimoiy tarmoq foydalanuvchilarining sharhlarini tahlil qilish va ularni real vaqt rejimidagi savdo natijalari bilan solishtirish imkonini beruvchi texnikalar majmuasi.

Ma'lumotlarni intellektual taxlil qilish (Data mining). Rag'batlanayotgan mahsulot yoki xizmat uchun iste'molchilarning eng sezgir toifalarini aniqlash, eng muvaffaqiyatli xodimlarning xususiyatlarini aniqlash va iste'molchilarning xatti-harakatlar modelini bashorat qilish imkonini beradigan texnikalar to'plami.

Ansamblni o'rganish (Ensemble learning). Ushbu usulda ko'plab bashoratlari modellar qo'llaniladi, bu esa qilingan bashoratlarning sifatini yaxshilaydi.

Genetik algoritmlar (Genetic algorithms). Ushbu tushuncha texnikada amalga oshirilib, xudi biologiyadagi kabi mumkin bo'lgan yechimlar birlasha oladigan va mutatsiyaga uchragan "xromosomalar" sifatida ifodalanadi. Tabiiy evolyutsiya jarayonida eng kuchli shaxs omon qolgini kabi, Big Data da eng kuchli algoritmlar tanlab olinadi.

Mashinali o'qitish (Machine learning). Empirik ma'lumotlarni tahlil qilish asosida o'z-o'zini o'rgatib borish algoritmlarini yaratishga qaratilgan informatikaning yo'nalishi bo'lib "sun'iy intellekt" tarkibida ko'rib chiqiladi. Bu masala Big Dataning asosiy yo'nalishlaridan biri.

Tabiiy tilni qayta ishlash (NLP- Natural language processing). Informatika va tilshunoslikdan olingan tabiiy tilni aniqlash usullari to'plami.

Tarmoq tahlili (Network analysis). Tarmoqlardagi tugunlar orasidagi bog'lanishlarni tahlil qilish usullari to'plami. Ijtimoiy tarmoqlarga kelsak, u alohida foydalanuvchilar, kompaniyalar, jamoalar va boshqalar o'rtasidagi munosabatlarni tahlil qilish imkonini beradi.

Optimallashtirish (Optimization). Bir yoki bir nechta ko'rsatkichlarni yaxshilash uchun murakkab tizimlar va jarayonlarni qayta loyihalashning raqamli usullari to'plami. Strategik qarorlarni qabul qilishda yordam beradi, masalan, bozorga kiritilgan mahsulot tarkibi, investitsiya tahlilini o'tkazish va hokazo.

Shaklni aniqlash (Pattern recognition). Iste'molchilarning xulq-atvor modelini bashorat qilish uchun o'z-o'zini o'rganish elementlari bo'lgan texnikalar to'plami.

Bashoratli modellashtirish (Predictive modeling). Voqealarning rivojlanishi uchun oldindan belgilangan ehtimolli stsenariyning matematik modelini yaratishga imkon beruvchi texnikalar to'plami. Masalan, CRM tizimining ma'lumotlar bazasini obunachilarni provayderlarni o'zgartirishga undaydigan sharoitlar uchun tahlil qilish.

Regressiya (Regression). Bir yoki bir nechta mustaqil o'zgaruvchilar va qaram o'zgaruvchilar o'rtasidagi o'zgarishlarni aniqlash uchun statistik usullar

to'plami. Ko'pincha prognozlash va bashorat qilish uchun ishlatiladi. Ma'lumotlarni qazib olishda qo'llaniladi.

Hissiyotlarni tahlil qilish (Sentiment analysis). Iste'molchi kayfiyatini baholash usullari insonning tabiiy tilini aniqlash texnologiyalariga asoslangan. Ular sizni qiziqtirgan mavzuga (masalan, iste'molchi mahsulotiga) tegishli xabarlarni umumiy axborot oqimidan ajratish imkonini beradi. Keyinchalik, hukmning polaritesini (ijobiy yoki salbiy), hissiylik darajasini va hokazolarni baholang.

Signalni qayta ishlash (Signal processing). Shovqin fonida signalni tanib olish va uni keyingi tahlil qilishga qaratilgan radiotexnikadan olingan texnikalar to'plami.

Fazoviy tahlil (Spatial analysis). Fazoviy ma'lumotlarni tahlil qilish uchun qisman statistikadan olingan texnikalar to'plami - er topologiyasi, geografik koordinatalar, ob'ektlarning geometriyasi. Bu holatda katta ma'lumotlarning manbai ko'pincha geografik axborot tizimlari (GIS) hisoblanadi.

Statistika (Statistics). Ma'lumotlarni to'plash, tartibga solish va sharhlash, shu jumladan so'rovnomalarni ishlab chiqish va tajribalar o'tkazish haqidagi fan. Statistik usullar ko'pincha ma'lum hodisalar o'rtasidagi munosabatlar haqida baho berish uchun ishlatiladi.

Nazorat ostida o'rganish (Supervised learning). Tahlil qilinadigan ma'lumotlar massivlarida funktsional munosabatlarni aniqlash imkonini beruvchi mashinani o'rganish texnologiyalariga asoslangan texnikalar to'plami.

Simulyatsiya (Simulation). Murakkab tizimlarning xatti-harakatlarini modellashtirish ko'pincha rejalashtirishda turli stsenariylarni bashorat qilish, bashorat qilish va ishlab chiqish uchun ishlatiladi.

Vaqt seriyasini tahlil qilish (Time series analysis). Vaqt o'tishi bilan takrorlanadigan ma'lumotlar ketma-ketligini tahlil qilish uchun statistika va raqamli signallarni qayta ishlashdan olingan usullar to'plami. Aniq foydalanishdan biri fond bozorini yoki bemorlarning kasallanishini kuzatishdir.

Nazoratsiz o'rganish (Unsupervised learning). Tahlil qilinayotgan ma'lumotlar to'plamlarida yashirin funktsional munosabatlarni aniqlash imkonini beruvchi mashinani o'rganish texnologiyalariga asoslangan texnikalar to'plami. Kalasterli taxlil (**Cluster Analysis**) bilan umumiy xususiyatlarga ega.

Vizualizatsiya (Визуализация). Metody graficheskogo predstavleniya natijalari tahlili bo'lishish dannyx v video diagrammasi yoki animirovanix izobrajeniy uchun uproshcheniya interpretacii obleggeniya ponimaniya poluchennyh natijalari.

1.3. Katta hajmdagi ma'lumotlarga qo'yiladigan talablar.

Foydalanuvchida yetarlicha mavjud bo'lgan barcha ma'lumotlarga qaramay, ko'plab tashkilotlar ko'proq ma'lumotlarni qayta ishlash bilan bog'liq muammoga duch kelayotganliklarini anglamaydilar yoki shunchaki o'ylab ko'rishga tayyor emaslar. Agar uning mavjud ilovalari va ma'lumotlar bazalari har qanday talab uchun ma'lumotlar miqdori yoki xilma-xilligining keskin o'sishini o'zlashtira olmasa va boshqara olmasa, tashkilot katta ma'lumotlar texnologiyasidan foydalanishi kerak bo'ladi.

Katta ma'lumotlar bilan ishlashga to'g'ri yondashuv o'z vaqtida topilmasa, bu yuqori xarajatlarga, shuningdek, operatsion samaradorlik va raqobatbardoshlikning pasayishiga olib kelishi mumkin. Aksincha, aqlli katta ma'lumotlar strategiyasi tashkilotga katta ma'lumotlar texnologiyalaridan foydalangan holda mavjud bo'lgankatta ish yuklarini oshirish va imkoniyatlardan foydalanish uchun yangi ilovalarni o'rnatish orqali xarajatlarni kamaytirishga qo'shimcha qo'shimcha operatsion foyda olishga yordam beradi.

Katta ma'lumotlar texnologiyalari ma'lumotlarni qayta ishlash siklining barcha bosqichlari uchun yangi vositalarni o'z ichiga olib, ulardan foydalanish texnik va moliyaviy nuqtai nazardan ancha qulaydir. Ushbu vositalardan foydalanib, katta ma'lumotlar to'plamlarini yig'ish va saqlash, shuningdek, yangi qimmatli tahliliy ma'lumotlarni olish uchun ularni qayta ishlash muammolarini hal qilishingiz mumkin. Ko'pgina hollarda, katta ma'lumotlar bilan ishlash standart jarayonni o'z ichiga oladi: dastlabki (xom) ma'lumotlarni to'plashdan toki foydalanishga yaroqli intellectual tahlilga asoslangan ma'lumotlarni olishgacha bo'lgan jarayonlarni qamram oladi.

Bu jarayonni tashkillash quyidagi talablarni bajarishni taqozo etadi:

- **Ma'lumotlarni yig'ish;**
- **Ma'lumotlarni saqlash;**
- **Qayta ishlash va tahlil;**
- **Foydalanish va vizualizatiya.**

Ma'lumotlarni yig'ish - dastlabki (xom) ma'lumotlarni yig'ish (tranzaksiyalar, jurnalga kirishlar, mobil qurilma voqealari va boshqalar orqali) katta ma'lumotlar bilan ishlashda tashkilotlar duch keladigan eng birinchi qiyinchilikdir. Sifatli katta ma'lumotlar platformasi ishlab chiquvchilarga real vaqt rejimidan to'liq ishlov berishgacha bo'lgan vaqtda tuzilgan va har qanday tezlikda keng turdagi ma'lumotlarni olish imkonini beruvchi imkoniyatlarni yaratish ushbu bosqichni asosiy talabidir.

Ma'lumotlarni saqlash - har qanday katta ma'lumotlar platformasi qayta ishlashdan oldin ham, keyin ham ma'lumotlarni saqlash uchun ishonchli, xavfsiz va

kengaytiriladigan omborni-saqlash joyini o'z ichiga olishi kerak. Muayyan talablarga qarab, ko'chirilayotgan ma'lumotlar uchun vaqtinchalik saqlash kerak bo'lgan joylarni yaratish ham shu bo'limni talablaridandir.

Qayta ishlash va tahlil -bosqichida ma'lumotlarni xom holatdan foydalanish mumkin bo'lgan formatga aylantiradi. Bunga odatda maxsus rivojlangan funksiyalar va algoritmlarni saralash, jamlash, birlashtirish yoki qo'llash orqali erishiladi. Olingan ma'lumotlar paketlari keyinchalik qayta ishlash uchun saqlanadi yoki biznes razvedkasi va vizualizatsiya vositalaridan foydalanish uchun taqdim etiladi. Sifatli qayta ishlash va aniq tahlil qilish katta ma'lumotlarni eng muhim talablaridan hisoblanadi.

Foydalanish va vizualizatsiya - katta ma'lumotlar bilan ishlashning asosiy talablaridan bo'lib, amaliy qo'llash uchun ular asosida qimmatli tahliliy xulosalarni olishdir. Ideal holda, katta ma'lumotlar barcha manfaatdor tomonlarga taqdim etilishi kerak, shunda ular o'z-o'ziga xizmat ko'rsatish biznes razvedkasi vositalari va maxsus vizualizatsiya yordamida ma'lumotlar paketlarini tez va oson o'rganishlari, o'zlashtirishlari mumkin. Tahlillarning turiga qarab, yakuniy foydalanuvchilarga statik "bashorat" ma'lumotlari (bashoratli tahlillar bo'lsa) yoki tavsiya etilgan harakatlar (reseptiv tahlillar bo'lsa) ko'rinishida taqdim etilishi mumkin bo'ladi.

Shuningdek, katta ma'lumotlarga qo'yiladigan quyidagi dolzarb talablarni ham ko'rib o'tish lozim:

- **Integratsiya;**
- **Tezkor kirish;**
- **Qamrashning kengligi va turli xil imkoniyatlar;**
- **Ishonchlilik va xavfsizlik;**
- **Yuzlab hamkorlar va yechimlar.**

Integratsiya - katta ma'lumotlar turli xil manbalar va ilovalardan ma'lumotlarni birlashtirishga imkon beradi. Ma'lumotlarni olish, o'zgartirish va yuklash (ETL) vositalari kabi ma'lumotlarni integratsiyalashning an'anaviy mexanizmlari vazifaga mos kelmaydi. Terabayt yoki hatto petabayt ma'lumotlar to'plamini tahlil qilish yangi strategiya va texnologiyalarni talab qiladi. Integratsiya bosqichida ma'lumotlar qo'shiladi, qayta ishlanadi va korporativ tahlilchilar uchun ular bilan ishlashda qulay bo'lgan ko'rinishda formatlanadi.

Tezkor kirish - katta ma'lumotlar texnologiyalari asosan yirik server klasterlarida ishlaydi, bu esa resurslarni tayyorlash va sozlashning uzoq davrlariga olib keladi. Bunda ma'lumotlarga kirishda turli muammolar, masalan qotib qolish, qayta ulanish va kirishni sekinlashishi kabi noqulayliklar yuzaga keldi. Tezkor kirish muhim talab bo'lib, bunday katta ma'lumotlarga misol qilib AWS platformasi bilan ishlashni ko'rsatish mumkin. U platformaga deyarli bir zunda kerakli

infratuzilmani joylashtirish mumkin. Tezkor kirish natijasida mehnat unumdorligi oshadi, yangi g'oyalarni sinovdan o'tkazish jarayoni soddalashtiriladi, loyihalarni ishga tushirish vaqti qisqaradi.

Qamrashning kengligi va turli xil imkoniyatlar- katta ma'lumotlarning ish yuklari tahlil qilish uchun mo'ljallangan ma'lumotlar to'plamlari kabi xilma-xildir. Keng qamrovli va imkoniyatlarga ega platformadan foydalanish hajmi, yangilanish chastotasi yoki ma'lumotlar xilma-xilligidan qat'i nazar, deyarli har qanday katta ma'lumotlar ilovasini yaratishingiz va turli ish yuklarini boshqarishingiz mumkinligini anglatadi. Har yili taqdim etiladigan 50 dan ortiq yangi xizmatlar va yuzlab yangi imkoniyatlar bilan AWS bulutdagi katta ma'lumotlarni yig'ish, saqlash, qayta ishlash, tahlil qilish va vizualizatsiya qilish uchun kerak bo'lgan hamma narsani taqdim etadi. AWS Big Data Platformasi haqida ko'proq bilib oling.

Ishonchlilik va xavfsizlik- katta ma'lumotlar tashqi ta'silarga ko'pincha juda ham sezgir bo'ladi. Shunday qilib, moslashuvchanlikni yo'qotmasdan ma'lumotlar to'plamining xavfsizligini va infratuzilmani himoya qilishni ta'minlash juda muhimdir. Uskuna, tarmoq, dasturiy ta'minot va biznes jarayonlarini qo'llab-quvvatlash sohalarida taqdim etilgan funktsionallik eng qat'iy talablarga javob berishi lozim. ISO 27001 va FedRAMP, US DoD SRG va PCI DSS kabi sertifikatlariga muvofiq ma'lumotlar xavfsizligini ta'minlashi lozim.

Yuzlab hamkorlar va yechimlar- agar katta ma'lumotlar bilan ishlashda va yechimlar topishda o'z mahoratingiz yetarli bo'lmasa, hamkorlar katta ma'lumotlarni tezroq ishga tushirishga, masalalarini yechishga va innovasiyalar olib kirishga yordam berishi mumkin. Katta ma'lumotlar sohasida yordam beradigan maslahatchi hamkorni topish yoki ma'lumotlarni boshqarish bo'yicha turli ilovalar va vositalardan foydalanish imkoniyati mavjud.

1.4.Katta hajmdagi ma'lumotlar bilan ishlaydigan axborot tizimlari.

Korxonalar va muassasalar, kichik va yirik kompaniyalar uchun tashkiliy o'sish muhim bo'lib, u osij jarayonida turli xil muammolarga duch keladi. Kengayish bilan ko'proq mas'uliyat, strategik rejalashtirish, integratsiya va yaxshi muloqot qilish zarurati paydo bo'ladi.

Masalan, boshlang'ishida 1 yoki 2 xodimdan 500 yagacha xodimlarni ishlatish imkoniga qadar rivojlanib borishga erishgan korxona uchun ma'lumotlarni qayta ishlay oladigan aniq funktsiyalarni bajarish va samaradorlikni optimallashtirish uchun yangi biznes birliklari yaratilishi talab etiladi. Menejerlar har bir xodimni o'z vaqtida maqsadlarga erishishini ta'minlash uchun barcha bo'limni kuzatishi shart bo'lib qoladi. Rahbar esa mijozlarni jalb qilishdan boshqa hech narsa istamaydigan

raqobatchilarni kuzatishi kerak bo'ladi. Bu esa qancha turdagi ko'plab ma'lumotlarni boshqarish degani.

Ushbu yangi talablar xodimlar va menejerlarga muloqot qilish, hamkorlik qilish va tranzaktsiyalarni tezda qayta ishlashga imkon berish uchun sun'iy intellektdan foydalanadigan axborot tizimiga ega bo'lishni talab qiladi.

Bundan tashqari, rivojlanayotgan va tashkil etilgan kompaniya har bir bo'linmaga tuzilgan qarorlarni qabul qilishda va muammolarni hal qilish ko'nikmalarini oshirishda yordam beradigan turli xil axborot tizimlaridan foydalanishdan foyda ko'rishi kerak.

Bu masalalar bugungi kunda Big Data masalalari hisoblanadi. Xo'sh Big Data ya'ni katta ma'lumotlar bilan ishlaydigan axborot tizimlari qaysi sohalar va yo'nalishlarda qo'llaniladi?

Javob oddiy bu deyarli barcha sohalarda deb aytish mumkin. Keling bunday axborot tizimlarini ko'rib chiqamiz:

Katta ma'lumotlar bilan ishlaydigan axborot tizimlarini quyidagi sohalarga bo'lib chiqishimiz mumkin:

Operatsion menejment-tranzaktsiyalarni qayta ishlash tizimi - ofislarni avtomatlashtirish tizimi va bilimlarni boshqarish tizimini o'z ichiga oladi, tuzilgan ma'lumotlarni boshqaradi va ishchilar va xodimlar tomonidan kundalik operatsiyalarni soddalashtirish uchun foydalaniladi

Taktik boshqaruv tizimi - biznes bo'limi menejerlari tomonidan qo'llaniladi, yarim tuzilgan ma'lumotlarni boshqaradi va boshqaruv axborot tizimlarini o'z ichiga oladi

Strategik menejment tizimi - rahbarlar tomonidan qo'llaniladi, tuzilmasiz ma'lumotlarni boshqaradi va qarorlarni qo'llab-quvvatlash tizimi va ijro etuvchi qo'llab-quvvatlash tizimini o'z ichiga oladi.

Aksariyat korxonalar oltita turli xil axborot texnologiyalari tizimlaridan foydalanadilar, ularning har biri ma'lum bir biznes birligini yoki tashkiliy darajasini boshqarishda yordam beradigan funktsional imkoniyatlarga ega.

Ishbilarmonlik muhiti ma'lumotlarga bo'lgan keng talablarga ega bo'lganligi sababli, biznes-razvedka texnologiyalari tizimlari har bir bo'limga barcha ma'lumotlarni boshqarish va tartibga solishda yordam beradi, bu esa birlik a'zolariga asosiy maqsadlarga erishishda yordam beradi.

Agar IS tomonidan to'plangan ma'lumotlar tegishli va aniq bo'lsa, tashkilot undan vazifalarni soddalashtirish, samarasizlikni aniqlash va mijozlarga xizmat ko'rsatishni yaxshilash uchun foydalanishi mumkin.

Muvaffaqiyatli kompaniyalar odatda tashkilot ma'lumotlarining har bir jihati to'g'ri boshqarilishini va qaror qabul qilish va muammolarni hal qilishni yaxshilash uchun ishlatilishini ta'minlash uchun oltita turli xil tizimlardan foydalanadilar.

Bu kompaniyaga raqobatbardoshlikni saqlab qolish, o'sish imkoniyatlarini topish va muvofiqlik maqsadida moliyaviy va tranzaktsion ma'lumotlarning aniq auditorlik izini saqlashga imkon beradi.

Olti turdagi axborot tizimlari va ularning qanday ishlashi haqida umumiy ma'lumotlar quyida keltirilgan:

1. Tranzaktsiyalarni qayta ishlash tizimlari (**Transaction Processing Systems**)
2. Ofisni avtomatlashtirish tizimlari (**Office Automation Systems**)
3. Bilimlarni boshqarish tizimlari (**Knowledge Management Systems**)
4. Boshqaruv axborot tizimlari (**Management Information Systems**)
5. Qarorlarni qo'llab-quvvatlash tizimlari (**Decision Support Systems**)
6. Ijro etishni qo'llab-quvvatlash tizimi (**Executive Support System**)

Tranzaktsiyalarni qayta ishlash tizimlari (Transaction Processing Systems).

Bitim barcha mahsulotlarni sotib olish va sotishni o'z ichiga oladi xizmatlar, har qanday kundalik biznes operatsiyalari yoki kompaniyani boshqarish uchun zarur bo'lgan faoliyat bilan bir qatorda.

Miqdori va amalga bitimlar turlari farq, sanoat va hajmiga qarab / kompaniya ko'lami. Odatda operatsiyalarga misol sifatida hisob-kitob mijozlari, bank depozitlari, yangi yollash ma'lumotlari, inventarizatsiyani hisoblash yoki mijoz va mijoz munosabatlarini boshqarish ma'lumotlarini qayd etish kiradi.

Tranzaktsiyalarni qayta ishlash tizimi barcha shartnomaviy, tranzaktsion va mijozlar bilan munosabatlar ma'lumotlari xavfsiz joyda saqlanishini va unga muhtoj bo'lgan har bir kishi uchun qulay bo'lishini ta'minlaydi. Shuningdek, u savdo buyurtmalarini kiritish, ish haqi, yuk tashish, sotishni boshqarish yoki operatsiyalarni bajarish uchun zarur bo'lgan boshqa muntazam operatsiyalarni qayta ishlashga yordam beradi.

TPS-dan foydalangan holda, tashkilotlar foydalanuvchi/mijoz ma'lumotlarida yuqori darajadagi ishonchlilik va aniqlikka ega bo'lishi mumkin, shu bilan birga inson xatosi potentsialini minimallashtiradi.

Ofisni avtomatlashtirish tizimlari (Office Automation Systems)

Ofisni avtomatlashtirish tizimi (OAT) - bu ish yuritish va boshqaruv vazifalarini bajarish uchun zarur bo'lgan turli xil vositalar, texnologiyalar va odamlar tarmog'i.

OAT tomonidan bajariladigan funktsiyalarning odatiy misollariga hujjatlarni chop etish, hujjatlarni pochta orqali yuborish, pochta orqali yuborish, kompaniya taqvimini yuritish va hisobotlarni tayyorlash kiradi. Birinchi navbatda, ofis avtomatlashtirish tizimi har bir kishi bir vazifani bajarish uchun hamkorlik mumkin,

shuning uchun turli idoralar o'rtasida muloqot oshirishga yordam.

OAT barcha aloqa ma'lumotlariga osongina kirish va bitta markazlashtirilgan joyda bo'lishini ta'minlash uchun elektron pochta yoki matnni qayta ishlash dasturlari bilan birlashishi mumkin. Ofisni avtomatlashtirish tizimidan foydalangan holda, korxonalar ishchilar o'rtasidagi aloqani yaxshilashi, boshqaruv faoliyatini soddalashtirishi va bilimlarni boshqarishni optimallashtirishi mumkin.

Bilimlarni boshqarish tizimlari (Knowledge Management Systems)

Bilimlarni boshqarish tizimi (BBT) foydalanuvchilarga o'z bilimlarini oshirishga va vazifalarni bajarish uchun hamkorlik harakatlarini optimallashtirishga yordam beradigan ma'lumotlarni saqlaydi va chiqaradi. Bilimlarni boshqarish tizimida topilgan hujjatlarga xodimlarni o'qitish materiallari, kompaniya siyosati va protseduralari yoki mijozlarning savollariga javoblar kiradi.

BBT tashkilot bilan bog'liq xodimlar, mijozlar, menejment va boshqa turli manfaatdor tomonlar tomonidan qo'llaniladi. Bu texnik qobiliyatlarni kompaniya bo'ylab birlashtirilishini ta'minlaydi, shu bilan birga xodimlarga ko'rgan ma'lumotlarini tushunishga yordam berish uchun ingl.

Ushbu axborot tizimi, shuningdek, o'z rollarini bajarish uchun tashqi bilimga muhtoj bo'lgan ishchilar talab qiladigan tashqi ma'lumotlarga intuitiv kirish imkoniyatini beradi. Misol uchun, bir BBT bir mijozga pitching qachon bir savdo jamoasi a'zosi uning/strategiyasini optimallashtirish yordam raqib ma'lumotlarni ushlab turing mumkin.

Chunki BBT tajriba almashadi va muhim savollarga javob beradi, ulardan foydalanish jamoa a'zolari o'rtasidagi aloqani yaxshilaydi va ishlash maqsadlariga erishishda barchaga yordam beradi.

Qarorlarni qo'llab-quvvatlash tizimlari (Decision Support Systems)

Qarorlarni qo'llab-quvvatlash tizimi (QQQT) boshqaruv qarorlarini qabul qilishda yordam berish uchun ma'lumotlarni qayta ishlaydi. U o'z vaqtida tegishli harakatlarni amalga oshirish uchun menejment uchun zarur bo'lgan ma'lumotlarni saqlaydi va to'playdi. Masalan, bank menejeri QQQT-dan foydalanib, aylanma kredit tendentsiyalarini baholab, qaysi yillik kredit maqsadlarini qondirishni aniqlaydi.

Qaror modellari katta hajmdagi ma'lumotlarni tahlil qilish va umumlashtirish va uni tushunarli qiladigan ko'rgazmaga qo'yish uchun dasturlashtirilgan.

QQQT interaktiv bo'lgani uchun menejment ma'lumotlarni osongina qo'shishi yoki o'chirishi va muhim savollarni berishi mumkin. Bu o'rta menejment uchun kompaniyaning maqsadlariga javob berishini ta'minlaydigan to'g'ri tanlov qilish uchun zarur bo'lgan dalillarni taqdim etadi.

Ijro etishni qo'llab-quvvatlash tizimi (Executive Support System)

Ijro etishni qo'llab-quvvatlash tizimlari (IEQQT) xuddi QQQT o'xshaydi, lekin birinchi navbatda ijro etuvchi rahbarlar va egalari tomonidan qaror qabul qilishni optimallashtirish uchun foydalaniladi.

Ekspert tizimi korxona rahbarlariga odatiy bo'lmagan savollarga javob topishda yordam beradi, shunda ular kompaniyaning dunyoqarashi va ish faoliyatini yaxshilaydigan tanlov qilishlari mumkin. QQQT farqli o'laroq, ijro etuvchi qo'llab-quvvatlash tizimi yaxshi telekommunikatsiya funktsiyalarini va katta hisoblash funktsiyalarini ta'minlaydi.

Grafik dasturiy ta'minot soliq qoidalari, yangi raqobatbardosh startaplar, ichki muvofiqlik muammolari va boshqa tegishli ijro ma'lumotlari haqidagi ma'lumotlarni namoyish qilish uchun ESS tarkibiga kiritilgan. Bu rahbarlarga ichki ko'rsatkichlarni kuzatish, raqobatni kuzatish va o'sish imkoniyatlarini aniqlash imkonini beradi.

1.5. Ma'lumotlar tahlilining asosiy tushunchalari

Sodda qilib aytganda, ma'lumotlarni tahlil qilish deganda ma'lumotlarni qayta ishlash algoritmlari bilan bog'liq bo'lgan va har bir kiruvchi ob'ekt haqidagi ma'lumotlarga aniq javob bermaydigan usullar va ilovalar to'plamini tushunish mumkin.

Bu ularni saralash yoki lug'at bo'yicha topishni amalga oshirish kabi klassik algoritmlardan farq qiladi. Klassik algoritmnini aniq amalga oshirish uning bajarilish vaqti va eggalagtan xotira hajmiga bog'liq, ammo uni qo'llashning kutilgan natijasi qat'iy belgilanadi.

Misol, raqamlarni taniydigan neyron tarmoqdan rasm bilan kiruvchi 6 raqamini aniqlashda 6-javobni qo'lda yozilgan oltita rasm bilan kiruvchi tasvirlardan olamiz, ammo bu aniq natijani beradi deb ayta olmaymiz. Bundan tashqari, har qanday (so'zning aniq ma'nosida) neyron tarmoq ba'zan to'g'ri kirish ma'lumotlarini ba'zi variantlarida xatoga yo'l qo'yadi.

Muammoning bunday bayonotini va uni yechishda ishlatiladigan usullar va algoritmlarni klassik (deterministik) dan farqli ravishda, noaniq (deterministik bo'lmagan) deb ataymiz. Yani, kiruvchi ma'lumotlardan talab qilingan aniq ma'lumotni ajratib olish murakkab masala sanaladi. Ushbu muammo ma'lumotlar tahlil orqali hal etiladi.

Ma'lumotlarni tahlil -bu eksperimental (keng ma'noda) ma'lumotlardan bilim olishning eng keng tarqalgan matematik usullari va hisoblash algoritmlarini qurish va tadqiq qilish bilan shug'ullanadigan, kerakli ma'lumotlarni olish va qaror qabul qilish uchun ma'lumotlarni o'rganish, saralash, o'zgartirish va modellashtirish jarayonini amalga oshiruvchi matematika va informatika sohasidir.

Ma'lumotlarni tahlil qilish ko'plab jihatlar va yondashuvlarga ega, fan va faoliyatning turli sohalarida turli usullarni qamrab oladi. Jumladan, katta ma'lumotlar (Big Data), ma'lumotlarni intellektual tahlil qilish (Data Mining), mashinali o'qitish (Machine Learning), algoritmlar, matematika kabi fan sohalarida bo'lishi mumkin.

Ma'lumotlarni intellektual tahlil qilish (Data Mining)-bu ma'lumotlarni tavsiflashga emas, balki modellashtirish va ochishga qaratilgan ma'lumotlarni tahlil qilishning o'ziga xos usulidir. Biznes tahlillari yig'ilishga asoslangan ma'lumotlarni tahlil qilishni o'z ichiga oladi. Statistik ma'noda, ba'zilar ma'lumotlar tahlilini tavsiflovchi statistikaga, tadqiqot ma'lumotlarini tahlil qilishga va statistik gipotezalarni tekshirishga ajratadilar. Tadqiqot ma'lumotlarini tahlil qilish ma'lumotlarning yangi xususiyatlarini kashf etish bilan shug'ullanadi va statistik gipotezalarni tekshirish mavjud gipotezalarni tasdiqlash yoki rad etish bilan bog'liq. Bashoratli tahlil bashorat qilish yoki tasniflash uchun statistik yoki tarkibiy modellarni qo'llashga qaratilgan, matn tahlili esa tuzilmagan ma'lumotlarga tegishli matn manbalaridan ma'lumotlarni olish va tasniflash uchun statistik, lingvistik va tarkibiy usullarni qo'llaydi. Bularning barchasi ma'lumotlarni tahlil qilishning turlari hisoblanadi.

Ma'lumotlar integratsiyasi ma'lumotlar tahlilining kashfiyotchisi va ma'lumotlar tahlilining o'zi ma'lumotlarni vizualizatsiya qilish va ma'lumotlarni tarqatish bilan chambarchas bog'liq. Ma'lumotlarni tahlil qilish atamasi ba'zan sinonim sifatida ishlatiladi ma'lumotlarni.

Umuman olganda, ma'lumotlarni tahlil qilish loyihasini ishlab chiqish tsikli quyidagichadir.

- Muammoning bayonini, mumkin bo'lgan ma'lumot manbalarini o'rganish.
- Matematik tilda qayta tuzish, bashorat qilish sifati ko'rsatkichlarini tanlash.
- Haqiqiy muhitda o'rganish va (hech bo'lmaganda sinov) foydalanish uchun quvvur liniyasini yozish.
- Muammoni hal qiluvchi Evristikani yoki oddiy mashinani o'rganish algoritmini yozish.
- Agar kerak bo'lsa, algoritm sifatini yaxshilash, o'lchovlarni aniqlashtirish va qo'shimcha ma'lumotlarni jalb qilish mumkin.

2-mavzu: Qarorni qo'llab quvvatlash tizimlari. OLAP- tizimlari.

(2 soat amaliy)

Reja:

1. Qarorni qo'llab quvvatlash tizimlarining asosiy tushunchalari.
2. OLAP vositalari va tizimlari.

3. Qaror daraxtini qurish.

1. Qarorni qo'llab quvvatlash tizimlarining asosiy tushunchalari.

Qarorlarni qo'llab-quvvatlash tizimi (inlizchada Decision Support System-DSS) qisqacha QQQT desak ham bo'ladi - bu kompaniyaning qaror qabul qilish imkoniyatlarini yaxshilash uchun foydalaniladigan kompyuter dasturi. U katta hajmdagi ma'lumotlarni tahlil qiladi va tashkilotga mavjud bo'lgan eng yaxshi variantlarni taqdim etadi. Qarorlarni qo'llab-quvvatlash tizimlari foydalanuvchilarga odatdagi hisobotlar va xulosalardan tashqari ma'lumotlarni taqdim etish uchun turli sohalar va manbalardan ma'lumotlar va bilimlarni birlashtiradi. Bu odamlarga ong-li qarorlar qabul qilishda yordam berish uchun mo'ljallangan. Qarorni qo'llab-quvvatlash ilovasi to'plashi va taqdim etishi mumkin bo'lgan odatiy ma'lumotlarga quyidagilar kiradi:

- bir hafta va keyingi hafta o'rtasidagi qiyosiy savdo ko'rsatkichlari;
- yangi mahsulotni sotish bo'yicha taxminlar asosida prognoz qilingan daromad ko'rsatkichlari; va
- turli qarorlarning oqibatlari.

Qarorlarni qo'llab-quvvatlash tizimi operatsion dasturdan farqli o'laroq, axborot ilovasidir. Axborot ilovalari foydalanuvchilarga turli xil ma'lumotlar manbalariga asoslangan tegishli ma'lumotlarni taqdim etadi, bu esa yaxshi ma'lumotga ega bo'lgan qarorlar qabul qilishni qo'llab-quvvatlaydi. Operatsion ilovalar, aksincha, biznes operatsiyalari tafsilotlarini, shu jumladan biznesning qarorlarni qo'llab-quvvatlash ehtiyojlari uchun zarur bo'lgan ma'lumotlarni yozib oladi.

Qarorlarni qo'llab-quvvatlash tizimining tarkibiy qismlari Oddiy QQQT uch xil qismdan iborat: ma'lumotlar bazasi, dasturiy ta'minot va foydalanuvchi interfeysi.

Bilimlar bazasi. Bilimlar bazasi qarorlarni qo'llab-quvvatlash tizimining ma'lumotlar bazasining ajralmas qismi bo'lib, ichki va tashqi manbalardan ma'lumotlarni o'z ichiga oladi. Bu muayyan mavzular bilan bog'liq ma'lumotlar kutubxonasi va DSSning bir qismi bo'lib, tizimning fikrlash mexanizmi tomonidan harakat yo'nalishini aniqlash uchun foydalaniladigan ma'lumotlarni saqlaydi.

Dasturiy ta'minot tizimi. Dasturiy ta'minot tizimi model boshqaruv tizimlaridan iborat. Model - bu tizim qanday ishlashini va uni qanday yaxshilash mumkinligini tushunish maqsadida real tizimning simulyatsiyasi. Tashkilotlar tizimdagi turli xil tuzatishlar bilan natijalar qanday o'zgarishini bashorat qilish uchun modellardan foydalanadilar. Misol uchun, modellar haqiqiy hayotda to'liq o'rganish uchun juda murakkab, juda qimmat yoki juda xavfli tizimlarni tushunish uchun

foydali bo'lishi mumkin. Ilmiy tadqiqotlar, muhandislik sinovlari, ob-havoni bashorat qilish va boshqa ko'plab ilovalar uchun ishlatiladigan kompyuter simulyatsiyalari ortidagi g'oya shu. Modellardan taklif qilinayotgan yangi texnologiya, rejalashtirilgan zavod yoki biznes ta'minot zanjiri kabi hali mavjud bo'lmagan tizimlarni ko'rsatish va o'rganish uchun ham foydalanish mumkin. Korxonalar, shuningdek, biznes qarorlarini qabul qilishda yordam berish uchun tizimdagi turli xil o'zgarishlar (masalan, siyosatlar, xavflar va qoidalar) natijalarini bashorat qilish uchun modellardan foydalanadilar.

Foydalanuvchi interfeysi. Foydalanuvchi interfeysi tizimni oson navigatsiya qilish imkonini beradi. Qaror qabul qilishni qo'llab-quvvatlash tizimining foydalanuvchi interfeysining asosiy maqsadi foydalanuvchiga unda saqlangan ma'lumotlarni boshqarishni osonlashtirishdir. Korxonalar oxirgi foydalanuvchilar uchun QQQT tranzaksiyalarining samaradorligini baholash uchun interfeysdan foydalanishlari mumkin. QQQT interfeyslariga oddiy oynalar, murakkab menyuga asoslangan interfeyslar va buyruq qatori interfeyslari kiradi.

Aqlli qarorlarni qo'llab-quvvatlash tizimi (AQQT) Foydalanuvchilar, shuningdek, sun'iy intellekti (AI) qarorlarni qo'llab-quvvatlash tizimlariga aylantirishi mumkin. Aqlli qarorlarni qo'llab-quvvatlash tizimlari (AQQT) deb ataladigan AI tushunchalarni olish va yaxshiroq qaror qabul qilish uchun tavsiyalar berish uchun katta hajmdagi ma'lumotlarni ishlab chiqaradi va qayta ishlaydi. U buni bir nechta ma'lumotlar manbalarini tahlil qilish va inson qarorlarini qabul qilish qobiliyatiga taqlid qilish uchun naqshlar, tendentsiyalar va uyushmalarni aniqlash orqali amalga oshiradi. Inson maslahatchisiga o'xshash harakat qilish uchun mo'ljallangan IDSS muammolarni aniqlash va bartaraf etish hamda mumkin bo'lgan yechimlarni taqdim etish va baholash orqali qaror qabul qiluvchilarni qo'llab-quvvatlash uchun ma'lumotlarni to'playdi va tahlil qiladi. DSS ning sun'iy intellekt komponenti kompyuter tizimi sifatida ma'lumotlarni yanada samarali qayta ishlash va tahlil qilish bilan birga, inson imkoniyatlarini iloji boricha taqlid qiladi. AQQQT bilimlar bazasi, mashinani o'rganish, ma'lumotlarni qidirish va foydalanuvchi interfeysi kabi ilg'or imkoniyatlarni o'z ichiga olishi mumkin. IDSS tatbiqlariga misollar orasida moslashuvchan yoki aqlli ishlab chiqarish tizimlari, marketing qarorlarini qo'llab-quvvatlashning aqlli tizimlari va tibbiy diagnostika tizimlari kiradi. Har xil turdagi qarorlarni qo'llab-quvvatlash tizimi kompaniyaning turli sohalarida qaror qabul qilish imkoniyatlarini yaxshilashi mumkin.

Qarorlarni qo'llab-quvvatlash tizimlarining turlari Qarorlarni qo'llab-quvvatlash tizimlarini har biri o'zining asosiy ma'lumot manbalariga asoslangan toifalarga bo'lish mumkin.

Ma'lumotlarga asoslangan QQQT. Ma'lumotlarga asoslangan QQQT - bu ichki ma'lumotlar bazalari yoki tashqi ma'lumotlar bazalari ma'lumotlari asosida

qarorlar qabul qiladigan kompyuter dasturi. Odatda, ma'lumotlarga asoslangan QQQT tendentsiyalarni va naqshlarni aniqlash uchun ma'lumotlarni qidirish usullaridan foydalanadi, bu esa kelajakdagi voqealarni bashorat qilish imkonini beradi. Korxonalar ko'pincha inventarizatsiya, sotish va boshqa biznes jarayonlari bo'yicha qaror qabul qilishda yordam berish uchun ma'lumotlarga asoslangan QQQT'lardan foydalanadilar. Ba'zilar kelajakda jinoiy xatti-harakatlar ehtimolini bashorat qilish kabi davlat sektorida qarorlar qabul qilishda yordam berish uchun ishlatiladi.

Modelga asoslangan QQQT. Asosiy qaror modeli asosida qurilgan, modelga asoslangan qarorlarni qo'llab-quvvatlash tizimlari ushbu talablarga javob beradigan turli stsenariylarni tahlil qilishda yordam berish uchun oldindan belgilangan foydalanuvchi talablari to'plamiga muvofiq moslashtirilgan. Misol uchun, modelga asoslangan QQQT moliyaviy hisobotlarni rejalashtirish yoki ishlab chiqishda yordam berishi mumkin.

Muloqotga asoslangan va guruhli DSS. Muloqotga asoslangan va guruh qarorlarini qo'llab-quvvatlash tizimi turli xil aloqa vositalaridan foydalanadi -- elektron pochta, lahzali xabar almashish yoki ovozli chat -- bitta vazifada bir nechta odamga ishlash imkonini beradi. Ushbu turdagi QQQTning maqsadi foydalanuvchilar va tizim o'rtasidagi hamkorlikni oshirish va tizimning umumiy samaradorligi va samaradorligini oshirishdir.

Bilimga asoslangan QQQT. Ushbu turdagi qarorlarni qo'llab-quvvatlash tizimida tizimni boshqaradigan ma'lumotlar bilimlarni boshqarish tizimi tomonidan doimiy ravishda yangilanadigan va qo'llab-quvvatlanadigan bilimlar bazasida joylashgan. Bilimga asoslangan QQQT foydalanuvchilarga kompaniyaning biznes jarayonlari va bilimlariga mos keladigan ma'lumotlarni taqdim etadi.

Hujjatga asoslangan QQQT. Hujjatga asoslangan QQQT - ma'lumotlarni olish uchun hujjatlardan foydalanadigan axborotni boshqarish tizimining bir turi. Hujjatga asoslangan QQQTlar foydalanuvchilarga veb-sahifalar yoki ma'lumotlar bazalarini qidirish yoki muayyan qidiruv so'zlarini topish imkonini beradi. Hujjatga asoslangan QQQT tomonidan foydalaniladigan hujjatlarga misollar siyosat va protseduralar, yig'ilish bayonnomalari va korporativ yozuvlarni o'z ichiga oladi.

Bilimlar bazasi bilimga asoslangan qarorlarni qo'llab-quvvatlash tizimining ajralmas qismidir.

Qarorlarni qo'llab-quvvatlash tizimiga misollar. Tashkilotlar qarorlarni qo'llab-quvvatlash tizimlaridan bir nechta turli kontekstlarda foydalanadilar, jumladan, quyidagilar:

- **GPS marshrutlash.** GPS marshrutini rejalashtirish odatiy DSSga misoldir. U masofa, haydash vaqti va narxi kabi omillarni hisobga olgan holda turli

marshrutlarni taqqoslaydi. GPS navigatsiya tizimi , shuningdek, foydalanuv-chilarga muqobil marshrutlarni tanlash, ularni xaritada ko'rsatish va bos-qichma-bosqich ko'rsatmalar berish imkonini beradi.

- **ERP boshqaruv paneli.** ERP (korxona resurslarini rejalashtirish) asboblari paneli ishlab chiqarish va biznes jarayonlaridagi o'zgarishlarni vizualizatsiya qilish, joriy biznes samaradorligini belgilangan maqsadlarga nisbatan monitoring qilish va yaxshilash uchun yo'nalishlarni aniqlash uchun qarorlarni qo'llab-quvvatlash tizimidan foydalanishi mumkin. ERP boshqaruv paneli biznes egalariga kompaniyaning eng muhim raqamlari va ko'rsatkichlarining suratini ko'rish imkonini beradi.

- **Klinik qarorlarni qo'llab-quvvatlash tizimi.** Klinik qarorlarni qo'llab-quvvatlash tizimi (KQQQT) shifokorlarga eng yaxshi tibbiy qarorlarni qabul qilishda yordam berish uchun ilg'or qaror qabul qilish algoritmlaridan foydalanadigan dasturiy ta'minot dasturidir. Sog'liqni saqlash mutaxassislari ko'pincha bemorlarning yozuvlari va test natijalarini sharhlash va eng yaxshi davolash rejasini hisoblash uchun ulardan foydalanadilar. Sog'liqni saqlash sohasidagi KQQQT provayderlarga ma'lum testlar paytida anormalliklarni aniqlashga yordam beradi, shuningdek, bemorlarda nojo'ya reaksiyalar bor-yoqligini aniqlash uchun ma'lum protseduralardan so'ng monitoringini olib boradi.

2. OLAP vositalari va tizimlari

OLAP yangi tushuncha emas va o'nlab yillar davomida ishlatilgan. Aslida, texnologiyaning kelib chiqishi 1962 yilda boshlangan. Ammo bu atama faqat 1993 yilda ma'lumotlar bazasi muallifi Ted Koddom tomonidan ishlab chiqilgan bo'lib, u mahsulot uchun 12 ta qoidalarini o'rnatgan. Ko'pgina boshqa dasturlarda bo'lgani kabi, kontseptsiya ham bir necha evolyutsiya bosqichlaridan o'tdi.

OLAP texnologiyasi tarixining o'zi 1970 yildan boshlanadi axborot resurslari Express va birinchi Olap server. Ular 1995 yilda Oracle tomonidan sotib olingan va keyinchalik taniqli kompyuter markasi o'z ma'lumotlar bazasida taqdim etgan ko'p o'lchovli hisoblash dvigatelini onlayn analitik qayta ishlashning asosiga aylandi. 1992 yilda yana bir taniqli onlayn tahliliy qayta ishlash mahsuloti Essbase Arbor Software tomonidan chiqarildi (2007 yilda Oracle tomonidan sotib olingan).

1998 yilda Microsoft MS Analysis Services-ni onlayn ma'lumotlarni qayta ishlash analitik serverini chiqardi. Bu texnologiyaning ommalashishiga hissa qo'shdi va boshqa mahsulotlarni ishlab chiqarishga turtki berdi. Bugungi kunda Olap dasturlarini taklif qiluvchi bir necha dunyoga mashhur etkazib beruvchilar mavjud, jumladan IBM, SAS, SAP, Essbase, Microsoft, Oracle, Iccube.

OLAP mexanizmi bugungi kunda ma'lumotlarni tahlil qilishning eng mashhur usullaridan biridir. Ushbu muammoni hal qilishda ikkita asosiy yondashuv mavjud. Birinchisi, ko'p o'lchovli OLAP (MOLAP) deb nomlanadi, bu mexanizmni server tomonidagi ko'p o'lchovli ma'lumotlar bazasidan foydalangan holda amalga oshiradi, ikkinchisi, Relational OLAP (ROLAP) esa kublarni tezda yaratadi SQL so'rovlari relyatsion DBMS-ga. Ushbu yondashuvlarning har biri o'zining ijobiy va salbiy tomonlariga ega. Ular qiyosiy tahlil ushbu maqola doirasidan tashqarida. Biz ish stoli ROLAP modulining yadrosini amalga oshirishni tasvirlaymiz. Ushbu muammo Borland Delphi-ga kiritilgan Decision Cube komponentlari asosida qurilgan ROLAP tizimining qo'llanilishidan keyin paydo bo'ldi. Afsuski, ushbu komponentlar to'plamidan foydalanish katta hajmdagi ma'lumotlarning yomon ishlashini ko'rsatdi. Ushbu muammoni kublar bilan oziqlantirishdan oldin iloji boricha ko'proq ma'lumotlarni qisqartirishga urinib, kamaytirish mumkin. Ammo bu har doim ham etarli emas. Internetda va matbuotda siz OLAP tizimlari haqida juda ko'p ma'lumotni topishingiz mumkin, ammo uning ichki ishlashi haqida deyarli hech qaerda aytilmagan. Shuning uchun, muammolarning aksariyati sinov va xato bilan hal qilindi.

OLAP tizimlaridan foydalanish tashkilotni boshqarishning strategik darajasini avtomatlashtirishga imkon beradi. OLAP (Onlayn analitik ishlov berish) - bu ma'lumotlarni qayta ishlash va tadqiq qilishning kuchli texnologiyasi. OLAP texnologiyasi asosida qurilgan tizimlar deyarli amal qiladi cheksiz imkoniyatlar hisobotlarni tuzish, murakkab analitik hisob-kitoblarni bajarish, prognozlar va stsenariylarni yaratish, rejalarning ko'plab variantlarini ishlab chiqish uchun.

To'liq huquqli OLAP tizimlari 90-yillarning boshlarida qarorlarni qo'llab-quvvatlovchi axborot tizimlarining rivojlanishi natijasida paydo bo'ldi. Ular turli xil, ko'pincha turli xil ma'lumotlarni foydali ma'lumotlarga aylantirish uchun mo'ljallangan. OLAP tizimlari ma'lumotlarni bir qator mezonlarga muvofiq tartibga solishi mumkin. Mezonlarning aniq xususiyatlarga ega bo'lishi shart emas.

OLAP tizimlari tashkilotning strategik boshqaruvining ko'plab masalalarida o'z ishini topdi: biznes samaradorligini boshqarish, strategik rejalashtirish, byudjetni tuzish, rivojlanishni bashorat qilish, moliyaviy hisobotlarni tayyorlash, ishlarni tahlil qilish, tashkilotning tashqi va ichki muhitini simulyatsiya qilish, ma'lumotlarni saqlash va hisobot berish.

OLAP tizimi ko'p o'lchovli ma'lumotlar massivlarini qayta ishlashga asoslangan. Ko'p o'lchovli massivlar shunday tuzilganki, massivning har bir elementi boshqa elementlar bilan ko'plab aloqalarga ega bo'ladi. Ko'p o'lchovli massivni yaratish uchun OLAP tizimi boshqa tizimlardan (masalan, ERP yoki CRM tizimlaridan) yoki tashqi kirish orqali ma'lumot olishlari kerak. OLAP tizimining

foydalanuvchisi o'z ma'lumotiga binoan kerakli ma'lumotlarni tuzilgan shaklda oladi. Belgilangan harakatlar tartibiga asoslanib, siz OLAP tizimining tuzilishini tasavvur qilishingiz mumkin.

Umuman olganda, OLAP tizimining tuzilishi quyidagilardan iborat quyidagi elementlar:

- malumotlar bazasi . Ma'lumotlar bazasi OLAP tizimining ishlashi uchun ma'lumot manbai hisoblanadi. Ma'lumotlar bazasi turi OLAP tizimi va OLAP server algoritmlari turiga bog'liq. Odatda relyatsion ma'lumotlar bazalari, ko'p o'lchovli ma'lumotlar bazalari, ma'lumotlar do'konlari va boshqalar ishlatiladi.

- OLAP-server. U ko'p o'lchovli ma'lumotlar tuzilishini va ma'lumotlar bazasi va OLAP tizimi foydalanuvchilari o'rtasidagi munosabatlarni boshqarishni ta'minlaydi.

- maxsus dasturlar... OLAP tizimining ushbu elementi foydalanuvchi so'rovlarini boshqaradi va ma'lumotlar bazasiga kirish natijalarini hosil qiladi (hisobotlar, grafikalar, jadvallar va boshqalar).

Ma'lumotlarni tartibga solish, qayta ishlash va saqlash uslubiga qarab, OLAP tizimlari amalga oshirilishi mumkin mahalliy kompyuterlar foydalanuvchilar yoki maxsus serverlardan foydalanish.

Ma'lumotlarni saqlash va qayta ishlashning uchta asosiy usuli mavjud:

- mahalliy. Ma'lumotlar foydalanuvchilarning kompyuterlariga joylashtirilgan. Ma'lumotlarni qayta ishlash, tahlil qilish va boshqarish mahalliy ish joylarida amalga oshiriladi. Ushbu OLAP tizim tuzilmasi ma'lumotlarni qayta ishlash tezligi, ma'lumotlar xavfsizligi va ko'p o'zgaruvchan tahlildan cheklangan foydalanish bilan bog'liq muhim kamchiliklarga ega.

- relyatsion ma'lumotlar bazalari... Ushbu ma'lumotlar bazalari CRM tizimi yoki ERP tizimi bilan OLAP tizimi bilan birgalikda ishlatiladi. Ma'lumotlar ushbu tizimlarning serverida relyatsion ma'lumotlar bazalari yoki ma'lumotlar omborlari shaklida saqlanadi. OLAP-server kerakli ko'p o'lchovli tuzilmalarni shakllantirish va tahlil qilish uchun ushbu ma'lumotlar bazalariga murojaat qiladi.

- ko'p o'lchovli ma'lumotlar bazalari... Bunday holda, ma'lumotlar ajratilgan serverda maxsus ma'lumotlar do'koni sifatida tashkil etiladi. Ma'lumotlarning barcha operatsiyalari ushbu serverda amalga oshiriladi, bu asl ma'lumotni ko'p o'lchovli tuzilmalarga o'zgartiradi. Bunday tuzilmalar deyiladi OLAP kubi... OLAP kubini shakllantirish uchun ma'lumotlar manbalari relyatsion ma'lumotlar bazalari va / yoki mijoz fayllari hisoblanadi. Ma'lumotlar serveri ma'lumotlarni oldindan tayyorlash va qayta ishlashni amalga oshiradi. OLAP-server OLAP kub bilan ma'lumotlar manbalariga (relyatsion ma'lumotlar bazalari, mijoz fayllari va hk) to'g'ridan-to'g'ri kirish huquqisiz ishlaydi.

Ma'lumotlarni saqlash va qayta ishlash uslubiga qarab barcha OLAP tizimlarini uchta asosiy turga bo'lish mumkin.

1. ROLAP (Relational OLAP - relyatsion OLAP tizimlari) - bunday OLAP tizimi relyatsion ma'lumotlar bazalari bilan ishlaydi. Ma'lumotlarga to'g'ridan-to'g'ri relyatsion ma'lumotlar bazasiga kirish mumkin. Ma'lumotlar relyatsion jadvallar sifatida saqlanadi. Foydalanuvchilar an'anaviy OLAP tizimlaridagi kabi ko'p o'zgaruvchan tahlillarni o'tkazish qobiliyatiga ega. Bunga SQL vositalari va vaqtincha so'rovlar yordamida erishiladi.

ROLAP-ning afzalliklaridan biri bu katta hajmdagi ma'lumotlarni yanada samarali ishlash qobiliyatidir. ROLAP-ning yana bir afzalligi - bu raqamli va matnli ma'lumotlarni samarali ishlash qobiliyatidir.

ROLAP-ning kamchiliklari past ko'rsatkichlarni (an'anaviy OLAP tizimlariga nisbatan) o'z ichiga oladi, chunki ma'lumotlar OLAP-server tomonidan qayta ishlanadi. Yana bir kamchilik - bu SQL ishlatilishi sababli cheklangan funksionallik.

2. MOLAP (Ko'p o'lchovli OLAP - ko'p o'lchovli OLAP tizimlari). Ushbu turdagi OLAP tizimlari an'anaviy tizimlarga tegishli. An'anaviy OLAP tizimining boshqa tizimlardan farqi ma'lumotlarni tayyorlash va optimallashtirishda. Ushbu tizimlar, qoida tariqasida, ma'lumotlarni oldindan qayta ishlashni amalga oshiradigan maxsus serverdan foydalanadilar. Ma'lumotlar ko'p o'lchovli massivlarda - OLAP kubiklarida shakllanadi.

MOLAP tizimlari ma'lumotlarni qayta ishlashda eng samarali hisoblanadi, chunki ular turli xil foydalanuvchi so'rovlari uchun ma'lumotlarni qayta tuzishni va tuzilishini osonlashtiradi. MOLAP analitik vositalari sizga murakkab hisob-kitoblarni amalga oshirishga imkon beradi. MOLAP-ning yana bir afzalligi - bu so'rovlarni tezda yaratish va natijalarga erishish qobiliyatidir. Bunga OLAP kublarini oldindan shakllantirish orqali erishiladi.

MOLAP tizimining kamchiliklariga qayta ishlangan ma'lumotlarning cheklanganligi va ma'lumotlarning ortiqcha bo'lishi kiradi ko'p o'lchovli kublarni hosil qilish uchun har xil jihatlar bo'yicha ma'lumotlar takrorlanishi kerak.

3. HOLAP (Hybrid OLAP - gibrid OLAP tizimlari). Gibrid OLAP tizimlari ROLAP va MOLAP tizimlarining birikmasidir. Gibrid tizimlarda biz ikkita tizimning afzalliklarini birlashtirishga harakat qildik: ko'p o'lchovli ma'lumotlar bazalaridan foydalanish va relyatsion ma'lumotlar bazalarini boshqarish. HOLAP tizimlari relyatsion jadvallarda katta hajmdagi ma'lumotlarni saqlashga imkon beradi va qayta ishlangan ma'lumotlar oldindan o'rnatilgan ko'p o'lchovli OLAP kubiklarida saqlanadi. Ushbu turdagi tizimning afzalliklari - bu ma'lumotlarning miqyosi, tezkor ishlov berish va ma'lumot manbalariga moslashuvchan kirish. OLAP tizimlarining boshqa turlari mavjud, ammo ular mustaqil OLAP tizimiga qaraganda ko'proq ishlab chiqaruvchilarning marketing hiyla-nayrangidir.

Ushbu turlarga quyidagilar kiradi:

- WOLAP (Veb OLAP). OLAP tizimining ko'magi bilan ko'rish veb-interfeys... Ushbu OLAP tizimlari veb-interfeys orqali ma'lumotlar bazalariga kirish imkoniyatiga ega.

- DOLAP (ish stoli OLAP). Ushbu turdagi OLAP tizimi foydalanuvchilarga mahalliy fayllarni yuklab olish imkoniyatini beradi ish joyi ma'lumotlar bazasi va u bilan mahalliy darajada ishlash.

- MobileOLAP. Bu ma'lumotlar bazasi bilan masofadan turib ishlashga imkon beruvchi OLAP tizimlarining vazifasidir mobil qurilmalar.

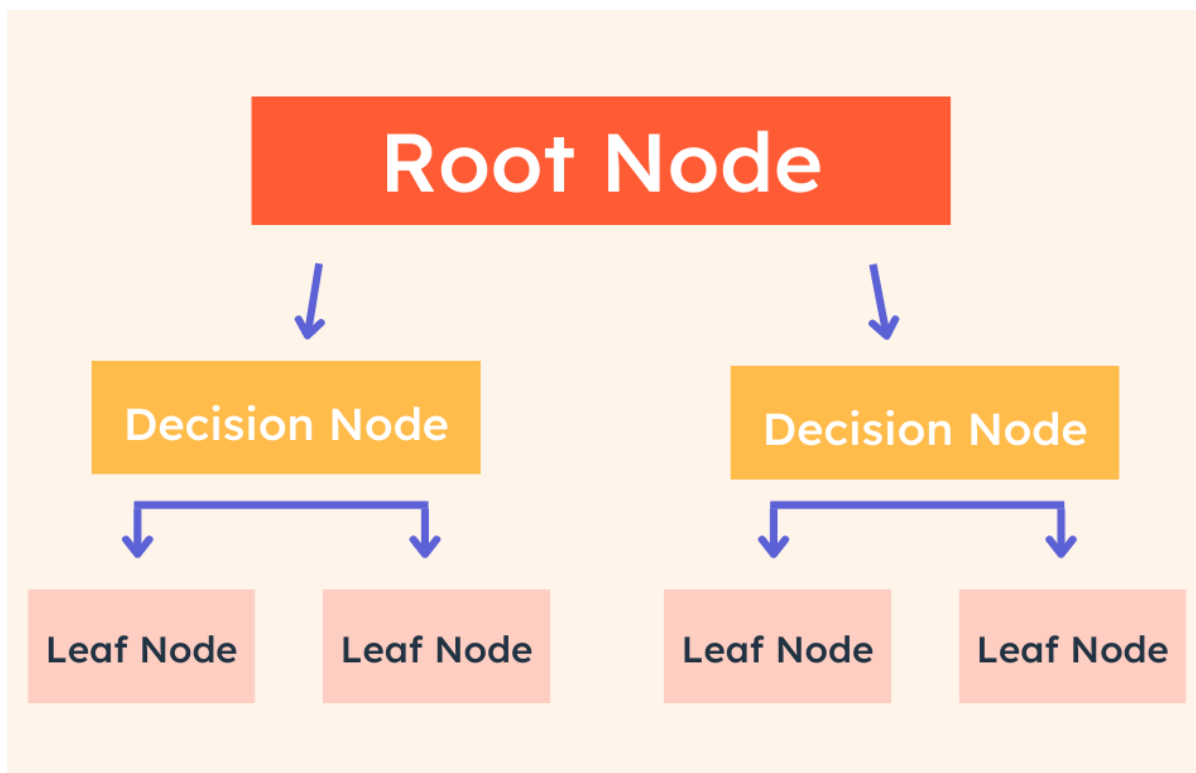
- SOLAP (fazoviy OLAP). Ushbu turdagi OLAP tizimlari fazoviy ma'lumotlarni qayta ishlashga mo'ljallangan. U geografik axborot tizimlari va OLAP tizimlarini birlashtirish natijasida paydo bo'ldi. Ushbu tizimlar ma'lumotlarni nafaqat alfasayısal formatda, balki vizual ob'ektlar va vektorlar shaklida ham qayta ishlashga imkon beradi.

3. Qaror daraxtini qurish

Qaror daraxti - bu bitta asosiy g'oya yoki savol bilan boshlanadigan va har bir qarorning potentsial natijalari bilan ajralib turadigan sxema. Qaror daraxtidan foydalanib, eng yaxshi harakat yo'nalishini aniqlashingiz mumkin.

Qarorlar daraxtining vizual elementi, agar siz bu haqda gapirgan bo'lsangiz, ko'proq potentsial harakatlar va natijalarni o'z ichiga oladi va kutilmagan oqibatlar xavfini kamaytiradi.

Bundan tashqari, diagramma sizga kichikroq tafsilotlarni kiritish va bosqichma-bosqich rejani yaratish imkonini beradi, shuning uchun siz o'z yo'lingizni tanlaganingizdan so'ng, u ergashish uchun allaqachon rejalashtirilgan.



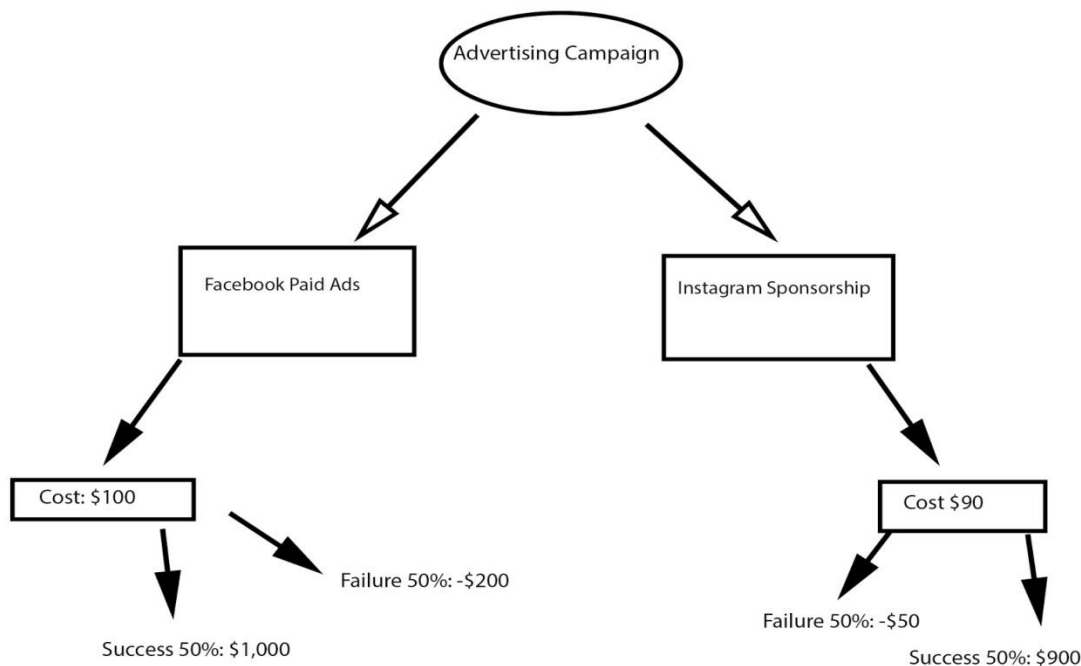
17-rasm.Qarorlar daraxtini tuzilishi (ildiz, qaror va barg).

Qaror daraxti to'rt elementni o'z ichiga oladi: ildiz tugunlari, qaror tugunlari, barg tugunlari va ularni bir-biriga bog'laydigan shoxlar.

- **Ildiz tuguni** daraxtning boshlanadigan joyidir. Bu siz ko'rib chiqayotgan katta muammo yoki qaror.
- Nomidan ko'rinib turibdiki, **qaror tugunlari** sizning daraxtingizdagi qarorni ifodalaydi. Ular sizning asosiy muammoingizni "hal qilish" mumkin bo'lgan yo'llardir.
- **Yetakchi tugunlar** qarorning mumkin bo'lgan natijalarini ifodalaydi. Misol uchun, agar siz tushlik uchun qayerda ovqatlanishni hal qilmoqchi bo'lsangiz, *McDonald'sda gamburger iste'mol qilish mumkin*. Tegishli barg tugunlari quyidagicha bo'lishi mumkin: *\$5 dan kam mablag' sarflab, pulni tejang*.
- **Filiallar** qaror daraxtidagi har bir elementni bog'laydigan o'qlardir. Har bir qarorning xavf-xatarlari va foydalarini tushunish uchun filiallarni kuzatib boring.

Qaror daraxti tahlili. Aytaylik, siz yangi kompaniyangizni qayerda reklama qilishni hal qilyapsiz:

Facebookda, pullik reklamalardan foydalanish yoki Instagramda, ta'sir qiluvchi homiyliklardan foydalanish. Oddiylik uchun ikkala variant ham sizning ideal demografiyangizga mos keladi va brendingiz uchun mantiqiy deb hisoblaymiz. Reklama kompaniyangiz uchun dastlabki qarorlar daraxti:



18-rasm. Qarorlar darxtiga misol.

Ko‘rib turganingizdek, siz o‘zingizning yakuniy maqsadingizni yuqoriga qo‘ymoqchisiz - bu holda, *Reklama kompaniyasi* siz qabul qilishingiz kerak bo‘lgan qarordir. Keyinchalik, har bir potentsial harakatga (barglaringiz) o‘qlarni (novdalarangizni) chizishingiz kerak bo‘ladi. Bizning misolimiz uchun, sizda faqat ikkita dastlabki harakat bor: Facebook pullik reklamalari yoki Instagram homiyliqi. Biroq, sizning daraxtingiz maqsadga qarab bir nechta muqobil variantlarni o‘z ichiga olishi mumkin. Endi siz xarajatlarni solishtirish uchun novdalar va barglarni chizishni xohlaysiz. Agar bu oxirgi qadam bo‘lsa, qaror aniq bo‘lar edi: Instagram 10 dollarga arzonroq, shuning uchun siz buni tanlagan bo‘lar edingiz. Biroq, bu oxirgi qadam emas. Muvaffaqiyat va muvaffaqiyatsizlik imkoniyatlarini aniqlab olishingiz kerak. Maqsadingizning murakkabligiga qarab, siz sanoatdagi mavjud ma'lumotlarni yoki kompaniyangizdagi oldingi loyihalarni, jamoangizning imkoniyatlarini, byudjetini, vaqt talablarini va bashorat qilingan natijalarni tekshirishingiz mumkin. Muvaffaqiyatga ta'sir qilishi mumkin bo‘lgan tashqi vaziyatlarni ham ko‘rib chiqishingiz mumkin.

Qarorlar daraxtini qanday yaratish kerak. Quyidagi amallarni bajarib, qarorlar daraxtini yaratishingiz mumkin. Esingizda bo'lsin: daraxtingizni tugatganingizdan so'ng, eng yaxshi harakat yo'nalishini topish uchun har bir qarorni tahlil qilishni boshlashingiz mumkin.

1. Asosiy fikringizni yoki savolingizni aniqlang. Birinchi qadam ildiz tuguningizni aniqlashdir. Bu siz o'rganmoqchi bo'lgan asosiy masala, savol yoki g'oya. Oqim diagrammangizning yuqori qismiga ildiz tuguningizni yozing.

2. Potentsial qarorlar va natijalarni qo'shing. Keyinchalik, potentsial qarorlarni qo'shish orqali daraxtingizni kengaytiring. Ushbu qarorlarni novdalar bilan ildiz tuguniga ulang. Bu erdan har bir qarorning aniq va potentsial natijalarini yozing.

3. Yakuniy nuqtalarga yetguncha kengaytiring. Daraxtingizdagi har bir qarorni e'tiborga olishni unutmang. Har bir qaror oxir-oqibat yakuniy nuqtaga etib borishi kerak, bu esa barcha natijalar yuzaga chiqishini ta'minlaydi. Boshqacha aytganda, kutilmagan hodisalar uchun joy yo'q.

4. Risk va mukofotni hisoblang. Endi raqamlarni sindirish vaqti keldi.

Eng samarali qaror daraxtlari miqdoriy ma'lumotlarni o'z ichiga oladi. Bu har bir qarorning kutilgan qiymatini hisoblash imkonini beradi. Eng keng tarqalgan ma'lumotlar puldir.

5. Natijalarni baholang. Oxirgi bosqich - natijalarni baholash. Ushbu bosqichda siz o'zingiz qabul qilishga tayyor bo'lgan tavakkalchilik miqdori asosida qaysi qaror eng ideal ekanligini aniqlaysiz. Esingizda bo'lsin, eng qimmatli qaror eng yaxshi harakat yo'nalishi bo'lmasligi mumkin. Nega? Bu yuqori mukofot bilan kelgan bo'lsa-da, u yuqori darajadagi xavfni ham keltirishi mumkin.

Byudjetingiz, vaqt jadvalingiz va boshqa omillar asosida eng yaxshi natijani aniqlash sizga va sizning jamoangizga bog'liq.

Data Miningda ma'lumotlar tahlilining vizualizatsiyasi.

Data Miningda har qanday ma'lumotlarni tahlil qilish usullari bilan olingan ma'lumotlarni vizual formatda taqdim etish uni tushunishni bir necha bor yaxshilaydi. Agar tahlilning ayrim usullarini amalga oshirish texnik jihatdan murakkab jarayon ekanligi haqida gapiradigan bo'lsak, unda uning natijalari deyarli har qanday axborot foydalanuvchisi uchun tushunarli bo'lishi mumkin va kerak. Bunda faqat vizualizatsiya samarali bo'lishi mumkin.

Ma'lumotlarni vizualizatsiya qilish va texnik bo'lmagan mutaxassislar bilan ishlash uchun maxsus yaratilgan professional analitik dasturlar Data Miningda vizualizatsiya yaratishga yordam beradi.

Analitik dasturlardan biri bu Tableau. Mahsulot biznes intellektini vizual, chiroyli va interaktiv qilishning oddiy va tezkor usulini taklif etadi. Dasturdagi

bo'limlarning kerakli hisobotlari bir necha soniyalarda yaratiladi va IT bilimi bo'lmagan mutaxassis Tableauni o'zlashtira oladi, chunki dastur interfeysi juda sodda va intuitivdir.

Topshiriqlar:

1-topshiriq. Fermer makkajo'xori yoki soya yetishtirishi mumkin. Ushbu ekinlar uchun kelajakdagi hosil narxining oshishi, o'zgarmasligi yoki pasayishi ehtimoli mos ravishda 0,25, 0,30 va 0,45 ni tashkil qiladi. Narxlar ko'tarilsa, makkajo'xori hosili 30 000 dollar, soya hosili esa 10 000 dollarni tashkil qiladi, agar narxlar o'zgarmasa, fermer faqat xarajatlarni qoplaydi. Ammo narxlar pasaysa, makkajo'xori va soya ekinlari mos ravishda 35 000 va 5 000 dollar yo'qotishlarga olib keladi.

- 1.Ushu masalani yechish qaror daraxtini yarating.
- 2.Femer qanday hosilni etishtirishi kerak?
- 3.Uning foydasining kutilayotgan qiymati qanday?

2-topshiriq. Siz yangi konsalting xizmatini yaratish istiqbollari ko'rib chiqyapsiz. Dastlabki bosqichda talab qilinadigan investitsiyalar hajmi 200 ming dollarni tashkil etadi, 1-yilda talab yuqori bo'lishi ehtimoli 60% ni tashkil qiladi.

Agar birinchi yilda talab yuqori bo'lsa, keyingi yillarda yuqori va past talabning ehtimoli mos ravishda 80% va 20% ni tashkil qiladi. Agar 1-yilda talab past bo'lsa, keyingi yillarda yuqori va past talabning ehtimoli mos ravishda 40% va 60% ni tashkil qiladi.

Yuqori talab bilan prognoz qilingan daromad 500 ming dollarni tashkil qiladi. yiliga; past talab bilan, bashorat qilingan daromad 300 ming dollar. yiliga. Siz istalgan vaqtda xizmatlar ko'rsatishni to'xtatishingiz mumkin. Kompyuterdan foydalanish bilan bog'liq bo'lmagan xarajatlar 140 000 dollarni tashkil qiladi. talab darajasidan qat'iy nazar yiliga.

Agar siz konsalting xizmatiga pul qo'ymaslikka qaror qilsangiz, uni yiliga 20% ga deyarli xavf-xatarsiz investitsiya qilishingiz mumkin. Agar konsalting xizmatini tashkil etishga qaror qilinsa, siz faoliyatning asosini tashkil etuvchi kompyuter hisob-kitoblarini amalga oshirish masalasini hal qilishingiz kerak bo'ladi.

Mumkin variantlardan biri server sotib olishdir. Uning eskirish muddati 5 yil. Xarajatlar 150 ming dollarlik boshlang'ich xarajat va 20 ming dollarlik yillik operatsion xarajatlardan iborat bo'ladi, kerak bo'lganda kompyuter resurslarini ijaraga olish mumkin. Bu holda, ijara xarajatlari talabga mutanosib bo'ladi va 140 ming kelishilgan sobit xarajatlar minus daromad 30% tashkil etadi, barcha hollarda, boshqa hech qanday xarajatlar bor.

- a. Ushbu variantlarni ko'rsatadigan va 3 yilga mo'ljallangan "qarorlar daraxti" ni yarating.
- b. Konsalting xizmatini tashkil etishga arziydimi yoki xavf-xatarsiz daromad foydaliroqmi? Ikki va uch yillik faoliyat natijalarini ko'rib chiqing.
- c. Qaysi biri yaxshiroq - kompyuter sotib olish yoki ijaraga olish?
- d. Faraz qilaylik, 3 yillik faoliyatdan so'ng siz xizmatni alohida biznes sifatida o'rtacha 350 ming dollarga sotishingiz mumkin. Yillik necha foiz o'sish siz olgan daromadga to'g'ri keladi?
- e. Siz qilishingiz kerak bo'lgan qo'shimcha taxminlarni aniq ko'rsating.

AMALIY MASHG'ULOT MATERIALLARI

1-amaliy mashg'ulot. Qarorni qo'llab quvvatlash tizimlari. OLAP-tizimlari.

1.1.Ma'lumot tushunchasi. Big Data da axborot va bilim tushunchasi.

Atrofimizdagi moddiy dunyo materiya, energiya va axborot kabi asosiy tarkibiy qismlardan iborat. Materiya, energiya tushunchalari fizika, geografiya kabi fanlarda ilmiy jihatdan o'rganilgan. Axborot tushunchasi informatika fanidan ma'lum. Axborotga informatika nuqtai nazaridan ta'rif berilgan, unga ko'ra axborot nomoddiy tushuncha sanaladi. Ya'ni moddiy narsalar kabi unga ta'sir qilish mumkin emas. Biroq bugungi kunda axborotni boshqarish eng dolzarb masala sanaladi. Uni chuqur o'rganish, tushunish va tasniflash talab qilinadi. Demak, axborot o'zi nima?

Axborot - bu shaxs yoki maxsus qurilmalar tomonidan aloqa jarayonida moddiy yoki ma'naviy olam faktlarining aksi sifatida qabul qilinadigan ma'lumotlar to'plamidir.

Shuningdek, faylasuflar, matematiklar, kompyuter olimlari va boshqarmutaxassislar tomonidan qo'llaniladigan boshqa ta'riflarni ham berish mumkin:

- Axborot real olamning in'ikosidir: bu bir real ob'ekt boshqa real ob'ekt haqida o'z ichiga olgan ma'lumotdir (materialistik falsafa);

- Axborot - bu ularni taqdim etish shaklidan qat'i nazar berilgan ma'lumotdir. (<https://ru.wikipedia.org/wiki/Information>);

- Axborot - ma'lum bir kontekstda o'ziga xos kontekstga ega bo'lgan faktlar, hodisalar, narsalar, g'oyalar va tushunchalar haqidagi bilimlardir (ISO/IEC 2382:2015).

Ko'rinib turibdiki, bu ta'riflardan axborot atamasining ma'nosi aniq berilgani shu deb aytib bo'lmaydi – axborot atamasi uning sinonimlari – “ma'lumot”, “bilim” va hokazolar bilan almashtiriladi, ular aslida ma'lumotdir. Gnostik ilon yana o'z dumini ushlab oldi.

O'z-o'zidan ma'lumotni matematik kabi mavhum tushunchalar sifatida tasniflash mumkin, ammo bir qator xususiyatlar uni moddiy ob'ektlarga yaqinlashtiradi. Demak, ma'lumotni olish, yozib olish, o'chirish, uzatish mumkin; ma'lumot yo'qdan kelib chiqishi mumkin emas.

Biroq, axborotni tarqatish jarayonida uning mulki namoyon bo'ladi, bu moddiy ob'ektlarga xos emas - saqlanish qonunining yo'qligi.

Axborotni bir tizimdan ikkinchisiga uzatishda, odatda qabul qiluvchi tizimda ko'paysa ham, uzatish tizimidagi axborot miqdori kamaymaydi.

Agar ma'lumot bunday xususiyatga ega bo'lmaganida, masalan, o'qituvchi talabalarga ma'ruza o'qiyotganda, ma'lumotni yo'qotib, johil bo'lib qolar edi.

Bernard Shou dedi: "Agar siz va menda bittadan g'oya bo'lsa, har birimizda bittadan g'oya bo'ladi. Agar biz ularni almashsak, ikkalamizda ikkitadan g'oya bo'ladi bo'ladi".

Axborotning bilvosita ta'rifiga qiziqarli yondashuv "Tartib - tartibsizlik" tushunchalarining dialektik juftligi dinamikasini tahlil qilishdan shakllanishi mumkin.

Shu nuqtai nazardan, ma'lumot dastlab xaotik muhitda tartib darajasi, ya'ni. entropiyaning o'zaro nisbati. Boshqacha qilib aytganda, bu dunyoni termal xaos holatidan olib chiqadigan, unga heterojen naqshlar va uyushgan tizimlarni kiritadigan ma'lumotdir.

Birlamchi toifalar ta'riflarining murakkabligi bilan bog'liq muammolardan mavzu sohasining muhandislik xususiyatlariga o'tib, biz ma'lumotlar va bilimlarning umumiy qabul qilingan ta'riflarini kiritamiz.

Ma'lumotlar rasmiylashtirilgan shaklda taqdim etilgan faktlar, tushunchalar yoki buyruqlar bo'lishi mumkin, bu ularni uzatish,

Megabayt:	1 MB = 10^6 bayt
Gigabayt:	1 GB = 10^9 bayt
Terabayt:	1 TB = 10^{12} bayt
Petabayt:	1 PB = 10^{15} bayt
Eksobayt:	1 EB = 10^{18} bayt
Zettabayt:	1 ZB = 10^{21} bayt

izohlash yoki qayta ishlashga imkon beradi [Ma'lumotlarni qayta ishlash. Lug'at. Asosiy shartlar. – 1992].

Ma'lumotlar raqamli massivlar, faktlar, matnlar, grafikalar, rasmlar, tovushlar, analog yoki raqamli video segmentlar bo'lishi mumkin.

Ular kuzatishlar, o'lchovlar, tajribalar, arifmetik va mantiqiy amallar natijasida olish mumkin.

Маълумот - бу бирор бир объект, жараён, ходиса ёки воқеликни ифодалаб (таснифлаб) берувчи белги ёки белгилар мажмуасидир. Берилган маълумот (белги)лар қандай қиймат қабул қилишига қараб маълумотларнинг бир қанча турлари мавжуд.

Dunyoda qancha raqamli ma'lumotlar to'plangan? 2015-16 yillarda insoniyat tomonidan yaratilgan raqamli axborot miqdori 4 zettabayt ma'lumotni tashkil etdi. 2020 yilgacha bu hajm 40 zettabaytgacha oshishi kutilmoqda.

1000 sahifali bitta elektron kitob taxminan 3 megabaytni tashkil qiladi, shuning uchun bu 2,73 TB ma'lumotni o'z ichiga olgan kutubxona taxminan bir kvadrillion jildni o'z ichiga olishi kerak.

Shu bilan birga, hozir dunyoda 130 millionga yaqin kitoblar mavjudligi ma'lum. Ularning axborot imkoniyatlarini baholash juda qiyin. Taxminan 10 million jild bor, ular haqiqatda foydali ma'lumotlarni o'z ichiga oladi. Elektron shaklda yozishda ular 30 Ptb oladi, ya'ni. barcha to'plangan ma'lumotlardan million marta

kamroq. Bu birlamchi ma'lumot va uni konsentratsiyalangan shaklga - bilimga tarjima qilish natijasi o'rtasidagi farqni ko'rsatadi.

Juda katta ma'lumotlar bazasi (VLDB) - bu jismoniy xotira qurilmasida juda katta hajmdagi joyni egallaydigan ma'lumotlar bazasi. Bu atama ma'lumotlar bazasining maksimal mumkin bo'lgan hajmlarini anglatadi, ular jismoniy ma'lumotlarni saqlash texnologiyalari va dasturiy ta'minot bilan ishlash texnologiyalaridagi so'nggi yutuqlar bilan belgilanadi.

“O'ta katta hajm” tushunchasining miqdoriy ta'rifi vaqt o'tishi bilan o'zgarib turadi. Shunday qilib, 1997 yilda dunyodagi eng katta matnli ma'lumotlar bazasi 7 terabayt hajmli Knight Ridderning DIALOG edi. 2001 yilda eng katta ma'lumotlar bazasi 10,5 terabayt, 2003 yilda - 25 terabayt deb hisoblangan. 2005 yilda dunyodagi eng katta ma'lumotlar bazalari yuzlab terabayt hajmli saqlash joyi deb hisoblangan. 2006 yilda Google qidiruv tizimi 850 terabaytlik ma'lumotlar bazasidan foydalangan

2010 yilga kelib, ultra-katta ma'lumotlar bazasi hajmi kamida petabaytlarda o'lchanishi kerak deb hisoblangan.

2014 yilga kelib, bilvosita hisob-kitoblarga ko'ra, Google o'z serverlarida jami 10-15 ekzabaytgacha ma'lumotlarni saqlagan.

Ba'zi hisob-kitoblarga ko'ra, 2025 yilga borib, genetiklar 100 milliondan 2 milliardgacha odamning genomlari to'g'risida ma'lumotlarga ega bo'lishadi va bu hajmdagi ma'lumotlarni saqlash uchun 2 dan 40 ekzabaytgacha kerak bo'ladi.

Shunday qilib, tahlilchilar qo'lida ishlab chiqarish jarayonlarini monitoring qilish natijasida olingan, ommaviy tibbiy ko'riklar jarayonida to'plangan yoki sotsiologik so'rovlar jarayonida to'plangan juda katta hajmdagi ma'lumotlar mavjud. Buning nima keragi bor? Ushbu ma'lumotlar o'z-o'zidan foydasiz va undan foydali bilimlarni olish uchun qayta ishlanishi kerak.

Bilim - bu ma'lum bir fan sohasidagi mutaxassislarning tajribasini rasmiylashtiradigan va dastlabki ma'lumotlarda aniq ko'rsatilmagan javoblar (yechimlar) berishga imkon beradigan faktlar va qoidalar.

Bilimning o'zi kerak emas, lekin boshqaruv qarorlarini shakllantirish uchun uni amalga oshirish mumkin.

O'z navbatida, qabul qilinayotgan qarorlarning sifati nazorat tizimlarining samaradorligi va ularni muayyan fan sohalarida amalga oshirish orqali baholanadi.

Masalan, sanoatda boshqaruv qarorlarining yakuniy sifati, qoida tariqasida, korxonaning iqtisodiy samaradorligi nuqtai nazaridan baholanadi.

Axborot tahlilining uch darajasi

Semiotika (imo-ishora tizimlari haqidagi fan) g'oyalari nuqtai nazaridan ma'lumotlarning adekvatligi

ma'lumot, uning mazmunining ko'rsatilgan ob'ekt tasviriga mos kelishi uchta

shaklda ifodalanishi mumkin:

- ☐ sintaktik;
- ☐ semantik;
- ☐ pragmatik.

Sintaktik adekvatlik aks ettirishning rasmiy tuzilmaviy xususiyatlarini takrorlash, semantik va iste'mol parametrlaridan abstraktsiyalash bilan bog'liq. Sintaktik darajada quyidagilar hisobga olinadi: media turi, tasvirlash usuli, uzatish va qayta ishlash tezligi, kod formati, konversiyaning ishonchliligi va aniqligi va boshqalar. Shu bilan birga, axborot o'z tashuvchisining energiya va fazo-vaqt xususiyatlariga nisbatan o'zgarmasdir. Xuddi shu ma'lumot turli xil kodlarda mavjud bo'lishi mumkin.

Semantik shakl tushuncha va g'oyalarning shakllanishini, ma'no, ma'lumot mazmunini aniqlashni ta'minlaydi. Xabardagi semantik ma'lumotlarning miqdori nisbiy qiymatdir: bir xil xabar vakolatli foydalanuvchi uchun semantik tarkibga ega bo'lishi va qobiliyatsiz foydalanuvchi uchun ma'nosiz (semantik shovqin) bo'lishi mumkin.

Axborotni ko'rib chiqishning pragmatik jihati uning qiymati, foydaliligi va tizim maqsadlariga erishish uchun amaliy foydalanish bilan bog'liq.

1.2. Ma'lumot turlari.

Demak, ma'lumotlar - kompyuter tomonidan boshqariladigan magnit, optik yoki mexanik vositalarda qayd etilgan elektr signallari (ipmuls) shaklida saqlanishi va uzatilishi mumkin bo'lgan miqdorlar, qiymatlar, belgilar yoki simvollar to'plamidir.

Big data - katta ma'lumotlar atamasi vaqt o'tishi bilan kattalashib boruvchi va eksponent ravishda o'sib boradigan ma'lumotlar to'plamini tavsiflash uchun ishlatiladi. Bunday katta hajmdagi ma'lumotlarni qayta ishlashni amalga oshirish mashinali o'qitishni o'rganmasdan juda ham murakkabdir. Big data quyidagi afzalliklarni taqdim etadi:

- Turli manbalardan ma'lumotlarni to'plash;
- Real vaqt rejimida analitik taxlil orqali ish jarayonlarini takomillashtirish;
- Katta hajmdagi ma'lumotlarni saqlash;
- Insayt, ya'ni uzoq muddatli ongsiz aqliy faoliyat natijasi bo'lgan muammoning yechimini to'satdan ongli ravishda aniqlash usuli bo'lib, katta ma'lumotlarda tarkiblangan va yarim tarkiblangan ma'lumotlar yordamida yashirin ma'lumotlarni aniqlash;

- Katta ma'lumotlar tegishli xavf tahlillari tufayli xavfni kamaytirishga va aqlli qarorlar qabul qilishga yordam beradi.

Data Science va katta ma'lumotlar sohasida har xil tipdagi ma'lumotlar mavjud bo'lib, ularning har biri o'z vositalari va usullarini talab qiladi. Ma'lumotlarning asosiy toifalari quyida keltirilgan:

- Tarkiblangan ma'lumotlar.
- Tarkiblanmagan ma'lumotlar.
- Tabiiy tildagi ma'lumotlar.
- Mashinali ma'lumotlar.
- Grafik ma'lumotlar.
- Audio, video va grafik ma'lumotlar.
- Oqimli ma'lumotlar.

Tarkiblangan ma'lumotlar

Tarkiblangan ma'lumotlar - modeliga bog'liq va yozuv ichidagi belgilangan maydonda saqlanadigan ma'lumotlardir. Shunga ko'ra, tarkiblangan ma'lumotlar ko'pincha jadvallarda, ma'lumotlar bazalarida yoki Excel fayllarida saqlash uchun qulaydir. SQL (Structured Query Language - tarkiblangan so'rovlar tili) ma'lumotlar bazalarida saqlangan ma'lumotlarga so'rovlarni yaratish va boshqarish uchun asosiy vositadir. Bundan tashqari, ba'zida an'anaviy relyatsion ma'lumotlar bazasida saqlash qiyin bo'lgan tarkiblangan ma'lumotlar mavjud (masalan, ierarxik ma'lumotlar, masalan, genealogik daraxt).

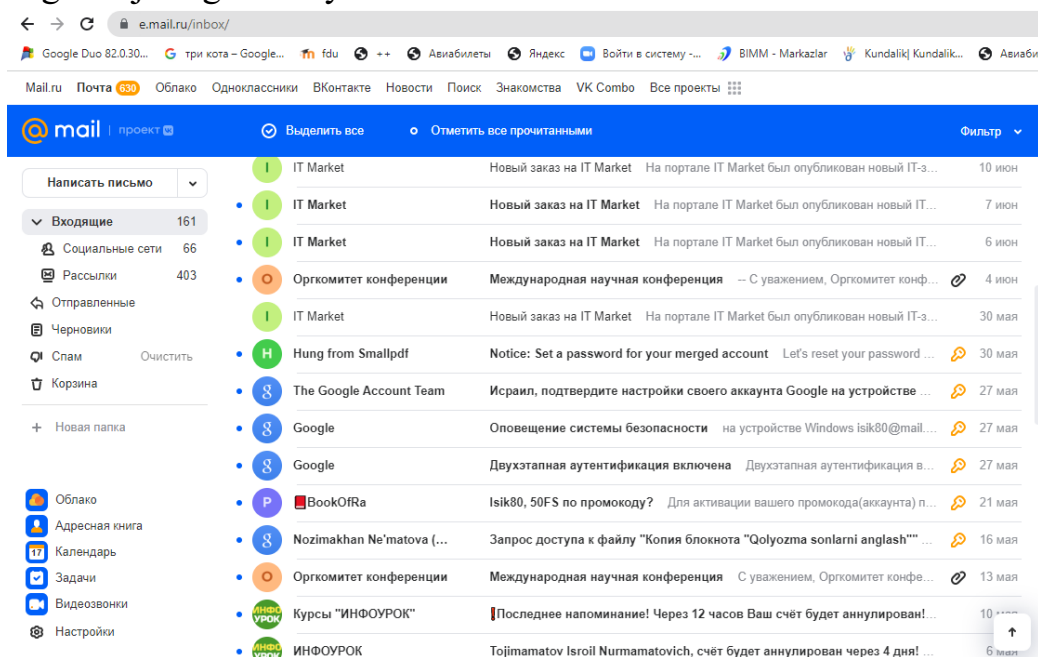
Biroq, dunyo tarkiblangan ma'lumotlardan iborat emas, shunchaki bu taqdim etishda inson va mashinalar uchun qulaydir. Ko'pincha haqiqiy ma'lumotlar tarkiblanmagan shaklda saqlanadi.

N	ФИО	Лавозим	Оклад	Иш кун	Хисоб и	Д-солип	ИНПС	К-у	Аванс	Жами ушланганла	Кулга тегиш	Жами	Хайри
1	Умаров А	Директор	195600	25	4890000	586800	97800	48900	1956000	2689500	2200500	4156500	97800
2	Каримова С	Бугалтер	185610	22	4083420	490010,4	81668,4	40834,2	1633368	2245881	1837539	3470907	81668,4
3	Анваров К	Шафёр	175620	21	3688020	442562,4	73760,4	36880,2	1475208	2028411	1659609	3134817	73760,4
4	Умаров А	Мутахассис	165630	21	3478230	417387,6	69564,6	34782,3	1391292	1913026,5	1565203,5	2956496	69564,6
5	Каримова С	Мутахассис	155640	21	3268440	392212,8	65368,8	32684,4	1307376	1797642	1470798	2778174	65368,8
6	Анваров К	Мутахассис	145650	21	3058650	367038	61173	30586,5	1223460	1682257,5	1376392,5	2599853	61173
7	Умаров А	Мутахассис	135660	21	2848860	341863,2	56977,2	28488,6	1139544	1566873	1281987	2421531	0
8	Каримова С	Мутахассис	125670	21	2639070	316688,4	52781,4	26390,7	1055628	1451488,5	1187581,5	2243210	0
9	Анваров К	Мутахассис	115680	21	2429280	291513,6	48585,6	24292,8	971712	1336104	1093176	2064888	0
10	Умаров А	Мутахассис	105690	21	2219490	266338,8	44389,8	22194,9	887796	1220719,5	998770,5	1886567	0
11	Каримова С	Мутахассис	95700	21	2009700	241164	40194	20097	803880	1105335	904365	1708245	0
12	Анваров К	Мутахассис	85710	21	1799910	215989,2	35998,2	17999,1	719964	989950,5	809959,5	1529924	0
13	Умаров А	Мутахассис	159600	21	3351600	402192	67032	33516	1340640	1843380	1508220	2848860	67032
14	Каримова С	Мутахассис	65730	21	1380330	165639,6	27606,6	13803,3	552132	759181,5	621148,5	1173281	0
15	Анваров К	Мутахассис	55740	21	1170540	140464,8	23410,8	11705,4	468216	643797	526743	994959	0
16	Умаров А	Мутахассис	45750	21	960750	115290	19215	9607,5	384300	528412,5	432337,5	816637,5	0
17	Каримова С	Мутахассис	260000	21	5460000	655200	109200	54600	2184000	3003000	2457000	4641000	109200
18	Анваров К	Фаррош	25770	30	773100	92772	15462	7731	309240	425205	347895	657135	0
19	Умаров А	Каровул	15780	60	946800	113616	18936	9468	378720	520740	426060	804780	0
20	Каримова С	Мутахассис	5790	60	347400	41688	6948	3474	138960	191070	156330	295290	0
21	Жами												

1-rasm. Tarkiblangan ma'lumotlarga doir misol

Tarkiblanmagan ma'lumotlar

Tarkiblanmagan ma'lumotlarni ma'lum bir ma'lumotlar modeliga moslashtirish qiyin, chunki uning mazmuni kontekstga bog'liq yoki o'zgaruvchidir. Tarkibi bo'lmagan ma'lumotlarning bir misoli oddiy elektron pochta xabarlaridir (2-rasm). Xabarda tuzilgan elementlar (jo'natuvchi, sarlavha, korpus) bo'lsada, bir xil vazifalarni turli yo'llar bilan bajarish mumkin, masalan, xabarlarda ma'lum bir shaxsni eslatishning son-sanoqsiz usullari mavjud. Muammo minglab tillar va dialektlarning mavjudligi bilan yanada murakkablashadi.



2-rasm. Pochta xabarlari

Inson tomonidan yozilgan elektron pochta xabari (2-rasm) ko'rsatilgandek ham tabiiy til ma'lumotlarining ideal namunasidir.

Tabiiy tildagi ma'lumotlar

Tabiiy til ma'lumotlari - tuzilmagan ma'lumotlarning maxsus turi bo'lib, bunday ma'lumotlarni qayta ishlash ancha murakkab, chunki u ham tilshunoslik, ham Datasceince ining maxsus usullarini bilishni talab qiladi.

Tabiiy tildagi ma'lumotlarni qayta ishlash hamjamiyati ob'ektni tanib olish, mavzu maydonini aniqlash, umumshtirish, matnni to'ldirish va hissiyotlarni tahlil qilishda muvaffaqiyatlarga erishdi, biroq bir domenga moslashtirilgan modellar boshqa domenlar uchun yaxshi umumshtirilmaydi. Hatto eng zamonaviy usullar ham o'zboshimchalik bilan matnning ma'nosini hal qila olmaydi. Va bu haqiqat hech kimni ajablantirmaydi; odamlar ham tabiiy tilni tushunishda muammolarga duch kelishadi. Bu tabiatan noaniq. Ma'no tushunchasining o'zi bahsli. Ikki kishi bitta suhbatni tinglayapti; Ular undan bir xil ma'noni oladilarmi? Bundan tashqari, alohida so'zlarning ma'nosi so'zlovchining kayfiyatiga qarab o'zgarishi mumkin. (3-rasm)



3-rasm. Ijtimoiy tarmoq yoki pochtdan kelgan habarlar.

Mashina ma'lumotlar.

Mashina ma'lumotlari inson aralashuvisiz kompyuter, jarayon, dastur yoki qurilma tomonidan avtomatik ravishda yaratilgan ma'lumotlarni anglatadi. Mashina ma'lumotlari asosiy ma'lumot manbalaridan biriga aylanib bormoqda va vaziyat o'zgarishi dargumon. Wikibon prognozlariga ko'ra, sanoat Internetining bozor qiymati (bu atama Frost & Sullivan tomonidan tarmoq sensorlari va dasturiy ta'minotga ega murakkab jismoniy qurilmalarga ishora qilish uchun yaratilgan) 2020 yilga kelib taxminan 540 milliard dollarni tashkil qiladi. IDC (International Data Corporation) hisob-kitoblariga ko'ra, 2020 yilga borib tarmoq tugunlari soni aholi sonidan 26 baravar oshadi. Ushbu tarmoq ko'pincha narsalar Interneti deb ataladi.


```

CSIPERF:TXCOMMIT;313236
2014-11-28 11:36:13, Info
69), objectname [6]"(null)"
2014-11-28 11:36:13, Info
result 0x00000000, handle @0x4e54
2014-11-28 11:36:13, Info
Beginning NT transaction commit...
2014-11-28 11:36:13, Info
trace:
CSIPERF:TXCOMMIT;273993
2014-11-28 11:36:13, Info
70), objectname [6]"(null)"
2014-11-28 11:36:13, Info
result 0x00000000, handle @0x4e5c
2014-11-28 11:36:13, Info
Beginning NT transaction commit...
2014-11-28 11:36:14, Info
trace:
CSIPERF:TXCOMMIT;396259
2014-11-28 11:36:14, Info
71), objectname [6]"(null)"
2014-11-28 11:36:14, Info
result 0x00000000, handle @0x4e5c
2014-11-28 11:36:14, Info
Beginning NT transaction commit...
2014-11-28 11:36:14, Info
trace:
CSIPERF:TXCOMMIT;375581

```

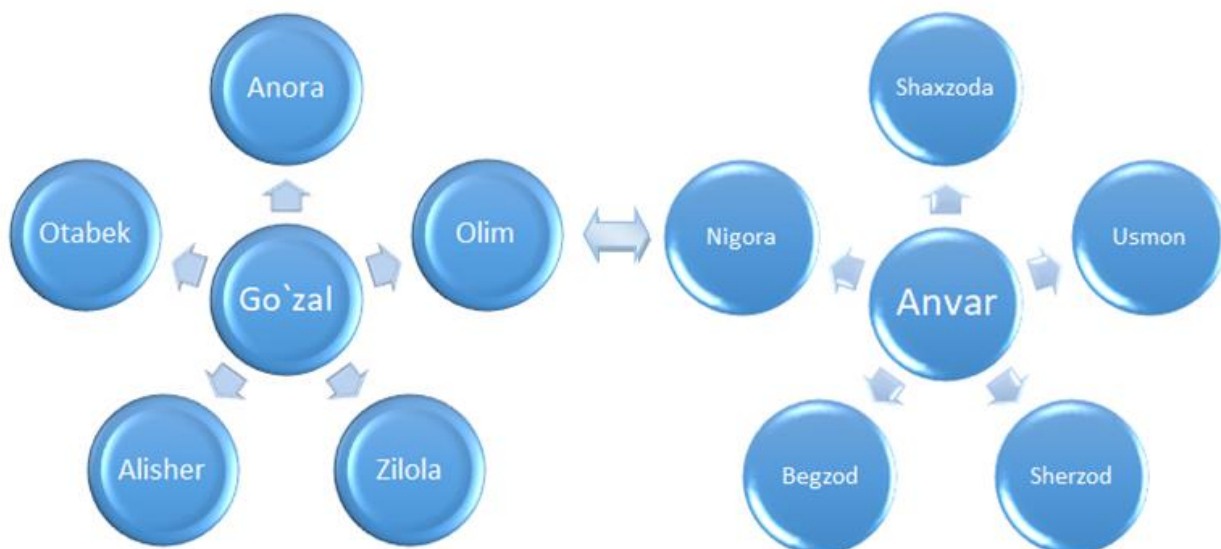
```

CSI 00000153 Creating NT transaction (seq
CSI 00000154 Created NT transaction (seq 69)
CSI 00000155@2014/11/28:10:36:13.471
CSI 00000156@2014/11/28:10:36:13.705 CSI perf
CSI 00000157 Creating NT transaction (seq
CSI 00000158 Created NT transaction (seq 70)
CSI 00000159@2014/11/28:10:36:13.764
CSI 0000015a@2014/11/28:10:36:14.094 CSI perf
CSI 0000015b Creating NT transaction (seq
CSI 0000015c Created NT transaction (seq 71)
CSI 0000015d@2014/11/28:10:36:14.106
CSI 0000015e@2014/11/28:10:36:14.429 CSI perf

```

4-rasm. Mashina ma'lumotlari.

Grafik ma'lumotlari - atamasi qaysidir ma'noda chalkash tushuncha bo'lishi mumkin, chunki har qanday ma'lumotni grafik ko'rinishda taqdim etilishi mumkin. Biroq Big Datada "grafik" deganda biz grafiklarning matematik nazariyasidan olingan grafik tushunchasini – ob'ektlar orasidagi juft munosabatlarni modellashtirish uchun matematik tuzilmani tushunamiz kerak bo'ladi. Qisqacha qilib aytganda, grafik yoki tarmoq ma'lumotlarida asosiy e'tibor ob'ektlarning o'zaro bog'liqligi yoki qo'shniligiga qaratiladi. Grafik ma'lumotlar tuzilmalari grafik ma'lumotlarni ko'rsatish va saqlash uchun tugunlar, qirralar va xususiyatlardan foydalanadi. Grafik ma'lumotlar ijtimoiy tarmoqlarni ifodalash uchun tabiiy mos keladi va uning tuzilishi ishtirokchilarning ta'siri va ikki kishi o'rtasidagi eng qisqa yo'l kabi aniq ko'rsatkichlarni hisoblash imkonini beradi.



5-rasm. Grafik ma'lumotlar. Ijtimoiy tarmoqda do'stlar o'rtasidai aloqalar.

Grafik ma'lumotlarga misollar ko'plab ijtimoiy tarmoq veb-saytlarida topilgan. Masalan, LinkedIn-da siz kompaniyada kimlarni bilishingizni ko'rishingiz mumkin. Sizning Twitter kuzatuvchilar ro'yxati ham grafik ma'lumotlarga misoldir. Bog'langan ma'lumotlarning kuchi va kuchi bir xil tugunlarda qurilgan bir nechta o'zaro bog'liq grafiklarni tahlil qilishda kuchga kiradi. Masalan, chetlari Facebook'dagi "do'stlar"ni ifodalaydi deb tasavvur qiling. Keling, xuddi shu odamlar bilan, lekin LinkedIn orqali biznes hamkasblarini bog'laydigan yana bir grafikni va #1Wxx filmlariga qiziqish asosida uchinchi grafikni olaylik. Ushbu uchta grafikni joylashtirish ko'plab qiziqarli savollarga javob beradi.

Audio, video va grafik ma'lumotlar.

Audio, video va grafiklar ma'lumotlar olimi uchun qiyin muammolarni keltirib chiqaradigan ma'lumotlar turlaridir. Inson nuqtai nazaridan ahamiyatsiz bo'lgan vazifalar (masalan, rasmdagi ob'ektni tanib olish) kompyuter uchun qiyin bo'lib chiqadi. 2014 yilda MLBAM (Major League Baseball Advanced Media) operativ tahlil qilish uchun har bir beysbol o'yini uchun video tasvirlarni taxminan 7 TBgacha kengaytirishini e'lon qildi. Stadionlardagi yuqori tezlikdagi kameralar, masalan, real vaqtda himoyachining traektoriyasini hisoblash uchun to'p va sportchilarning harakatlarini yozib oladi. Yaqinda DeepMind video o'yinlarni o'ynashni o'rganishi mumkin bo'lgan algoritm yaratishga muvaffaq bo'ldi. Algoritm o'z vaqtida ekranning mazmunini oladi va bu ma'lumotlarni murakkab chuqur o'rganish jarayonida qanday izohlashni o'rganadi. Bu ajoyib yutuq va Google sun'iy intellektni rivojlantirish uchun DeepMind-ni sotib oldi. O'rganish algoritmi kompyuter o'yini tomonidan yaratilgan ma'lumotlarni, ya'ni oqim ma'lumotlarini oladi.

Oqimli ma'lumotlar

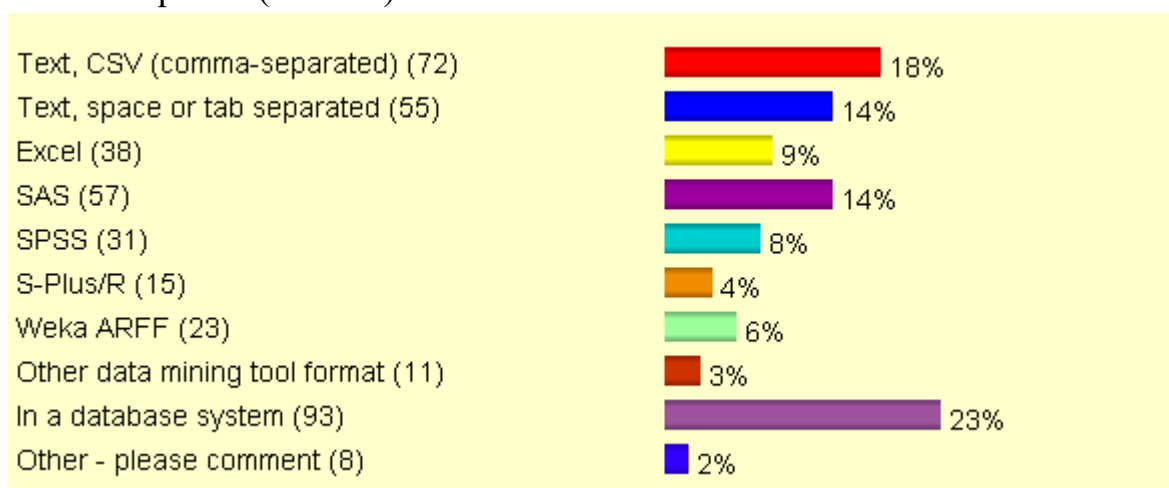
Oqim ma'lumotlari deyarli har qanday shaklni olishi mumkin, ammo ular bitta qo'shimcha xususiyatga ega. Ma'lumotlar katta massivlarda ma'lumotlar omboriga yuklangandan ko'ra, ma'lum hodisalar sodir bo'lganda tizimga kiradi. Garchi ular texnik jihatdan alohida turdagi ma'lumotlar bo'lmasa-da, biz ularni maxsus toifaga ajratamiz, chunki siz o'zingizning ish oqimingizni oqimli ma'lumotlar bilan ishlashga moslashingiz kerak bo'ladi. Oqimli ma'lumotlarga misol sifatida "Nima bo'lyapti?" bo'limi bo'lishi mumkin. Twitterda, jonli sport va musiqa tadbirlari va birja kotirovkalari.

1.3. Ma'lumotlarni saqlash formatlari.

Yuqorida ta'kidlanganidek, katta ma'lumotlarning asosiy xususiyatlaridan biri ko'p ma'lumotlarni turli va xar xil to'plamlardan iborat ekanligidir. Umuman? ma'lumotlar bilan ishlashning to'rt xil o'ziga xos jihati mavjud:

- ma'lumotlarni aniqlash;
- hisoblash;
- manipulyatsiya qilish;
- qayta ishlash (yig'ish, uzatish va boshqalar).

Ma'lumotlarni manipulyatsiya qilishda "fayl" tipidagi ma'lumotlar strukturasidan foydalaniladi. Bunda fayllar turli formatlarda bo'lishi mumkin. Xozirgi vaqtda ma'lumotlar yaratilish va foydalanish maqsadiga ko'ra turli formatlarda saqlanadi(6 -rasm).



6-rasm. Fayl tipida saqlanadigan ma'lumotlar ulushi (% da)

Ma'lumotlarning eng ko'p qismi (32%) ma'lumotlarni o'zlari foydalanadigan ma'lumotlar bazasi formatida saqlashni afzal ko'radi. Matn, CSV formatida - ma'lumotlarning 18%, Matn 14% har biri ma'lumotlarni, bo'sh joy yoki tabdan ajratilgan va SAS formatlarida saqlaydi; Excel formatida - 16%, SPSS - 8%, S-Plus / R - 4%, Weka ARFF - 6%, boshqa vosita formatlarida - 2%

Big Data fayl formatlari haqida shuni aytishimiz mumkinki, bugun biz chiziqli va ustun turlari o'rtasidagi farqni ko'rib chiqamiz, shuningdek, Apache Hadoop, Kafka, Spark, Flume bilan ishlashda AVRO, Sequence, Parquet, ORC va RCFile o'rtasida qanday tanlov qilish haqida gaplashamiz. Hive, Drill, Druid va boshqa katta ma'lumotlar vositalari. Shunday qilib, Big Data fayl formatlari bilan tanishib chiqamiz.

Big Data fayl formatlarining barcha xilma-xilligi (AVRO, Sequence, Parquet, ORC, RCFile) 2 toifaga bo'linishi mumkin: chiziqli (satr) va ustunli (ustun).

Row-oriented: rows stored sequentially in a file

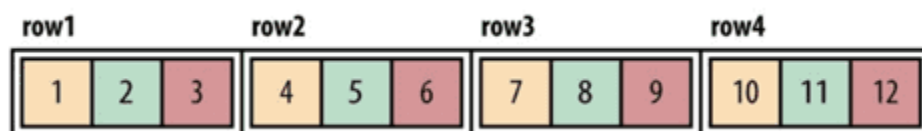
Key	Fname	Lname	State	Zip	Phone	Age	Sales
1	Bugs	Bunny	NY	11217	(123) 938-3235	34	100
2	Yosemite	Sam	CA	95389	(234) 375-6572	52	500
3	Daffy	Duck	NY	10013	(345) 227-1810	35	200
4	Elmer	Fudd	CA	04578	(456) 882-7323	43	10
5	Witch	Hazel	CA	01970	(567) 744-0991	57	250

Column-oriented: each column is stored in a separate file
Each column for a given row is at the same offset.

Key	Fname	Lname	State	Zip	Phone	Age	Sales
1	Bugs	Bunny	NY	11217	(123) 938-3235	34	100
2	Yosemite	Sam	CA	95389	(234) 375-6572	52	500
3	Daffy	Duck	NY	10013	(345) 227-1810	35	200
4	Elmer	Fudd	CA	04578	(456) 882-7323	43	10
5	Witch	Hazel	CA	01970	(567) 744-0991	57	250

7-rasm. Big Data fayl formatlari.

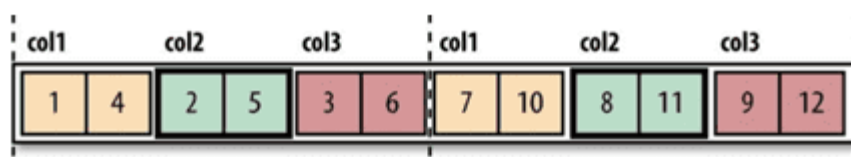
Chiziqli formatlarda (AVRO, Sequence) bir xil turdagi ma'lumotlar qatorlari qo'shni saqlashni tashkil qilish uchun birga saqlanadi. Agar satrdan faqat ba'zi qiymatlarni olish kerak bo'lsa ham, butun satr diskdan xotiraga o'qiladi. Ma'lumotlarni saqlashning chiziqli usuli o'qish operatsiyalari tezligini pasaytiradi va tanlangan so'rovlarni bajarishga, shuningdek, disk maydonining ko'proq sarflanishiga olib keladi. Chiziqli formatlar, ustunli formatlardan farqli o'laroq, qattiq yozilmaydi: masalan, Apache AVRO har bir maydonning turi haqidagi ma'lumotlarni sxema bilan birga metama'lumotlar bo'limida saqlaydi, shuning uchun ketma-ketlashtirilgan ma'lumotlar ma'lumotlarini o'qish sxema haqida oldindan ma'lumotni talab qilmaydi. Apache Hadoop ekotizimida katta ma'lumotlarni ketma-ket kalit/qiymat juftlari sifatida saqlash uchun ikkilik ketma-ketlik fayl formati fayl sarlavhasida ham metama'lumotlarni o'z ichiga oladi. E'tibor bering, Sequence File formati MapReduce vazifalarini bajarishda parallellik uchun juda yaxshi, chunki bir xil faylning turli qismlarini ochish va mustaqil ravishda ishlatish mumkin. Biroq satr formatlarida axborotni siqish darajasi ustun formatlariga qaraganda pastroq. Biroq, bu oqimli yozish uchun eng mos keladigan chiziqli formatlardir, chunki. nosozlik bo'lsa, ma'lumotni oxirgi sinxronlash nuqtasidan tiklash (qayta sinxronlash) mumkin.



8-rasm. Ustunga yo'naltirilgan formatlar.

Ustunga yo'naltirilgan formatlarda (Parquet, RCFile, ORCFile) fayl birgalikda saqlanadigan, lekin bir-biridan mustaqil ravishda qayta ishlanishi mumkin bo'lgan bir nechta ma'lumotlar ustunlariga bo'linadi. Ma'lumotni saqlashning ushbu usuli

ma'lumotlarni o'qishda keraksiz ustunlarni o'tkazib yuborishga imkon beradi, bu ma'lumotlarni o'qishni sezilarli darajada tezlashtiradi va sizga kichik qatorlar kerak bo'lganda yoki Apache Hive DBMS kabi tanlangan so'rovlarni bajarish juda yaxshi. Ammo bu o'qish/yozish formati RAMda ko'proq joy egallaydi, chunki har bir satr bir nechta satr ustunini olish uchun keshlanadi. Bundan tashqari, ustunga yo'naltirilgan formatlar oqimni qayta ishlash muhitlarida (Apache Kafka, Flume) ishlatilmaydi, chunki yozish xatosidan so'ng, yo'qolgan sinxronlash nuqtalari tufayli joriy faylni tiklab bo'lmaydi. Biroq, ustunlar bo'yicha ma'lumotlarning yanada samarali siqilishi tufayli ustunli fayllar qattiq diskda kamroq joy egallaydi [1].



9-rasm. Ustunlar bo'yicha ma'lumotlarni nusxali saqlash tamoili.

BIG DATA UCHUN QAYSI FAYL FORMATINI TANLASH UCHUN

Shunday qilib, Big Data fayllarining ustunli va chiziqli formatlari o'rtasidagi yuqorida tavsiflangan arxitektura farqlari ularning ma'lumotlarni saqlash, o'qish va yozishdagi turli samaradorligini aniqlaydi. Shuning uchun, ma'lumotni qayta ishlash tezligini u yoki bu formatda boshqalarga nisbatan ibtidoiy taqqoslash butunlay to'g'ri emas, chunki. nafaqat formatlar, balki ulardan foydalanadigan dastur tizimlarining algoritmik sifati ham baholanadi. E'tibor bering, Big Data mahsulotining har bir yetkazib beruvchisi, birinchi navbatda, o'zining ma'lumotlar formatlarini va uning yechimi bilan foydalanish uchun eng mos bo'lganlarini targ'ib qilishga intiladi, masalan, Cloudera'dan Apache Hadoop tarqatish uchun Data Engineer sertifikat imtihoni (CHD, Cloudera's Distribution of Hadoop) paraket bo'yicha mavzularni o'z ichiga oladi, Hortonworks Data Platformidagi shunga o'xshash sertifikat ORC formatiga ko'proq e'tibor beradi. Paraket, shuningdek, Apache Impala, Drill va Big Data MAPR platformalarida keng qo'llaniladi. Va Apache Kafkaning tijorat versiyasini ishlab chiqaruvchi Confluent Corporation AVRO formatini afzal ko'radi.

Shunday qilib, ma'lumotlar formatini tanlashda, birinchi navbatda, ma'lum bir Big Data tizimining funksionalligi doirasida uning yordami bilan hal qilinishi kerak bo'lgan amaliy muammolarni yodda tutish kerak [6]. Masalan, chiziqli AVRO formati ma'lumot yozishning yuqori tezligini ta'minlaydi va shuning uchun u Apache Kafka, Flume va korporativ ma'lumotlar ko'llarida (Data Lake) Big Data oqimlarini qayta ishlash uchun juda mos keladi, shuningdek, barcha maydonlarni to'liq o'qish muammosini yaxshi hal qiladi. ETL- omborlarda va ma'lumotlar martlarida talab qilinadigan yozuvning.

O‘z navbatida, ustunga yo‘naltirilgan formatlar (Parquet, RCFile, ORCFile) keraksiz ustunlarni o‘tkazib yuborish orqali fayldan ma'lumotlarni tezroq o‘qiydi va diskda kamroq joy egallaydi. Shuning uchun ustun formatlari fayl omborlarida va Apache Hadoop-ga asoslangan DBMSda, xususan, Hive-da faol qo‘llaniladi. SQL so‘rovlaridan foydalanish kontekstida, boshqa ustun fayllariga nisbatan ORC formatining ba'zi asosiy xususiyatlarini ta'kidlash kerak:

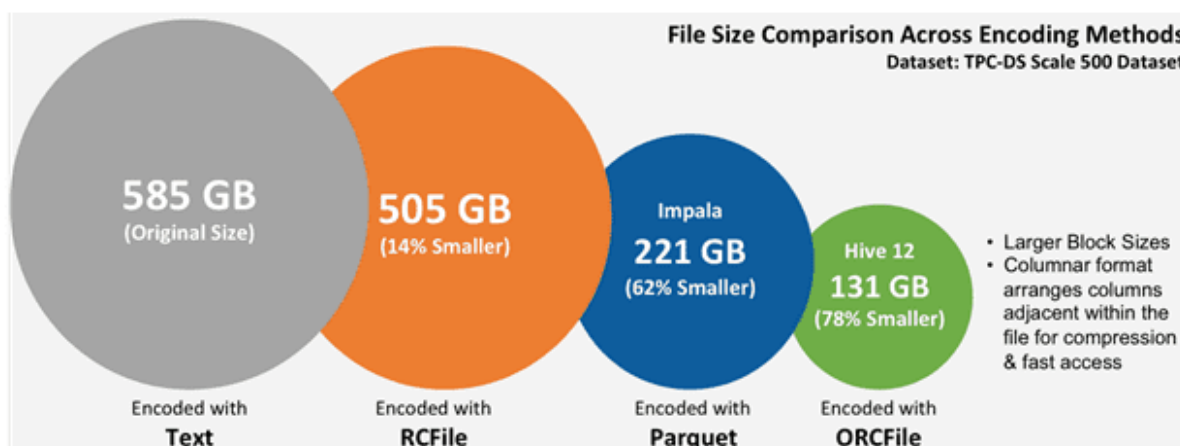
kiritish-chiqarish operatsiyalarini tezlashtiradigan har bir ustunni blok indekslash;

ustun darajasida metama'lumotlarni o‘qish SQL so‘rovlarini optimallashtirish imkonini beradi;

ACID tranzaksiya talablarini qo‘llab-quvvatlash (Atomicity - Atomicity, Consistency - Consistency, izolyatsiya - izolyatsiya, chidamlilik - chidamlilik);

fayllarni yuqoriroq siqish qattiq diskdagi joyni tejaydi .

Va Apache AVRO formatining o‘ziga xos afzalliklari orasida odamlar o‘qiy oladigan JSON formati kiradi, bu moslikni ta'minlaydi va o‘zgartirishlar alohida maydonlarni o‘tkazib yuborish, qo‘shish yoki o‘zgartirish orqali qayta ishlanganda ma'lumotlar sxemalarining evolyutsiyasini qo‘llab-quvvatlaydi.



10-rasm. Big Data formatlari nusxa saqlashda eng yaxshi siquvchi ORC fayllaridir

Shunday qilib, eng keng tarqalgan Big Data fayl formatlarining qiyosiy tavsifi yana bir bor eng yaxshi samara aniq vazifaga qaratilgan vositadan foydalanish ekanligini tasdiqlaydi. Masalan, Apache Spark yuqoridagi barcha ma'lumotlar formatlarini (AVRO, Sequence, Parket, ORC, RCFile) qo‘llab-quvvatlaydi, lekin onlayn tahlil rejimida ustunli fayllar va axborot oqimlarini qayta ishlashda string fayllari bilan yaxshi ishlaydi.

Big Data loyihangiz uchun qaysi formatni tanlashni va u bilan qanday ishlashni IT-mutaxassislarini tayyorlash va malakasini oshirish bo‘yicha litsenziyalangan o‘quv markazida menejerlar, arxitektorlar, muhandislar, ma’murlar, tahlilchilar va Data Sciens uchun katta ma'lumotlar bo‘yicha amaliy kurslarimizda bilib oling. Moskvada:

HDDE: Ma'lumotlar muhandislari uchun Hadoop
KAFKA: Kafka klasterini boshqarish
SPARK: Apache Spark klaster ma'muri
DEVKA1: Ishlab chiquvchilar uchun Kafka oqimlari
DEVKA2: Ishlab chiquvchilar uchun Kafka integratsiyasi.

2. KATTA MA'LUMOTLARDA DATA SCIENCE

2.1. Data Science va uning asosiy tushunchalari.

Hozirgi kunda Data Science deyarli barcha ma'lumotlar tarmoqlarida - chakana savdo, elektron tijorat, moliya, sog'liqni saqlash, ta'lim, va telekommunikatsiya sohalarida qo'llaniladi. Bundan tashqari internet olamida katta kuchga ega bo'lgan *Google, Apple, Facebook, Amazon* va boshqa bir qancha yirik kompaniyalar data sciencedan foydalanib ulkan ishlar qilmoqdalar. Ushbu kompaniyalar o'z bizneslarini rivojlantirish uchun aynan data sciencega murojat qiladilar va aniq va to'g'ri bo'lgan datalar orqali o'z qarorlarini joriy etadilar. Data Science nima? Data Science bu — malumot olish, tayyorlash, tahlil qilish, vizualizatsiya va ma'lumotlarni yetkazib berishdir. Bu ko'p qirrali soha bo'lib, ilmiy tahlillar, dasturiy algoritmlar, va kopleks jarayonlarni o'z ichiga oladi.

Hozirgi rivojlanayotgan texnologiya asrida, datalar ko'payib bormoqda va uni tahlil qiladigan, ular bilan ishlaydigan sohaga e'tibor ham kuchayyabti. Data Scienceni bilgan mutahassislarga "Data Scientist"larga ish bozorida ehtiyoj ham kundan-kun oshib boryabti. Data Scientistlar statistika va machine learning qurilmalaridan malumotlarni olish va tahlil qilish bilan shug'ullanib qolmay, machine learning algoritmlardan foydalanib kelajakdagi ehtimolli o'sish va pasayish jarayonlarini aniq datalar yordamida oldindan ko'rib, tahlil qilib berishyabti.

Ma'lumotlar muhandisligi juda katta soha va albatta sohaga qadam qo'yishdan avval bevosita ma'lumotlarni yig'ish va ularga ishlov berishni o'rganish zarur. Ma'lumotlar matn, jadval, rasm, audio/video, arxiv qog'ozlari va boshqa ko'rinishda bo'lishi mumkin. Sizning birlamchi vazifangiz esa bu ma'lumotlarga ishlov berishdan avval ularni bir ko'rinishga keltirish, bo'shliqlarni to'ldirish, xato ma'lumotlarni o'chirish va hokazolardan iborat bo'ladi.

Buning uchun oz bo'lsada turli ma'lumot turlari bilan ishlaydigan dasturlarini bilishimiz kerak. Masalan, jadvallar bilan ishlash uchun Excel, pdf yoki rasm ko'rinishidagi hujjatlarni matn ko'rinishiga keltiruvchi dasturlar va hokazo.

Data science – bu ma'lumotlar haqidagi fan bo'lib, sohalararo ilmiy usulublar, jarayonlar, algoritmlar va tizimlardan tashkil topgan tarkiblangan va tarkiblanmagan bilim va g'oyalarni o'z ichiga oluvchi fandir. **Data science** - ma'lumotlarni intellectual tahlil qilish, mashinani o'rganish va katta ma'lumotlar (Big data) bilan bog'liqdir.

Data science -bu ma'lumotlar yordamida "haqiqiy dunyo hodisalarini tushunish va tahlil qilish" maqsadida "statistika, ma'lumotlarni tahlil qilish va tegishli texnikalarni birlashtirish" konsepsiyalaridir. U matematika, statistika, informatika va informatika fan bilimlari kontekstida ko'plab sohalardan olingan usullar va nazariyalardan foydalanadi.

Turing mukofoti sovrindori Jim Grey Data science fani "to'rtinchi paradigma" (empirik, nazariy, hisoblash va hozirda ma'lumotlarga asoslangan) fan sifatida taqdim etdiladi hamda "barcha fanlardagi hamma narsalar axborot texnologiyalari ta'siri tufayli o'zgarib boradi" va ma'lumotlar oqimining kopayishini organadi – deb ta'kidlaydi. Ma'lumotlar haqidagi fan - data science – bazan datalogy - inglizcha so'zlardan olingan.

Bu fan raqamli shaklda ma'lumotlarni tahlil qilish, qayta ishlash va taqdim etish muammolarini o'rganadigan informatika sohasidir. Katta hajmli va yuqori darajadagi parallel tarzda ma'lumotlarni qayta ishlash usullarini, statistik usullarni, ma'lumotlarni intellektual qidirish tizimi usullarini va ma'lumotlar bilan ishlash uchun sun'iy intellekt ilovalarini, shuningdek ma'lumotlar bazalarini loyihalash va ishlab chiqish usullarini birlashtiradigan fandir.

2.2. Data science jarayonlari

Data science – bu katta hajimli turli sohalar va jarayonlar, narsalarga doir xilma-xil ma'lumotlarni yuqori tezlikda qayta ishlaydigan fan sohasidir

Data science jarayonlar odatda 6 bosqichda amalga oshiriladi. Data science amalga oshirish jarayoni Data science jarayon deb ataladi. Big Data ni tashkil topishi bevosita Data science jarayonlar bilan bog'liq

Data science jarayonlari

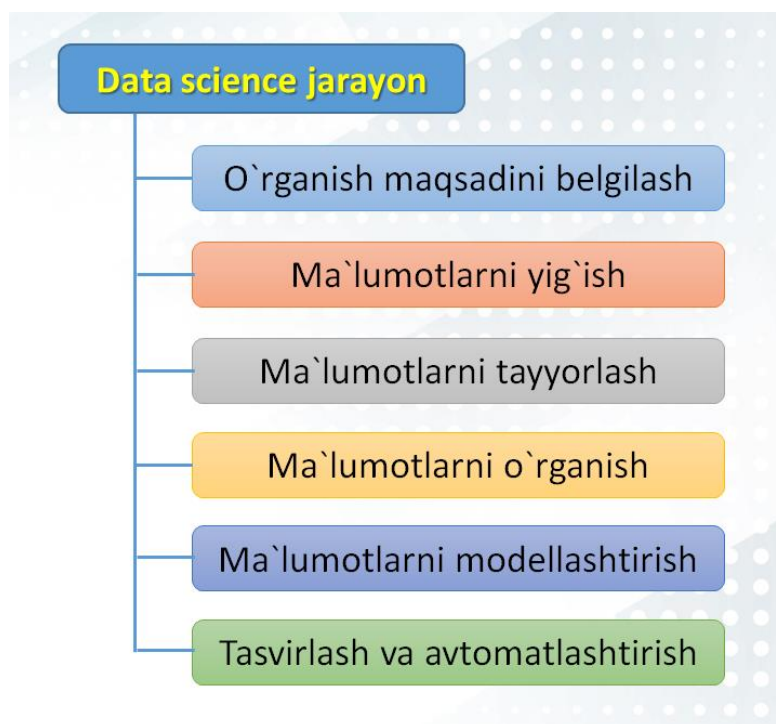
1. O'rganish maqsadini belgilash
2. Ma'lumotlarni yig'ish
3. Ma'lumotlarni o'rganish
4. Ma'lumotlarni tayyorlash
5. Ma'lumotlarni modellashtirish
6. Tasvirlash va avtomatlashtirish

Ma'lumotlar ombori

Ma'lumotlar bilan ishlash uchun o'z-o'zidan ma'lumotlarni biror joyda saqlash, kerak vaqtda unga murojaat qilish, ma'lumotlar orasidan o'zimizga kerak qismini filtrlab, ajratib olishni ham bilishimiz kerak.

Bugungi kunda SQL (va unga o'xshash) ma'lumotlar omboridan foydalanish ma'lumotlarni saqlashning eng ommabop usuli hisoblanadi. Demak siz ham SQL

ma'lumotlar ombori bilan ishlashni minimal darajada bilishingiz kerak bo'ladi. Jumladan, omborga murojaat qilish, undagi ma'lumotlarni ma'lum parametr yoki filtr asosida chaqirib olish, ma'lumotlarni tartiblash, ularni qaytarib omborga yuklash va hokazo.



11-rasm. Data science jarayonni umumiy ko'rinishi.

2.3. Ma'lumotlarni jamlash va ishlov berish

Ma'lumotlar muhandisligi juda katta soha va albatta sohaga qadam qo'yishdan avval bevosita ma'lumotlarni yig'ish va ularga ishlov berishni o'rganish zarur.

Ma'lumotlar matn, jadval, rasm, audio/video, arxiv qog'ozlari va boshqa ko'rinishda bo'lishi mumkin. Sizning birlamchi vazifangiz esa bu ma'lumotlarga ishlov berishdan avval ularni bir ko'rinishga keltirish, bo'shliqlarni to'ldirish, xato ma'lumotlarni o'chirish va hokazolardan iborat bo'ladi.

Buning uchun oz bo'lsada turli ma'lumot turlari bilan ishlaydigan dasturlarini bilishingiz kerak. Masalan, jadvallar bilan ishlash uchun Excel, pdf yoki rasm ko'rinishidagi hujjatlarni matn ko'rinishiga keltiruvchi dasturlar va hokazo.

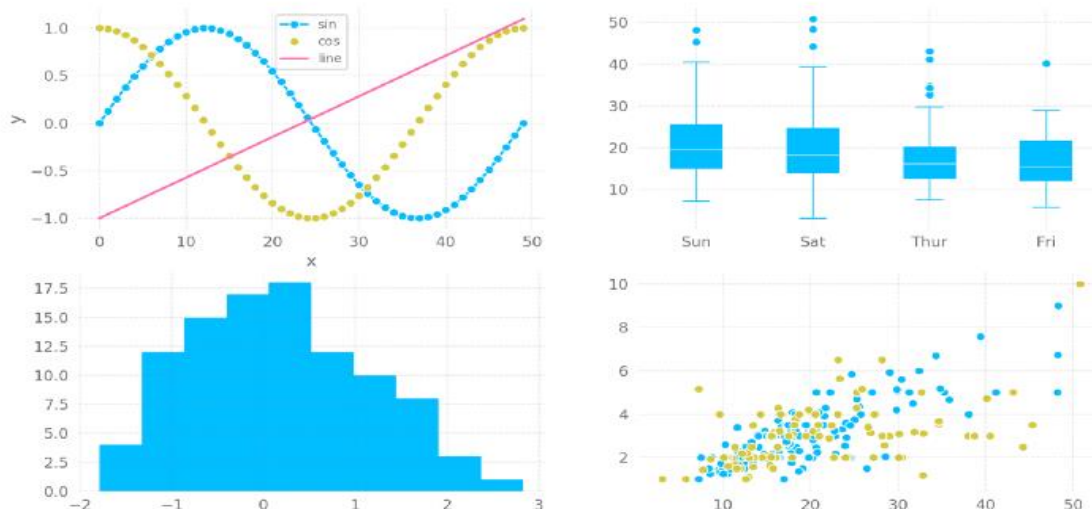
Ma'lumotlarga bevosita Pythonda ishlov berish uchun esa pandas kutubxonasi bilan yaqindan tanishib chiqishni tavsiya qilaman.

Shuningdek bu qadamning muhim qismlaridan biri turli manbaalarga murojaat qilish orqali loyihangiz uchun kerakli ma'lumotlar to'plamini topish yoki jamlash.

Odatda bunday to'plamlar **datasets** deb ataladi va onlaynda turli mavzulardagi datasetlar yig'ilgan saytlar talaygina.

2.4. Ma'lumotlarni vizuallash

Ma'lumotlarni tushunarli ko'rinishga keltirish, mijozlarga yoki hamkasblarga taqdim qilish uchun ularni quruq jadval va sonlar ko'rinishida emas, chiroyli grafiklar ko'rinishida ko'rsatish o'ta muhim. Bu ma'lumotlarni tahlil qilishni ham bir necha barobar osonlashtiradi.



12-rasm. Ma'lumotlarni vizuallashga misol.

Texnologik jarayonlarni avtomatlashtirish tizimlarini loyihalashda loyiha hujjatlarining sifatini oshirish, ularning hajmini va muddatini qisqartirish uchun avtomatlashtirish sohasida ilg'or sanoat tajribalarini o'zida mujassamlashtirgan instruktiv va normativ materiallarga asoslanish, shuningdek, umumsanoat va tarmoq xarakteriga ega bo'lgan normativ materiallardan foydalanish kerak. Texnologik jarayonlarni avtomatlashtirish tizimlarining loyihalarini yaratishda tipaviy loyihalar, yechimlar, konstruksiyalar va shu kabilardan maksimal darajada foydalanish kerak. Avtomatlashtirish tizimlarini loyihalash murakkab va sermehnat talab jarayon bo'lsa, unda ijodiy ish (muhandislik tahlili, yechimlar variantlarini tayyorlash) tipaviy-loyihaviy yechimlardan foydalanish bilan qo'shib olib borilgani uchun ko'pchilik jamoalarning kuchi avtomatik loyihalash tizimlarini (ALT), avtomatlashtirish tizimlarini ishlab chiqish bilan bog'liq masalalarni hal etishga qaratilgan. Bunda ALT deganda loyihalashning turli bosqichlarida masalalarni bosqichma-bosqich hal etishni ta'minlovchi EHM lar uchun hisoblash dasturlari to'plami tushuniladi. Bu ishlarni bajarishning birinchi bosqichi tarmoq loyiha tashkilotlarida tarmoqda foydalaniladigan avtomatlashtirishning texnik vositalari nomenklaturasini aks ettiruvchi axborot hisoblash bazasini yaratish hisoblanadi. Hozirgi paytda avtomatlashtirish tizimlarini loyihalashning noijodiy qismi ma'lum darajada rasmiylashtirilgan va zamonaviy HT vositalaridan foydalanib hal qilinmoqda, avtomatlashtirish elementlari va vositalarini hisoblash, AHV ni tahlil va sintez qilish, loyihalashning matn va chizma qismlarini rasmiylashtirish

(bezash). Loyihalashni avtom atlashtirish, loyihaviy hujjatlarni ishlab chiqish muddatlarini kamaytiradi va uning sifatini oshiradi.

2.5. Mashinali o‘qitish (Machine Learning)

Mana endi bevosita qiziq joyiga keldik. Ma’lumotlaringiz tayyor bo‘lsa endi bu ma’lumotlar yordamida kompyuterni tarbiyalash va turli foydali dasturlar yaratish mumkin.

Machine Learning juda katta yo‘nalish bo‘lib, supervised learning (boshqaruv ostida tarbiyalash), unsupervised learning (mustaqil tarbiyalash), reinforcement learning kabi katta sohalardan iborat.

Deep Learning

Deep Learning ham aslida yuqoridagi Machine Learning yo‘nalishing bir qismi hisoblanib, Neural Network yordamida murakkab muammolarga yechim topuvchi modellarni yaratishga qaratilgan. Jumladan, Deep Learning yordamida odam yuzini tanish, tasvirlardagi obyektlarni klassifikasiya qilish, ovozdan tanish, tabiiy tillar bilan ishlash (chatbotlar) kabi dasturlar yaratiladi.

Bu qadamda ham yuqorida zikr qilingan kreas, Tensorflow va PyTorch frameworklaridan foydalaniladi.

Natural Language Processing (Nlp)

Tabiiy tillar bilan ishlash yo‘nalishi. Bu qadam mutlaqo ixtiyoriy hisoblanib, o‘z darajasini yana bir pog‘ona oshirishni istaganlarga mo‘ljallangan.

Bu yo‘nalish nutq va matnlarni tahlil qilish orqali odam tilini tushunuvchi aqlli dasturlarni yaratishga qaratilgan. Zamonaviy telefon va kompyuterlardagi “Ok Google” yoki “Siri” xizmatlari, turli saytlardagi aqlli chatbotlar aynan shunday dasturlar turkumiga kiradi.

Bu yo‘nalish aynan qaysi tilda dasturlar yaratishingizga qarab juda oson (ingliz, rus) yoki o‘ta murakkab (o‘zbek) bo‘lishi mumkin.

2-Amaliy mashg'ulot. Katta hajimli ma'lumotlarni qayta ishlashda qo'llanlayotgan texnologiyalar. (2 soat amaliy)

Reja

1. Katta hajimli ma'lumotlarni qayta ishlashda qo'llanlayotgan texnologiyalar.
2. NoSQL, MapReduce, Hadoop, R, ERP, SAP.
3. Matnlarni chuqur va ma'lumotlarni vizual tahlil qilish.

1. Katta hajimli ma'lumotlarni qayta ishlashda qo'llanlayotgan texnologiyalar.

Big Data texnologiyasining foydalanish ko'lamida juda katta. Big Data yordamida siz mijozlarning hohish-istaklari, marketing kompaniyalarining samaradorligi to'g'risida ma'lumot olishingiz yoki xavf tahlili o'tkazishingiz mumkin. Quyida IBM instituti tomonidan kompaniyalarda Big Data-dan foydalanish sohalari bo'yicha o'tkazilgan so'rov natijalari keltirilgan.

Big Datadan mijozlarga xizmat ko'rsatish sohasida foydalanishadi, ikkinchi eng mashhur yo'nalish operatsion samaradorlik, axborot xavfsizligi sohasida Big Data hozirgi paytda uchinchi o'rinda. Katta ma'lumotlarni yig'ish va qayta ishlashda foydalaniladigan texnologiyalarni 3 guruhga bo'lish mumkin:

- Dasturiy ta'minot;
- Uskunalar;
- Xizmatlar

Ma'lumotni qayta ishlashning eng keng tarqalgan usullari (dasturiy ta'minot) quyidagilarni o'z ichiga oladi:

SQL-ma'lumotlar bazalari bilan ishlashga imkon beradigan tuzilgan so'rovlar tili. SQL-dan foydalanib, siz ma'lumotlarni yaratishingiz va o'zgartirishingiz mumkin va tegishli ma'lumotlar bazasini boshqarish tizimi ma'lumotlar qatorini boshqarish uchun javobgardir.

NoSQL-atama nafaqat SQL (SQLemas) degan ma'noni anglatadi. Bu ma'lumotlar bazasini amalga oshirishga yo'naltirilgan bir qator yondashuvlarni o'z ichiga oladi, ular an'anaviy, bog'langan MBBLlarda foydalaniladigan modellardan farq qiladi. Ular doimo o'zgarib turadigan ma'lumotlar tuzilishi bilan foydalanish uchun qulaydir. Masalan, ijtimoiy tarmoqlarda ma'lumot to'plash va saqlashda.

Mapreduce-hisob-kitoblarni taqsimlash modeli. U juda katta ma'lumot to'plamlari (petabaytlar) bo'yicha parallel hisoblashlarda qo'llaniladi. Dastur interfeysida ma'lumotlar qayta ishlash uchun dasturga emas, balki dastur ma'lumotlarga uzatiladi. Shunday qilib, so'rov alohida dastur bo'lib, jarayon prinsipi –ikki usulida ma'lumotlarni ketma-ket qayta ishlaydi, Map (Xarita) dastlabki ma'lumotlarni tanlaydi va Reduce ularni agregatlarini qisqartiradi.

Hadoop-Facebook, eBay, Amazon va boshqalarga yuklangan saytlarni qidirish va kontekstual mexanizmlarini amalga oshirishda foydalaniladi. O'ziga xos xususiyati shundaki, tizim biron bir klaster tugunlarining ishdan chiqmasidan himoyalangan, chunki har bir blok tugunda boshqa ma'lumotlarning kamida bitta nusxasi bor.

SAP XANA-ma'lumotlarni saqlash va qaytaishlash uchun yuqori samarali NewSQL platformasi. Yuqori tezlikda so'rovlarni qayta ishlashni ta'minlaydi. Yana bir o'ziga xos jihat shundaki, SAPXANA analitik tizimlarni qo'llab-quvvatlash xarajatlarini kamaytirgan holda, tizim landshaftini soddalashtiradi. Shuningdek, Big Data tegishli vositalar Hive, HBase, MapReduce, Spark RDD, Spark Streaming, SparkSQL, SparkR, MLlib, Flume, Sqoop, Oozie, Kafka, Data freym'lari va GraphX vositalari bilan ishlaydi. Texnologik uskunalar quyidagilarni o'z ichiga oladi:

1. Serverlar-ma'lumotlar omborini o'z ichiga oladi.
2. Infratuzilma uskunalari-infratuzilma uskunalari platformani tezlashtirish vositalari, uzluksiz quvvat manbalari, server konsollari to'plamlari va boshqalarni o'z ichiga oladi.
3. Xizmatlar-xizmatlar ma'lumotlar bazasi arxitekturasini yaratish, infratuzilmani tashkil qilish va optimallashtirish, ma'lumotlarni saqlashni ta'minlashni o'z ichiga oladi. Big Data ko'plab biznes sohalarida keng qo'llaniladi. Ular sog'liqni saqlash, telekommunikatsiya, savdo, logistika, moliyaviy kompaniyalarda, shuningdek davlat boshqaruvida qo'llaniladi. Quyida ba'zi sohalarida Big Data qo'llanishiga misollar keltiramiz:

Savdoda. Chakana savdo do'konlarining ma'lumotlar bazasida mijozlar, inventarizatsiyani boshqarish tizimi, tijorat mahsulotlarini yetkazib berish to'g'risida ko'plab ma'lumotlar to'planishi mumkin. Ushbu ma'lumotlarni do'konlarning barcha sohalarida foydali bo'lishi mumkin. Shunday qilib, to'plangan ma'lumotlar yordamida siz tovarlarni yetkazib berish, uni saqlash va sotishni boshqarishingiz mumkin. Yig'ilgan ma'lumotlarga asoslanib, siz tovarlarga talab va taklifni taxmin qilishingiz mumkin. Shuningdek, ma'lumotlarni qayta ishlash va tahlil qilish tizimi chakana sotuvchining boshqa muammolarini hal qilishi mumkin, masalan xarajatlarni optimallashtirishda yoki hisobotlarni tayyorlashda.

Sog'liqni saqlashda - bu katta hajmdagi ma'lumotlarni qayta ishlashga majbur bo'lgan yana bir sohadir. Quyida katta ma'lumotlarning sog'liqni saqlash tizimiga yordam berishi mumkin bo'lgan ba'zi usullari mavjud: Katta ma'lumotlar davolanish xarajatlarini kamaytiradi, chunki keraksiz tashxis qo'yish imkoniyatini kamaytiradi. Oldindan epidemiyani bashorat qilishda va uning oqibatlarini minimallashtirish uchun qanday profilaktika choralarini ko'rish bo'yicha yordam beradi. Katta ma'lumotlar kasalliklarni dastlabki bosqichlarida aniqlash orqali oldini olishga yordam beradi. Sog'liqni saqlash sohasida foylanidigan asboblardan va datchiklarga joriy etilsa, ular bemorning elektron ma'lumotnomasini real vaqt rejimida yetkazib bera oladi. Bunday texnologiyalardan biri Apple kompaniyasiga tegishlidir. Apple HealthKit, CareKit va ResearchKit-ni ishlab chiqdi. Asosiy maqsad -iPhone foydalanuvchilari o'zlarining real vaqt rejimidagi sog'liqni saqlash yozuvlarini (kassalik tarixini) telefonlarida saqlash va ularga kirish huquqini berishdir.

Ma'lumotlar qaror qabul qilishda muhim rol o'ynaydigan bugungi dunyoda matn ma'lumotlarini vizuallashtirish va katta hajimli axborotni tahlil qilish va

tushunish uchun muhim vositaga aylanmoqda. Ushbu usullar nafaqat ma'lumotlarni vizual shaklda taqdim etish imkonini beradi, balki uning sifatini yaxshilaydi va turli vazifalarni bajarish uchun yanada foydali bo'ladi. Bugungi kunda katta hajmdagi kengaytirilgan, ammo keraksiz ma'lumotlardan xoli va 1-5 jumladan iborat asl so'rovga iloji boricha yaqinroq yaratish kifoya.

Data Mining usullaridan foydalangan holda ma'lumotlarni tahlil qilish natijasida olingan natijalar har doim ham inson idroki uchun qulay emas. Ko'pgina tasniflash yoki ijtimoiy qoidalarda, matematik formulalarda, odam uchun yangi va foydali bilimlarni tez va oson topish juda qiyin. Axborotning murakkabligi sababli, bu qarorlar daraxtlari, datagramlar, ikki o'lchovli grafiklar va boshqalar kabi bilimlarni ifodalashning eng oddiy grafik turlarida har doim ham mumkin emas. Shu munosabat bilan tahlilni ko'rsatishning yanada murakkab vositalariga ehtiyoj bor. Bularga vizual ma'lumotlarni tahlil qilish vositalari kiradi, ular xorijiy adabiyotlarda ko'pincha Visual Mining atamasi deb ataladi.

Vizual ma'lumotlarni tahlil qilishning asosiy g'oyasi ma'lumotlarni qandaydir vizual shaklda taqdim etishdan iborat bo'lib, odamga ma'lumotlarga kirish, uning vizual tasviri bilan ishlash, uning mohiyatini tushunish, xulosalar chiqarish va ma'lumotlar bilan bevosita o'zaro aloqada bo'lish imkonini beradi.

Ma'lumotlar bazalarida saqlanadigan tuzilgan ma'lumotlarni tahlil qilish dastlabki ishlov berishni talab qiladi: ma'lumotlar bazasini loyihalash, ma'lum qoidalarga muvofiq ma'lumotlarni kiritish, uni maxsus tuzilmalarga joylashtirish (masalan, relyatsion jadvallar) va hokazo. Shunday qilib, bevosita ushbu ma'lumotlarni tahlil qilish va undan yangi bilimlarni olish uchun qo'shimcha harakat talab qiladi. Biroq, ular har doim ham tahlil bilan bog'liq emas va kerakli natijaga olib kelishi shart emas. Shu sababli, tuzilgan ma'lumotlarni tahlil qilish samaradorligi pasayadi.

Bundan tashqari, ma'lumotlarni tahlil qilishda barcha turdagi foydali ma'lumotlarni yo'qotmasdan tuzilishi mumkin emas. Masalan, matnning semantikasini va ob'ektlar o'rtasidagi munosabatlarni yo'qotmasdan, matnli hujjatlarni jadval ko'rinishiga aylantirish deyarli mumkin emas. Shu sababli, bunday hujjatlar ma'lumotlar bazasida matn maydonlari (BLOB maydonlari) kabi o'zgartirilmasdan saqlanadi.

Shu bilan birga, matnda juda katta miqdordagi ma'lumotlar yashiringan, ammo uning tuzilmaganligi Data Mining algoritmlaridan foydalanishga imkon bermaydi. Strukturalanmagan matnni tahlil qilish usullari ushbu muammoni hal qiladi. Adabiyotlarda bunday tahlil Text Mining -matnni qazib olish deb ataladi.

Tarkibi tuzilmagan matnlarni tahlil qilish usullari bir nechta sohalarning kesishmasida joylashgan: ma'lumotlarni qazib olish, tabiiy tilni qayta ishlash, axborotni qidirish, ma'lumot olish va bilimlarni boshqarish.

Quydagı savollarga javob bering:

(Savollarga javoblar yozma kompyuterda tayyorlanadi)

1. Big Data fani nimani o'rganadi?
2. Big Data VVV xususiyati nimani anglatadi?
3. Katta ma'lumot bo'lish uchun hajm qay darajada katta bo'lishi kerak.
4. Katta ma'lumotlarga doir qnday atamalarni bilasiz?
5. Katta hajmdagi ma'lumotlarga qo'yiladigan talablar qanday bo'ladi?
6. Katta hajmdagi ma'lumotlar bilan ishlaydigan qanday axborot tizimlarini bilasiz?
7. Ma'lumotlarni tahlil qilishning qanday yondashuvlari bor?
8. Ma'lumotlarni tahlil qilishda vizualizatsiya nima?
9. Ma'lumotlarni tahlil qilishda qanday matematik usullardan foydalaniladi?
- 10.10.O'zbekistonda qanday Big Data tizimlari mavjud?
- 11.Big Data ma'lumot tushunchasi va uning mohiyatini tushuntiring.
- 12.Axborot ta'riflarini ayting.
- 13.Bilim nima?
- 14.Bilm haqida Bernard Shouning qanday iqtibosi bor?
- 15.Ma'lumot nima?
- 16.Bugungi kunda dunyo miqyosida qancha elektron ma'lumot mavjud?
- 17.Ma'lumotni fizik o'lchovidan tashqari qanday o'lchovlari mavjud?
- 18.Katta ma'lumotlarda qanday ma'lumot turlarini bilasiz?
- 19.Ma'lumotlarni qanday saqlash formatlari bor?
- 20.Katta ma'lumotlarni umumiy qabul qilingan yagona formati mavjudmi?
- 21.Data Science nima?
- 22.Data Science jarayon qanday amalga oshiriladi?
- 23.Data Science jarayon necha bosqichdan iborat?
- 24.O'rganish maqsadini belgilashda qanday ishlar amalga oshiriladi?
- 25.Ma'lumotlarni yig'sh jarayonida qanday amallar bajariladi?
- 26.Ma'lumotlarni tayyorlash jarayonida qanday amallar bajariladi?
- 27.Ma'lumotlarni o'rganish jarayonida qanday amallar bajariladi?
- 28.Ma'lumotlarni modellashtirish jarayonida qanday amallar bajariladi?
- 29.Ma'lumotlarni tasvirlash jarayonida qanday amallar bajariladi?
- 30.Ma'lumotlarni avtomatlashtirish jarayonida qanday amallar bajariladi?
- 31.Data Mining tushunchasi nima?
- 32.Data Mining texnologiyalari haqida gapiring?
- 33.Data Miningni qo'llash sohalari haqida tushuntiring?
- 34.OLAP nima?
- 35.Tibbiyotda Data Mining haqida tushuntiring?

36. Iqtisodiyotda Data Mining haqida tushuntiring?
37. Data Mining usullari bilan aniqlanadigan beshta standartni ayting.
38. Assotsiatsiya nima?
39. Klasterlash qanday amalga oshiriladi?
40. Tasniflash qanday amalga oshiriladi?
41. Data Mining nima uchun qayerlarda qo'llaniladi?
42. Data Mining texnologiyasidan foydalanishni aylanish sxemasi qanday?
43. Data Miningni qo'llash sohalarini sanang.
44. Data Mining metodlarini ayting.
45. Data Mining bosqichlari qanday bo'ladi?
46. Iqtisodiyotda Data Mining qanday ahamiytga ega?
47. Aloqa va telekommunikatsiyada Data Mining o'rni qanday?
48. Meditsina, molikulyar genetika va gen muhandisligida Data Mining qanday ahamiytga ega?
49. Ma'lumotni qazib olish bo'yicha haqiqiy hayot misollar keltiring.
50. Data Miningda ma'lumotlar tahlilining vizualizatsiyasini tushuntiring.

V.KEYSLAR TESTLAR

Big Data va malumotlar tahlili

1-variant

1. Ko'p o'lchovli MBga ega MBBT ning matematik asosini kim ishlab chiqqan?

- A) Don Nelson B) Kennet Kukker
- C) Fon Neyman D) Edgar Kodda

2. Katta ma'lumotlarni asosan uchta V bilan hususiyatlari belgilanadi. Bu qaysi hususiyatlar?

- A) Visual – Variety – Velocity (viziuallik-xilma xillik-tezlik)
- B) Volume – Variety – Velocity (hajm-xilma xillik-tezlik)
- C) Volume – Virtual – Velocity (hajm-virtuallik-tezlik)
- D) Volume – Variety – Web (hajm-xilma xillik-web)

3. MapReduce printsipi bu ...

- A) Mijozlar, ular haqida ma'lum bo'lgan barcha ma'lumotlarga asoslanib, ishonchli mijozlarni taxlil qiladi
- B) ma'lumotlar dastlab saqlangan tugunlarda hisob-kitoblarni amalga oshirish, saqlash tizimlarining hisoblash quvvatidan foydalanish
- C) massivda parallel ishlov berish muammolarini hal qilish uchun funktsional dasturlashdan foydalanish
- D) Mijozlar, ular haqida ma'lum bo'lgan barcha ma'lumotlarga asoslanib, ishonchsiz mijozlarni taxlil qiladi

4. Kompyuter tomonidan harakatlarni yartuvchi, unga oldindan joylashtirilmagan bilim va ma'lumotlarni avtomat hosil qilishni tashkil etuvchi tadqiaot muhiti bu ...

- A) Ma'lumotlar taxlili
- B) Neyron tarmoq
- C) Mashinali o'qitish
- D) Ma'lumotlarni intellektual qidirish

5. Katta ma'lumotlar toplami bilan ishlaganda dasturchilar uchun qanday tavsiyalar beriladi?

- A) Bajarilgan ishlarni takrorlamang, qurilmani barcha imkoniyatidan foydalaning, hisoblash resurslarini tejang
- B) Bajarilgan ishlarni takrorlamang, qurilmani barcha imkoniyatidan foydalanish juda ham shert emas, hisoblash resurslarini tejang
- C) Bajarilgan ishlarni takrorlash mumkin, qurilmani barcha imkoniyatidan foydalaning, hisoblash resurslarini tejang
- D) Bajarilgan ishlarni takrorlamang, qurilmani barcha imkoniyatidan foydalaning, hisoblash resurslarini kattasini tanlang

6. Ma'lumotlar modellari yoki sxemalariga mos kelmaydigan va tarkibi bo'lmagan ma'lumotlar hamda korxona ma'lumotlarining 80% ni tashkil qiluvchi ma'lumotlar ...

- A) relatsion ma'lumotlar deb ataladi
- B) tarkiblanmagan ma'lumotlar deb ataladi
- C) tarkiblangan ma'lumotlar deb ataladi
- D) tarmoqli ma'lumotlar deb ataladi

7. Apache Hadoop loyihasiga qanday asosiy modullar kiritilgan?

- A) MapReduce, Hadoop Common, Ozone, Storm, HDFS
- B) MapReduce, Hadoop Common, Ozone, YARN, HDFS
- C) Spark, Hadoop Common, Ozone, YARN, HDFS
- D) MapReduce, Hadoop Common, Ozone, YARN, Pig

8. Kubda ma'lumotlarni bir yoki bir nechta o'lchov bilan filtrlash nima deb ataladi?

- A) Kub tomoni
- B) Kub manbasi
- C) Kubni kesish
- D) Kub proyeksiyasi

9. Ma'lumotlarni ierarxik, tarmoq va relytsion modelini mulohazali yaratish bu ...

- A) Modellashtirish
- B) Kontseptual modellashtirish
- C) Fizik modellashtirish
- D) Mantiqiy modellashtirish

10. Xopfild tarmog'i Xemming tarmog'i bilan qachon almashtiriladi?

- A) eslab qolingani namunalari sonini ko'paytirganda
- B) tarmoq namunalarni aniq saqlanganda va chiqarishga hojat qolmaganda
- C) tarmoqning konvergensiya vaqtini tezlashtirganda
- D) tarmoqning barqarorligi ta'minlanganda

11. Atributlar orasidagi bog'liqlikni topish bu ...

- A) Tasniflash (Klassifikatsiyalash) ("O'qituvchi bilan ta'lim olish")
- B) Bashoratlash (Prognozlash) ("O'qituvchi bilan o'rganish")
- C) Birlashmalar (Assosatsiya) ("O'qituvchisiz o'rganish")
- D) Klasterlash ("O'qituvchisiz o'rganish")

12. Matematik model va biologik neyron tarmoqlar asosida, tirik organizm nerv hujayralari tarmoqlarini tashkil etish va ishlash tamoyiliga qurilgan dasturiy yoki apparat ta'minoti bu ...

- A) Qarorlarni q'ollab quvvatlash tizimi - Online analytical processing (OLAP)
- B) Neyron tarmoq (sun'iy neyron tarmoq, SNT)
- C) Tranzaksiya tizimi - Online Transaction Processing (OLTP)

D) Obsessive-Compulsive Disorder (OCD)

13. Raqamli shaklda ma'lumotlarni tahlil qilish, qayta ishlash va taqdim etish muammolarini o'rganadigan informatika sohasi.

- A) Data mining
- B) Data source
- C) Data engineer
- D) Data science

14. "OLAP" so'zi kengaytirilganda qanday ma'noni anglatadi?

- A) OLAP (inglizcha, Online analytical processing, interaktiv analitik ishlov berish) -ko'p o'lchovli printsiptga muvofiq tuzilgan katta ma'lumotlar massivlari asosida umumlashtirilgan (jamlangan) ma'lumotlarni tayyorlash, qayta ishlash texnologiyasi.
- B) OLAP (inglizcha, Multidimensional Online Analytical, interaktiv analitik ishlov berish) -ko'p o'lchovli printsiptga muvofiq tuzilgan katta ma'lumotlar massivlari asosida umumlashtirilgan (jamlangan) ma'lumotlarni tayyorlash, qayta ishlash texnologiyas
- C) OLAP (inglizcha, Obsessive-Compulsive Disorder, interaktiv analitik ishlov berish) -ko'p o'lchovli printsiptga muvofiq tuzilgan katta ma'lumotlar massivlari asosida umumlashtirilgan (jamlangan) ma'lumotlarni tayyorlash, qayta ishlash texnologiyas
- D) OLAP (inglizcha, Online Transaction Processing, interaktiv analitik ishlov berish) -ko'p o'lchovli printsiptga muvofiq tuzilgan katta ma'lumotlar massivlari asosida umumlashtirilgan (jamlangan) ma'lumotlarni tayyorlash, qayta ishlash texnologiyas

15. Klassik Hadoop MapReduce-ning asosiy afzalliklaridan biri bu ...

- A) Bir nechta fayllarni tezda qayta ishlash qobiliyati
- B) Klaster rejimida ishlash qobiliyati
- C) Open-source holati
- D) hisoblash resurslariga yuqori talablar yo'qligi va deyarli har qanday mashinada ishlash qobiliyati

16. Kontseptual ko'p o'lchovli tasvir, shaffoflik, mavjudligi, hisobotni ishlab chiqishda barqaror samaradorlik, mijoz-server arxitekturas, umumiy ko'p o'lchovlilik, ajratilgan matritsalarini dinamik boshqarish, ko'p foydalanuvchilarni qo'llab-quvvatlash, cheksiz o'zaro operatsiyalar, intuitiv ma'lumotlarni manipulyatsiya qilish, hisobotlar olishning moslashuvchan imkoniyatlari, cheksiz o'lcham va yig'ish darajalari soni - OLAP ning qaysi qoidasida beriladi?

- A) Umumiy qoidalari
- B) Yahlilik qoidasi
- C) Tarqoqlik qoidasi
- D) Kodda qoidalari

17. Katta ma'lumotlarning to'rtta asosiy xususiyatini ko'rsating

- A) Virtualization, Volume, Variability, Vehicle
- B) Video, Value, Variety, Volume

- C) Verification, Volume, Velocity, Visualization
D) Variety, Velocity, Volume, Value

18. Katta ma'lumotlar tendentsiyasi ortidagi yurituvchi nima? (1) yirik korporatsiyalarning marketing kampaniyalari, (2) saqlash xarajatlarini kamaytirish

(3) oqimli ma'lumotlarni qayta ishlash uchun yangi texnologiyalarning paydo bo'lishi, (4) xotiradagi ma'lumotlar bazasini chiqarish

- A) 1,3 B) 1,2
C) 3,4 D) 2,4

19. Faktlarga asoslangan model bu ...

- A) RM/T model B) ER model
C) ORM model D) Funktsional model

20. Big Data so`zini lug`aviy ma`nosini belgilang

- A) Katta ma'lumot (inglizcha big data, ['big 'deɪtə])
B) Katta ma'lumot (italyancha big data, ['big 'deɪtə])
C) Katta ma'lumot (fransuzcha big data, ['big 'deɪtə])
D) Katta ma'lumot (ispancha big data, ['big 'deɪtə])

21. IBS kompaniyasi tahlilchilarini ma'lumotlariga ko`ra 2015 yilda ma'lumotlarning butun dunyo hajmini qanchani tashkil qilar edi?

- A) 5 ekzabayt ma'lumotlar (1 EB = 1 milliard gigabayt)
B) 6,5 zettabaytdan yuqori
C) 0,18 zettabayt (1 zb = 1024 ekzabayt)
D) 40-44 zettabayt

22. Ikki yoki undan ortiq paydo bo'lgan populyatsiyalarni (guruhlarini) "ajraladigan" (diskriminatsiya qiluvchi) atributlarni aniqlash bu ...

- A) Klasterlash ("O'qituvchisiz o'rganish")
B) Diskriminant tahlili ("O'qituvchisiz o'rganish")
C) Birlashmalar (Assosasiya) ("O'qituvchisiz o'rganish")
D) Bashoratlash (Prognozlash) ("O'qituvchi bilan o'rganish")

23. Ma'lumotlarni bir yoki bir nechta o'lchamlar bo'yicha yig'ish (kubning yiqilishi) nima deb ataladi?

- A) Kubni kesish B) Kub proyeksiyasi
C) Kub tomoni D) Kub manbasi

24. Katta ma'lumotlar va uning tahlili ko'plab odamlarni o'yinlar bilan o'ziga jalb qilish imkonini berdi. Bugungi kunda dunyo bo'ylab qancha odam kompyuter o'yinlarini o'ynaydi?

- A) 1,5 milliard
B) 1 milliard
C) 0,5 milliard
D) 2,5 milliard

25. Ob'ektlarni oldindan belgilangan toifalarga ajratish bu ...

- A) Bashoratlash (Prognozlash) ("O'qituvchi bilan o'rganish")
- B) Birlashmalar (Assosasiya) ("O'qituvchisiz o'rganish")
- C) Klasterlash ("O'qituvchisiz o'rganish")
- D) Tasniflash (Klassifikasiyalash) ("O'qituvchi bilan ta'lim olish")

2-variant

1. Xususiyatlari bo'yicha qolganlardan sezilarli darajada farq qiladigan ob'ektlarni qidirish bu ...

- A) Bashoratlash (Prognozlash) ("O'qituvchi bilan o'rganish")
- B) Istisnolarni aniqlash ("O'qituvchi bilan va o'qituvchisiz o'rganish")
- C) Birlashmalar (Assosasiya) ("O'qituvchisiz o'rganish")
- D) Tasniflash (Klassifikasiyalash) ("O'qituvchi bilan ta'lim olish")

2. Mashinali o'itish tushunchasi kim tomonidan fanga kiritilgan?

- A) Mayk Robert
- B) Don Nelson
- C) Edgar Kodda
- D) Artur Semyuel

3. Bugungi kunda katta ma'lumotlar mashinani o'rganish uchun ham faol foydalanilmoqda - "oqilona" algoritmlar (BERT, GPT-3 va boshqalar) shunday paydo bo'ldi, ular The "Guardian" da ustunlar yozadi va inson nutqini taniy oladi. Ushbu modellar yana nima qila oladi?

- A) Begona ovozlarni aniqlash
- B) Van Goga usulida chizish
- C) Van Goga usulida chizish va begona ovozlarni aniqlash
- D) Boshqa xech narsa qilmaydi

4. Qaysi ma'lumotlar hajmidan boshlab ma'lumotlarni saqlash uchun Hadoop klasteridan foydalanish maqsadga muvofiq?

- A) 100 GB va 1 TB
- B) 1 TB va 500 MB
- C) 1 TB va 250 GB
- D) 1 TB va 1 PB

5. HOLAP qanday tizim hisoblanadi?

- A) Ko'p o'lchovli OLAP tizim
- B) Bunday tizim mavjud emas
- C) Relyatsion OLAP tizim
- D) Gibril OLAP tizim

6. Ontologiya – bu ...

- A) ma'lum bir til yoki ular tavsiflangan tillardan mustaqil bo'lgan mavjud tushunchalar haqidagi nazariya
- B) ma'lumotlar tuzilmalari va ma'lumotlarni boshqarish bo'yicha operatsiyalarni tavsiflashning qulay usuli

- C) modelda ma'lumotlar va ular orasidagi munosabatlar ikki o'lchovli jadvallarni aloqalari ko'rinishida ifodalanadi
D) tarmoqlar sun'iy intellektning eng muhim sohalaridan biri

7. Koxonen neyroni qchon "g'olib" deb hisoblanadi?

- A) NET ning maksimal qiymati
B) OUT qiymatining maksimal qiymati
C) NET ning minimal qiymati
D) OUT qiymatining minimal qiymati

8. OLAP nechta turga bo'linadi?

- A) 2 turga B) 5 turga
C) Faqat bitta turi mavjud
D) 3 turga

9. O'yin sanoatidagi katta ma'lumotlar birinchi navbatda pul ishlashning bir usuli hisoblanadi. Avvaliga o'yin bepul bo'lishi mumkin, lekin siz unga tahlil qiladigan ma'lumotlarni berasiz, keyin esa doimiy ravishda o'zingizni eslatib, pullik variantlarni taklif qiladi va daromad keltiradi. Bunday modelning nomi nima?

- A) O'yin xuddi uchuvchi kabi (Game as a pilot)
B) O'yin xuddi xizmat kabi (Game as a service)
C) O'yin xuddi o'rganish kabi (Game as a learning)
D) O'yin xuddi maxsulot kabi (Game as a product)

10. Mashinali o'qitish - ...

- A) modelda ma'lumotlar va ular orasidagi munosabatlar ikki o'lchovli jadvallarni aloqalari ko'rinishida ifodalanadi
B) ma'lumotlar tuzilmalari va ma'lumotlarni boshqarish bo'yicha operatsiyalarni tavsiflashning qulay usuli
C) ma'lum bir til yoki ular tavsiflangan tillardan mustaqil bo'lgan mavjud tushunchalar haqidagi nazariya
D) bu ma'lumot olish va ulardan axborotlarni chiqrib olish jarayonida kompyuterning aniqligi yaxshilanadigan jarayondir

11. Relyatsion modelda ma'lumotlar va ular orasidagi munosabatlar ikki o'lchovli jadvallarni aloqalari ko'rinishida ifodalangan model bu ...

- A) RM/T model B) ORM model
C) ER model D) Funktsional model

12. Tarmoq "falaji" qachon sodir bo'lishi mumkin?

- A) vazn qiymatlari juda kichik bo'lganda
B) vazn qiymatlari juda katta bo'lganda
C) qadam hajmi juda kichik bo'lganda
D) qadam hajmi juda katta bo'lganda

13. Ixtiyoriy ma'lumotlar bazasida (massiv, to'plam, ma'lumotlar bazasi) berilgan xususiyatga ega bo'lgan bir yoki bir nechta elementlarni aniqlash usullari bu ...

- A) Ma'lumotlarni belgilash
- B) Ma'lumotlarni qidirish
- C) Ma'lumotlarni saralash
- D) Ma'lumotlarni taxlil qilish

14. Mashinani o`qitish bir qator qiyinchiliklarga ega. Kirish ma'lumotlari uchun u yoki bu doimiy raqamli qiymatni bashorat qilishga qaratilgan usulning nomi nima?

- A) Qayta o`qish
- B) Regressiya
- C) Klassifikatsiya
- D) Klasterlash

15. 1995 yilda E. Codda tamoyillari asosida nima deb nomlangan test paydo bo`ldi?

- A) CODDA
- B) BIGDATA
- C) OLTP
- D) FASMI

16. 2012 yil uchun insoniyat tomonidan to'plangan raqamli ma'lumotlar hajmi qanchaga yetgan?

- A) zettabayt
- B) petabayt
- C) ekzabaytlar
- D) yottabayt

17. Bu ma'lumotlar modeli ma'lumotlar qanday taqdim etilishini belgilaydi va ma'lumotlar bazasini yaratish uchun zarur bo'lgan barcha tafsilotlarni o'z ichiga oladi.

- A) Mantiqiy modellashtirish
- B) Modellashtirish
- C) Fizik modellashtirish
- D) Kontseptual modellashtirish

18. NoSQL ni asosiy afzalliklari berilgan javobni belgilang.

- A) Atomarlik (Atomicity),
Kelishuvchanlik (Consistency),
Ajratilganlik (Isolation),
Uzoq muddatlilik (Durability)
- B) Atomarlik (Atomicity),
Kelishuvchanlik (Consistency),
Ajratilganlik (Isolation),
Uchuvilik (Pilot)
- C) Atomarlik (Atomicity),
Quvatlilik (Powerful),
Tezlik (Speed),
Uchuvilik (Pilot)
- D) Atomarlik (Atomicity),
Quvatlilik (Powerful),
Ajratilganlik (Isolation),

Uchuvilik (Pilot)

19. Katta ma'lumotlar tendentsiyasining umidsizliklari qanday aniqlanadi?

- A) fuqarolarning shaxsiy hayotiga tahdid tufayli ma'lumotlarni yig'ish tartib-qoidalari soddalashtiriladi, bu esa katta ma'lumotlar qiymatining pasayishiga olib keladi
- B) fuqarolarning shaxsiy hayoti (maxfiyligi) xavfsizligiga tahdid tufayli ma'lumotlarni yig'ish tartib-qoidalari murakkablashadi, bu esa katta ma'lumotlar qiymatining pasayishiga olib keladi
- C) Katta ma'lumotlar sohasida bunday tushuncha mavjud emas
- D) umidsizlik deyarli mavjud emas

20. IBS kompaniyasi tahlilchilarini ma'lumotlariga ko'ra 2025 yilda ma'lumotlarning butun dunyo hajmini qanchani tashkil qilishi bashorat qilingan?

- A) 40-44 zettabayt
- B) 0,18 zettabayt (1 zb = 1024 ekzabayt)
- C) 400-500 zettabayt
- D) 6,5 zettabaytdan yuqori

21. Apache NiFi nima uchun ishlatiladi?

- A) Big Data oqimlarini yo'naltirish va ETL konveyrlarini qurish uchun
- B) Katta ma'lumotlarni samarali saqlash uchun
- C) Tahlil natijalarini vizualizatsiya qilish uchun
- D) DWH uchun SQL so'rovlarini optimallashtirish uchun

22. Katta ma'lumotlarni to'plash uchun etarli emas - ular xavfsiz joyda saqlanishi kerak. Bugungi kunda bulutli serverlar buning uchun juda yaxshi usul. Ularning qanday afzalliklari bor?

- A) Internet aloqasi mavjudligigina, ma'lumotlarga kirishni ta'minlash
- B) Ma'lumotlarni bir nechta nusxada mustaqil serverlarda saqlash orqali xatolarga chidamlilikni ta'minlash
- C) Internet aloqasi mavjudligidan qat'i nazar, ma'lumotlarga kirishni ta'minlash
- D) Ma'lumotlarni bir nechta nusxada yagona serverda saqlash orqali xatolarga chidamlilikni ta'minlash

23. IBS kompaniyasi tahlilchilarini ma'lumotlariga ko'ra 2008 yilda ma'lumotlarning butun dunyo hajmini qanchani tashkil qilar edi?

- A) 6,5 zettabaytdan yuqori
- B) 5 ekzabayt ma'lumotlar (1 EB = 1 milliard gigabayt)
- C) 40-44 zettabayt
- D) 0,18 zettabayt (1 zb = 1024 ekzabayt)

24. Tasvirlarning tasnifi, Klasterlash, Funksiyaning approksimasiyasi, Bashoratlash, Optimallashtirish, Assotsiativ xotira, Boshqaruv kabi vazifalar nima yordamida hal qilinadi?

- A) Data engineer yordamida hal qilinadi
- B) Data science yordamida hal qilinadi

- C) SNT yordamida hal qilinadi
D) Data mining yordamida hal qilinadi

25. Atrofimizdagi jismoniy va ijtimoiy muhitning ayrim jihatlarini va muloqotni tushunish uchun rasmiy ta'riflash faoliyati bu ..

- A) Fizik modellashtirish
B) Kontseptual modellashtirish
C) Modellashtirish
D) Mantiqiy modellashtirish

JAVOBLARI

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1 - variant	A	B	B	C	A	B	B	C	D	B	C	B	D	A	D	D	D	B	C	A	B	B	B	D	D
2 - variant	B	D	C	D	D	A	A	D	B	D	A	B	B	C	D	A	C	A	B	C	A	B	D	C	B

IV.GLOSSARIY

ABS	Raqamni absolyut qiymatini belgilovchi funktsiya. Misol uchun, ABS (-22) 22 ga teng funktsiya bir raqamni boshqasiga. Kattalik jihatidan qanchalik yaqin ekanligini aniqlash uchun foydalaniladi. Misol uchun, IFABS (TARGET GUESS) < 3 raqamlar bilan birdek yaxshi ishlaydi, bunda qiymatlarning farqi +1, +2, -1 yoki -2 ga teng.
A-bus	shinalar to‘plami, chiqish - kirish qurilmasi va hotirani, hamda ALU blok mikroprotsessorni ichidan bog‘lovchi hisoblanadi. A-bus shinalari asosiy hisoblanib boshqa shinalar mikroprotsessori boshqa qurilmalar bilan bog‘laydi va buffer mikroshemalariga mos keladi, bu orqali A-bus shinalari va boshqa ichki shinalar bilan bog‘laniladi.
Algoritm	bu oddiy, bu matematik formula yoki ma'lumotlarni tahlil qilish uchun ishlatiladigan statistik jarayonning bir turi. Ushbu atama va algoritmning o‘zi katta ma'lumotlarda keng qo‘llaniladi.
Analitika	ma'lumotlarni o‘rganish jarayoni. Masalan, 1960 yildan 2017 yilgacha bo‘lgan ma'lumotlarni o‘laylik, qaysi mamlakat terma jamoasi xokkey bo‘yicha jahon chempionatida g‘olib chiqqan. Keling, ushbu 57 qatordan g‘olib jamoalar ro‘yxatini tuzamiz va jamoaning necha marta chempion bo‘lish foizini hisoblaymiz. Asl ma'lumotlarga asoslangan ma'lumotlar to‘plamidan boshqa natijalarni oladigan bo‘lsak, bu jarayon analitik deb ataladi.
Bashoratli tahlil	Bashoratli tahlil - masalan, siz bir necha yil davomida turli bo‘limlar bo‘yicha chop etish hajmlari haqidagi ma'lumotlarni tahlil qildingiz. Endi siz ushbu tahlildan printerlar va MFPlar uchun yangi kartridlar sotib olish nuqtai nazaridan kelgusi yil uchun byudjetingizni rejalashtirish uchun foydalanishingiz mumkin. Bu bashoratli tahlillar, keyingi qiymatlarni bashorat qilish uchun ishlatiladigan tahlillardir.
BABT	(British Approval Board for Telecommunications) –Buyuk Britaniyadagi telefon liniyalariga qurilmalar ulashga ruxsat beruvchi tashkilot. Bunday ruxsatnomaga ega bo‘lgan qurilmalarga maxsus yashil “stiker” yopishtiriladi.
BACS	(Bankers Automated Clearance System) kompyuter aloqalari qo‘llaniladigan banklararo o‘zaro kelishuv metodi.

BAK	(BAK file) – fayl bo‘lib, u oldingi fayl talqinining aynan o‘sha nomi bilan nomlanib, lekin bir oz kengaytirilgan ko‘rinishda. Misol uchun, CHAP1. BAK fayli CHAP 1.TXT faylining zahiradagi nusxasi hisoblanadi. Ba’zi bir dasturlarda zaxiradagi nusxalar uchun kengaytirish sifatida BAK o‘rniga \$\$\$ simvoli ishlatiladi.
Big Blue	IBM ning sleng nomlanishi bo‘lib, IBM firma markasining tashqi ko‘rinishidan kelib chiqadi.
Big Data	“Katta ma'lumotlar” umumiy atamasi ma'lumotlarni qayta ishlashning an'anaviy vositalari (masalan, RDBMS - relyatsion ma'lumotlar bazasini boshqarish tizimlari) tomonidan qayta ishlanishi uchun etarlicha katta va murakkab bo‘lgan har qanday ma'lumotlar to‘plami hisoblanadi.
DBMS	(database management system) – ma'lumotlar bazasini amalda boshqaruvchi tizim, dasturlar to‘plami bo‘lib, u ma'lumotlar bazasini tuzishga va unda ishlashga imkoniyat beradi.
Data mining	ma'lum namunalar asosida ma'lumotlarni tahlil qilish va katta ma'lumotlar to‘plamlarida ular uchun moslikni topish jarayonidir. Bu atama "Analitika" tushunchasi bilan chambarchas bog'liq bo‘lib, ma'lumotlar olimlari ma'lumotlar o‘rtasidagi munosabatlarni qidirish uchun matematik va statistik algoritmlardan foydalanadilar. Bunday holda, qidiruv asosan ma'lumotlar orasidagi yashirin aloqalar yo‘nalishi bo‘yicha olib boriladi.
Data Scientist	zamonaviy dunyodagi eng ko‘p ishlatiladigan mutaxassisliklardan biri. Bu katta ma'lumotlarni tushunadigan va qayta ishlaydigan mutaxassisdir. Ma'lumotlar modellarini yaratadi, ular o‘rtasidagi munosabatlarni belgilaydi, yashirin bog'liqliklarni qidiradi.
Data Lake	korxona miqyosidagi ma'lumotlar omborlari, ma'lumotlar tuzilmagan (xom) shaklda saqlanadi. Dastlabki ishlov berish va tozalashdan o‘tgan ma'lumotlar qat'iy tuzilgan shaklda saqlanadi. Ma'lumotlar ko‘llari - bu ma'lumotlar umumiy yig'maga tashlanadigan yoki ma'lum qismlarga parchalanadigan, lekin hali ham asl shaklida saqlanadigan omborlar. Bunday ma'lumotlarga kirish uchun ularning tuzilishi va ularni qanday qayta ishlash kerakligini aniq tushunish kerak.
DCE	(data communications equipment) – ma'lumotlar uzatuvchi qurilma.

CAD	(computer-aided design) – avtomat loyihalashtirish tizimi.
CAD/CAM	(computer-aided design and computer-aided manufacture) – avtomat loyihalashtirish va ishlab chiqarish tizimi; mahsulot ishlab chiqarish va qayta yo‘llash uchun dasturlar to‘plami.
CAI	(computer-aided instruction) – o‘qituvchilar uchun dasturlar to‘plami undan darsga tayyorgarlikda, yoki testlar tuzishda va boshqa zaruriy yordamchi materiallar tayyorlashda faoydalanishlari mumkin.
CAL	(computer-assisted learning) – kompyuter yordamida o‘qitish; bunda kompyuter bilimni tekshirishga imkoniyat beradi va ma’lumotlarni displeyga chiqarib beradi; kelgusi taraqqiyot o‘quv kompyuterlariga ovozni taniydigan va nutqni sintez qiladigan qurilma taqdim etishi kutilmoqda, shu bilan klaviaturaga bo‘lgan ehtiyoj yo‘qolmoqda.
CON	(console) kompyuter tizimi monitorini va klaviaturani identifikatsiya qilishda foydalaniladigan so‘z.
COM1	AUX 1 ga qarang.
COM	MS-DOS operatsion tizimi uchun buyruq protsessori; MS-DOS va CP/M da ham foydalaniladigan fayllarni kengaytirish. Bu kengaytirish bilan dasturlar 64 k va undan kam bo‘lgan hotiraga kiritiladi va yuklanadi.
ECHO	MS-DOS buyruq faylida foydalaniladigan buyruq ECHO OFF buyruq faylining buyruqlari ekranga chiqarilmasligi uchun foydalaniladi. Bu o‘rinda ECHO OFF buyrug‘i o‘zi chiqariladi. Uning o‘zi ham chiqarilmasligi uchun keyinchalik MS-DOS talqinida ECHO OFF buyrug‘i qo‘llaniladi.
EDP	(elektronik data processing) – kompyuter yordamida ma’lumotlarni ishlash. Bunday ishlov ma’lumotlarni kiritishdan, ularni saqlashdan, ular ustidan ma’lum bir dastur harakatini amalga oshirish va ekranga chiqarishdan iborat.

<i>EEMS</i>	(extended expanded memory system) – yordamchi va kengaytirilgan xotiradan foydalaniladigan xotira EVM kompyuterlarida va ularga, mos keluvchi mashinalarda qo'llanilgan standart xotira. Kengaytirilgan xotira yordamida ko'pgina dasturlarni kiritishni amalga oshirish mumkin.
<i>FDISK</i>	qattiq diskni tahminiy formatlashda foydalanadigan va MS-DOS uning bir bo'lagi hisoblanuvchi dastur.
<i>FET</i>	(field-effect transistor) – hoshiyali tranzistor. FF – betni o'tkazish (u betdan bu betga o'tish).
<i>FIFO</i>	(first-in, first-out) – ma'lumotlar kiritilishining tartibi. "birinchi kirgan –birinchi chiqadi".
<i>GOTO</i>	"belgi" —o'tish buyrug'i do'lib, u deyarli barcha dasturlash tillarida foydalaniladi. "Belgi" qaerga o'tish kerakligini ko'rsatadi. Beysikda GOTO qator raqami bilan qo'llaniladi. To'g'ri yo'l buyrug'i tez – tez qo'llash dasturni qiyin o'qiladigan ba ishlamaydigan qilib qo'yadi.
<i>grep</i>	UNIX operatsion tizimi buyrug'i bo'lib, ba'zi bir fayllarda yoki barcha fayllarda ketma – ket kiritilgan harflarni qidirishni amalga oshiradi. MS-DOS da uning ekvivalenti FIND buyrug'i hisoblanadi, lekin bu buyruqning imkoniyatlari kam.
<i>GUI</i>	(graphical user interface) — foydalanuvchining grafik interfeysi; WIMP tizimining boshqa nomi.
<i>gulp</i>	Bir butun deb ishlanadigan, ba'zan baytlar guruhini ifodalash uchun foydalanadigan termin.
<i>GW-BASIC</i>	Beysik interpretatorining turi, MS-DOS bilan shaxsiy kompyuterlar uchun taqdim etiladi. GW-BASIC MS-DOS 5.0 va bir muncha keyingi talqini uchun yanada yahshi tuzilgan Beysir interpretatoriga o'z o'rnini bo'shatib berdi.
<i>IDE</i>	(integrated device electronics) — AT ko'rinishidagi shaxsiy kompyuterlar diskovodini nazorat qiluvchi qurilma. Bu nazoratchiga tizimli blok bilan aloqa qilish uchun juda katta bo'lmagan interfeys talab qilinadi.
<i>IEC FITTING</i>	To'g'riburchakli shtirdan foydalanuvchi tarmoq ulanish qurilmasi. Bu ulanish qurilmasi javob beruvchi qismga ulanganda tok keluvchi elementlarni to'liq yopadi.Xuddi shunday ulanish qurilmasi blok tizimidan monitorga quvvat manbaini o'tkazish uchun qo'llaniladi.

IEE	(Institution of Electrical Engineers) — Elektrotexnika bo'yicha injenerlar instituti, u Velikobritaniyada elektrotexnika va elektronika bo'yicha injenerlar faoliyatini nazorat qilish va rasmiylashtirish funksiyalarini amalga oshiradi; AQSH da IEEE — elektrotexnika va elektronika bo'yicha injenerlar instituti faoliyat yuritadi.
IEEE-488	IEEE (AQSH elektrotexnika va elektronika bo'yicha injenerlar instituti) tomonidan yaratilgan, standart ketma – ket interfeys. Undan faqat ma'lumotlarni uzatish va aloqani tasdiqlashda foydalaniladi. Tizim birinchi navbatda kompyuterlarni o'lchash asboblari ga ulash uchun foydalaniladi. SHunga qaramay ko'pgina mikrokompyuterlar bu interfeys yordamida kompyuterning qo'shimcha qurilmalariga ulanishi mumkin bo'ladi, agarda ularning kirishi va chiqishi IEEE-488 ga javob bersa.
IF	Dasturlash tillarida harf kiritish uchun qo'llaniladigan so'z. IF ni qo'llaydigan test ikki natijaga ega bo'lishi mumkin — HAQIQAT yoki YOLG'ON, va barcha dastur ikkala holatda ham ma'lum bir harakat amalga oshirilishi mumkin bo'lgan holda ishlab chiqilishi kerak.
POP	(hamda PULL) — tarmoqdan qandaydir ma'lumotni olish bo'yicha yoki baytlarni tarmoqdan registr yo registrlarga yoki baytlarni joylashtirish bo'yicha Assembler tili buyrug'i. Masalan, Z-80 uchun POP AF buyrug'i ifodalanadi, stekdan ikki bayt olinishi va A va F registrlariga joylashtirilishi kerak.
POST	1. Va'lumotlar fayliga ma'lumotlarni yozib kiritish. 2. (Power-on self test) — harakat, ko'pgina zamonaviy kompyuter tomonidan amalga oshiriladi, u orqali xotira holati tekshiriladi va operatsion tizimga yuklatishdan oldin mikrosxema harakati.
soft	(Hisoblash tizimi qismi haqida) – oson o'zgarilishga xos; masalan, termin xotiradagi dasturga taalluqli bo'lishi mumkin.
sort	Utilita, MS-DOS uchun utilitlar to'plami uchun qism hisoblanadi; fayllarni alfavit yoki sonli tartibda joylashtirish imkonini beradi.
SOS	Xususan, harbiy ehtiyojlar uchun integral sxemasini juda qattiq talablar bilan yaratish metodi.
type	Ekrandagi faylni bosib chiqarishni amalga oshiruvchi MS-DOS buyrug'i. Bu dastur faqat matnli fayllar uchun foydalanishi kerak va bu fayllarni chiqarish ekran bo'yicha matnni aylantirish bilan vujudga keladi. Qulay bo'lishi uchun bu buyruqni yaxshisi TYPE/MORE ko'rinishida foydalanish

Volume	kerak , unda xulosa betma-bet hosil bo‘ladi va boshqa betga o‘tish klavishni bosish bilan amalga oshadi Ma`lomitlar hajmi to‘plamda qancha ma`lumot mavjudligi.
Velocity	Tezlik – yangi ma`lumotlar qanday tezlikda generatsiya bo‘lishi.
Variety	Xilma-xillik – ma`lumotlarni tipi bir-biridan qanchaga farq qilishi.

VII.FOYDALANILGAN ADABIYOTLAR RO‘YXATI

I. O‘zbekiston Respublikasi Prezidentining asarlari

1. Мирзиёев Шавкат Миромонович. Қонун устуворлиги ва инсон манфаатларини таъминлаш – юрт тараққиёти ва халқ фаровонлигининг гарови. Ўзбекистон Республикаси Конституцияси қабул қилинганининг 24 йиллигига бағишланган тантанали маросимдаги маъруза. 2016 йил 7 декабр /Ш.М.Мирзиёев. – Тошкент: “Ўзбекистон”, 2017.
2. Мирзиёев Шавкат Миромонович. Эркин ва фаровон, демократик Ўзбекистон давлатини биргаликда барпо этамиз. Ўзбекистон Республикаси Президенти лавозимига киришиш тантанали маросимига бағишланган Олий Мажлис палаталарининг қўшма мажлисидаги нутқ / Ш.М. Мирзиёев. – Тошкент: Ўзбекистон, 2016.

II. Normativ-huquqiy hujjatlar

1. O‘zbekiston Respublikasining Konstitutsiyasi. – T.: O‘zbekiston, 2018.
2. O‘zbekiston Respublikasining 2020 yil 23 sentabrda qabul qilingan “Ta’lim to‘g‘risida”gi O‘RQ-637-sonli Qonuni.
3. O‘zbekiston Respublikasi Prezidentining 2022 yil 22 yanvar “2022 - 2026 yillarga mo‘ljallangan Yangi O‘zbekistonning taraqqiyot strategiyasi to‘g‘risida”gi PF-60-son Farmoni.
4. O‘zbekiston Respublikasi Prezidentining 2017 yil 26 sentabr“Oliy ta’lim muassasalariga kirish uchun nomzodlarni maksadli tayyorlash tizimini yanada takomillashtirish to‘g‘risida”gi PQ-3290-sonli Qarori..
5. O‘zbekiston Respublikasi Prezidentining 2020 yil 3 dekabrda “Iqtidorli yoshlarni saralab olish tizimi va akademik litseylar faoliyatini takomillashtirish chora-tadbirlari to‘g‘risida” PQ-4910-son Qarori.
6. O‘zbekiston Respublikasi Vazirlar Mahkamasining 2022 yil 1 iyundagi “Akademik litseylar rahbar va pedagog xodimlarining uzluksiz malakasini oshirish tizimini joriy etish to‘g‘risida” 296-son Qarori.

III. ASOSIY ADABIYOTLAR

1. Дэви Силен, Арно Мейсман, Мохамед Али. Основы Data Science и Big Data. Python и наука о данных. СПб.: Питер, 2017. -336 с.(Серия «Библиотека программиста»).
2. Хенрик, Ричардс Джозеф, Феверолф Марк, Машинное обучение, СПб.: Питер, 2017. -336 с.(Серия «Библиотека программиста»).
3. В. Е. Туманов. Проектирование хранилищ данных для приложений систем деловой осведомленности (Business Intelligence System). изд. ИНТУИТ. 2016г.
4. П.Флах. Наука и искусство построения алгоритмов, которую извлекают знания из данных. Издание Cambridge University Press, 2012 г.
5. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И., Методы и модели анализа данных: OLAP и Data Mining. СПб.: БХВ-Петербург, 2004. - 336 с.
6. Елманова Н., Федоров А. Введение в OLAP-технологии Microsoft. СПб.: БХВ-Петербург, 2014.-232 с.

QO‘SHIMCHA ADABIYOTLAR

7. Эрик Томпсон. OLAP решения: Разработка многомерных информационных систем. изд. Уилли. 2002 г.
8. Джеми Макленнан, Жаохуи Танг, Богдан Криват. Интеллектуальный анализ данных с Microsoft SQL Server 2008. изд. Уилли. 2008 г. - 636р
9. Джордж Споффорд, Сивакумар Харинат, Кристофер Вебб, Дилан Хаи Хуанг, Франческо Сиварди. Решения MDX на Microsoft SQL Server Analysis Services. изд. Уилли. 2006 г. — 716 стр
- 10.Фрэнкс Б. Революция в аналитике. Как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики / Б. Фрэнкс. (М.: Альпина Диджитал, 2014. (370 с.
- 11.Моррисон А. Большие Данные: как извлечь из них информацию. Технологический прогноз / Ежеквартальный журнал. – 2010. – №3. – С. 22–29.
- 12.Михнев И.П. Информационная безопасность в современном экономическом образовании // Международный журнал прикладных и фундаментальных исследований. – 2013. – №4. – С. 111–113.
- 13.Михнев И.П. Обучение и контроль знаний студентов с помощью UniTest // Фундаментальные исследования. – 2008. – №1. – С. 94–95.
- 14.Михнев И.П. Информационная безопасность на просторах мобильного интернета // Образовательные ресурсы и технологии. – 2015. – №4(12). – С. 66–70.

15. Черняк Л. Большие Данные – новая теория и практика // Открытые системы. СУБД. – 2011. – №10. – С. 36–41.
16. Что такое Big data: собрали всё самое важное о больших данных // RUSBASE [Электронный ресурс]. – Режим доступа: <https://rb.ru/howto/что-такое-big-data/> (дата обращения 11.03.2018).
17. Big Data: проблема, технология, рынок [Электронный ресурс]. – Режим доступа: <http://compress.ru/Article.aspx?id=22725> (дата обращения: 10.03.2018).

ELEKTRON TA'LIM RESURSLARI:

18. www.gov.uz
19. www.lex.uz
20. www.ziyonet.uz
21. www.edu.uz
22. www.tdpu.uz
23. <https://ru.wikipedia.org>
24. Серьёзно о технологиях для Больших Данных | Открытые системы. СУБД | Издательство «Открытые системы» osp.ru
25. Свежий взгляд на Большие Данные | Открытые системы. СУБД | Издательство «Открытые системы» osp.ru