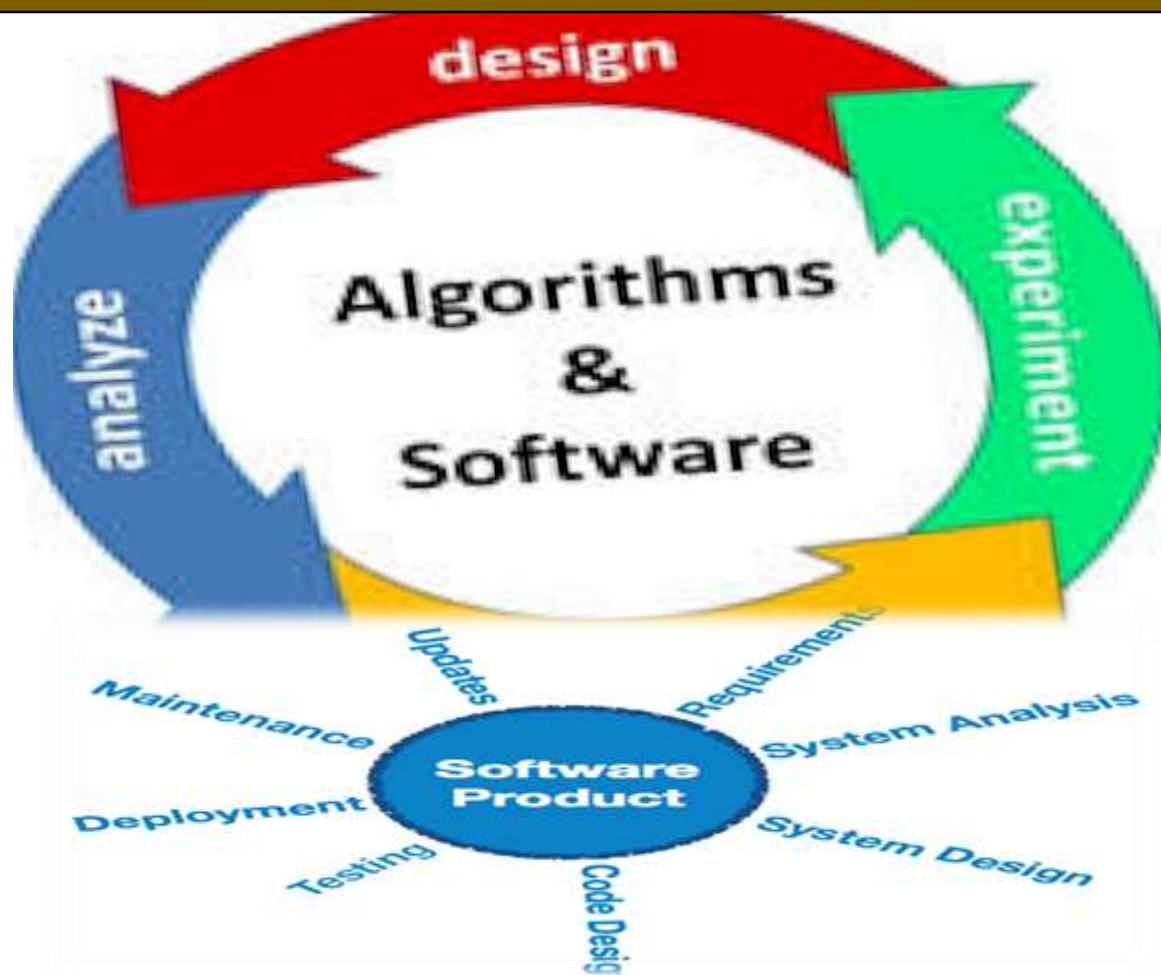


ЎЗБЕКИСТОН РЕСПУБЛИКАСИ
ОЛИЙ ВА ЎРТА МАХСУС ТАЪЛИМ ВАЗИРЛИГИ

МУҲАММАД АЛ-ХОРАЗМИЙ НОМИДАГИ ТОШКЕНТ АХБОРОТ
ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ ҲУЗУРИДАГИ ПЕДАГОГ КАДРЛАРНИ
ҚАЙТА ТАЙЁРЛАШ ВА УЛАРНИНГ МАЛАКАСИНИ ОШИРИШ
ТАРМОҚ МАРКАЗИ



ЎҚУВ – УСЛУБИЙ МАЖМУА

КАТТА МАЪЛУМОТЛАРНИ
ҚАЙТА ИШЛАШ УСУЛ ВА ВОСИТАЛАРИ

“Дастурий инжиниринг” йўналиши

**ЎЗБЕКИСТОН РЕСПУБЛИКАСИ
ОЛИЙ ВА ЎРТА МАХСУС ТАЪЛИМ ВАЗИРЛИГИ**

**ОЛИЙ ТАЪЛИМ ТИЗИМИ ПЕДАГОГ ВА РАЎБАР КАДРЛАРИНИ
ҚАЙТА ТАЙЁРЛАШ ВА УЛАРНИНГ МАЛАКАСИНИ ОШИРИШНИ
ТАШКИЛ ЭТИШ БОШ ИЛМИЙ - МЕТОДИК МАРКАЗИ**

**МУҲАММАД АЛ-ХОРАЗМИЙ НОМИДАГИ ТОШКЕНТ АХБОРОТ
ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ ҲУЗУРИДАГИ ПЕДАГОГ
КАДРЛАРНИ ҚАЙТА ТАЙЁРЛАШ ВА УЛАРНИНГ МАЛАКАСИНИ
ОШИРИШ ТАРМОҚ МАРКАЗИ**

“Дастурий инжиниринг” йўналиши

**“Катта маълумотларни қайта ишлаш усул ва
воситалари»
МОДУЛИ БЎЙИЧА**

Ў Қ У В – У С Л У Б И Й М А Ж М У А

Тошкент – 2022

**Мазкур ўқув-услубий мажмуа Олий ва ўрта махсус таълим
вазирлигининг 2020 йил 7 декабрдаги 648-сонли буйруғи билан
тасдиқланган ўқув режа ва дастур асосида тайёрланди.**

Тузувчилар: Муҳаммад ал-Хоразмий номидаги ТАТУ ҳузуридаги
ПКҚТваУМО тармоқ маркази директори, п.ф.д., проф.
Ф.Закирова,
Муҳаммад ал-Хоразмий номидаги ТАТУ, “Ахборот
технологиялари” кафедраси катта ўқитувчиси И.Юсупов.

Такризчилар: Беларусь-Ўзбекистон кўшма тармоқлараро амалий
техник квалификациялар институти, илмий ишлар ва
инновациялар бўйича директор ўринбосари в.б., доц.
Л.Набиулина,
Муҳаммад ал-Хоразмий номидаги ТАТУ, “Ахборот
технологиялари” кафедраси мудири, проф.
Х.Зайнидинов.

**Ўқув -услубий мажмуа Муҳаммад ал-Хоразмий номидаги Тошкент
ахборот технологиялари университети Кенгашининг қарори билан
нашрга тавсия қилинган (2020 йил 29 октябрдаги 3(705)-сонли
баённомаси)**

МУНДАРИЖА

I. Ишчи дастур	5
II. Модулни ўқитишда фойдаланиладиган интерфаол методлар	10
III. Назарий материаллар.....	17
IV. Амалий машғулот материаллари.....	47
V. Кейслар банки.....	80
VI. Глоссарий	84
VII. Адабиётлар рўйхати.....	90

І БЎЛІМ

ИШЧИ ДАСТУР

I. ИШЧИ ДАСТУР

Кириш

Дастур Ўзбекистон Республикасининг 2020 йил 23 сентябрда тасдиқланган “Таълим тўғрисида”ги Қонуни, Ўзбекистон Республикаси Президентининг 2017 йил 7 февралдаги “Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида”ги ПФ-4947-сон, 2019 йил 27 августдаги “Олий таълим муассасалари раҳбар ва педагог кадрларининг узлуксиз малакасини ошириш тизимини жорий этиш тўғрисида”ги ПФ-5789-сон, 2019 йил 8 октябрдаги “Ўзбекистон Республикаси олий таълим тизимини 2030 йилгача ривожлантириш концепциясини тасдиқлаш тўғрисида”ги ПФ-5847-сон ва 2020 йил 29 октябрдаги “Илм-фанни 2030 йилгача ривожлантириш концепциясини тасдиқлаш тўғрисида”ги ПФ-6097-сонли Фармонлари ҳамда Ўзбекистон Республикаси Вазирлар Маҳкамасининг 2019 йил 23 сентябрдаги “Олий таълим муассасалари раҳбар ва педагог кадрларининг малакасини ошириш тизимини янада такомиллаштириш бўйича қўшимча чора-тадбирлар тўғрисида”ги 797-сонли Қарорида белгиланган устувор вазифалар мазмунидан келиб чиққан ҳолда тузилган бўлиб, у олий таълим муассасалари педагог кадрларининг касб маҳорати ҳамда инновацион компетентлигини ривожлантириш, соҳага оид илғор хорижий тажрибалар, янги билим ва малакаларни ўзлаштириш, шунингдек амалиётга жорий этиш кўникмаларини такомиллаштиришни мақсад қилади.

Қайта тайёрлаш ва малака ошириш йўналишининг ўзига хос хусусиятлари ҳамда долзарб масалаларидан келиб чиққан ҳолда дастурда тингловчиларнинг мутахассислик фанлар доирасидаги билим, кўникма, малака ҳамда компетенцияларига қўйиладиган талаблар такомиллаштирилиши мумкин.

Модулнинг мақсади ва вазифалари

“Катта маълумотларни қайта ишлаш усул ва воситалар” модулининг мақсади: катта маълумотлар билан ишлаш технологиялари бўйича олий таълим муассасалари педагог кадрларининг касбий компетентлигини ошириш.

Модулнинг вазифалари: олий таълим муассасалари педагог кадрларида катта маълумотлар билан ишлаш технологиялари ҳақида назарий ва амалий билимларни, кўникма ва малакаларни шакллантиришдан иборат.

Модул бўйича тингловчиларнинг билими, кўникмаси, малакаси ва компетенцияларига қўйиладиган талаблар

“Катта маълумотларни қайта ишлаш усул ва воситалар” модулини ўзлаштириш жараёнида амалга ошириладиган масалалар доирасида:

Тингловчи:

- катта маълумотлар (Big Data)нинг хусусиятлари, катта маълумотларни қайта ишлаш усуллари, Hadoop тизими имкониятларини *билиши* керак.
- катта маълумотлар сўровларнинг Hive тизими орқали амалга ошириш, SQL тилига кириш, SQL дан HiveQLга, HIVEe дан HIVEQL га ўтиш; Hadoop

тизимидан файлларни сўраш учун HIVE дан фойдаланиш; Hadoop тизимида маълумотларни ўтказиш, Hadoop тизимида файлларни қайта ишлаш, Hadoop тизимидан маълумотларни олиш **кўникмаларига** эга бўлиши лозим.

- катта маълумотлар билан Hadoop тизимида ишлаш **малакаларига** эга бўлиши лозим.

- катта маълумотлар асосида ахборот тизимларини бошқариш **компетенцияларига** эга бўлиши лозим.

Модулни ташкил этиш ва ўтказиш бўйича тавсиялар

“Катта маълумотларни қайта ишлаш усул ва воситалар” модули маъруза ва амалий машғулотлар шаклида олиб борилади.

Модулни ўқитиш жараёнида таълимнинг замонавий методлари, педагогик технологиялар ва ахборот-коммуникация технологиялари қўлланилиши назарда тутилган:

- маъруза дарсларида замонавий компьютер технологиялари ёрдамида презентацион ва электрон-дидактик технологиялардан;

- ўтказиладиган амалий машғулотларда техник воситалардан, экспресс-сўровлар, тест сўровлари, ақлий ҳужум, гуруҳли фикрлаш, кичик гуруҳлар билан ишлаш, коллоквиум ўтказиш, ва бошқа интерактив таълим усуллари қўллаш назарда тутилади.

Модулнинг ўқув режадаги бошқа модуллар билан боғлиқлиги ва узвийлиги

“Катта маълумотларни қайта ишлаш усул ва воситалар” модули мазмуни ўқув режадаги “Компьютер инжиниринги”, “Булутли ҳисоблаш технологиялари“, “Компьютер тармоқлари хавфсизлиги” ўқув модуллари билан узвий боғланган ҳолда педагогларнинг таълим жараёнида булутли ҳисоблаш, катта маълумотлар ва виртуал реаллик тизимларидан фойдаланиш бўйича касбий педагогик тайёргарлик даражасини оширишга хизмат қилади.

Модулнинг олий таълимдаги ўрни

Модулни ўзлаштириш орқали тингловчилар электрон ҳукуматни жорий этишни ўрганиш, амалда қўллаш ва баҳолашга доир касбий компетентликка эга бўладилар.

Модул бўйича соатлар тақсимоги

№	Модуль мавзулари	Аудитория укув юкламаси			
		Жами	жумладан		
			Назарий	Амаий машғулот	Кўчма машғулот
1.	Катта маълумотлар (Big Data) тушунчаси	2	2		
2.	Катта маълумотларнинг хусусиятлари	2	2		
3	Катта маълумотларни қайта ишлаш усуллари	2	2		
4	Катта маълумотлар сўровларнинг Hive тизими орқали амалга ошириш. SQL тилига кириш, SQL дан HiveQLга, HIVEe дан HIVEQL га ўтиш.	4		4	
5	HIVE e HIVEQL га кириш, Hadoop файлларни сураш учун Hive-дан фойдаланиш	2		2	
6	Hadoop тизимига маълумотларни ўтказиш. Hadoop тизимида файлларни қайта ишлаш. Hadoop тизимидан маълумотларни олиш.	4		4	
	Жами:	16	6	10	

НАЗАРИЙ МАШҒУЛОТЛАР МАЗМУНИ

1-маъруза: Катта маълумотлар (Big Data) тушунчаси (2 соат)

Big data га кириш. Катта маълумотлар тушунчаси, таърифлари. Big data: асосий вазифалар.

2-маъруза. Катта маълумотларнинг хусусиятлари (2 соат)

Катта маълумотларнинг афзалликлари. Таҳлил қилинадиган маълумотларнинг ҳажми, анъанавий маълумотлар таҳлил тизимлари ошиб кетганда, катта маълумотларни ҳисоблаш тизимини жорий етиш.

3-маъруза. Катта маълумотларни қайта ишлаш усуллари (2 соат).

Google File System (асосий характеристикалари, ишлаш тамойиллари). Hadoop тақсимланган файл тизими (асосий характеристикалари, ишлаш тамойиллари). MapReduce (асосий характеристикалари, ишлаш тамойиллари).

АМАЛИЙ МАШҒУЛОТЛАР МАЗМУНИ

1-амалий иш. Катта маълумотлар сўровларнинг Hive тизими орқали амалга ошириш. SQL тилига кириш, SQL дан HiveQLга, HIVEe дан HIVEQL га ўтиш. (4 соат)

2-амалий иш. HIVE e HIVEQL га кириш, Hadoop файллари сураш учун Hive-дан фойдаланиш (2 соат)

3-амалий иш. Hadoop тизимида маълумотларни ўтказиш. Hadoop тизимида файллари қайта ишлаш. Hadoop тизимидан маълумотларни олиш. (4 соат)

ЎҚИТИШ ШАКЛЛАРИ

Мазкур модул бўйича қуйидаги ўқитиш шаклларида фойдаланилади:

- маърузалар, амалий машғулотлар (маълумотлар ва технологияларни англаб олиш, ақлий қизиқишни ривожлантириш, назарий билимларни мустаҳкамлаш);
- давра суҳбатлари (қўрилаётган лойиҳа ечимлари бўйича таклиф бериш қобилиятини ошириш, эшитиш, идрок қилиш ва мантиқий хулосалар чиқариш);
- баҳс ва мунозаралар (лойиҳалар ечими бўйича далиллар ва асосли аргументларни тақдим қилиш, эшитиш ва муаммолар ечимини топиш қобилиятини ривожлантириш).

II БЎЛИМ

МОДУЛНИ ЎҚИТИШДА
ФОЙДАЛАНИЛАДИГАН
ИНТЕРФАОЛ ТАЪЛИМ
МЕТОДЛАРИ

II. МОДУЛНИ ЎҚИТИШДА ФОЙДАЛАНИЛАДИГАН ИНТЕРФАОЛ ТАЪЛИМ МЕТОДЛАРИ

«Блум кубиги» методи

Методнинг мақсади: Мазкур метод тингловчиларда янги ахборотлар тизимини қабул қилиш ва билимларни ўзлаштирилишини енгиллаштириш мақсадида қўлланилади, шунингдек, бу метод тингловчилар учун “Очиқ” саволлар тузиш ва уларга жавоб топиш машқи вазифасини белгилайди.

Методни амалга ошириш тартиби:

1. Ушбу методни қўллаш учун, оддий куб керак бўлади. Кубнинг ҳар бир томонида кўйидаги сўзлар ёзилади:
 - **Санаб беринг, таъриф беринг (оддий савол)**
 - **Нима учун (сабаб-оқибатни аниқлаштирувчи савол)**
 - **Тушинтириб беринг (муаммони ҳар томонлама қараш саволи)**
 - **Таклиф беринг (амалиёт билан боғлиқ савол)**
 - **Мисол келтиринг (ижодкорликни ривожлантирувчи савол)**
 - **Фикр беринг (таҳлил қилиш ва баҳолаш саволи)**
2. Ўқитувчи мавзуни белгилаб беради.
3. Ўқитувчи кубикни столга ташлайди. Қайси сўз чиқса, унга тегишли саволни беради.

“KWLH” методи

Методнинг мақсади: Мазкур метод тингловчиларда янги ахборотлар тизимини қабул қилиш ва билимларни тизимлаштириш мақсадида қўлланилади, шунингдек, бу метод тингловчилар учун мавзу бўйича кўйидаги жадвалда берилган саволларга жавоб топиш машқи вазифасини белгилайди.

Изоҳ. KWLH:

Know – нималарни биламан?

Want – нимани билишни хоҳлайман?

How - қандай билиб олсам бўлади?

Learn - нимани ўрганиб олдим?.

“KWL” методи	
1. Нималарни биламан: -	2. Нималарни билишни хоҳлайман, нималарни билишим керак: -
3. Қандай қилиб билиб ва топиб оламан: -	4. Нималарни билиб олдим: -

“5W1H” методи

Методнинг мақсади: Мазкур метод тингловчиларда янги ахборотлар тизимини қабул қилиш ва билимларни тизимлаштириш мақсадида қўлланилади, шунингдек, бу метод тингловчилар учун мавзу бўйича қўйидаги жадвалда берилган олти саволларга жавоб топиш машқи вазифасини белгилайди.

What?	Нима? (таърифи, мазмуни, нима учун ишлатилади)	
Where?	Қаерда (жойлашган, қаердан олиш мукин)?	
What kind?	Қандай? (параметрлари, турлари мавжуд)	
When?	Қачон? (ишлатилади)	
Why?	Нима учун? (ишлатилади)	
How?	Қандай қилиб? (яратилади, сақланади, тўлдирилади, таҳрирлаш мумкин)	

“SWOT-таҳлил” методи.

Методнинг мақсади: мавжуд назарий билимлар ва амалий тажрибаларни таҳлил қилиш, таққослаш орқали муаммони ҳал этиш йўллари топишга, билимларни мустаҳкамлаш, такрорлаш, баҳолашга, мустақил, танқидий фикрлашни, ностандарт тафаккурни шакллантиришга хизмат қилади.

S – (strength)	• кучли томонлари
W – (weakness)	• заиф, кучсиз томонлари
O – (opportunity)	• имкониятлари
T – (threat)	• хавфлар

“БЕЕР” методи

Методнинг мақсади: Бу метод мураккаб, кўптармоқли, мумкин қадар, муаммоли характеридаги мавзуларни ўрганишга қаратилган. Методнинг моҳияти шундан иборатки, бунда мавзунинг турли тармоқлари бўйича бир хил ахборот берилди ва айна пайтда, уларнинг ҳар бири алоҳида аспектларда муҳокама этилади. Масалан, муаммо ижобий ва салбий томонлари, афзаллик, фазилат ва камчиликлари, фойда ва зарарлари бўйича ўрганилади. Бу интерфаол метод танқидий, таҳлилий, аниқ мантиқий фикрлашни муваффақиятли ривожлантиришга ҳамда ўқувчиларнинг мустақил ғоялари, фикрларини ёзма ва оғзаки шаклда тизимли баён этиш, ҳимоя қилишга имконият яратади. “Беер” методидан маъруза машғулотида индивидуал ва жуфтликлардаги иш шаклида, амалий ва семинар машғулотида кичик гуруҳлардаги иш шаклида мавзу юзасидан билимларни мустаҳкамлаш, таҳлил қилиш ва таққослаш мақсадида фойдаланиш мумкин.

Методни амалга ошириш тартиби:



тренер-ўқитувчи иштирокчиларни 5-6 кишидан иборат кичик гуруҳларга ажратади;



тренинг мақсади, шартлари ва тартиби билан иштирокчиларни таништиргач, ҳар бир гуруҳга умумий муаммони таҳлил қилиниши зарур бўлган қисмлари туширилган тарқатма материалларни тарқатади;



ҳар бир гуруҳ ўзига берилган муаммони атрофлича таҳлил қилиб, ўз мулоҳазаларини тавсия этилаётган схема бўйича тарқатмага ёзма баён қилади;



навбатдаги босқичда барча гуруҳлар ўз тақдимотларини ўтказадилар. Шундан сўнг, тренер томонидан таҳлиллар умумлаштирилади, зарурий ахборотлар билан тўлдирилади ва мавзу яқунланади.

Муаммоли савол					
1-усул		2-усул		3-усул	
афзаллиги	камчилиги	афзаллиги	камчилиги	афзаллиги	камчилиги
Хулоса:					

“Кейс-стади” методи

«Кейс-стади» - инглизча сўз бўлиб, («case» – аниқ вазият, ҳодиса, «stady» – ўрганмоқ, таҳлил қилмоқ) аниқ вазиятларни ўрганиш, таҳлил қилиш асосида ўқитишни амалга оширишга қаратилган метод ҳисобланади. Мазкур метод дастлаб 1921 йил Гарвард университетиде амалий вазиятлардан иқтисодий бошқарув фанларини ўрганишда фойдаланиш тартибида қўлланилган. Кейсда очик ахборотлардан ёки аниқ воқеа-ҳодисадан вазият сифатида таҳлил учун фойдаланиш мумкин.

“Кейс методи” ни амалга ошириш босқичлари

Иш босқичлари	Фаолият шакли ва мазмуни
1-босқич: Кейс ва унинг ахборот таъминоти билан таништириш	<ul style="list-style-type: none"> ✓ якка тартибдаги аудио-визуал иш; ✓ кейс билан танишиш(матнли, аудио ёки медиа шаклда); ✓ ахборотни умумлаштириш; ✓ ахборот таҳлили; ✓ муаммоларни аниқлаш
2-босқич: Кейсни аниқлаштириш ва ўқув топшириғни белгилаш	<ul style="list-style-type: none"> ✓ индивидуал ва гуруҳда ишлаш; ✓ муаммоларни долзарблик иерархиясини аниқлаш; ✓ асосий муаммоли вазиятни белгилаш
3-босқич: Кейсдаги асосий муаммони таҳлил этиш орқали ўқув топшириғининг ечимини излаш, ҳал этиш йўллари ишлаб чиқиш	<ul style="list-style-type: none"> ✓ индивидуал ва гуруҳда ишлаш; ✓ муқобил ечим йўллари ишлаб чиқиш; ✓ ҳар бир ечимнинг имкониятлари ва тўсиқларни таҳлил қилиш; ✓ муқобил ечимларни танлаш
4-босқич: Кейс ечимини ечимини шакллантириш ва асослаш, тақдимот.	<ul style="list-style-type: none"> ✓ якка ва гуруҳда ишлаш; ✓ муқобил вариантларни амалда қўллаш имкониятларини асослаш; ✓ ижодий-лойиҳа тақдимотини тайёрлаш; ✓ якуний хулоса ва вазият ечимининг амалий аспектларини ёритиш

“Ассесмент” методи

Методнинг мақсади: мазкур метод таълим олувчиларнинг билим даражасини баҳолаш, назорат қилиш, ўзлаштириш кўрсаткичи ва амалий кўникмаларини текширишга йўналтирилган. Мазкур техника орқали таълим олувчиларнинг билиш фаолияти турли йўналишлар (тест, амалий кўникмалар, муаммоли вазиятлар машқи, қиёсий таҳлил, симптомларни аниқлаш) бўйича ташҳис қилинади ва баҳоланади.

Методни амалга ошириш тартиби:

“Ассесмент”лардан маъруза машғулотларида талабаларнинг ёки қатнашчиларнинг мавжуд билим даражасини ўрганишда, янги маълумотларни баён қилишда, семинар, амалий машғулотларда эса мавзу ёки маълумотларни ўзлаштириш даражасини баҳолаш, шунингдек, ўз-ўзини баҳолаш мақсадида индивидуал шаклда фойдаланиш тавсия этилади. Шунингдек, ўқитувчининг ижодий ёндашуви ҳамда ўқув мақсадларидан келиб чиқиб, ассесментга қўшимча топшириқларни киритиш мумкин.

Ҳар бир катакдаги тўғри жавоб 5 балл ёки 1-5 балгача баҳоланиши мумкин.



Тест



Муаммоли вазият



**Тушунча таҳлили
(симптом)**



Амалий вазифа

“Инсерт” методи

Методни амалга ошириш тартиби:

- ўқитувчи машғулотга қадар мавзунинг асосий тушунчалари мазмуни ёритилган матнни тарқатма ёки тақдимот кўринишида тайёрлайди;
- янги мавзу моҳиятини ёритувчи матн таълим олувчиларга тарқатилади ёки тақдимот кўринишида намойиш этилади;
- таълим олувчилар индивидуал тарзда матн билан танишиб чиқиб, ўз шахсий қарашларини махсус белгилар орқали ифодалайдилар. Матн билан ишлашда талабалар ёки қатнашчиларга қуйидаги махсус белгилардан фойдаланиш тавсия этилади:

Белгилар	Матн
“V” – таниш маълумот.	
“?” – мазкур маълумотни тушунмадим, изоҳ керак.	
“+” бу маълумот мен учун янгилик.	
“– ” бу фикр ёки мазкур маълумотга қаршиман?	

Белгиланган вақт якунлангач, таълим олувчилар учун нотаниш ва тушунарсиз бўлган маълумотлар ўқитувчи томонидан таҳлил қилиниб, изоҳланади, уларнинг моҳияти тўлиқ ёритилади. Саволларга жавоб берилади ва машғулот якунланади.

Ш БЎЛИМ

НАЗАРИЙ
МАТЕРИАЛЛАР

III. НАЗАРИЙ МАТЕРИАЛЛАР

1-маъруза: Катта маълумотлар (Big Data) тушунчаси.

Режа:

- 1.1. Big data га кириш
- 1.2. Big data: асосий фазифалар.

Таянч иборалар: *Big Data – катта маълумотлар, Hadoop, Modeling, Data Mining, Data File Sources, File Exporting, Баиоратли дастурлар (Predictive Applications).*

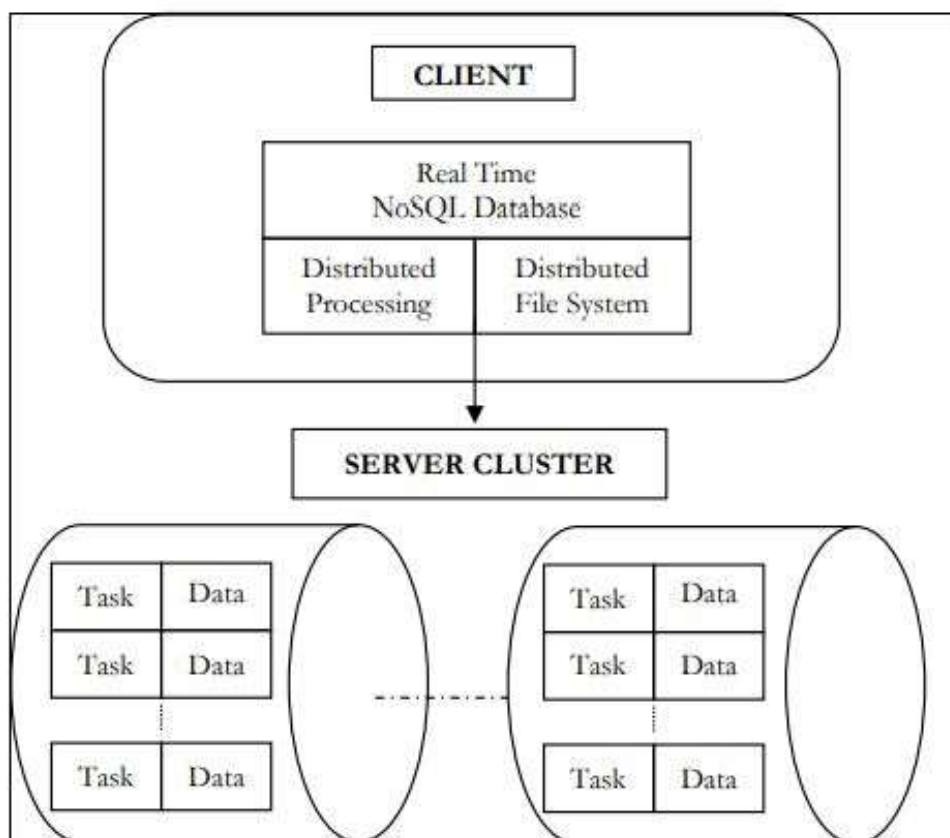
1.1. Big Data га кириш.

Big Data – катта маълумотлар - бу катта ҳажм, тезлик, хилма-хиллик ва ишонччилик билан бошқариладиган маълумотлар тўпламидир. Бу маълумотларга ишлов беришнинг анъанавий дастурий таъминоти томонидан ҳал қилиниши мумкин бўлмаган жуда катта ёки мураккаб бўлган маълумотларни таҳлил қилиш, маълумотни мунтазам равишда чиқариб олиш ёки маълумотлар тўплamlари билан муомала қилиш усуллари билан ишлайдиган майдон ҳисобланади. Улар турли хил манбалардан - интернет, мобил қурилмалар, ижтимоий медиа, геокосмик қурилмалар, сенсорлар ва бошқа машина томонидан олинган маълумотлардан ташкил топган. МББТ ва маълумотлар омборидан фойдаланган ҳолда анъанавий маълумотларни қайта ишлаш ва тузилган маълумотларни таҳлил қилиш энди Катта маълумотлар муаммоларини хал қила олмайди. Катта маълумотлар технологиялари очик манбали дастурий таъминот ва оммавий равишда тақсимланган ишлов бериш платформаларини қамраб олади. Иқтисодиётнинг ўзгариши билан бир каторда, технология асосий (mainframe) компьютер, шахсий компьютер, мижоз-сервер компьютерлари, Интернет, cloud computing, мобил

компьютерлар ва ижтимоий тармоқларда ривожланмоқда. Катта маълумотлар мукамал бўрон деб таърифлаган технологиянинг учта йўналишини бирлаштирган эволюциянинг сўнгги босқичи сифатида пайдо бўлади: ҳисоблаш, маълумотлар ва конвергенция.

Катта маълумотлар таъсири кўплаб тармоқлар ва соҳаларда сезилмоқда. Катта маълумотлар деганда одатдаги маълумотлар базасининг имкониятларидан ташқари бўлган маълумотлар тўпламлари тушунилади. Катта маълумотлар кенг тарқалган технология бўлиб, сақлаш ва қайта ишлаш қобилиятига егадир.

Катта маълумотларни йиғиш, қайта ишлаш ва таҳлил қилиш қобилияти динамик бозор шароитлари ва мижозлар еҳтиёжларига жавоб бериш қобилиятини ошириш орқали улкан рақобат устунлигини таъминлаши мумкин.



1.1-расм. Катта маълумотлар учун мижоз-сервер архитектураси

Мижозлар даражасидаги архитектура NoSQL маълумотлар базалари, тақсимланган файл тизимлари ва тақсимланган ишлов бериш тизимларидан иборат. NoSQL одатда "Фақат SQL эмас" деб талқин қилинади. NoSQL маълумотлар базалари - ўзаро боғлиқ бўлмаган, SQL-га асосланган эмас ва маълумотни ўзаро боғлиқ бўлмаган маълумотлар билан яхши ишлайдиган калит-қиймат жуфтликларида сақлайди. NoSQL маълумотлар базалари Катта маълумотлар учун кенг тарқалган, кенгайтириладиган маълумотларни сақлашни таъминлайди.

1.2. Big data: асосий вазифалар.

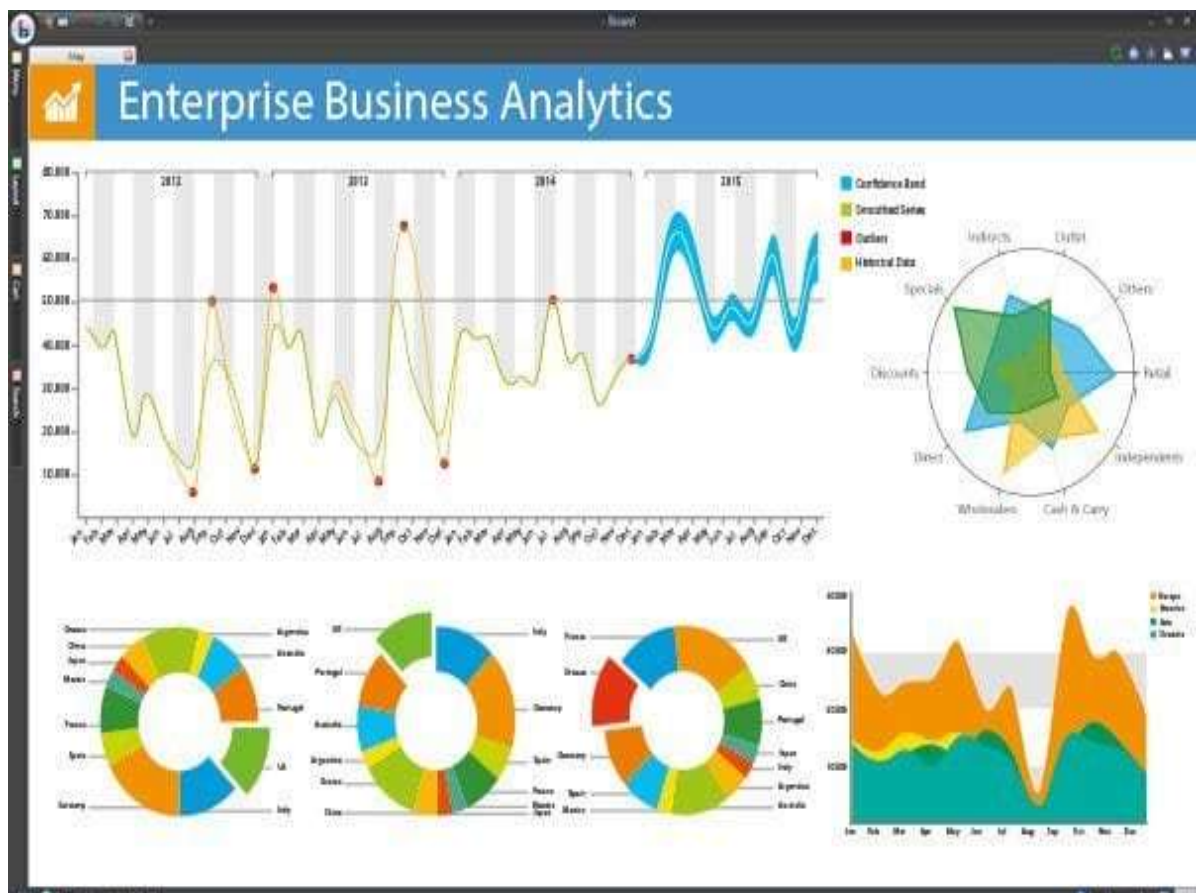
Катта маълумотларнинг асосий функцияларини келтирамыз:

- Маълумотларни қайта ишлаш:
 - Modeling,
 - Data Mining,
 - Data File Sources,
 - File Exporting,
- Башоратли Дастурлар (Predictive Applications);
- Таҳлил:
 - Хавф таҳлили,
 - Қарорларни бошқариш,
 - Контентни таҳлил қилиш,
 - Статистик таҳлил,
 - Башоратли таҳлил,
 - Ижтимоий Медиа таҳлили,
- Ҳисобот хусусиятлари:
- Хавфсизлик хусусиятлари;
 - Ягона кириш,
 - Маълумотни шифрлаш,
- Технологияларни қўллаб-қувватлаш:

- A / B синов,
- Nadoop билан интеграция.

Маълумотларни қайта ишлаш

Маълумотни қайта ишлаш хусусиятлари, маълумотларни тўплаш ва ташкиллаштиришни ўз ичига олади. Маълумотни моделлаштириш - мураккаб маълумотлар тўпламлари ва уларни визуал диаграмма намоиш этишдир. Қарор қабул қилиш учун ушбу маълумотлардан фойдаланишга ҳаракат қилаётган фойдаланувчилар учун ишнинг боришини осонлаштиради. Маълумотларни ишлаб чиқариш фойдаланувчиларга маълумотларни ҳар хил нуқтаи назардан ажратиб олиш ва таҳлил қилиш ва уларни амалдаги тушунчаларга умумлаштириш имконини беради. Бу вақт давомида тўпланган катта тузилмали маълумотлар тўпламларида айниқса фойдалидир.



1.2-расм. Board BI томонидан тақдим етилган визуал маълумот модели

Катта маълумотларни таҳлил воситалари Microsoft Access, Microsoft Excel, матнли файллар ва бошқа матн файллар каби манбалардан маълумотларни импорт қилишни ёқиши керак. Бир нечта манбалардан ва бир нечта форматдаги маълумотларни бирлаштириш имкониятига ега бўлиш, маълумотни ўзгартириш заруриятини олдини олиш орқали меҳнатни камайтиради ва тизимга тўғридан-тўғри импорт қилиш орқали бутун жараёни тезлаштиради. Экспорт имкониятлари ҳақида ҳам шундай дейиш мумкин - визуализация қилинган маълумотлар тўпламини олиш ва уларни PDF, Excel файллари, Word файллари ва .dat файллари каби экспорт қилиш аввалги жараёнларда тўпланган маълумотларнинг фойдалилиги ва ўтказувчанлиги учун муҳимдир.

Башоратли дастурлар (Predictive Applications)

Идентификацияни бошқариш (ёки идентификация ва киришни бошқариш) - бу сизнинг маълумотларингизга кириш ҳуқуқини бошқариш учун ташкилий жараён. Идентификацияни бошқариш функционал имкониятлари тизимга кириш ҳуқуқига эга бўлган барча фойдаланувчилар, шу жумладан шахсий фойдаланувчилар, компьютер ускуналари ва дастурий таъминотни идентификациялаш маълумотларини бошқаради. Идентификация менежменти, шунингдек фойдаланувчиларнинг кириш ҳуқуқи билан идентификацияни қандай олишлари, ушбу идентификаторларни ҳимоя қилиш ва тармоқ протоколлари ва пароллари каби бошқа тизим ҳимояларини қўллаб-қувватлаш масалалари билан шуғулланади. Бу фойдаланувчи тизимга кириш ҳуқуқига эга ёки йўқлигини ва фойдаланувчидан фойдаланиш ҳуқуқига эга эканлигини аниқлайди. Идентификация менежменти иловалари тизимга ва тасдиқланган фойдаланувчилар фақат сизнинг тизимингизга ва кенгайтирилган маълумотларга кира олишларини таъминлашга қаратилган. Бу ҳар қандай

ташкilotнинг хавфсизлик режасининг муҳим элементи бўлиб, реал вақт режимида хавфсизлик ва фирибгарликларни таҳлил қилиш имкониятларини ўз ичига олади.

Таҳлил

Big Data Analytics воситалари фойдаланувчиларга турли хил таҳлил пакетлари ва модулларини таклиф қилади. Масалан, риск таҳлиллари ҳар қандай берилган ҳаракатлар атрофидаги ноаниқликни ўрганишдир. Ундан келажакдаги воқеаларнинг салбий таъсирини минималлаштириш учун прогнозлаш билан биргаликда фойдаланиш мумкин. Хавфларни таҳлил қилиш фойдаланувчиларга ташкilotнинг сабр-тоқатлилиги ва хавф-хатарига аниқ жавоб бериш ва тушуниш орқали ушбу хавфларни камайтиришга имкон беради. Қарорларни бошқариш бизнес юритиш учун қарор қабул қилиш жараёнларини ўз ичига олади. Қарорларни бошқариш модуллари қарорларга фойдаланиладиган активлар сифатида қарашади. У қарорларни қабул қилиш жараёнининг қисмларини автоматлаштириш учун муҳим нуқталарда технологияни ўз ичига олади. Матнни таҳлил қилиш бу мижозлар томонидан ёзилган ёки ёзилган матнни ўрганиш жараёни. Таҳлил дастури ушбу матнда нақшларни топишга ёрдам беради ва ўрганган нарсангиз асосида бажарилиши мумкин бўлган ҳаракатларни таклиф қилади. Ушбу турдаги таҳлиллар, айниқса мижозларингизнинг еҳтиёжлари ва еҳтиёжлари тўғрисида тўғридан-тўғри ташкilotингиз билан ўзаро алоқада бўлганликлари тўғрисида маълумот олиш учун фойдалидир. Таркибни таҳлил қилиш матнни таҳлил қилиш билан жуда ўхшаш, аммо аудио, видео, расм ва ҳоказоларни, шу жумладан ҳужжатларнинг барча форматларини таҳлил қилишни ўз ичига олади. Ижтимоий медиа таҳлиллари - бу сизнинг фойдаланувчи базангизнинг ижтимоий медиада ўз брендингиз билан қандай муносабатда бўлишига қаратилган контент таҳлилининг бир шакли. Статистик таҳлиллар рақамлардан иборат маълумотлар тўпламларини

тўплайди ва таҳлил қилади. Мақсад жами аҳолининг вакили бўлган умумий маълумотларнинг намунасини олишдир. Статистик таҳлил беш босқичда амалга оширилади:

- 1) Маълумотларнинг моҳиятини тавсифлаш, маълумотларни тақдим этган шакл билан боғлиқликни ўрганиш;
- 2) Уланишларни умумлаштириш учун модель яратиш;
- 3) Тўғрилигини исботлаш;
- 4) Рад этиш;
- 5) Қарорларни бошқариш учун башоратли таҳлилларни қўллаш.

Ҳисобот хусусиятлари

Ҳисобот функциялари фойдаланувчиларни ўз бизнесларининг юқори қисмида ушлаб туради. Ҳақиқий вақтда ҳисоботларни маълумотларни тўплайди ва сизга одатда интуитив бошқарув панели шаклида етказди. Бу фойдаланувчиларга оғир вақт шароитида тезкор қарорлар қабул қилишга ва рақобатдош бўлишга имкон беради.

Хавфсизлик хусусиятлари

Муваффақиятли бизнес учун тизимингизни хавфсиз сақлаш жуда муҳимдир. Big Data таҳлил воситалари хавфсизликни таъминлаш учун хавфсизлик хусусиятларини таклиф қилиши керак. Бундай хусусиятлардан бири бу битта тизимга кириш ёки SSO ҳам дейилади. Бу фойдаланувчиларга бир нечта дастурларга кириш учун кириш маълумотларини битта тўпламини тайинлайдиган аутентификация хизмати. У охириги фойдаланувчи рухсатларини тасдиқлайди ва бир сеанс давомида бир неча марта киришга эҳтиёжни йўқ қилади. Шунингдек, у тизимда ким нима қилаётганини кузатиб бориш учун фойдаланувчи фаолияти ва қайд ёзувларини қайд қилиши ва кузатиши мумкин. Катта маълумотларнинг таҳлил платформалари томонидан таклиф қилинадиган яна бир хавфсизлик хусусияти маълумотларни

шифрлашдир. Маълумотни шифрлаш алгоритмлар ёки кодлар ёрдамида электрон маълумотни ўқиб бўлмайдиган форматга ўзгартиришни ўз ичига олади.

Технологияларни қўллаб-қувватлаш

Сизнинг таҳлил дастурий таъминотингиз сизга фойдали бўлиши мумкин бўлган турли хил технологиялар ва вазифаларни қўллаб-қувватлаши керак. А / Б синовлари бунга битта мисолдир. А / Б тести веб-саҳифа ёки дастурнинг иккита версиясини солиштириб, қайси бири яхшироқ ишлашини аниқлайди. Фойдаланувчилар веб-саҳифанинг иккала версияси билан ўзаро қандай ишлашини каталоглаштирадилар ва ушбу ўзгартириш учун қайси версия энг яхши ишлашини аниқлаш учун ушбу натижалар бўйича статистик таҳлил ўтказадилар.

Маълумотни таҳлил қилишнинг яна бир катта хусусияти - бу Hadoop билан интеграция. Hadoop - бу маълумотлар таҳлилини амалга ошириш учун таянч вазифасини ўташи мумкин бўлган очиқ манбали дастурлар тўплами. У тўрт модулдан иборат:

1. Тақсимланган файл тизими (Distributed File System): маълумотни уланган сақлаш мосламалари тизими орқали кириш мумкин бўлган форматда сақлашга имкон беради.
2. MapReduce: ушбу файл тизимидаги маълумотларни ўқийди ва фойдаланувчилар шархлашлари мумкин бўлган визуализацияларга форматлайди.
3. Hadoop Common: фойдаланувчи компьютерлари учун файл тизимида сақланган маълумотни ўқиш учун зарур бўлган Java воситалари тўплами.
4. YARN: маълумотларни сақлаш ва таҳлил қилиш тизимлари ресурсларини бошқаради.

Ушбу модуллар билан интеграция фойдаланувчиларга Hadoop-дан тўпланган натижаларни бошқа тизимларга юбориш имкониятини беради. У ташкилот ичида ҳам, ташкилотлар ўртасида ҳам ўзаро ва мослашувчанликни, шунингдек алоқаларни ривожлантиришга ёрдам беради.

Hadoop

Hadoop катта маълумотларнинг асосий қисми ёки асосидир. Hadoop - бу арзон машиналар кластерида катта миқдордаги маълумотлар тўпламини тақсимланган ҳолда сақлаш технологияси ҳисобланади.

Нима учун Hadoop ихтиро қилинди?

- Катта маълумотлар базаларини сақлаш. Анъанавий МББТ катта ҳажмдаги маълумотларни сақлашга қодир эмас. Мавжуд МББТда маълумотларни сақлаш қиймати жуда катта. Бу аппарат ва дастурий таъминот учун ҳам қимматга тушади.
- Турли хил форматларда маълумотларни қайта ишлаш. МББТ маълумотларни тузилган форматда сақлаш ва бошқариш имкониятига эга. Аммо реал дунёда биз маълумотлар билан тузилган, тузилмаган ва ярим тузилмали форматда ишлашимиз керак.

Маълумотлар юқори тезликда олинади. Маълумотлар ҳар куни тера-пета байтгача тартибда чиқарилади. Шундай қилиб, биз бир неча сония ичида реал вақт режимида маълумотларни қайта ишлаш учун тизимга муҳтожмиз. Анъанавий МББТ реал вақт режимида катта тезликда ишлашни таъминлай олмайди.

Назорат саволлари:

1. Big data деганда нимани тушинасиз?
2. Катта маълумотларнинг асосий вазифалари нимада?
3. Hadoop – бу нима?

4. Modeling тушинчасига таъриф беринг.
5. Data Mining тушинчасига таъриф беринг.
6. Data File Sources тушинчасига таъриф беринг.
7. File Exporting тушинчасига таъриф беринг.
8. Башоратли дастурлар (Predictive Applications) вазифасини тушинтириб беринг.
9. Катта маълумотлар таҳлилининг қандай турлари билан ишлайди?
10. Катта маълумотлар технологияларини қўллаб-қўвватлаш вазифасини тушинтириб беринг.

Адабиётлар ва интернет сайтлари:

1. Радченко И., Николаев И., Технологии и инфраструктура Big data: Учебное пособие. – Санк-петербург: ИТМО, 2018.
2. <https://computingforgeeks.com/data-mining-your-clicks>
3. <https://www.selecthub.com/big-data-analytics/big-data-analytics-requirements/>
4. <https://data-flair.training/blogs/hadoop-ecosystem-components/>

2-маъруза. Катта маълумотларнинг хусусиятлари.

Режа:

2.1. Катта маълумотларнинг афзалликлари.

2.2. Таҳлил қилинадиган маълумотларнинг ҳажми, анъанавий маълумотлар таҳлил тизимлари ошиб кетганда, катта маълумотларни ҳисоблаш тизимини жорий этиш.

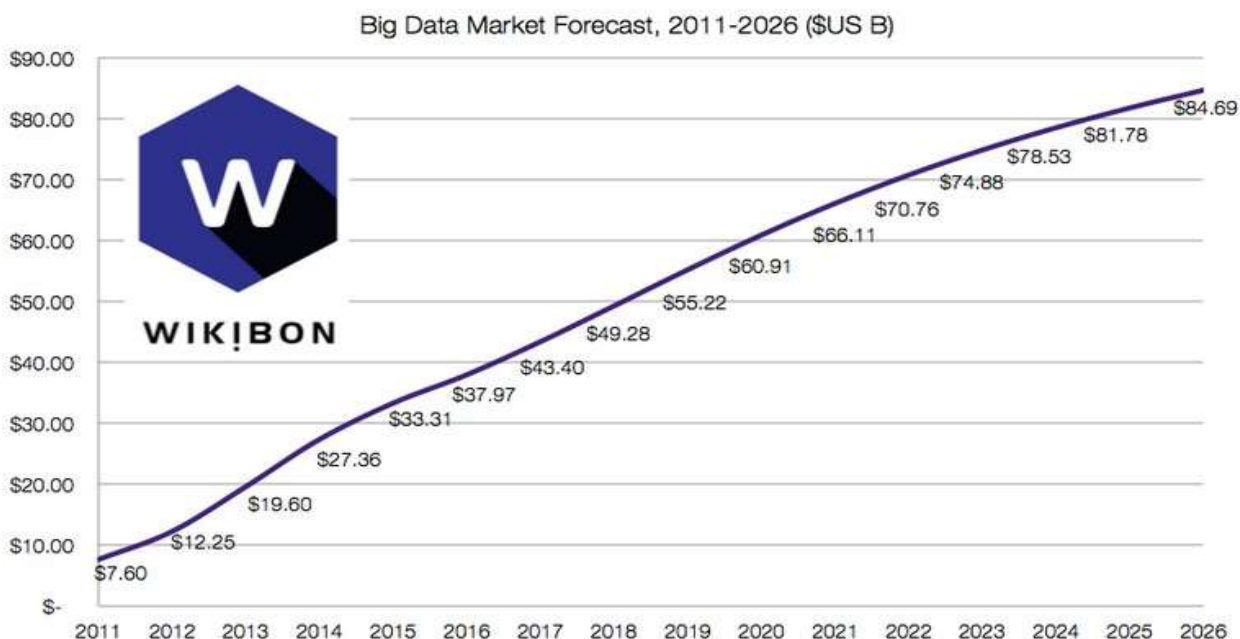
Таянч иборалар: *Big Data analysis, катта маълумотлар технологиялари, катта маълумотлар афзалликлари.*

2.1. Катта маълумотларнинг афзалликлари.

Big data — бу структураланган ва структураланмаган маълумотларни, конкрет масалалар ва мақсадларда уларни қўллаш учун, ишлов бериш методлари, турли инструментлар ва ёндашувлардир. Структураланмаган маълумотлар - бу маълум тартибда ташкиллаштирилмаган ёки олдиндан аниқ структурага эга бўлмаган ахборот.

«Катта маълумотлар» терминини Nature журналининг редактори Клиффорд Линч 2008 йилда, дунёда ахборот ҳажмларининг ўсишига бағишланган маҳсус нашрида киритган эди. Шунга қарамасдан, албатта «Катта маълумотлар» олдинроқ ҳам мавжуд эди.

Мутахассисларнинг фикрича Big data даражасига кунига 100 Гб ортиқ барча маълумотлар оқими киради. Бугунда бу оддий термин остида иккитагина сўз ётади – маълумотларга ишлов бериш ва сақлаш. Замонавий дунёда Big data - катта миқдордаги маълумотларни таҳлил қилиш учун янги технологиялар пайдо бўлиши билан боғлиқ ижтимоий-иқтисодий феномен.



2.1-расм. Big data ўсиш кўрсаткичи

Инсон аниқ ва унга керакли бўлган натижаларни олиш учун ва уларни келажакда самарали қўллаши учун катта хажмдаги ахборотларга ишлов берилади. Big data - бу муаммони ечими ва анъанавий маълумотларни бошқариш тизимларига альтернативдир.

Катта маълумотлар бу жуда хилма-хил, тез ўзгариб турадиган ёки анъанавий технологиялар, маҳорат ва инфратузилмани самарали ечиш учун катта хажмдаги маълумотларни қамраб оладиган технологиялар ва ташаббусларга тегишли. Аммо ҳозирги кунда янги технологиялар ёрдамида катта маълумотлар қийматини англаш жуда осон, масалан, харидорлар томонидан харид қилинадиган харидорларнинг ахлоқ тузатиш тенденцияларини аниқлаш, маҳсулотларнинг нархини белгилаш шулар жумласидандир.

АТ соҳасидаги устунликлар:

Кўпгина ески АТ-компаниялар тўлиқ реализация вақтида носозликлар ва муаммоларнинг асл сабабларини аниқлаш ва антик кодли базаларни аниқлаш

орқали ўзларининг эскирган асосий кадрларини модернизация қилиш учун катта маълумотларга тўла боғлиқ.

Маълумотларнинг катта ечимлари битта машинадан минглаб машиналаргача бўлган ҳар бири маҳаллий ҳисоблаш ва сақлашни таклиф қиладиган *hadoop*-га асосланади, бундан ташқари у "бепул" очик манбали платформалар бўлиб, янги ташкилотни сотиб олишга сармоя киритишни минималлаштиришга имкон беради.

Катта маълумот технологияларининг ёрдами билан АТ-компаниялар учинчи томон маълумотларини тезкор равишда қайта ишлашга қодир.

Бизнесдаги катта маълумотларнинг афзалликлари:

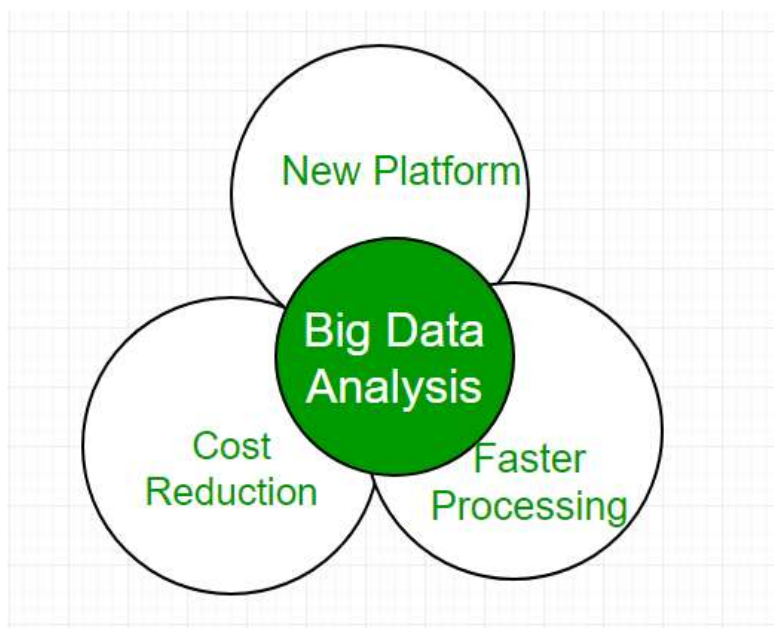
Нақд пулни тўлдириш, тўлиқсиз ёки ноаниқ кредит лимитлари ёки нархлар тўғрисидаги маълумот мижозларга хизмат кўрсатишнинг йўқолишига олиб келади ёки даромадни камайтиради ёки хизмат нархини ошириши мумкин, катта маълумотлар технологиялари ва турли хил алгоритмларни тезроқ ишлатиш қобилияти билан маълумотлар янгиланиши мумкин. кун давомида мунтазам равишда янгиланади.

Маълумотларнинг тизимли таҳлили ёки мавжуд ҳолатларга мувофиқ тўғри бизнес қарорларини қабул қилишга олиб келадиган маълумотларнинг умумий ҳолатини баҳолаш учун ишлатилади, чунки баъзида нотўғри маълумотлар нотўғри бошқаришга олиб келса, бизнес қарорлари нотўғри маълумотларга асосланади ва у бизнес қулайди.

Катта маълумотлар учун «Уч V» деб номланувчи анъанавий аниқловчи характеристикаларни ажратиш мумкин.

- **Volume** — физик хажмнинг катталиги.
- **Velocity** — натижаларни олиш учун тезликнинг ошиши ва тезкор ишлов беришнинг тезлиги.
- **Variety** — турли типдаги маълумотларга бир вақтда ишлов бериш имконияти.

Турли маълумотлар ҳажми ва тез келиб тушадиган сонли ахборотларга анъанавий инструментлар билан ишлов бериш имконига эга эмас. Маълумотлар таҳлилининг ўзи инсон кўра олмайдиган аниқ ва сезилмас қонуниятларни кўриш имконини беради. Бу бизнинг ҳаётимизда барча соҳаларни – давлат бошқарувидан то ишлаб чиқариш ва текоммуникацияларнинг оптималлаштириш имконини беради.

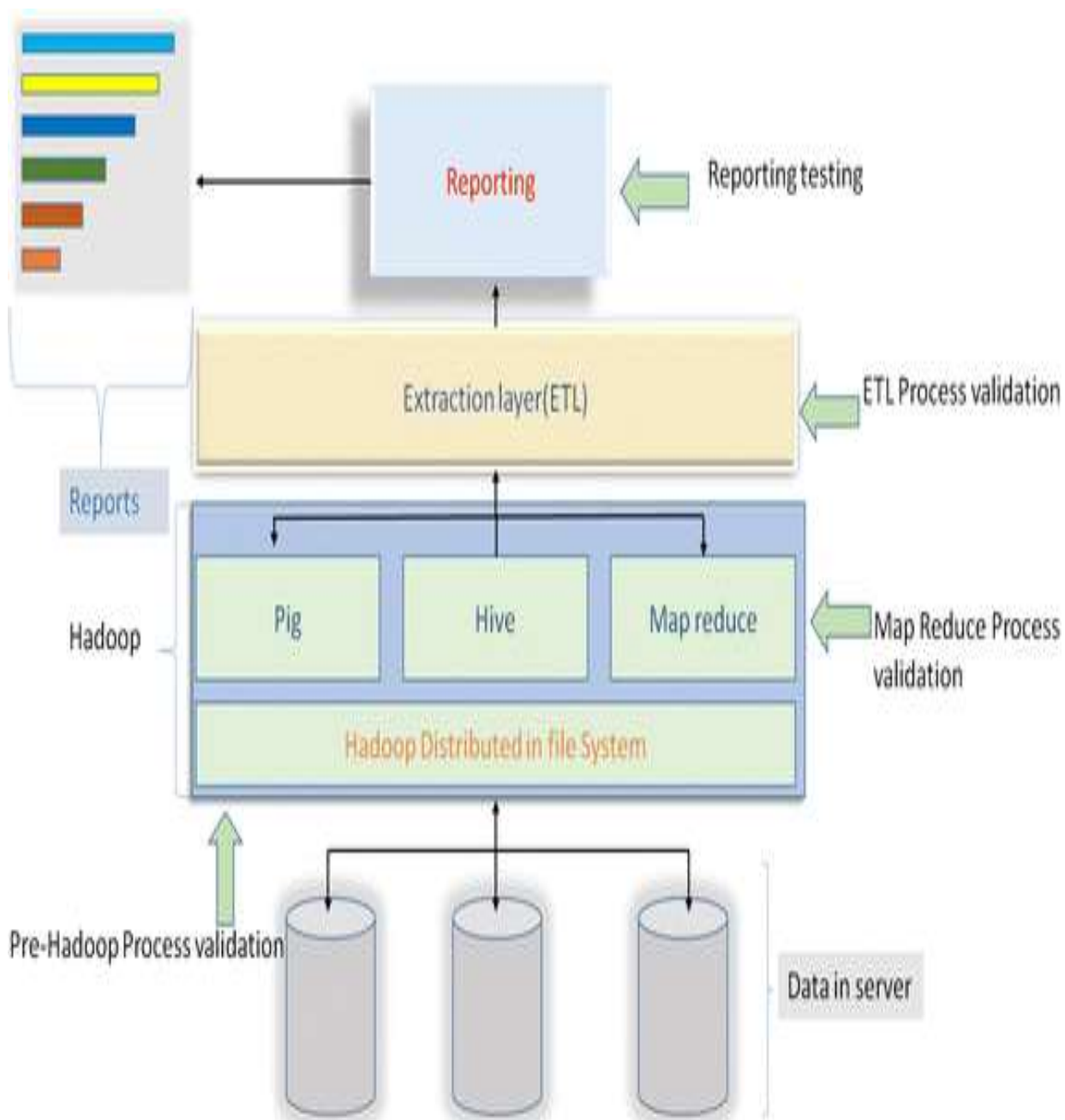


2.1-расм. Big Data analysis.

2.3.Таҳлил қилинадиган маълумотларнинг ҳажми, анъанавий маълумотлар таҳлил тизимлари ошиб кетганда, катта маълумотларни ҳисоблаш тизимини жорий етиш.

Корхонада катта маълумотларнинг афзалликлари:

Катта маълумотлар компанияга ўз маҳсулотлари, манбалари ёки миқдорлари бўйича триллионлаб ёки миллиардлаб реал вақтда маълумот пунктларини тўплашига имкон беради ва кейин миқдорлар тажрибасини оптималлаштириш учун уларни дарҳол қайта тўплаши мумкин.



2.2-расм. Катта маълумотлар технологиялари

Катта маълумотлар технологияларидан фойдаланган ҳолда маълумотларнинг янгиланиши тезлиги корхоналарга мижозларнинг талабларига тез ва аниқ жавоб беришга имкон беради. Масалан, MetLife MongoDB-дан мижозлар маълумотларини 70 дан ортиқ турли манбаларда тезда бирлаштириш ва ягона, тез янгиланадиган кўринишни тақдим этиш

учун фойдаланган. Катта маълумотлар корхоналарга рақобатчиларига караганда ўзгаришларга тезроқ мослашишларига имкон берадиган даражада харакатланишга ёрдам беради.

Бошқа соҳалардаги катта маълумотларнинг афзалликлари:

Катта маълумотлар технологиялари истеъмолчилар учун турли компанияларнинг "сотиб олиш" ва "сотиш" қарорларини башорат қилиш учун ишлатилади.

Search-Engine катта маълумотлар технологияларидан фойдаланган ҳолда турли хил маълумотлар базаларидан сонияларнинг сонияларида кўп сонли маълумотларни олиш. Масалан, Google MapReduce алгоритмидан берилган сўровни қидириш учун фойдаланади. MapReduce вазифани кичик қисмларга ажратади ва ушбу қисмларни тармоқ орқали уланган кўплаб компьютерларга тайинлайди ва натижани якуний натижани шакллантириш учун тўплайди. Молиявий хизматлар ташкилотлари миждозларнинг ўзаро муносабатлари тўғрисидаги маълумотларни қидириш учун катта маълумотлардан фойдаланиб, фойдаланувчиларни нозик сегментларга ажратиб олишади, бу эса тобора долзарб ва мураккаб таклифларни яратишга ёрдам беради.

Кластерлардан фойдаланиш кластер аъзолигини бошқариш, ресурсларни тақсимлашни мувофиқлаштириш ва алоҳида тугунларда ҳақиқий ишларни режалаштириш учун йечим талаб қилади. Кластерга аъзолик ва ресурсларни тақсимлаш Hadoop-нинг YARN (бошқа манбалар музокарачиси деган маънони англатади) ёки Apache Mesos каби дастурлар томонидан бошқарилиши мумкин.

Кластерли ҳисоблаш - Катта маълумотларнинг сифати туфайли таҳлил қилинадиган маълумотларнинг ҳажми, анъанавий маълумотлар таҳлил тизимлари ошиб кетганда, катта маълумотларни ҳисоблаш тизимини жорий этиш шахсий компютерлар кўп босқичларда маълумотларни қайта ишлаш учун кўпинча йетарли эмас. Катта маълумотларнинг юқори сақлаш ва ҳисоблаш еҳтиёжларини яхшироқ ҳал қилиш учун компютер кластерлари янада мос келади.

Ресурсларни тўплаш: Маълумотни сақлаш учун мавжуд сақлаш майдонини бирлаштириш аниқ фойда келтиради, аммо процессор ва хотирани бирлаштириш жуда муҳимдир. Катта маълумотлар тўпламини қайта ишлаш учун ушбу учта манбанинг кўпи керак.

Юқори фойдаланиш имконияти: Кластерлар турли даражадаги хатоларга чидамликни ва аппарат ёки дастурий таъминотнинг ишламай қолишини олдини олиш учун маълумотлар ва ишлов беришга таъсир қилишини таъминлайди. Ҳақиқий вақтда таҳлил қилиш муҳимлигини таъкидлашда давом етар еканмиз, бу тобора муҳим аҳамият касб етмоқда.

Осон миқёслилик: Кластерлар гуруҳга қўшимча машиналарни қўшиш орқали горизонтал равишда ўлчовни осонлаштиради. Бу шунинг билан, тизим манба талабларининг ўзгаришига машинада жисмоний ресурсларни кенгайтмасдан жавоб бера олади.

Назорат саволлари:

1. Big data технологиялари қаерда ва қандай қўлланилади? Мисоллар келтиринг.
2. Big data технологияларининг афзалликларини тушинтириб беринг.
3. Таълим тизимида Big data технологияларидан қандай фойдаланиш мумкин?

Адабиётлар ва интернет сайтлари:

1. Радченко И., Николаев И., Технологии и инфраструктура Big data: Учебное пособие. – Санкт-петербург: ИТМО, 2018.
2. <https://computingforgeeks.com/data-mining-your-clicks>
3. <https://www.selecthub.com/big-data-analytics/big-data-analytics-requirements/>
4. <https://data-flair.training/blogs/hadoop-ecosystem-components/>
5. <https://www.digitalocean.com/community/tutorials/an-introduction-to-big-data-concepts-and-terminology>

3-маъруза. Катта маълумотларни қайта ишлаш усуллари.

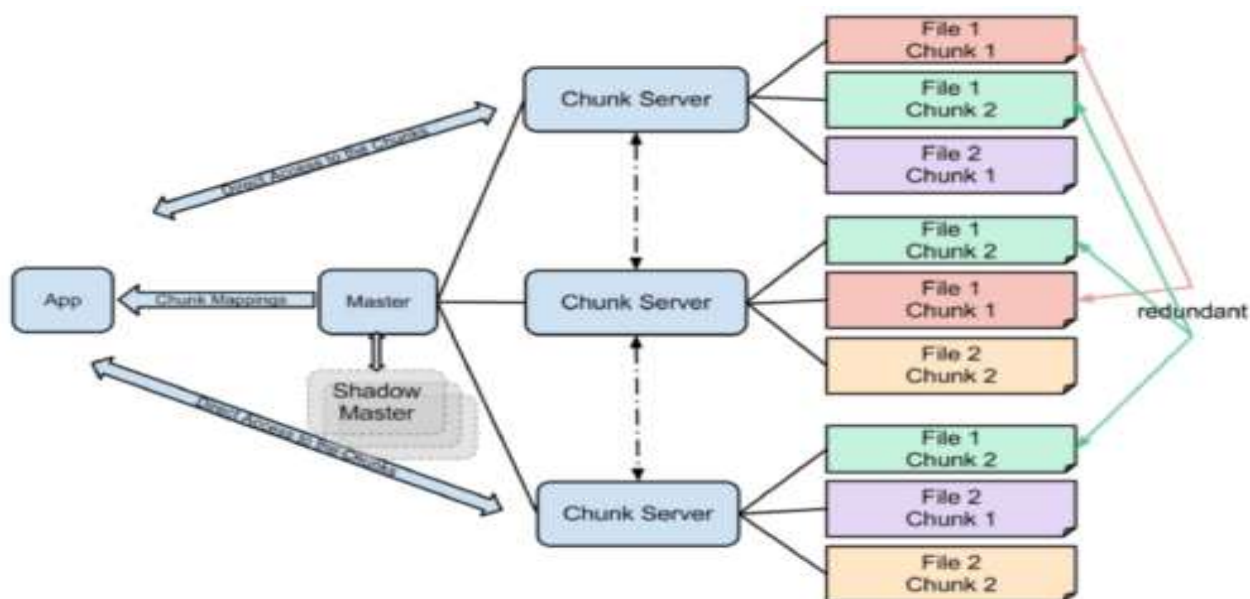
Режа:

- 3.1. Google File System
- 3.2. Hadoop тақсимланган файл тизими.
- 3.3. MapReduce.

Таянч иборалар: *Big Data, HIVE e HIVEQL, Google File System, Hadoop, Hadoop файллари, MapReduce.*

3.1. Google File System.

Кўпгина маълумотлар тўпламлари битта машинага ўрнатиш учун жуда катта ҳисобланади. Маълумотлар базасига тузилмаган маълумотларни киритиш осон бўлмаслиги мумкин. Тақсимланган (distributed) файл тизимлари маълумотларни кўп сонли серверларда сақлайди. Google Fayl System (GFS) бу 2000-йилларнинг бошларида Google томонидан ишлатиладиган тақсимланган файл тизими. Бу кўп сонли арзон серверларда ишлаш учун мўлжалланган.



3.1-расм. Google File System.

GFT мақсади - катта файлларни сақлаш ва уларга кириш имконияти еди, ва умуман айтганда битта қаттиқ дискда сақланиб бўлмайдиган файлларни назарда тутилади. Ҳозир бу файлларни бошқариладиган 64 МБ ҳажмдаги бўлақларга бўлиш ва ушбу бўлимларни бир нечта тугунларда сақлаш, шу билан бирга файллар тизимида сақланадиган қисмлар ўртасида харитани тузишдир.

GFT, кўпинча муваффақиятсиз бўлиши мумкин бўлган жуда арзон товар таркибий қисмларида ишлайди, шунинг учун муваффақиятсиз мониторинг ва тикланишни амалга ошириши керак. У бир вақтнинг ўзида кўплаб йирик файлларни сақлаши мумкин ва уларга икки хил ўқиш имкониятини беради: кичик тасодифий ўқишлар ва катта оқим оқимлари. Файлларни қайта ёзиш ўрнига, GFT тизимдаги мавжуд файлларга маълумотларни қўшиш учун оптималлаштирилган. GFT бош тугмаси файллар индексини сақлайди, GFT тармоқ серверлари еса бир нечта Linux тугунларида файл тизимларида ҳақиқий қисмларни сақлайди. GFT -да сақланадиган қисмлар кўпайтирилади, шунинг учун тизим сервернинг ишдан чиқишига тоқат қилиши мумкин. Текширув варақалари ёрдамида маълумотларнинг бузилиши ҳам аниқланади ва GFT ушбу ҳодисаларни имкон қадар тезроқ қоплашга ҳаракат қилади.

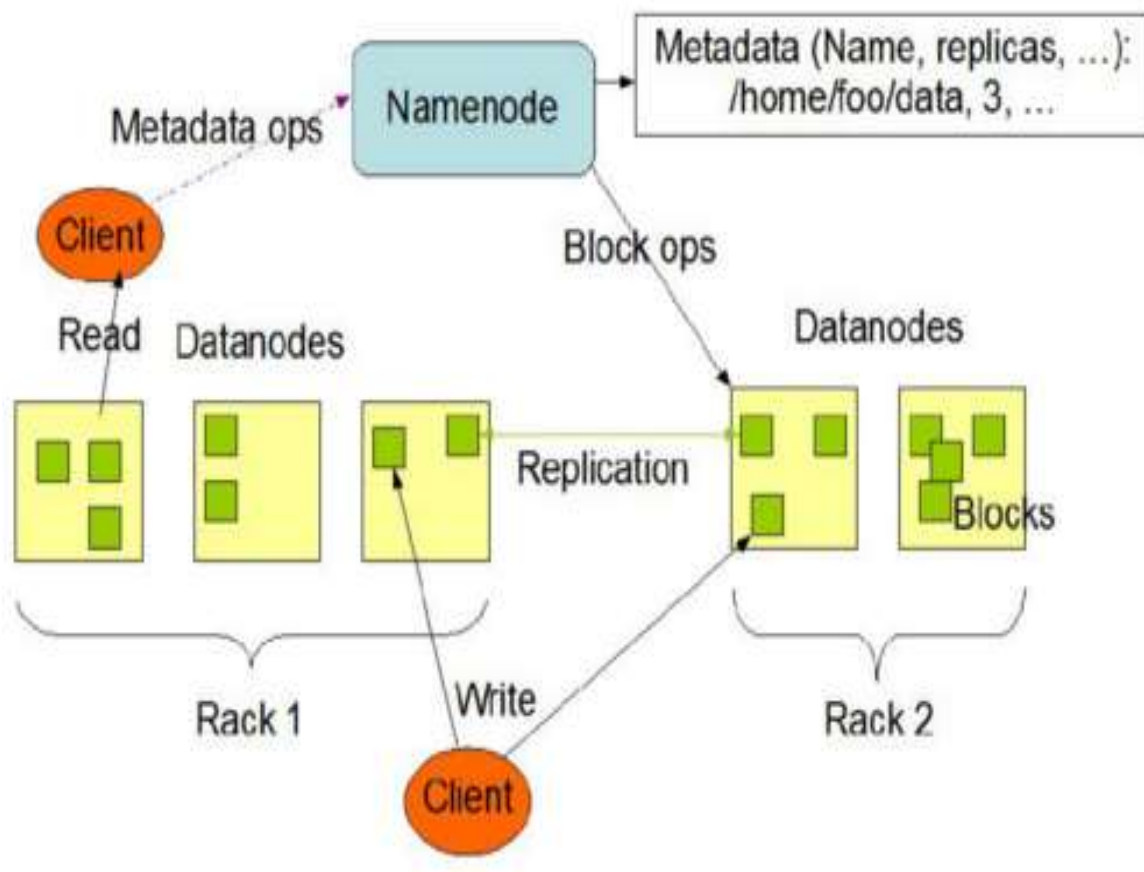
Google fayl тизимининг қисқача тарихи:

- 2003: Google File System чиқди.
- 2004: MapReduce асослари чиқарилди. Бу кластерда параллел, тақсимланган алгоритмли катта маълумотлар тўпламларини қайта ишлаш ва яратиш учун дастурий модел ва тегишли дастур.
- 2006: MapReduce дастурий моделидан фойдаланган ҳолда катта маълумотларни тарқатиш ва қайта ишлаш учун дастурий таъминотни тақдим етувчи Hadoop ташкил етилди. Hadoop-даги барча модулар аппаратдаги узилишлар тез-тез учрайдиган ҳодиса ва автоматик равишда тизим томонидан ишлов берилиши керак деган қатъий тахмин билан ишлаб чиқилган.

- 2007: Google Bigtable- дан кейин моделлаштирилган ва Жава-да ёзилган HBase, очиқ манбали, алоқасиз, тақсимланган маълумотлар базаси. У Apache Hadoop лойиҳасининг бир қисми сифатида ишлаб чиқилган ва HDFS юқори қисмида ишлайди.
- 2008 йил: Hadoop TeraSort танловида ғолиб чиқди. TeraSort бу таниқли таққословчи бўлиб, берилган компютер тизимида тасодифий тақсимланган маълумотларнинг бир терабайтини саралаш учун вақтни ўлчайди.
- 2009 йил: очиқ манба бўлиб тарқатилган, умумий мақсадлар учун мўлжалланган кластер-ҳисоблаш базаси. Маълумотлар мутаносиблиги ва хатоларга чидамлилиги билан бутун кластерларни дастурлаш учун интерфейсни таъминлайди.
- 2010: маълумотларни сўраш ва таҳлил қилиш учун Apache Hadoop-нинг тепасида қурилган маълумотлар омбори дастурининг лойиҳаси яратилди. Hadoop билан бирлаштирилган турли хил маълумотлар базалари ва файл тизимларида сақланадиган маълумотларга сўров бериш учун SQL-га ўхшаш интерфейсни беради.

3.2. Hadoop тақсимланган файл тизими.

Hadoop тақсимланган файл тизими (HTFT) - бу товар ускуналарида ишлашга мўлжалланган тақсимланган файл тизими. Мавжуд тақсимланган файл тизимлари билан жуда кўп ўхшашликларга эга. Бироқ, бошқа тарқатилган файл тизимларидан фарқлар катта. HTFT жуда кам ишларга бардошли ва арзон нархлардаги ускуналарга ўрнатилиши учун мўлжалланган. HTFT дастур маълумотларини юқори даражада ўтказиш имкониятини таъминлайди ва катта маълумотлар тўпламига эга бўлган дастурлар учун мос келади. Аслида, HTFT 1000 дан ортиқ тугунларининг жойлаштирилиши мавжуд.



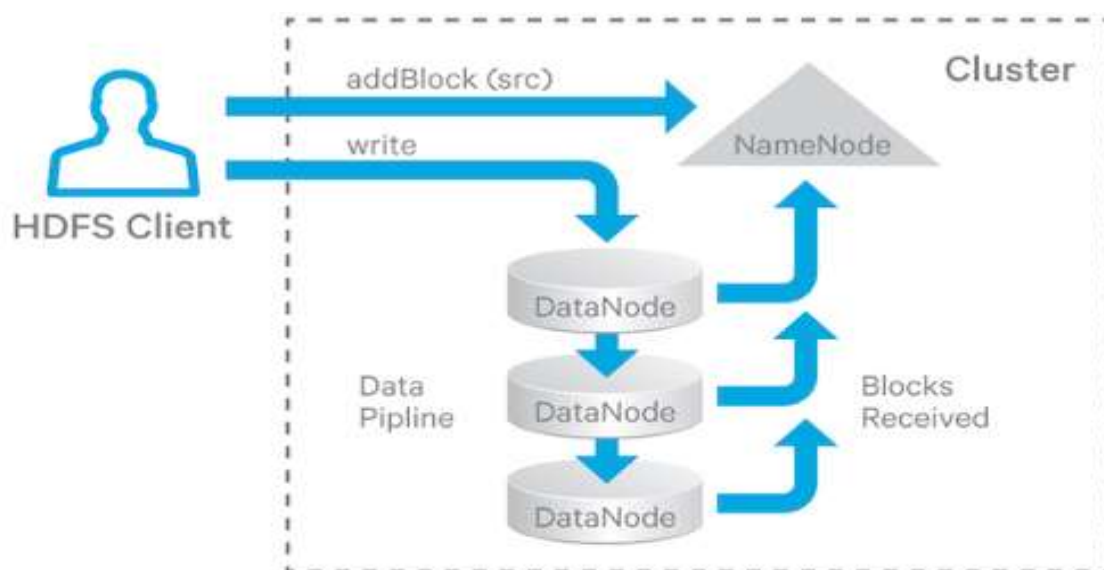
3.2.-расм. Hadoop тақсимланган файл тизими.

HTFT - да файллар блокларга бўлинади ва файлларга кириш кўп ўқийдиган ва битта-ёзувчи семантикасига мос келади. Хатоларга бардошлилик талабини қондириш учун турли хил DataNode-да файллар блокларга бўлинади ва файлларга кириш кўп ўқийдиган ва битта-ёзувчи семантикасига мос келади. Хатоларга бардошлилик талабини қондириш учун турли хил DataNode қувур линиясини яратади. (Репликация коэффициенти одатда қувурлар ичидаги DataNode сонини аниқлайди.) Кейинчалик ушбу блокга қувур линияси орқали ўтилади. Ўқиш операциялари учун мижоз блок нусхасини ушлаб турган DataNode -дан бирини танлайди ва ундан маълумот узатишни талаб қилади.

3.3. MapReduce.

MapReduce бу харитани ёзиш ва қисқартириш функцияларидан иборат дастурий модел. Харита калит / қиймат жуфтлигини қабул қилади ва калит / қиймат жуфтликлари кетма-кетлигини ҳосил қилади. Кейин маълумотлар гуруҳли калитларга бирлаштирилади. Шундан сўнг, қабул қилинган қийматларни бир хил калит билан қисқартирамиз ва янги калит / қиймат жуфтлигини ҳосил қиламиз.

Бажариш жараёнида Map маълумотлари кириш маълумотлари асосида машиналарга берилади. Кейин ушбу Map вазифалари ўз натижаларини беради. Кейинчалик, хариталаш натижаси аралаштирилади ва тартибланади. Кейин, қисқартириш вазифалари режалаштирилган ва бажарилади. Кичиклаштириш натижаси дискка сақланади.



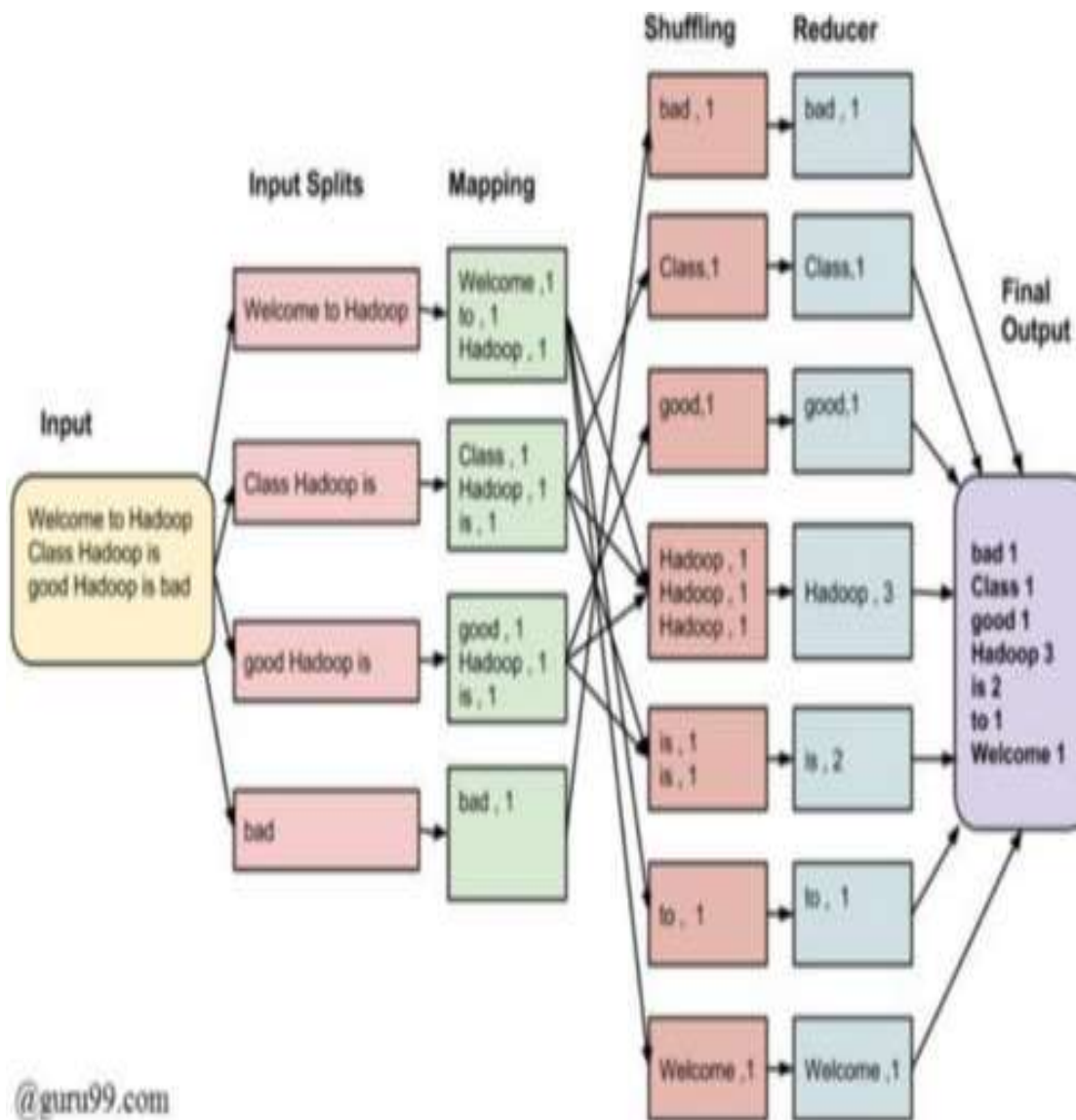
3.3-расм. MapReduce.

Pythonda MapReduce

Келинг, баъзи кодларни кўриб чиқайлик.

Қуйидаги код - Харита функциясидир. У STDIN- дан маълумотларни

Ўқийди, уни сўзларга ажратади ва сўзларни STDOUT- га (оралик) ҳисоблаш учун хариталар рўйхатини чиқаради. Map скриптида сўзларнинг юзага келишининг ўртача (оралик) йиғиндиси ҳисобланмайди. Бунинг ўрнига, у дарҳол <word> 1 tuple чиқаради, гарчи маълум бир сўз киритишда бир неча бор пайдо бўлса ҳам. Бизнинг ҳолда, биз кейинги қисқартириш босқичига яқуний суммани ҳисоблашга имкон берамиз.



3.4-расм. Pythonда MapReduce

```

import sys
# input comes from STDIN (standard input)
for line in sys.stdin:
    # remove leading and trailing whitespace
    line = line.strip()
    # split the line into words
    words = line.split()
    # increase counters
    for word in words:
        # write the results to STDOUT (standard output);
        # what we output here will be the input for the
        # Reduce step, i.e. the input for reducer.py
        # tab-delimited; the trivial word count is 1
        print '%s\t%s' % (word, 1)

```

Куйидаги код - камайтириш функцияси. У STDIN- дан харитадаги кадам натижаларини ўқийди ва ҳар бир сўзнинг ҳолатини яқуний ҳисобланишгача йиғади ва натижаларини STDOUT- га чиқаради.

```

import sys
current_word = None
current_count = 0
word = None
# input comes from STDIN
for line in sys.stdin:
    # remove leading and trailing whitespace
    line = line.strip()
    # parse the input we got from mapper.py
    word, count = line.split('\t', 1)

```

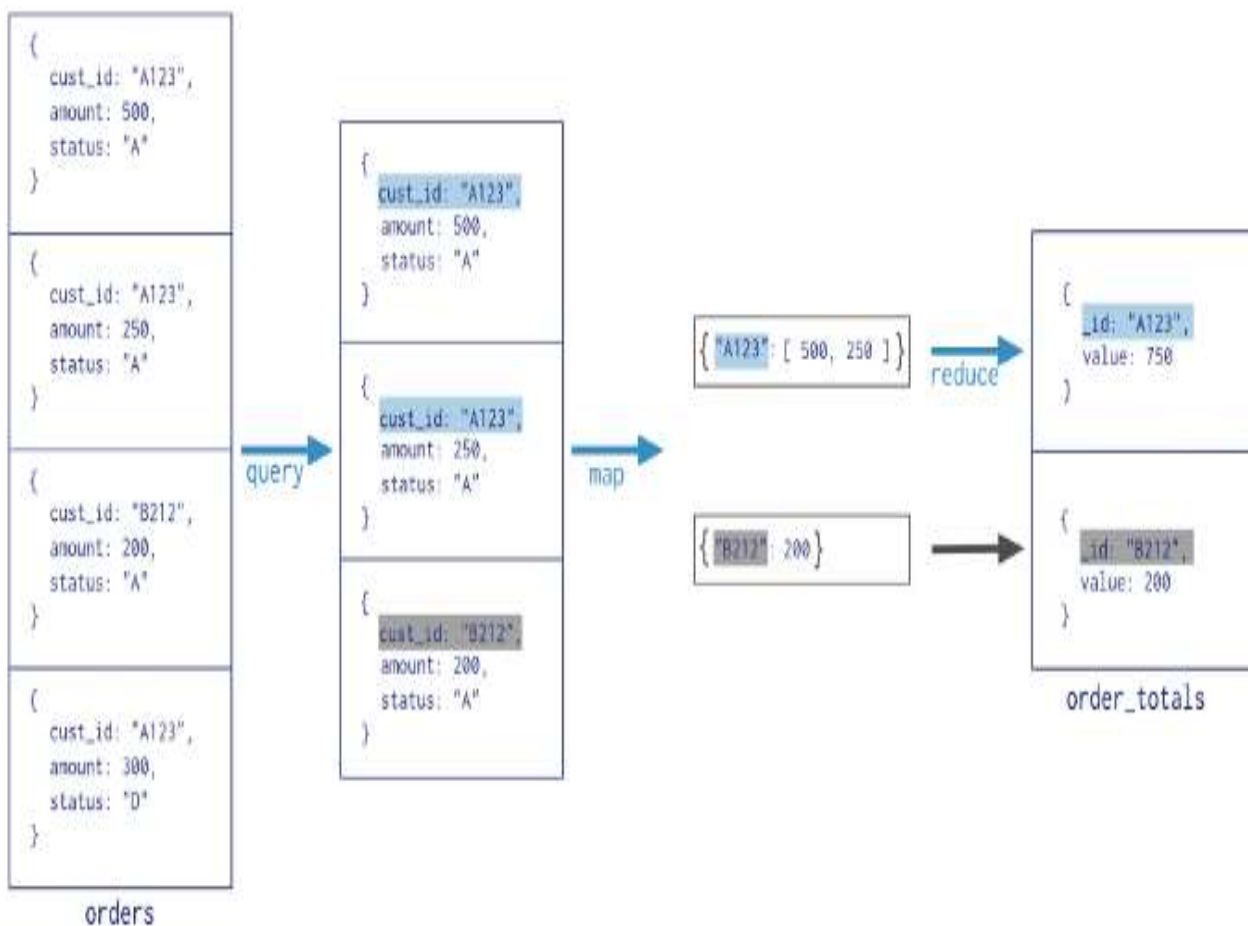
```

# convert count (currently a string) to int
try:
    count = int(count)
except ValueError:
    # count was not a number, so silently
    # ignore/discard this line
    Continue

```

MongoDB-da MapReduce

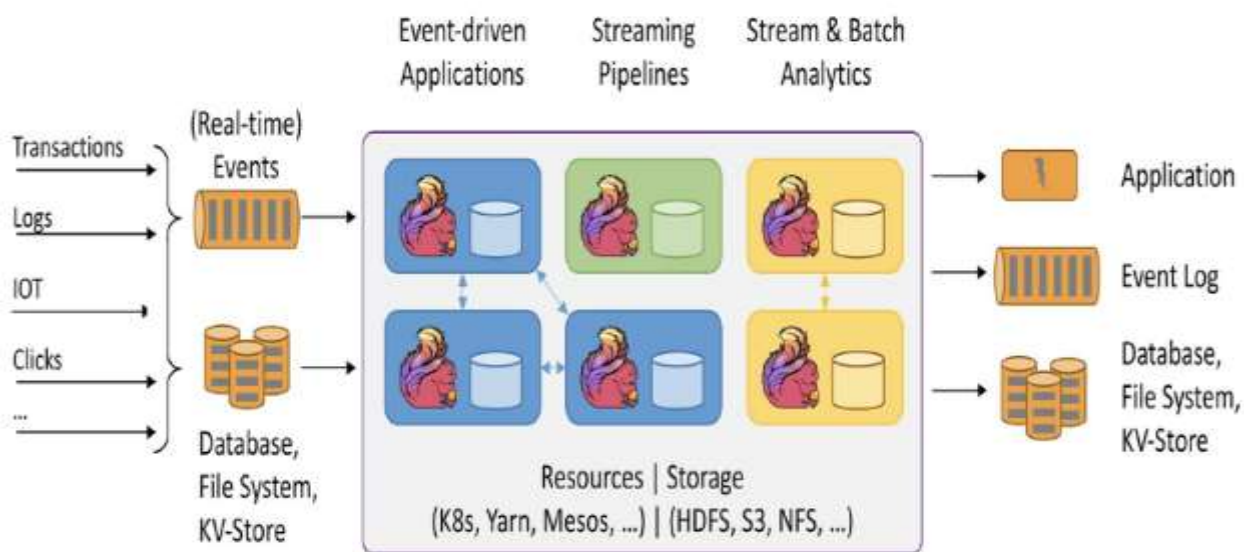
Биз шунингдек MapReduce маълумотлар базаси буйруғи орқали MongoDB-da Map-Reduce- дан фойдаланишимиз мумкин. Қуйидаги хариталарни қисқартириш амалиётини кўриб чиқинг:



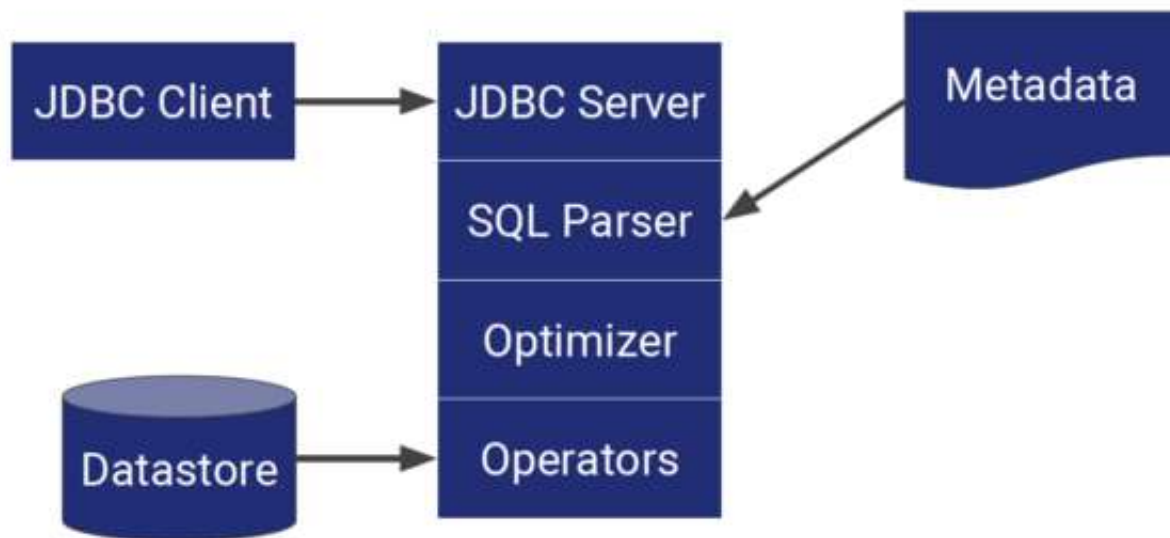
3.5-расм. MongoDB-da MapReduce

Бу харитани қисқартириш ишида MongoDB ҳар бир кириш ҳужжатиға (яъни сўров шартларига мос келадиган тўпладиги ҳужжатлар) харита фазасини қўллайди. Харита функцияси калит қиймат жуфтлигини чиқаради. Бир нечта қийматга эга бўлган ушбу калитлар учун MongoDB йиғилган маълумотларни тўплайдиган ва сиқиб чиқарадиган камайтириш фазасини қўллайди. Кейин MongoDB натижаларни тўпламда сақлайди. Ихтиёрий равишда, қисқартириш функциясининг чиқиши, кейинчалик йиғиш ёки йиғиш натижаларини қайта ишлаш учун яқунлаш функциясидан ўтиши мумкин.

MongoDB- да харитани қисқартиришнинг барча функциялари JavaScript- дир ва монгод жараёнида ишлайди. Харитани қисқартириш операциялари битта тўплам ҳужжатларини кириш сифатида қабул қилади ва харита босқичини бошлашдан олдин ихтиёрий тартибланиш ва чеклашни амалга ошириши мумкин. mapReduce харитани қисқартириш операциясининг натижаларини ҳужжат сифатида қайтариши ёки натижаларни тўпламларга ёзиши мумкин. Кириш ва чиқиш тўпламлари ўзгартирилиши мумкин.



3.6-расм. MongoDB-да MapReduce технологияси.



3.7-расм. MongoDB-da MapReduce тузилмаси.

Бошқа томондан, Салсите архитектураси миждозни, серверни ва тахлил қилувчини олиб ташлайди ва оптимизаторга метадата ишлашда оғир ишларни бажаришга имкон беради. Қалцит оптимизатори сўровларни оптималлаштириш учун 100 дан ортиқ қайта ёзиш қоидаларидан фойдаланади. Сўровлар реляцион алгебрадан фойдаланади, аммо рационал бўлмаган алгебрада ишлаши мумкин. Қалцит сўровни бажариш учун енг кам харажатли усулни топишга қаратилган.

Назорат саволлари:

1. Big data ни қайта ишлаш усуллари санаб беринг.
2. Google File System ишлаш принципи нимада?
3. Hadoop тақсимланган файл тизими нима?
4. Hadoop тақсимланган файл тизими нима учун керак?
5. Hadoop тақсимланган файл тизими қандай ишлайди?

6. MapReduce нима?
7. MapReduce нима учун керак?
8. MapReduce қандай ишлайди?

Адабиётлар ва интернет сайтлари:

1. Радченко И., Николаев И., Технологии и инфраструктура Big data: Учебное пособие. – Санкт-Петербург: ИТМО, 2018.
2. <https://computingforgeeks.com/data-mining-your-clicks>
3. <https://www.selecthub.com/big-data-analytics/big-data-analytics-requirements/>
4. <https://data-flair.training/blogs/hadoop-ecosystem-components/>
5. <https://www.digitalocean.com/community/tutorials/an-introduction-to-big-data-concepts-and-terminology>
6. <https://medium.com/cracking-the-data-science-interview/an-introduction-to-big-data-distributed-data-processing-36654202c6ce>

IV БЎЛИМ

АМАЛИЙ МАШҒУЛОТ
МАТЕРИАЛЛАРИ

IV. АМАЛИЙ МАШҒУЛОТ МАТЕРИАЛЛАРИ

1-амалий иш. Катта маълумотлар сўровларнинг Hive тизими орқали амалга ошириш. SQL тилига кириш, SQL дан HiveQLга, HIVEe дан HIVEQL га ўтиш. (4 соат)

Ишнинг мақсади: Катта маълумотлар сўровларнинг Hive тизими орқали амалга ошириш. SQL тилига кириш, SQL дан HiveQLга, HIVEe дан HIVEQL га ўтиш бўйича махсус компетенцияларни такомиллаштириш.

Назарий маълумот.

Hive - бу Hadoop Distributed File System (HDFS) устида ишлаб чиқилган ETL ва маълумотларни сақлаш воситаси ҳисобланади. Hive нинг асосий вазибаларидан қуйидагилар: Data encapsulation, Ad-hoc queries, Analysis of huge datasets каби операцияларни бажаришда ишни осонлаштиради.

Hive нинг муҳим хусусиятлари:

- ✓ Hive да аввал жадваллар ва маълумотлар базалари яратилади, сўнгра маълумотлар ушбу жадвалларга юкланади.
- ✓ Hive маълумотлар омбори сифатида жадвалларда сақланадиган маълумотларни бошқариш ва сўров ҳосил қилиш учун мўлжалланган.
- ✓ Структураланган маълумотлар билан ишлашда Map Reduceda UDF каби оптималлаштириш ва фойдаланиш хусусиятлари мавжуд емас, аммо Hive фреамворкда мавжуд. Сўровларни оптималлаштириш кўрсаткичлар бўйича сўровларни бажаришнинг самарали усулини билдиради.

Бу маълумотлар базаси дунёдаги таниш тушунчаларни, масалан, жадвалларни, сатрларни, устунларни ва схемаларни ва бошқаларни ўрганишда қулайлик учун қайта ишлатади. Hive нинг янги ва муҳим таркибий қисми, яъни схема маълумотларини сақлаш учун ишлатиладиган Metastore. Ушбу Metastore одатда реляцион маълумотлар базасида

жойлашган.

Биз Hive билан Web GUI ва Java Database Connectivity (JDBC) интерфейси каби усуллардан фойдаланишимиз мумкин:

Одатда, HQL синтаксиси кўпгина маълумотлар таҳлилчилари таниш бўлган SQL синтаксисига ўхшайди. Қуйидаги намуна сўровида кўрсатилган жадвал номидаги барча ёзувлар акс еттирилган.

Масалан : `Select * from <TableName>`

Hive TEXTFILE, SEQUENCEFILE, ORC ва RCFILE (Ёзув устунли файл) тўртта файл форматини қўллаб-қувватлайди.

Биз Hive га ўтишдан олдин, SQL га қисқача назар ташлаймиз ва MySQL МББТ дан фойдаланиб ишларни амалга оширамиз. SQL - бу маълумотлар базаларида маълумотларни сақлаш, бошқариш ва олиш учун стандарт тил.

DDL ва DML командаларидан фойдаланамиз.

Амалий вазифалар:

Сўров: `show databases;`

```
Enter password: *****
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 9
Server version: 5.6.16 MySQL Community Server (GPL)

Copyright (c) 2000, 2014, Oracle and/or its affiliates. All rights reserved.
Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| fqbfbf |
| fqbzbz |
| mysql |
| performance_schema |
| sakila |
| tatu |
| test |
| world |
+-----+
9 rows in set (0.00 sec)
```

Мавжуд маълумотлар базаларининг рўйхати

Сўров: create database фойдали_қазилма_бойитиш_заводи;

Натижа:

```
mysql> create database foydali_qazilma_boyitish_zavodi;
Query OK, 1 row affected (0.00 sec)

mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| foydali_qazilma_boyitish_zavodi |
| fqbfb |
| fqbzb |
| mysql |
| performance_schema |
| sakila |
| tatu |
| test |
| world |
+-----+
10 rows in set (0.00 sec)
```

Янги фойдали_қазилма_бойитиш_заводи номли базасининг яратилиши.

1. Олтин_рудаси номли жадвал яратамиз:

Сўров: create table олтин_рудаси (id int, konsentratsiyasi double, qazib_olingani varchar(20));

Натижа:

```
mysql> use foydali_qazilma_boyitish_zavodi;
Database changed
mysql> create table olting_rudasi ( id int, konsentratsiyasi double, qazib_olingani varchar(20));
Query OK, 0 rows affected (0.31 sec)

mysql> show tables from foydali_qazilma_boyitish_zavodi;
+-----+
| Tables_in_foydaali_qazilma_boyitish_zavodi |
+-----+
| olting_rudasi |
+-----+
1 row in set (0.00 sec)

mysql>
```

Олтин_рудаси номли жадвал яратилиши.

2. Жадвални кўрамиз:

Сўров: desc oltin_rudasi;

Натижа:

```
mysql> desc oltin_rudasi;
+-----+-----+-----+-----+-----+-----+
| Field          | Type          | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| id             | int(11)       | YES  |     | NULL    |       |
| konsentratsiyasi | double        | YES  |     | NULL    |       |
| qazib_olingani | varchar(20)   | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
3 rows in set (0.01 sec)
```

Яратилган Олтин_рудаси жадвалини кўриниши.

**Сўров: Alter table oltin_rudasi add sanasi
varchar(10);**

Натижа:

```
mysql> alter table oltin_rudasi add sanasi varchar(10);
Query OK, 0 rows affected (0.47 sec)
Records: 0 Duplicates: 0 Warnings: 0

mysql> desc oltin_rudasi;
+-----+-----+-----+-----+-----+-----+
| Field          | Type          | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| id             | int(11)       | YES  |     | NULL    |       |
| konsentratsiyasi | double        | YES  |     | NULL    |       |
| qazib_olingani | varchar(20)   | YES  |     | NULL    |       |
| sanasi         | varchar(10)   | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
4 rows in set (0.01 sec)
```

Олтин_рудаси жадвалига ALTER оператори орқали устун қўшилиши.

**Сўров: Alter table oltin_rudasi modify sanasi
varchar(30);**

Натижа:

```
mysql> alter table oltin_rudasi modify sanasi varchar(30);
Query OK, 0 rows affected (0.48 sec)
Records: 0 Duplicates: 0 Warnings: 0

mysql> desc oltin_rudasi;
+-----+-----+-----+-----+-----+-----+
| Field          | Type          | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| id             | int(11)       | YES  |     | NULL    |       |
| konsentratsiyasi | double        | YES  |     | NULL    |       |
| qazib_olingani | varchar(20)   | YES  |     | NULL    |       |
| sanasi         | varchar(30)   | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
4 rows in set (0.01 sec)
```

Олтин_рудаси жадвалида ALTER оператори орқали мавжуд устун тоифаси о`згартирилиши.

**Сўров: Alter table oltin_rudasi change konsentratsiyasi
ulushi double;**

Натижа:

```
mysql> alter table oltin_rudasi change konsentratsiyasi ulushi double;
Query OK, 0 rows affected (0.09 sec)
Records: 0 Duplicates: 0 Warnings: 0

mysql> desc oltin_rudasi;
+-----+-----+-----+-----+-----+-----+
| Field          | Type          | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| id             | int(11)       | YES  |     | NULL    |       |
| ulushi         | double        | YES  |     | NULL    |       |
| qazib_olingani | varchar(20)   | YES  |     | NULL    |       |
| sanasi         | varchar(30)   | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
4 rows in set (0.01 sec)
```

Олтин_рудаси жадвалида ALTER оператори орқали мавжуд устун номини ўзгартириш.

Назорат саволлари:

1. CREATE буйругининг вазифалари нималардан иборат?
2. ALTER буйругига мисол келтиринг. DROP va DELETE буйруқларининг фарқи нимада?
3. DELETE буйругининг структурасини кўрсатинг.

Адабиётлар ва интернет сайтлари:

1. Радченко И., Николаев И., Технологии и инфраструктура Big data: Учебное пособие. – Санкт-петербург: ИТМО, 2018.
2. <https://computingforgeeks.com/data-mining-your-clicks>
3. <https://www.selecthub.com/big-data-analytics/big-data-analytics-requirements/>
4. <https://data-flair.training/blogs/hadoop-ecosystem-components/>
5. <https://www.digitalocean.com/community/tutorials/an-introduction-to-big-data-concepts-and-terminology>

2-амалий иш. HIVE e HIVEQL га кириш, Hadoop файлларни сураш учун Hive-дан фойдаланиш (2 соат)

Ишнинг мақсади: тингловчиларнинг HIVE e HIVEQL га кириш, Hadoop файлларни сураш учун Hive-дан фойдаланиш бўйича омптенецияларини такомиллаштириш.

Назарий маълумот

Hadoop тизимларида ишлайдиган Apache Hive фоситаси HDFS-да сақланган маълумотни худди SQL билан боғлиқ маълумотлар базаси каби сўроқ қилишга имкон беради. Hive - бу MapReduce- нинг юқори қисмидаги абстракция, бу HiveQL деб номланган SQL- га жуда ўхшаш тилда баёнотлар ёрдамида ишларни яратишга имкон беради. Hive-дан фойдаланиш Java-да ўзингизнинг MapReduce ишларини ёзишдан кўра анча тезроқ ва осонроқдир. Шунини ёдда тутишимиз керакки, Hive HDFS маълумотларини реляцион маълумотлар базасига айлантирмайди, бу уларга фақат HiveQL сўровлари ёрдамида кириш имконини беради. Hive ни қандай ишлатиш керак?

Уяни қандай ишлатиш керак? Hue (Hadoop User Experience) ёрдамида веб-браузерда Hive- дан фойдаланишимиз мумкин, буйруқ сатри орқали "Билайн" воситаси ёрдамида ёки дастурий равишда бунинг учун мос JDBC ёки ODBC улагичидан фойдаланинг. Ҳар ҳолда, IP-манзилни, портни ва Hive- га кириш маълумотларини кўрсатиш керак бўлади. Hive лар жадваллар билан ишлайди. Шундай қилиб, биринчи нарса, бу жадвалларни Hive-да белгилашдир. Жадвал яратиш учун биз унга ном беришимиз ва ҳар бир устуннинг майдонларини кўрсатишимиз керак (худди SQL маълумотлар базасида бўлгани каби). Бундан ташқари, биз юкланадиган маълумотлар жойлашган каталогни ва улар қайси форматда сақланишини белгилашимиз керак (сукут бўйича у `"/ user / hive / warehouse / <tablename>" katalogida bo'ladi`). Жадвал яратишда биз белгилайдиган майдонларнинг тузилиши

бизнинг файлларимиздаги маълумотлар майдонларига мос келишига еҳтиёт бўлишимиз керак. Шу тарзда, Hive маълумотлар тузилишини ва ушбу маълумотлар қайерда сақланишини кўрсатадиган метадата яратади, шунинг учун биз хоҳлаганча маълумот сўрашимиз мумкин. Худди шу тарзда, реляцион маълумотлар базасида биз жадвалларни бирлаштириш ва баъзи SQL функцияларидан фойдаланиш орқали мураккаб сўровларни бажаришимиз мумкин.

Амалий вазифалар

Beeline мисолида

```
beeline -n <usr> -p <pwd> \  
-u jdbc:hive2://<host>:<port>/<db>
```

Beeline билан HiveQL сўровини ишга тушириш

```
beeline -n <usr> -p <pwd> \  
-u jdbc:hive2://<host>:<port>/<db>  
-e "<query>"
```

HiveQL сўровини " Beeline " билан файлдан ишга тушириш

```
beeline -n <usr> -p <pwd> \  
-u jdbc:hive2://<host>:<port>/<db>  
-f <file.hql>
```

HiveQL командалар Show tables

```
SHOW TABLES;
```

Жадвал таснифи

```
DESCRIBE <table>;
```

Comments (only Hue and Scripts)

Using “–“

Hive функцияларини кўрсатиш

```
SHOW FUNCTIONS;
```

Функция таснифи

```
DESCRIBE FUNCTION <function>;
```

Функцияни тўлиқ таснифи

```
DESCRIBE FUNCTION EXTENDED <function>;
```

Кўп фойдаланувчи командалар

```
ROUND(<number>,<n-decimals>)
```

```
CEIL(<number>)
```

```
FLOOR(<number>)
```

```
YEAR(<timestamp>)
```

```
SUBSTRING(<string>,<ini>,<end>)
```

```
TO_UTC_TIMESTAMP(<timestamp>,'UTC')
```

```
CAST(<val> as <type>)
```


Агрегат функциялар GROUP BY

COUNT(*)

COUNT(<col>)

COUNT(DISTINCT <col>)

MAX(<col-number>)

MIN(<col-number>)

SUM(<col-number>)

AVG(<col-number>)

Hive да устун тоифалари

STRING

BOOLEAN

TIMESTAMP

INT

BIGINT

FLOAT

DOUBLE

Hive да жадвал яратиш

CREATE TABLE <table> (

<name> <type>,

...

);

Бошқа майдон ажратувчиси билан жадвал яратинг

```
CREATE TABLE <table> (  
    <name> <type>,  
    ...  
)  
  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY '<symb>';
```

SequenceFile файлидан жадвал яратиш

```
CREATE TABLE <table> (  
    <name> <type>,  
    ...  
)  
  
STORED AS SEQUENCEFILE;
```

Avro файлидан жадвал яратиш

```
CREATE TABLE <table> (  
    <name> <type>,  
    ...  
)  
  
STORED AS AVRO;
```

Жадвални ўчириш

```
DROP TABLE <table>;
```

Агар жадвал мавжуд бўлса жадвални ўчириш

```
DROP TABLE IF EXISTS <table>;
```

HDFS каталогини кўрсатадиган жадвал яратинг

```
CREATE TABLE <table> (
```

```
  <name> <type>,
```

```
  ...
```

```
)
```

```
LOCATION '<dir>;'
```

Ташқи жадвал яратиш

```
CREATE EXTERNAL TABLE <table> (
```

```
  <name> <type>,
```

```
  ...
```

```
)
```

```
LOCATION '<dir>;'
```

Жадвал каталогдаги жадвалларни нусхалаш файлларини тўлдириш

```
hdfs dfs -mv <file> <table-dir>
```

Жадвални HiveQL буйруғи билан тўлдириш

```
LOAD DATA INPATH '<file>' INTO TABLE <table>;
```

Мавжуд жадвалга қатор қўшиш

```
INSERT INTO TABLE <table>
```

```
<query>
```

```
CREATE TABLE <table> AS
```

```
<query>
```

```
CREATE TABLE <table>
```

```
STORED AS <format>
```

```
AS
```

```
<query>
```

```
1 |
2 | user@ubuntu1:~$ hive --define address='country, state, city, post'
3 | hive> select ${address} from temp_user limit 5;
4 | OK
5 | country      state  city      post
6 | AU           TA     Leith     7315
7 | AU           QL     Proston   4613
8 | AU           WA     Hame      1 6215
9 | AU           NS     Talmalmo 2640
10 | Time taken: 0.57 seconds, Fetched: 5 row(s)
11 | hive>
```

```
1
2 user@ubuntu1 $ hive --database testdb
3
4         or
5
6 hive> use testdb;
```

```
1
2 $ export VAR1='London'
3 $ hive -S -e "SELECT firstname, state, city FROM temp_user WHERE city='${VAR1}'" > temp
```

```
user@tri03ws-386:~$ export VAR1='London'
user@tri03ws-386:~$ hive -S -e "SELECT firstname, state, city FROM temp user WHERE city='${VAR1}'" > temp_out.txt
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/user/bigdata/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/user/bigdata/apache-hive-0.14.0-bin/lib/hive-jdbc-0.14.0-standalone.jar!/org/apache/hive/jdbc/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/user/bigdata/hbase-0.98.8-hadoop2/lib/phoenix-4.2.2-client.jar!/org/apache/derby/impl/client/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
user@tri03ws-386:~$ cat temp_out.txt
Laurena ON London
Felix ON London
Hershel ON London
Nan ON London
Bette ON London
Effie ON London
Tamar OH London
user@tri03ws-386:~$
```

```
1
2 user@tri03ws-386:~$ hive -h 172.16.3.78 -p 10000
3
4 Logging initialized using configuration in jar:file:/home/user/bigdata/apache-hive-0.14.0-bin/lib/hive-common-0.14.0.jar!/hive-log4j.properties
5 .....
6 [172.16.3.78:10000] hive> show tables;
7 adapter
8 mytest
9 page_view
10 page_view14
11 emp
12 .....
13 [172.16.3.78:10000] hive> dfs -ls / ;
14 Found 15 items
15 -rw-r--r-- 1 tri03 supergroup 138 2014-10-22 13:15 /Sample.txt
16 drwxr-xr-x - tri03 supergroup 0 2014-11-05 12:32 /adapter
17 drwxr-xr-x - tri03 supergroup 0 2014-11-05 11:31 /adapter3
18 drwxr-xr-x - tri03 supergroup 0 2014-11-05 12:24 /adapter4
19 .....
```

```
1
2 $ hive -f /home/user/test.hql
```

```
1
2 $ hive -f hdfs://localhost:9000/test.hql
```

Мустақил ишлаш вазифалари:

1. DML комалдаларини Hive да бажаринг.
2. Агрегат функцияларидан фойдаланиб сўров ёзинг.
3. Жадвал устунлари тоифасини ўзгартирувчи сўров ёзинг.

Адабиётлар ва интернет сайтлари:

1. Радченко И., Николаев И., Технологии и инфраструктура Big data: Учебное пособие. – Санкт-петербург: ИТМО, 2018.
2. <https://computingforgeeks.com/data-mining-your-clicks>

3. <https://www.selecthub.com/big-data-analytics/big-data-analytics-requirements/>
4. <https://data-flair.training/blogs/hadoop-ecosystem-components/>
5. <https://www.digitalocean.com/community/tutorials/an-introduction-to-big-data-concepts-and-terminology>

3-амалий иш. Hadoop тизимига маълумотларни ўтказиш. Hadoop тизимида файлларни қайта ишлаш. Hadoop тизимидан маълумотларни олиш (4 соат)

Ишнинг мақсади: Hadoop тизимига маълумотларни ўтказиш, Hadoop тизимида файлларни қайта ишлаш ва Hadoop тизимидан маълумотларни олиш бўйича махсус компетенцияларни такомиллаштириш.

Назарий маълумот.

Hadoop -ни GNU / Linux платформаси қўллаб-қувватлайди, шунинг учун Hadoop муҳитини ўрнатиш учун Linux операцион тизимини ўрнатишимиз керак. Агар Linuxдан бошқа операцион тизим мавжуд бўлса, унда бизга Virtualbox дастурини биз фойдаланаётган OT га ўрнатишимиз ва Virtualboxга Linux OT ни ўрнатиш талаб етилади. Hadoop-ни Linux муҳитига ўрнатишдан олдин ssh (Secure Shell) ёрдамида Linux ни ўрнатишимиз керак. Linux муҳитини созлаш учун қуйида келтирилган амалларни бажарамиз.

1. Фойдаланувчини яратиш: Аввало, Hadoop файл тизимини Unix файл тизимидан ажратиш учун Hadoop учун алоҳида фойдаланувчи яратиш тавсия етилади.

Фойдаланувчини яратиш учун қуйидаги амалларни бажарамиз.

- `su` командаси билан `root` ни очамиз,
- "`Useradd username`" буйруғи ёрдамида `root` дан фойдаланувчи яратамиз.
- "`su username`" буйруғи ёрдамида мавжуд фойдаланувчини ишга тушириш мумкин.

Linux терминалини очамиз ва фойдаланувчи яратиш учун қуйидаги буйруқларни бажарамиз:

```
$ su
```

```
password:
```



```
# useradd hadoop
# passwd hadoop
New passwd:
Retype new passwd
```

SSH созламалари ва калитларни яратиш.

SSH созламаларида ишга тушириш, тўхтатиш, тарқатилган демон қобиғи операциялари каби турли хил операцияларни бажариш талаб қилинади. Hadoop турли хил фойдаланувчиларининг ҳақиқийлигини текширишда Hadoop фойдаланувчиси учун очик / шахсий калит жуфтлигини тақдим этиш ва уни турли фойдаланувчилар билан бўлишиш талаб қилинади.

```
$ ssh-keygen -t rsa
$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
$ chmod 0600 ~/.ssh/authorized_keys
```

Қуйидаги командалар орқали Hadoop ни ўрнатамиз.

```
$ su
password:
# cd /usr/local
# wget http://apache.claz.org/hadoop/common/hadoop-2.4.1/
hadoop-2.4.1.tar.gz
# tar xzf hadoop-2.4.1.tar.gz
# mv hadoop-2.4.1/* to hadoop/
# exit
```

Hadoop ишлаш тартиби.

Hadoop- ни юклаб олганингиздан сўнг, Hadoop кластерингизни қўллаб-

қувватланадиган учта режимдан бирида ишлатишингиз мумкин бўлади.

Local/ мустақил режим - тизимга Hadoop-ни юклаб олгандан сўнг, у мустақил режимда ишлатилиши мумкин.

Pseudo тақсимланган режим - бу битта машинада тақсимланган симуляция ҳисонланади.

Тўла тақсимланган режим - ушбу режим кластер сифатида камида икки ёки ундан ортиқ машиналар билан тўлиқ тақсимланади

Hadoopни Local//мустақил режимда ўрнатиш. Бунда биз Hadoop 2.4.1 ни ўрнатамиз.

~ / .Bashrc файлига қуйидаги буйруқларни қўшиш орқали Hadoop муҳит ўзгарувчиларини ўрнатиш мумкин.

```
export HADOOP_HOME=/usr/local/hadoop
```

Ўрнатишни давом етишдан олдин, Hadoop яхши ишлаётганига ишонч ҳосил қилиш керак.

```
$ hadoop version
```

```
Hadoop 2.4.1
```

```
Subversion https://svn.apache.org/repos/asf/hadoop/common -r 1529768
```

```
Compiled by hortonmu on 2013-10-07T06:28Z
```

```
Compiled with protoc 2.5.0
```

```
From source with checksum 79e53ce7994d1628b240f09af91e1af4
```

Hadoop нинг оддий мисолда текшириб кўрайлик. Hadoop ўрнатилиши MapReduce .jar файлини тақдим этади, бу MapReduce-нинг асосий функциясини таъминлайди ва Pi қийматини, файлларнинг маълум бир рўйхатидаги сўзларни ҳисоблашда ва бошқаларни ҳисоблашда ишлатилади.

```
$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.2.0.jar
```

Файллардаги сўзларнинг умумий сонини ҳисоблаш масаласини кўриб чиқамиз.

Сўзларнинг умумий сонини ҳисоблаш учун .jar файлида сўзларни ҳисоблаш учун дастур мавжуд бўлса, MapReduce-ни ёзишимиз шарт эмас. Худди шу .jar файлидан фойдаланиб, бошқа мисолларни синаб кўришингиз мумкин. MapReduce функционал дастурларини *hadoop-mapreduce-sample-2.2.0.jar* файли орқали текшириш учун қуйидаги буйруқлар бажарилади:

```
$ hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-  
apreduceexamples-2.2.0.jar
```

1- қадам.

Кириш каталогида вақтинчалик таркиб файлларини яратамиз. Биз ушбу каталогни хоҳлаган жойда яратишимиз мумкин.

```
$ mkdir input  
$ cp $HADOOP_HOME/*.txt input  
$ ls -l input
```

Бу бизга кириш каталогимизда қуйидаги файлларни беради.

```
total 24  
-rw-r--r-- 1 root root 15164 Feb 21 10:14 LICENSE.txt  
-rw-r--r-- 1 root root 101 Feb 21 10:14 NOTICE.txt  
-rw-r--r-- 1 root root 1366 Feb 21 10:14 README.txt
```

2- кадам.

Кириш каталогида мавжуд бўлган барча файллардаги сўзларнинг умумий сонини ҳисоблаш учун Hadoop жараёнини бошлаймиз.

```
$ hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-  
mapreduceexamples-2.2.0.jar wordcount input output
```

3- кадам.

Иккинчи кадамда керакли ишларни бажаради ва чиқаришни *output / part-r
00000* файлида сақлайди.

```
$cat output/*
```

Қуйида барча сўзлар ва уларнинг каталогидаги барча файллардаги умумий сонлар рўйхати келтирилган.

"AS 4

"Contribution" 1

"Contributor" 1

"Derivative 1

"Legal 1

"License" 1

"License"); 1

"Licensor" 1

"NOTICE" 1

"Not 1

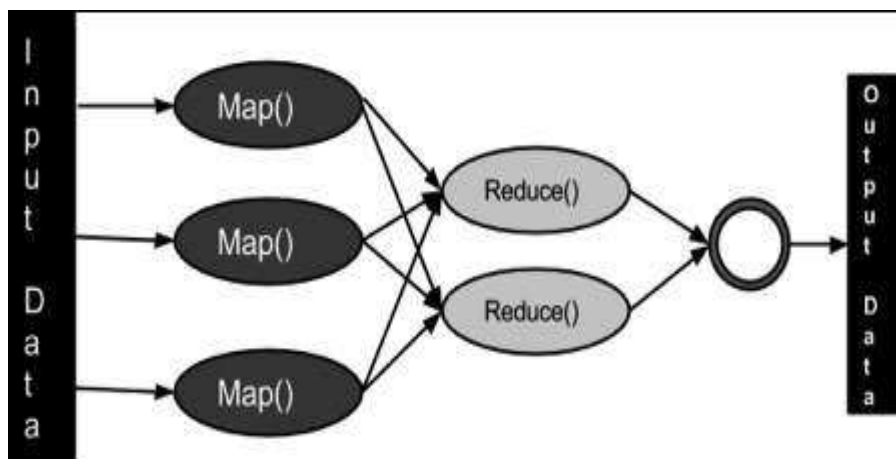
"Object" 1

"Source" 1

"Work" 1
"You" 1
"Your") 1
"[]" 1
"control" 1
"printed" 1
"submitted" 1
(50%) 1
(BIS), 1
(C) 1
(Don't) 1
(ECCN) 1
(INCLUDING 2
(INCLUDING, 2

Ушбу жараён шу билан яқунланади.

MapReduce - бу ишлов бериш техникаси ва жава асосида тарқатилган ҳисоблаш учун дастур модели ҳисобланади. MapReduce алгоритми иккита муҳим вазифани ўз ичига олади, яъни Map ва Reduce. Map маълумотлар тўпламини олади ва уларни бошқа маълумотлар тўпламига ўзгартиради, бу йерда алоҳида элементлар катакчаларга бўлинади (калит / қиймат жуфтлари). Иккинчидан, Map дан чиқишни кириш сифатида қабул қиладиган ва маълумотлар катакчаларини кичикроқ катакчаларга бирлаштириш вазифани бажаради. MapReduce номининг кетма-кетлигини назарда тутганидек, камайтириш(Reduce) вазифаси ҳар доим харита ишидан кейин бажарилади.



1 – расм. Алгоритм.

Кириш ва чиқариш.

MapReduce framework $\langle \text{key}, \text{value} \rangle$ жуфтлари устида ишлайди, яъни framework ишга киришни $\langle \text{key}, \text{value} \rangle$ жуфтлари тўплами сифатида кўриб чиқади ва ишнинг натижаси сифатида $\langle \text{key}, \text{value} \rangle$ жуфтлари тўпамини ҳосил қилади. Калит ва қиймат синфлари framework бўйича кетма-кетликда бўлиши керак. Бундан ташқари, асосий синфлар framework бўйича саралашни осонлаштириш учун Writable-Comparable интерфейсини амалга ошириши керак. MapReduce ишининг кириш ва чиқиш турлари - (Кириш) $\langle k1, v1 \rangle \rightarrow \text{map} \rightarrow \langle k2, v2 \rangle \rightarrow$ камайтириш $\rightarrow \langle k3, v3 \rangle$ (Чиқиш).

	Input	Output
Map	$\langle k1, v1 \rangle$	list ($\langle k2, v2 \rangle$)
Reduce	$\langle k2, \text{list}(v2) \rangle$	list ($\langle k3, v3 \rangle$)

Қуйидаги мисолда биз ташкилотнинг электр энергиясини истеъмол қилишига оид маълумотларни келтириб ўтганмиз. Унда ойлик электр энергиясини

истеъмол қилишни ва ҳар хил йиллар учун йиллик ўртача қийматни ўз ичига олади.

Yilla	Ya	Fe	M	Ap	Ma	Iyu	Iyu	Av	Se	Ok	No	De	O'rt
r	n	b	ar	r	y	n	l	g	n	t	y	k	acha
2010	23	23	2	43	24	25	26	26	26	26	25	26	25
2011	26	27	28	28	28	30	31	31	31	30	30	30	29
2012	31	32	32	32	33	34	35	36	36	34	34	34	34
2013	39	38	39	39	39	41	42	43	40	39	38	38	40
2014	38	39	39	39	39	41	41	41	00	40	39	39	45

Агар юқоридаги маълумотлар кириш маълумоти сифатида берилган бўлса, биз уни қайта ишлаш учун дастур ёзишимиз ва максимал фойдаланиш йилини, минимал фойдаланиш йилини ва бошқаларни топишимиз керак бўлади. Лекин бизга катта ташкилотнинг бир неча йиллик маълумотлари, электр истемоли ҳақида катта маълумотлар устида ишлашга тўғри келса, Уларни бажаришга кўп вақт талаб етилади. Маълумотни манбадан тармоқ серверига кўчиришда ва ҳоказоларда кўп вақт истемоли бўлади. Қуйида MapReduce framework дан фойдаланган ҳолда намунавий маълумот учун дастур коди берилган.

```
package hadoop;
import java.util.*;
import java.io.IOException;
```

```

import java.io.IOException;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.conf.*;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;
import org.apache.hadoop.util.*;
public class ProcessUnits {
    //Mapper class
    public static class E_EMapper extends
MapReduceBase implements
Mapper<LongWritable /*Input key Type */
Text, /*Input value Type*/
Text, /*Output key Type*/
IntWritable> /*Output value Type*/
{
    //Map function
    public void map(LongWritable key, Text
value,
OutputCollector<Text, IntWritable> output,
Reporter reporter) throws IOException {
        String line = value.toString();
        String lasttoken = null;
        StringTokenizer s = new
StringTokenizer(line, "\t");
        String year = s.nextToken();
        while(s.hasMoreTokens()) {
            lasttoken = s.nextToken();
        }
    }
}

```



```

        int avgprice =
Integer.parseInt(lasttoken);
        output.collect(new Text(year), new
IntWritable(avgprice));
    }
}
//Reducer class
public static class E_EReduce extends
MapReduceBase implements Reducer< Text,
IntWritable, Text, IntWritable > {

//Reduce function
public void reduce( Text key, Iterator
<IntWritable> values,
OutputCollector<Text, IntWritable> output,
Reporter reporter) throws IOException {
    int maxavg = 30;
    int val = Integer.MIN_VALUE;

    while (values.hasNext()) {
        if((val = values.next().get())>maxavg) {
            output.collect(key, new
IntWritable(val));
        }
    }
}
}
//Main function

```

```

    public static void main(String args[])throws
Exception {
        JobConf conf = new
JobConf(ProcessUnits.class);
        conf.setJobName("max_electricityunits");
        conf.setOutputKeyClass(Text.class);
        conf.setOutputValueClass(IntWritable.class);
        conf.setMapperClass(E_EMapper.class);
        conf.setCombinerClass(E_EReduce.class);
        conf.setReducerClass(E_EReduce.class);
        conf.setInputFormat(TextInputFormat.class);

conf.setOutputFormat(TextOutputFormat.class);

        FileInputFormat.setInputPaths(conf, new
Path(args[0]));
        FileOutputFormat.setOutputPath(conf, new
Path(args[1]));
        JobClient.runJob(conf);
    }
}

```

Юқоридаги дастур кодини ProcessUnits.java сифатида сақланг. Дастурнинг компиляцияси ва бажарилиши эса қуйидагича бўлади. Жараён бирликлари дастурини тузиш ва бажариш. Биз тасаввур қилайлик Hadoop фойдаланувчисининг уй каталогидамиз. (масалан /home / hadoop).

Юқоридаги дастурни компиляция қилиш ва бажариш учун қуйида келтирилган амалларни бажаришимиз зарур.

1- кадам

Қуйидаги буйрук - тузилган жадвал синфларини сақлаш учун каталог яратиш.

```
$ mkdir units
```

2- кадам

MapReduce дастурини компиляция қилиш ва бажариш учун ишлатиладиган Hadoop-core-1.2.1.jar дастурини юклаб олинг. Банкни юклаб олиш учун қуйидаги ҳавола mvnrepository.com га ташриф буюринг. Юкланган папка / home / hadoop / деб тахмин қилайлик.

3- кадам

ProcessUnits.java дастурини компиляция қилиш ва дастур учун жар яратиш учун қуйидаги буйруқлардан фойдаланилади.

```
$ javac -classpath hadoop-core-1.2.1.jar -d units ProcessUnits.java
```

```
$ jar -cvf units.jar -C units/ .
```

```
$HADOOP_HOME/bin/hadoop fs -mkdir input_dir
```

```
$HADOOP_HOME/bin/hadoop fs -ls input_dir/
```

```
$HADOOP_HOME/bin/hadoop jar units.jar hadoop.ProcessUnits input_dir  
output_dir
```

Амалга оширилгандан сўнг, қуйида кўрсатилгандек, чиқиш/киришнинг бўлиниш сони, Мар вазифалари сони, редуктор вазифалари сони ва бошқаларни кўриш мумкин бўлади.

```
INFO mapreduce.Job: Job job_1414748220717_0002  
completed successfully  
14/10/31 06:02:52
```

```
INFO mapreduce.Job: Counters: 49
    File System Counters
FILE: Number of bytes read = 61
FILE: Number of bytes written = 279400
FILE: Number of read operations = 0
FILE: Number of large read operations = 0
FILE: Number of write operations = 0
HDFS: Number of bytes read = 546
HDFS: Number of bytes written = 40
HDFS: Number of read operations = 9
HDFS: Number of large read operations = 0
HDFS: Number of write operations = 2 Job
Counters
    Launched map tasks = 2
    Launched reduce tasks = 1
    Data-local map tasks = 2
    Total time spent by all maps in occupied
slots (ms) = 146137
    Total time spent by all reduces in occupied
slots (ms) = 441
    Total time spent by all map tasks (ms) =
14613
    Total time spent by all reduce tasks (ms) =
44120
    Total vcore-seconds taken by all map tasks =
146137
    Total vcore-seconds taken by all reduce tasks
= 44120
```

Total megabyte-seconds taken by all map tasks
= 149644288

Total megabyte-seconds taken by all reduce
tasks = 45178880

Map-Reduce Framework

Map input records = 5

Map output records = 5

Map output bytes = 45

Map output materialized bytes = 67

Input split bytes = 208

Combine input records = 5

Combine output records = 5

Reduce input groups = 5

Reduce shuffle bytes = 6

Reduce input records = 5

Reduce output records = 5

Spilled Records = 10

Shuffled Maps = 2

Failed Shuffles = 0

Merged Map outputs = 2

GC time elapsed (ms) = 948

CPU time spent (ms) = 5160

Physical memory (bytes) snapshot = 47749120

Virtual memory (bytes) snapshot = 2899349504

Total committed heap usage (bytes) =
277684224

File Output Format Counters

Bytes Written = 40

Чиқиш папкасида олинган файлларни текшириш учун қуйидаги буйруқ ишлатилади.

```
$HADOOP_HOME/bin/hadoop fs -ls output_dir/
```

Part-00000 файлида чиқишни кўриш учун қўйидаги буйруқ ишлатилади:

```
$HADOOP_HOME/bin/hadoop fs -cat output_dir/part-00000
```

Қуйида MapReduce дастури томонидан ишлаб чиқарилган маҳсулот келтирилган.

	2010		34
	2011		40
	2012		45

Барча Hadoop буйруқлари \$ HADOOP_HOME / bin / hadoop буйруғи билан чақирилади.

Статусни кўриш учун эса:

```
$ $HADOOP_HOME/bin/hadoop job -status <JOB-ID>
```

ёки

```
$ $HADOOP_HOME/bin/hadoop job -status job_201310191043_0004
```

Таҳрирни кўриш учун:

```
$ $HADOOP_HOME/bin/hadoop job -history <DIR-NAME>
```

ёки

```
$ $HADOOP_HOME/bin/hadoop job -history /user/expert/output
```

буйруқлари ишлатилади.

Назорат саволлари:

1. Hadoop нима?
2. Hadoop нинг асосий вазифаларини келтиринг.
3. Hadoop дан текшириш буйруқлари қандай амалга оширилади?

4. MapReduce дастурини компиляция қилиш ва бажариш деганда нимани тушунасиз?

Адабиётлар ва интернет сайтлари:

1. Радченко И., Николаев И., Технологии и инфраструктура Big data: Учебное пособие. – Санкт-петербург: ИТМО, 2018.
2. <https://computingforgeeks.com/data-mining-your-clicks>
3. <https://www.selecthub.com/big-data-analytics/big-data-analytics-requirements/>
4. <https://data-flair.training/blogs/hadoop-ecosystem-components/>
5. <https://www.digitalocean.com/community/tutorials/an-introduction-to-big-data-concepts-and-terminology>

У БЎЛИМ

КЕЙСЛАР БАНКИ

V. КЕЙСЛАР БАНКИ

1-кейс.

“Катта маълумотларни ўқув жараёнида қўллаш усуллари”

Вазият тавсифи: Катта маълумотлардан ўқув жараёнида фойдаланиш концепциясини ишлаб чиқиш вазифасини олдингиз. Нима киласиз?

Кейс саволлари:

- 1) Катта маълумотлар тушунчасига таъриф беринг?
- 2) Катта маълумотларнинг хусусиятларини санаб беринг?
- 3) Катта маълумотларни таълим тизимида қўллашга мисоллар келтиринг:

<i>№</i>	<i>Мисол</i>	<i>Натижаси</i>	<i>Изоҳ</i>
1			
2			
3			
4			

5			
6.			
7.			

4) Катта маълумотлардан ўқув жараёнида фойдаланиш концепциясини ишлаб чиқиш.

5) Катта маълумотлардан ўқув жараёнида фойдаланиш концепциянгизни яхшилаш учун SCAMPER усули асосида саволларга жавоб беринг.

SUBSTITUTE (АЛМАШТИРИШ)	Нима билан алмаштириш мумкин?	
COMBINE (БИРЛАШТИРИШ)	Нималарни бирлаштириш мумкин?	
ADAPT (МОСЛАШТИРИШ)	Нимага мослаштириш мумкин?	
MODIFY (МОДИФИКАЦИЯ)	Қандай яхшилашим мумкин? (ортиқча ишланган, етмаяпти)	
PUT TO OTHER USES (БОШҚА СОҲАЛАРДА)	Нима ўзгариши мумкин? (маълумотни шакли, белгилар, ранг ва	

ҚЎЛЛАШ)	бошқалар)	
ELEMINATE (ҚИСҚАРТИРИШ)	Яна қандай ҳолда қўллаш мумкин?	
REARRANGE/REVERSE (ТАРТИБИНИ ЎЗГАРТИРИШ)	Нимани қайта тиклаш мумкин? (Буюртмани ўзгартириш, компонентларни алмаштириш)	

VI БЎЛИМ

ГЛОССАРИЙ

VI. ГЛОССАРИЙ

Тушунча ўзбек тилида	Таъриф	Тушунча инглиз тилида
Big Data	бу катта ҳажм, тезлик, хилма-хиллик ва ишончлилиқ билан бошқариладиган маълумотлар тўпламидир	Big Data
NoSQL маълумотлар базалари	ўзаро боғлиқ бўлмаган, SQL-га асосланган эмас ва маълумотни ўзаро боғлиқ бўлмаган маълумотлар билан яхши ишлайдиган калит-қиймат жуфтликларида сақлайди	NoSQL databases
Modeling, Data Mining, Data File Sources, File Exporting	Маълумотларни қайта ишлаш технологиялари	Modeling, Data Mining, Data File Sources, File Exporting
Big Data Analytics воситалари	фойдаланувчиларга турли хил таҳлил пакетлари ва модулларини таклиф қилади	Big Data Analytics tools
Статистик таҳлилнинг 1-чи босқичи	Маълумотларнинг моҳиятини тавсифлаш, маълумотларни тақдим этган шачс билан боғлиқликни ўрганиш;	Phase 1 of statistical analysis
Статистик таҳлилнинг 2-чи босқичи	Уланишларни умумлаштириш учун модель яратиш;	Phase 2 of statistical analysis

Статистик таҳлилнинг 3-чи босқичи	Тўғрилигини исботлаш;	Phase 3 of statistical analysis
Статистик таҳлилнинг 4-чи босқичи	Рад этиш;	Phase 4 of statistical analysis
Статистик таҳлилнинг 5-чи босқичи	Қарорларни бошқариш учун башоратли таҳлилларни қўллаш.	Phase 5 of statistical analysis
Тақсимланган файл тизими (Distributed File System):	маълумотни уланган сақлаш мосламалари тизими орқали кириш мумкин бўлган форматда сақлашга имкон беради.	Distributed File System:
Volume	физик хажмнинг катталиги.	Volume
Velocity	натижаларни олиш учун тезликнинг ошиши ва тезкор ишлов беришнинг тезлиги.	Velocity
Variety	турли типдаги маълумотларга бир вақтда ишлов бериш имконияти.	Variety
Google Fayl System (GFS)	Google томонидан ишлатиладиган тақсимланган файл тизими	Google File System (GFS)
2003 йилда	Google File System чиқди.	In 2003

2004 йилда	MapReduce асослари чиқарилди.	In 2004
2006 йилда	MapReduce дастурий моделидан фойдаланган ҳолда катта маълумотларни тарқатиш ва қайта ишлаш учун дастурий таъминотни тақдим етувчи Hadoop ташкил етилди	In 2006
2008 йилда	Hadoop TeraSort танловида ғолиб чикди	In 2008
2010 йилда	маълумотларни сўраш ва таҳлил қилиш учун Apache Hadoop-нинг тепасида қурилган маълумотлар омбори дастурининг лойиҳаси яратилди	In 2010
HTFT	Hadoop тақсимланган файл тизими	HTFT
MapReduce	бу харитани ёзиш ва қисқартириш функцияларидан иборат дастурий модел	MapReduce
Hive	бу Hadoop Distributed File System (HDFS) устида ишлаб чиқилган ETL ва маълумотларни сақлаш воситаси	Hive
Дуг Каттинг	Hadoop асосчиси	Dug Katting

Жадвал таснифи	DESCRIBE <table>;	Table classification
Функция таснифи	DESCRIBE FUNCTION <function>;	Function classification
Жадвални ўчириш	DROP TABLE <table>;	Delete the table
Жадвални HiveQL буйруғи билан тўлдириш	LOAD DATA INPATH '<file>' INTO TABLE <table>;	Fill the table with the HiveQL command
Hadoop	маълумотлар таҳлилини амалга ошириш учун таянч вазифасини ўташи мумкин бўлган очик манбали дастурлар тўплами	Hadoop
MapReduce	ушбу файл тизимидаги маълумотларни ўқийди ва фойдаланувчилар шархлашлари мумкин бўлган визуализацияларга форматлайди.	MapReduce
Hadoop Common	фойдаланувчи компьютерлари учун файл тизимида сақланган маълумотни ўқиш учун зарур бўлган Java воситалари тўплами.	Hadoop Common
YARN	маълумотларни сақлаш ва таҳлил қилиш тизимлари ресурсларини бошқаради.	YARN

VII БЎЛИМ

АДАБИЁТЛАР
РЎЙХАТИ

VII. АДАБИЁТЛАР РЎЙХАТИ

I. Ўзбекистон Республикаси Президентининг асарлари

1. Мирзиёев Ш.М. Буюк келажакимизни мард ва олижаноб халқимиз билан бирга қурамыз. – Т.: “Ўзбекистон”, 2017. – 488 б.
2. Мирзиёев Ш.М. Миллий тараққиёт йўлимизни қатъият билан давом эттириб, янги босқичга кўтарамиз. 1-жилд. – Т.: “Ўзбекистон”, 2017. – 592 б.
3. Мирзиёев Ш.М. Халқимизнинг розилиги бизнинг фаолиятимизга берилган энг олий баҳодир. 2-жилд. Т.: “Ўзбекистон”, 2018. – 507 б.
4. Мирзиёев Ш.М. Нияти улуғ халқнинг иши ҳам улуғ, ҳаёти ёруғ ва келажак фаёвон бўлади. 3-жилд.– Т.: “Ўзбекистон”, 2019. – 400 б.
5. Мирзиёев Ш.М. Миллий тикланишдан – миллий юксалиш сари. 4-жилд.– Т.: “Ўзбекистон”, 2020. – 400 б.

II. Норматив-ҳуқуқий ҳужжатлар

6. Ўзбекистон Республикасининг Конституцияси. – Т.: Ўзбекистон, 2018.
7. Ўзбекистон Республикасининг 2020 йил 23 сентябрда қабул қилинган “Таълим тўғрисида”ги ЎРҚ-637-сонли Қонуни.
8. Ўзбекистон Республикаси Президентининг 2015 йил 12 июнь “Олий таълим муасасаларининг раҳбар ва педагог кадрларини қайта тайёрлаш ва малакасини ошириш тизимини янада такомиллаштириш чора-тадбирлари тўғрисида”ги ПФ-4732-сонли Фармони.
9. Ўзбекистон Республикаси Президентининг 2017 йил 7 февраль “Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида”ги 4947-сонли Фармони.
10. Ўзбекистон Республикаси Президентининг 2017 йил 20 апрель “Олий таълим тизимини янада ривожлантириш чора-тадбирлари тўғрисида”ги ПҚ-2909-сонли Қарори.
11. Ўзбекистон Республикаси Президентининг 2018 йил 21 сентябрь “2019-2021 йилларда Ўзбекистон Республикасини инновацион ривожлантириш стратегиясини тасдиқлаш тўғрисида”ги ПФ-5544-сонли Фармони.
12. Ўзбекистон Республикаси Президентининг 2018 йил 19 февраль

“Ахборот технологиялари ва коммуникациялари соҳасини янада такомиллаштириш чора-тадбирлари тўғрисида”ги ПФ-5349-сонли Фармони.

13. Ўзбекистон Республикаси Президентининг 2019 йил 27 май “Ўзбекистон Республикасида коррупцияга қарши курашиш тизимини янада такомиллаштириш чора-тадбирлари тўғрисида”ги ПФ-5729-сон Фармони.

14. Ўзбекистон Республикаси Президентининг 2019 йил 17 июнь “2019-2023 йилларда Мирзо Улуғбек номидаги Ўзбекистон Миллий университетда талаб юқори бўлган малакали кадрлар тайёрлаш тизимини тубдан такомиллаштириш ва илмий салоҳиятини ривожлантириш чора-тадбирлари тўғрисида”ги ПҚ-4358-сонли Қарори.

15. Ўзбекистон Республикаси Президентининг 2019 йил 27 август “Олий таълим муассасалари раҳбар ва педагог кадрларининг узлуксиз малакасини ошириш тизимини жорий этиш тўғрисида”ги ПФ-5789-сонли Фармони.

16. Ўзбекистон Республикаси Президентининг 2019 йил 8 октябрь “Ўзбекистон Республикаси олий таълим тизимини 2030 йилгача ривожлантириш концепциясини тасдиқлаш тўғрисида”ги ПФ-5847-сонли Фармони.

17. Ўзбекистон Республикаси Вазирлар Маҳкамасининг 2019 йил 23 сентябрь “Олий таълим муассасалари раҳбар ва педагог кадрларининг малакасини ошириш тизимини янада такомиллаштириш бўйича қўшимча чора-тадбирлар тўғрисида”ги 797-сонли Қарори.

18. Ўзбекистон Республикаси Президентининг 2019 йил 21 май “«Электрон ҳукумат» тизими доирасида ахборот-коммуникация технологиялари соҳасидаги лойиҳаларни ишлаб чиқиш ва амалга ошириш сифатини яхшилаш чора-тадбирлари тўғрисида”ги ПҚ-4328-сонли Қарори.

19. Ўзбекистон Республикаси Президенти Шавкат Мирзиёевнинг 2020 йил 25 январдаги Олий Мажлисга Мурожаатномаси.

20. Ўзбекистон Республикаси Президентининг 2020 йил 29 октябрь “Илм-фанни 2030 йилгача ривожлантириш концепциясини тасдиқлаш тўғрисида”ги ПФ-6097-сонли Фармони.

21. Ўзбекистон Республикаси Президентининг 2020 йил 5 октябрь “Рақамли

Ўзбекистон-2030” Стратегиясини тасдиқлаш ва уни самарали амалга ошириш чора-тадбирлари тўғрисида”ги ПФ-6079-сонли Фармони.

Ш. Махсус адабиётлар

22. Радченко И., Николаев И., Технологии и инфраструктура Big data: Учебное пособие. – Санкт-Петербург: ИТМО, 2018.

IV. Интернет сайтлар

23. [http:// www.mitc.uz](http://www.mitc.uz) - Ўзбекистон Республикаси ахборот технологиялари ва коммуникацияларини ривожлантириш вазирлиги

24. <http://lex.uz> – Ўзбекистон Республикаси Қонун ҳужжатлари маълумотлари миллий базаси

25. <http://lib./bimm.uz> – Олий таълим тизими педагог ва раҳбар кадрларини қайта тайёрлаш ва уларнинг малакасини оширишни ташкил этиш бош илмий-методик маркази

26. <http://ziyonet.uz> – Таълим портали Ziyonet

27. <http://natlib.uz> – Алишер Навоий номидаги Ўзбекистон Миллий кутубхонаси

28. [http:// www.tuit.uz](http://www.tuit.uz) - Муҳаммад ал-Хоразмий номидаги Тошкент ахборот технологиялари университети

29. Большие данные.

<https://uchebnik.mos.ru/composer3/document/44251549/view>

30. <https://computingforgeeks.com/data-mining-your-clicks>

2. <https://www.selecthub.com/big-data-analytics/big-data-analytics-requirements/>

3. <https://data-flair.training/blogs/hadoop-ecosystem-components/>

4. Big Data от А до Я. Часть 1: Принципы работы с большими данными, парадигма MapReduce. <https://habr.com/ru/post/267361/>

5. Big Data от А до Я. Часть 2: Hadoop. <https://habr.com/ru/company/dca/blog/268277/>

6. <https://data-flair.training/blogs/hadoop-ecosystem-components/>

7. <https://www.digitalocean.com/community/tutorials/an-introduction-to-big-data-concepts-and-terminology>

РЕЦЕНЗИЯ

на учебно-методический комплекс, составленный проф. Ф.Закировой и преподавателем И.Юсуповым по модулю «Методы и средства обработки больших данных» для курсов повышения квалификации и переподготовки педагогических кадров высших образовательных учреждений

Учебно-методический комплекс по модулю «Методы и средства обработки больших данных» составлен для курсов повышения квалификации и переподготовки педагогических кадров высших образовательных учреждений и содержит в себе программу курсов, рекомендованные педагогические технологии, тексты лекций, материалы для практических занятий, кейсы, глоссарий и список рекомендованной литературы и интернет сайтов.

Программа модуля соответствует содержанию типовой программы и включает в себя введение, цели и задачи модуля, требования к знаниям, умениям, навыкам и компетенциям слушателей, рекомендации к проведению занятий, разбивка часов по темам, краткое содержание теоретических и практических занятий, список рекомендованной литературы и интернет сайтов. В теоретических материалах раскрываются такие вопросы, как большие данные и их характеристики, свойства больших данных, методы обработки больших данных. В практических работах описываются этапы работы с большими данными на примере Hadoop.

Разработанный авторами учебно-методический комплекс по модулю «Методы и средства обработки больших данных» соответствует содержанию типовой и учебной программы, часы распределены соответственно часам, указанным в учебном плане.

Подводя итог, можно сказать, что учебно-методический комплекс по модулю «Методы и средства обработки больших данных» может быть рекомендован к использованию на курсах повышения квалификации и переподготовки педагогических кадров высших образовательных учреждений, а также его можно рекомендовать к публикации.

И.о. заместителя директора по научной работе и инновациям Совместного Белорусско-Узбекского межотраслевого института прикладных технических квалификаций, к.п.н.



Набиуллина

*является завершено,
канцелярией
Ибрагимовна
16.12.2020г.*

**ОЛИЙ ТАЪЛИМ МУАССАСАЛАРИ ПЕДАГОГ КАДРЛАРИНИ
ҚАЙТА ТАЙЁРЛАШ ВА МАЛАКАСИНИ ОШИРИШ КУРСИ УЧУН
ТАЙЁРЛАНГАН “КАТТА МАЪЛУМОТЛАРНИ ҚАЙТА ИШЛАШ УСУЛ
ВА ВОСИТАЛАРИ” МОДУЛИНИНГ ЎҚУВ-УСЛУБИЙ МАЖМУАСИГА
ТАҚРИЗ**

Ўқув-услубий мажмуа “Катта маълумотларни қайта ишлаш усул ва воситалари” модули бўйича қайта тайёрлаш ва малака ошириш тингловчилари учун яратилган. Қайта тайёрлаш ва малака ошириш йўналишининг ўзига хос хусусиятлари ҳамда долзарб масалаларидан келиб чиққан ҳолда ўқув-услубий мажмуада тингловчиларнинг “Катта маълумотларни қайта ишлаш усул ва воситалари” модули доирасидаги билим, кўникма, малака ҳамда компетенцияларига қўйиладиган талаблар асосида ўқув-услубий мажмусида берилган материаллар ушбу мақсадга йўналтирилиб, катта маълумотлар, уларни қайта ишлаш усул ва воситаларини ўрганиш, катта маълумотлар технологияларни таълим жараёнига қўллаш бўйича назарий ва амалий маълумотлар келтирилган.

Ўқув-услубий мажмуа доирасида берилаётган мавзулар педагог кадрларни қайта тайёрлаш ва малакасини ошириш мазмуни, сифати ва уларнинг тайёргарлигига қўйиладиган умумий малака талаблари, ўқув режалари ва дастурлари асосида шакллантирилган бўлиб, олий таълим муассасалари педагог кадрларининг соҳага оид замонавий таълим ва инновация технологиялари, илғор хорижий тажрибалардан самарали фойдаланиш, катта маълумотларни қайта ишлаш усул ва воситалари амалиётга кенг татбиқ этиш билан боғлиқ компетенцияларга эга бўлишлари таъминланади.

Умуман олганда, “Катта маълумотларни қайта ишлаш усул ва воситалари” модули бўйича яратилган ўқув-услубий мажмуа барча талабларга жавоб беради ва уни ўқув жараёнида қўллаш ва чоп этиш учун тавсия этиш мумкин.

Муҳаммад Ал-Хоразмий номидаги
ТАТУ “Ахборот технологиялари” кафедраси
мудир, профессор



Х.Зайнидинов

