

МАЪЛУМОТЛАР ОМБОРИНИ ҚУРИШДА ЗАМОНАВИЙ АРХИТЕКТУРАЛАРНИ ҚЎЛЛАШ УСУЛЛАРИ

Эргашев Аслон Акрамович

Бухоро давлат университети E-mail: aslonakromovich@gmail.com

KEYWORDS

маълумотлар архитектураси, лямбда архитектураси, каппа архитектураси, тўпламли қатлам, тезлик қатлам, хизмат қатлами, Big Data, Apache Kafka.

ABSTRACT

Ушбу мақолада лямбда ва каппа маълумотлар архитектурасини ахборот тизимлари воситасида таққослаш ўрганилган. Ҳар бир архитектуранинг асосий тамойиллари тавсифланиб, турлича фойдаланиш ҳолатларида уларнинг афзалликлари ва камчиликларини таҳлил қилинган. Apache Kafkaга каппа архитектурасининг асосий компоненти сифатида алоҳида эътибор берилиб, унинг реал вақт режимида оқимли маълумотларни қайта ишлашни таъминлашдаги роли таъкидлаб ўтилган. Лойиҳанинг ўзига хос эҳтиёжларига қараб тегишли архитектурани танлаш бўйича таҳлиллар қилинган. Мақола маълумотлар архитектураси соҳасида лямбда ва каппа ёндашувлари ўртасидаги танловни яхшироқ тушунишга ёрдам беради.

Кириш.

Бизнес тобора кўпроқ маълумотларга ва унинг таҳлилига боғлиқ бўлган бугунги ахборотли жамиятда маълумотлар архитектураси ҳар қандай ташкилотнинг муваффақиятида асосий рол ўйнамоқда. Бу шунчаки маълумотлар омбори эмас, балки катта ҳажмдаги маълумотларга тезкор кириш ҳуқуқига эга бўлиш, унинг хавфсизлиги ва таҳлилий ишлов беришни таъминлайдиган мураккаб тизим ҳисобланади [1].

Сўнгги йилларда маълумотлар ландшафтининг ўзгариши натижасида маълумотлар билан ишлашнинг янги имкониятлар пайдо бўлди. Маълумотларни сақлаш қиймати сезиларли даражада камайди, шу билан бирга маълумотларни йиғиш учун маблағ қиймати ўсишда давом этмоқда. Баъзи маълумотлар тезкор суръатда олиниб, бу уларни доимий йиғиш ва таҳлил қилишни талаб қилади. Бошқа маълумотлар кечроқ олиниши, лекин жуда катта блокларда, кўпинча ўн йиллик журнал жадваллардаги маълумотларни ўз ичига олади. Бундай шароитда илғор таҳлиллар ёки машинали ўқитишдан фойдаланиш зарурати билан боғлиқ муаммолар пайдо бўлиши мумкин.

Катта маълумотларни қайта ишлаш архитектураси айнан шу вазифаларни ҳал қилиш учун мўлжалланган [1,3].

Катта маълумотларни қайта ишлашнинг ечимлари сифатида одатда қуйидаги иш юкламаларидан бири ёки бир нечтаси фойдаланилган бўлиши мумкин [2]:

- фаол бўлмаган катта маълумотлар манбаларини тўпламли қайта ишлаш;
- катта маълумотларни реал вақт режимида қайта ишлаш;
- катта маълумотларни интерактив ўрганиш.
- башоратли таҳлил ва машинали ўқитиш.

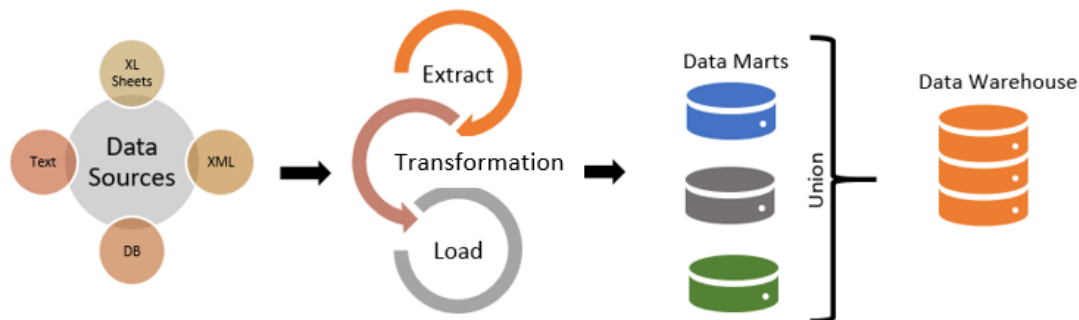
Маълумотлар билан ишлаш архитектураси ахборот тизимининг асоси бўлиб, бунда маълумотларни йиғиш, сақлаш, қайта ишлаш ва узатиш жараёнларини белгилаб беради. Маълумотлар омбори архитектураси қуйидаги босқичлар орқали ривожланиб келмоқда [4]:

- Файл тизимлари
- Реляцион маълумотлар базалари
- Булутли технологиялар
- Микросервис архитектураси
- Воқеаларга асосланган архитектура (Event-Driven Architecture, EDA)

• Серверсиз(Serverless) архитектура

Бу ўзига хос хусусиятлар ва афзалликларга эга бўлган маълумотларни

сақлаш тизимларини куришда турли ёндашувларнинг пайдо бўлишига олиб келди.



1-расм. Маълумотларни сақлаш учун Инмон модели

Ушбу ёндашувларнинг энг оммабопи Уильям Инмон ва Ральф Кимбалл томонидан ишлаб чиқилган. Улар маълумотни самарали бошқариш ва таҳлил қилиш учун маълумотлар архитектурасини ташкил қилиш кераклиги хақида фикрларни таклиф қилишган.

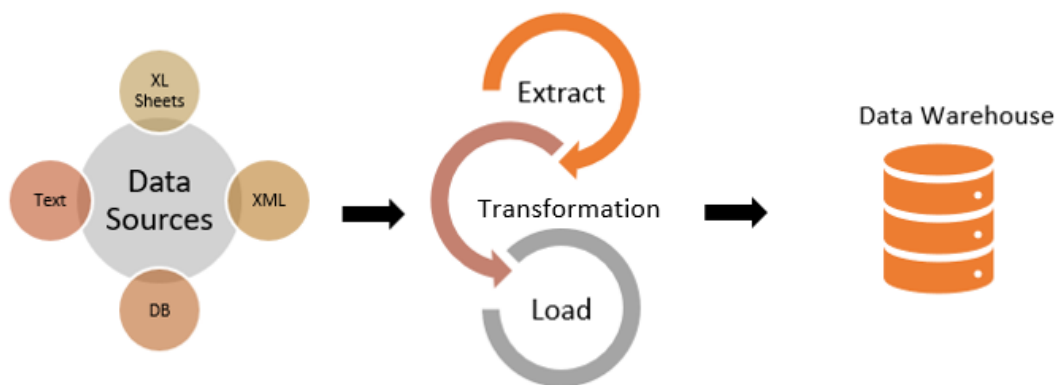
Бунга кўра маълумотлар архитектурасининг 2 та асосий тури мавжуд:

- Инмон турли мақсадларда ишлатилиши мумкин бўлган ягона маълумотлар омборини яратишни таклиф қилган (1-расм).

- Кимбалл эса таҳлил учун ишлатилиши мумкин бўлган кўп ўлчовли маълумотлар моделларини яратишни илгари сурган (2-расм) [5, 6, 7].

Маълумотлар омбори бир неча қатламлардан ташкил топган бўлиши мумкин, бунда ҳар бир қатлам ўз функциясини бажаради.

Масалан, биринчи қатлам манба маълумотларини сақлаш учун, иккинчи қатлам маълумотларни таҳлил қилиш, учинчи қатлам эса таҳлил натижаларини кўрсатиш учун ишлатилиши мумкин.



2-расм. Маълумотларни сақлаш учун

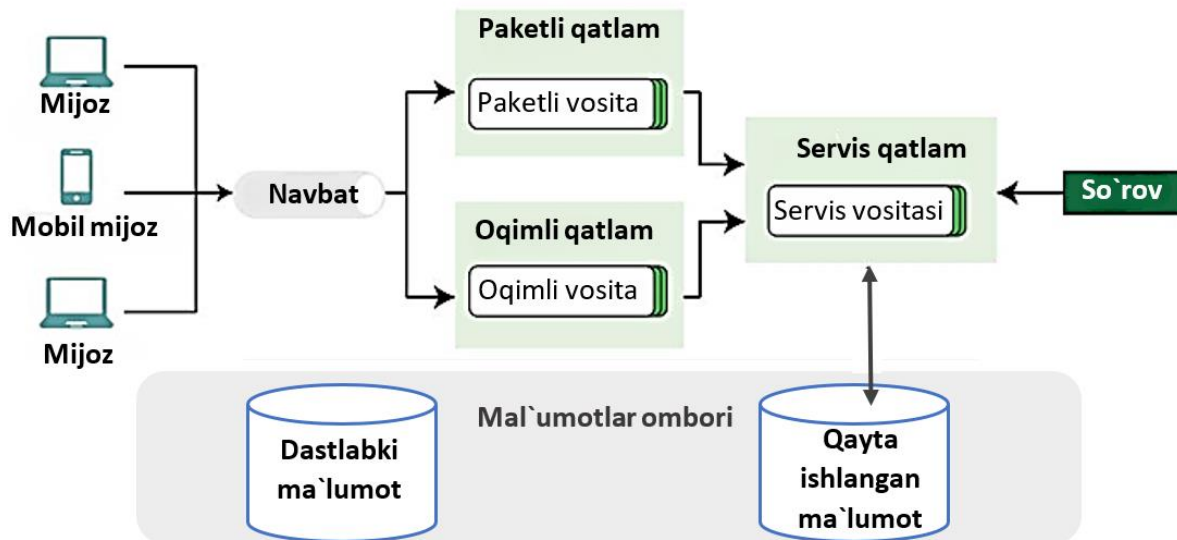
Кимбалл модели Замонавий технологиялар бизга турли мақсадларда, шу жумладан машинавий ўқитиш учун ишлатилиши мумкин бўлган маълумотлар омборларини яратишга имкон беради. Бу маълумотларни чуқурроқ ва аниқроқ таҳлил

қилиш имконини беради, бу эса янада самарали ечимларга олиб келади.

Замонавий маълумотлар архитектураларига Лямбда ва Каппа ёндашувлари ўзига хос хусусиятлари билан

ажралиб туради, айниқса ETL (тўплаш, трансформация, юклаш) конвейрларини куриш контекстида маълумотларни қайта ишлашнинг асосий элементи ҳисобланади (3-расм).

Лямбда архитектураси хизматга асосланган ёндашув сифатида бозордаги ўзгаришларга тезда мослаша олади ва долзарб таклифлар учун кўпроқ ечим бера олади.



3-расм. Лямбда архитектураси

Бу айниқса мижозларнинг тарихий маълумотларини, уларнинг ҳозирги жойлашуви каби маълумотларни ҳисобга олиш керак бўлганда ва мослаштирилган чегирма таклифларини юбориш каби вазифалар учун жуда муҳим рол ўйнайди.

3-расмда кўрсатилганидек, лямбда архитектурасининг таркибий қисмлари қуйидагича [8]:

1. **Тўпламли қатлам:** бунда маълумотлар асл шаклида сақланади ва кечикиш билан қайта ишланади. Бу қатламда дастлабки қайта ишланмаган маълумотлар, шунингдек, камдан-кам ўзгарадиган маълумотлар киради. Бунда машинали ўқитиш усулларидадан фойдаланган ҳолда, мижозларни сегментлаш ва башоратли моделларни яратиш учун тарихий маълумотлар таҳлил қилинади.

2. **Тезлик қатлами:** бу қатлам минимал кечикиш билан реал вақт режимида маълумотларни таҳлил қилишни таъминлайди. Қисқа ҳаёт циклига эга маълумотларни қайта ишлаш учун Apache Kafka, Spark, Storm ёки Flink каби

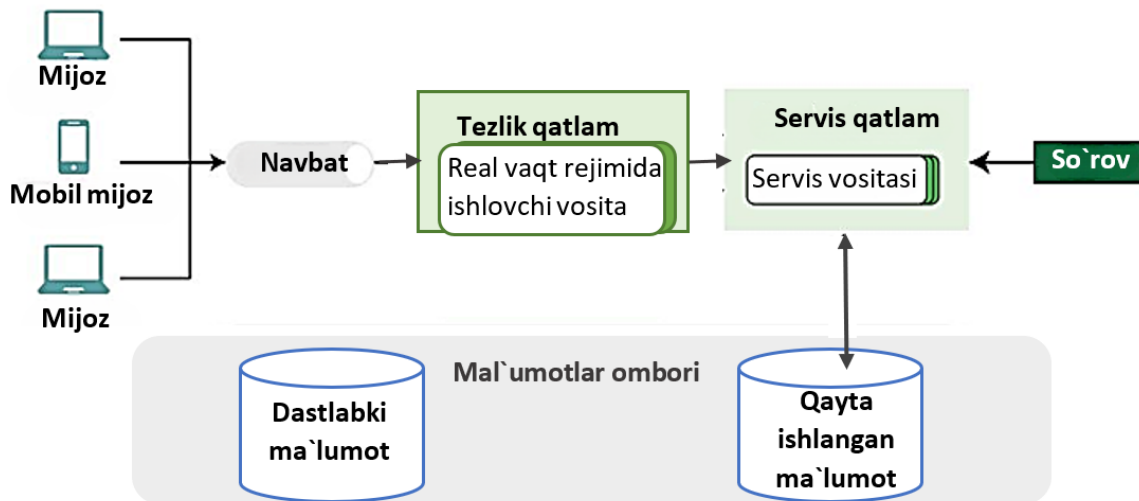
маълумотларини қайта ишлаш фреймворкларидадан фойдаланилади. Ушбу қатламда тезлик ва аниқлик ўртасидаги келишувга эришилади.

3. **Хизмат қатлами:** хизмат қатлами тўплам ва тезлик қатламлидан маълумотларни бирлаштириш учун интерфейсни тақдим этади. Бу жамланган ва долзарб маълумотларни таҳлил қилиш учун самарали фойдаланишни таъминлайди.

Лямбда архитектурасини катта ҳажмдаги Big Data амалиётида қўллаш носозликларга бардошлилик ва кенгайиш хусусияти маълумотларни минимал кечикиш билан қайта ишлашни таъминлайди. Бунга тарихий маълумотларнинг яхлитлиги ва очиқлиги сақланган ҳолда тўпламли ва тезлик қатламлини бирлаштириш орқали эришилади. Бундай сифатлар Лямбда архитектурасидан Twitter, Netflix ва Yahoo каби йирик компанияларнинг Big Data лойиҳаларида қўлланилаётгани мисол бўлади[8].

Лямбда архитектураси учун асосий фойдаланиш ҳолатларига қуйидагилар киради [9]:

- Ўзгармас маълумотлар омборидаги сўровларни қайта ишлаш.



4-расм. Каппа архитектураси

Big Data соҳасида Каппа архитектураси: соддалаштириш ва самарадорликни янада яхшилаш. 4-расмда кўрсатилганидек, Каппа архитектураси-бу барча маълумотларни кетма-кет ҳаракатлар оқими сифатида қараланадиган маълумотларни қайта ишлашнинг тизимли моделидир. Ушбу ҳаракатлар мунтазам равишда ишончли ҳодисалар журналида тартибга солинади, ҳар бир янги ҳаракат мавжуд ҳолатни янгилайди. Лямбда архитектурасидан фарқли ўларок, Каппа реал вақтда маълумотларни қайта ишлашнинг тўплам даражасидан воз кечган ҳолда доимий равишда янгилашиб турадиган маълумотлар омборининг кўриниш объектлари сифатида сақлаш имконини беради.

Каппа архитектураси тўпламли ишлов бериш заруратини бартараф этиш ва барча амаллар учун асос бўлиб хизмат қиладиган ўзгармас ҳодисалар журналига эътибор қаратиш орқали Big Data тизимларининг лойиҳаларини соддалаштиради. Ушбу ёндашув оқимли ҳисоблаш тизимларига маълумотларни тўғридан-тўғри журналдан қайта ишлашга имкон беради, зарурият туғилганда қўшимча маълумотлар омборидан керакли маълумотлар билан тўлдиради [9].

- Тезкор жавоб бериш қайтариш ва янги маълумот оқимлари интеграцияси.
- Тарихий маълумотларни сақлаган ҳолда маълумотларни кўшиш ва янгилаш имконияти.

Ҳозирда каппа архитектурасида ишлатиладиган технологияларга қуйидагилар мисол бўла олади :

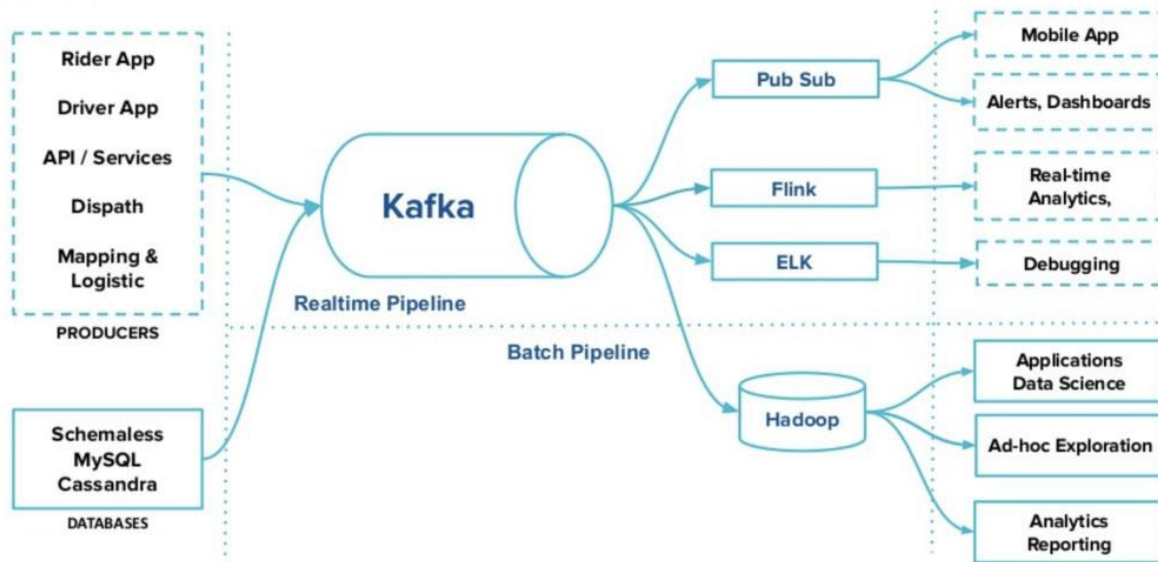
- **Apache Kafka** ва **Amazon Kinesis** каби доимий ҳаракатларни қайд этиш тизимлари.
- Оқимли ҳисоблаш фреймворклари, жумладан **Apache Spark** ва **Flink**.
- Хизмат даражасидаги тўлиқ матнли кидирув ечимларига эга бўлган резидентли маълумотлар базасидан тортиб токи ихтисослаштирилган тизимларгача(5-расм).

Каппа архитектурасини қўллаш ва Apache Kafka роли. Каппа архитектураси турли сценарийларда қўлланилади, жумладан:

- Таксимланган файл тизимида ҳаракатлар навбати ва сўровларни бошқариш.
- Воқеалар тартиби аниқланмаган ҳолда сценарийлар ва оқимли фреймворклар исталган вақтда маълумотларни ўзаро алмашиши мумкин.
- Юқори барқарорлик ҳолатида маълумотларни қайта ишлаш тизимнинг ҳар бир тугунида амалга оширилади.

Apache Kafka kuchli xabar vositachisi sifatida ma'lumotlarni yig'ish va agregatlash uchun tezkor, ishonchli va kengaytirilgan platformani takdim etish orqali ushbu talablarga javob beradi. Bu LinkedIn kabi juda kўp bir xil sўrovlariga

xizmat qilish uchun katta hajmdagi ma'lumotlarni qayta ishlash va saqlashni talab qiladigan loyihalarda Kafka asosidagi Kappa arxitekturasini jozibadorligini ta'minlaydi [10].



5-рasm. ETL қуришда Каппа архитектураси

1-жадвалда илова қилинган қиёсий таҳлил ҳар бир архитектуранинг асосий жиҳатларини таққослаш ва уларнинг

фарқларини аниқлашга асосланган бўлиб, бу маълумотларни қайта ишлаш учун мос танловни аниқлашга ёрдам беради [11].

1-жадвал

Лямбда ва каппа архитектуралари ўртасидаги асосий тафовутлар

Тавсиф	Лямбда архитектурасининг тавсифи	Каппа архитектурасининг тавсифи
Маълумотларни қайта ишлаш	Маълумотларни қайта ишлашнинг иккита паралел оқимларидан фойдаланади: тўпلامли (batch) ва реал вақт режимли (real-time).	Реал вақт режимида маълумотларни қайта ишлашнинг ягона оқимидан фойдаланади.
Компонентлар архитектураси	Икки қатламдан иборат: "юмшок" (Batch Layer) ва "қаттиқ" (Speed Layer).	Алоҳида қатламларга ажратмайди балки, барча маълумотлар реал вақтда қайта ишланади.
Мураккаблиги	Маълумотларни қайта ишлашнинг икки хил оқимини сақлаш ва созлаш зарурати туфайли архитектурани янада мураккаблаштиради.	Оддий ва соддалаштирилган архитектура, чунки у фақат реал вақтда маълумотларни қайта ишлайди.
Қайта ишлашдаги тўхталишлар	Тўпلامли қайта ишлашда кечикишлар бўлиши мумкин. Қайта ишлаш вақти "қаттиқ" қатламнинг созланмаси ва самарадорлигига боғлиқ.	Барча маълумотлар реал вақтда қайта ишлангани учун минимал кечикишлар амалга оширилади.
Ресурслардан	Маълумотларни қайта ишлашнинг икки хил оқимини сақлаш ва созлаш зарурати туфайли кўп ресурс талаб қилади.	Реал вақт режимида битта оқим қайта ишлангани учун жуда кам ресурс сарфлайди.

фойдалан иш		
Мослик	Тўпламли ва оқимли маълумотларни қайта ишлаш сценарийларига мос келади, масалан Big Data таҳлиллари каби.	Мониторинг ва реал вақт режимидаги таҳлил сценарийларига мос келади.
Воситалар	Тўпламли қатламда-Hadoop MapReduce Тезлик қатламида-Apache Storm, Apache Samza ва Spark Streaming Apache Spark, Apache Flink-тўплам ва тезлик қатламларида ишлатилиши мумкин	Apache Storm, Apache Samza ва Spark Streaming

Хулоса. Замоनावий маълумотларни таҳлил қилиш архитектураларини синтезлаш. Лямбда архитектураси, каппа архитектураси ва микросервисларни ўз ичига олган турли хил маълумотлар архитектураларини ўрганиш ахборотни бошқариш ва таҳлил қилиш учун кенг имкониятларни очиб беради. Ушбу ёндашувлар бизнес жараёнларини ва қарорларни қабул қилишни ўзгартира оладиган кучли ва мослашувчан маълумотларни қайта ишлаш тизимлари учун асос яратади.

Қуйидаги ҳолларда катта маълумотларни қайта ишлаш учун мазкур архитектуралардан фойдаланиш мақсадга мувофиқдир:

- маълумотлар анъанавий маълумот базаси сақлаш имкониятларидан юқори ҳажми ташкил қилганда уни йиғиш ва қайта ишлаш зарурияти пайдо бўлганда;
- кейинги таҳлил ва ҳисобот учун тузилмани бўлмаган маълумотларни қайта ишлашда;
- бир-бирига боғлиқ бўлмаган маълумотлар оқимларини реал вақт режимида ёки минимал кечикиш билан ёзиб олиш, қайта ишлаш ва таҳлилларни олиб бориш учун.

Ушбу технологияларни ўзлаштириш нафақат вақт ва ақлий куч, балки махсус билимларни ҳам талаб қилади. Шу нуқтаи назардан, маълумотлар архитектурасини амалга ошириш ва қўллаб-қувватлаш учун тўғри ҳамкор ёки платформани танлаш муваффақиятнинг асосий омили бўлиши мумкин.

Маълумотлар архитектурасини, хусусан лямбда ва каппа архитектураларини

ўрганиш замоनावий ахборот технологиялари соҳасида катта ўзгаришларга олиб келди. Тўпламли ва оқимли маълумотларни қайта ишлашни бирлаштирган Лямбда архитектураси юқори мослашувчанликни ва ҳар хил турдаги маълумотлар билан ишлаш имкониятини таъминлайди. Бошқа томондан, каппа архитектураси реал вақт режимида маълумотларни қайта ишлашга эътибор қаратиб, маълумотларни тезкор қайта ишлашни таъминлайди, бу айниқса мониторинг, реал вақтдаги таҳлиллар ва мижоз ҳаракатларига жавоб қайтариш каби талаблар учун муҳимдир. Мавжуд архитектуралар орасидаги танлов ташкилотнинг ўзига хос эҳтиёжлари ва мақсадларига мувофиқ бўлиши керак.

Адабиётлар.

1. Эргашев А., Эшанкулов Ҳ.И. (2021). Иқтисодий бошқарув қарорларини қабул қилишда Business Intelligence тизимларининг устунлик жиҳатлари. Педагогик маҳорат журнали.
2. Azure Architecture Center | Microsoft Learn - <https://learn.microsoft.com/en-us/azure/architecture/databases/guide/big-data-architectures>
3. Eshankulov Hamza Ilhomovich.(2023) Business Intelligence dasturiy ta'minotlarning xususiyatlari va tuzilmasi. Volume 1, issue 3,
4. October 2023 Raqamli Transformatsiya va Sun'iy Intellekt ilmiy jurnali.
5. <https://habr.com/ru/articles/795297/> - Памятка РП. Микросервисная архитектура
6. Breslin, M. (2004). Data Warehousing Battle of the Giants: Comparing the Basics of Kimball and Inmon Models. Business

- Intelligence Journal, Winter 2004; 9(1), pp.6-19.
7. Jukic, N. (2006). Modelling strategies and alternatives for data warehousing projects. Communications of the ACM, 49(4), pp.83-88
 8. L. Yessad and A. Labiod, “Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault,” in 2016 International Conference on System Reliability and Science (ICSRS), 2016, pp. 95–99
 9. Kiran, Mariam et al.: Lambda architecture for cost-effective batch and speed big data processing. In: Big Data. IEEE, pp. 2785–2792, 2015.
 10. Lin, Jimmy: The Lambda and the Kappa. IEEE Internet Computing, 21(5):60–66, 2017
 11. Zhelev, Svetoslav & Rozeva, Anna: Big data processing in the cloud-Challenges and platforms. In: AIP Conference Proceedings. AIP Publishing, 2017
 12. <https://medium.com/towards-data-engineering/data-processing-architectures-lambda-vs-kappa-for-big-data-8cc9a7edefd> - Data processing architectures — Lambda vs Kappa for Big Data.