

## APACHE KAFKA VA SPARK STREAMING ASOSIDA OQIMLI MA'LUMOTLARNI QAYTA ISHLASH

*Farg'ona davlat texnika universiteti talabasi*

***Tursunov Ikromjon Ibrohimjon o'g'li***

*tursunovikromjon90@gmail.com*

*Tel: +998944974346*

*Fan o'qituvchisi: Xolmatov Abdurashid*

**Annotatsiya:** Ushbu ishda Apache Kafka va Spark Streaming texnologiyalari asosida oqimli ma'lumotlarni real vaqt rejimida yig'ish, uzatish va qayta ishlash jarayoni yoritilgan. Tizim katta hajmdagi ma'lumotlarni uzluksiz tahlil qilish, filtrlash va natijalarni tezkor olish imkonini beradi. Ushbu yondashuv sanoat, moliya, telekommunikatsiya va IoT sohalarida samaradorlikni oshirish hamda qaror qabul qilish jarayonlarini tezlashtirishga xizmat qiladi.

**Аннотация:** В данной работе рассматриваются технологии Apache Kafka и Spark Streaming для сбора, передачи и обработки потоковых данных в режиме реального времени. Система обеспечивает непрерывный анализ больших объемов данных, их фильтрацию и получение результатов с высокой скоростью. Такой подход повышает эффективность и ускоряет процесс принятия решений в промышленности, финансовой сфере, телекоммуникациях и IoT.

**Annotation:** This study focuses on the use of Apache Kafka and Spark Streaming technologies for real-time collection, transmission, and processing of streaming data. The system enables continuous analysis of large data flows, efficient filtering, and rapid generation of analytical results. This approach improves efficiency and accelerates decision-making processes in industries such as finance, telecommunications, manufacturing, and IoT.

### Kirish

Zamonaviy axborot texnologiyalari rivojlanishi natijasida turli tizimlar tomonidan sekundiga millionlab ma'lumotlar hosil qilinmoqda. Bunday oqimli ma'lumotlarni an'anaviy usullar yordamida saqlash va tahlil qilish samarali natija bermaydi. Shu sababli real vaqt rejimida ishlov berish imkoniyatiga ega dasturiy platformalar zaruriyatga aylandi. Apache Kafka va Spark Streaming texnologiyalari ana shunday talablarni qondiruvchi kuchli vositalardan hisoblanadi.

Apache Kafka oqimli ma'lumotlarni ishonchli qabul qilish, saqlash va uzatishni ta'minlasa, Spark Streaming ushbu ma'lumotlarga tezkor tahliliy ishlov berish imkonini yaratadi. Ushbu texnologiyalar majmuasi moliya, telekommunikatsiya, sanoat, IoT va boshqa ko'plab sohalarda real vaqt rejimida monitoring, nazorat hamda

analitik qarorlar qabul qilishni sezilarli darajada soddalashtiradi. Mazkur mavzuda ushbu texnologiyalarning imkoniyatlari, afzalliklari va amaliy qo'llanilishi yoritib beriladi.

Zamonaviy raqamli davrda turli sohalarda axborot oqimi keskin oshib bormoqda. Korxonalar, internet xizmatlari, bank tizimlari, IoT qurilmalari va telekommunikatsiya tarmoqlari sekundiga millionlab ma'lumotlar generatsiya qilmoqda. An'anaviy ma'lumotlarni qayta ishlash usullari bunday tezlik va hajmdagi oqim bilan ishlash imkoniyatiga ega emas. Shuning uchun real vaqt rejimida ma'lumotlarni yig'ish, uzatish va tahlil qilishni ta'minlaydigan kuchli platformalarga ehtiyoj paydo bo'ldi. Shu nuqtada Apache Kafka va Spark Streaming texnologiyalari muhim rol o'ynaydi.

Apache Kafka — bu yuqori tezlikda kelayotgan oqimli ma'lumotlarni ishonchli tarzda qabul qilish, vaqtincha saqlash va boshqa tizimlarga uzatish imkoniyatiga ega bo'lgan distribyutiv platforma. Kafka ma'lumotlarni "topic"larga ajratgan holda ularni tartibli saqlaydi va kerakli ilovalar uchun uzluksiz yetkazib berishni ta'minlaydi. Uning asosiy afzalligi — barqarorlik, yuqori ishlash samaradorligi, nosozliklarga chidamlilik va katta yuklamalarda ham samarali ishlay olishidir. Shuning uchun u bugungi kunda dunyodagi yirik kompaniyalar tomonidan faol qo'llanilmoqda.

Spark Streaming esa oqimli ma'lumotlarga real vaqtning o'zida analitik ishlov beradigan kuchli hisoblash tizimidir. U ma'lumotlarni mikro-paketlarga (mini-batch) bo'lib qayta ishlaydi va murakkab hisob-kitoblarni qisqa vaqt ichida amalga oshiradi. Spark Streaming yordamida ma'lumotlarni filtrlash, guruhlash, tahlil qilish, statistik natijalar olish, grafik va hisobotlar shakllantirish mumkin. Ushbu platforma yuqori tezlikda ishlashi, massiv ma'lumotlar bilan ishlash qobiliyati hamda moslashuvchanligi bilan ajralib turadi.

Kafka va Spark Streaming birgalikda ishlaganda juda kuchli real vaqtli tahliliy tizimni hosil qiladi. Jarayon odatda quyidagicha tashkil etiladi: bir nechta manbalardan ma'lumotlar real vaqt rejimida Apache Kafka tizimiga keladi. Kafka ularni saqlab, tartiblaydi va Spark Streaming tizimiga uzatadi. Spark Streaming esa ushbu ma'lumotlarni qayta ishlab, kerakli natijalarni ma'lumotlar bazasi, vizualizatsiya tizimi yoki analitik platformaga yuboradi. Natijada foydalanuvchi real vaqtning o'zida qulay, tahliliy va aniq ma'lumotlarga ega bo'ladi.

Ushbu texnologiyalar bir qancha sohalarda keng qo'llaniladi. Bank va moliya tizimlarida tranzaksiyalarni nazorat qilish, firibgarlik holatlarini aniqlash, risklarni baholashda juda muhimdir. Telekommunikatsiyada tarmoq trafikini kuzatish, uzilishlarni aniqlash va xizmat sifatini oshirishda yordam beradi. Sanoat va IoT sohasida sensorlardan kelayotgan signal va ko'rsatkichlarni boshqarish, ishlab chiqarishni nazorat qilishda qo'llaniladi. Internet platformalarida esa foydalanuvchi faoliyatini kuzatish, tavsiyalar tizimini yaratish, reklama tizimini takomillashtirishda

samarali natija beradi. Shuningdek, transport, logistika, xavfsizlik tizimlari va davlat boshqaruvi sohalarida ham ulardan keng foydalanilmoqda.

Kafka va Spark Streaming asosida oqimli ma'lumotlarni qayta ishlashning asosiy afzalliklari quyidagilardan iborat: real vaqt rejimida ishlash imkoniyati, yuqori tezlik va unumdorlik, ishonchlilik, tizimni istalgancha kengaytirish imkoniyati, katta hajmdagi ma'lumotlarni barqaror boshqarish va tahlil qilish imkoniyati. Bularning barchasi zamonaviy axborot tizimlari uchun juda muhim omillardir.

Xulosa qilib aytganda, Apache Kafka va Spark Streaming texnologiyalari oqimli ma'lumotlarni qayta ishlashning eng samarali va ishonchli usullaridan biri hisoblanadi. Ularning qo'llanilishi natijasida real vaqt rejimida monitoring, nazorat va analitik jarayonlar samarali tashkil etiladi, bu esa qaror qabul qilish jarayonlarini tezlashtiradi, xizmat sifatini oshiradi va tizimlarning ishonchliligini ta'minlaydi. Ushbu texnologiyalar kelajakda yanada rivojlanib, katta ma'lumotlar bilan ishlashning ajralmas qismiga aylanishi shubhasiz.

### **Xulosa**

Apache Kafka va Spark Streaming asosida oqimli ma'lumotlarni qayta ishlash texnologiyalari bugungi raqamli dunyoning ajralmas qismiga aylandi. Real vaqt rejimida ma'lumotlarni yig'ish, uzatish va tahlil qilish imkoniyati turli sohalarda tezkor qaror qabul qilish, muammolarni erta aniqlash va samaradorlikni oshirishga xizmat qilmoqda. Agar ilgari ma'lumotlarni yig'ib, so'ngra tahlil qilish uchun ko'p vaqt talab etilgan bo'lsa, endilikda ushbu texnologiyalar yordamida hammasi bir zumda amalga oshmoqda.

Masalan, bank tizimini olaylik. Mijoz kartasidan shubhali tranzaksiya amalga oshirilsa, Kafka va Spark Streaming yordamida bu holat darhol aniqlanadi va tizim bir necha soniya ichida xavfni bartaraf etish choralarini ko'radi. Shuningdek, internet platformalarida foydalanuvchi qanday videoni ko'rayotgani yoki nimalarga qiziqayotgani bir vaqtning o'zida tahlil qilinib, unga mos tavsiyalar taklif etiladi — bu esa xizmat sifatini sezilarli oshiradi.

Telekommunikatsiya sohasida ham ushbu texnologiyalar juda muhim. Masalan, tarmoqda yuklama oshib ketayotganini tizim real vaqtning o'zida aniqlaydi va avtomatik ravishda yuklamani boshqa serverlarga taqsimlaydi. Natijada foydalanuvchilar uzilishsiz aloqa va barqaror internet xizmatidan foydalanadi. Sanoatda esa ishlab chiqarish liniyalaridagi sensorlar doimiy ravishda ma'lumot uzatadi va tizim har qanday nosozlikni erta aniqlab, ishlab chiqarishning to'xtab qolishining oldini oladi.

Shu kabi hayotiy misollar shuni ko'rsatadiki, Apache Kafka va Spark Streaming nafaqat texnik yechim, balki inson hayotini yengillashtiruvchi, xavfsizlikni oshiruvchi, xizmatlarni yanada sifatli qiluvchi muhim vositadir. Ular yordamida katta hajmdagi

oqimli ma'lumotlar tartibli boshqariladi, tezkor tahlil qilinadi va ishonchli natijalar olinadi.

Yakuniy xulosa sifatida aytish mumkinki, ushbu texnologiyalar zamonaviy axborot tizimlarining tayanch poydevorlaridan biriga aylangan. Kelajakda ham ularning qo'llanilish doirasi kengayib, iqtisodiyot, ta'lim, tibbiyot, transport va boshqa ko'plab sohalarda yangi imkoniyatlar yaratishda muhim rol o'ynaydi.

### **Foydalanilgan adabiyotlar**

Ushbu mavzuni yoritishda bir qator ilmiy manbalar va rasmiy hujjatlardan foydalanildi. Avvalo, oqimli ma'lumotlarni qayta ishlash texnologiyalarini chuqur o'rganishda **Apache Software Foundation** tomonidan taqdim etilgan *Apache Kafka Documentation* va *Spark Streaming Programming Guide* rasmiy hujjatlari muhim nazariy va amaliy manba bo'lib xizmat qildi. Kafka arxitekturasini, uning ishlash tamoyillari va qo'llanilish sohalari bo'yicha Kreps J., Narkhede N. va Rao J. tomonidan yozilgan **“Kafka: a Distributed Messaging System for Log Processing” (2011)** asari asosiy texnik yo'nalishlarni tushuntirib berdi.

Spark Streaming texnologiyasining nazariy asoslari va uning ishonchli ishlash mexanizmlari M. Zaharia va hamkorlari tomonidan yozilgan **“Discretized Streams: Fault-Tolerant Streaming Computation at Scale” (2013)** ilmiy ishida keng yoritilgan. Spark texnologiyasining yuqori samaradorlikda ishlashi va funksional imkoniyatlari Karau H. va Warren R.ning **“High Performance Spark” (2017)** asarida batafsil tushuntirib berilgan.

Kafka platformasining chuqur tahlili va amaliyotdagi qo'llanilishi Shapira G., Palino T. va Seidman T. muallifligidagi **“Kafka: The Definitive Guide” (2017)** kitobida keng ochib berilgan. Shuningdek, Spark Streaming va Structured Streaming imkoniyatlari Vohra D.ning **“Practical Apache Spark” (2018)** asarida amaliy misollar asosida ko'rsatib o'tilgan. Real vaqt rejimida ma'lumotlarni qayta ishlash masalalarini yoritishda Chen L. va Suneja S.ning **IEEE konferensiya maqolasi (2019)** ham muhim ilmiy manbalardan biri bo'ldi.

Katta ma'lumotlarni tahlil qilishga oid umumiy nazariy va amaliy bilimlar Guller M.ning **“Big Data Analytics with Spark” (2015)** hamda Marz N. va Warren J.ning **“Big Data: Principles and Best Practices of Scalable Real-Time Data Systems” (2015)** asarlarida keng qamrab olingan.