

OQIMLI PROTSSESSORLAR, REAL VAQT REJIMIDA MA'LUMOTLARNI QAYTA ISHLASH

Shokirov Shodmon Shoyimovich

Muhammad al-Xorazmiy nomidagi TATU

Katta o'qituvchi

Annotatsiya

Ushbu maqola oqimli protsessorlar (stream processors) va ularning real vaqt rejimida ma'lumotlarni qayta ishlashdagi ahamiyatini tahlil qiladi. Oqimli protsessorlar yordamida ma'lumotlar oqimi doimiy ravishda kuzatiladi va tahlil qilinadi, bu esa katta hajmdagi ma'lumotlar bilan ishlashda katta afzalliklar beradi. Maqolada oqimli qayta ishlashning asosiy kontseptlari, afzalliklari va qo'llanilishi haqida batafsil ma'lumot beriladi. Shuningdek, texnik detallar, jumladan, ma'lumotlar oqimi modellari, qayta ishlash semantiklari va turg'unlik boshqaruvi kabi masalalar ham ko'rib chiqiladi. Oqimli protsessorlarning moliyaviy soha, sog'liqni saqlash, ijtimoiy tarmoqlar va telekom sohalaridagi qo'llanilishi haqida misollar keltirilgan. Ushbu maqola oqimli protsessorlar va ularning zamonaviy texnologiyalar bilan integratsiyasiga qiziquvchilar uchun mo'ljallangan.

Kalit so'zlar: oqimli protsessorlar, ma'lumotlarni real vaqt rejimida qayta ishlash, oqim (stream), oqim qayta ishlash dasturlari, oqim qayta ishlash dasturiy platformalari, apache kafka, apache flink, apache storm, spark streaming, reaktivlik, shkalalanish, kechikishlarni minimallashtirish, ma'lumotlar oqimi modellari

Oqimli protsessorlar (Stream Processors) - bu ma'lumotlarni real vaqt rejimida qayta ishlash imkonini beruvchi maxsus qurilmalar yoki dasturiy ta'minot komponentlari. Oqimli protsessorlar yordamida ma'lumotlar oqimi (stream) doimiy ravishda kuzatiladi va tahlil qilinadi, bu esa katta hajmdagi ma'lumotlar bilan ishlashda katta afzalliklar beradi. Quyida oqimli protsessorlar haqida batafsilroq ma'lumotlar keltirilgan:

Oqimli Ma'lumotlarni Qayta Ishlash (Stream Processing)

Asosiy Kontseptlar:

- Oqim (Stream): Bu, odatda, doimiy ravishda kiritiladigan va yangilanadigan ma'lumotlar ketma-ketligidir. Masalan, sensorlar, log fayllar yoki ijtimoiy tarmoqlardan olinadigan ma'lumotlar oqimi.
- Oqim Qayta Ishlash Dasturlari (Stream Processing Applications): Bu dasturlar oqimdagi ma'lumotlarni real vaqt rejimida tahlil qiladi va qayta ishlaydi. Oqim qayta ishlashning asosiy vazifalari ma'lumotlarni filtratsiyalash, transformatsiya qilish va agregatsiyalashdan iborat.
- Oqimli Qayta Ishlash Dasturiy Platformalari (Stream Processing Frameworks): Apache Kafka, Apache Flink, Apache Storm va Spark Streaming kabi platformalar oqimli ma'lumotlarni qayta ishlash uchun keng qo'llaniladi.

Oqimli Ma'lumotlarni Qayta Ishlashning Afzalliklari:

1. Reaktivlik (Reactivity): Oqimli qayta ishlash ma'lumotlarga real vaqt rejimida javob berish imkonini beradi. Bu hodisalar sodir bo'lgan vaqtda darhol tahlil qilish

imkoniyatini beradi.

2. Shkalalanish (Scalability): Oqimli qayta ishlash tizimlari katta hajmdagi ma'lumotlarni samarali qayta ishlash uchun shkalalanadi.

3. Kechikishlarni Minimallashtirish (Low Latency): Oqimli qayta ishlash tizimlari ma'lumotlarni deyarli kechiktirmasdan qayta ishlaydi.

Qo'llanilishi:

1. Moliyaviy Soha: Moliyaviy tranzaksiyalarni real vaqt rejimida kuzatish va firibgarliklarni aniqlash.

2. Sog'liqni Saqlash: Sensorlar va tibbiy qurilmalardan olinadigan ma'lumotlarni real vaqt rejimida qayta ishlash va kuzatish.

3. Ijtimoiy Tarmoqlar: Foydalanuvchilarning faoliyatini kuzatish va real vaqt rejimida tahlil qilish.

4. Telekom: Tarmoq trafiginu kuzatish va optimallashtirish.

Texnik Detallar:

1. Ma'lumotlar Oqimi Modellar: Oqim modeli asosan ikkita usulga bo'linadi - micro-batching va event-by-event. Micro-batchingda ma'lumotlar kichik bloklarga bo'linadi va har bir blok alohida qayta ishlanadi. Event-by-event usulida esa har bir hodisa (event) alohida qayta ishlanadi.

2. Qayta Ishlash Semantikolari: Exactly-once, at-least-once, va at-most-once semantikolari oqimli qayta ishlashda qo'llaniladi. Bu semantikalar ma'lumotlarni necha marta qayta ishlash kerakligini belgilaydi.

3. Turg'unlik (State Management): Oqimli qayta ishlash tizimlari ma'lumotlar oqimidan turg'unlikni saqlash va tiklash uchun turg'unlikni boshqarish usullarini qo'llaydi. Bu tizimlar qayta ishga tushganda yoki nosozlik yuzaga kelganda muhimdir.

Oqimli protsessorlar ma'lumotlarni real vaqt rejimida tahlil qilish imkoniyatini taqdim etadi, bu esa zamonaviy texnologiyalar va katta ma'lumotlar dunyosida juda muhimdir. Ular turli sohalarda qo'llaniladi va turli xil texnikalar bilan birgalikda ishlatiladi, bu esa ularga katta moslashuvchanlik va kuchli imkoniyatlar beradi.

Foydalanilgan Adabiyotlar:

- 1.Kreps, J., Narkhede, N., & Rao, J. (2011). Kafka: A Distributed Messaging System for Log Processing. Proceedings of the NetDB, Athens, Greece.
- 2.Carbone, P., Katsifodimos, A., Ewen, S., Markl, V., Haridi, S., & Tzoumas, K. (2015). Apache Flink: Stream and Batch Processing in a Single Engine. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering.
- 3.Toshniwal, A., Taneja, S., Shukla, A., Ramasamy, K., Rao, S., & Jha, G. (2014). Storm @ Twitter. Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, New York, NY, USA.
- 4.Zaharia, M., Das, T., Li, H., Hunter, T., Shenker, S., & Stoica, I. (2013). Discretized Streams: Fault-Tolerant Streaming Computation at Scale. Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles, New York, NY, USA.
- 5.Garg, R. (2020). Real-Time Analytics with Storm and Kafka: Real-Time Analytics on Streaming Data with Storm and Kafka. Packt Publishing Ltd.
- 6.Kleppmann, M. (2017). Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. O'Reilly Media.
- 7.Munshi, A. (2018). Advanced Analytics with Spark: Patterns for Learning from Data at Scale. O'Reilly Media.
- 8.Lorido-Bofran, T., Miguel-Alonso, J., & Lozano, J. A. (2014). A Review of Auto-scaling Techniques for Elastic Applications in Cloud Environments. Journal of Grid Computing.