

№12 Дәріс. Нақты уақыттағы деректерді өңдеу әдістері мен технологиялары

Дәріскер: Даркенбаев Д.К. PhD, қауымдастырылған
профессор



Дәріс мақсаты

Нақты уақыттағы деректерді өңдеу түсінігін, оның түрлерін, әдістерін және заманауи технологияларын түсіндіру

Нақты уақыт жүйелерінде қолданылатын архитектуралық тәсілдер мен платформаларды меңгерту.

Кілт сөздер

Нақты уақыт жүйесі

stream processing

real-time processing

event-driven architecture

latency

in-memory computing

complex event processing (CEP)

message queue

Технологиялар

Kafka

Spark Streaming

Flink

Дәріс жоспары

01

Нақты уақыттағы
деректерді өңдеуге кіріспе

02

Нақты уақыт жүйелерінің
сипаттамалары

03

Нақты уақыттағы
деректерді өңдеу
модельдері

04

Event-driven архитектура

05

Нақты уақыттағы
деректерді өңдеу әдістері

06

Заманауи технологиялар:
Kafka, Spark Streaming,
Flink, Storm

07

In-Memory және CEP
жүйелері

08

Қолданылу салалары

09

Қорытынды



Кіріспе

1. Нақты уақыттағы деректерді өңдеуге кіріспе

Нақты уақыттағы деректерді өңдеу (Real-time data processing) — деректерді олардың келу сәтінде өңдеу және минималды кідіріс (latency)пен жауап беру процесі.

Мақсат: оқиға болған сәтте тез шешім қабылдау.

Қажет салалар:

Индустрия 4.0

Қаржы мониторингі
және алаяқтықты
анықтау

Автономды көліктер

Интернет заттар
жүйелері (IoT)

Онлайн медицина
және телемедицина

1. Нақты уақыт жүйелерінің негізгі сипаттамалары

Қасиет	Түсініктеме
Төмен кідіріс (Low latency)	Жауап мс-секунд ішінде
Сенімділік	Қате болмауы тиіс
Нақты уақыттық операциялар	Оқиғалар ағымда өңделеді
Масштабталу	Үлкен ағындармен жұмыс
Қолжетімділік	Үзіліссіз жұмыс (24/7)

1. Нақты уақыттағы деректерді өңдеу модельдері

Batch Processing (Топтық)

- Үлкен деректерді кезең бойынша өңдеу
- Мысалы: Hadoop MapReduce

Stream (Real-Time) Processing

- Деректер келген сайын өңдеу
- Мсек уақыт масштабы

Near-real-time

- Жауап 1–5 секунд кешігумен



Stream processing мысалы:

- Төлем транзакцияларын тексеру
- Видеокамерадан жол қозғалысын талдау

1. Event-Driven Architecture (Оқиғаға бағытталған архитектура)

Компоненттер оқиғалармен байланысады:



Event producers

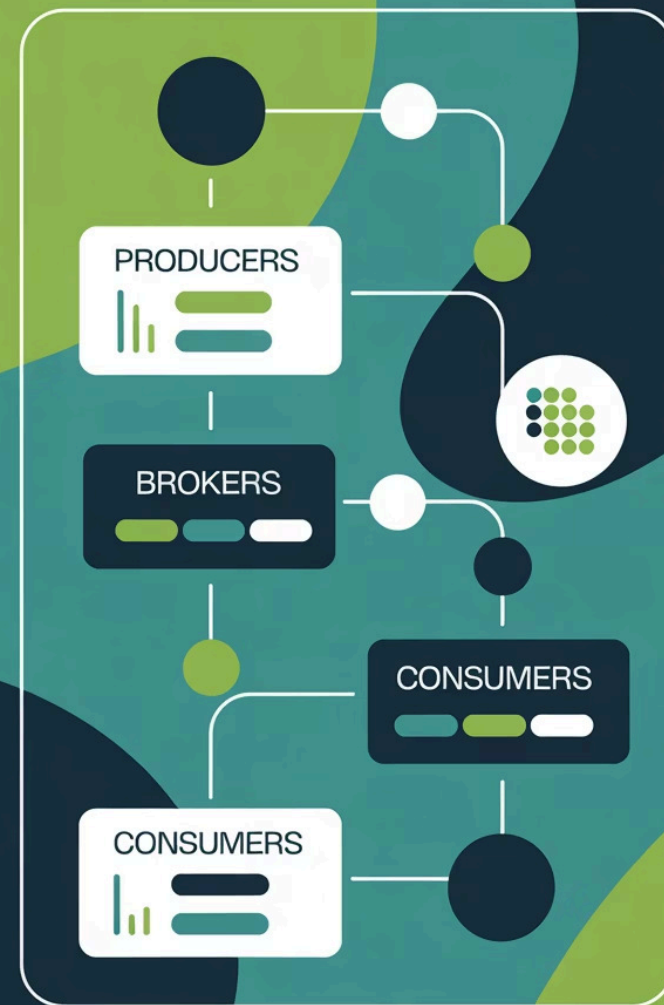
Дерек жібереді (сенсор, қолданба)

Event broker

Kafka, RabbitMQ — оқиғаларды маршрутизациялайды

Event consumers

Оқиғаны өңдейді



1. Нақты уақыттағы деректерді өңдеу әдістері

Әдіс	Түсініктеме
Micro-batch processing	Өте шағын топтарда өңдеу (Spark Streaming)
Event-by-event processing	Әр оқиғаны жеке өңдеу (Flink, Storm)
CEP (Complex Event Processing)	Күрделі үлгіні тану, оқиғалар комбинациясы
In-memory processing	RAM-да жылдам есептеу (Redis, Ignite)
Message-driven processing	Кезек арқылы өңдеу

1. Нақты уақыт технологиялары



Kafka

Жоғары жылдамдықты деректер шинасы



Apache Spark Streaming

Micro-batch stream processing



Apache Flink

Нағыз event-based stream processing



Apache Storm

Үздіксіз stream өңдеу



RabbitMQ

Message-queue брокері



Redis Streams

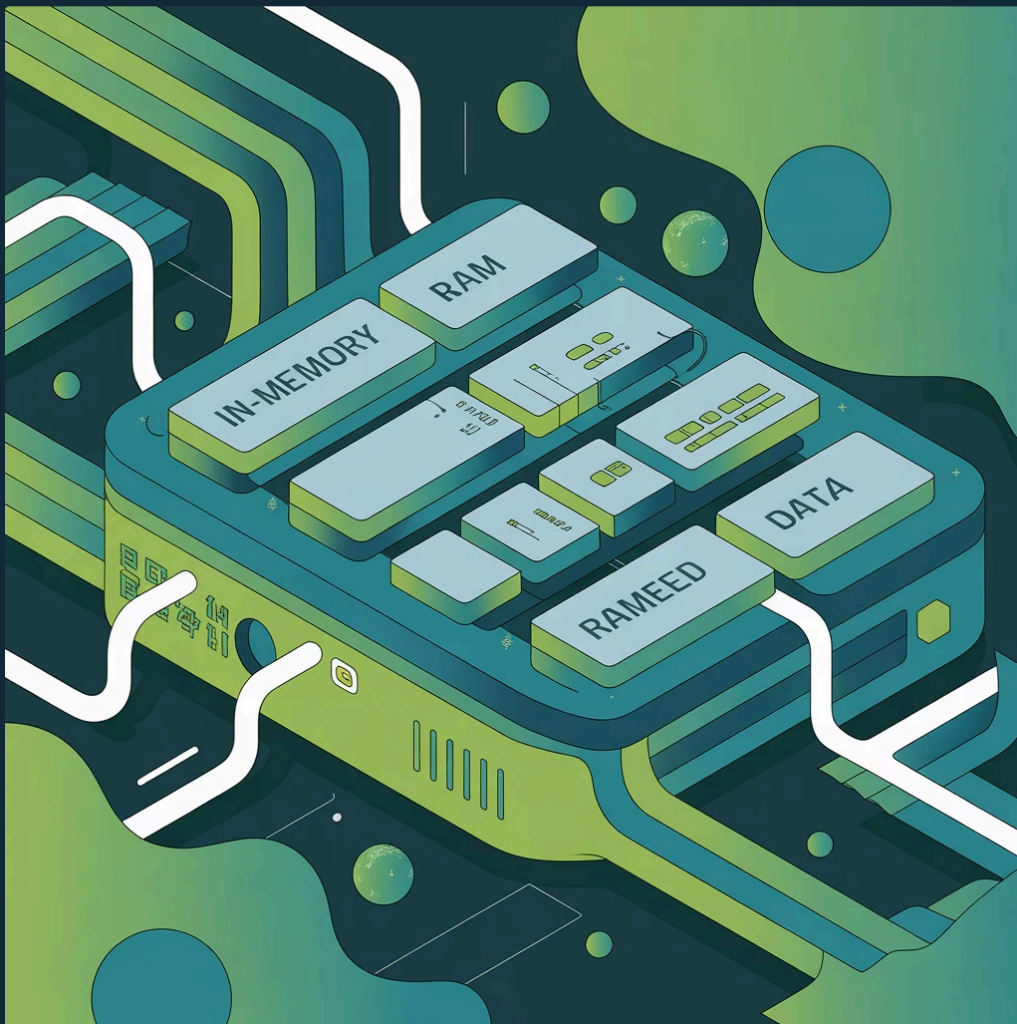
In-memory real-time queue

1. In-Memory және CEP жүйелері

In-memory processing жүйелері

RAM-да жұмыс істеу арқылы жылдамдықты арттыру:

- Redis
- Apache Ignite
- Memcached



CEP платформалары

Күрделі оқиғаларды талдау:

- Esper
- WSO2 CEP
- Flink CEP модулі



1. Қолданылу салалары

Сала	Қолданылуы
Қаржы	Онлайн транзакция, алаяқтықты анықтау
IoT	Сенсорлардан үздіксіз ақпарат
Көлік	Автономды көліктер, GPS-мониторинг
Медицина	Қашықтан пациент мониторингі
Өндіріс	Робототехника, автоматтандыру
Киберқауіпсіздік	Желілік шабуылдарды ерте тану

Қорытынды

Нақты уақыттағы деректерді өңдеу технологиялары қазіргі заманғы кибер-физикалық жүйелердің негізі болып табылады.

Stream-processing, CEP, in-memory есептеу және хабар алмасу брокерлері сияқты компоненттер жүйенің жылдамдығы мен сенімділігін қамтамасыз етеді.

Бұл тәсілдер экономиканың барлық секторларында, әсіресе цифрлық трансформация мен Industry 4.0 жағдайында аса маңызды.



Бақылау сұрақтары

- 1 Нақты уақыттағы деректерді өңдеу дегеніміз не?
- 2 Real-time және batch өңдеу айырмашылығы?
- 3 Event-driven архитектура принципі?
- 4 Spark Streaming және Flink айырмашылықтары?
- 5 CEP қандай міндеттер үшін қажет?

Пайдаланылған әдебиеттер

1. Tanenbaum A.S., Van Steen M. Distributed Systems.
2. G. Coulouris, Distributed Systems: Concepts and Design.
3. Kleppmann M. Designing Data-Intensive Applications.
4. Apache Kafka, Spark, Flink ресми құжаттары
5. Real-Time Streaming Architecture — ACM Digital Library
6. Marz & Warren. Big Data: Lambda Architecture

