

УДК 004.6

А. Зелінський, В. Лісовський

Тернопільський національний технічний університет імені Івана Пулюя, Україна

ПРОЕКТУВАННЯ ХМАРНОГО РІШЕННЯ ДЛЯ АНАЛІЗУ ПОТОКОВИХ ДАНИХ НА ОСНОВІ AWS KINESIS

A. Zeliniskyi, V. Lisovskyi

DESIGN OF A CLOUD SOLUTION FOR ANALYSIS OF STREAMING DATA BASED ON AWS KINESIS

У теперішній інформаційний час, щодня генерується більше ніж 1000 петабайт даних починаючи від мобільних пристроїв і закінчуючи розумним холодильником. Велика частка цих даних генеруються неперервним потоком. Потоківі дані - це дані, які постійно генеруються різними джерелами. Такі дані слід обробляти поступово за допомогою методів потокової обробки без доступу до всіх даних. Крім того, слід враховувати, що в даних може статися дрейф концепції, що означає, що властивості потоку можуть змінюватися з часом. Зазвичай потік даних використовується в контексті великих даних, де вони генеруються багатьма різними джерелами з високою швидкістю.

Потік даних також можна пояснити як технологію, яка використовується для доставки вмісту на пристрої через Інтернет, і вона дозволяє користувачам отримати доступ до вмісту негайно, а не чекати, поки його завантажуть. Великі дані змушують багато організацій зосереджуватися на витратах на зберігання, що викликає інтерес до озер даних і потоків даних. Озеро даних відноситься до зберігання великої кількості неструктурованих даних і є корисним через збільшення великих даних, оскільки їх можна зберігати таким чином, щоб фірми могли зануритися в озеро даних і отримати те, що їм потрібно в даний момент. Тоді як потік даних може виконувати аналіз поточкових даних у реальному часі, і він відрізняється від озер даних швидкістю та безперервним характером аналізу без необхідності попереднього зберігання даних [1].

Для обробки великої кількості поточкових даних без розгортання власних серверів використано сервіс Kinesis на хмарній платформі Amazon Web Services.

Amazon Kinesis - хмарний сервіс для обробки великої кількості розподілених потоків даних у режимі реального часу. Сервіс входить в інфраструктуру Amazon Web Services. Дозволяє розробникам витягувати будь-яку кількість даних з будь-якої кількості джерел, збільшуючи або зменшуючи кількість джерел при необхідності. Має деяку схожість по функціоналу з Apache Kafka [2].

До сервісів AWS Kinesis входить:

1. Amazon Kinesis Data Streams
2. Amazon Kinesis Video Streams
3. Amazon Kinesis Data Firehose
4. Amazon Kinesis Data Analytics

Kinesis включає в собі набір сервісів які можна поєднувати між собою та з іншими сервісами утворюючи послідовність опрацювання даних починаючи від

вхідного потоку і закінчуючи озером даних, кінцевим користувачем, структурованим сховищем даних, або все одразу [3] (див. рис 1).



Рисунок 1 - Приклад використання сервісів Kinesis

Отже використовуючи сервіси Amazon Kinesis, можна досягти ефективного опрацювання потоку даних, його масштабованість та стабільність оскільки частину роботи по підтримці на себе бере хмарне рішення.

Перелік використаної літератури:

1. What is streaming data? <https://aws.amazon.com/streaming-data/>
2. Ultimate Guide to Data Streaming with AWS Kinesis. <https://www.udemy.com/course/ultimate-guide-to-data-streaming-with-aws-kinesis>
3. Amazon Kinesis. Easily collect, process, and analyze video and data streams in real time, 2016. <https://aws.amazon.com/kinesis/>