

Есе на тему «One of Big Data's V - Velocity»

Для того щоб зрозуміти, що таке BigData без чіткої межі для розмірів «даних» варто розглянути проблеми котрі допомагає вирішити BigData, котрі можуть виступати у ролі векторів оцінки. А саме основні Vs: Velocity, Volume, Value, Veracity, Variety. Я вирішив розглянути саме Velocity, з котрою стикався найбільше разів. Однією з проблем котрі дозволяє вирішити BigData є швидкість (Velocity). А саме міру того, як швидко дані будуть надходити, оброблятися, зберігатися і як швидко можна буде отримати до них доступ після цього. Тобто це швидкість накопичення даних.

Важливість швидкості – пов'язана зі зростаючою швидкістю з якою надходять дані до організації без підмножин предметної області. Проблеми котрі раніше існували тільки для окремих предметних областей організації стали важливими для більшого контексту. Розвиток інтернету та мобільних девайсів вплинув на те, наскільки отримання продуктів та сервісів стало високо інструментальним процесом. Наприклад, значний ріст електронної комерції підвищив потребу у конкурентній перевазі. Перевагою є відповідність потребам та запитам покупців. Для оцінки цього е-комерція може збирати абсолютно усю взаємодію клієнта з системою починаючи від наповнення кошика покупок користувача, чи рухів мишкою, закінчуючи швидкістю скролінгу на певних сторінках. Проблема є не тільки у швидкості вхідних даних і їх збереженні, а й у потоковій передачі у масове сховище з подальшим пакетним використанням та отриманням. Дані мають найбільшу цінність, коли вони свіжі. Ця проблема є важливою, адже на прикладі навіть однієї лікарні можна побачити масивний і постійний потік даних, що створюється та передається до певного місця. Дані можуть надходити з

різних місць, включаючи медичні апарати, датчики дверей, контролю повітря, системи відеонагляду і навіть зі смартфонів. Дуже часто потрібно отримати певну інформацію (output) з цих даних в режим реального часу. Чим тісніша петля зворотнього зв'язку, тим корисніша система, чи у прикладі електронної комерції більша конкурентна перевага.

На мою думку до появи big data цю проблему вирішували шляхом створення величезної кількості мікро сервісів, котрі окремо відповідали за збір своїх даних предметної області, збереження (Data Mart), подальшу обробку і дуже обмежений вивід інформації. Це потребувало великих зусиль для синхронізації даних, унеможливлювало об'єднання даних з різних предметних областей. Також могла використовуватися Data Warehouse система із дуже складним ETL процесом.

Аналітика великих даних виконує пакетний аналіз і обробку збережених даних - це дані на рівні функцій або хмарних сховищ великих даних, таких як Amazon S3 і Azure Blob Storage. Big Data аналітика зазвичай використовується для узагальнення спостережень, аналізу шаблонів і виявлення інцидентів. Дані ефективно обробляються історично (пакетно проти warehouse).

Мав практику цієї проблеми якраз з електронною комерцією. Потрібно було збирати й обробляти інформацію про дії користувача на платформі. Тобто не кінцеву інформацію у вигляді успішних замовлень чи популярності товарів певних категорій, а саме інтерактивну взаємодію користувачів з елементами платформи, кліки скролінг, час перегляду певних елементів, реакцію потенційного покупця на акції, знижки, формат подання цих новин. Оскільки GTM не надавав усіх достатніх даних та можливостей, було вирішено звернутися до Contentsquare. Це платформа для оптимізації досвіду для бізнесу, яка фіксує статистику поведінки відвідувачів і розроблена для

збільшення конверсій, що також використовує штучний інтелект для надання автоматичних рекомендацій. Якщо чесно, отримавши доступ до статистики, логіки збору та прикладів зібраної був спантеличений та частково нажаханий. Але з іншого боку – це корисна інформація котра допомогла зменшити bounce rate, покращити механізм рекомендацій, поліпшити таргетинг та значно збільшити конверсію.