**1.Wstęp**

Rozwój Internetu jaki można zaobserwować w przeciągu ostatnich 20 lat, spowodował powstanie wiele systemów webowych, posiadających ogromne ilości danych dotyczących różnych aspektów życia. Każda ze stworzonych aplikacji ma na celu ułatwienie oraz przeniesienie podstawowych czynności do internetu. W bardzo szybkim tempie, pojawiły się witryny oferujące usługi w dziedzinie handlu, bankowości czy sportu. Jednakże najpopularniejszą grupą aplikacji, z których korzystają użytkownicy, stały się portale społecznościowe. Dane, jakie są dostępne za pomocą tych portali, są to informacje dotyczące życia większości osób, używających sieć internetową. Każdy z użytkowników, posiadających konto w którymś z portali społecznościowych, udostępnia swoje zdjęcia, opinie oraz informacje dotyczące wydarzeń, w jakich bierze udział. Takie działanie, prowadzi do sytuacji, gdzie serwery tych witryn, posiadają ogromną wiedzę na temat użytkowników w sieci. Jednym z celów tej pracy jest pokazanie, jakie informacje można odczytać z analiz danych, które są zawarte na publicznych profilach stron społecznościowych.

Biorąc pod uwagę ilość kont oraz informacji dostępnych na serwerach, aktualnie najczęściej używanym portalem społecznościowym, jest aplikacja stworzona przez Marka Zuckerberga o nazwie Facebook. Użytkownicy tej aplikacji, tworzą bogaty profil swojej osoby, udostępniając zdjęcia oraz różne informacje na temat życia, takie jak wykształcenie czy rzeczy, które użytkownik lubi. Dodając inne osoby do siatki swoich znajomych, profile zawierają informacje, na podstawie których, można zbudować pełen obraz danej osoby.

Drugą aplikacją internetową, pod względem popularności jest Instagram. Główną funkcjonalnością tego produktu, jest wgrywanie zdjęć, opisanych „hashtagami”, które pokazują wydarzenie, które ma lub miało miejsce. Dziennie na serwery Instagrama, wgrywa się więcej niż 80 milionów zdjęć dziennie. Liczba ta pokazuje jak dokładnie użytkownicy dokumentują swoje życie, poprzez rozbudowę profilu instagramowego. Podobnie jak w Facebooku, istnieje możliwość połączenia profilów ze znajomymi. Poprzez obserwacje danego użytkownika, dostaje się aktualną informacje o nowo wrzuconych zdjęciach interesujących nas osób.

Ostatnim narzędziem internetowym, który zostanie poddany analizie w aplikacji pisanej w ramach tej pracy, jest Twitter. Aplikacja ta, skupia się na publikowaniu krótkich treści, opisujących różne wydarzenia ze świata. W Twitterze, użytkownicy opisują temat swoich publikacji, podobnie jak w Instagramie, za pomocą hashtagów. Taki opis tematu, pozwala w prosty sposób pogrupować Tweety i ułatwić użytkownikom wyszukiwanie opinii na konkretne wydarzenie. By otrzymywać informacje o nowo wrzuconych Tweetach konkretnej osoby, użytkownik może zacząć obserwować daną osobę, w celu otrzymywania aktualnych informacji na jej temat, na stronie swojego konta. Aktualnie Twittera używa około 650 miliona internautów. W Polsce aplikacja ta nie jest używana na tak dużą skalę, jak za granicą, jednakże treści jakie są publikowane pozwalają na stworzenie raportów, które obrazują podejście oraz zdanie ludzi na różne tematy związane z życiem.

Wzrastająca ilość danych w sieci, stała się przyczyną wielu problemów, związanych z przetrzymywaniem oraz analizą tych danych. Ilość transakcji bazodanowych oraz czas w jakim te transakcje były przeprowadzane były niedopuszczalne z perspektywy klienta, używającego danego systemu. W tym celu, zaczęto rozwijać rozwiązania z grupy Big Data oraz Business Intelligence, które mają starają się zdefiniować struktury, które z powodzeniem przetworzą ogromną ilość danych w akceptowalnym czasie. W szybkim czasie powstały mechanizmy takie jak Hadoop czy Casandra oraz zdefiniowane zostały różne sposoby zbudowania hurtowni danych, których zadaniem jest przedstawienie raportów na podstawie dostępnej dużej ilości, zróżnicowanych danych. Jednakże, stworzenie takich raportów jest czasochłonne i dlatego wymyślona została struktura Business Intelligence – Architektura Lambda. Twórcą tej architektury jest Nathan Marz. Celem mechanizmów Business Intelligence, zawartych w tej architekturze, jest przyspieszenie dostępu do aktualnych danych oraz raportów, stworzony w ramach aktualnych informacji.

W ramach tej pracy, zostanie zbudowana aplikacja analizująca dane z trzech, wyżej wymienionych portali społecznościowych – Facebook, Twitter oraz Instagram, za pomocą mechanizmów architektury wywodzącej się z dziedziny Business Intelligence – Architektury Lambda, w celu otrzymania raportów na temat użytkowników tych portali społecznościowych.