Podstawową technologią wykorzystywaną obecnie do analizy dużych wolumenów danych, rozwijaną intensywnie w ciągu ostatnich 20 lat, jest technologia inteligencji biznesowej (BI – business intelligence), na którą składa się technologia hurtowni danych i systemy przetwarzania analitycznego OLAP. Powyższe technologie dostarczają podstawowych narzędzi do efektywnej wielowymiarowej analizy danych w celu wydobycia z przechowywanych danych użytecznej wiedzy dla celów wspomagania podejmowania decyzji. Uzupełnieniem systemów OLAP są systemy eksploracji danych, które w przeciwieństwie do systemów OLAP, w których analiza danych jest sterowana zapytaniami użytkowników, pozwalają na automatyczną analizę danych i odkrywanie wiedzy. Tradycyjne systemy hurtowni danych i przetwarzania analitycznego OLAP są przeznaczone głównie do przetwarzania danych transakcyjnych mających postać zbiorów niezależnych danych opisujących. Przykładem klasycznej analizy OLAP danych transakcyjnych jest zapytanie o sprzedaż różnych typów produktów AGD (lodówki, pralki, odkurzacze), w poszczególnych kwartałach danego roku z podziałem na województwa.

Jednakże, aktualnie, bardzo wiele systemów i aplikacji generuje i przetwarza dane, których podstawową cechą jest uporządkowanie elementów składających się na te dane. Taki typ danych nosi nazwę danych sekwencyjnych i stanowi, jak powiedzieliśmy, jeden z najpopularniejszych typów danych spotykanych aktualnie w wielu systemach informatycznych: systemy biletowe (analiza historii podróżowania pasażerów komunikacji miejskiej), systemy zarządzania przepływem pracy, aplikacje medyczne, systemy analizy zachowania użytkowników w serwisie WWW, inteligentne systemy transportowe bazujące na technologii RFID, systemy zarządzania infrastrukturą (np. inteligentne budynki, systemy zdalnego pomiaru konsumpcji energii, gazu, wody), bioinformatyka (analiza sekwencji DNA). Przetwarzanie danych sekwencyjnych było i pozostaje nadal ważnym obszarem badawczym w zakresie systemów baz danych, systemów przepływu pracy, bioinformatyki. Uzyskano w tym obszarze szereg istotnych wyników w zakresie: transakcyjnego przetwarzania danych sekwencyjnych, optymalizacji planów ich wykonywania, nowych struktur indeksowych wspierających różne typy zapytań do sekwencyjnych baz danych. Niestety, znacznie mniej zrobiono dotychczas w zakresie przetwarzania analitycznego OLAP i magazynowania danych sekwencyjnych w hurtowniach danych. Dostępne dzisiaj systemy OLAP i hurtownie danych, jak wspomnieliśmy wyżej, były projektowane z myślą o przetwarzaniu danych transakcyjnych i, stąd, nie wspierają przetwarzania analitycznego i eksploracyjnego danych sekwencyjnych. W ostatnim czasie, wraz ze wzrostem popularności danych sekwencyjnych, nastąpił gwałtowny wzrost zainteresowania problematyką analitycznego przetwarzania danych sekwencyjnych – tak zwane systemy SOLAP.

Przetwarzanie analityczne danych sekwencyjnych jest zagadnieniem trudnym ze względu na różnorodność typów sekwencyjnych danych (dyskretne, interwałowe, ciągłe) jak i złożoność agregacji danych sekwencyjnych. Dane sekwencyjne mogą być agregowane w oparciu: o wartości atrybutów opisujących elementy sekwencji, sekwencje, czy podsekwencje zawarte w sekwencjach. Dodatkowo, systemy SOLAP pozwalają na grupowanie sekwencji w oparciu o wzorce częste występujące w tych sekwencjach (np. podaj liczbę pasażerów podróżujących środkami komunikacji publicznej wg. wzorca dom-praca-praca-dom - XYYX). Ten rodzaj zapytań nazywamy zapytaniami typu „pattern-based”.

Celem projektu jest analiza i zaproponowanie rozwiązań w zakresie problematyki zaawansowanego przetwarzania analitycznego oraz eksploracji danych sekwencyjnych w hurtowniach danych. W ramach projektu analizowane będą dwa zasadnicze typy danych sekwencyjnych – sekwencje danych kategorycznych, nazywane również dyskretnymi sekwencjami danych, oraz sekwencje danych interwałowych. W ramach projektu jest planowana realizacja następujących zadań szczegółowych: (1) opracowanie formalnych modeli danych pozwalających na przetwarzanie dyskretnych oraz interwałowych sekwencji danych, (2) opracowanie języka zapytań dla potrzeb analitycznego przetwarzania danych sekwencyjnych, (3) opracowania i analiza architektur składowania danych sekwencyjnych wspierających przetwarzanie analityczne, (4) opracowanie struktur indeksowych wspierających efektywne przetwarzania analityczne danych sekwencyjnych, (5) opracowanie i ocena efektywności nowych algorytmów eksploracji danych sekwencyjnych różnego typu, w szczególności, algorytmów klasyfikacji i grupowania, oraz (6) opracowanie i implementacja systemu hurtowni danych sekwencyjnych i systemu OLAP udostępnionych jako usługa publiczna w architekturze SaaS (software as a service).