
ANALIZA REALIZACJI PIERWSZEGO ETAPU WDROŻENIA HURTOWNI DANYCH GLOBAL TERRORISM

Kraków, 10.12.2019

Autorzy opracowania:

Ania Gajda

Ania Gogól

Spis treści

1. Wstęp	4
1.1. Cele hurtowni „Global Terrorism”	4
1.2. Terroryzm	4
1.3. Zasady wdrażania hurtowni danych i raporty	4
2. Harmonogram i organizacja projektu.....	5
2.1. Lista zadań i harmonogram	5
2.2. Struktura organizacyjna projektu	6
2.2.1. Proponowana struktura organizacyjna projektu i schemat zespołu.....	6
2.2.2. Podział ról w zespole tworzącym i obsługującym hurtownię danych	6
2.2.3. Nadzór projektu	7
2.2.4. Kierownik projektu po stronie Bunny Hop	7
2.2.5. Kierownik projektu po stronie YYY	7
2.3. Obszary odpowiedzialności	8
2.3.1. Obszar odpowiedzialności za dane źródłowe	8
2.3.2. Obszar odpowiedzialności za integrację i uporządkowanie danych	8
2.3.3. Obszar odpowiedzialności za nadawanie uprawnień.....	8
2.4. Rodzaje dokumentów	8
3. Wstęp do specyfikacji interfejsu hurtowni danych	9
3.1. Model interfejsów	9
3.2. Spis interfejsów	9
3.3. Specyfikacja interfejsów.....	9
4. Analiza wymagań biznesowych.....	11
4.1. Wymagania funkcjonalne.....	11
4.2. Wymagania нефункционалне.....	11
4.3. Przypadki użycia.....	11
5. Aspekty techniczne	14
5.1. Warstwa zasilania	14
5.1.1. Szczegóły dotyczące warstwy zasilania	14
5.1.2. Szczegóły bazy danych z Wikipedii.....	14
5.1.3. Szczegóły bazy danych Global Terrorism Database	14
5.1.4. Porównanie obydwu źródeł danych.....	15
5.1.5. Web scraping	15
5.2. Praca z Azure DevOps	16
5.2.1. Aktualizacja i integracja.....	16

5.2.2. Organizacja zadań	16
5.2.3. Nadzór nad systemem kontroli wersji Git	17
5.3 Warstwa prezentacji i analizy.....	17
5.3.1. Power BI	17
5.3.2. SQL Server Analysis Services.....	17
5.3.3. Artificial Intelligence.....	17
6. Szczegóły dotyczące wdrożenia modułu „Hurtownia danych” w YYY.....	18
6.1. OLAP: Atak terrorystyczny.....	18
6.2. OLAP: Poszkodowani.....	19
6.3. OLAP: Sprawcy Ataków	20
7. Czas i koszty wdrożenia.....	22
7.1. Czas wdrożenia i licencje.....	22
7.2. Wynagrodzenie zespołu	23

1. Wstęp

1.1. Cele hurtowni „Global Terrorism”

Celem stworzenia hurtowni danych o nazwie Global Terrorism jest łatwy i szybki dostęp do danych potrzebnych następnie do analizy dla potrzeb instytucji rządowych, którym przesyłane są wyniki w formie comiesięcznych raportów.

1.2. Terroryzm

Z uwagi na powagę zagadnienia chcę zająć się usystematyzowaniem pozyskiwanych przeze mnie danych i analizowaniem ich tak aby móc przedstawić spójne, łatwe do odczytania dla wszystkich informacje. Będą one pomocne do rozpowszechniania wiedzy na temat możliwości zapobiegania zdarzeniom, które są istotne z punktu widzenia bezpieczeństwa oraz jako pomoce naukowe podczas szkoleń. W średniej i długiej perspektywie może się okazać, że brak świadomości zagrożenia może skutkować brakiem możliwości nawiązania współpracy ze społeczeństwem w walce z terroryzmem tam, gdzie to konieczne. Terroryzm wszedł już do kanonu stosunków międzynarodowych i niezbędne jest rozpowszechnianie wiedzy na temat sposobów radzenia sobie z tym zagrożeniem.

Firma YYY pozyskuje dane o atakach terrorystycznych na całym świecie, aby następnie je opracować i rozpowszechniać na stronach rządowych.

1.3. Zasady wdrażania hurtowni danych i raporty

Jest kilka podstawowych zasad, na jakich oparty jest efektywny proces budowy hurtowni danych. Należą do nich:

- zapewnienie strategicznego kontekstu analiz dostarczanych przez hurtownię danych
- przyrostowa, iteracyjna budowa systemu
- oparcie budowy składników hurtowni danych na definicjach architektury
- ścisła współpraca z użytkownikami.

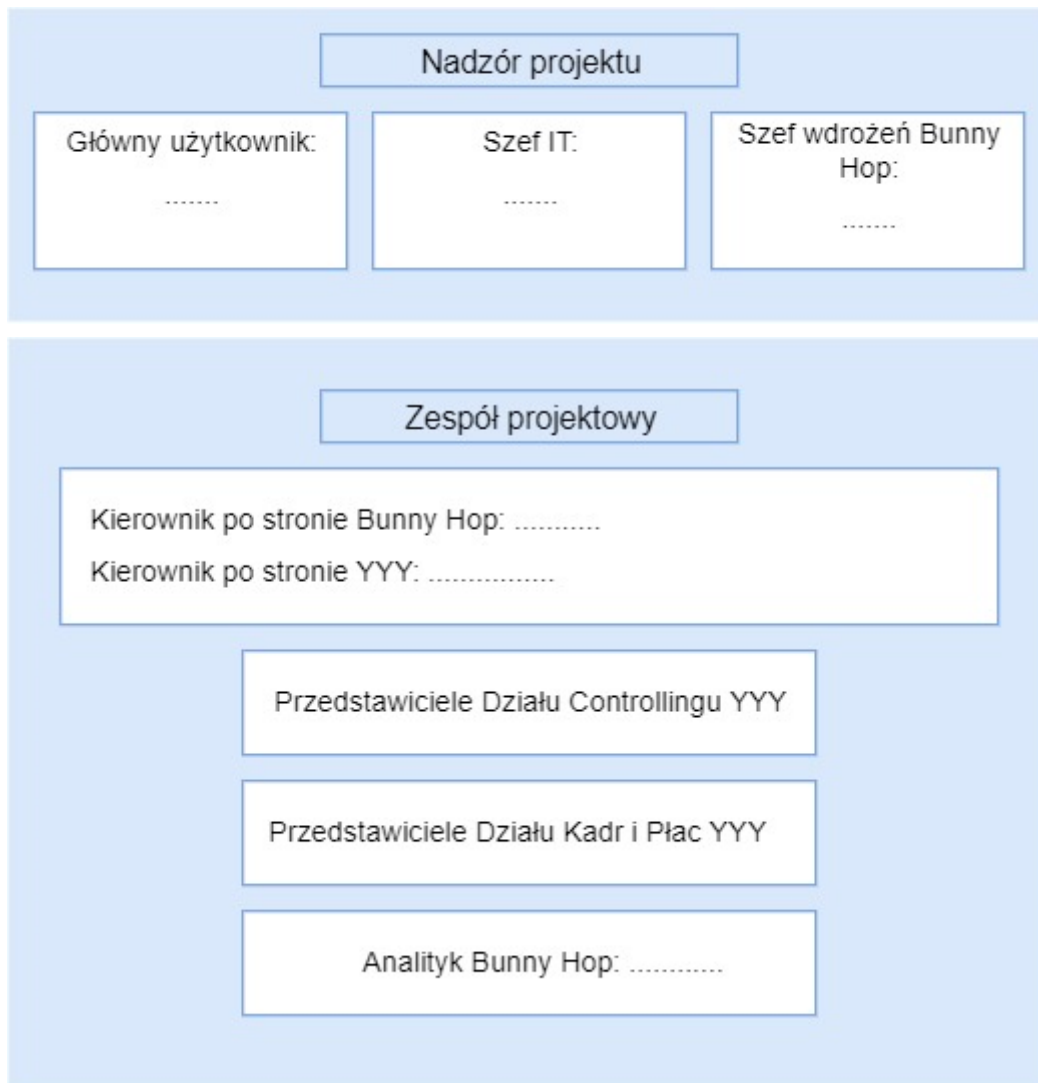
2. Harmonogram i organizacja projektu

2.1. Lista zadań i harmonogram

Nazwa zadania	Osoby odpowiedzialne	Data rozp. plan.	Data zak. plan.
Wdrożenie modułu Hurtowni Danych w YYY	Bunny Hop, YYY	17/01/2020	20/03/2020
Prace analityczne	Bunny Hop	19/01/2020	02/02/2020
Specyfikacja interfejsów	Bunny Hop	19/01/2020	24/01/2020
Analiza przedwdrożeniowa	Bunny Hop	28/01/2020	02/02/2020
Prace wdrożeniowe	Bunny Hop	17/01/2020	20/03/2020
Wykonanie interfejsów	Bunny Hop	26/01/2020	02/03/2020
Interfejs Organizacja	Bunny Hop	26/01/2020	15/02/2020
Interfejs Lokalizacja	Bunny Hop	26/01/2020	15/02/2020
Interfejs Typ	Bunny Hop	15/02/2020	02/03/2020
Interfejs Broń	Bunny Hop	15/02/2020	02/03/2020
Interfejs Cel	Bunny Hop	15/02/2020	02/03/2020
Interfejs Data	Bunny Hop	26/01/2020	15/02/2020
Obszar Zamachowy	Bunny Hop	17/01/2020	20/03/2020
Wykonanie kostki Atak terrorystyczny	Bunny Hop	04/03/2020	16/03/2020
Wykonanie kostki Poszkodowani	Bunny Hop	04/03/2020	16/03/2020
Wykonanie kostki Sprawcy Ataków	Bunny Hop	04/03/2020	16/03/2020
Integracja danych z dwóch źródeł	Bunny Hop	17/01/2020	20/03/2020

2.2. Struktura organizacyjna projektu

2.2.1. Proponowana struktura organizacyjna projektu i schemat zespołu.



2.2.2. Podział ról w zespole tworzącym i obsługującym hurtownię danych

Hurtownia danych nie jest produktem gotowym, który można kupić „z półki”. Budowa hurtowni danych wymaga powołania zespołu tworzącego hurtownię oraz jej eksploatację. Osoby zaangażowane w proces tworzenia i eksploatacji hurtowni danych to między innymi:

- administrator hurtowni danych,
- kierownik zespołu tworzącego hurtownię danych,
- analityk biznesowy,
- architekt systemowy,
- specjalista ETL,
- projektant interfejsu użytkownika,
- specjalista OLAP,
- projektant modelu danych.

Administrator HD jest odpowiedzialny za integrację i koordynację metadanych i danych z różnych źródeł danych oraz za zarządzanie źródłami danych, fizyczny projekt BD, tworzenie kopii zapasowych i odzyskiwanie bazy po awarii, bezpieczeństwa, wydajności i strojenia. Kierownik zespołu projektowego jest odpowiedzialny za zarządzanie zespołem, za przestrzeganie wymagań biznesowych i koordynację planów strategicznych, za rozwój projektu hurtowni danych, oraz za wybór członków zespołu projektującego HD o określonych kompetencjach i umiejętnościach do wykonywania zadań. Analityk biznesowy jest odpowiedzialny za określenie jakie informacje z hurtowni danych są potrzebne dla celów zarządzania. Architekt systemowy jest odpowiedzialny za rozwój i wdrażanie infrastruktury technicznej HD, od konfiguracji sprzętowej i programowej serwera bazy danych po stanowisko użytkownika końcowego.

Metodyka tworzenia hurtowni danych:

Faza	Wejście	Wyjście	Wykonawca
Analiza systemu informacyjnego	Istniejąca dokumentacja	Schemat operacyjnej bazy danych	Projektant modelu bazy danych, menedżer systemu informacyjnego
Specyfikacja wymagań	Schemat operacyjnej bazy danych	Tabela faktów	Projektant modelu bazy danych, użytkownicy końcowi
Projektowanie koncepcyjne	Tabela faktów	Schemat wymiarów	Projektant HD
Walidacja schematu wymiarowego	Schemat wymiarowy	Model logiczny	Specjalista ETL, Użytkownik końcowy
Projekt logiczny HD	Schemat wymiarowy, model logiczny	Logiczny schemat HD	Projektant HD

2.2.3. Nadzór projektu

Ścisła współpraca z użytkownikami jest podstawowym zadaniem w każdym projekcie informatycznym, jednak w przypadku budowy hurtowni danych wzrasta zarówno rola, jak i wyzwanie niesione przez ten postulat. "Ścisła współpraca" to w tym przypadku postulat dotyczący użytkownika, którego czas jest wyjątkowo cenny. Użytkownik ten jest dla nas źródłem dwóch rodzajów wymagań.

1. Przegląd dokumentów projektu: harmonogramu, notatek ze spotkań, raportów okresowych, analizy, rejestru zagrożeń, kluczowych procedur
2. Podejmowanie decyzji w sytuacji, gdy kierownik projektu sygnalizuje opóźnienie zadań, za które odpowiada

2.2.4. Kierownik projektu po stronie Bunny Hop

- Koordynowanie zadań wykonywanych po stronie Bunny Hop
- Uczestniczenie we wszystkich spotkaniach roboczych w YYY a w sytuacji gdy jest to niemożliwe wyznaczanie zastępców
- Tworzenie i aktualizacja dokumentacji projektu: harmonogramu, notatek ze spotkań, raportów okresowych, analizy
- Ścisła współpraca z kierownikiem po stronie YYY

2.2.5. Kierownik projektu po stronie YYY

- Koordynowanie zadań wykonywanych po stronie YYY

- Koordynowanie zadań wykonywanych po stronie dostawców systemów IT, z którymi łączyć się będzie hurtownia
- Organizowanie spotkań roboczych w YYY oraz uczestniczenie w nich, a w sytuacji gdy jest to niemożliwe wyznaczanie zastępców
- Ścisła współpraca z kierownikiem po stronie Bunny Hop

2.3. Obszary odpowiedzialności

Kluczowe dla projektu wdrożenia hurtowni danych jest wyodrębnienie następujących obszarów odpowiedzialności:

2.3.1. Obszar odpowiedzialności za dane źródłowe

Odpowiedzialność po stronie dostawców systemów źródłowych. Dotyczy jakości i rzetelności danych wprowadzanych do systemów, z których później przekazywane są do hurtowni. Hurtownia danych poprzez reguły walidacyjne zapewni odpowiednią jakość danych.

2.3.2. Obszar odpowiedzialności za integrację i uporządkowanie danych

Odpowiedzialność po stronie Bunny Hop. Dotyczy wszelkiej obróbki danych źródłowych do postaci, w której są one przetwarzane i prezentowane użytkownikom końcowym.

2.3.3. Obszar odpowiedzialności za nadawanie uprawnień

Odpowiedzialność po stronie KLIENTA. Dotyczy przyporządkowywania użytkowników końcowych do ról, bądź administrowania struktura organizacyjną z której wynikają uprawnienia dynamiczne

2.4. Rodzaje dokumentów

Głównym dokumentem projektu jest analiza przedwdrożeniowa, która jest cały czas uaktualniana.

Pozostałe dokumenty:

Notatki ze spotka

Każde spotkanie jest dokumentowane notatką wysyланą mailem do uczestników oraz kierownictwa projektu

Raporty miesięczne

Do 5 dnia kolejnego miesiąca sporządzany jest raport miesięczny, który zawiera:

- Omówienie harmonogramu (co się udało zrobić, czego się nie udało),
- Aktualizację analizy przedwdrożeniowej (wraz z harmonogramem),
- Listę kluczowych zagrożeń dla projektu (rejestr zagrożeń).

Rejestr zagrożeń

Lista spraw, które mogą zagrozić terminowej realizacji projektu. Każdej ze spraw przypisywana jest waga i stopień prawdopodobieństwa wystąpienia. W przypadku zagrożeń przekraczających określony próg, informowany jest nadzór projektu.

3. Wstęp do specyfikacji interfejsu hurtowni danych

W przypadku wszystkich systemów działających w ramach YYY wymiana danych będzie miała charakter jednostronny – dane będą przepływać ze źródeł zewnętrznych do hurtowni danych. Dodatkowo interfejs analityczny powinien być możliwie wygodny w użyciu i łatwy w obsłudze także dla osób, które mogą nie być bezpośrednio związane z informatyką, ale za to potrafiące bardzo dobrze analizować informacje przechowywane w tym systemie.

3.1. Model interfejsów

Dane będą wymieniane w ramach interfejsów. Pod pojęciem tym kryje się sposób, rozumiany jako ściśle określony zestaw reguł i ich opisów, w jaki programy komputerowe komunikują się między sobą. Zadaniem interfejsu programowania aplikacji jest dostarczenie odpowiednich specyfikacji podprogramów, struktur danych, klas obiektów i wymaganych protokołów komunikacyjnych. W ramach danego interfejsu dane będą:

- Udostępniane w postaci tabelarycznej jako odpowiedź zapytań SQL
- Udostępniane w postaci raportów
- Segregowane według kryteriów wybranych przez konkretnego użytkownika
- Importowane na serwer hurtowni danych (za pomocą usług SSIS)
- Przetwarzane i wstawiane do bazy hurtowni danych

3.2. Spis interfejsów

W ramach integracji z systemami w YYY:

Nazwa interfejsu	Rodzaj wymienianych danych	Częstotliwość
ORGANIZACJA	Szczegóły o organizacjach terrorystycznych.	Raz / dobę
LOKALIZACJA	Szczegóły o lokalizacji ataku.	Raz / dobę
TYP	Typ ataku terrorystycznego.	Raz / dobę
BROŃ	Typ broni użytej podczas ataku.	Raz / dobę
CEL	Cel ataku.	Raz / dobę

3.3. Specyfikacja interfejsów

ORGANIZACJA

Nazwa pola	Format	Uwagi
ID	Nvarchar	Unikalny identyfikator rekordu
NAZWA	Nvarchar	
AKTYWNA_OD	Datetime	
AKTYWNA_DO	Datetime	
AKTYWNA	Bit	0 – Nie; 1 – Tak
OBSZAR DZIAŁALNOŚCI	Nvarchar	LOKALIZACJA_ID
WIELKOŚĆ	Float	
PODORGANIZACJA	Nvarchar	Częścią jakiej większej organizacji.

LOKALIZACJA

Nazwa pola	Format	Uwagi
ID	Nvarchar	Unikalny identyfikator rekordu
KRAJ	Nvarchar	
MIASTO	Nvarchar	
LOKACJA	Nvarchar	Rodzaj budynku itp.

TYP_ATAKU

Nazwa pola	Format	Uwagi
ID	Nvarchar	Unikalny identyfikator rekordu
TYP	Nvarchar	Sposób ataku
PODTYP	Nvarchar	
GRUPA	Bit	0 – Nie; 1 – Parę osób
GRUPA_WIELKOŚĆ	Float	Wielkość grupy

BRONĀ

Nazwa pola	Format	Uwagi
ID	Nvarchar	Unikalny identyfikator rekordu
TYP	Nvarchar	
PODTYP	Nvarchar	
SZCZEGÓŁY	Nvarchar	

CEL

Nazwa pola	Format	Uwagi
ID	Nvarchar	Unikalny identyfikator rekordu
TYP	Nvarchar	Jednostka organizacyjna
PODTYP	Nvarchar	Budynek, miejsce
NARODOWOŚĆ	Nvarchar	

4. Analiza wymagań biznesowych

4.1. Wymagania funkcjonalne

1. tworzenie raportów o procentowym udziale ataków na danym terytorium (kontynencie, stanie) kraju
2. tworzenie raportów o czasie dokonywanych ataków
3. tworzenie raportów o skali wybuchu, ilości poniesionych strat (materialnych i osobowych) do jednego zamachu
4. możliwość przeglądania danych ze względu na parametry
5. możliwość sortowania danych według kryteriów
6. możliwość archiwizacji danych
7. wspomaganie decyzji
8. agregacja wielu zamachów przy braku jednolitej struktury
9. wsparcie dla zewnętrznych systemów np po przez dobieranie strategii do walki z terroryzmem

4.2. Wymagania нефunkcjonalne

1. Tworzenie raportu nie może trwać dłużej niż 20 sekund
2. Raporty powinny tworzyć się płynnie, bezproblemowo
3. Raporty tworzone są przez program Excel

4.3. Przypadki użycia

1. **Przypadek użycia:**
 - Utworzenie raportu o procentowym udziale ataków na danym terytorium:
- Aktorzy:**
 - Użytkownik, System
- Scenariusz Główny:**
 - Użytkownik zaznaczając wybiera z listy wyboru interesujące go dane:
 - Country
 - Region, State, City
 - Latitude, Longitude
 - Location
 - Summary
 - Success
 - Suicide
 - Attactype
 - Target, Target subtype
 - Motive
 - Weapon type, Weapon subtype, Weapon detail
 - Hostiges
 - Loss
 - System wyświetla zakres dat (dzień-miesiąc-rok) o atakach terrorystycznych.
 - Użytkownik wybiera zakres dat z którego chce utworzyć raport.
 - System pyta użytkownika czy chce dodatkowo posortować dane-rekordy.
 - Użytkownik nie wybiera sortowania danych-rekordów.
 - System wyświetla zakładkę z jednostką, w której ma być wykonany raport (wartość liczbowa, lub procenty).

- Użytkownik ustala jednostkę w tym przypadku procenty w której generowany jest raport ataków.
- System wyświetla poglądowy raport utworzony przez użytkownika.
- Użytkownik sprawdza czy przykładowy raport spełnia jego oczekiwania.
- System wyświetla okienko z akceptacją lub odrzuceniem raportu.
- Użytkownik akceptuje utworzony raport.
- Użytkownik z zakładki plik klika utwórz raport.

Scenariusz alternatywny:

- Użytkownik zaznaczając wybiera z listy wyboru interesujące go dane:
 - Country
 - Region, State, City
 - Latitude, Longitude
 - Location
 - Summary
 - Success
 - Suicide
 - Attactype
 - Target, Target subtype
 - Motive
 - Weapon type, Weapon subtype, Weapon detail
 - Hostiges
 - Loss
- System wyświetla zakres dat (dzień-miesiąc-rok) o atakach terrorystycznych.
- Użytkownik wybiera zakres dat z którego chce utworzyć raport.
- System wyświetla komunikat o błędzie, zwracając wiadomość, że w tym okresie nie wystąpił żaden atak terrorystyczny.

2. Przypadek użycia:

- Przeglądanie danych ze względu na różne parametry:

Aktorzy:

- Użytkownik, System

Scenariusz Główny:

- Użytkownik otwiera hurtownię danych.
- Użytkownik wybiera z panelu widoku opcję przeglądania danych.
- System wyświetla formularz do wyboru.
- Użytkownik wybiera dane, przeklikując cały formularz i zaznaczając interesujące go rekordy.
- Użytkownik po wyborze klika akceptuj na końcu formularza.
- System wyświetla zaznaczone przez użytkownika rekordy.

3. Przypadek użycia:

- Archiwizacja danych i raportów:

Aktorzy:

- Użytkownik, System

Scenariusz Główny:

- Użytkownik otwiera hurtownię danych.
- Użytkownik wybiera z panelu widoku opcję archiwizowania danych.
- System wyświetla zakres danych do zaznaczenia, które użytkownik chce zarchiwizować
- Użytkownik wybiera dane zaznaczając je po przez kliknięcie w interesujące go rekordy.

- System pyta użytkownika czy chce zarchiwizować zaznaczone dane.
- Użytkownik akceptuje wybór.
- System pyta użytkownika w jakim formacie chce zapisać dane.
- Użytkownik wybiera rozszerzenie obsługiwane przez aplikację.
- System generuje raport historyczny.

5. Aspekty techniczne

5.1. Warstwa zasilania

5.1.1. Szczegóły dotyczące warstwy zasilania

Będziemy korzystać z dwóch stron do pobierania interesujących nas danych. Jedna jest uzupełniana raz do roku, jednak jest obszerniejsza w szczegółowe dane. Druga jest uzupełniana codziennie jednak ma bardziej ogólne dane.

Do pobierania danych będziemy używać web scrapingu, w tym przypadku wykorzystamy darmową wtyczkę do przeglądarki Chrome – Web Scraper.

5.1.2. Szczegóły bazy danych z Wikipedii

Będziemy korzystać ze strony https://en.wikipedia.org/wiki/List_of_terrorist_incidents#1970%E2%80%93present, podstrony Wikipedii. Są na niej dane incydentów terrorystycznych, które nie zostały przeprowadzone przez państwo lub jego siły. Zabójstwa są wymienione na liście zamordowanych ludzi.

Definicje terroryzmu są różne, dlatego wymienione tu incydenty są ograniczone do tych, które są godne uwagi i opisane jako „terroryzm” na podstawie konsensusu wiarygodnych źródeł.

5.1.3. Szczegóły bazy danych Global Terrorism Database

Dodatkowo połączymy dane z Wikipedii z danymi z darmowej bazy danych ze strony <https://www.start.umd.edu/gtd/>. Global Terrorism Database (GTD)™ to najobszerniejsza na świecie niesklasyfikowana baza danych ataków terrorystycznych. Narodowe konsorcjum do spraw badań nad terroryzmem i odpowiedziami na terroryzm (START) udostępnia GTD™ za pośrednictwem tej strony, aby zwiększyć zrozumienie przemocy terrorystycznej, aby można ją było łatwiej studiować i pokonać. GTD™ jest produkowany przez oddany zespół naukowców i pracowników technicznych, wspieranych przez absolwentów naukowców. Zarządzanie bazą danych zapewnia rada doradcza, w skład której wchodzi jeden z najbardziej szanowanych naukowców w dziedzinie badań nad terroryzmem.

GTD™ to baza danych o otwartym kodzie źródłowym, która zawiera informacje o krajowych i międzynarodowych wydarzeniach terrorystycznych na całym świecie od 1970 r., a teraz obejmuje ponad 190 000 wydarzeń. Dla każdego wydarzenia dostępny jest szeroki zakres informacji, w tym data i miejsce incydentu, zastosowana broń, charakter celu, liczba ofiar oraz - jeśli to możliwe - odpowiedzialna grupa lub osoba.

Aktualne źródło, zawierające informacje o incydentach o podłożu terrorystycznym, które miały miejsce we wszystkich krajach świata, na przestrzeni lat 1970 -2016. Zawiera dużo szczegółowych informacji na temat każdego rekordu, takich jak motyw ataku, rodzaj użytej broni, cel, liczba ofiar itd.

Baza danych GTD jest uzupełniana raz do roku, jednak zawiera dużo dodatkowych informacji takich jak na przykład: szczegółowy cel ataku, motyw, typ użytej broni, szczegóły u użytej broni, zakładnicy, straty.

Na dzień dzisiejszy posiadamy bazę danych GTD do końca 2018 roku, która będzie uzupełniana raz w roku. W przeciwieństwie do bazy danych pobieranej z Wikipedii aktualizowanej codziennie, która będzie naszą główną bazą danych.

5.1.4. Porównanie obydwu źródeł danych

Z Wikipedii pobieramy dane:

1. data,
2. typ ataku,
3. martwi (plus ile martwych należało do organizacji terrorystycznej),
4. poszkodowani,
5. lokalizacja,
6. opis,
7. sprawca,
8. czy atak należy do bardziej ogólnej akcji i jakiej.

Z GTD pobieramy dane:

1. id,
2. rok, miesiąc, dzień,
3. państwo,
4. region, stan, Miasto,
5. dokładne dane geograficzne,
6. lokacja,
7. podsumowanie,
8. sukces,
9. samobójstwo,
10. typ ataku,
11. cel,
12. motyw,
13. szczegóły na temat użytej broni,
14. zakładnicy,
15. straty.

5.1.5. Web scraping

Do pobierania danych będziemy korzystać z Web Scraper, darmowego rozszerzenia przeglądarkowego Chrome. Jest to narzędzie do ekstrakcji danych z witryny internetowej. Bez oprogramowania do pobrania, nie jest potrzebne Python / PHP / JS.

Korzystając z tego rozszerzenia, możesz stworzyć plan (mapę witryny), w jaki sposób należy przeglądać witrynę internetową i co należy wyodrębnić. Za pomocą tych map witryn Skrobak internetowy będzie odpowiednio nawigować w witrynie i wyodrębniać wszystkie dane. Zeskrobane dane można później wyeksportować jako CSV.

5.2. Praca z Azure DevOps

W naszym projekcie korzystamy z DevOps – metodyka zespolenia rozwoju (ang. development) i eksploatacji (ang. operations) oraz zapewnienia jakości (ang. quality assurance). Metodyka ta kładzie nacisk na ścisłą współpracę i komunikację profesjonalistów z zakresu utrzymania IT (administratorów) oraz specjalistów od rozwoju oprogramowania (programistów). Uwzględnia współzależność rozwoju i utrzymania IT. Skraca czas wdrożenia funkcji w oprogramowaniu.

Metoda rozwoju oprogramowania DevOps jest wskazana dla firm, w których częstotliwość edycji jest stosunkowo wysoka. W naszym projekcie w zależności od wydarzeń ze świata może dojść do realizacji jednego wdrożenia dziennie.

5.2.1. Aktualizacja i integracja

Aktualizacja i integracja nowych danych będzie się odbywać za pomocą Azure DevOps, dzięki temu klient codziennie będzie dostawał najaktualniejsze dane.

Skrypt na Azure DevOps każdego dnia będzie sprawdzał stronę internetową Wikipedii i w przypadku różnicy w zamieszczonych tam informacjach zaciągał różniące się dane i dzięki procesowi ETL przesyłał je narzędziem SQL Integration Services do hurtowni zamieszczonej w chmurze - Azure Cloud Services.

W przypadku danych z GTD skrypt na Azure DevOps raz w roku ściąga nowy plik Excel z uzupełnionymi danymi i porównuje go do aktualnego. Różnice danych przesyła do hurtowni danych procesem ETL. Jeśli hurtownia znajdzie dane do dodania w wymiarach robi to jeśli może.

5.2.2. Organizacja zadań

Połączenie planowania przebiegu typu „przeciągnij i upuść”, elastyczne śledzenie elementów roboczych i rozbudowane możliwości śledzenia. Zadania zespołu wykonującego pracę nad hurtownią rozpisane w wbudowanych tablicach Scrum i narzędziach planistycznych, aby ułatwić swojemu zespołowi przeprowadzanie przebiegów oraz spotkań aktualizacyjnych i planistycznych.

Na Azure DevOps uwzględnione zostały:

- nazwa zadania
- opis zadania
- osoby przypisane do zadania
- planowana data rozpoczęcia prac nad zadaniem
- planowana data zakończenia prac nad zadaniem
- priorytet zadania
- etap na jakim jest zadanie (to do, doing, done).

Dzięki połączeniu repozytorium GitHub z usługą Azure Boards możliwe jest łączenie zatwierdzeń i żądania ściągnięcia z elementami roboczymi.

5.2.3. Nadzór nad systemem kontroli wersji Git

Dzięki połączeniu systemu Azure Repos z systemem kontroli wersji Git możliwe jest, podczas planowania i rozpisywania zadań w harmonogramie, uwzględnianie konkretnych zadań, wycinków kodu z GitHuba. Synchronizacja polega na łączeniu elementów pracy z obiektami tworzonymi podczas programowania, takimi jak gałęzie, zatwierdzenia, żądania pobierania i kompilacje.

5.3 Warstwa prezentacji i analizy

Do prezentacji danych wykorzystywana jest warstwa narzędzi raportujących takich jak Power BI oraz do analizy danych wykorzystana jest warstwa analityczna. Dla obliczeń predykcyjnych zostanie użyty język skryptowy Python.

5.3.1. Power BI

Proces budowy raportu w Power BI wymaga pobrania i przekształcenia danych w Edytorze zapytań > budowy modelu danych i dodatkowych miar (nie jest to jednak obowiązkowe) > stworzenia wykresów i tabel składających się na raporty powiązane z interakcją > przygotowany raport możemy jednym kliknięciem opublikować na witrynie Power BI.com.

Power BI jest aplikacją typu self-service, co oznacza, że każdy jej użytkownik powinien być w stanie samodzielnie połączyć się z danymi i zbudować raport.

- projektowania dashboardów,
- interaktywności analizy,
- oraz daje możliwość rozwoju systemu raportowania (od rozsyłania plików do raportów dostępnych przez przeglądarkę i urządzenia mobilne)
- projektowania dashboardów,
- interaktywności analizy,
- oraz daje możliwość rozwoju systemu raportowania (od rozsyłania plików do raportów dostępnych przez przeglądarkę i urządzenia mobilne).

5.3.2. SQL Server Analysis Services

Umożliwia tworzenie analitycznych modeli relacyjnych (Tabular) oraz analizowanie danych za pomocą języka DAX. Narzędzie to znane jest od dawna jako część rodziny usług Microsoft SQL Server, gdzie służyło jako serwer analityczny dla Reporting Services i PowerPivot dla Excela.

5.3.3. Artificial Intelligence

Do projektu zostanie użyta sztuczna inteligencja mająca na celu określić sezonowość ataków terrorystycznych na określonej szerokości geograficznej. Do tego celu wykorzystamy język Python. Sprawdzimy do jak wielu ataków dochodzi w określonych przedziałach czasowych i jakie to okresy (czy związane ze świętami, miesiącami, wakacjami). Dodatkowo sprawdzimy jaki to typ ataków.

6. Szczegóły dotyczące wdrożenia modułu „Hurtownia danych” w YYY

Hurtownia danych dla organizacji rządowej będzie zasilana z dwóch źródeł o różnej ilości kolumn. System zostanie wdrożony dla 6 wymiarów. Faktem w tym przypadku jest atak terrorystyczny, poszkodowani i sprawcy ataków. Fakty opisane są wymiarami: Data, Lokalizacja, Organizacja, Typ (zamachu), Broń, Cel. Miary to ilość zabitych w ataku, ilość poszkodowanych, ilość zamachowców i straty (w dolarach). Poniżej szczegółowa specyfikacja obszarów (kostek), które zostaną wykonane w ramach wdrożenia.

6.1. OLAP: Atak terrorystyczny

Raportowanie będzie uwzględniać następujące wymagania związane z obszarem zamachowym:

- System będzie dawał możliwość raportowania danych, które zawierają raporty miesięczne wysyłane w połowie każdego miesiąca dla różnego rodzaju organizacji rządowych
- Analiza ataków terrorystycznych na każdy dzień lub za określony przedział czasu, miesięcznie i narastająco z podziałem na:
 - A. Miejsce ataku terrorystycznego,
 - B. Dokładne data przeprowadzenia ataku,
 - C. Sprawca ataku
- Analiza ataku ma odbywać się przez następujące wartości:
 - A. Ilość poniesionych strat,
 - B. Ilość zgonów,
- Raporty dynamiczne będą budowane w oparciu o kostkę OLAP „Atak terrorystyczny”

Charakterystyka obszaru: Atak terrorystyczny		
Miary i wymiary		
	Nazwa	Informacje szczegółowe
Miary	Straty	Sposób agregacji: SUMA
	Martwi	Sposób agregacji: SUMA
Wymiary	1. Data	<ul style="list-style-type: none"> • Kolumna klucza: Id • Kolumna nazwy: Data_Atku • Relacje Parent – child: Nie • Widoczne hierarchie: Data_Kalendarz • Nazwy poziomów: <ul style="list-style-type: none"> A. Rok B. Miesiąc C. Dzień • Dodatkowe widoczne atrybuty: <ul style="list-style-type: none"> A. Data_Rok B. Data_Miesiąc C. Data_Rok • Postępowanie z nieznanym członkiem wymiaru: Pokaż w pozycji „Nieznany”

Charakterystyka obszaru: Atak terrorystyczny		
	2. Państwo	<ul style="list-style-type: none"> • Kolumna klucza: Id • Kolumna nazwy: Panstwo • Relacje Parent – child: Nie • Widoczne hierarchie: Panstwo_Podzial • Nazwy poziomów: <ul style="list-style-type: none"> D. Region E. Stan F. Miasto • Dodatkowe widoczne atrybuty: <ul style="list-style-type: none"> D. Panstwo_Region E. Panstwo_Stan F. Panstwo_Miasto • Postępowanie z nieznanym członkiem wymiaru: Pokaż w pozycji „Nieznany”
	3. Sprawca	<ul style="list-style-type: none"> • Kolumna klucza: Id • Kolumna nazwy: Sprawca_Atku • Relacje Parent – child: Nie • Dodatkowe widoczne atrybuty: <ul style="list-style-type: none"> G. Sprawca_Uzyta_Bron H. Sprawca_Motyw I. Sprawca_Sukces • Postępowanie z nieznanym członkiem wymiaru: Pokaż w pozycji „Nieznany”
Uprawnienia	Uprawnienia będą nadawane w zależności od np. miejsca w strukturze.	

6.2. OLAP: Poszkodowani

- Raportowanie będzie uwzględniać następujące wymogi związane z obszarem zamachowym:
- Analiza osób poszkodowanych na każdy dzień lub za określony przedział czasu, miesięcznie i narastająco z podziałem na:
 - D. Miejsce ataku terrorystycznego,
 - E. Dokładne dane geograficzne ataku,
 - F. użytą broń
- Analiza poszkodowanych w ataku ma odbywać się przez następujące wartości:
 - C. Ilość poniesionych strat,
 - D. Ilość zgonów,
 - E. Ilość osób uwięzionych (zakładników),
- Raporty dynamiczne będą budowane w oparciu o kostkę OLAP „Poszkodowani”.

Charakterystyka obszaru: Poszkodowani		
Miary i wymiary		
	Nazwa	Informacje szczegółowe
Miary	Zakładnicy	Sposób agregacji: SUMA
	Straty	Sposób agregacji: SUMA
	Martwi	Sposób agregacji: SUMA
Wymiary	1. Państwo	<ul style="list-style-type: none"> • Kolumna klucza: Id • Kolumna nazwy: Panstwo • Relacje Parent – child: Nie • Widoczne hierarchie: Panstwo_Podzial • Nazwy poziomów: <ul style="list-style-type: none"> G. Region H. Stan I. Miasto • Dodatkowe widoczne atrybuty: <ul style="list-style-type: none"> J. Panstwo_Region K. Panstwo_Stan L. Panstwo_Miasto • Postępowanie z nieznanym członkiem wymiaru: Pokaż w pozycji „Nieznany”
	2. Dokładne dane geograficzne	<ul style="list-style-type: none"> • Kolumna klucza: Id • Kolumna nazwy: DaneGeo • Relacje Parent – child: Nie • Dodatkowe widoczne atrybuty: <ul style="list-style-type: none"> A. DaneGeo_SzerokoscGeo B. DaneGeo_DlugoscGeo • Postępowanie z nieznanym członkiem wymiaru: Pokaż w pozycji „Nieznany”
	3. Użyta broń	<ul style="list-style-type: none"> • Kolumna klucza: Id • Kolumna nazwy: Bron • Relacje Parent – child: Nie • Dodatkowe widoczne atrybuty: <ul style="list-style-type: none"> A. Bron_Rodzaj • Postępowanie z nieznanym członkiem wymiaru: Pokaż w pozycji „Nieznany”
Uprawnienia	Uprawnienia będą nadawane w zależności od np. miejsca w strukturze.	

6.3. OLAP: Sprawcy Ataków

- Raportowanie będzie uwzględniać następujące wymogi związane z obszarem zamachowym:
- Analiza sprawców z podziałem na:
 - G. Motyw działania,
 - H. Odniesiony sukces,
- Raporty dynamiczne będą budowane w oparciu o kostkę OLAP „Sprawcy Ataków”.

Charakterystyka obszaru: Sprawcy Ataków		
Miary i wymiary		
	Nazwa	Informacje szczegółowe
Miary	Poszkodowani	Sposób agregacji: SUMA
	Straty	Sposób agregacji: SUMA
	Martwi	Sposób agregacji: SUMA
Wymiary	1. Samobójstwo	<ul style="list-style-type: none"> • Kolumna klucza: Id • Kolumna nazwy: Samobojstwo • Relacje Parent – child: Nie • Dodatkowe widoczne atrybuty: <ul style="list-style-type: none"> M. Samobojstwo_Wystapienie N. Samobojstwo_Miesiac O. Samobojstwo_Dzien • Postępowanie z nieznanym członkiem wymiaru: Pokaż w pozycji „Nieznany”
	2. Typ ataku	<ul style="list-style-type: none"> • Kolumna klucza: Id • Kolumna nazwy: Typ • Relacje Parent – child: Nie • Dodatkowe widoczne atrybuty: <ul style="list-style-type: none"> J. Typ_Bron K. Typ_Organizacja L. Typ_Panstwo M. Typ_Miasto • Postępowanie z nieznanym członkiem wymiaru: Pokaż w pozycji „Nieznany”
	4. Sukces	<ul style="list-style-type: none"> • Kolumna klucza: Id • Kolumna nazwy: Sukces_Ataku • Relacje Parent – child: Nie • Dodatkowe widoczne atrybuty: <ul style="list-style-type: none"> P. Sukces_Rezultat Q. Sukces_Sprawca • Postępowanie z nieznanym członkiem wymiaru: Pokaż w pozycji „Nieznany”
	5. Motyw	<ul style="list-style-type: none"> • Kolumna klucza: Id • Kolumna nazwy: Motyw_Ataku • Relacje Parent – child: Nie • Dodatkowe widoczne atrybuty: <ul style="list-style-type: none"> R. Motyw_Cel • Postępowanie z nieznanym członkiem wymiaru: Pokaż w pozycji „Nieznany”
Uprawnienia	Uprawnienia będą nadawane w zależności od np. miejsca w strukturze.	

7. Czas i koszty wdrożenia

System budowany w cyklu przyrostowym wymaga dobrze zdefiniowanej architektury, inaczej niemożliwe jest zachowanie kontroli nad jego rozwojem. Chaotyczna rozbudowa hurtowni danych, brak punktu odniesienia umożliwiającego zaplanowanie kolejnych przyrostów i modyfikacji sprawia, że adaptacja systemu do zmian biznesowych staje się coraz trudniejsza i bardziej kosztowna.

W przypadku hurtowni danych definicja architektury obejmuje dokumentację strategii biznesowych, korporacyjny model danych, dokumentację systemów źródłowych, model systemu zasilania i technologii wykorzystywanej do implementacji i wdrożenia systemu. Model technologii jest szczególnie istotny w przypadku zastosowań korporacyjnych, wymagających integracji rozwiązań różnych dostawców - od środowiska raportowania i analiz interakcyjnych, (OLAP) poprzez bazę danych dla korporacyjnej hurtowni danych, do systemu transformacji i czyszczenia danych źródłowych.

W dłuższej perspektywie zdefiniowanie architektury zwiększa wydajność procesu budowy hurtowni dzięki temu, że w poszczególnych iteracjach zminimalizowany jest obszar prac koncepcyjnych na korzyść implementacji funkcjonalności rozwiązania.

W zależności od wielkości przedsięwzięcia, wdrożenie hurtowni danych może trwać od kilku tygodni do kilkunastu miesięcy. Warto zauważyć, że większość przedsiębiorstw, po wdrożeniu systemu widząc szereg korzyści i zalet z zaimplementowanego rozwiązania, zaczyna dostrzegać nowe możliwości rozwoju i decyduje się na rozszerzenie systemu o kolejne obszary i funkcjonalności.

7.1. Czas wdrożenia i licencje

W przypadku hurtowni danych Global Terrorism DW szacujemy, że wdrożenie jej zajmie 2 miesiące (8 tygodni). W tym czasie firma Bunny Hop będzie korzystała z narzędzi:

Nazwa	Czas użytkowania (w tyg.)	Cena (PLN)	Cel
Microsoft Azure	8	-	zestaw usług do przetwarzania w chmurze
SQL Server	8	55658,28	zarządzanie relacyjnymi bazami danych
Integration Services	8	-	do budowy przepływów, integrujących dane z różnych systemów
Data Quality Services	8	-	do badania jakości danych i budowy automatycznych mechanizmów czyszczących
Data Master Services	8	-	dzięki któremu użytkownicy mogą z poziomu dedykowanego interfejsu webowego, czy też Excela, zarządzać tzw. master danymi
SQL Server Analysis Services	4	-	narzędzie do tworzenia i zarządzania hurtowniami danych
SQL Server Reporting Services	4	-	tworzenie prostych raportów
Power BI Premium	4	19501,48	zaawansowana analiza, obsługa danych big data oraz raportowanie w środowisku lokalnym i chmurze

Nazwa	Czas użytkowania (w tyg.)	Cena (PLN)	Cel
Microsoft Office 2019 Professional Plus	8	1249,00	oficjalne dokumenty firmowe, Word, Excel
Visual Studio Professional	8	175,69 (msc)	środowisko Visual Studio Professional IDE
Suma cen licencji:		76,760.14	

7.2. Wynagrodzenie zespołu

Menadżer działu IT: 22000 zł brutto/msc

Szef wdrożeń: 20000 zł brutto/msc

Główny Analityk: 18000 zł brutto/msc

Deweloperzy: 30000 zł brutto/msc

Wynagrodzenie zespołu za cały okres pracy przy projekcie (2 miesiące): 180,000 zł brutto