

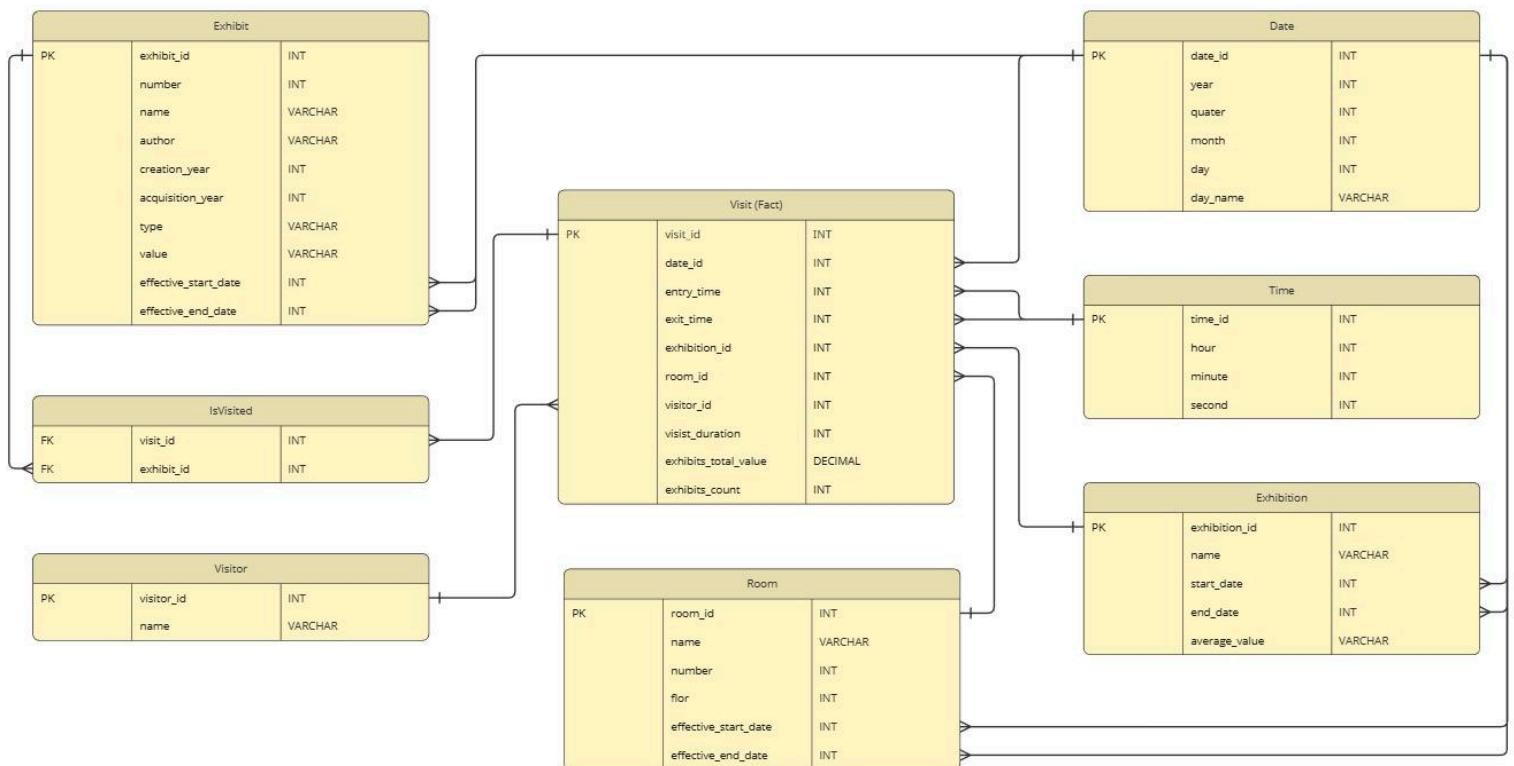
Muzeum – projekt hurtowni danych

Oliwier Komorowski 197808, Jakub Barwa 198360

1. Proces biznesowy

Hurtownia danych jest projektowana pod proces biznesowy odwiedzenia muzeum. Ten proces jest opisany w dokumencie *ProcessSpecification.pdf*.

2. Schemat Relacyjnej Bazy Danych



Visit (Fakt)		
Opisuje fakt odwiedzenia wystawy		
Atrybut	Typ	Opis
<u>visit_id</u>	INT	PK (Klucz surogatowy)
date_id	INT	FK Date Data odwiedzin
entry_time	INT	FK Time Czas wejścia
exit_time	INT	FK Time Czas wyjścia
exhibition_id	INT	FK Exhibition Wystawa
room_id	INT	FK Room Pokój
visitor_id	INT	FK Visitor Odwiedzający
visit_duration	INT	Całkowity czas odwiedzin
exhibits_total_value	DECIMAL	Suma wartości zobaczytych eksponatów
exhibits_count	INT	Ilość zobaczytych eksponatów

Date		
Opisuje konkretny dzień		
Atrybut	Typ	Opis
<u>date_id</u>	INT	PK
year	INT	Rok
quarter	INT	Kwartał
month	INT	Miesiąc
day	INT	Dzień
day_name	VARCHAR	Nazwa dnia tygodnia. Dozwolone: Poniedziałek, Wtorek, Środa, Czwartek, Piątek, Sobota, Niedziela

Time		
Opisuje konkretną godzinę (niezależnie od daty)		
Atrybut	Typ	Opis
<u>time_id</u>	INT	PK
hour	INT	Godzina
minute	INT	Minuta
second	INT	Sekunda

Room		
Opisuje konkretny pokój (SCD2)		
Atrybut	Typ	Opis
<u>room_id</u>	INT	PK
name	VARCHAR	Nazwa pokoju
number	INT	BK Numer pokoju
flor	INT	Piętro
effective_start_date	INT	FK Date Początek trwania okresu ważności danych
effective_end_date	INT	FK Date Koniec trwania okresu ważności danych

Visitor		
Opisuje odwiedzającego		
Atrybut	Typ	Opis
<u>visitor_id</u>	INT	PK
name	varchar	Imię

Exhibition		
Opisuje konkretną wystawę z uwzględnieniem przedziału średniej wartości eksponatów		
Atrybut	Typ	Opis
<u>exhibition_id</u>	INT	PK
name	VARCHAR	Nazwa wystawy
start_date	INT	FK Date Data otwarcia wystawy
end_date	INT	FK Date Data zamknięcia wystawy
avarage_value	VARCHAR	Zakres średniej wartości eksponatów. Dozwolone wartości: <10000, <100 000, <1 000 000, >1 000 000

IsVisited (Fakt)		
Opisuje fakt odwiedzin konkretnej wystawy		
Atrybut	Typ	Opis
<u>visit_id</u>	INT	FK Visit Odwiedziny
<u>exhibit_id</u>	INT	FK Exhibit Odwiedzana wystawa

Exhibit		
Opisuje konkretny eksponat z jego kategorią wartości, rokiem utworzenia i nabycia oraz typem		
Atrybut	Typ	Opis
<u>exhibit_id</u>	INT	PK
number	INT	BK Numer inwentaryzacyjny
name	VARCHAR	Nazwa eksponatu
author	INT	Autor
creation_year	INT	Rok powstania dzieła
acquisition_year	INT	Rok zakupu dzieła
type	VARCHAR	Typ dzieła. Dozwolone wartości: Obraz, Rzeźba, Fotografia, Instalacja, Grafika
value	VARCHAR	Zakres wartości eksponatu. Dzwolone wartości:<10000, <100 000, <1 000 000, >1 000 000
effective_start_date	INT	FK Date Początek trwania okresu ważności danych
effective_end_date	INT	FK Date Koniec trwania okresu ważności danych

3. Model wymiarów

a) Definicje faktów

Fakt 1 Wizyta: Odwiedzenie konkretnego pokoju muzeum (w każdym pokoju znajduje się dokładnie jedna wystawa i każda wystawa znajduje się w dokładnie jednym pokoju), konkretnego dnia, o konkretnej godzinie wejścia i wyjścia. Wizyta odbywa się przez konkretną osobę odwiedzającą.

Tabela faktu: Visit

Ziarnistość:

- określona wizyta,
- określonego odwiedzającego,
- w określonej wystawie,
- w określonym pokoju,
- w określonym dniu,
- o określonej godzinie wejścia i wyjścia,
- z określonym czasem trwania wizyty,
- z określoną łączną wartością oraz liczbą obejrzanych eksponatów.

Miary i funkcje agregujące:

- Liczba wizyt - COUNT (1),
- Czas trwania wizyty - SUM, AVG, MIN, MAX (visit_duration),
- Łączna wartość obejrzanych eksponatów - SUM, AVG, MIN, MAX (exhibits_total_value),
- Liczba obejrzanych eksponatów - SUM, AVG, MIN, MAX (exhibits_count)

Fakt 2 Zobaczenie eksponatu: Zobaczenie konkretnego eksponatu, przez konkretną osobę, podczas konkretnej wizyty. Fakt wynikający ze związku "wiele do wielu"

Tabela faktu: IsVisited

Ziarnistość:

- konkretny eksponat z konkretną wartością rynkową w konkretnym czasie
- konkretna wizyta w konkretnym czasie

Miary i funkcje agregujące:

- Liczba zobaczeń konkretnego eksponatu - COUNT (1).

b) Definicje wymiarów

Wymiar/Atrybut wymiaru	Tabela/Kolumna	Typ
Data wizyty	Date	Dimension
Rok wizyty	Date.year	Dimension attribute
Kwartał wizyty	Date.quarter	Dimension attribute
Miesiąc wizyty	Date.month	Dimension attribute
Dzień wizyty	Date.day	Dimension attribute
Nazwa dnia wizyty	Date.day_name	Dimension attribute

Czas wejścia	Time	Dimension
Godzina wejścia	Time.hour	Dimension attribute
Minuta wejścia	Time.minute	Dimension attribute
Sekunda wejścia	Time.second	Dimension attribute
Czas wyjścia	Time	Dimension
Godzina wyjścia	Time.hour	Dimension attribute
Minuta wyjścia	Time.minute	Dimension attribute
Sekunda wyjścia	Time.second	Dimension attribute
Wystawa	Exhibition	Dimension
Nazwa wystawy	Exhibition.name	Dimension attribute
Kategoria średniej wartości eksponatów na wystawie	Exhibition.avarage_value	Dimension attribute
Data otwarcia wystawy	Exhibition.start_date	Dimension attribute
Data zamknięcia wystawy	Exhibition.end_date	Dimension attribute
Pokój	Room	Dimension (SCD2)
Nazwa pokoju	Room.name	Dimension attribute
Numer pokoju	Room.number	Dimension attribute
Piętro pokoju	Room.floor	Dimension attribute
Początek trwania okresu ważności danych o pokoju	Room.effective_start_date	Dimension attribute
Koniec trwania okresu ważności danych o pokoju	Room.effective_end_date	Dimension attribute
Odwiedzający	Visitor	Dimension
Imię odwiedzającego	Visitor.name	Dimension attribute
Eksponat	Exhibit	Dimension
Nazwa eksponatu	Exhibit.name	Dimension attribute
Autor eksponatu	Exhibit.author	Dimension attribute
Rok powstania eksponatu	Exhibit.creation_year	Dimension attribute
Rok zakupu eksponatu	Exhibit.acquisition_year	Dimension attribute
Typ eksponatu	Exhibit.type	Dimension attribute

Kategoria wartości eksponatu	Exhibit.value	Dimension attribute
Początek trwania okresu ważności danych o eksponacie	Exhibit.effective_start_day	Dimension attribute
Koniec trwania okresu ważności danych o eksponacie	Exhibit.effective_end_day	Dimension attribute
Hierarchia czasu	<ul style="list-style-type: none"> • Time.hour •• Time.minute ••• Time.second 	Hierarchical dimension
Hierarchia daty wizyty	<ul style="list-style-type: none"> • Date.year •• Date.month ••• Date.day 	Hierarchical dimension
Hierarchia dnia tygodnia	<ul style="list-style-type: none"> • Date.year •• Date.quarter ••• Date.day_name 	Hierarchical dimension

Wymiar/Atrybut wymiaru	Tabela/Kolumna	Typ
Wizyta	Visit	Dimension
Numer wizyty	Visit.visit_id	Dimension attribute
Eksponat	Exhibit	Dimension
Nazwa eksponatu	Exhibit.name	Dimension attribute
Autor eksponatu	Exhibit.author	Dimension attribute
Rok powstania eksponatu	Exhibit.creation_year	Dimension attribute
Rok zakupu eksponatu	Exhibit.acquisition_year	Dimension attribute
Typ eksponatu	Exhibit.type	Dimension attribute
Kategoria wartości eksponatu	Exhibit.value	Dimension attribute
Początek trwania okresu ważności danych o eksponacie	Exhibit.effective_start_day	Dimension attribute
Koniec trwania okresu ważności danych o eksponacie	Exhibit.effective_end_day	Dimension attribute

4. Sprawdzanie wykonalności zapytań opartych na modelu wielowymiarowym

1. Porównaj średni czas przebywania w każdej sali w obecnym i poprzednim miesiącu.

Miara: Średni czas odwiedzania pokoju

Wymiar: Pokój (room_id, name, flor)

Wymiar: Data wizyty (year, month)

2. Które sale wykazały największą zmianę w średnim czasie przebywania po zmianie rozmieszczenia eksponatów?

Miara: Różnica w średnim czasie odwiedzania pokoju (obecny i poprzedni miesiąc)

Wymiar: Pokój (room_id, name, flor)

Wymiar: Data wizyty (year, month)

3. Czy nowe rozmieszczenie eksponatów miało wpływ na amplitudę średniej czasu przebywania w salach?

Miara: Średni czas odwiedzania pokoju

Wymiar: Pokój (room_id, name, flor)

Wymiar: Data wizyty (year, month)

4. Czy sale z eksponatami o wysokiej wartości przyciągnęły więcej uwagi i spowodowały dłuższy czas przebywania?

Miara: Średni czas odwiedzin pokoju

Miara: Ilość odwiedzin pokoju

Wymiar: Pokój(room_id, name, flor)

Wymiar: Eksponat (value)

Wymiar: Data wizyty (year, month)

5. Jak zmiana rodzaju eksponatów (np. obrazy vs rzeźby) wpłynęła na średni czas przebywania w poszczególnych salach?

Miara: Średni czas odwiedzin pokoju

Wymiar: Pokój(room_id, name, flor)

Wymiar: Eksponat (type)

Wymiar: Data wizyty (year, month)

6. Porównaj liczbę wejść do każdej sali w obecnym i poprzednim miesiącu.

Miara: Ilość odwiedzin pokoju

Wymiar: Pokój (room_id, name, flor)

Wymiar: Data wizyty (year, month)

7. Które sale odnotowały największy wzrost, a które największy spadek liczby odwiedzin po rotacji wystaw?

Miara: Różnica ilości odwiedzin pokoju

Wymiar: Pokój (room_id, name, flor)

Wymiar: Data wizyty (year, month)

8. Jak liczba odwiedzin sal zmienia się w zależności od wieku eksponatów prezentowanych w danej sali?

Miara: Ilość odwiedzin

Wymiar: Pokój (room_id, name, flor)

Wymiar: Eksponat (creation_year)

Wymiar: Data wizyty (year, month)

9. Czy sale z eksponatami nabytymi w ostatnich latach przyciągają więcej zwiedzających w porównaniu do sal z eksponatami starszymi?

Miara: Ilość odwiedzin

Wymiar: Pokój(room_id, name, flor)

Wymiar: Eksponat (acquisition_year)

Wymiar: Data wizyty (year, month)

10. W jaki sposób rozmieszczenie wystaw między piętrami wpłynęło na średnią liczbę wejść do poszczególnych sal?

Miara: Ilość odwiedzin

Wymiar: Pokój (flor, room_id, name)

Wymiar: Data wizyty (year, month)

5. Sprawdzanie, czy w źródłach danych znajdują się dane potrzebne do wypełnienia magazynu danych

Visit	
Opisuje fakt odwiedzenia wystawy	
Kolumna	Źródło
<u>visit_id</u>	Klucz surogatowy generowany przez bazę danych.
date_id	Data wizyty (ID). Kluczy obcy z tabeli wymiaru. Bazowany na <i>visit_date</i> w tabeli <i>Visitors</i> w źródle <i>VisitorTrack</i> .
entry_time	Czas wejścia (ID). Kluczy obcy z tabeli wymiaru. Bazowany na <i>entry_time</i> w tabeli <i>Exhibition_visits</i> w źródle <i>VisitorTrack</i> .
exit_time	Czas wyjścia (ID). Kluczy obcy z tabeli wymiaru. Bazowany na <i>exit_time</i> w tabeli <i>Exhibition_visits</i> w źródle <i>VisitorTrack</i> .
exhibition_id	Wystawa (ID). Kluczy obcy z tabeli wymiaru. Bazowany na wystawie w tabeli <i>Exhibitions</i> w źródle <i>VisitorTrack</i> .

room_id	Odwiedzony pokój (ID). Klucz obcy z tabeli wymiaru. Bazowany na pokoju w tabeli Rooms w źródle <i>VisitorTrack</i> .
visitor_id	Odwiedzający (ID). Klucz obcy z tabeli wymiaru. Bazowany na odwiedzającym w tabeli <i>Visitor</i> w źródle <i>VisitorTrack</i> .
visit_duration	Całkowity czas odwiedzin wyliczony podczas ETL na podstawie <i>entry_time</i> i <i>exit_time</i> w tabeli <i>Exhibition_visits</i> w źródle <i>VisitorTrack</i> .
exhibits_total_value	Suma wartości zobaczonych eksponatów wyliczona podczas ETL na podstawie danych z zasobu CSV z kolumny value z wierszy o wspólnych numerach inwentaryzacyjnych z eksponatami w tabeli <i>Exhibits</i> , które były na wystawie z tabeli <i>Exhibitions</i> odwiedzonych podczas wizyty.
exhibits_count	Ilość zobaczonych eksponatów wyliczona podczas ETL zliczając eksponaty w tabeli <i>Exhibits</i> , które były na wystawie z tabeli <i>Exhibitions</i> odwiedzonych podczas wizyty.

Room	
Opisuje konkretny pokój (SCD2)	
Atrybut	Źródło
<u>room_id</u>	Klucz surogatowy generowany przez bazę danych
name	Nazwa pokoju z tabeli <i>Rooms</i> w zasobie <i>VisitorTrack</i>
number	Numer pokoju z tabeli <i>Rooms</i> w zasobie <i>VisitorTrack</i>
flor	Piętro z tabeli <i>Rooms</i> w zasobie <i>VisitorTrack</i>
effective_start_date	Klucz obcy z tabeli wymiaru. Początek trwania okresu ważności danych (implementacja SCD2)
effective_end_date	Klucz obcy z tabeli wymiaru. Koniec trwania okresu ważności danych (implementacja SCD2)

Date
Opisuje konkretny dzień
Wszystkie dane generowane są na podstawie kalendarza przed procesem ETL

Time
Opisuje konkretną godzinę (niezależnie od daty)
Wszystkie dane generowane są na podstawie zegara przed procesem ETL

Visitor	
Opisuje odwiedzającego	
Atrybut	Źródło
<u>visitor_id</u>	Klucz surogatowy generowany przez bazę danych
name	Imię z tabeli <i>Visitors</i> w zasobie <i>VisitorTrack</i>

Exhibition	
Opisuje konkretną wystawę z uwzględnieniem przedziału średniej wartości eksponatów	
Atrybut	Opis
<u>exhibition_id</u>	Klucz surogatowy generowany przez bazę danych
name	Nazwa wystawy z tabeli <i>Exhibitions</i> w zasobie <i>VisitorTrack</i>
start_date	Klucz obcy z tabeli wymiaru. Data otwarcia wystawy. Bazowany na <i>exhibition_start</i> z tabeli <i>Exhibitions</i> w bazie <i>VisitorTrack</i> .
end_date	Klucz obcy z tabeli wymiaru. Data zamknięcia wystawy. Bazowany na <i>exhibition_end</i> z tabeli <i>Exhibitions</i> w bazie <i>VisitorTrack</i> .
avarage_value	Zakres średniej wartości eksponatów. Dozwolone wartości: <10000, <100 000, <1 000 000, >1 000 000

IsVisited (Fakt)	
Opisuje fakt odwiedzin konkretnej wystawy	
Atrybut	Opis
<u>visit_id</u>	Klucz obcy z tabeli wymiaru <i>Visit</i>
<u>exhibit_id</u>	Klucz obcy z tabeli wymiaru <i>Exhibit</i>

Exhibit	
Opisuje konkretny eksponat z jego kategorią wartości, rokiem utworzenia i nabycia oraz typem	
Atrybut	Źródło
<u>exhibit_id</u>	Klucz surogatowy generowany przez bazę danych
number	Klucz biznesowy z zasobu CSV z kolumny <i>id</i>
name	Nazwa eksponatu z zasobu CSV z kolumny <i>name</i>
author	Autor z zasobu CSV z kolumny <i>author</i>
creation_year	Rok powstania dzieła z zasobu CSV z kolumny <i>creation_year</i>
acquisition_year	Rok zakupu dzieła z zasobu CSV z kolumny <i>acquisition_year</i>
type	Typ dzieła. Dozwolone wartości: Obraz, Rzeźba, Fotografia, Instalacja, Grafika z zasobu CSV z kolumny <i>type</i>
value	Zakres wartości eksponatu. Dozwolone wartości:<10000, <100 000, <1 000 000, >1 000 000. Bazowane na kolumnie value z zasobu CSV
effective_start_date	Klucz obcy z tabeli wymiaru. Początek trwania okresu ważności danych (implementacja SCD2)
effective_end_date	Klucz obcy z tabeli wymiaru. Koniec trwania okresu ważności danych (implementacja SCD2)