



Muthupandian Śnieżek

ANALIZA RELACYJNYCH STRUKTUR BAZ  
DANYCH DLA POTRZEB GROMADZENIA  
I PRZETWARZANIA HETEROGENICZNYCH  
POMIARÓW HYDROMETEOROLOGICZNYCH

praca magisterska

studia dzienne

kierunek studiów: **informatyka**

specjalność: **informatyka stosowana w inżynierii środowiska**

promotor: **dr inż. Robert Szczepanek**

nr pracy: **2209**

Składam serdeczne podziękowania

**Kasi i Selvi** oraz  
**dr inż. Robertowi Szczepankowi**  
za wsparcie i cenne uwagi w trakcie pisania pracy.

Pracę tę dedukuję moim **Rodzicom**.

**SPIS TREŚCI**

|   |           |
|---|-----------|
| <b>1. WSTĘP .....</b>                                       | <b>2</b>  |
| 1.1. Cel pracy .....  | 3         |
| 1.2. Zakres pracy .....                                     | 4         |
| <b>2. POMIARY HYDROMETEOROLOGICZNE.....</b>                 | <b>6</b>  |
| 2.1. Historia pomiarów meteorologicznych .....              | 7         |
| 2.2. Pomiar hydrometeorologiczne.....                       | 8         |
| 2.3. Ewolucja pomiarów .....                                | 9         |
| 2.4. Heterogeniczny charakter danych .....                  | 9         |
| <b>3. RELACYJNE BAZY DANYCH.....</b>                        | <b>12</b> |
| 3.1. Relacyjny model baz danych .....                       | 13        |
| 3.2. Normalizacja baz danych .....                          | 15        |
| 3.3. System MySQL oraz typy danych .....                    | 16        |
| <b>4. ANALIZA WYMAGAŃ BAZY .....</b>                        | <b>20</b> |
| 4.1. Rodzaj gromadzonych danych .....                       | 21        |
| 4.2. Dane nietypowe .....                                   | 25        |
| 4.3. Założenia oraz funkcje bazy .....                      | 26        |
| <b>5. PROPONOWANE STRUKTURY BAZ DANYCH.....</b>             | <b>28</b> |
| 5.1. Podstawy teoretyczne dla stworzonych struktur.....     | 29        |
| 5.2. Wspólna część prezentowanych struktur.....             | 30        |
| 5.3. Baza metadanych pomiarów hydrometeorologicznych .....  | 37        |
| 5.4. Struktura 1 – rodzaj wykonywanego pomiaru.....         | 38        |
| 5.5. Struktura 2 – wartość pomiaru .....                    | 43        |
| 5.6. Struktura 3 – podobieństwa pomiarów .....              | 47        |
| 5.7. Problem przechowywania terminu wykonania pomiaru ..... | 54        |
| <b>6. ANALIZA I OCENA PROPONOWANYCH SCHEMATÓW .....</b>     | <b>59</b> |
| 6.1. Analiza wielkości wypełnionych baz .....               | 61        |
| 6.2. Analiza szybkości wykonywania zapytań .....            | 65        |
| 6.3. Ocena zaimplementowanych struktur .....                | 70        |
| <b>7. WNIOSKI .....</b>                                     | <b>75</b> |
| <b>8. BIBLIOGRAFIA .....</b>                                | <b>78</b> |
| <b>9. SPISY .....</b>                                       | <b>80</b> |
| 9.1. Spis ilustracji .....                                  | 81        |
| 9.2. Spis tabel .....                                       | 81        |
| <b>Abstrakt .....</b>                                       | <b>83</b> |

# **1. WSTĘP**

Zjawiska hydrometeorologiczne, zmiany zachodzące w atmosferze i związane z nimi zagrożenia wzbudzają na całym świecie coraz większe zainteresowanie. Zagrożenia lokalne takie jak powodzie, czy zagrożenia globalne jak ocieplenia klimatu, prognozowane są dzięki modelom matematycznym. Modele te tworzone są na podstawie analizy danych pochodzących z pomiarów hydrometeorologicznych. Inwestycje w rozwój technologii pomiarowych oraz obniżenie kosztów czujników automatycznych spowodowało, że zwiększana jest ilość, dokładność i częstotliwość prowadzenia pomiarów. Wszystko po to, by jak najlepiej modelować zmiany w atmosferze, by zebrać możliwie jak najwięcej informacji i jak najszybciej ostrzegać przed zagrożeniami.

Na przestrzeni lat zmieniały się metody prowadzenia pomiarów. Zmiany te miały wpływ na sposób przechowywania danych. Pomiarów prowadzone przez obserwatorów i zapisywane w dziennikach pomiarowych nie mogły być przechowywane razem z pomiarami wykonywanymi przez automatyczne czujniki. Pomiarów wykonywane tradycyjnymi metodami, przez urządzenia samopiszzące, czy pomiary wykonywane automatycznie różnią się od siebie pod względem formy w jakiej są prezentowane. Mają one więc charakter heterogeniczny i duży problem stanowi zapisanie ich w jednej bazie.

### **1.1. Cel pracy**

Celem mojej pracy jest zaprojektowanie kilku alternatywnych struktur baz danych, które będą zdolne do gromadzenia i przetwarzania heterogenicznych danych pochodzących z pomiarów hydrometeorologicznych. Ważnym aspektem pracy jest analiza zaproponowanych rozwiązań i próba odnalezienia optymalnego rozwiązania ze względu na szybkość wykonywania zapytań i wielkości baz.

## 1.2. Zakres pracy

Na początku pracy przedstawiłem historię pomiarów hydrometeorologicznych, ich ewolucję w czasie i odpowiedziałem na pytanie – co stanowi o ich heterogeniczności. Następnie przybliżyłem teorię relacyjnych baz danych i opisałem system bazodanowy MySQL, w którym zaimplementowane zostały zaprojektowane struktury.

W kolejnym etapie mojej pracy przedstawiłem dane, dla jakich utworzone zostały struktury. Zapoznanie się z nimi oraz dogłębna analiza jest podstawą do stworzenia odpowiedniej struktury. Dane, dla których stworzyłem rozwiązania są liczbami lub tekstem. W niniejszym opracowaniu nie został ujęty problem przechowywania obiektów binarnych w bazie danych, ale poruszyłem ten problem i zaproponowałem pewne uniwersalne rozwiązanie. Po zapoznaniu się z danymi postarałem się odpowiedzieć na pytanie - jakie mogą być podstawowe funkcje projektowanej bazy.

Główną częścią mojej pracy jest przedstawienie zaprojektowanych przeze mnie struktur. Zaprezentowałem trzy różne struktury, przeanalizowałem ich budowę, opisałem tabele tworzące bazę oraz przedstawiłem relacje między nimi. Pokazałem wady i zalety zastosowanych w nich rozwiązań. Dla lepszego zrozumienia tematu opisałem kilka przykładowych rekordów z bazy oraz mechanizm wykonywania zapytań w bazie.

Kolejny etap mojej pracy to analiza przedstawionych struktur a także ich ocena. Na potrzeby analizy wygenerowałem dane i zaimportowałem je do struktur zaimplementowanych w MySQL. Wyniki analizy poszczególnych struktur porównałem ze sobą i odnalazłem najlepsze rozwiązanie. Opisałem także kilka przykładowych pomysłów na ulepszenie zaprojektowanych struktur.

Ostatni etap to podsumowanie mojej pracy oraz przedstawienie wniosków płynących z analizy problemu integracji danych heterogenicznych.

## **2. POMIARY HYDROMETEOROLOGICZNE**



Na początku procesu budowy bazy danych należy przeprowadzić analizę wymagań oraz poznać dane, które będą zawarte w bazie. Aby tego dokonać trzeba zapoznać się z tematem jakim są pomiary hydrometeorologiczne. Wiele aspektów tego problemu wymaga gruntownego rozpoznania. Zacznę od genezy danych jakie zostały umieszczone w bazie, gdyż na ich podstawie zaprojektowana została struktura. Co więc stanowi o ich niejednorodności? Żeby odpowiedzieć na to pytanie należy przyjrzeć się historii prowadzenia pomiarów i ich ewolucji w czasie.

## **2.1. Historia pomiarów meteorologicznych**

Obserwacje zjawisk pogody sięgają swoimi początkami starożytności. Wzmianki świadczące o zainteresowaniu tym problemem można znaleźć już w starożytnej Grecji, Mezopotamii, Egipcie, Chinach i Indiach. Najstarszą znaną kroniką pogody są zapiski spostrzeżeń meteorologicznych dokonywanych w latach 1337-1344 w Anglii przez Williama Merle. W Polsce pierwsze obserwacje prowadzili profesorowie Wszechnicy Krakowskiej na przełomie XV i XVI wieku. Za czas prawdziwego przełomu należy uznać wiek XVII, kiedy to wprowadzono do użycia takie instrumenty pomiarowe jak termometr, deszczomierz, czy barometr. [imgw.pl, 2008]

Znaczny wpływ na rozwój meteorologii miał rozwój telekomunikacji. Za pomocą telegrafu można było przysyłać dane na znaczne odległości, z różnych stacji meteorologicznych. Ogromne znaczenie dla pomiarów meteorologicznych miało wprowadzenie technologii komputerowej do tej dziedziny nauki. Umożliwiło to szybkie przesyłanie narastającej ilości danych niezależnie od formatu. Wymiana danych w postaci graficznej, alfanumerycznej czy binarnej przestała być problemem.

## 2.2. Pomiary hydrometeorologiczne

Hydrometeorologia to nauka o związkach zachodzących pomiędzy atmosferyczną i lądową fazą cyklu hydrologicznego (parowanie, kondensacja pary wodnej w atmosferze, opad atmosferyczny). [Magnuszewski, 2001]

Wyniki obserwacji oraz pomiarów hydrometeorologicznych są zasadniczym źródłem poznawania zmian zjawisk i procesów zachodzących w atmosferze. Obserwacje dokonywane na stacjach pomiarowych przeprowadza się zarówno przez pomiary instrumentalne, jak poprzez spostrzeżenia wzrokowe obserwatorów. Warto także zaznaczyć, że bardziej rozbudowana sieć ośrodków pomiarowo – obserwacyjnych, ma bezpośredni wpływ na dokładność odwzorowania stanu aktualnego. [Janiszewski, 1988]

Bardzo istotną rolę w aspekcie pomiarów hydrometeorologicznych odgrywają obserwatorzy. Są oni specjalistami posiadającymi szczegółową wiedzę i doświadczenie, a ich praca jest istotnym elementem tworzenia zbioru danych pomiarów. Coraz częściej używa się urządzeń, które prowadzą pomiary automatycznie i są w stanie samodzielnie je rejestrować. Rola obserwatora nigdy jednak nie zostanie wyeliminowana, ponieważ pomiary prowadzone metodami tradycyjnymi zawsze będą wykonywane jako referencyjne do pomiarów automatycznych. Ponadto nie na wszystkich stacjach dostępne są nowoczesne urządzenia samorejestrujące, a w przypadku awarii takiego urządzenia pomiar musi zostać wykonany przez obserwatora.

Obecnie pomiary prowadzone są według międzynarodowych standardów, w określonych przedziałach czasowych. Umożliwia to globalną wymianę informacji i wpływa na tworzenie dokładniejszych (lepiej oddających stan rzeczywisty) modeli pogodowych.

### **2.3. Ewolucja pomiarów**

Przez lata w prowadzonych pomiarach hydrometeorologicznych następowały znaczące zmiany. Wprowadzano nowe metody pomiarów, przeprowadzano je z coraz większą dokładnością. Rola obserwatorów ograniczyła się do odczytywania wyników z coraz bardziej zaawansowanych urządzeń pomiarowych. Wraz z rozwojem technologii informatycznej pomiary hydrometeorologiczne przeżyły prawdziwy skok. Obserwatorów zaczęły wspierać automatyczne czujniki, samodzielnie dokonujące pomiarów z większą częstotliwością i automatycznie zapisujące wyniki. Zwiększenie ilości danych wymusiło potrzebę wprowadzenia sposobu ich katalogowania. Stworzono bazy danych, które otrzymywały pomiary bezpośrednio z automatycznych czujników.

Skok ten wprowadził jednak granicę między pomiarami tradycyjnymi i pomiarami automatycznymi. Tomy dzienników i kronik, pasków z pomiarami, zapisów z urządzeń samopiszących a wraz z nimi miliony bezcennych, historycznych już pomiarów, pozostawały w archiwach niszcząc. Wprowadzono je więc do baz danych, lecz ze względu na różnice dokładności, formatu zapisu i sposobu wprowadzania, nie mogły być przechowywane razem z aktualnie pozyskiwanymi danymi.

### **2.4. Heterogeniczny charakter danych**

Pomiary, które mają być umieszczone w bazie, mają charakter heterogeniczny. Oznacza to, że różnią się między sobą pod względem postaci w jakiej są reprezentowane. W środowisku informatycznym pojęcie to oznacza zróżnicowaną wielkość struktur danych, różną funkcjonalność, różne modele danych. [wikipedia.pl, 2009] Najstarsze pomiary hydrometeorologiczne zapisywane były w dziennikach pomiarowych, następnie wykorzystywano urządzenia samopiszące.

Dane rejestrowane automatycznie zapisywane są bezpośrednio w postaci binarnej w bazach danych. Należy pamiętać także o tym, że pomimo komputeryzacji i automatyzacji pomiarów, w niektórych ogródkach meteorologicznych jak i stacjach hydrologicznych, pomiary nadal są wykonywane przez obserwatorów a ich wyniki początkowo zapisywane w dzienniku pomiarowym, są ręcznie wprowadzone do bazy. Warto także zauważyć, że istnieją takie pomiary, w których komputer nie jest w stanie jeszcze zastąpić człowieka – np. określenie rodzaju chmur.

Obecnie stosuje się pewne standardy pomiarowe na całym świecie. Jednak gdy telekomunikacja nie była jeszcze tak rozwinięta jak teraz, a państwa nie współpracowały ze sobą, istniały spore rozbieżności w sposobie zapisywania pomiarów. Nie było ujednoczonych terminów dokonywania pomiarów.

Kilka powodów powstania problemu heterogenicznych danych:

- nie wszystkie ośrodki pomiarowe dokonywały tych samych pomiarów,
- nie wszystkie stosowały takie same urządzenia pomiarowe i stąd możliwa różnica w dokładności pomiarów,
- zapis pomiarów nie miał ujednoczonego charakteru,
- w przypadku awarii automatycznego czujnika, pomiaru dokonywał obserwator, jednak tak uzyskane dane nie były automatycznie wprowadzone do bazy.

Wszystkie te powody mają bezpośredni wpływ na taką różnorodność danych, dlatego stworzenie jednolitej bazy danych, która będzie w stanie przechowywać wszystkie te informacje stanowi ogromne wyzwanie.

W kolejnym rozdziale przedstawię model relacyjnych baz danych, opiszę także platformę bazodanową MySQL, której użyłem do zaimplementowania zaprojektowanych struktur.

### **3. RELACYJNE BAZY DANYCH**

Bazy danych stały się integralną częścią życia każdego człowieka, zwykle nawet bez naszej wiedzy. Informacje o nas samych znajdują się w kilkunastu a raczej kilkudziesięciu bazach już nie tylko w Polsce, ale i na całym świecie. Czym więc są bazy danych? Definicji jest tyle ile osób je wypowiadających. [Suehring, 2002]

Baza danych to zbiór danych zapisany w ściśle określony sposób w strukturach odpowiadających założonemu modelowi danych. Operuje głównie na danych tekstowych i liczbowych, lecz większość współczesnych baz umożliwia przechowywanie danych binarnych np. grafiki [wikipedia.pl, 2009]. Struktury baz danych, które przedstawię w mojej pracy oparte są na modelu relacyjnym.

### **3.1. Relacyjny model baz danych**

Model relacyjny bazy danych to najprościej mówiąc zbiór danych w postaci tabel, które są ze sobą połączone relacjami. Twórcą teorii relacyjnych baz danych jest nieżyjący już Edgar Frank Codd, który swoje postulaty opublikował w opracowaniu *A Relational Model of Data for Large Shared Data Banks*.

Dane w tym modelu reprezentowane są jako zbiór krotek, czyli uporządkowany i skończony zbiór elementów. Zbiory danych powiązane są ze sobą za pomocą encji w logiczny sposób. W ten sposób uniezależnia się widziany przez użytkownika obraz bazy danych od jej postaci fizycznej. [wikipedia.pl, 2009]

W skład modelu relacyjnego wchodzi trzy podstawowe elementy: struktura, integralność oraz manipulacja. Struktura znana jest w postaci tabel. Jest sposobem pokazania relacji zdefiniowanej w oparciu o dziedziny (zbiór wartości do którego mogą należeć wartości atrybutów w relacjach). Integralnością nazywamy ograniczenia, które są nałożone na bazę przez model relacyjny. Wyróżniamy dwie podstawowe reguły: integralność encji (klucz

główny nie może mieć wartości NULL) oraz integralność odwołań (nie mogą istnieć niedopasowane wartości klucza obcego). Trzecia składowa modelu relacyjnego, czyli element manipulacji, to zbiór operatorów relacyjnych oraz relacyjny operator przypisania, pozwalający na przypisanie relacji wyniku powstałego z wyrażenia relacyjnego opartego na algebrze relacyjnej.

Relacje to połączenia między tabelami. Umożliwiają jednoczesny odczyt informacji z wielu tabel. Pozwalają tak projektować strukturę bazy danych aby zminimalizować ilości przechowywanych informacji. Wyróżniamy trzy typy relacji. Najczęściej występującym typem relacji jest „jeden do wielu”. „Jeden” oznacza, że w danej tabeli jest tylko jeden rekord z daną. „Wielu” ponieważ w tabeli rekord może się wielokrotnie powtórzyć. Kolejnym typem relacji jest „jeden do jednego” – występuje wtedy, gdy z pewnych względów informacje, które powinny się znaleźć w jednej tabeli, są rozdzielone pomiędzy dwie tabele. Relacji tych używa się najczęściej, żeby usprawnić proces dostępu do danych lub ze względów bezpieczeństwa. Ostatnim typem relacji, który jest stosowany to relacja „wiele do wielu”. W praktyce ten typ relacji jest dwoma połączonymi relacjami typ „jeden do wielu” z wprowadzoną pomiędzy tymi relacjami tabelą łączącą. [Szczepanek, 2004] Typy relacji zostaną przedstawione na przykładach w rozdziale prezentującym zaprojektowane struktury.

Ważnym pojęciem w aspekcie relacyjnego modelu bazy danych jest pojęcie klucza. Jest to jedno lub kilka pól, które jednoznacznie identyfikują rekord. Wartość klucza dla każdego rekordu jest unikalna. Pola jednoznacznie odróżniające rekordy nazywane są Kluczem Głównym (Primary Key). Dodatkowo zdefiniowane jest pojęcie Klucza Obcego – kolumny tabeli są Kluczem Obcym, jeżeli nie są kluczem głównym tej tabeli ale ich wartości jest Kluczem Głównym



w innej tabeli. Kolumny kluczowe stanowią szczególnie ważny element relacyjnej bazy danych, ponieważ służą do tworzenia relacji między tabelami i pomagają we wprowadzeniu różnych poziomów integralności danych. [Szczepanek, 2004]

### **3.2. Normalizacja baz danych**

Normalizacja bazy danych to proces, który ma za zadanie wyeliminować powtarzające się dane w modelu relacyjnym. Idea polega na trzymaniu danych w jednym miejscu, a w razie potrzeby łączenia ich z innymi danymi. Innymi słowy rozbijamy tabele z danymi na mniejsze tabele. Tworząc bazę danych w ten sposób zwiększane jest bezpieczeństwo danych i zmniejszane ryzyko powstania niespójności.

Proces normalizacji struktury bazy danych polega na prowadzeniu jej do postaci normalnej. Wyróżniamy pięć postaci normalnych. Każda postać wprowadza coraz większe ograniczenia do struktury przy czym piąta postać normalna jest najbardziej wymagająca.

Pierwsza postać normalna zachodzi wtedy, gdy wszystkie pola w tabelach są atrybutami elementarnymi. Druga postać występuje, gdy wszystkie atrybuty są w pełni funkcjonalnie zależne od kluczy. Trzecia postać normalna pojawia się, gdy atrybuty nie wchodzące w skład klucza nie są przechodnio funkcjonalnie zależne od klucza. Czwarta i piąta postać występują niezwykle rzadko.

Nie należy jednak sztywno przestrzegać reguł normalizacji, gdyż nie zawsze daje to spodziewane efekty w postaci poprawy wydajności, czytelności oraz odporności na anomalie.

Projektując bazę danych należy kierować się raczej intuicją połączoną z doświadczeniem i zdrowym rozsądkiem niż sformalizowanymi regułami. [Szczepanek, 2004]

### 3.3. System MySQL oraz typy danych

System MySQL jest jednym z najbardziej popularnych rozwiązań bazodanowych opartych na licencji Open Source. Został stworzony przez szwedzka firmę MySQL AB, która tworzy również pozostałe narzędzia, które zostały użyte w tym opracowaniu, jak MySQL Administrator oraz MySQL Query Browser.

MySQL cieszy się opinią jednego z szybszych serwerów bazodanowych, dzięki czemu nadaje się jako serwer dla często odwiedzanych witryn www. Rozwiązania implementowane od wersji 5 zbliżają go pod względem kompatybilności z językiem SQL do PostgreSQL. [wikipedia.pl, 2009]

Poniżej przedstawię kilka powodów dla których zdecydowałem się na wybór właśnie tej platformy do zaimplementowania zaproponowanych rozwiązań:

- szybkość działania – jest uważany przez developerów za jeden z najszybszych systemów bazodanowych dostępnych na rynku,
- łatwy w użyciu – MySQL to relatywnie prosty w obsłudze system, jest zdecydowanie łatwiejszy w instalacji i administracji niż większe, komercyjne systemy,
- wsparcie dla języka zapytań – „rozumie” SQL (Structured Query Language), język używany obecnie we wszystkich nowoczesnych systemach bazodanowych,
- MySQL Server jest wielowątkowy – wielu użytkowników może podłączyć się do niego w tym samym czasie, każdy klient może korzystać z wielu baz danych symultanicznie,
- wspiera zakodowane połączenia używając protokołu SSL (Secure Sockets Layer),
- działa na nowoczesnych serwerach jak również na komputerach osobistych,

- dostępność i koszty – MySQL jako projekt Open Source jest dostępny dla użytkowników prywatnych bez opłat,
- ogromna baza pomocy online – na bieżąco rozbudowywana, także przez użytkowników prywatnych. [DuBois, 2009]

Wszystkie te zalety zdecydowały, że w niniejszym opracowaniu użyte zostały rozwiązania zaproponowane w MySQL. Oczywiście równie dużo zalet posiadają inne rozwiązania Open Source jak PostgreSQL czy SQLite – lecz dostępność materiałów oraz pomocy online przemówiły na korzyść MySQL.

Istotnym aspektem w projektowaniu struktur baz danych jest oszczędność przestrzeni dyskowej. Także w mojej pracy bardzo ważne jest odpowiednie dobranie formatu danych. Każda tabela składa się z pól. Każde pole w tabeli ma odpowiedni format odpowiadający danym, które mają się w niej znaleźć. W Tabeli 1. przedstawię zestawienie tych formatów na przykładzie zastosowanej przez mnie platformy bazodanowej MySQL.

Tabela 1. Format pól stosowanych w MySQL [MySQL.com, 2009]

| Typ                          | Typ       | Zakres  |
|------------------------------|-----------|---|
| INTEGER (całkowite)          | BIGINT    | Od -9223372036854775808 do 9223372036854775807  |
|                              | INTEGER   | Od -2147483648 do 2147483647  |
|                              | MEDIUMINT | Od -8388608 do 8388607  |
|                              | SMALLINT  | Od -32768 do 32767  |
|                              | TINYINT   | Od -128 do 127  |
| DECIMAL (zmiennoprzecinkowe) | FLOAT     | -3.402823466 * 10 <sup>38</sup> do 1.175494351 * 10 <sup>-38</sup>  |
|                              | DECIMAL   | -1.7976931348623157 * 10 <sup>308</sup> do 2.2250738585072014 * 10 <sup>-308</sup> (zapis w postaci CHAR)     |
|                              | DOUBLE    | -1.7976931348623157 * 10 <sup>308</sup> do 2.2250738585072014 * 10 <sup>-308</sup> (zapis w postaci binarnej) |
| DATE (czasowe)               | DATE      | RRRR-MM-DD  |
|                              | DATETIME  | RRRR-MM-DD GG:MM:SS   |
|                              | TIMESTAMP | RRRR-MM-DD GG:MM:SS<br>(od 1970-01-01 00:00:01 do 2038-01-09 03:14:07)  |
|                              | TIME      | -838:59:59 do 838:59:59   |
|                              | YEAR      | RRRR  |
| STRING (znakowe)             | BLOB      | Duże obiekty binarne. Obiekty te mogą przechowywać każdy rodzaj danych np. obraz, dźwięk                      |
|                              | CHAR(M)   | Łańcuch znaków o stałej długości M mogącej przyjmować wartości od 1 do 255                                    |
|                              | TEXT      | Pole tekstowe o zwykłej długości  |
|                              | VARCHAR   | Łańcuch znaków o różnej długości  |
| BOOLEAN                      | 1/0       | 1 dla wartości TRUE<br>0 dla wartości FALSE   |

Reasumując – jeżeli baza danych to zbiór uporządkowanych danych, to praca projektanta sprowadza się do utworzenia struktury, w której przechowywane będą dane, oraz wyboru informacji, które powinny znaleźć się w bazie danych. Proces tworzenia struktury dla bazy danych należy rozpocząć od analizy wymagań bazy, czyli rozpoznania ilości i typu danych, które wprowadzimy do bazy, postawienia kilku założeń, które baza musi spełniać oraz zaproponowania kilku podstawowych funkcji bazy.

## **4. ANALIZA WYMAGAŃ BAZY**

Pierwszym etapem tworzenia bazy danych jest zbadanie jakie dane mają się w bazie znaleźć. Przedstawię tylko podstawowe mierzone wielkości. Nie będę się sztywno trzymał opisanych danych, gdyż prezentowane przez mnie struktury będą miały możliwość dodawania różnych rodzajów pomiarów, nawet tych nie ujętych na etapie projektowania. Opiszę także kilka podstawowych warunków jakie muszą zająć, aby projektowana struktura spełniała założenia modelu relacyjnego. Przedstawię także zasadnicze funkcje bazy czyli zadania jakie ma wykonywać.

#### **4.1. Rodzaj gromadzonych danych**

Rozpoznanie rodzaju gromadzonych danych jest kluczowym aspektem, który należy dogłębnie przeanalizować dla stworzenia optymalnej struktury bazy danych. Temat mojej pracy mówi o heterogenicznym charakterze danych. Projektowane struktury mają za zadanie zintegrować wszystkie dane pomiarowe niezależnie od ich formatu, bez podziału na dane automatyczne czy tradycyjne.

Stworzenie struktury bazy tylko dla danych pochodzących z pomiarów automatycznych jest zadaniem o mniejszej skali trudności. Pomiarów takie są wykonywane regularnie, z tą samą dokładnością i jesteśmy w stanie wprowadzać je do bazy automatycznie. Ze względu na jednorodność danych struktury takie są łatwe w projektowaniu.

Jeżeli jednak chcemy do pomiarów automatycznych wprowadzić także pomiary wykonywane metodą tradycyjną, to zadanie stworzenia takiej struktury znacznie się utrudnia. W przypadku pomiarów tradycyjnych problemem jest integralność danych. Mogą zdarzać się przypadki, gdy osoba odpowiedzialna za przeprowadzenie pomiaru zapomni wpisać daty, lub choćby zapomni podpisać się na wyniku, a każdy taki brak informacji zmienia właściwości danej.

Należy także pamiętać o tym, że niektóre informacje zapisane w dziennikach pomiarowych mogą być nieczytelne. Pełną informację o rodzaju i typie danych będziemy mieć dopiero wtedy, gdy przejrzymy wszystkie wyniki pomiarów jakie powinny się znaleźć w naszej bazie.

Trzeba założyć jednak sytuację, w której po rozpoznaniu wszystkich danych i po stworzeniu dla nich bazy, dostajemy nową porcję danych o zupełnie odmiennej strukturze od tych już istniejących w bazie. W związku z tym struktury zaprezentowane w tej pracy stworzyłem nie znając szczegółowych informacji o wszystkich danych jakie będą do niej wprowadzane. Pozwoliło mi to zaprojektować trzy struktury, z których każda cechuje się elastycznością w celu przyjmowania danych heterogenicznych. Warto jednak poznać podstawowe dane dla jakich została zaprojektowana struktura.

#### **4.1.1. Dane meteorologiczne**

Meteorologia to dziedzina nauki o atmosferze, jej budowie, właściwościach i zachodzących w niej procesach fizycznych. Bada, jak te procesy wpływają na stan pogody na danym obszarze. Wyniki pomiarów meteorologicznych są podstawą do wielu różnorodnych analiz wykorzystywanych m.in. do prognozy pogody. [imgw.pl, 2008]

W Tabeli 2. przedstawię krótkie zestawienie pomiarów meteorologicznych jakie są obecnie wykonywane na terenie naszego kraju.



Tabela 2. Zestawienie parametrów meteorologicznych wraz z dokładnością pomiaru (1) [imgw.pl, 2008]

| Parametr  | Dokładność pomiaru |
|---|--------------------|
| Temperatura powietrza   | 0,1 [°C]           |
| Parametry wilgotności powietrza (termometrem zwykłym i zwilżonym) | 0,1 [°C]           |
| Ciśnienie atmosferyczne   | 0,1 [hPa]          |
| Kierunek wiatru   | 5 [°]              |
| Prędkość wiatru   | 1 [m/s]            |
| Widzialność pozioma   | [m]                |
| Wysokość podstawy chmury  | [m]                |
| Rodzaj chmur  | Opis obserwatora   |
| Wielkość zachmurzenia   | x/8                |
| Pogoda aktualna/ pogoda ubiegła                                   | Klucz              |

Za pomocą liczb klucza pogody bieżącej (ww) możliwe jest przekazanie informacji o występowaniu zjawisk pogodowych zarówno w chwili obserwacji, jak i w ciągu ostatniej godziny (np. zamiecie, błyskawice, mgły, deszcz). Informacja o pogodzie panującej w okresie minionym (maksymalnie za okres sześciu godzin) przekazywana jest za pomocą liczb klucza pogody ubiegłej ( $W_1W_2$ ). [imgw.pl, 2008]

Ponadto w określonych godzinach, zwanych terminami głównymi i pośrednimi, stacja meteorologiczna mierzy i wylicza zestawienia (Tabela 3):

Tabela 3. Zestawienie parametrów meteorologicznych wraz z dokładnością pomiaru (2) [imgw.pl, 2008]

| Parametr                           | Dokładność       |
|------------------------------------|------------------|
| Temperatura maksymalna powietrza   | 0,1 [°C]         |
| Temperatura minimalna powietrza    | 0,1 [°C]         |
| Temperatura w gruncie              | 0,1 [°C]         |
| Temperatura minimalna przy gruncie | 0,1 [°C]         |
| Wysokość opadu                     | 0,1 [mm]         |
| Usłonecznienie                     | 1 minuta na dobę |
| Grubość pokrywy śnieżnej           | 1 [cm]           |
| Stan gruntu                        | Klucz            |

Dodatkowo wyliczane są średnie dobowe, miesięczne i roczne dla kilku z podanych dotychczas parametrów. Jednak te dane nie muszą być przechowywane w bazie, gdyż mogą zostać wyliczone na podstawie danych cząstkowych. Należy także zaznaczyć, że wybrane stacje meteorologiczne dokonują różnych dodatkowych pomiarów parametrów nie ujętych w zestawieniu.

#### 4.1.1. Dane hydrologiczne

Hydrologia to dział nauki zajmujący się badaniem obiegu wody w przyrodzie, wodach powierzchniowych oraz wodach zawartych w litosferze i atmosferze. Hydrologia zajmuje się każdym aspektem obiegu wody. Interesuje ją nie tylko informacja, że woda płynie, ale również ilość przepływającej wody, jej prędkość oraz amplituda wahań poziomu wody w danym przekroju poprzecznym. Z pomiarów hydrologicznych możemy uzyskać informacje o stanie wody, prędkości przepływu oraz ilości przepływającej wody w jednostce czasu. [imgw.pl, 2008]

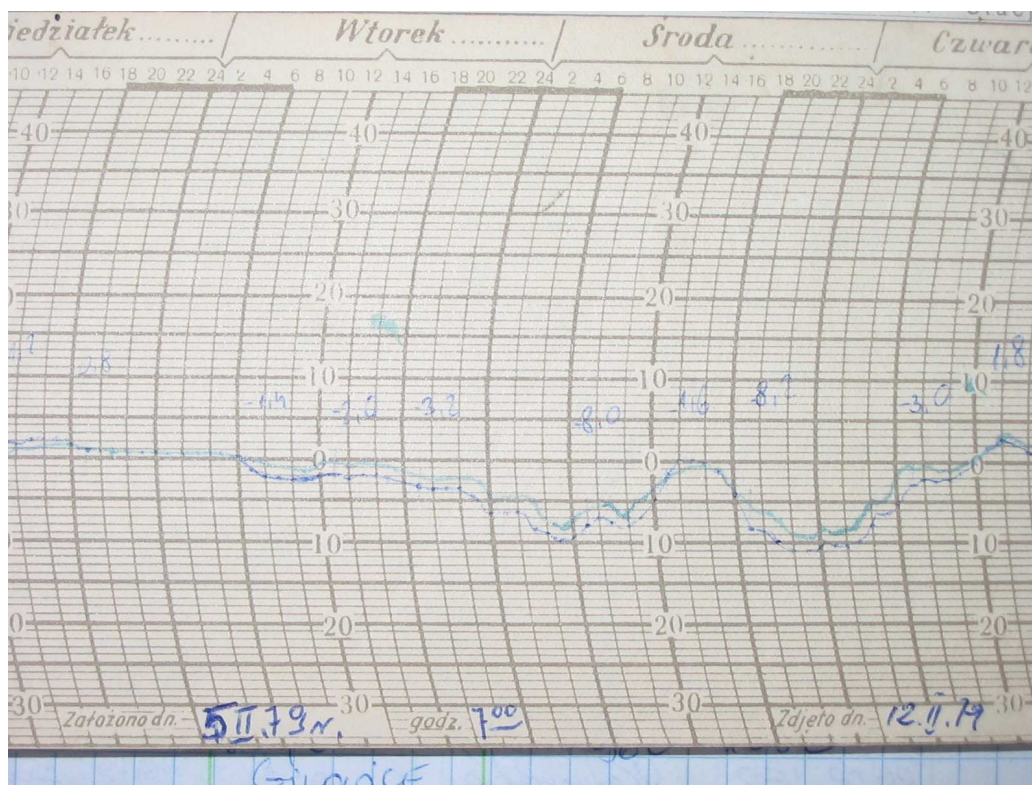
Tabela 4. Zestawienie parametrów hydrologicznych wraz z dokładnością pomiaru [imgw.pl, 2008]

| Parametr         | Dokładność                   |
|------------------|------------------------------|
| Stan wody        | 1 [cm]                       |
| Przepływ         | wyrażany w m <sup>3</sup> /s |
| Zjawiska lodowe  | 1 [cm] / 10 [%]              |
| Temperatura wody | 0,2 [°C]                     |

## 4.2. Dane nietypowe

Dane hydrometeorologiczne nie składają się jednak tylko z danych zapisywanych w postaci cyfrowej. Wyniki pomiarów z takich urządzeń jak limnigraf czy pluwiograf są wykresami mierzonej wielkości w sposób ciągły, najczęściej dla 24 godzin. Można spróbować odczytać wartości z tych wykresów np. w pełnych godzinach, ale wtedy traci się cenne informacje. Dlatego sporym problemem jest umieszczenie ich w bazie, gdyż muszą się w niej znaleźć jako obiekty binarne, które zajmują dużo przestrzeni dyskowej. Dodatkowo w interfejsie użytkownika musi znaleźć się narzędzie do prezentacji wizualnej tych pasków.

Do lepszego zobrazowania problemu na Rys.1 przedstawiam zdjęcie paska termogramu. Widoczna na nim linia to pomiar temperatury wykonany za pomocą termografu. Wyraźnie widoczne jest jaką trudność może powodować odczytanie pojedynczej wartości z tego wykresu. Możemy próbować odczytać wartości o wyznaczonych godzinach i wtedy wprowadzenie ich do projektowanych struktur nie stanowiłoby problemu. W tym przypadku należy założyć jednak pewien margines błędu.



Rysunek 1. Fotografia paska termogramu [Kinga Hajduk, 2006]

Problem ten jest na tyle złożony, że mógłby być tematem do osobnego opracowania. W mojej pracy zajmę się projektowaniem struktur zdolnych do przechowywania danych liczbowych lub tekstowych nie rozpatrując złożoności problemu jakim jest umieszczanie danych binarnych w bazie. Uznałem jednak, że warto wspomnieć o tym problemie.

### 4.3. Założenia oraz funkcje bazy

Aby zaprojektować strukturę bazy, która spełniać ma założenie modelu relacyjnego należy przedstawić kilka podstawowych założeń. Istotne jest takie zaprojektowanie struktury bazy danych aby:

- każdy rodzaj informacji, który może się potencjalnie pojawić, mógł zostać w bazie zapisany,
- każdy rodzaj informacji, który potencjalnie może być konieczny do uzyskania z bazy, mógł zostać uzyskany,

- zagwarantowana była jednoznaczność zapisanej informacji,
- możliwy był łatwy zapis, modyfikacja i usunięcie bazy informacji z bazy,
- struktura bazy nie powodowała redundancji czyli nadmiarowości (nie należy zapisywać większej ilości informacji niż jest to niezbędne). [Szczepanek, 2004]

Wszystkie te punkty należy na bieżąco weryfikować podczas procesu projektowania struktury.

Głównym zadaniem bazy jest gromadzenie danych z pomiarów hydrometeorologicznych. Oczywiście dane byłyby bezwartościowe gdyby nie można było z nich korzystać. Należy więc dodać, że zadaniem bazy jest umożliwienie zapisy i odczytu danych pochodzących z pomiarów hydrometeorologicznych. Zdolność modelowania zmian w atmosferze, przedstawiania zestawu danych na wykresach, nie leży w gestii samej bazy, lecz interfejsu lub systemu, który zostanie z nią połączony. Niemniej jednak struktury, które przedstawię muszą być tak zbudowane, by móc w jednoznaczny sposób i bezbłędnie dostarczać wszystkie wymagane informacje.

Przedstawiłem już założenia, jakie stawiane są przed projektowanymi strukturami. W kolejnym rozdziale przejdę do przedstawienia propozycji rozwiązania tego złożonego problemu - jak zintegrować heterogeniczne dane w jednej bazie.

## **5. PROPONOWANE STRUKTURY BAZ DANYCH**

Kluczowym problem przy tworzeniu każdej bazy danych jest odpowiednie skonstruowanie struktury danych. Jej konstrukcja ma duże znaczenie dla przejrzystości bazy, czytelnego ukazania zależności między danymi, szybkości działania bazy.

Według koncepcji relacyjnych baz danych, powinienem zacząć od umieszczenia wszystkich danych w jednej tabelce, następnie poprzez normalizację pierwszego, drugiego i trzeciego stopnia dojść do optymalnego rozwiązania. Niestety, dla danych heterogenicznych nawet najprostsze rozwiązanie stworzenia jednej tabeli i umieszczenia tam wszystkich danych, jest pomysłem nastroczającym wiele problemów. W związku z tym przedstawię trzy struktury bazy danych zdolnych do przechowywania danych heterogenicznych bez szczegółowego opisu procesu normalizacji.

W rozdziale tym przedstawię jedynie budowę i koncepcję działania proponowanych struktur. Wszelkich porównań i analiz dokonam w kolejnym, szóstym rozdziale.

## **5.1. Podstawy teoretyczne dla stworzonych struktur**

W poprzednim rozdziale przedstawiłem dane pochodzące z pomiarów hydrometeorologicznych. Każdy wykonany pomiar oprócz swojej wartości posiada rząd atrybutów, które go opisują i sprawiają, że pomiar ten jest unikalny. Te atrybuty to:

- miejsce, w którym dokonano pomiaru – każdy pomiar musi zawierać informację w jakim miejscu został przeprowadzony, czyli szczegółowe dane o stacji pomiarowej,
- jaki rodzaj pomiaru jest wykonywany – pomiar to jedynie nie mówiąca liczba, dopiero gdy będziemy wiedzieć jaką wielkość opisuje i w jakiej jednostce, pomiar zyskuje wartość,

- czujnik dokonujący pomiaru – ta wartość w niektórych przypadkach może być zbędna, bo możemy pomiar przypisać do miejsca w jakim się go wykonuje, jednak gdy w jednym miejscu mierzy się jakąś wartość dwoma metodami np. automatycznie i przez obserwatora, wartość ta jest niezbędna,
- czas, w jakim go przeprowadzono – każdy pomiar musi mieć informacje o czasie, kiedy został zrobiony. Wartość ta jest niezbędna do modelowania zmian klimatycznych i ich tempa. Należy w tym miejscu także dodać, że sposób w jaki wprowadzamy ten czynnik do bazy ma bardzo duży wpływ na wielkość bazy,
- informacja o osobie dokonującej pomiaru lub czy pomiar jest wykonywany automatycznie (jeżeli pomiaru dokonał obserwator informacja o nim także jest ważną wartością opisującą pomiar).

Powyższe informacje połączone ze sobą pozwalają jednoznacznie opisać pomiar. Według wszystkich tych cech możliwe jest projektowanie struktur. Oczywiście niektóre atrybuty lepiej i dokładniej opisują dane niż inne. Również muszę zaznaczyć, że nie jesteśmy w stanie stworzyć struktury bazy danych według wszystkich tych atrybutów.

Najistotniejszą kwestią w mojej pracy jest organizacja danych heterogenicznych, dlatego też dla informacji, które muszą zawierać się w bazie a nie są istotne dla samej organizacji danych, zastosuję pewne uogólnienia. Uproszczenia te pozwolą stworzyć bardziej przejrzystą strukturę, łatwiejszą do opisu i mniej złożoną.

## **5.2. Wspólna część prezentowanych struktur**

W tej części przedstawię wspólne tabele dla wszystkich struktur, aby usprawnić proces ich prezentacji. Wszystkie struktury



mają zbliżone do siebie schematy. Różnią się one jednak znacząco w najistotniejszej kwestii organizacji danych heterogenicznych, czyli w sposobie przechowywania wyników pomiarów. Tabele które stanowią część wspólną, to te z informacjami stałymi, takimi jak: spis miejsc, w których wykonuje się pomiary, rodzaje pomiarów jakie się wykonuje, spis obserwatorów dokonujących pomiarów.

Tabela 5. Tabela Miejsce

|                        |   |
|------------------------|---|
| ID_Miejsce<br>(PK)     | Indeks miejsca - unikalny numer identyfikacyjny dla danego miejsca  |
| Nazwa                  | Nazwa miejscowości gdzie przeprowadzane są pomiary  |
| Województwo            | Nazwa województwa gdzie znajduje się dana miejscowość   |
| Szerokość geograficzna | Informacja niezbędna do jednoznacznego zlokalizowania zestawu czujników   |
| Długość geograficzna   | Informacja niezbędna do jednoznacznego zlokalizowania zestawu czujników   |
| ...                    | W kolejnych polach tej tabeli możemy przedstawić pozostałe informacje wymagane do określenia miejsca. Mogą to być takie dane jak adres, położenie nad poziomem morza, informacja jakiej jednostce podlega dana lokalizacja itp. |

Tabela Miejsce zawiera wszystkie informacje o położeniu czujników. Oczywiście informacje zaprezentowane w powyższej tabeli są przykładowe. Ilość pól nie jest w żaden sposób ograniczona, ale musi być zdefiniowana przy tworzeniu bazy. Warto także wprowadzić pole informacyjne czy dane miejsce prowadzi nadal pomiary, czyli czy dana stacja pomiarowa jest czynna. Dodatkowo można wprowadzić pole uwagi, gdzie będzie można wpisywać wszystkie dodatkowe informacje. Jednak tak jak zaznaczyłem na początku zastosowałem pewne uogólnienia, aby prezentowana struktura była bardziej przejrzysta. Kluczem Głównym w tej tabeli będzie pole ID\_Miejsce, które jest unikalną wartością i jednoznacznie określa położenie czujników.

Tabela 6. Tabela Obserwator

|                       |   |
|-----------------------|---|
| ID_Obserwator<br>(PK) | Indeks obserwatora - unikalny numer identyfikacyjny dla danej osoby   |
| Imię                  | Imię obserwatora prowadzącego pomiary   |
| Nazwisko              | Nazwisko obserwatora prowadzącego pomiary   |
| ...                   | Kolejne pola tej tabeli zawierać będą informacje o danej osobie. Mogą to być takie dane jak posiadany tytuł, telefon kontaktowy, adres itp. |

W powyższej tabeli zapisane będą wszystkie informacje o obserwatorach dokonujących pomiarów. Jest to niezwykle istotna wartość dla pomiarów wykonywanych tradycyjnymi metodami.

Tabela 7. Tabela Rodzaj Pomiaru

|                    |  |
|--------------------|--|
| ID_Pomiar<br>(PK)  | Indeks pomiaru - unikalny numer identyfikacyjny dla rodzaju wykonywanego pomiaru   |
| Nazwa pomiar       | Nazwa mierzonej wielkości np. temperatura, ciśnienie   |
| Jednostka          | Jednostka wykonywanego pomiaru np. [K],[hPa]   |
| Liczba/Tekst       | Pole formatu 0/1 – informacja czy rodzaj pomiaru wyraża się liczbą, czy tekstem  |
| Minimalna wartość  | Minimalna wartość jaka jest mierzona dla danego pomiaru  |
| Maksymalna wartość | Maksymalna wartość jaka jest mierzona dla danego pomiaru   |
| Dokładność         | Dokładność z jaką wykonywany jest pomiar. Dla wyniku pomiaru zapisywanego tekstowo domyślna wartość w tym polu wynosi 1. |

Tabela ta jest zbiorem wszystkich wykonywanych pomiarów. Takie rozwiązanie daje nam elastyczność i skalowalność. Nie jesteśmy ograniczeni do stosowania jednej jedynej jednostki dla pomiarów tej samej wielkości. Oczywiście wykonywaniem pomiarów i ich zapisem rządzą pewne standardy, lecz zastosowanie takiego rozwiązania powoduje przydatność bazy do przechowywania wyników pomiarów, nawet jeżeli te standardy ulegną zmianie w przyszłości. Dzięki zastosowaniu takiego rozwiązania będziemy mogli dodawać nowe

rodzaje pomiarów, nawet jeśli jednostka w której wykonujemy pomiary zmieni się w przyszłości. Możemy dla jednej mierzonej wielkości zastosować dwie różne jednostki. Jeżeli w przyszłości do pomiarów dojdzie jeszcze jedna wielkość, musimy jedynie wprowadzić kolejny rekord do tabeli.

Na podstawie powyższych tabel możemy stworzyć tabelę czujników, która będzie łączyła ze sobą wszystkie przedstawione tabele.

Tabela 8. Tabela Czujnik

|             |  |
|-------------|--|
| ID_Czujnik  | Indeks czujnika - unikalny numer identyfikacyjny dla danego czujnika   |
| ID_Miejsce  | Indeks miejsca z tabeli Miejsce, jednoznacznie określający gdzie dany czujnik się znajduje   |
| ID_Pomiar   | Indeks pomiaru z tabeli Rodzaj Pomiaru, jednoznacznie określający jakie dane czujnik wykonuje pomiary i w jakiej jednostce.  |
| Auto/Obserw | Jest to pole typu 1/0. Informacja czy pomiar został wykonany automatycznie czy przez obserwatora.  |
| Rodzaj      | Informacja jaki rodzaj czujnika jest zainstalowany – czy jest to zwykły termometr czy np. czujnik który wykonuje pomiary automatyczne. W tym przypadku możemy tu wpisać np. numer seryjny. |
| ...         | Pozostałe pole tabeli Czujnik będą się różnić w zależności od zaproponowanego rozwiązania i będą wówczas szczegółowo omawiane  |

Dzięki zastosowaniu takiego rozwiązania każdy czujnik jest jednoznacznie określony. Pole Auto/Obserw informuje o tym, czy dany pomiar wykonany jest automatycznie (pole przyjmie wartość 1), czy pomiar został przeprowadzony przez obserwatora (pole przyjmie wartość 0). Jeżeli pomiar został wykonany przez obserwatora to informacja o nim musi się również znaleźć w bazie. Nie można jednak dopisać bezpośrednio osoby do czujnika, gdyż takie rozwiązanie

ogranicza możliwość łączenia kilku osób z danym czujnikiem. Dlatego konieczne jest wprowadzenie zmodyfikowanego rozwiązania.

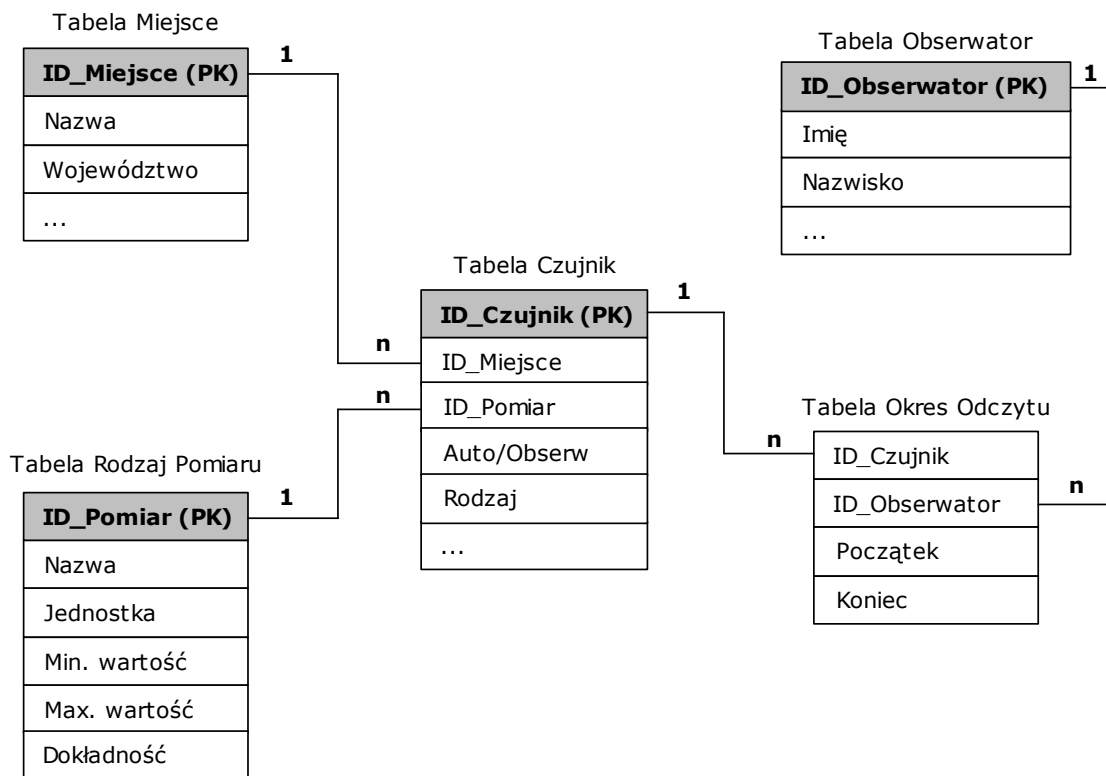
W poszukiwaniach odpowiedniego rozwiązania napotkałem na problem serii. Czy jest sens dla każdego pomiaru dopisywać informację o osobie, która go wykonała? Jeżeli w danym miejscu ze wszystkich czujników przez okres dwóch miesięcy jeden obserwator przeprowadzał wszystkie pomiary, to wpisywanie tego samego identyfikatora obserwatora dla wszystkich pomiarów nie ma sensu. Można zastosować rozwiązanie, które powie – wszystkie pomiary w tym miejscu przez dwa miesiące wykonywał dany obserwator. Dla zastosowanie tego rozwiązania wprowadzę tabelę Okres Odczytu. Jej struktura przedstawia się następująco:

Tabela 9. Tabela Okres Odczytu

|               |   |
|---------------|---|
| ID_Obserwator | Indeks obserwatora z tabeli Obserwator, który wykonywał serię pomiarów          |
| ID_Czujnik    | Indeks czujnika z tabeli Czujnik, z którego obserwator prowadził serię pomiarów |
| Początek      | Data kiedy dany obserwator rozpoczął odczyty z danego czujnika                  |
| Koniec        | Data kiedy dany obserwator zakończył odczyty z danego czujnika                  |

Takie rozwiązanie wpłynie na znaczne zmniejszenie wielkości bazy i może również skrócić czas zapisu do bazy. Jednak w przypadku odczytu z bazy czas ten może zostać wydłużony w porównaniu z przypadkiem kiedy do każdego rekordu z pomiarem dodamy informacje o obserwatorze, który tego pomiaru dokonał.

Schemat relacji tabel stanowiących wspólną część dla wszystkich struktur przedstawia się następująco:



Rysunek 2. Schemat wspólnej części bazy dla wszystkich struktur

Na rysunku 2 widzimy tabele składające się z pól oraz relacje między nimi. PK oznacza wspomniany wcześniej Primary Key czyli Klucz Główny. Linia ciągłą oznaczone są relacje między tabelami. Wszystkie relacje zastosowane dla struktury wspólnej części baz są typu „1 – n” czyli relacje „jeden do wielu”. Typy relacji zostały już opisane we wstępie teoretycznym, ale teraz ten problem szczegółowo opiszę na kilku przykładowych rekordach z tabel Miejsce oraz Czujnik. Dla bardziej przejrzystego opisanie problemu przykładowe rekordy zawierają jedynie pola najistotniejsze dla rozważanego problemu.

Tabela 10. Przykładowe rekordy tabeli Miejsce

| ID_Miejsce | Nazwa    | Województwo | ... |
|------------|----------|-------------|-----|
| 121        | Ślemień  | Śląskie     | ... |
| 122        | Szczyrk  | Śląskie     | ... |
| 123        | Gilowice | Śląskie     | ... |

Pole ID\_Miejsce jest Kluczem Głównym tej tabeli. Jest to unikalny numer i jego wartość nie może się powtórzyć w kolejnych rekordach, dlatego na schemacie struktury jest to relacja „1”.

Tabela 11. Przykładowe rekordy tabeli Czujnik

| ID_Czujnik | ID_Miejsce | ID_Pomiar | Auto/Obserw |
|------------|------------|-----------|-------------|
| 1211       | 121        | 14        | 1           |
| 1212       | 122        | 14        | 1           |
| 1213       | 122        | 14        | 0           |
| 1214       | 122        | 13        | 1           |

W tabeli Czujnik pole ID\_Czujnik jest Kluczem Głównym, dlatego jego wartość nie może się powtórzyć. Można jednak zauważyć, że wartości w polu ID\_Miejsce się powtarzają. Powyższe rekordy należy interpretować następująco: Czujnik o indeksie 1212 znajduje się w miejscu 122 i mierzy wartość 14. Jest to pomiar tradycyjny o czym świadczy wartość 1 w polu Auto/Obserw. Natomiast czujnik o indeksie 1213 także znajduje się w miejscu 122 i także mierzy wartość 14. Jednak jest to pomiar prowadzony automatycznie. Widzimy, że wartości w polu ID\_Miejsce w tabeli Miejsce nie mogą się powtarzać. Natomiast wartości pola ID\_Miejsce w tabeli Czujnik może wystąpić wiele razy. Tak właśnie należy interpretować relacje „jeden do wielu”.

Przedstawione rozwiązanie w sposób jednoznaczny określa pomiar. Ujęte zostały wszystkie aspekty związane z opisaniem pomiaru, choć nie ma jeszcze możliwości przechowywania wyników pomiarów.

### 5.3. Baza metadanych pomiarów hydrometeorologicznych

Powyższa struktura nie zawiera jeszcze miejsca na pomiary. Można jednak zauważyć, że w jej skład wchodzi wszystkie wartości opisujące pomiar. W tej postaci jest ona strukturą bazy metadanych. Najprościej można to pojęcie zdefiniować, jako dane o danych. W bazach danych metadanymi są definicje tabel, widoków, kluczy, natomiast danymi są zawartości tych tabel, widoków, czyli rekordy.

Dla lepszego zdefiniowania pojęcia przedstawię przykładową kartę metadanych, która została przygotowana przez Kingę Hajduk z Politechniki Krakowskiej w pracy magisterskiej „Baza metadanych pomiarów hydrometeorologicznych Zakładu Hydrologii Politechniki Krakowskiej”

Tabela 12. Przykładowa karta - Pomiar opadu atmosferycznego deszczomierzem Hellmanna na posterunku Ślemień [Kinga Hajduk, 2006]

| Mierzony parametr                      | Opad atmosferyczny             |         |
|--|--------------------------------|---------|
| Przyrząd pomiarowy                     | Deszczomierz Hellmanna         |         |
| Rodzaj zapisu danych: samopis/dziennik | Dziennik                       |         |
| Format gromadzonych danych             | Miesięczny wykaz opadowy       |         |
| Cechy wykresu danych                   | Suma opadów za dobę            |         |
| Jednostka pomiarowa                    | [mm]                           |         |
| Dokładność pomiarowa                   | 1 mm                           |         |
| Skala pomiarowa                        | (zależna od rodzaju przyrządu) |         |
| Lokalizacja                            | Ślemień                        |         |
| Miejsce                                | -                              |         |
| Obserwator                             | Znamirowski                    |         |
| Okres pomiarowy                        | 01-1978                        | 02-1991 |
| Częstotliwość pomiaru                  | 1 raz dziennie                 |         |
| Stacja pomiarowa: czynna/nieczynna     | Brak danych                    |         |
| Uwagi                                  | Brak                           |         |

Wszystkie informacje zawarte w powyższej karcie można wprowadzić do zaprezentowanej struktury. W ten sposób, nie posiadając jeszcze informacji o sposobie uporządkowania danych, stworzona została baza metadanych.

Baza ta może mieć bardzo szerokie zastosowanie. Poniżej przedstawię kilka podstawowych informacji jakie można uzyskać z zaprezentowanej bazy metadanych:

- spis stacji pomiarowych, nie tylko obecnie działających, ale wszystkich nawet tych już nie prowadzących pomiarów,
- lista wszystkich urządzeń stosowanych do pomiarów,
- lista stacji, na których prowadzone są pomiary tradycyjne,
- lista osób pełniących rolę obserwatorów na przestrzeni lat,
- na podstawie lokalizacji można tworzyć mapy ze stacjami pomiarowymi.

Oczywiście powyższe propozycje są tylko przykładem jak szerokie może być zastosowanie utworzonej bazy metadanych.

Jednak celem mojej pracy jest zaprojektowanie struktur baz danych, przechowujących nie tylko metadane ale także wyniki pomiarów. Kolejnym krokiem będzie zatem rozbudowanie struktury bazy metadanych o tabele przechowujące wyniki pomiarów.

#### **5.4. Struktura 1 – rodzaj wykonywanego pomiaru**

W pierwszej strukturze, którą przedstawię, dane zostaną podzielone ze względu na rodzaj wykonywanego pomiaru, a następnie przyporządkowane do odpowiedniej tabeli. Przykładowo, jeżeli zmierzona wartość jest temperaturą, zostanie ona umieszczona w tabeli Temperatura. Zacznę od opisanie tabel i relacji w nich występujących aby lepiej przedstawić pomysł.



W tym rozwiązaniu należy stworzyć tyle tabel ile rodzajów pomiarów jest wykonywanych. Przedstawię dwie przykładowe tabele: Pomiary Temperatury i Pomiary Stanu Wody.

Tabela 13. Tabela Pomiary Temperatury

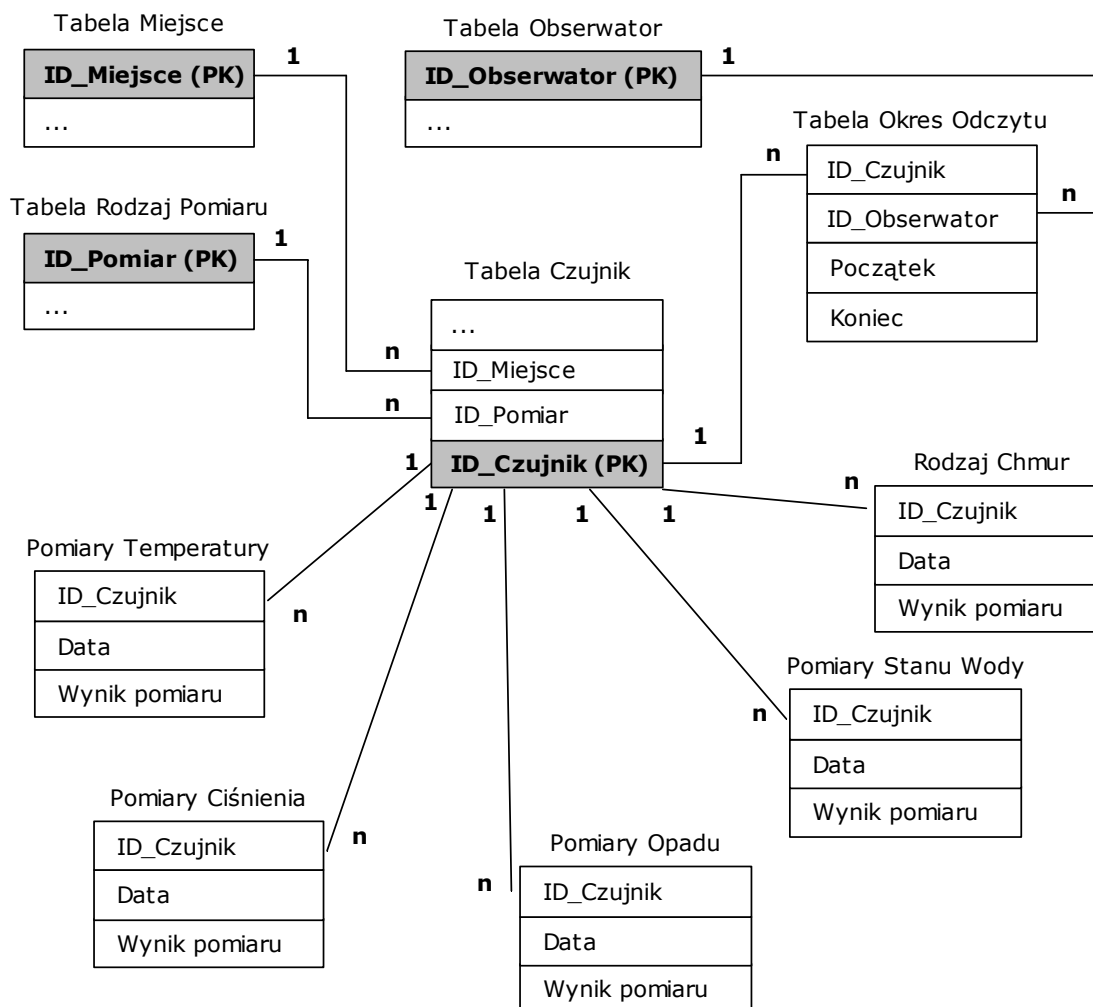
|            |  |
|------------|--|
| ID_Czujnik | Indeks czujnika, który wykonał pomiar lub z którego pomiar został odczytany przez obserwatora. Pole to jest połączone z tabelą czujniki.   |
| Czas       | Data i godzina wykonania odczytu   |
| Wynik      | Wynik pomiaru. Pole to w zależności od odczytywanej wielkości będzie posiadało odpowiednie typy. W tym przypadku - tabeli dla temperatury wyrażonej w [°C] zapisywanej do jednego miejsca po przecinku, typ tego pola będzie zmiennoprzecinkowy. |

Tabela 14. Tabela Pomiary Stanu Wody

|            |  |
|------------|--|
| ID_Czujnik | Indeks czujnika, który wykonał pomiar lub z którego pomiar został odczytany przez obserwatora. Pole to jest połączone z tabelą czujniki. |
| Czas       | Data i godzina wykonania odczytu   |
| Wynik      | Wynik pomiaru. W tym przypadku dla wartości pomiaru stanu wody wyrażonej w [cm], typ tego pola będzie całkowity.                         |

Obie tabele mają podobną budowę lecz różnią się w typie pola, jakie jest zastosowane dla pola Wynik. W zależności od rodzaju pomiaru jaki jest wykonywany, przyjmie ono odpowiedni typ.

Struktura zaproponowanego rozwiązania będzie wyglądać następująco:



Rysunek 3. Struktura 1 – baza z danymi podzielonymi ze względu na mierzoną wielkość

Zaprezentowana wcześniej struktura wspólnej części (Rys. 2) została rozbudowana o tabele, w których przechowywane będą wyniki pomiarów. Przedstawione rozwiązanie (Rys. 3) jest jedynie zarysem pomysłu na przechowywanie danych heterogenicznych. Aby móc wprowadzić wszystkie pomiary, należy dodać do tej struktury kolejne tabele dla każdego rodzaju pomiaru. Te przykładowe tabele jasno przedstawiają pomysł a struktura jest dzięki temu przejrzysta.

Tabele przechowujące wyniki pomiarów wyglądają bardzo podobnie lecz w rzeczywistości różnią się znacznie od siebie w najistotniejszym miejscu. Różnią się formatem pola Wynik pomiaru. Dla przykładu pole Wynik pomiaru z tabeli Rodzaj Chmur będzie miało format tekstowy, z tabeli Stan Wody będzie miało format całkowity a z tabeli Temperatura będzie miało format zmiennoprzecinkowy.

Poniżej przedstawię na przykładzie rozwiązanie, jakie zaimplementowane są w tej bazie. Opiszę proces odnajdywania wszystkich własności przykładowego pomiaru.

Tabela 15. Przykładowe rekordy tabeli Temperatura dla struktury 1

| ID_Czujnik | Czas                | Pomiar |
|------------|---------------------|--------|
| 111201     | 06/08/1980 12:00:00 | 21,5   |
| 110020     | 06/08/1980 12:00:00 | 19,3   |
| 111201     | 07/08/1980 12:00:00 | 18,4   |
| 110021     | 07/08/1980 12:00:00 | 12,6   |

Pomiary są w ten sposób zapisywane w bazie dla każdego czujnika. Jednak niewiele mówią one o miejscu czy sposobie wykonania pomiaru – a te informacje muszą być możliwe do odczytania z bazy.

W Tabeli 15 widzimy, że czujnik o danym indeksie wykonał pomiar o danej godzinie a jego wartość wynosiła x. Niestety informacje w tej formie nie mówią nam za wiele. Wiemy jednak, że wszystkie informacje są zawarte w naszej bazie i muszą one zostać w odpowiedni sposób z niej wydobyte. Widzimy, że pierwszy pomiar został wykonany przez czujnik o indeksie 111201. W tym momencie przeszukiwana jest tabela Czujnik gdzie znajdujemy rekord:

Tabela 16. Przykładowy rekord tabeli Czujnik

| ID_Czujnik | ID_Miejsce | ID_Pomiar | Auto/Obserw |
|------------|------------|-----------|-------------|
| 111201     | 212        | 16        | 1           |

Następnie w tabeli Miejsce szukamy rekordu 212, z którego uzyskujemy następujące informacje:

Tabela 17. Przykładowy rekord tabeli Miejsce

| ID_Miejsce | Nazwa   | Województwo | Szerokość   | Długość     | ... |
|------------|---------|-------------|-------------|-------------|-----|
| 212        | Szczyrk | Śląskie     | 49°43'16" N | 19°00'21" E | ... |

Uzyskaliśmy szczegółowe informacje o położeniu czujnika a na podstawie rekordu o indeksie 16 z tabeli Rodzaj Pomiaru wiemy, że wartość 21,5 jest temperaturą mierzoną w °C.

Ostatnie pole w tabeli Czujnik mówi nam, że czujnik ten nie wykonuje pomiarów automatycznych. Gdyby jednak znalazła się tam wartość 0 to dowiemy się, że pomiar wykonany został automatycznie i nie musimy wydobywać z bazy kolejnych informacji. Wartość 1 w tym polu wskazuje na to, że odczyt pomiaru został wykonany przez obserwatora. W związku z tym musimy uzyskać informację o tym obserwatorem. Wiemy, że pomiaru dokonał czujnik 111202 o godzinie 06/08/1980 12:00:00. Przeszukiwana jest tabela Okres Odczytu. Gdy odnaleziony jest rekord z indeksem czujnika 111202 sprawdzane jest, czy data pomiaru zawiera się między datami podanymi w polach Początek i Koniec. Znajdujemy następujący rekord:

Tabela 18. Przykładowy rekord tabeli Okres Odczytu

| ID_Czujnik | Początek            | Koniec              | ID_Osoby |
|------------|---------------------|---------------------|----------|
| 111202     | 01/07/1980 06:00:00 | 31/08/1980 00:00:00 | 2343     |

Taki rekord zawiera informacje, że odczytu z czujnika 11202 06/06/1980 dokonała osoba o numerze identyfikacyjnym 2343 i na tej podstawie w tabeli Obserwator łatwo znaleźć wszystkie informacje, na temat danego obserwatora które są w bazie przechowywane.

Zaproponowana struktura jest najbardziej podstawowym rozwiązaniem. Baza ta przechowuje każdy typ pomiaru w osobnej

tabeli. Łatwo będzie w ten sposób poradzić sobie z heterogenicznym charakterem danych. Jeżeli nawet jakikolwiek pomiar był w historii przedstawiany w innych jednostkach niż obecnie, lub jeżeli przedstawiany był za pomocą tekstu, wystarczy stworzyć dla tego pomiaru nową tabelę.

Podsumowując zaprezentowaną strukturę zestawię jej mocne i słabe strony, które widoczne są już na tym etapie projektowania.

Plusy:

- przejrzysta struktura – dla każdego rodzaju pomiaru tworzona jest osobna tabela,
- łatwa implementacja – aby przygotować odpowiednią ilość tabel wystarczy poznać wszystkie typy pomiarów jakie mają się znaleźć w bazie.
- jest to dobre rozwiązanie do analizy danych jednego typu wykonywanego pomiaru

Minusy:

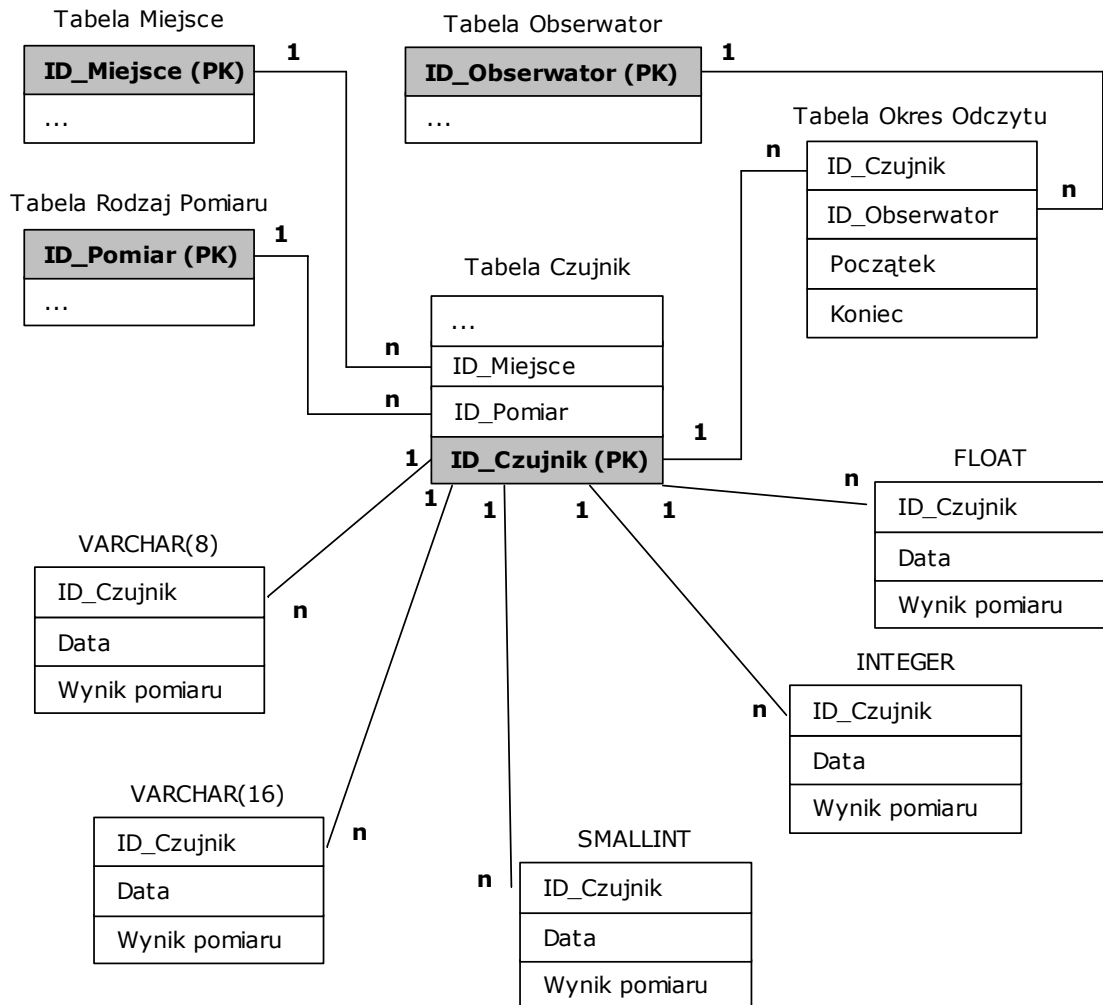
- gdy wprowadzamy nowy pomiar zachodzi potrzeba stworzenia nowej tabeli – jeżeli w przyszłości do standardów pomiarów hydrometeorologicznych zostanie wprowadzona nowa mierzona wielkość, w celu zapisania jej do stworzonej struktury, zachodzi konieczność stworzenia nowej tabeli dla tej wielkości.
- jeżeli analizujemy dane dla określonych chwil czasowych przeszukanych musi zostać wiele tabel.

## **5.5. Struktura 2 – wartość pomiaru**

Kolejna struktura, którą zaprojektowałem, przyporządkowuje dane do odpowiednich tabel ze względu na format danej. Pomysł polega na tym, aby zmierzona wartość wpisana została do tabeli o odpowiednim formacie pola Wynik pomiaru. Rozwiązanie to oparte

jest na typach danych stosowanych w MySQL i przedstawionych w rozdziale 3.3.

Aby zastosować takie rozwiązanie, należy stworzyć odpowiednią liczbę tabel, różniących się między sobą formatem pola Wynik pomiaru. Struktura tego rozwiązania przedstawia się następująco:



Rysunek 4. Struktura 2 – baza z danymi podzielonymi ze względu na wartość pomiaru

Tabela 19. Przykładowa tabela przechowująca pomiary dla struktury 2

|               |  |
|---------------|--|
| ID_Czujnik    | Indeks czujnika, który przeprowadził pomiar  |
| Data          | Data wykonania pomiaru   |
| Wynik pomiaru | Pole, w które wpisywane będą zmierzone wielkości. W zależności od wielkości pomiaru jaki będzie wprowadzany, pole to będzie różnić się formatem. |

Pole Wynik pomiaru będzie miało taki format, jak jest to zaznaczone w nazwie poszczególnych tabel. Rozwiązania zastosowane w strukturze 2 przedstawię na przykładowych scenariuszach wprowadzania danych do bazy.

Scenariusz pierwszy przewiduje wykonanie pomiaru temperatury. Zmierzona wartość wynosi 3,1 °C. Oczywiście pomiar ten jest przypisany do pewnego czujnika. Za pomocą relacji wiemy jaki pomiar jest wykonany i z jaką dokładnością. W tym przypadku dokładność wynosi 0,1. System dzieli wartość temperatury przez dokładność i otrzymuje wynik 31. Następnie badana jest ta wartość. Należy ona do przedziału od 0 do 256, więc może zostać wpisana do tabeli TINYINT, który takie właśnie wartości przyjmuje. Jeżeli chcemy wartość tą odczytać z bazy, mnożymy ją przez dokładność i otrzymujemy z powrotem wynik 3,1.

Scenariusz drugi to ten sam czujnik mierzący temperaturę, jednak w okresie cieplejszym. Wynik pomiaru to 31,3. Dzieląc tak jak przedtem tą wartość przez dokładność otrzymujemy 313, co zostanie automatycznie przypisane do tabeli SMALLINT.

Zastosowanie takiego rozwiązania wpływa na wielkość przestrzeni dyskowej potrzebnej do zapisania wszystkich pomiarów. Wartości pomiarów są katalogowane ze względu na format jaki reprezentują. Jest to bardzo dobre rozwiązanie jeżeli chodzi o oszczędność miejsca na dysku. Niestety wydłuża proces zapisu do bazy i odczytu z bazy wszystkich wartości gdyż np. dla pomiaru

temperatury przeszukane muszą zostać dwie tabele. Jeżeli szukamy zmierzonej wartości z drugiego scenariusza przeszukana musi zostać tabela TINYINT a dopiero następnie tabela SMALLINT, w której znajduje się szukana wartość. Dodam, że zastosowanie takiego rozwiązania wpłynie na niewielką ilość tabel w bazie, ale zarazem na ogromne ilości rekordów w tych tabelach.

Chciałbym także w tym miejscu uzasadnić zastosowanie dwóch pól w tabeli Pomiary czyli Max wartość i Min wartość. Dla przykładu rozpatrzę wielkość jaką jest kierunek wiatru. Wartości tego pomiaru zawierają się w przedziale od  $0^{\circ}$  do  $360^{\circ}$ . Dokładność jest równa  $5^{\circ}$ , czyli wielkość ta potrzebuje jedynie 72 liczb aby można była ją zapisać. Dlatego wszystkie wartości pomiarukierunku wiatru możemy zapisać do tabeli TINYINT. Jeżeli chcemy odczytać szukaną wartość, system najpierw sprawdzi jaki zakres jest przeznaczony dla tego pomiaru. Widząc, że dla zapisu tej wartości potrzeba jedynie 72 liczb, jedyną tabelą jaką będzie przeszukiwał w poszukiwaniu pomiaru będzie tabela TINYINT.

Takie samo rozwiązanie można zastosować do pomiarów, których wyniki są przechowywane w postaci tekstu. Po zbadaniu długości wprowadzonego tekstu wartość byłaby wprowadzana albo do tabeli VARCHAR(8) dla tekstu od 1 do 8 znaków albo do tabeli VARCHAR(16) dla tekstu od 8 do 16 znaków. Oczywiście podane wartości są przykładowe, gdyż nowoczesne systemy bazodanowe pozwalają nam dowolnie dobierać ograniczenia długości znaku.

Zaprezentowana struktura wymaga naturalnie rozpoznania wszystkich danych, jakie mają zostać wprowadzone do bazy, aby można było przygotować odpowiednią ilość tabel. Poniżej przedstawię zestawienie mocnych i słabych stron struktury 2:



**Plusy:**

- mała ilość tabel – przejrzysta struktura bazy,
- oszczędność miejsca na dysku – każdy pomiar jest odpowiednio przygotowany a następnie przypisany do właściwej tabeli, aby jak najbardziej zmniejszyć rozmiar bazy.

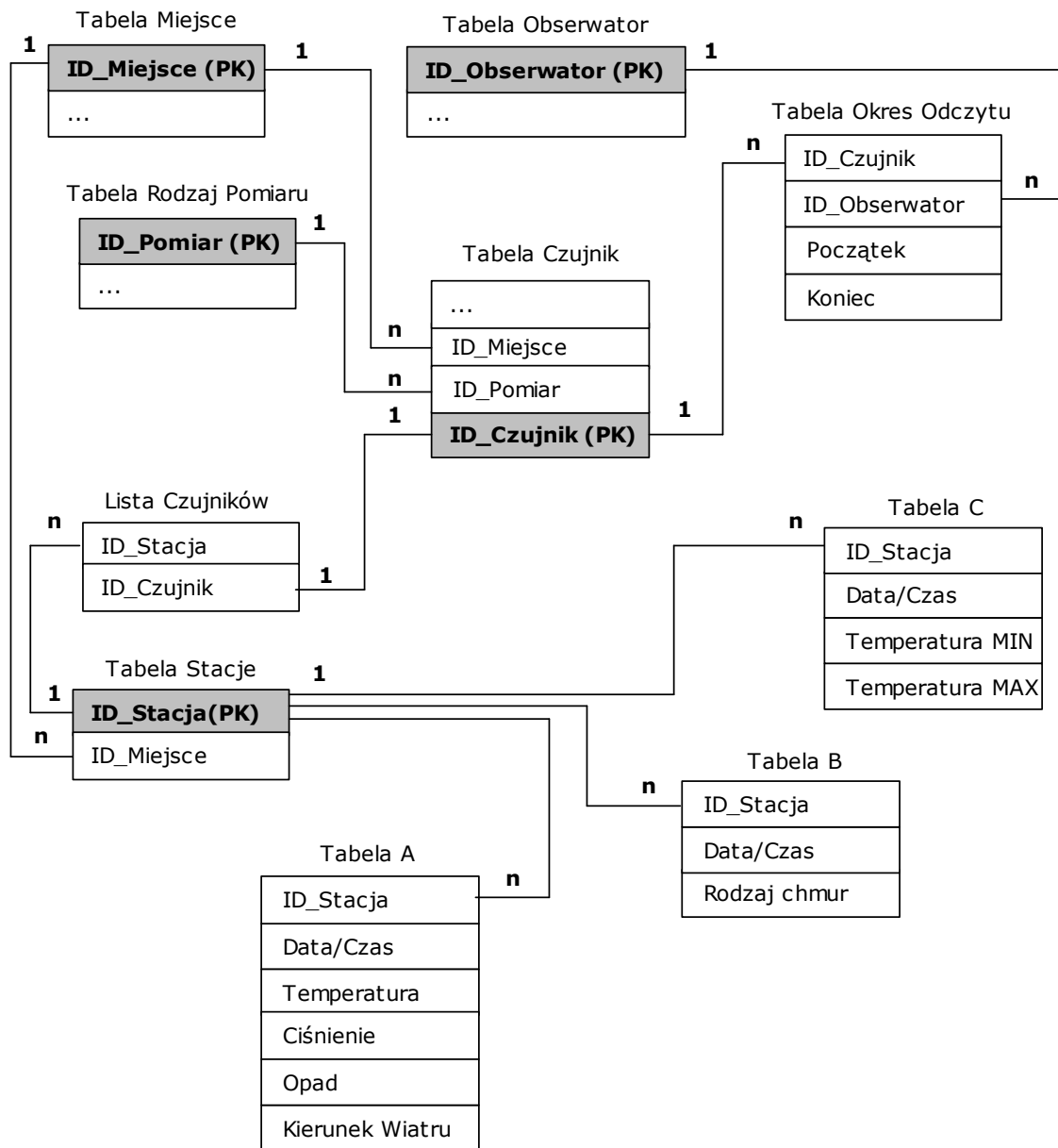
**Minusy:**

- wydłużony czas zapisu do bazy – każda wartość wprowadzona do bazy musi zostać odpowiednio przygotowana a następnie zbadana do jakiej tabeli ma zostać przyporządkowana,
- wydłużony czas odczytu z bazy – pomiary z jednego czujnika mogą znajdować się w różnych tabelach i w celu ich odczytania należy przeszukać wszystkie tabele, w których może potencjalnie znaleźć się pomiar,
- duża ilość rekordów w tabelach także wpłynie negatywnie na szybkość wykonywania zapytań.

**5.6. Struktura 3 – podobieństwa pomiarów**

Ostatnia prezentowana struktura kataloguje dane w bazie ze względu na podobieństwa wykonywania pomiarów przez różne stacje obserwacyjne. Pomysł polega na analizie wszystkich danych archiwalnych, które mają znaleźć się w naszej bazie oraz zapoznanie się z aktualnie panującymi standardami pomiarów. Na podstawie dokonanej analizy wyszukane zostaną podobieństwa między stacjami pomiarowymi (np. dwie różne stacje wykonują takie same rodzaje pomiarów i o tej samej porze). Dla takich podobnych stacji zostanie utworzona odpowiednia tabela.

Schemat struktury 3 wygląda następująco:



Rysunek 5. Struktura 3 – baza z danymi podzielonymi ze względu na podobieństwa

Do zaprezentowanej wcześniej wspólnej części struktury dodano tabelę Stacje oraz tabelę Lista Czujników, które połączone ze sobą będą zdefiniowane jako zbiór wszystkich czujników w danym miejscu.

Tabela 20. Tabela Stacje

|            |   |
|------------|---|
| ID_Stacja  | Indeks stacji - unikalny numer identyfikacyjny dla danej stacji |
| ID_Miejsce | Indeks miejsca, w którym dana stacja się znajduje               |

Tabela 21. Tabela Lista czujników

|            |                 |
|------------|-----------------|
| ID_Stacji  | Indeks stacji   |
| ID_Czujnik | Indeks czujnika |

W ten sposób stworzona została lista czujników, jakie znajdują się na danej stacji. Zostanie to wykorzystane w celu katalogowania danych.

Na Rys. 5 warto zwrócić uwagę na relację panującą między tabelą Czujnik a tabelą Lista Czujników. Jest to relacja typu „jeden do jednego”. Oznacza to, że w obydwu tabelach wartość pola ID\_Czujnik jest unikalna i nie może się powtarzać.

Wyniki pomiarów będą przechowywane w stworzonych tabelach. Na schemacie widoczne są przykładowe tabele jakie wprowadziłem do tej struktury. Cel i sens zastosowania takiego rozwiązania zaprezentuję na przykładowym scenariuszu znajdującym zastosowanie w tej strukturze.

Założmy, że stacja 1 mierzy następujące wielkości: temperaturę, ciśnienie, opad, kierunek wiatru, rodzaj chmur. Stacja 2 wykonuje podobnie pomiary temperatury, ciśnienia, opadu, kierunku wiatru ale nie jest na niej rozpatrywany rodzaj chmur.

Pomysł polega na tym, aby grupować ze sobą podobne stacje i stworzyć dla nich odpowiednie tabele. Na powyższym przykładzie widać, że stacje 1 i 2 wykonują cztery takie same pomiary. Dla tych pomiarów stworzona zostanie odpowiednia tabela, której struktura wygląda następująco:

Tabela 22. Przykładowa tabela z danymi po zgrupowaniu stacji 1 i 2

|                 |                         |
|-----------------|-------------------------|
| ID_Stacji       | Indeks stacji           |
| Data            | Data wykonania pomiaru  |
| Temperatura     | Wartość temperatury     |
| Ciśnienie       | Wartość ciśnienia       |
| Opad            | Wartość opadu           |
| Kierunek wiatru | Wartość kierunku wiatru |

Oczywiście każde pole będzie miało odpowiedni format dla zapisywanej danej.

Tabela 23. Przykładowe rekordy tabeli z pomiarami

| ID_Stacji | Data                 | Temperatura | Ciśnienie | Opad | Kierunek wiatru |
|-----------|----------------------|-------------|-----------|------|-----------------|
| 1         | 15/01/80<br>12:00:00 | 5,4         | 1023      | 0    | 235             |
| 2         | 15/01/80<br>18:00:00 | 1,2         | 1012      | 0,1  | 180             |
| 1         | 16/01/80<br>12:00:00 | 6,7         | 1020      | 0    | 205             |
| 2         | 16/01/80<br>18:00:00 | 0,5         | 1020      | 3,4  | 175             |

Co zatem z rodzajem chmur, które także są obserwowane w stacji 1? Zostanie dla ich utworzona osobna tabela. Załóżmy jednak, że istnieje stacja 3 na której obserwowany jest tylko rodzaj chmur.

Tabela 24. Przykładowa tabela przechowująca pomiary rodzaju chmur dla struktury 3

|              |                        |
|--------------|------------------------|
| ID_Stacji    | Indeks stacji          |
| Data         | Data wykonania pomiaru |
| Rodzaj chmur | Wynik pomiaru          |

Tabela 25. Przykładowe rekordy tabeli przechowującej pomiary rodzaju chmur dla struktury 3

| ID_Stacji | Data              | Rodzaj chmur |
|-----------|-------------------|--------------|
| 1         | 15/01/80 12:00:00 | X            |
| 3         | 15/01/80 18:00:00 | X            |
| 3         | 16/01/80 06:00:00 | X            |
| 1         | 16/01/80 12:00:00 | X            |

Zastosowanie takiego rozwiązania może być niezwykle korzystne w przypadku pomiarów automatycznych. Dzięki przyjętym standardom wykonywania pomiarów hydrometeorologicznych duża ilość stacji pomiarowych wykonuje takie same pomiary regularnie w tych samych godzinach.

Aby lepiej wyszukiwać podobieństwa należy wprowadzić pewne rozbudowane rozwiązanie mówiące, które pomiary z jaką częstotliwością są wykonywane. Dzięki temu będziemy je w stanie lepiej dobierać według podobieństwa. Należy lekko zmodyfikować tabelę Czujnik. Dodamy do niej pole informujące, ile pomiarów danego parametru jest wykonywanych w ciągu 24 godzin.

Tabela 26. Zmodyfikowana tabela Czujnik dla schematu 3

|              |  |
|--------------|--|
| ID_Czujnik   | ...  |
| ID_Miejsce   | ...  |
| ID_Pomiar    | ...  |
| Auto/Obserw  | ...  |
| Rodzaj       | ...  |
| Ile pomiarów | Pole informujące o tym, ile pomiarów wykonuje dany czujnik w ciągu 24 godzin |
| ...          | ...  |

Oprócz zmodyfikowania tabeli Czujnik musimy także posiadać informacje, o których godzinach dany pomiar jest wykonywany, aby móc ze sobą zestawić pomiary, które wykonywane są z tą samą częstotliwością i o tych samych godzinach.

Tabela 27. Tabela Kiedy Pomiar zastosowana dla struktury 3

|            |   |
|------------|---|
| ID_Czujnik | Indeks czujnika, który wykonuje pomiary |
| Godzina    | Godzina wykonywania pomiaru             |

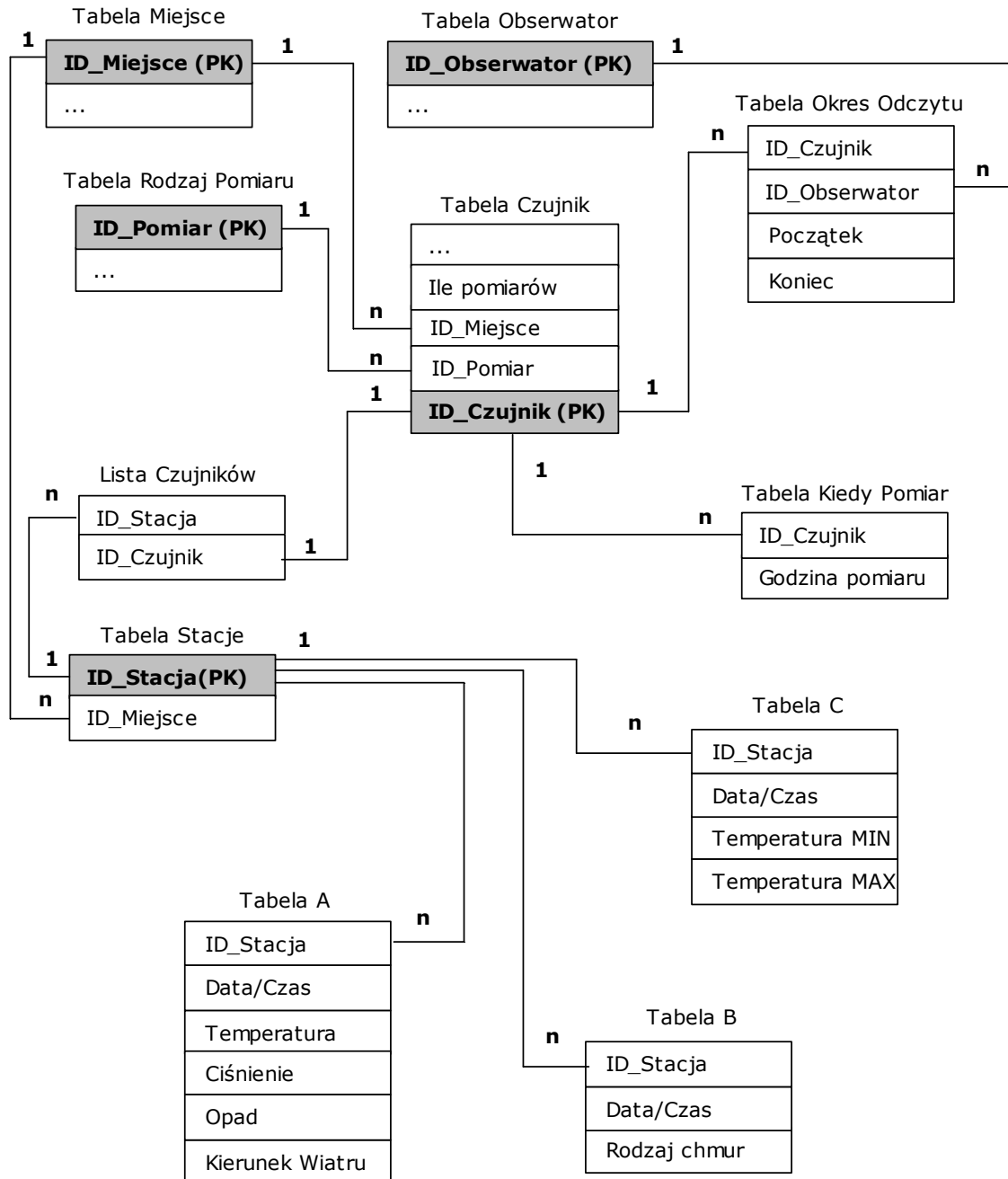
Oczywiście jest to tylko przykładowe rozwiązanie. Innym pomysłem jest umieszczanie w tabeli tylko godziny jednego pomiaru. Jeżeli pomiar temperatury wykonywany jest 4 razy w ciągu doby, a pierwszy pomiar jest wykonywany o 00:00, to z posiadanych informacji można wnioskować, że kolejne pomiary odbywają się o 06:00, 12:00 i 18:00. W ten sposób oszczędzana jest przestrzeń dyskowa ale wpływa to negatywnie na szybkość zapisu i przejrzystość struktury. Rozwiązanie, które zaproponowałem wydaje się być bardziej uniwersalne, ponieważ można w nim uwzględnić nawet pomiary nieregularne.

Tabela 28. Przykładowe rekordy tabeli Kiedy Pomiar dla struktury 3

| ID_Czujnik | Godzina |
|------------|---------|
| 112002     | 00:00   |
| 112002     | 06:00   |
| 112002     | 12:00   |
| 112002     | 18:00   |

Zastosowane rozwiązanie pozwoli jednoznacznie określić podobieństwa między czujnikami. Zbadane zostanie jakie czujniki znajdujące się w danym miejscu wykonują pomiary w tym samym czasie. Jeżeli takie czujniki się znajdują, zostaną połączone ze sobą przez przyporządkowanie ich do jednej stacji.

Schemat zmodyfikowanej struktury 3 przedstawia się następująco:



Rysunek 6. Schemat zmodyfikowanej struktury 3

Już na etapie projektowania możliwe jest krótkie zestawienie mocnych i słabych punktów prezentowanej struktury:

#### Plusy

- w zależności od poziomu stosowania standardów pomiarów struktura może zyskać na wartości,
- dane wprowadzane są do bazy seriami, co wpływa pozytywnie na czas zapisu.

#### Minusy:

- musimy znać wszystkie dane, jakie chcemy wprowadzić do bazy, przed jej implementacją,
- problem jeżeli pomiary wykonywane są o różnych porach, np. na stacji mierzone są 4 parametry o różnych godzinach, więc dla każdego parametru musi zostać stworzona osobna tabela,
- problem jeżeli jakiś pomiar nie zostanie wykonany – jeżeli z jakiś przyczyn, np. awarii jednego z czujników, pomiar nie zostanie wykonany, to jego pole pozostanie puste a więc następuje strata przestrzeni dyskowej.

Zanim przejdę do implementacji struktur oraz do ich szczegółowej analizy, chciałbym omówić niezwykle istotny problem, jakim jest sposób przechowywania daty i czasu wykonania pomiaru w zaprezentowanych strukturach.

### **5.7. Problem przechowywania terminu wykonania pomiaru**

W zaproponowanych strukturach 1 i 2 data i czas wykonania pomiaru są bezpośrednio przypisane do wykonanego pomiaru. Wartość ta ma więc duży wpływ na wielkość bazy, co jest jednym z najistotniejszych czynników przy jej ocenie.



Sposoby rozwiązania tego problemu, które zostaną zaproponowane, są jedynie przykładem podejścia do problemu przechowywania czasu prowadzenia pomiarów. Nie były one przedstawione przy okazji prezentacji struktur aby rozwiązania były bardziej czytelne. Rozwiązania te nie zostaną zastosowane w implementowanych bazach, gdyż nie ten problem jest najistotniejszy w mojej pracy. W każdej strukturze format czasu będzie następujący: RRRR-MM-DD GG:MM:SS (w MySQL taki typ danej to DATETIME). RRRR oznacza rok wykonania pomiaru (np. 2009), MM to miesiąc (np. 03), a DD to dzień (np. 30). Podobnie zapisywany jest czas – GG to godzina, MM to minuty a SS oznaczają sekundy (np. 11:05:20)

Dlaczego ten problem jest więc taki istotny. Jeżeli w bazie znajduje się milion rekordów z pomiarami to oznacza to, że milion razy musimy wpisać informację o terminie wykonania pomiaru. Jeśli natomiast mamy 100 stacji w systemie, każda wykonuje 10 pomiarów w tym samym czasie to oznacza to, że po jednej serii pomiarów informacja o godzinie wykonania tych pomiarów (dla każdego pomiaru jest to ta sama godzina) znajdzie się w bazie 1000 razy. Dlatego rozbudowanie rozwiązania pozwalające na bardziej logiczne podejście do problemu jest jak najbardziej na miejscu. Warto także porównać wyniki przedstawione w kolejnym rozdziale zawarte w tabelach 35 oraz 38.

Pierwszym sposobem który pozwoli na zmniejszenie wielkości bazy jest rozdzielenie daty i czasu. Pole typu DATETIME zajmujące 8 bajtów zostanie podzielone na pole typu DATE (zawierające informacje o dacie) o wielkości 3 bajtów oraz pole typu TIME (zawierające informacje o czasie) także o wielkości 3 bajtów. Rozwiązanie takie pozwoli na zmniejszenie miejsca potrzebnego na

przechowywanie informacji o czasie wykonania pomiaru o 25 procent, lecz wpłynie negatywnie na czas wykonywania zapytań.

Kolejnym pomysłem na rozwiązanie tego problemu jest zastosowanie innego typu pola dla rozpatrywanych danych. Typ ten zmieniłby się z DATETIME (zapisywanego w formacie RRRR-MM-DD GG:MM:SS) do typu TIMESTAMP (także zapisywanego w formacie RRRR-MM-DD GG:MM:SS). Różnica między tymi typami to zakres dat z jakiego możemy skorzystać. Dla pola typu DATETIME zakres dat z jakiego możemy skorzystać to od 1000-01-01 00:00:00 do 9999-12-31 23:59:59, natomiast pole typu TIMESTAMP może zawierać jedynie daty z zakresu od 1970-01-01 00:00:01 do 2038-01-09 03:14:07. Jeżeli więc nasze pomiary mieszczą się w tych granicach warto zastosować pole TIMESTAMP. W porównaniu do pola typu DATETIME (8 bajtów) TIMESTAMP zajmuje jedynie 4 bajty, co czyni to rozwiązanie bardziej oszczędnym, a taka zmiana typu pola nie wpłynie na czas wykonywania zapytań. [MySQL.com, 2009]

Kolejne rozwiązanie, jakie zaproponuję, oparte jest na regularności z jaką wykonywane są pomiary automatyczne. Pomysł polega na tym, aby dla danego pomiaru umieścić informacje jedynie o dacie wykonania pomiaru, a czas wykonania tego pomiaru byłby zapisany za pomocą odpowiedniego identyfikatora.

Jeżeli dany czujnik wykonuje regularne pomiary np. co 10 minut w ciągu dnia, to przez całą dobę wykona tych pomiarów 144. Pole typu TINYINT będzie zatem wystarczające by opisać wszystkie czasy wykonania pomiaru. Aby zastosować takie rozwiązanie należy do Tabeli Czujnik dodać pola informacyjne o czasie wykonania pierwszego pomiaru oraz o czasie, jaki występuje między dwoma kolejnymi pomiarami. Na tej podstawie zostanie utworzona tabela z godzinami pomiarów oraz o odpowiadającym im identyfikatorze czasu.

Tabela 29. Przykładowe rekordy tabeli Czas wykonania pomiaru

|          |     |
|----------|-----|
| 00:00:00 | 0   |
| 00:10:00 | 1   |
| ...      | ... |
| 11:50:00 | 71  |
| 12:00:00 | 72  |
| 12:10:00 | 73  |

Dla struktury 1 data i czas przechowywane są w polu typu DATETIME które zajmuje 8 bajtów przestrzeni dyskowej. Gdy zostanie zaimplementowane zaproponowane rozwiązanie data i czas będą przechowywane w polu DATE które zajmuje 3 bajty oraz w polu ID\_Czasu, które jest typu TINYINT i zajmuje 1 bajt. Rozwiązanie to zmniejsza więc ilość miejsca którą musi poświęcić na przechowywanie daty i czasu o połowę.

Niestety takie rozwiązanie jest trudne do zaimplementowania w przypadku pomiarów wykonywanych bardzo nieregularnie. Także czas zapisu jak i odczytu informacji z bazy jest zwiększony. Warto jednak takie rozwiązanie zastosować dla pomiarów wykonywanych regularnie.

W tabelach 30 – 33 przedstawię porównanie przykładowego rekordu ze struktury 1 dla tabeli Temperatura dla wszystkich trzech zaproponowanych ulepszeń.

Tabela 30. Przykładowy rekord tabeli Temperatura dla struktury 1

| ID_Czujnik | Data i czas (DATETIME) | Pomiar |
|------------|------------------------|--------|
| 101        | 2008-01-01 12:30:00    | -4.1   |
|            | 8 bajtów               |        |

Tabela 31. Przykładowy rekord tabeli Temperatura dla struktury 1 z zastosowaniem pól DATE oraz TIME

| ID_Czujnik | Data (DATE) | Czas (TIME) | Pomiar |
|------------|-------------|-------------|--------|
| 101        | 2008-01-01  | 12:30:00    | -4.1   |
|            | 3 bajty     | 3 bajty     |        |

Tabela 32. Przykładowy rekord tabeli Temperatura dla struktury 1 z zastosowaniem pola TIMESTAMP

| ID_Czujnik | Data i czas (TIMESTAMP) | Pomiar |
|------------|-------------------------|--------|
| 101        | 2008-01-01 12:30:00     | -4.1   |
|            | 4 bajty                 |        |

Tabela 33. Przykładowy rekord tabeli Temperatura dla struktury 1 z rozwiązaniem opartym na identyfikatorach czasu

| ID_Czujnik | Data (DATE) | ID_Czasu (TINYINT) | Pomiar |
|------------|-------------|--------------------|--------|
| 101        | 2008-01-01  | 75                 | -4.1   |
|            | 3 bajty     | 1 bajt             |        |

Zaprezentowane rozwiązania są jedynie przykładowym pomysłem podejścia do problemu przechowywania daty i czasu w bazie. W zależności od danych jakie będą wprowadzane do bazy, dokładność z jaką musi być określony czas pomiaru, oraz czy w projektowanej bazie priorytetem ma być pojemność bazy czy szybkość wykonywania zapytań, można zastosować różne kombinacje powyższych rozwiązań.

W kolejnym rozdziale przeanalizuję przedstawione rozwiązania na podstawie zaimplementowanych struktur oraz przykładowych zestawów danych. Postaram się także ocenić ich funkcjonalność, szybkość wymiany danych oraz wielkość bazy.

## **6. ANALIZA I OCENA PROPONOWANYCH SCHEMATÓW**

Kolejnym etapem tworzenia baz danych jest jej implementacja, testowanie oraz ewentualne wprowadzanie poprawek. Dla przygotowania analizy przeprowadzono symulację danych. Przyjęte zostało założenie, że analiza zostanie przeprowadzona na przykładowym zestawie danych pochodzących z regularnych pomiarów hydrometeorologicznych.

Symulacje zostały przeprowadzone na następującej platformie: Procesor Intel Core 2 Duo 1.83 GHz, 2 GB RAM, 80 GB FUJITSU (mhv2080bh) na systemie operacyjnym Windows XP Home z zainstalowanym Service Pack 3. Baza została stworzona w MySQL 5.1.31. Dodatkowo użyte zostały następujące narzędzia: MySQL Administrator 1.2.16 dla obliczenia przybliżonej wielkości bazy oraz MySQL Query Browser 1.2.16 dla symulacji zapytań oraz obliczenia czasu ich wykonywania.

Do zaimplementowanych struktur baz zostały automatycznie zaimportowane dane z plików CSV, wygenerowane przez prosty skrypt napisany w języku C. Zasymulowane wyniki pomiarów spełniają pewne z góry przyjęte założenia – temperatura nie może się zmieniać więcej niż 2 stopnie między kolejnymi dwoma pomiarami, maksimum osiąga w miesiącach letnich a minimum w zimowych itp.

Badaniom poddano małą sieć obserwacyjną składającą się z 10 stacji oraz dużą sieć składającą się z 50 stacji. Każda z nich wykonuje regularnie co 10 minut 8 pomiarów – 3 rodzaje temperatury, ciśnienie atmosferyczne, wilgotność powietrza, wysokość opadu, kierunek i prędkość wiatru. Dla obu sieci pomiarowych zasymulowano pomiary na przestrzeni jednego miesiąca, jednego roku oraz pięciu lat.

Na podstawie wygenerowanych pomiarów, zaimplementowano odpowiednie rozwiązania we wszystkich strukturach.

Tabela 34. Liczba wykonywanych pomiarów przez czujniki na przestrzeni  
zadanych okresów symulacji

|                           | 1 miesiąc | 1 rok      | 5 lat       |
|---------------------------|-----------|------------|-------------|
| 8 czujników (1 stacja)    | 35 712    | 420 480    | 2 102 400   |
| 80 czujników (10 stacji)  | 357 120   | 4 204 800  | 21 024 000  |
| 400 czujników (50 stacji) | 1 785 600 | 21 024 000 | 105 120 000 |

W bazie dla struktury 1 zostało utworzonych 8 tabel – osobno dla każdego pomiaru. Naturalnie dla każdej tabeli pole przechowujące wyniki pomiaru różniło się ze względu na typ - dla pomiarów prędkości czy kierunku wiatru zastosowano pole typu całkowitego, dla pomiarów temperatur czy ciśnienia zastosowano pole typu zmiennoprzecinkowego.

Dla bazy opracowanej na podstawie struktury 2 zostały utworzone tylko dwie tabele przechowujące wyniki pomiarów. Po analizie danych pomiarowych, które miały zostać wprowadzone do bazy okazało się, że aby przechować wyniki wszystkich 8 mierzonych wielkości wystarczy, aby w bazie zaimplementowane były tabele typu TINYINT oraz SMALLINT.

Ze względu na to, że pomiary jakie zostaną przeanalizowane pochodzą z sieci czujników automatycznych wykonujących regularne pomiary, baza oparta na strukturze 3 zyskała jedynie jedną tabelę zdolną do przechowywania wszystkich pomiarów.

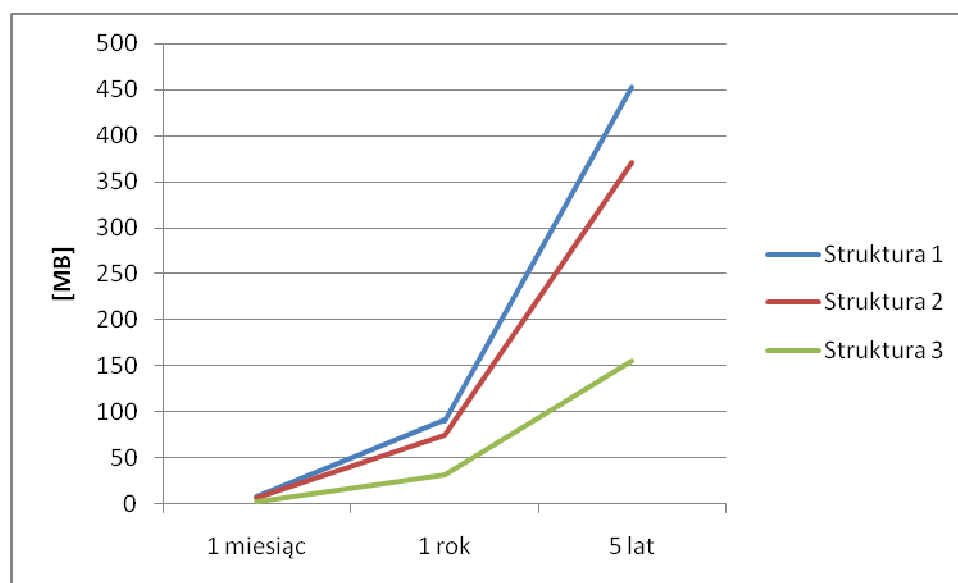
### **6.1. Analiza wielkości wypełnionych baz**

Jednym z dwóch najistotniejszych czynników wpływających na ocenę bazy danych jest jej rozmiar. Do każdej z zaimplementowanych struktur wprowadziłem ten sam zestaw zasymulowanych wyników pomiarów, aby określić, które rozwiązanie najlepiej wpływa na oszczędność przestrzeni dyskowej.

Dla lepszego uwidocznienia różnic w objętości podane zostały tylko wielkości tabel przechowujących wyniki pomiarów. Części wspólne zawierające szczegółowe informacje o lokalizacji czy rodzaju czujnika wykonującego pomiar zostały pominięte. W tabeli 35. przedstawiono zestawienie wielkości bazy dla poszczególnych struktur dla sieci obserwacyjnej składającej się z 10 stacji – każdej wykonującej 8 typów pomiarów.

Tabela 35. Objętość wypełnionych baz danych dla danych pochodzących z 10 stacji [MB]

|             | 1 miesiąc | 1rok | 5lat  |
|-------------|-----------|------|-------|
| Struktura 1 | 7,7       | 90,7 | 453,6 |
| Struktura 2 | 6,4       | 74,1 | 370,4 |
| Struktura 3 | 2,6       | 31,1 | 155,4 |



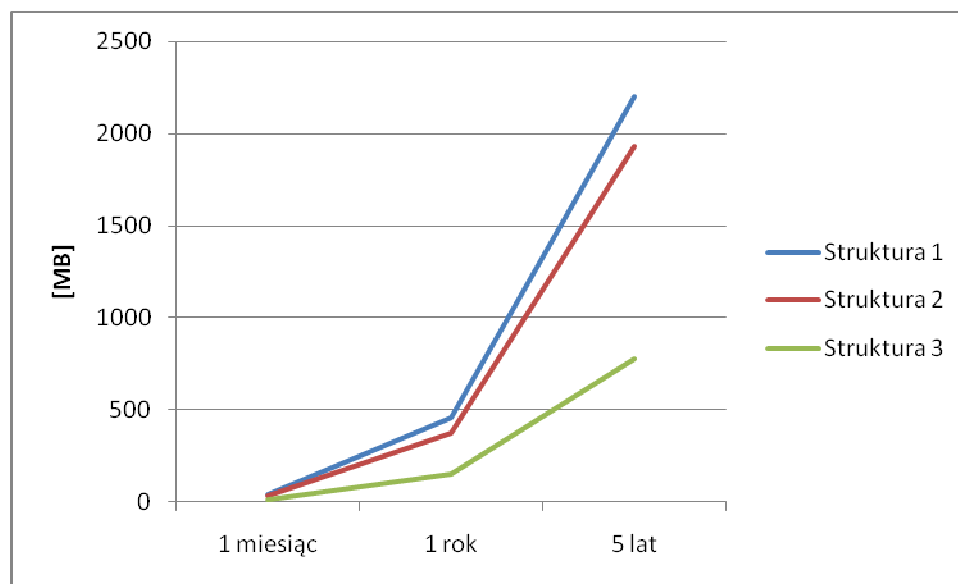
Rysunek 7. Wielkości baz dla pomiarów pochodzących z 10 stacji

W tabeli 36. przedstawię zestawienie dla rozbudowanej sieci obserwacyjnej w skład której wchodzi 50 stacji pomiarowych:



Tabela 36. Objętość wypełnionych baz danych dla danych pochodzących z 50 stacji [MB]

|             | 1 miesiąc | 1rok  | 5lat    |
|-------------|-----------|-------|---------|
| Struktura 1 | 38,5      | 453,6 | 2 200,0 |
| Struktura 2 | 31,8      | 370,4 | 1 930,0 |
| Struktura 3 | 13,2      | 155,4 | 776,9   |



Rysunek 8. Wielkości baz dla pomiarów pochodzących z 50 stacji

Na powyższych zestawieniach wyraźnie widoczna jest przewaga struktury 3 nad pozostałymi. Taki rezultat jest wynikiem różnicy w sposobie przechowywania danych. Struktura 3 przechowuje dane seriami.

Dla jednej stacji wykonującej 8 pomiarów w tym samym czasie, informacja o numerze stacji, z której czujnik wykonał pomiar, jak i o czasie wykonania pomiaru, znajduje się tylko raz. Dla struktur 1 oraz 2 dla tych samych 8 pomiarów informacja o czujnikach, które wykonały pomiary oraz o czasie wykonania pomiaru znajdują się 8 razy – osobno dla każdego czujnika.

Budowa tabeli przechowującej wyniki pomiarów dla struktury 3 prezentuje się następująco:

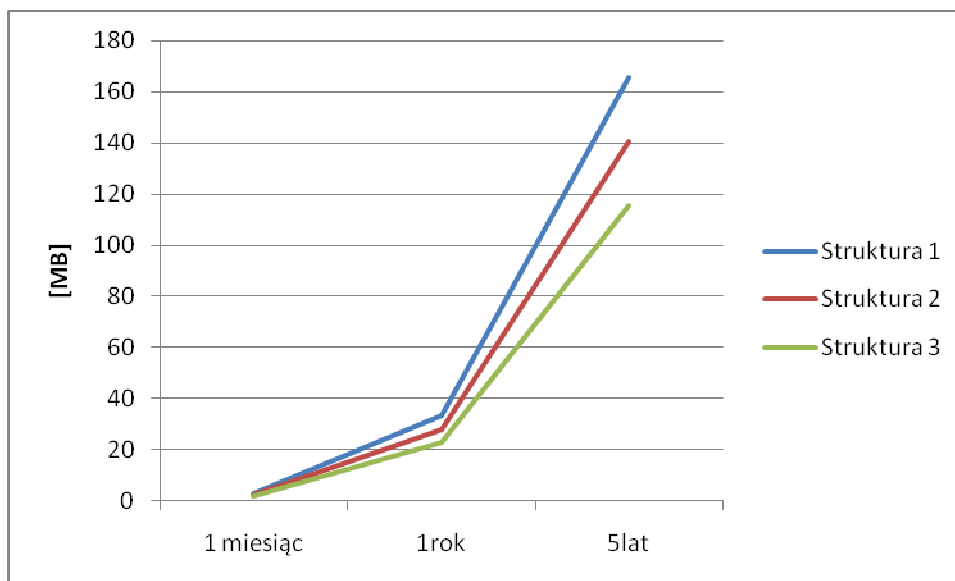
Tabela 37. Tabela Pomiary dla struktury 3

|                 |
|-----------------|
| ID_Stacji       |
| Data/Czas       |
| Temperatura 1   |
| Temperatura 2   |
| Temperatura 3   |
| Ciśnienie       |
| Opad            |
| Wilgotność      |
| Kierunek wiatru |
| Prędkość wiatru |

Dla lepszego zobrazowania problemu podam zestawienie przybliżonej wielkości baz, w których zawarte są jedynie wyniki pomiarów, nie zawierające informacji o indeksie czujnika oraz o terminie wykonania pomiaru dla 10 stacji.

Tabela 38. Objętości wypełnionych baz danych bez informacji o indeksie czujnika i o terminie wykonania pomiaru [MB]

|             | 1 miesiąc | 1rok | 5lat  |
|-------------|-----------|------|-------|
| Struktura 1 | 2,8       | 33,1 | 165,4 |
| Struktura 2 | 2,4       | 28,1 | 140,4 |
| Struktura 3 | 2,0       | 23,1 | 115,3 |



Rysunek 9. Wielkości baz danych dla 10 stacji pomiarowych nie zawierających informacji o ID czujnika oraz terminie wykonania pomiaru

Wyraźnie widoczne jest jak wiele na takim zabiegu zyskały struktury 1 i 2. Wielkości baz stworzonych na ich podstawie zmniejszyły się ponad dwukrotnie, podczas gdy pojemność bazy opartej na strukturze 3 zmniejszyła się nieznacznie. Na powyższym przykładzie widoczne jest jak istotny jest problem sposobu przechowywania daty i czasu w bazie, opisany dokładniej w rozdziale 5.7.

Należy jednak zaznaczyć, że wraz ze wzrostem heterogeniczności wprowadzanych danych pomiarowych, ich niejednorodności, stopień możliwości wprowadzania danych seriami będzie malał, co wpłynie bezpośrednio na zwiększanie wielkości bazy dla struktury 3 oraz zmniejszanie się różnicy w wielkości bazy dla poszczególnych struktur.

## 6.2. Analiza szybkości wykonywania zapytań

Drugim najistotniejszym czynnikiem wpływającym na ocenę bazy jest szybkość wykonywania zapytań. Bezpośredni wpływ na ten

czynnik ma także wielkość bazy – im więcej rekordów w tabeli tym dłuższy czas wykonania zapytań.

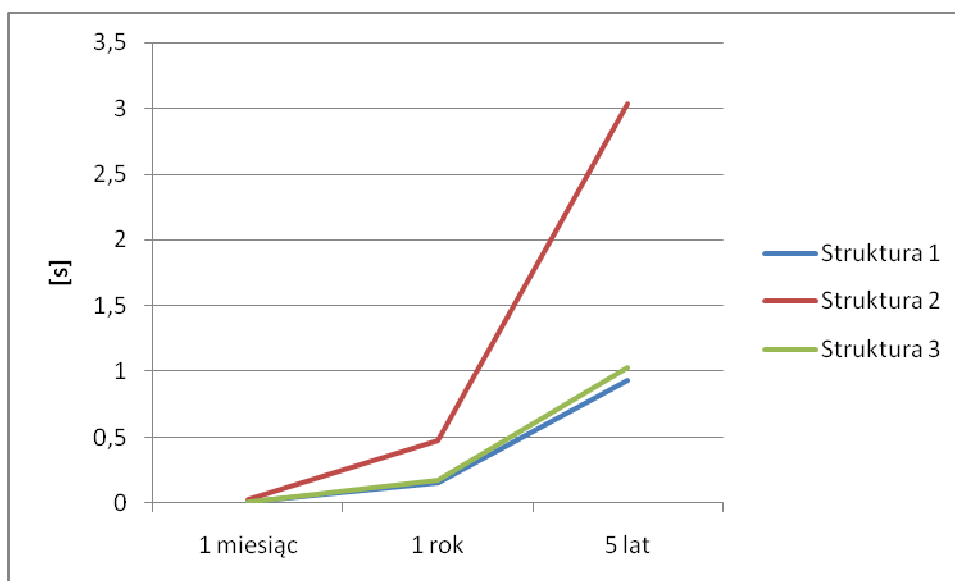
Do analizy szybkości wykonywania zapytań w poszczególnych strukturach użyłem dwóch zapytań. Dla struktury 1 i 2 wszystkie rekordy w odpowiednich tabelach zostają przeszukane według indeksu (numeru identyfikacyjnego) danego czujnika. Dla struktury 3 przeszukana jest najpierw tabela Lista Czujników aby odnaleźć indeks stacji do której przypisany jest czujnik. Dopiero gdy ID\_Stacji zostanie odnalezione może zostać wykonane właściwe zapytanie.

Zapytanie 1 polegało na odnalezieniu w bazie wszystkich pomiarów temperatur dla danego czujnika o wartości większej niż 10. Wartość ta została wybrana celowo by jak najlepiej odzwierciedlić rozwiązania zawarte w strukturze 2, przechowującej pomiary z uwzględnieniem wielkości pomiaru. Według rozwiązań zaimplementowanych w tej strukturze wartości temperatury zawierające się w przedziale od -12.8 do 12.7 powinny się znaleźć w tabeli zawierającej wartości TINYINT, natomiast pozostałe wartości będą zapisane w tabeli SMALLINT. Aby więc uzyskać wszystkie wyniki pomiaru temperatury większe niż 10°C, muszą zostać przeszukane obydwie tabele.

Zestawienia z tabel 39. oraz 40. prezentują czas wykonywania zapytań dla sieci obserwacyjnej składającej się z 10 stacji pomiarowych oraz dla sieci składającej się z 50 stacji. Wyniki pomiarów podane są w sekundach.

Tabela 39. Czasy wykonania zapytania 1 dla 10 stacji [s]

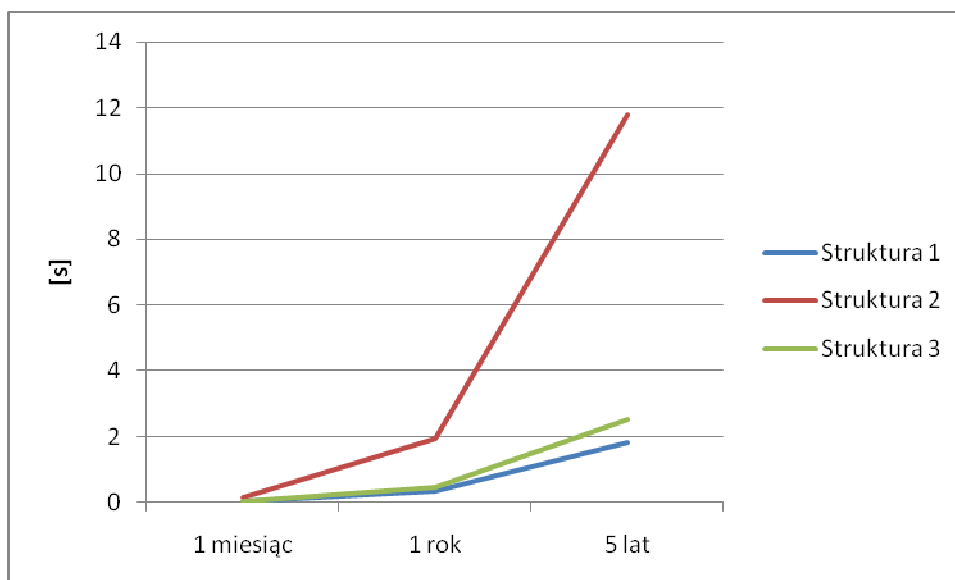
|             | 1 miesiąc | 1rok   | 5lat   |
|-------------|-----------|--------|--------|
| Struktura 1 | 0,0086    | 0,1567 | 0,9325 |
| Struktura 2 | 0,0284    | 0,4781 | 3,0385 |
| Struktura 3 | 0,0092    | 0,1713 | 1,0348 |



Rysunek 10. Czasy wykonania zapytania 1 dla 10 stacji

Tabela 40. Czasy wykonania zapytania 1 dla 50 stacji [s]

|             | 1 miesiąc | 1rok   | 5lat    |
|-------------|-----------|--------|---------|
| Struktura 1 | 0,0319    | 0,3550 | 1,8069  |
| Struktura 2 | 0,1432    | 1,9402 | 11,7778 |
| Struktura 3 | 0,0440    | 0,4606 | 2,5259  |



Rysunek 11. Czasy wykonania zapytania 1 dla 50 stacji

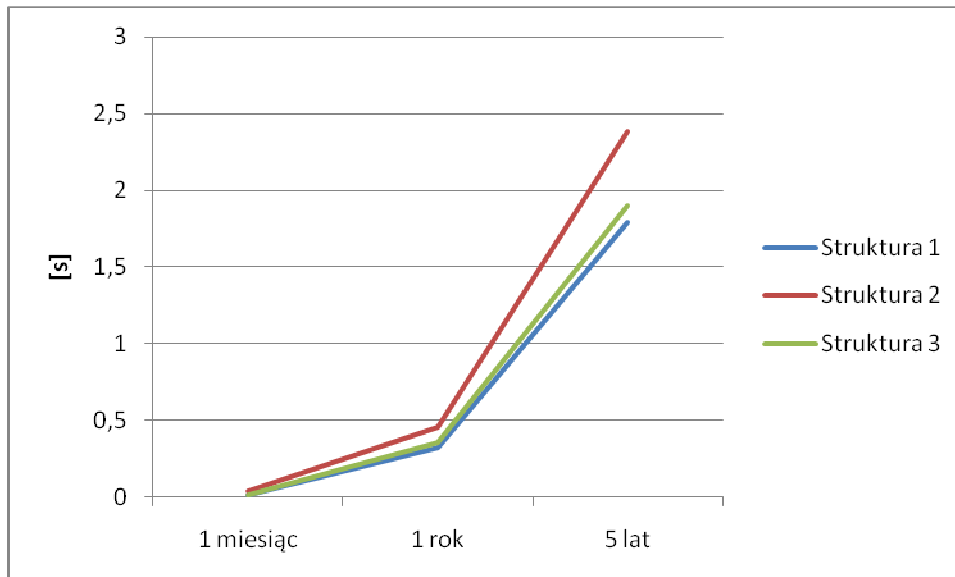
Dla struktur 1 i 3 czasy wykonania zapytania 1 są zbliżone. Natomiast dla struktury 2 czas ten jest znacznie wydłużony.

Bezpośredni wpływ na to mają rozwiązania zastosowane w tej strukturze. Aby uzyskać wszystkie wyniki należy przeszukać wszystkie rekordy w obu tabelach w bazie 2. W przypadku 50 stacji wykonujących pomiary przez 5 lat musi zostać przeszukanych 105 120 000 rekordów. Dla porównania dla pozostałych struktur musi zostać przeszukanych jedynie 13 140 000 rekordów.

Drugie zapytanie jakie zostanie wysłane do bazy będzie miało za zadanie odnalezienie wszystkich pomiarów wykonanych przez zadany czujnik.

Tabela 41. Czasy wykonania zapytania 2 dla 10 stacji [s]

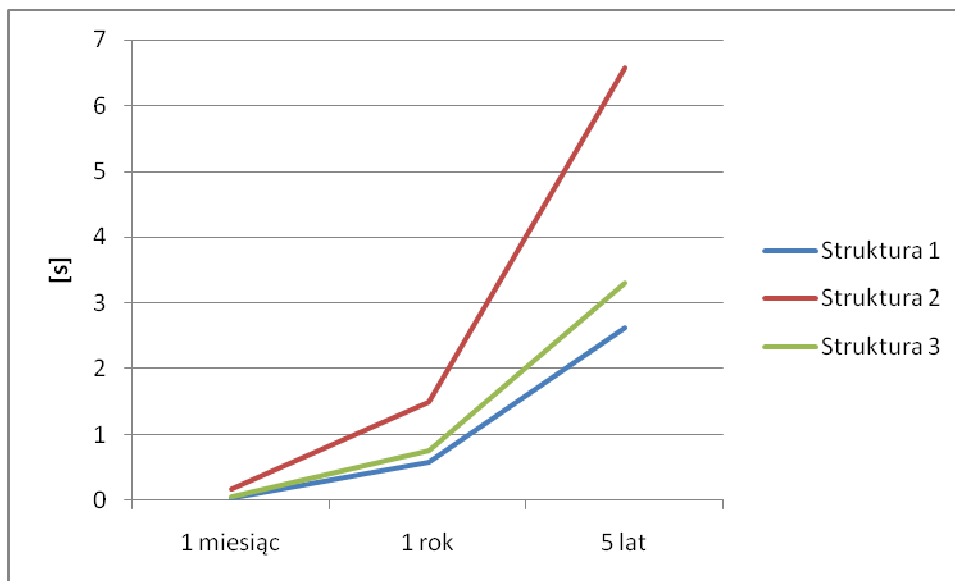
|             | 1 miesiąc | 1 rok  | 5 lat  |
|-------------|-----------|--------|--------|
| Struktura 1 | 0,0165    | 0,3280 | 1,7934 |
| Struktura 2 | 0,0373    | 0,4572 | 2,3834 |
| Struktura 3 | 0,0192    | 0,3574 | 1,8978 |



Rysunek 12. Czasy wykonania zapytania 2 dla 10 stacji

Tabela 42. Czasy wykonania zapytania 2 dla 50 stacji [s]

|             | 1 miesiąc | 1rok   | 5lat   |
|-------------|-----------|--------|--------|
| Struktura 1 | 0,0386    | 0,5847 | 2,6162 |
| Struktura 2 | 0,1766    | 1,5025 | 6,5841 |
| Struktura 3 | 0,0579    | 0,7590 | 3,3063 |



Rysunek 13. Czasy wykonania zapytania 2 dla 50 stacji

W zapytaniu 2 poszukiwane są wszystkie wartości dla danego czujnika mierzącego ciśnienie. Dla struktury 2 wszystkie wartości tego pomiaru znajdują się w tabeli SMALLINT, czyli tylko ta jedna tabela musi zostać przeszukana. Mimo, że czas wykonania tego zapytania nadal jest największy dla struktury 2, to różnice od pozostałych struktur są znacznie mniejsze niż w przypadku wykonania zapytania 1.

Postaram się poniżej ocenić zaprojektowane struktury oraz zaproponować ewentualne poprawki, które moim zdaniem powinny ulepszyć te struktury.

### 6.3. Ocena zaimplementowanych struktur

Każda zaproponowana struktura bazy danych posiada mocne, jak i słabe strony. Wszystkie zdolne są do przechowywania danych heterogenicznych pochodzących z pomiarów hydrometeorologicznych. Na etapie projektowania struktur trudno jest dokonać szczegółowej analizy szybkości zapisu czy odczytu. Dopiero po etapie implementacji przychodzi czas na testowanie rozwiązań i ewentualne wprowadzanie poprawek. Trudno je także obiektywnie ocenić na podstawie danych pochodzących z regularnych pomiarów automatycznych. Przedstawię jednak podstawową analizę projektów i postaram się odpowiedzieć na pytanie, która struktura wydaje się być najbardziej optymalną i w jakim przypadku. Postaram się także wstępnie ocenić zaproponowane struktury.

Struktura 1 porządkowała dane podzielone ze względu na mierzony parametr. Wszystkie pomiary zapisywane były do tabeli odpowiadającej rodzajowi pomiaru. Takie rozwiązanie nie wymaga dużej ilości czasu, jaką należałoby poświęcić na rozpoznawanie danych. Już na podstawie informacji zawartych w bazie metadanych można rozpoznać ilość i format tabel, jakie musimy stworzyć dla pomiarów. Schemat ten jest także łatwy w implementacji. Wszystkie tabele przechowujące pomiary mają taką samą budowę a różnią się jedynie typem pola, do jakiego wprowadzane zostają pomiary. Warto także w przypadku tej struktury wprowadzić alternatywną metodę przechowywania daty i czasu wykonania pomiaru, gdyż w testach na wielkość bazy struktura ta wypadła najgorzej. Minusem tego rozwiązania jest także fakt, że gdy zostanie wprowadzony nowy pomiar, należy dodać do bazy dodatkową tabelę zdolną przechowywać wyniki tego pomiaru.



Struktura 1 jest więc dobrym rozwiązaniem, gdy zależy nam na szybkości wprowadzania danych do bazy, jak i szybkości wykonywania zapytań. Zdecydowaną zaletą jest także przejrzysta struktura bazy i łatwa implementacja zaproponowanego rozwiązania.

Kolejny schemat, który zaprezentowałem, dzielił wyniki pomiarowe ze względu na wartość danej i przyporządkowywał do odpowiednich tabel. Rozwiązanie to podobnie do poprzedniego posiada przejrzystą strukturę, łatwą do zaimplementowania. Problem w implementacji mogą natomiast stanowić niektóre rozwiązania potrzebne do strukturalizowania danych. Każdy pomiar wprowadzany do bazy musi zostać podzielony przez dokładność zadaną dla tego pomiaru. Następnie system musi zbadać z jakiego zakresu jest to wartość i dopiero zapisywana jest ona do odpowiedniej tabeli. Taki proces wpływa negatywnie na czas zapisu danych do bazy.

Zdecydowanym pozytywem tego schematu jest oszczędność miejsca na dysku. Każdy pomiar jest odpowiednio sprawdzany przed wprowadzeniem do tabeli, by trafił do bazy z odpowiednim dla niego formatem. W ten sposób wprowadzone zostaje rozwiązanie, które np. pomiaru temperatury o wartości 3,0 °C nie wprowadzi do pola formatu FLOAT (które w MySQL zajmuje 8 bajtów), ale do pola formatu TINYINT (w MySQL to zaledwie 1 bajt). Na tym przykładzie widać, że w ten sposób zostaje zaoszczędzonych 7 bajtów. W przypadku bazy, która przyjmuje dane z tysięcy czujników, takie oszczędności są niezwykle istotne.

Rozwiązanie takie negatywnie wpływa na odczyt pomiarów z bazy z dwóch przyczyn. Pierwsza to skomplikowany sposób katalogowania danych, druga to ogromne ilości rekordów w niektórych tabelach. Jeżeli baza otrzymuje zapytanie o jakiś pomiar, najpierw sprawdzany jest zakres pomiaru by sprawdzić w jakich tabelach pomiar ten może się znajdować. Jeśli okaże się, że

jeden pomiar może znajdować się w dwóch lub więcej tabel, to proces wyszukiwania tego rekordu znacznie się wydłuża. Pewnym sposobem na zmniejszenie czasu wyszukiwania może być podział tabeli jednego formatu na kilka tabel o zadanym zakresie. Przykładowo dzielimy tabelę TINYINT na cztery różne tabele. Każda przyjmuje dane formatu TINYINT ale dodatkowo każda tabela ma zadany zakres wielkości, jaki może być do niej wprowadzony. W tym przypadku tabela 1 przyjmowałaby wartości od 0 do 63, tabela 2 wartości od 64 do 127, tabela 3 od 128 do 191 a tabela 4 od 192 do 255. Wprowadzenie takiego rozwiązania zwiększyłoby ilość tabel ale ograniczona zostałaby ilość rekordów w tych tabelach. Niestety wprowadzenie takiego rozwiązania musi być poprzedzone szczegółowymi analizami i symulacjami, aby skalkulować, w którym momencie dalsze dzielenie tabel wpływałoby negatywnie na szybkość działania bazy.

Podsumowując rozwiązania zawarte w strukturze 2 należy wyróżnić ją jako bardziej oszczędną, jeżeli bierzemy pod uwagę zajmowanie przestrzeni dyskowej. Podobnie jak dla struktury 1 warto zaimplementować w niej alternatywne sposoby przechowywania daty i czasu. Wadą jest wydłużony czas odczytu z bazy, ale jak opisałem powyżej istnieją proste rozwinięcia tego schematu, które mogą problem szybkości wykonywania zapytań znacznie zredukować.

W ostatnim zaprezentowanym schemacie tabele przechowujące dane pomiarowe tworzone były na podstawie podobieństw występujących między stacjami pomiarowymi. Ważną rolę w tej strukturze odgrywają standardy prowadzenia pomiarów. Jeżeli założymy, że dwie stacje pomiarowe wykonują te same pomiary i wszystkie pomiary w danej stacji wykonywane są o tej samej godzinie, to zaproponowane rozwiązanie zyskuje na wartości.

Rozwiązanie to ma jednak najbardziej złożoną strukturę i aby możliwa była jej implementacja potrzebna jest wnikliwa analiza danych pomiarowych przed przystąpieniem do projektowania struktury, w celu znalezienia podobieństw.

Dane pomiarowe w tym projekcie są wprowadzane do bazy seriami, co pozytywnie wpływa na czas zapisu informacji do bazy. Czas wykonywania zapytań jest tylko nieznacznie dłuższy niż w przypadku struktury 1. Na początku czujnik, który wykonywał pomiar musi zostać dopasowany do stacji, następnie wyszukiwany jest odpowiedni rekord z serią danych i w końcu na podstawie definicji tabeli, pomiar musi zostać odczytany z odpowiedniego pola. Wszystkie te czynności wydłużają odczyt danych z bazy.

Poważnym problemem jest także możliwość awarii czujnika, kiedy nie będzie można dostarczyć do bazy wyniku pomiaru. Rekord w odpowiedniej tabeli i tak zostanie stworzony ze względu na to, że inne czujniki z tej samej stacji działają poprawnie i zapisują dane. Pole, gdzie dany czujnik ma wprowadzić daną pozostanie puste. Załóżmy teraz, że ze względu na awarię, pomiaru dokonał obserwator. Musi on wprowadzić wartość do odpowiedniego pola stworzonego rekordu, dlatego bardzo ważna jest przejrzysta i jednoznaczna możliwość edycji rekordów.

Schemat 3 jest więc dobrym rozwiązaniem gdy bierzemy pod uwagę szybkość zapisu do bazy, gdyż dane wprowadzane są seriami. Zaprezentowane rozwiązanie traci jednak na wartości w przypadku wykonywania pomiarów nie według przyjętych standardów. Również możliwość awarii czujników zmniejsza ocenę tej struktury. Schematy 1 i 2 w przypadku awarii czujnika po prostu nie wpisywały rekordu do bazy. Struktura 3 taki rekord musi wprowadzić ze względu na pozostałe sprawne czujniki, które są w stanie wykonać pomiar. Pole z czujnika wadliwego zostanie puste.

Analiza zaproponowanych rozwiązań wyraźnie pokazuje, że nie jest łatwo na etapie projektowania struktury znaleźć najlepsze rozwiązanie dla problemu, który rozpatrywałem. Dokonałem jednak wstępnej oceny schematów oraz zaprezentowałem kilka możliwości dalszego rozwoju i rozbudowy wszystkich trzech schematów.

Aby stworzyć idealną strukturę bazy danych dla danych heterogenicznych, należy poznać wszystkie wyniki pomiarów, które zostaną do bazy wprowadzone. Jest to żmudna i czasochłonna praca. Dodatkowo gdy baza ma przechowywać dane zarówno z pomiarów tradycyjnych jak i z pomiarów automatycznych, tylko dogłębna analiza wyników symulacji potrafiłaby określić jakie rozwiązanie jest najlepsze i dla jakiej ilości danych należałoby zastosować inne rozwiązanie.

## **7. WNIOSKI**

Celem mojej pracy było stworzenie kilku alternatywnych rozwiązań dla problemu jakim jest integracja danych heterogenicznych pochodzących z pomiarów hydrometeorologicznych. Przeanalizowałem trzy różne metody podejścia do tego problemu a także zaprezentowałem wynik tych analiz, czyli struktury baz danych zdolne do przyjmowania wyników pomiarów prowadzonych metodami tradycyjnymi oraz tych prowadzonych przez zautomatyzowane czujniki.

W procesie tworzenia bazy dla danych heterogenicznych została stworzona baza metadanych zawierająca wszystkie wartości opisujące pomiary. Przedstawiłem jak szerokie zastosowanie może mieć taka baza.

Najważniejszym wnioskiem płynącym z mojej analizy jest jednak to, że możliwe jest przechowywanie danych heterogenicznych w jednej bazie. Dzięki temu możliwe jest tworzenie różnych zestawień – jaki wpływ ma opad na poziom wody w rzece dla danego obszaru, jaki wpływ ma temperatura powietrza na temperaturę przy gruncie, jaki wpływ ma temperatura powietrza na poziom zlodowacenia na rzece, i tym podobne. Na podstawie danych zawartych w bazie można tworzyć wykresy zmian wartości danego pomiaru, a dzięki szczegółowym informacjom o lokalizacji można tworzyć mapy pogodowe.

Oczywiście są to tylko przykładowe zastosowania gdyż jest ich znacznie więcej. Najistotniejszy jest jednak fakt, że wszystkie te zestawienia, wykresy czy symulacje mogą być tworzone dla zintegrowanych danych tradycyjnych i automatycznych. Znacznie wydłuża to okres dla którego tworzone są zestawienia i uwiarygodnia otrzymane wyniki. Dzięki integracji danych jesteśmy w stanie modelować zmiany klimatyczne z większą dokładnością.

Zaprojektowane struktury zostały poddane analizie pod względem wielkości bazy i szybkości wykonywania zapytań. Po wygenerowaniu zestawu danych i wprowadzeniu ich do baz okazało się, że struktura 3 jest najbardziej oszczędną, gdy bierzemy pod uwagę zajmowaną przestrzeń dyskową. Dla pomiarów pochodzących z 50 stacji obserwacyjnych na przestrzeni 5 lat rozmiar tej bazy okazał się ponad dwukrotnie mniejszy niż w przypadku pozostałych dwóch struktur. Warto więc ją zastosować, gdy tworzymy bazę dla ogromnej ilości danych.

Po analizie kilku przykładowych zapytań doszedłem do wniosku, że baza stworzona na podstawie struktury 1 wykonuje zapytania najszybciej, jednak jej przewaga nad strukturą 3 jest minimalna. Struktura 1 jest więc najlepszym rozwiązaniem w przypadku, gdy wymagany jest szybki dostęp do danych. Jeżeli jednak weźmiemy pod uwagę oba czynniki (wielkość bazy, szybkość wykonywania zapytań) to struktura 3 jest najlepszym wyborem. Muszę jednak dodać, że wybór optymalnej struktury bazy będzie w znacznej mierze zależał od zadań jakie ma spełniać.

Cel mojej pracy został więc osiągnięty a jej efektem są trzy projekty struktur baz danych, które są w stanie przechowywać dane heterogeniczne pochodzące z pomiarów hydrometeorologicznych. Przedstawione propozycje wykorzystania tych struktur pokazują, jak szerokie zastosowania mają takie bazy. Tworzenie takich baz jest bardzo ważne, gdyż umożliwiają one chronienie danych historycznych – tak istotnych dla naszego i przyszłych pokoleń.

## **8. BIBLIOGRAFIA**



1. Byczkowski Andrzej, „Hydrologia”, Wydawnictwo SGGW, Warszawa 1996
2. DuBois Paul, „MySQL”, Addison-Wesley, 2009
3. Hajduk Kinga, „Baza metadanych pomiarów hydrometeorologicznych zakładu hydrologii Politechniki Krakowskiej”, Politechnika Krakowska, Kraków 2007 – praca magisterska
4. Instytut Meteorologii i Gospodarki Wodnej - oficjalna strona (26.10.2008)  
[http://www.imgw.pl/wl/internet/zz/wiedza/hydro/enc\\_hydro.html](http://www.imgw.pl/wl/internet/zz/wiedza/hydro/enc_hydro.html)  
[http://www.imgw.pl/index.php?option=com\\_content&view=article&id=90&Itemid=114](http://www.imgw.pl/index.php?option=com_content&view=article&id=90&Itemid=114)
5. Janiszewski Feliks, „Instrukcja dla stacji meteorologicznych” Wydawnictwo Geologiczne, Warszawa 1988
6. Magnuszewski Artur, „Międzynarodowy słownik hydrologiczny”, Wydawnictwo Naukowe PWN SA, Warszawa 2001
7. MySQL 5.0 Reference Manual (20.01.2009)  
<http://dev.mysql.com/doc/refman/5.0/en/index.html>
8. Schneider Robert D. „MySQL® Database Design and Tuning”, Sams Publishing, 2005
9. Suehring Steve, „MySQL Bible”, Wiley Publishing Inc., New York 2002
10. Szczepanek Robert, „Relacyjne bazy danych”, „Normalizacja baz danych”, „MySQL - wprowadzenie”, Politechnika Krakowska, Kraków 2004 – niepublikowane materiały z wykładów
11. Welling Luke & Thomson Laura, „PHP and MySQL Web Development”, Sam Publishing, 2005
12. Wikipedia (20.01.2009)  
[http://pl.wikipedia.org/wiki/Relacyjna\\_baza\\_danych](http://pl.wikipedia.org/wiki/Relacyjna_baza_danych)  
<http://pl.wikipedia.org/wiki/Mysql>
13. Zabrzaska-Gąsiorek Barbara, Wróbel Andrzej „Wykorzystanie systemu zarządzania bazami danych FoxPro do gromadzenia i przetwarzania informacji z zakresu monitoringu środowiska”, Archiwum Fotogrametrii, Kartografii i Teledetekcji, Vol. 2, Kraków 1994  
[http://home.agh.edu.pl/~zfiit/publikacje\\_pliki/ZabrzaskaGasiorek\\_Wrobel\\_1994.pdf](http://home.agh.edu.pl/~zfiit/publikacje_pliki/ZabrzaskaGasiorek_Wrobel_1994.pdf) (16.10.2008)

## **9. SPISY**

## 9.1. Spis ilustracji

|             |   |    |
|-------------|---|----|
| Rysunek 1.  | Fotografia paska termogramu [Kinga Hajduk, 2006] .....  | 26 |
| Rysunek 2.  | Schemat wspólnej części bazy dla wszystkich struktur .....  | 35 |
| Rysunek 3.  | Struktura 1 – baza z danymi podzielonymi ze względu na mierzoną wielkość.....   | 40 |
| Rysunek 4.  | Struktura 2 – baza z danymi podzielonymi ze względu na wartość pomiaru.....   | 44 |
| Rysunek 5.  | Struktura 3 – baza z danymi podzielonymi ze względu na podobieństwa .....   | 48 |
| Rysunek 6.  | Schemat zmodyfikowanej struktury 3 .....  | 53 |
| Rysunek 7.  | Wielkości baz dla pomiarów pochodzących z 10 stacji .....   | 62 |
| Rysunek 8.  | Wielkości baz dla pomiarów pochodzących z 50 stacji .....   | 63 |
| Rysunek 9.  | Wielkości baz danych dla 10 stacji pomiarowych nie zawierających informacji o ID czujnika oraz terminie wykonania pomiaru ..... | 65 |
| Rysunek 10. | Czasy wykonania zapytania 1 dla 10 stacji .....   | 67 |
| Rysunek 11. | Czasy wykonania zapytania 1 dla 50 stacji .....   | 67 |
| Rysunek 12. | Czasy wykonania zapytania 2 dla 10 stacji .....   | 68 |
| Rysunek 13. | Czasy wykonania zapytania 2 dla 50 stacji .....   | 69 |

## 9.2. Spis tabel

|            |  |    |
|------------|--|----|
| Tabela 1.  | Format pól stosowanych w MySQL [MySQL.com, 2009] .....   | 18 |
| Tabela 2.  | Zestawienie parametrów meteorologicznych wraz z dokładnością pomiaru (1) [imgw.pl, 2008] .....                             | 23 |
| Tabela 3.  | Zestawienie parametrów meteorologicznych wraz z dokładnością pomiaru (2) [imgw.pl, 2008] .....                             | 24 |
| Tabela 4.  | Zestawienie parametrów hydrologicznych wraz z dokładnością pomiaru [imgw.pl, 2008] .....                                   | 25 |
| Tabela 5.  | Tabela Miejsce .....   | 31 |
| Tabela 6.  | Tabela Obserwator .....  | 32 |
| Tabela 7.  | Tabela Rodzaj Pomiaru .....  | 32 |
| Tabela 8.  | Tabela Czujnik.....  | 33 |
| Tabela 9.  | Tabela Okres Odczytu.....  | 34 |
| Tabela 10. | Przykładowe rekordy tabeli Miejsce .....   | 36 |
| Tabela 11. | Przykładowe rekordy tabeli Czujnik .....   | 36 |
| Tabela 12. | Przykładowa karta - Pomiar opadu atmosferycznego deszczomierzem Hellmanna na posterunku Ślęmień [Kinga Hajduk, 2006] ..... | 37 |
| Tabela 13. | Tabela Pomiary Temperatury .....   | 39 |
| Tabela 14. | Tabela Pomiary Stanu Wody .....  | 39 |
| Tabela 15. | Przykładowe rekordy tabeli Temperatura dla struktury 1.....  | 41 |
| Tabela 16. | Przykładowy rekord tabeli Czujnik.....   | 41 |
| Tabela 17. | Przykładowy rekord tabeli Miejsce.....   | 42 |
| Tabela 18. | Przykładowy rekord tabeli Okres Odczytu.....   | 42 |
| Tabela 19. | Przykładowa tabela przechowująca pomiary dla struktury 2.....  | 45 |
| Tabela 20. | Tabela Stacje .....  | 49 |
| Tabela 21. | Tabela Lista czujników .....   | 49 |
| Tabela 22. | Przykładowa tabela z danymi po zgrupowaniu stacji 1 i 2.....   | 50 |
| Tabela 23. | Przykładowe rekordy tabeli z pomiarami .....   | 50 |
| Tabela 24. | Przykładowa tabela przechowująca pomiary rodzaju chmur dla struktury 3.....  | 50 |

---

|            |  |    |
|------------|--|----|
| Tabela 25. | Przykładowe rekordy tabeli przechowującej pomiary rodzaju chmur dla struktury 3.....                           | 51 |
| Tabela 26. | Zmodyfikowana tabela Czujnik dla schematu 3 .....  | 51 |
| Tabela 27. | Tabela Kiedy Pomiar zastosowana dla struktury 3 .....  | 52 |
| Tabela 28. | Przykładowe rekordy tabeli Kiedy Pomiar dla struktury 3 .....  | 52 |
| Tabela 29. | Przykładowe rekordy tabeli Czas wykonania pomiaru.....   | 57 |
| Tabela 30. | Przykładowy rekord tabeli Temperatura dla struktury 1 .....  | 57 |
| Tabela 31. | Przykładowy rekord tabeli Temperatura dla struktury 1 z zastosowaniem pól DATE oraz TIME .....                 | 58 |
| Tabela 32. | Przykładowy rekord tabeli Temperatura dla struktury 1 z zastosowaniem pola TIMESTAMP .....                     | 58 |
| Tabela 33. | Przykładowy rekord tabeli Temperatura dla struktury 1 z rozwiązaniem opartym na identyfikatorach czasu .....   | 58 |
| Tabela 34. | Liczba wykonywanych pomiarów przez czujniki na przestrzeni zadanych okresów symulacji .....                    | 61 |
| Tabela 35. | Objętość wypełnionych baz danych dla danych pochodzących z 10 stacji [MB] .....                                | 62 |
| Tabela 36. | Objętość wypełnionych baz danych dla danych pochodzących z 50 stacji [MB] .....                                | 63 |
| Tabela 37. | Tabela Pomiary dla struktury 3 .....   | 64 |
| Tabela 38. | Objętości wypełnionych baz danych bez informacji o indeksie czujnika i o terminie wykonania pomiaru [MB] ..... | 64 |
| Tabela 39. | Czasy wykonania zapytania 1 dla 10 stacji [s] .....  | 66 |
| Tabela 40. | Czasy wykonania zapytania 1 dla 50 stacji [s] .....  | 67 |
| Tabela 41. | Czasy wykonania zapytania 2 dla 10 stacji [s] .....  | 68 |
| Tabela 42. | Czasy wykonania zapytania 2 dla 50 stacji [s] .....  | 69 |

## **Abstrakt**

Pomiary hydrometeorologiczne na przestrzeni ostatnich kilkunastu lat przeżyły rewolucję. Zmiany sposobu prowadzenia pomiarów, częstotliwości ich wykonywania czy sposobu zapisu miały bezpośredni wpływ na problem katalogowania ich w jednym miejscu.

Niniejsza praca opisuje problem przechowywania heterogenicznych danych w jednej bazie. Jej celem jest zaprojektowanie struktur baz danych zdolnych do gromadzenia i przetwarzania niejednorodnych danych hydrometeorologicznych.

Efektom jest projekt trzech struktur relacyjnych baz danych. Szczegółowo opisano budowę wszystkich tabel i relacje między nimi panujące. Na przykładowych scenariuszach dostępu do bazy opisano zastosowane rozwiązania. Wszystkie zaprojektowane struktury zostały zaimplementowane w MySQL, a następnie poddane badaniom pod względem pojemności bazy oraz szybkości wykonywania zapytań. Wnioski z analizy zostały szczegółowo opisane. Przedstawiono także propozycje rozwoju struktur i ulepszenia ich funkcjonalności.

**Abstract**

Hydrometeorological monitoring has undergone a considerable revolution in the recent years. The changes in the methods of monitoring, their frequency and the manner of recording have had a direct impact upon the problem of single space storage.

This paper presents the issue of storing heterogeneous data in a single database. The aim is to develop the structure of a database capable of storing and processing non-homogeneous hydrometeorological data.

The above resulted in the design of three relational databases. The construction and relation between all of the tables is presented in detail. All of the developed structures have been implemented in MySQL and then analyzed with respect to the database storage capacity and time of request. The analysis conclusions are described in detail. Furthermore, a structure development and functionality enhancement proposal is offered.