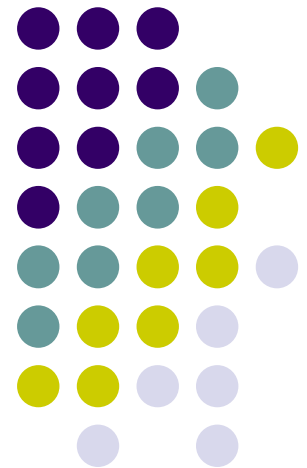
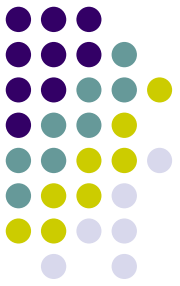


Bazy danych „Pi of the Sky”

Marek Biskup
Uniwersytet Warszawski
mbiskup@mimuw.edu.pl



Zawartość baz danych



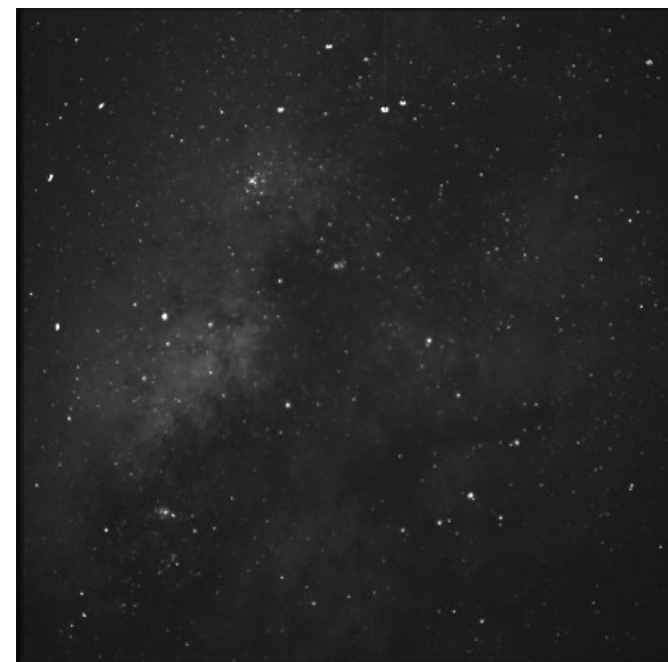
- Pomiar jasności gwiazd
 - Czas wykonania pomiaru
 - Jasność
 - Dodatkowe parametry
 - Numer kamery
- Uśrednione informacje o gwiazdach
 - Średnia jasność
 - Średnie położenie
 - Nazwa (astronomiczna)



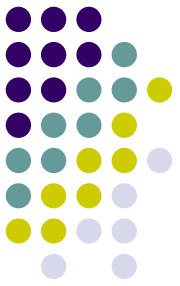
Rozmiar bazy danych



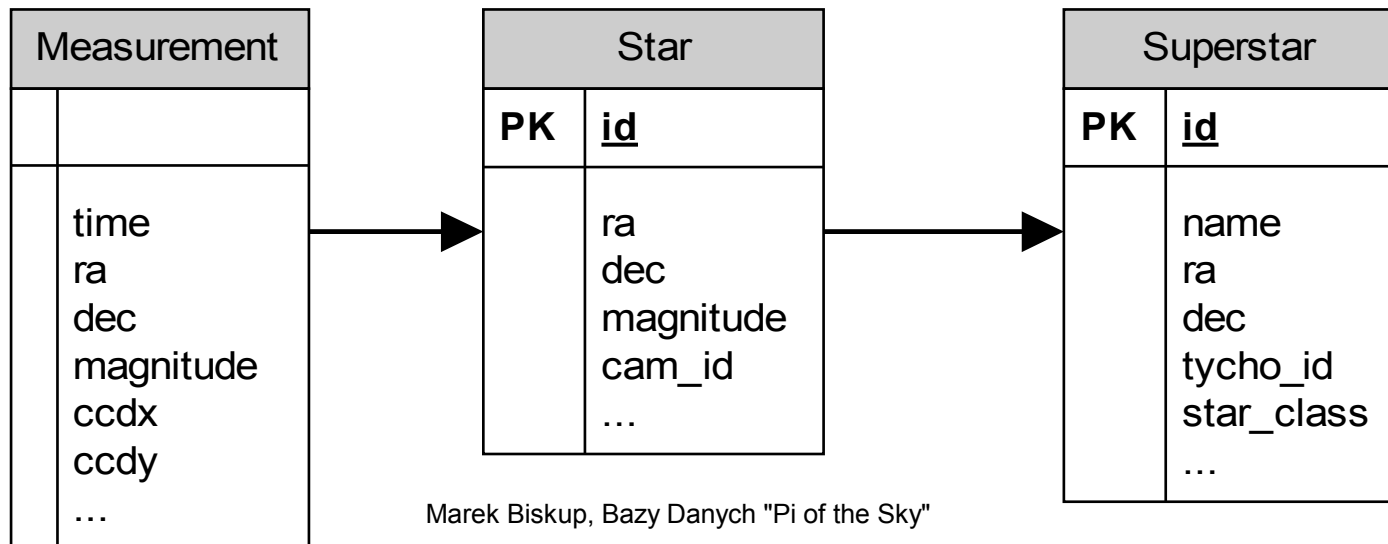
- Około 6 mln pomiarów na dobę
 - Zdjęcia co 12 s., sumowane po 20
 - ~20000 gwiazd / zdjęcie
 - ~10 godzin obserwacji / noc
 - 2 kamery
- Ponad 1 mld pomiarów na rok
 - ~200GB
- Docelowo 32 kamery
 - 100 mln pomiarów na dobę
 - 20 mld pomiarów na rok
 - Nawet 20x więcej jeśli obrabiane będą pojedyncze klatki



Struktura bazy danych



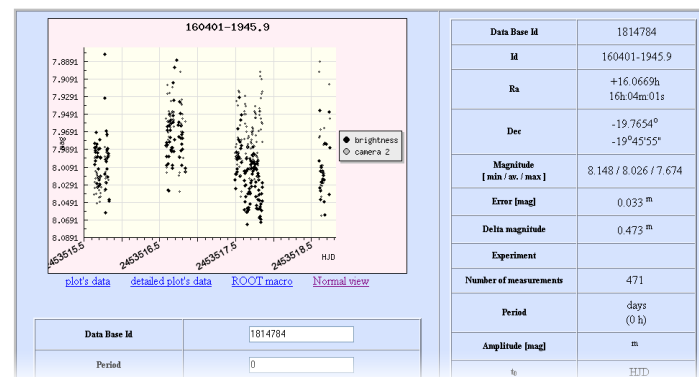
- Podstawowe tabele
 - **Measurements** – pomiary jasności
 - **Stars** – uśrednione pomiary z jednej kamery dla każdej gwiazdy
 - **Superstars** – uśrednione pomiary ze wszystkich kamer dla każdej gwiazdy



Implementacja bazy danych



- Engine: PostgreSQL
- Kilka baz danych gwiazd
 - Oddzielne bazy danych dla różnych okresów
 - Bazy danych z informacjami z katalogów gwiazd
 - Robocze bazy danych dla aktualnej fazy eksperymentu (fragmenty danych)
- Skrypty do ładowania z plików z danymi
- Interfejs WWW (php)



Problemy do rozwiązania



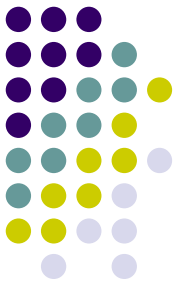
- Wiele baz danych
 - bałagan (trudności w utrzymaniu)
 - brak możliwości łączenia danych do zapytań SQL
- PostgreSQL – brak:
 - narzędzi dostępnych dla komercyjnych baz danych
 - ręcznej optymalizacji zapytań
 - rozproszenia bazy danych
- Niedoskonałe „środowisko pracy naukowca”

Plan zmian baz danych



- **Migracja bazy** do DB2 + integracja baz danych
- Opracowanie strategii backupów, fail-over, partycjonowania, utrzymania środowiska roboczego
- Stworzenie bazy buforowej dla nowych danych
- Migracja oprogramowania do DB2
 - Biblioteki C++, interfejsy PHP, procedury PL/SQL
- Analiza kosztów zapytań, klastrowanie, materialized views, **optymalizacja** bazy i zapytań SQL

Dodatkowe tematy

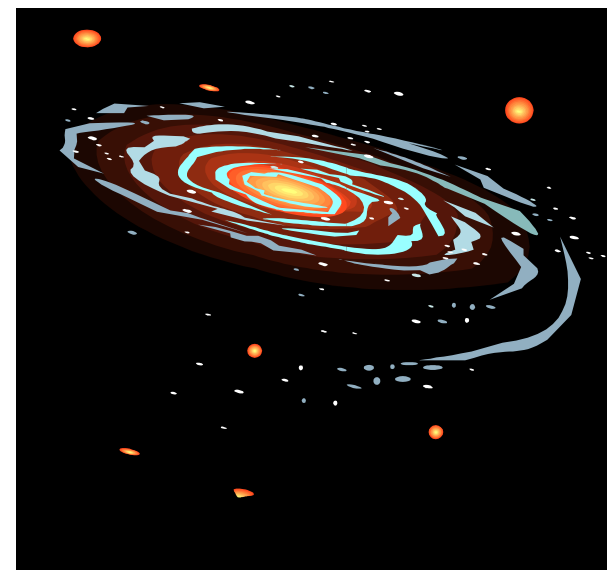


- **System kontroli eksperymentu**
 - Parametry pracy urządzeń zapisywane do bazy danych
 - Interfejs do kontroli
 - Analiza danych – możliwość pisania pluginów
 - Technologie: np. Ruby on Rails, GWT, Ajax, Django
- Interfejs przeglądania gwiazd w stylu „Google Maps” (płynne przewijanie, zoom, zdjęcia, etc.)

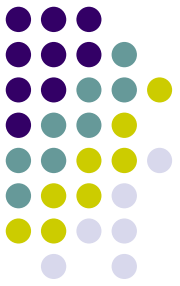
Możliwości



- Uczestnictwo w **eksperymentach naukowym**
- Praca z „produkcyjną” bazą danych średniego i dużego (w przyszłości) rozmiaru
- Poznawanie **nowych technologii**
 - w szczególności DB2 i jej narzędzi
 - i PostgreSQL, aby zmigrować dane
- Starcie z typowymi problemami **inżynierii oprogramowania**



Podsumowanie



- W „Pi of the Sky” pracuje się z rzeczywistymi i sporymi bazami danych
- „Pi” jest otwarte na współpracę ze studentami
- „Pi” planuje migrację do DB2 i integrację baz danych
 - Przy tym potrzebna jest pomoc
- Nawiązanie współpracy z IBM otwiera możliwość poznania technologii tej firmy

Dodatkowe informacje



- Strona projektu „Pi of the Sky”
 - <http://grb.fuw.edu.pl>
- DB2 partitioning features
 - <http://www.ibm.com/developerworks/db2/library/techarticle/dm-0608mcinerney/>
- Kontakt: mbiskup@mimuw.edu.pl