Bazy danych i usługi sieciowe

Podsumowanie

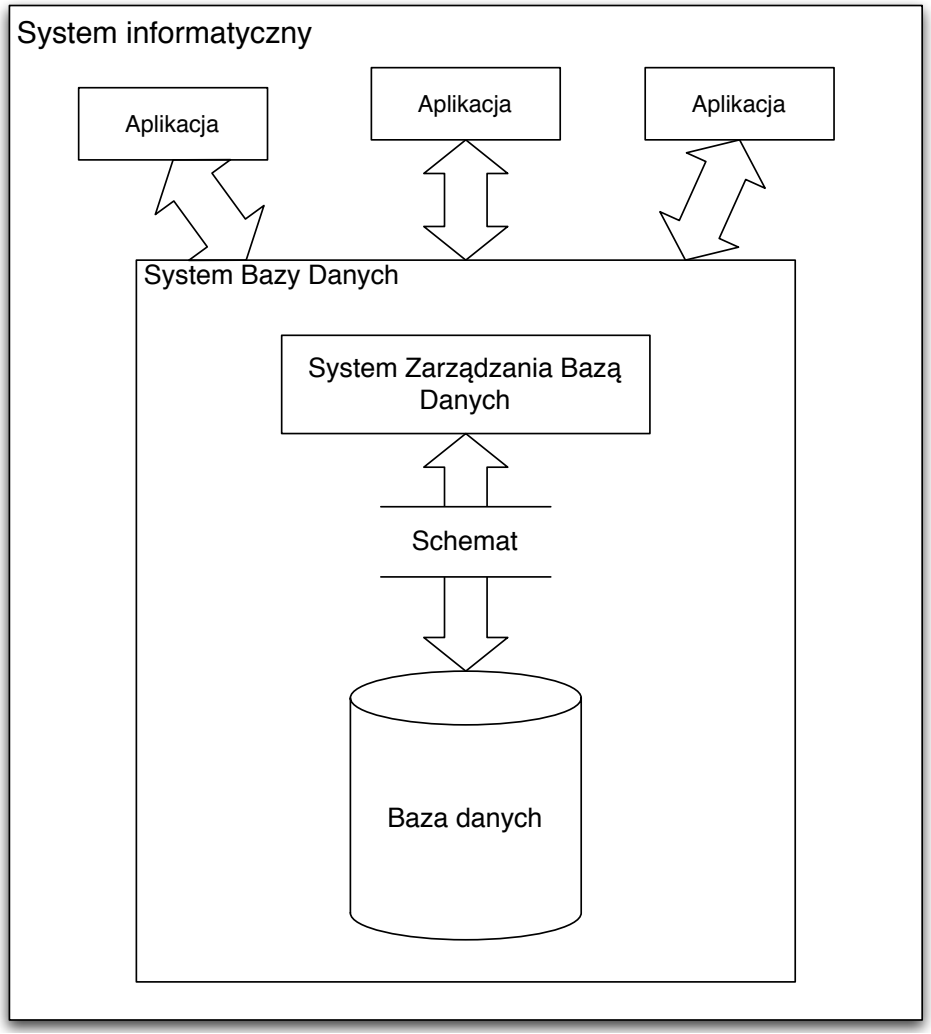
Paweł Daniluk

Wydział Fizyki

Jesień 2014




Użytkownicy


System informatyczny
Aplikacja
Aplikacja
Aplikacja
System Bazy Danych
System Zarządzania Bazą Danych
Schemat
Baza danych

Bazy danych

Wymagania

- Dane są niezależne od korzystających z nich aplikacji
- Długi czas przechowywania (kilkadziesiąt lat)
- Równoczesny dostęp wielu użytkowników
- Wydajność
- Bezpieczeństwo

Spójność bazy danych

- Zgodność danych z rzeczywistością
- Zgodność zależności między danymi z przyjętym modelem
- Odporność na błędy i awarie
- Brak anomalii związanych ze współbieżnym dostępem
- Odporność na błędy użytkowników

Bazy danych c.d.

Transakcje

- Atomic – wykonują się w całości albo wcale
- Consistent – nie naruszają spójności bazy danych
- Isolated – są od siebie niezależne
- Durable – wynik zakończonej transakcji nie może zostać utracony

Zastosowania

- Przetwarzanie transakcyjne (On-Line Transaction Processing) - OLTP
  - systemy ewidencyjne
- Przetwarzanie analityczne (On-Line Analytical Processing) - OLAP
  - hurtownie danych
- Inne
  - Computer Aided Desing - CAD
  - Geographical Information Systems - GIS
  - Protein Data Bank - PDB

Modelowanie

Modelowanie polega na odwzorowaniu obiektów świata rzeczywistego w systemie informatycznym(bazie danych).

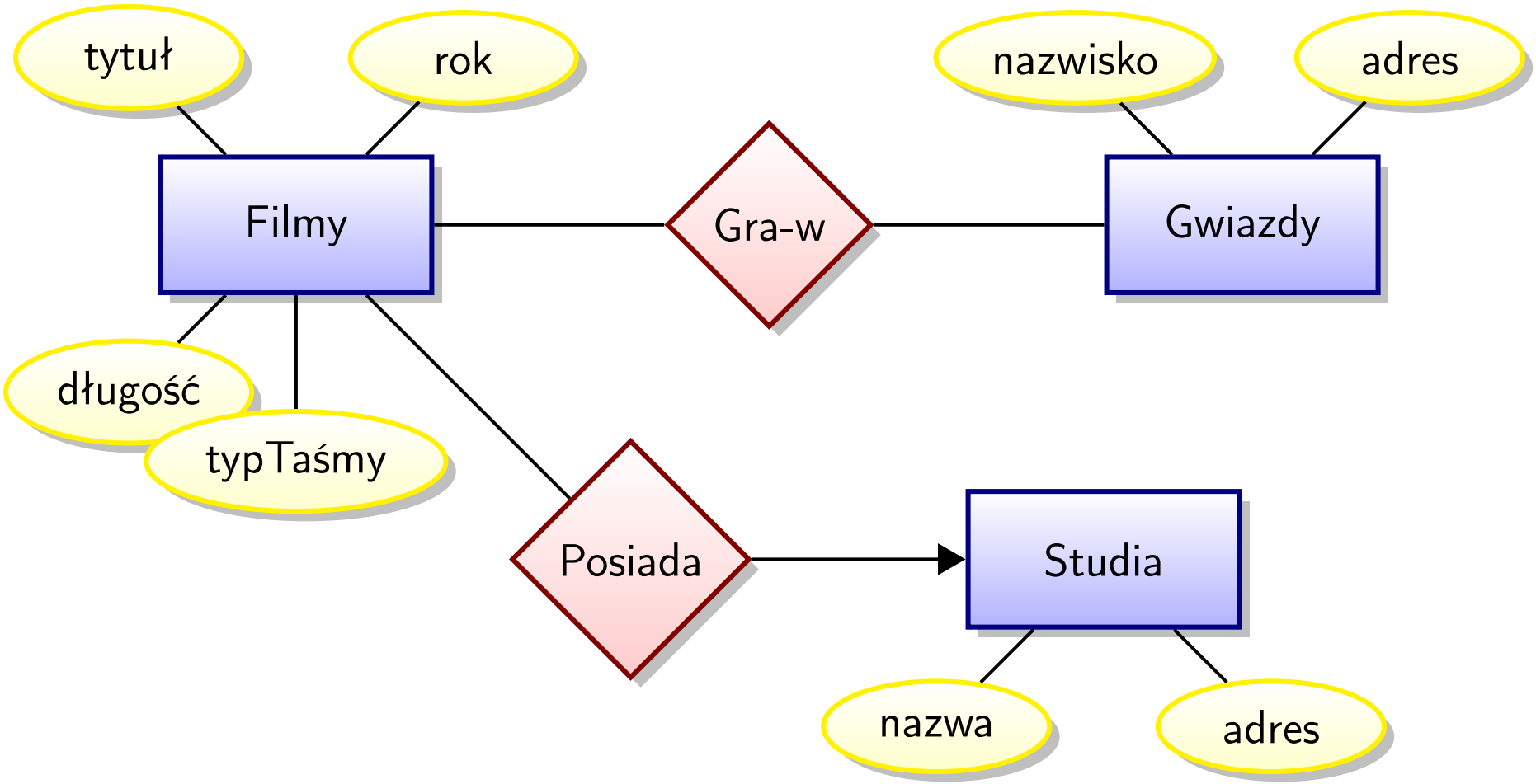
Modele konceptualne

- uniwersalne
- niezależne od modelu implementacyjnego
- np. model związków encji

Modele implementacyjne

- wykorzystywane do implementacji modeli konceptualnych
- np. model relacyjny

Diagram związków encji



Spójność bazy danych

Klucze

- W każdym zbiorze encji musi być klucz.
- Klucz może składać się z więcej niż jednego atrybutu.
- Zbiór encji może mieć więcej niż jeden potencjalny klucz. Zazwyczaj wyróżniamy klucz główny.
- Jeżeli zbiór encji należy do hierarchii "isa", wszystkie atrybuty klucza muszą należeć do korzenia hierarchii.

Spójność bazy danych c.d.

Więzy

- Więzy jednoznaczności – wymaganie, aby wartość w określonym kontekście była unikatowa: klucze, związki wiele do jeden
- Więzy integralności referencyjnej – odwołania muszą wskazywać na obiekty, które faktycznie znajdują się w bazie
- Więzy domenowe – wymaganie, aby wartość należała do określonego zbioru lub zakresu
- Więzy zasadnicze – inne arbitralnie narzucone warunki

Model relacyjny

- Relacja to podzbiór iloczynu kartezjańskiego
- Dwuargumentowa $R \subset A \times B$
- n-argumentowa $R \subset A_1 \times A_2 \times \ldots A_n$
- Reprezentacja jako dwuwymiarowa tabela

Relacja *Filmy*

tytuł	rok	długość	typFilmu
Gwiezdne Wojny	1977	124	kolor
Potężne Kaczory	1991	104	kolor
Świat Wayne'a	1992	95	kolor

Anomalie

Filmy

tytuł	rok	długość	typFilmu	nazwaStudia	nazwiskoGwiazdy
Gwiezdne Wojny	1977	124	kolor	Fox	Carrie Fisher
Gwiezdne Wojny	1977	124	kolor	Fox	Mark Hamill
Gwiezdne Wojny	1977	124	kolor	Fox	Harrison Ford
Potężne Kaczory	1991	104	kolor	Disney	Emilio Estevez
Świat Wayne'a	1992	95	kolor	Paramount	Dana Carvey
Świat Wayne'a	1992	95	kolor	Paramount	Mike Meyers

- Redundancja
- Anomalie przy aktualizacji
- Anomalie przy usuwaniu
- Anomalie przy wstawianiu

Normalizacja

- Wyeliminowanie redundancji
- Wyeliminowanie anomalii
- *Dekompozycja* - rozkład relacji na mniejsze, które nie mają redundancji i anomalii
- Potrzebna informacja o zależnościach funkcyjnych

Własności dla normalizacji

- Zachowanie atrybutów
- Zachowanie informacji
- Zachowanie zależności funkcyjnych w relacjach po normalizacji

Pierwsza postać normalna - 1NF

1NF

Schemat relacji R znajduje się w pierwszej postaci normalnej (1NF), jeżeli wartości atrybutów są atomowe (niepodzielne).

Przykład

tytuł	rok	długość	typFilmu	nazwaStudia	gwiazdy
Gwiezdne Wojny	1977	124	kolor	Fox	Carrie Fisher, Mark Hamill, Harrison Ford
Potężne Kaczory	1991	104	kolor	Disney	Emilio Estevez
Świat Wayne'a	1992	95	kolor	Paramount	Dana Carvey, Mike Meyers

Filmy(tytuł, rok, długość, nazwaStudia, Gwiazdy(nazwisko))

Druga postać normalna - 2NF

2NF

Schemat relacji R znajduje się w drugiej postaci normalnej (2NF), jeżeli żaden atrybut niekluczowy tej relacji nie jest częściowo zależny funkcyjnie od żadnego z kluczy relacji R.

Filmy

tytuł	rok	długość	typFilmu	nazwaStudia	nazwiskoGwiazdy	adresGwiazdy
Gwiezdne Wojny	1977	124	kolor	Fox	Carrie Fisher	Malibu
Gwiezdne Wojny	1977	124	kolor	Fox	Mark Hamill	Brentwood
Gwiezdne Wojny	1977	124	kolor	Fox	Harrison Ford	Hollywood
Potężne Kaczory	1991	104	kolor	Disney	Emilio Estevez	West Hollywood
Świat Wayne'a	1992	95	kolor	Paramount	Dana Carvey	Beverly Hills
Świat Wayne'a	1992	95	kolor	Paramount	Mike Meyers	Carlsbad

Zależności funkcyjne

tytuł rok nazwiskoGwiazdy $\rightarrow$ *długość typFilmu nazwaStudia adresGwiazdy*
tytuł rok $\rightarrow$ *długość typFilmu nazwaStudia*
nazwiskoGwiazdy $\rightarrow$ *adresGwiazdy*

Trzecia postać normalna - 3NF

3NF

Schemat relacji R znajduje się w trzeciej postaci normalnej (3NF), jeżeli dla każdej zależności funkcyjnej $\{A_1, \ldots, A_n\} \rightarrow B$ zbiór $\{A_1, \ldots, A_n\}$ jest nadkluczem lub B jest elementem pewnego klucza.

StudiaFilmowe

tytuł	rok	długość	typFilmu	nazwaStudia	adresStudia
Gwiezdne Wojny	1977	124	kolor	Fox	Hollywood
Potężne Kaczory	1991	104	kolor	Disney	Buena Vista
Świat Wayne'a	1992	95	kolor	Paramount	Hollywood

Zależności funkcyjne

tytuł rok $\rightarrow$ *długość typFilmu nazwaStudia adresStudia*
nazwaStudia $\rightarrow$ *adresStudia*

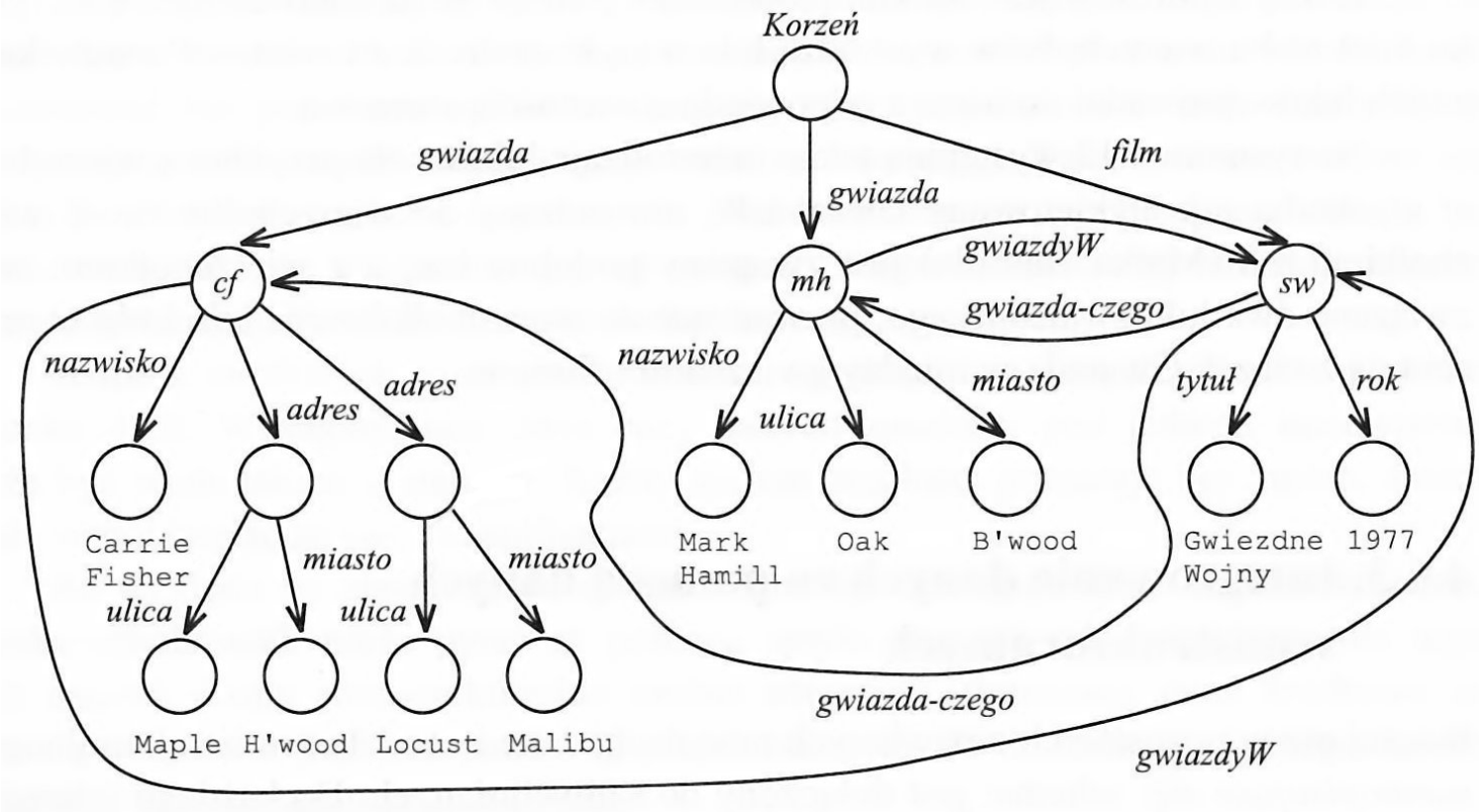
Postać normalna Boyce'a-Codda - BCNF

BCNF

Schemat relacji R znajduje się w postaci normalnej Boyce'a-Codda (BCNF), jeżeli dla każdej zależności funkcyjnej $\{A_1, \ldots, A_n\} \rightarrow B$ zbiór $\{A_1, \ldots, A_n\}$ jest nadkluczem.

- Warunek BCNF jest dostateczny dla 3NF, ale nie konieczny.
- BCNF $\Longrightarrow$ 3NF

Semistrukturalny model danych



Nie zawsze jest potrzeba albo możliwość posiadania ściśle zdefiniowanego schematu danych.

W modelu semistrukturalnym dane określają schemat.

Semistrukturalny model danych c.d.

Zalety

- Swoboda w rozszerzaniu i adaptowaniu do nowych potrzeb.
- Brak ograniczeń wynikających z modelu danych.

Wady

- Trudność w formułowaniu zapytań.
- Brak wydajnych implementacji pozwalających na przechowywanie dużych ilości danych.

Zastosowania

- Integracja danych pochodzących z różnych źródeł.
- Adaptacja "starych" baz danych do nowych potrzeb.
- Opis dokumentów.

NoSQL (*not only SQL*)

Twierdzenie CAP (Brewer 2000, Gilbert&Lynch 2002)

W systemie rozproszonym nie da się zapewnić równocześnie:

- spójności (*Consistency*) – wszystkie węzły mają te same dane
- dostępności (*Availability*) – gwarancji, że na każde żądanie zostanie udzielona odpowiedź
- odporności na podział (*Partition tolerance*) – poprawnego działania mimo utraty komunikatów

NoSQL (*not only SQL*)

Twierdzenie CAP (Brewer 2000, Gilbert&Lynch 2002)

W systemie rozproszonym nie da się zapewnić równocześnie:

- spójności (*Consistency*) – wszystkie węzły mają te same dane
- dostępności (*Availability*) – gwarancji, że na każde żądanie zostanie udzielona odpowiedź
- odporności na podział (*Partition tolerance*) – poprawnego działania mimo utraty komunikatów

NoSQL

- brak ustalonych tabel
- brak złączeń
- skalowalność horyzontalna
- brak gwarancji spójności
- lepsza wydajność przy dużych operacjach we/wy

NoSQL (*not only SQL*) c.d.

Techniki

- tablice asocjacyjne (klucz-wartość)
- XML – XQuery

Zastosowania

- Digg – green badges (3TB)
- Facebook – inbox search (50 TB)

UnQL (Unstructured Query Language)

```
CREATE COLLECTION abc;
```

```
INSERT INTO abc VALUE 1234;
INSERT INTO abc VALUE 3.141592653;
INSERT INTO abc VALUE "This is a string";
INSERT INTO abc VALUE ["this", "is", "an", "array"];
INSERT INTO abc VALUE {type: "message", content: "This is an object"

INSERT INTO abc VALUE {
  type: "nested",
  content: {
    content: "nested object",
    x: 1,
    y: {str: "hi", str2: "there"},
    z: true
  }
};
```

UnQL (Unstructured Query Language) c.d.

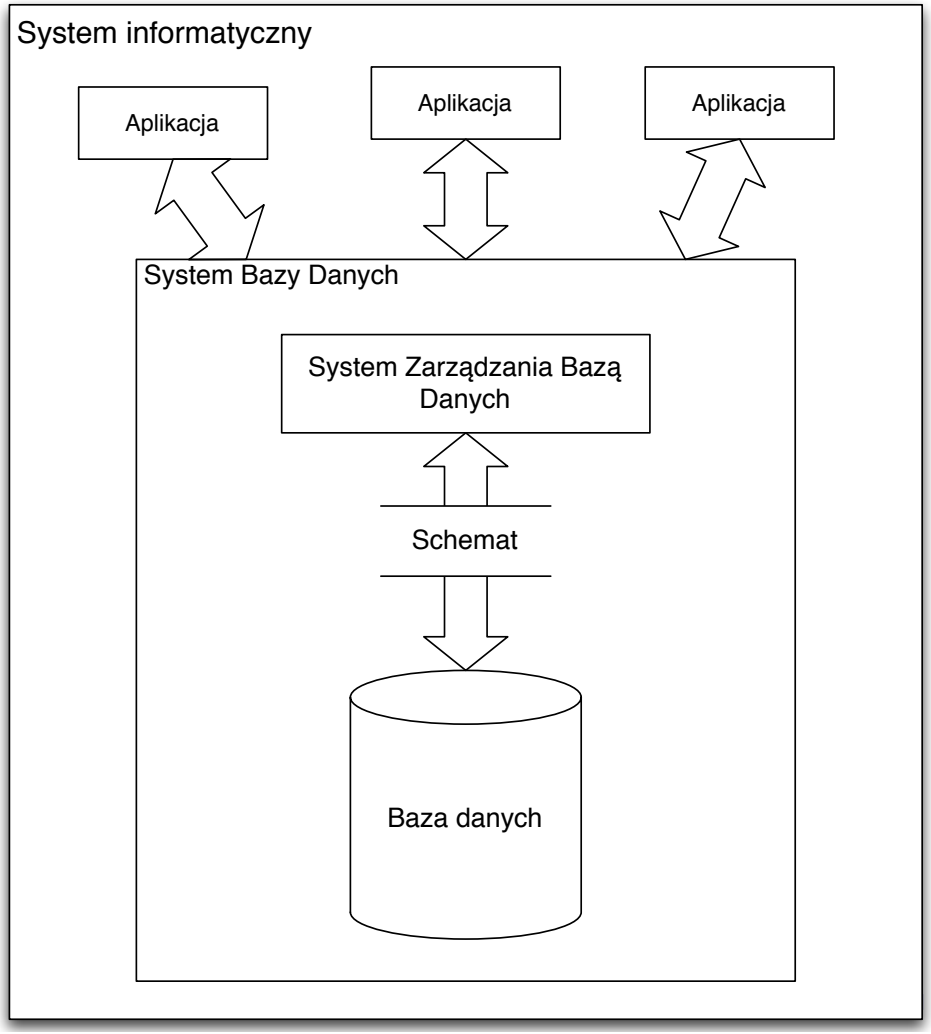
```
SELECT FROM abc;
```

```
1234
3.141592653
"This is a string"
["this","is","an","array"]
{"type":"message","content":"This is an object"}
{"type":"nested","content":
    {"content":"nested object","x":1,"y":
        {"str":"hi","str2":"there"},"z":true}}
```

```
SELECT { x:abc.type, y:abc.content.x, z:abc.content.x+50 } FROM abc;
```

Wielodostępność





Wielodostępność

1. Wiele terminali podłączonych do jednego komputera



Wielodostępność

1. Wiele terminali podłączonych do jednego komputera



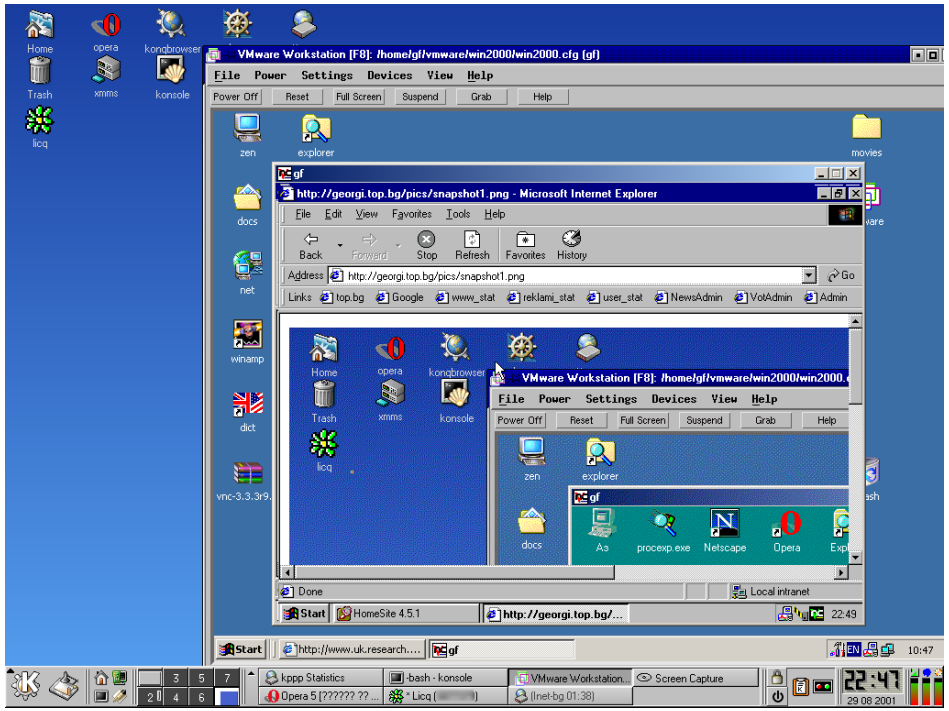
Wielodostępność

1. Wiele terminali podłączonych do jednego komputera
2. Zdalny terminal
   - tekstowy (xterm)



Wielodostępność

1. Wiele terminali podłączonych do jednego komputera
2. Zdalny terminal
   - tekstowy (xterm)
   - graficzny (VNC, thin-client)



Wielodostępność

1. Wiele terminali podłączonych do jednego komputera
2. Zdalny terminal
   - tekstowy (xterm)
   - graficzny (VNC, thin-client)



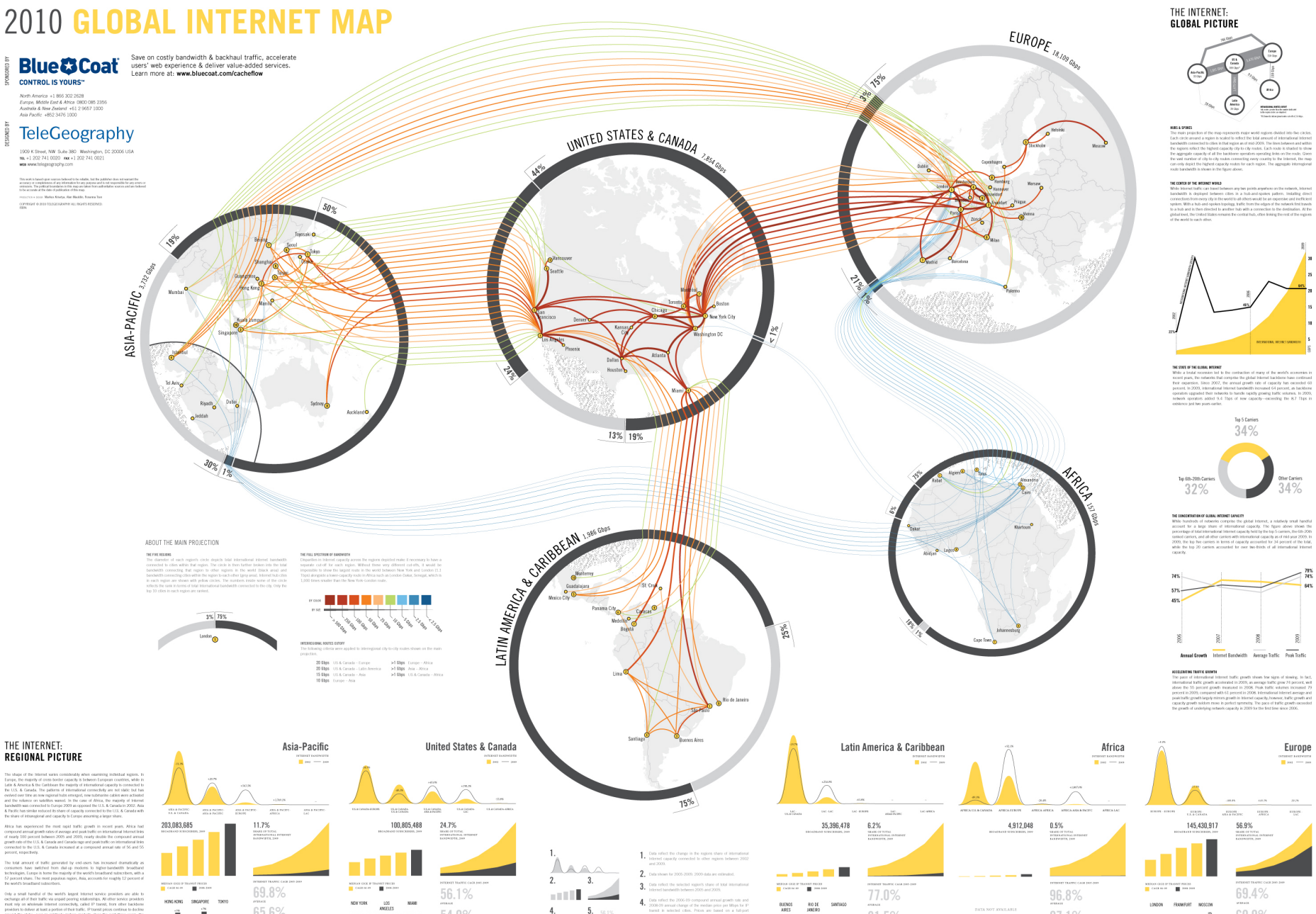
Wielodostępność

1. Wiele terminali podłączonych do jednego komputera
2. Zdalny terminal
   - tekstowy (xterm)
   - graficzny (VNC, thin-client)
3. Dedykowana aplikacja (Płatnik)

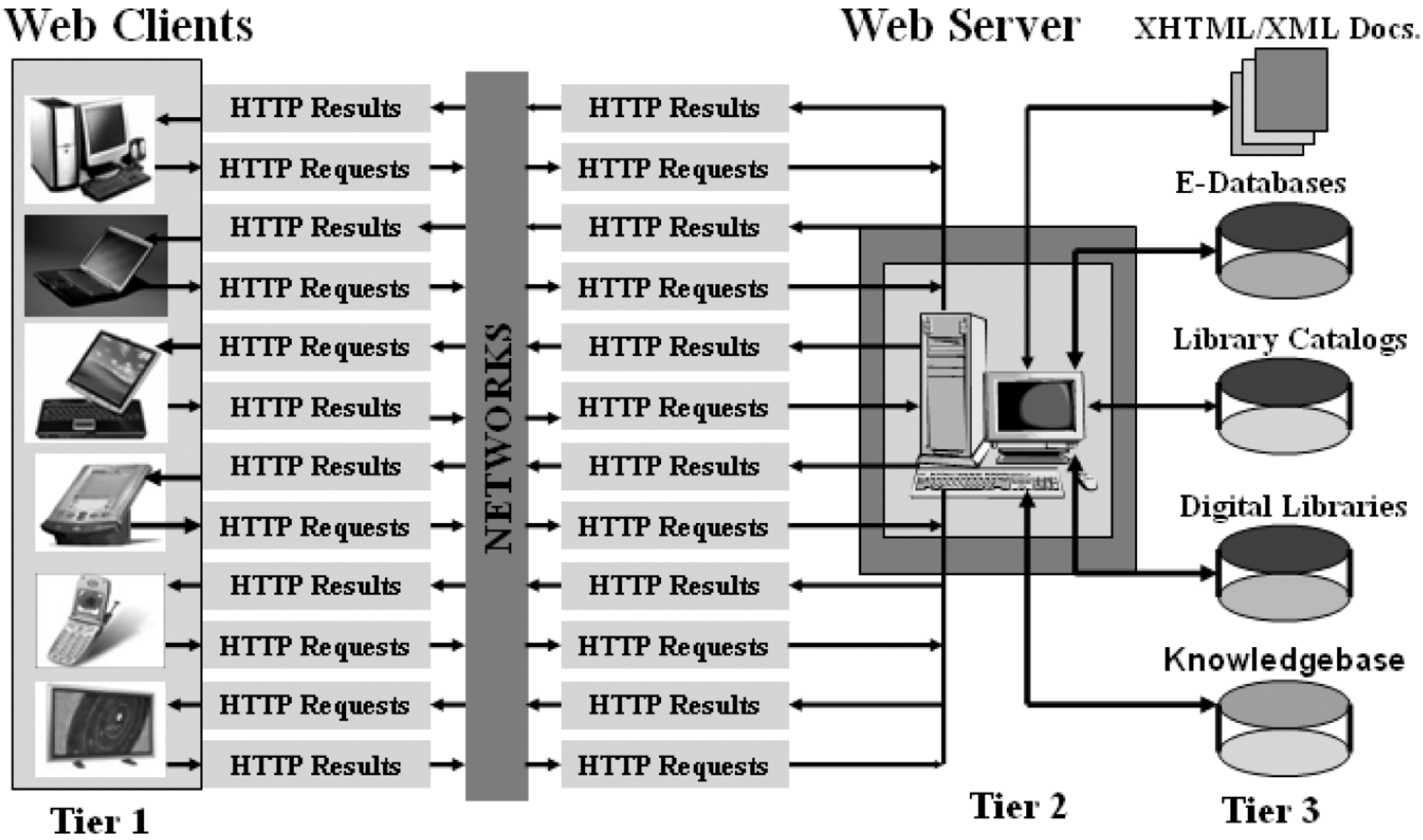
Wielodostępność

1. Wiele terminali podłączonych do jednego komputera
2. Zdalny terminal
   - tekstowy (xterm)
   - graficzny (VNC, thin-client)
3. Dedykowana aplikacja (Płatnik)
4. Aplikacja WWW

Internet


2010 GLOBAL INTERNET MAP
SPONSORED BY
BlueCoat
CONTROL IS YOURS™
Save on costly bandwidth & backhaul traffic, accelerate users' web experience & deliver value-added services. Learn more at: www.bluecoat.com/cacheflow
North America +1 866 302 2628
Europe, Middle East & Africa 0800 085 2355
Australia & New Zealand +61 2 9657 1000
Asia Pacific +852 3476 1000
DESIGNED BY
TeleGeography
1909 K Street, NW Suite 380 Washington, DC 20006 USA
TEL +1 202 741 0020 FAX +1 202 741 0021
WEB www.telegeography.com
ASIA-PACIFIC 3,232 Gbps
UNITED STATES & CANADA 7,854 Gbps
EUROPE 18,109 Gbps
LATIN AMERICA & CARIBBEAN 1,966 Gbps
AFRICA 157 Gbps
ABOUT THE MAIN PROJECTION
THE FULL SPECTRUM OF BANDWIDTH
INTERREGIONAL ROUTES CUTOFF
THE INTERNET:
GLOBAL PICTURE
THE INTERNET:
REGIONAL PICTURE
Asia-Pacific
United States & Canada
Latin America & Caribbean
Africa
Europe

Model klient-serwer



Aplikacje WWW

Warstwa danych

Serwer bazy danych (np. MySQL).

Warstwa aplikacji

- PHP – PHP: Hypertext Preprocessor
- CGI – Common Gateway Interface
- Java Servlets, JSP

Warstwa prezentacji

- HTML (HyperText Markup Language) – treść
- CSS (Cascading Style Sheets) – wygląd

Komunikacja zwrotna

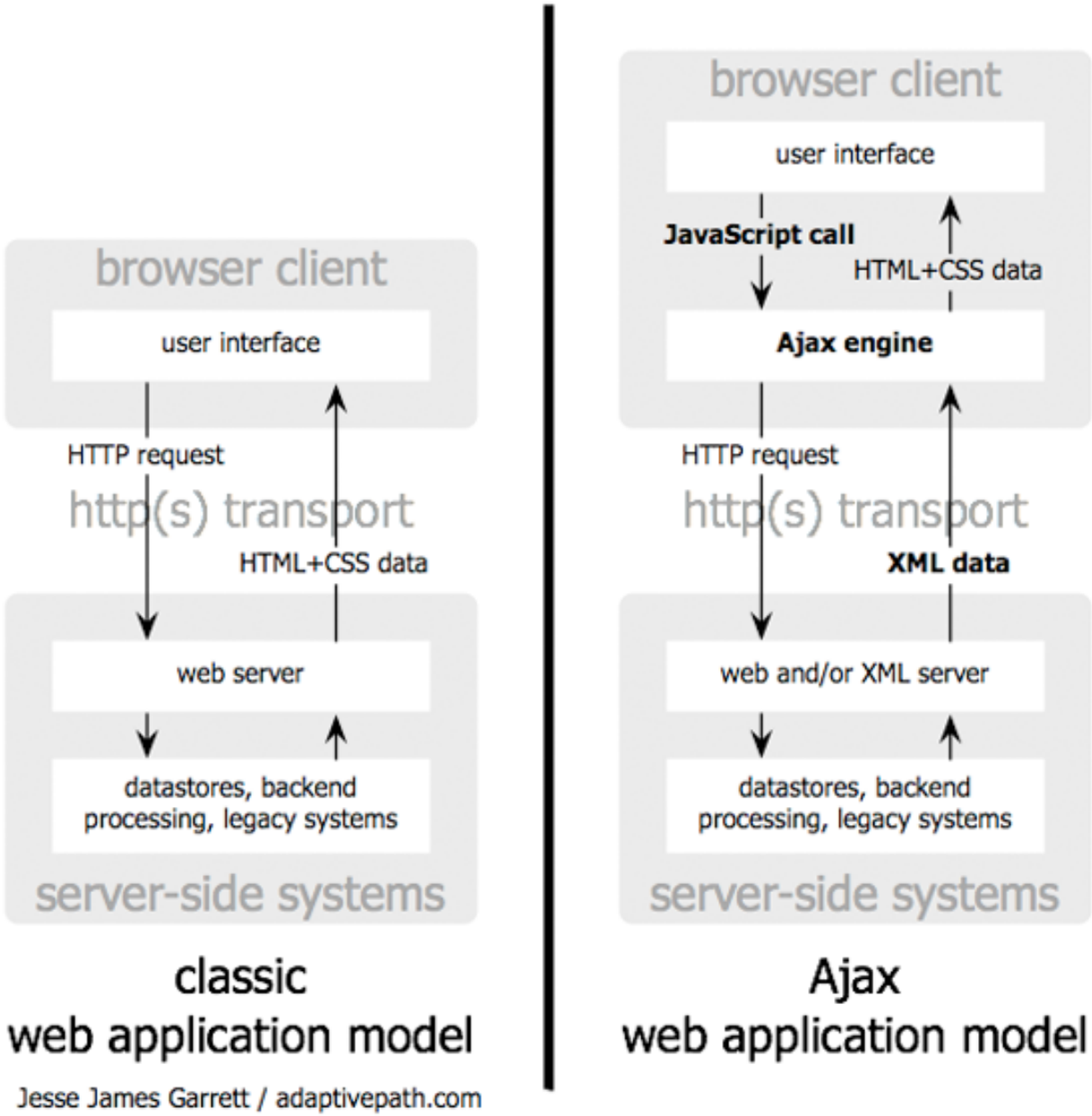
Metody GET i POST protokołu HTTP.

Aplikacje WWW – AJAX

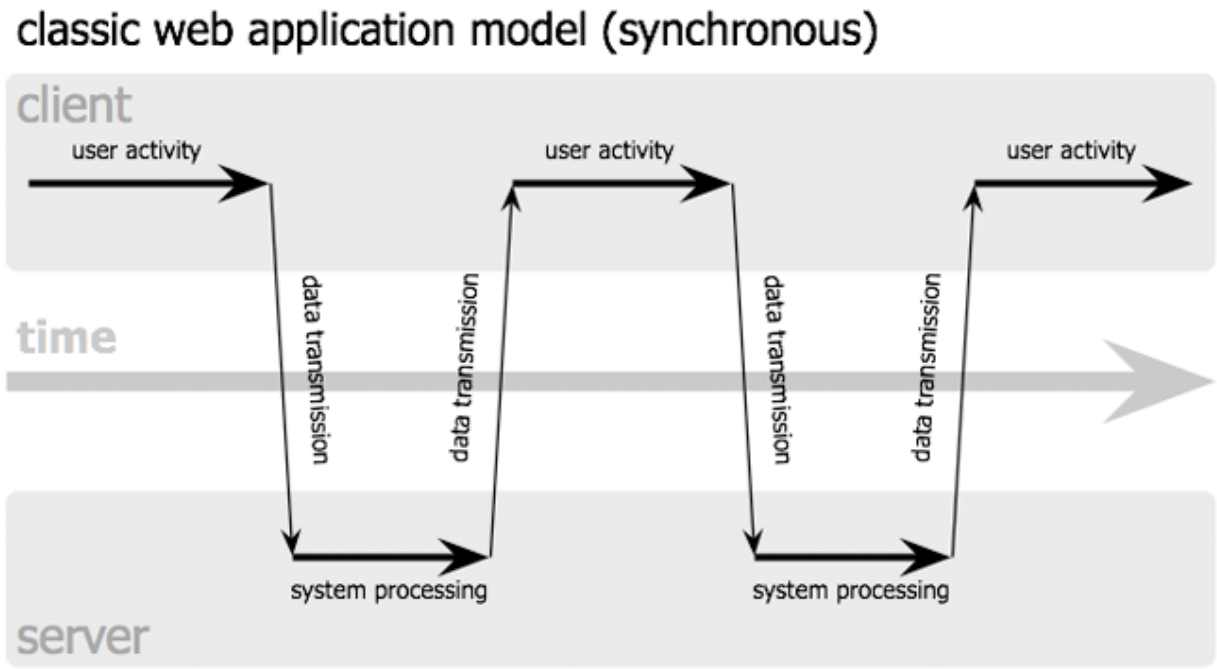
Asynchronous **J**avaScript **a**nd **X**ML

- prezentacja – XHTML i CSS
- dynamiczne modyfikacje – Document Object Model
- wymiana i manipulacja danymi – XML and XSLT
- komunikacja asynchroniczna – XMLHttpRequest
- JavaScript

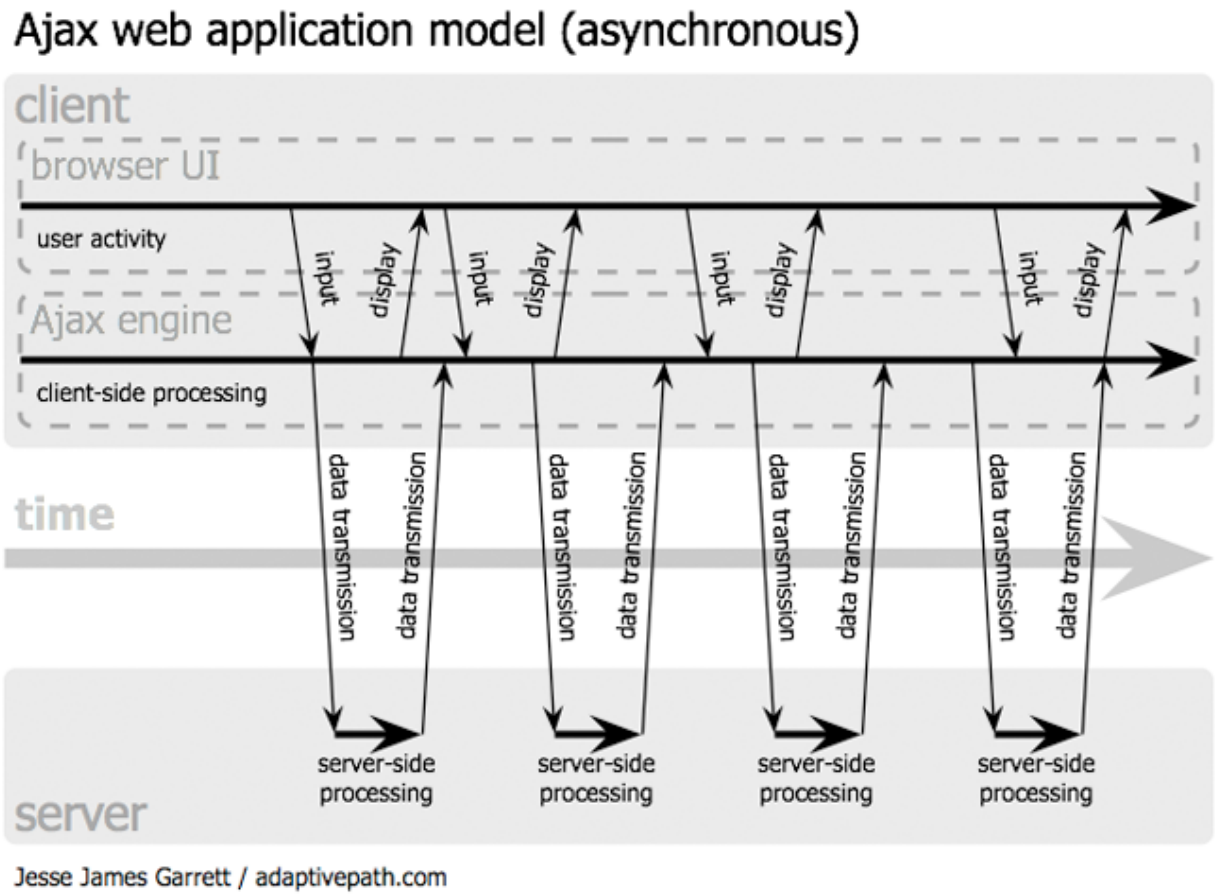
AJAX c.d.


browser client
user interface
HTTP request
http(s) transport
HTML+CSS data
web server
datastores, backend
processing, legacy systems
server-side systems
classic
web application model
Jesse James Garrett / adaptivepath.com
browser client
user interface
JavaScript call
HTML+CSS data
Ajax engine
HTTP request
http(s) transport
XML data
web and/or XML server
datastores, backend
processing, legacy systems
server-side systems
Ajax
web application model

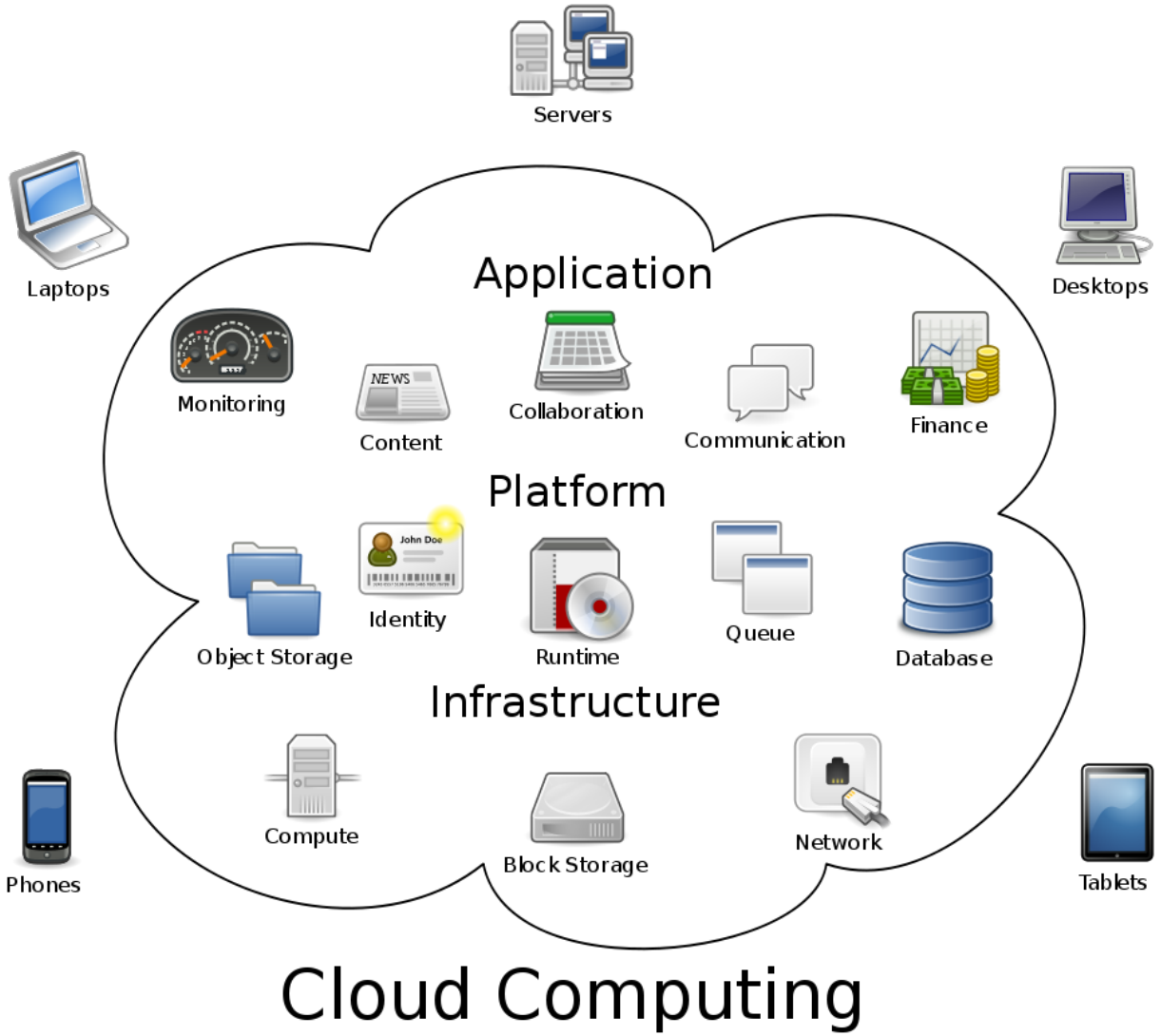
AJAX c.d.


classic web application model (synchronous)
client
user activity
user activity
user activity
data transmission
data transmission
data transmission
data transmission
time
system processing
system processing
server

AJAX c.d.


Ajax web application model (asynchronous)
client
browser UI
user activity
input
display
Ajax engine
client-side processing
data transmission
time
server-side processing
server
Jesse James Garrett / adaptivepath.com

Cloud computing



Cloud computing – c.d.

Warstwy

- Klient
- Usługi (SaaS – Software as a Service) – np. Google Apps, iCloud
- Platformy (PaaS – Platform as a Service) – np. Google App Engine, serwisy hostingowe
- Infrastruktura (IaaS – Infrastructure as a Service) – np. Amazon EC2
- Serwery

Cloud computing – wady i zalety

Zalety

- skalowalność
- niezależność od miejsca
- niższy koszt (efekt skali, inna kwalifikacja kosztu)
- łatwość konserwacji
- bezpieczeństwo
- daje większe możliwości użytkownikom końcowym

Cloud computing – wady i zalety

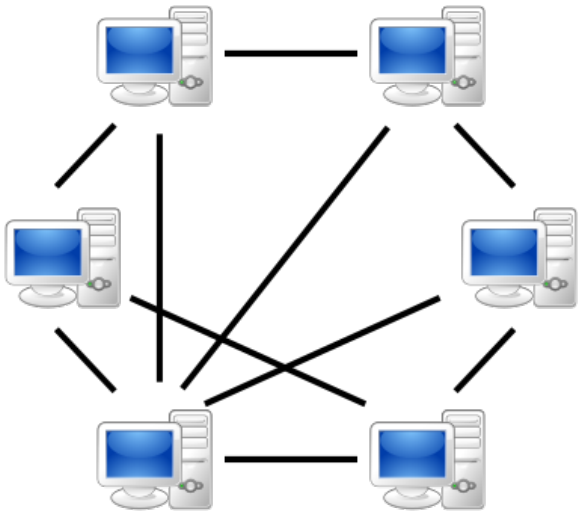
Zalety

- skalowalność
- niezależność od miejsca
- niższy koszt (efekt skali, inna kwalifikacja kosztu)
- łatwość konserwacji
- bezpieczeństwo
- daje większe możliwości użytkownikom końcowym

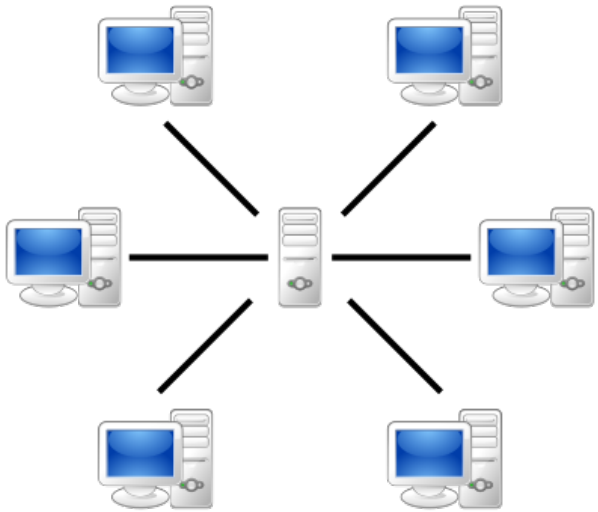
Wady

- brak kontroli nad danymi (data escrow)
- mniejsza możliwość dostosowania do szczególnych wymagań
- wielość standardów
- problemy prawne
- bezpieczeństwo

Peer to Peer (P2P)



vs.



Peer to Peer (P2P) – c.d.

P2P

- podejście czyste – wszystkie węzły są jednakowe
- podejście hybrydowe – superwęzły
- podejście scentralizowane

Peer to Peer (P2P) – c.d.

P2P

- podejście czyste – wszystkie węzły są jednakowe
- podejście hybrydowe – superwęzły
- podejście scentralizowane

Zalety

- lepsze wykorzystanie zasobów (przepustowości)
- łatwość konfiguracji
- niezawodność

Peer to Peer (P2P) – c.d.

P2P

- podejście czyste – wszystkie węzły są jednakowe
- podejście hybrydowe – superwęzły
- podejście scentralizowane

Zalety

- lepsze wykorzystanie zasobów (przepustowości)
- łatwość konfiguracji
- niezawodność

Wady

- niebezpieczeństwo "wstrzyknięcia" błędnych treści
- brak kontroli nad zawartością
- problemy prawne (cenzura internetu)
- blokowanie przez ISP (net neutrality?)