

**CASCON** [2009]



# Deploying MediaWiki on IBM DB2 in the Cloud

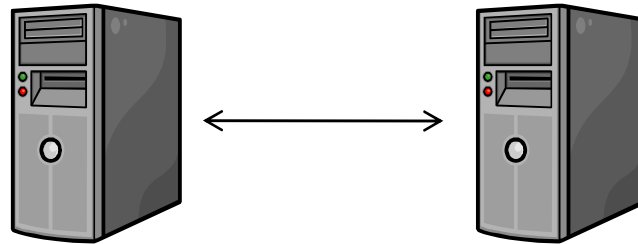
Leons Petrazickis  
IBM Toronto Laboratory  
leonsp@ca.ibm.com



# What We Want

Rightscale

Amazon EC2



MediaWiki server

DB2 server

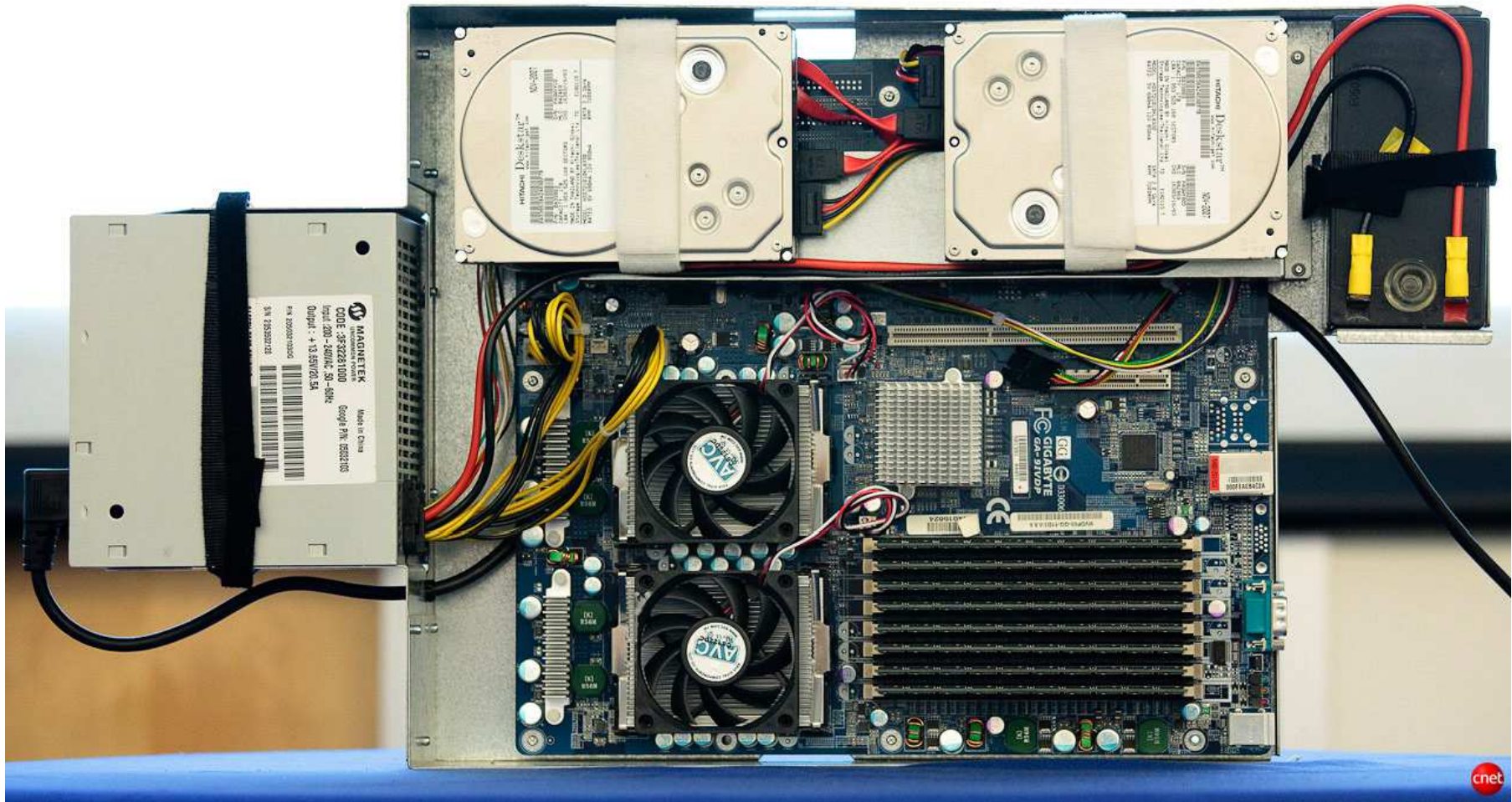
# What is Cloud Computing?

- Illusion of infinite resources available on demand
  - Infrastructure – Infinite servers
    - e.g. Amazon EC2
  - Platform – Infinite computing capacity
    - e.g. Google AppEngine
  - Software – Infinite scalability
    - e.g. SalesForce.com
-

# Traditional Server



# Google Server



# Google Rack

## **One server**

DRAM: 16GB, 100ns, 20GB/s

Disk: 2TB, 10ms, 200MB/s

## **Local rack (80 servers)**

DRAM: 1TB, 300us, 100MB/s

Disk: 160TB, 11ms, 100MB/s

## **Cluster (30+ racks)**

DRAM: 30TB, 500us, 10MB/s

Disk: 4.80PB, 12ms, 10MB/s



# Google Data Center

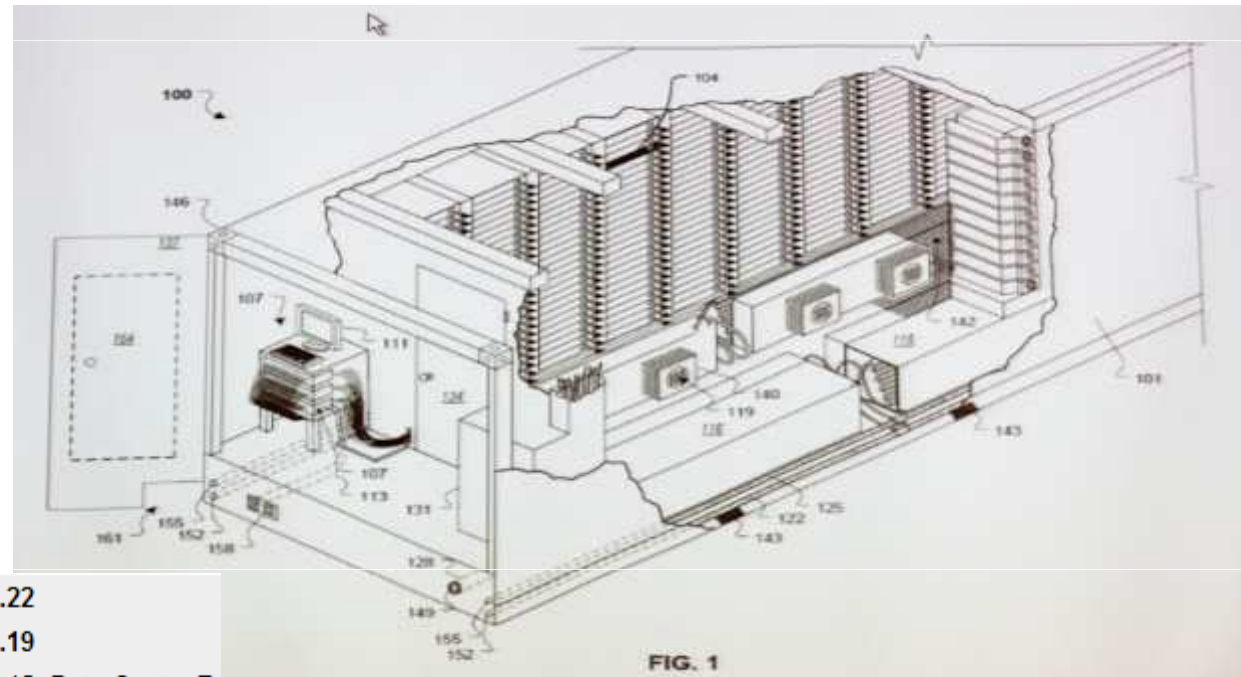
- Containerized data center since 2005
- 1160 servers per standard shipping container

$$\text{PUE} = \frac{\text{Total Facility Power}}{\text{IT Equipment Power}}$$

Scenario	PUE
Current Trends	1.9
Improved Operations	1.7
Best Practices	1.3
State-of-the-Art	1.2

Figure 1: EPA Estimated PUE Values in 2011 <sup>5</sup>

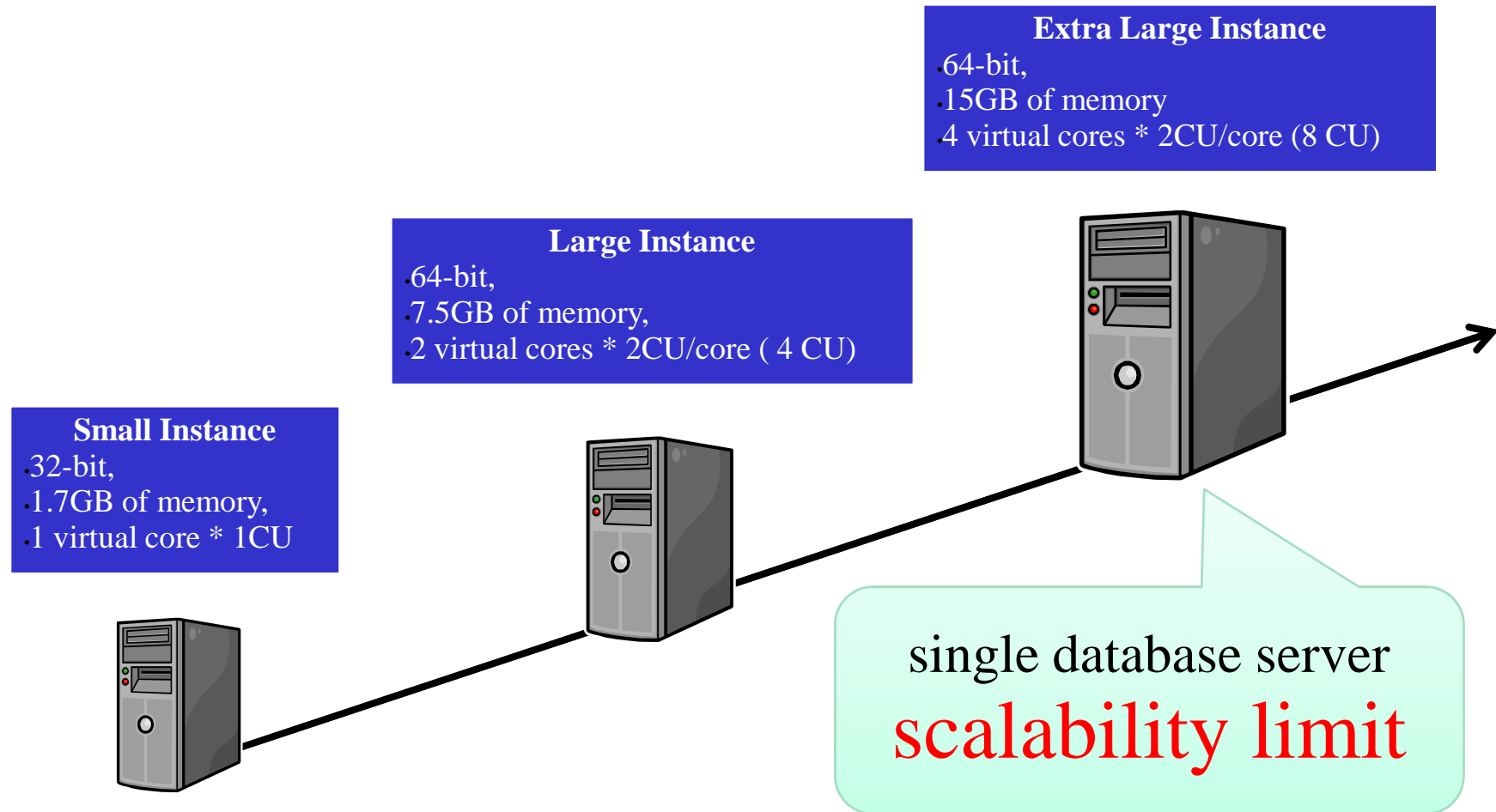
Quarterly energy-weighted average PUE:	<b>1.22</b>
Trailing twelve-month energy-weighted avg. PUE:	<b>1.19</b>
Individual facility minimum quarterly PUE:	1.15, Data Center B
Individual facility minimum TTM PUE*:	1.14, Data Center B
Individual facility maximum quarterly PUE:	1.33, Data Center H
Individual facility maximum TTM PUE*:	1.21, Data Center A



# Business Case for Cloud

- Cheaper than traditional approaches
    - No initial capital expense
    - Only pay for what you use
  - Best practices without the learning curve
  - Instant scalability
    - Bring up 10 more web servers because everyone linked to you at once
    - Bring up 1000 test servers to test a networking app, shut them down right after
-

# Vertical Scalability on Amazon EC2



# Good and Bad Uses

## Good

- Proof of Concept
- **Development and Test**
- Public facing web presence
- Departmental systems
- Applications with highly variable workloads
- Seasonal applications
- Disaster recovery and backups

## Bad

- Highly transactional workloads
- Data with high security/privacy demands
- Applications with complex regulatory requirements



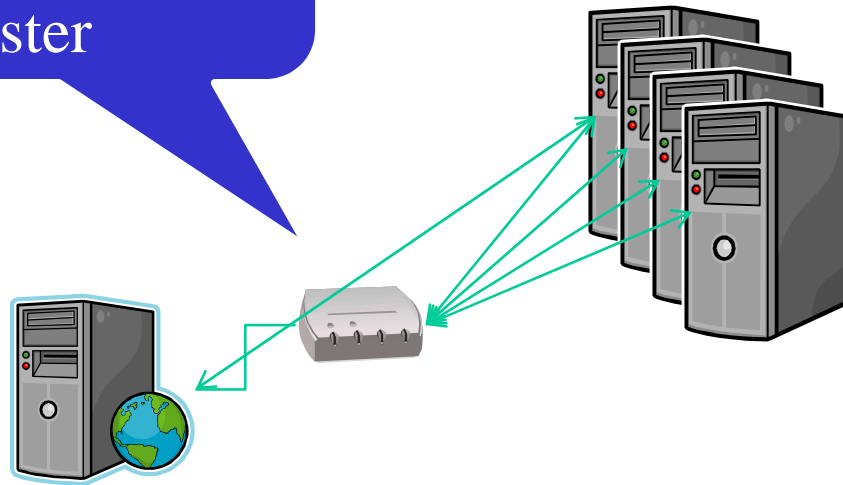
# What is DB2?

- IBM DB2 V9.7 database server
- Free DB2 Express-C edition
  - <http://ibm.com/db2/express/>
- Transactional, ACID-compliant
- Stored procedures in both SQL and Java
- pureXML, Full Text Search, etc

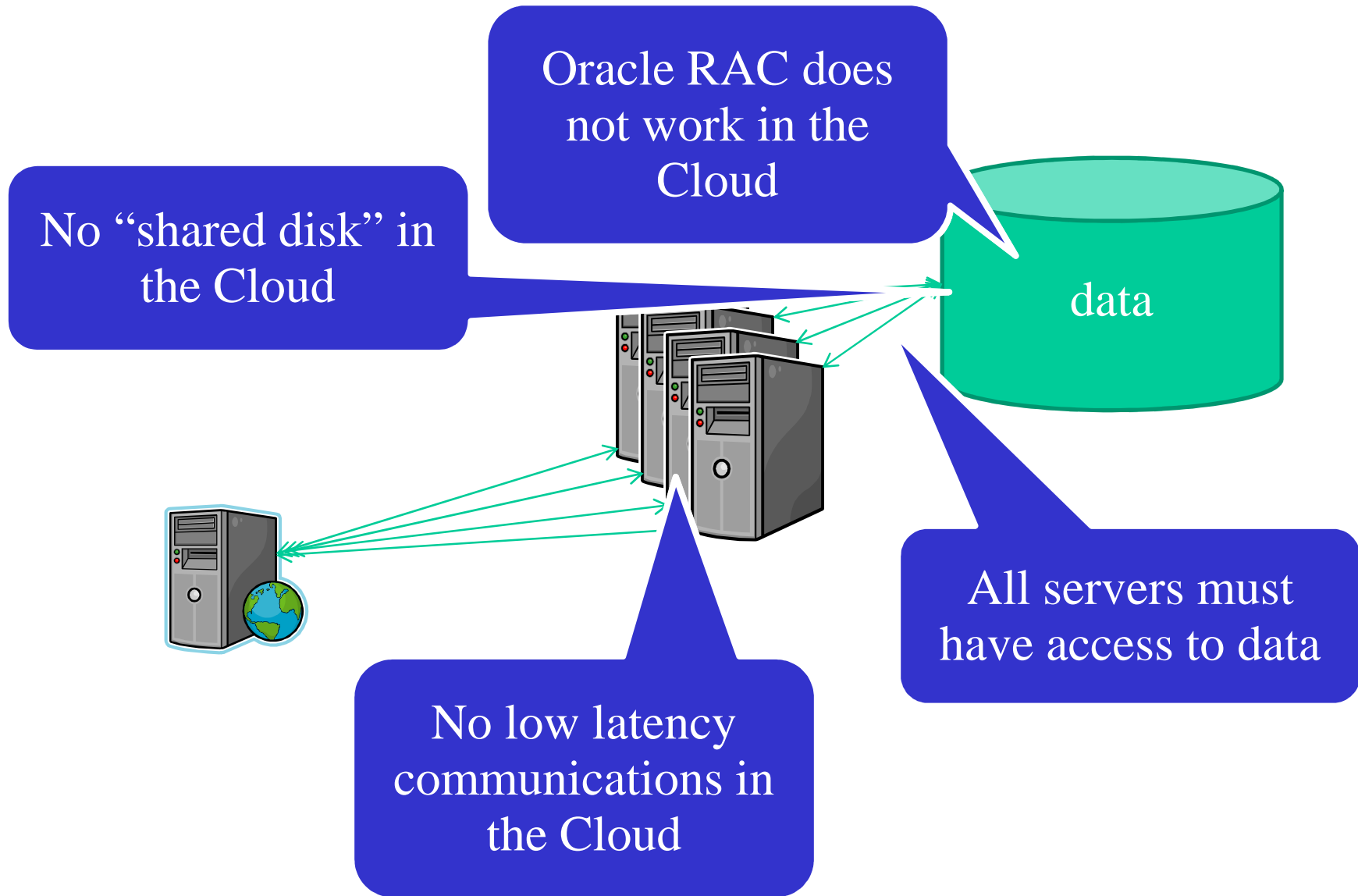


# Database Architecture in the Cloud

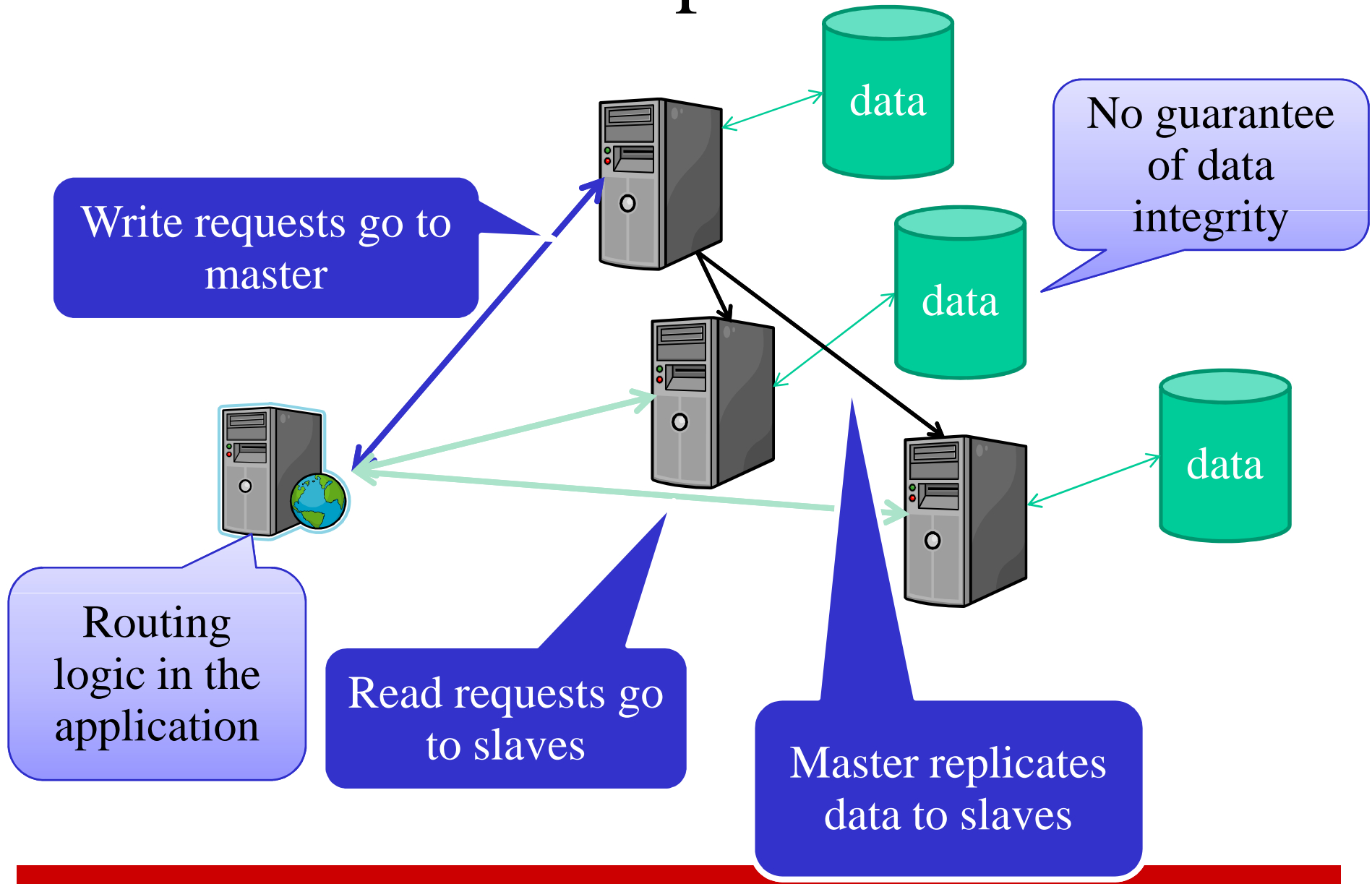
load balancer to  
create a single  
system image for the  
cluster



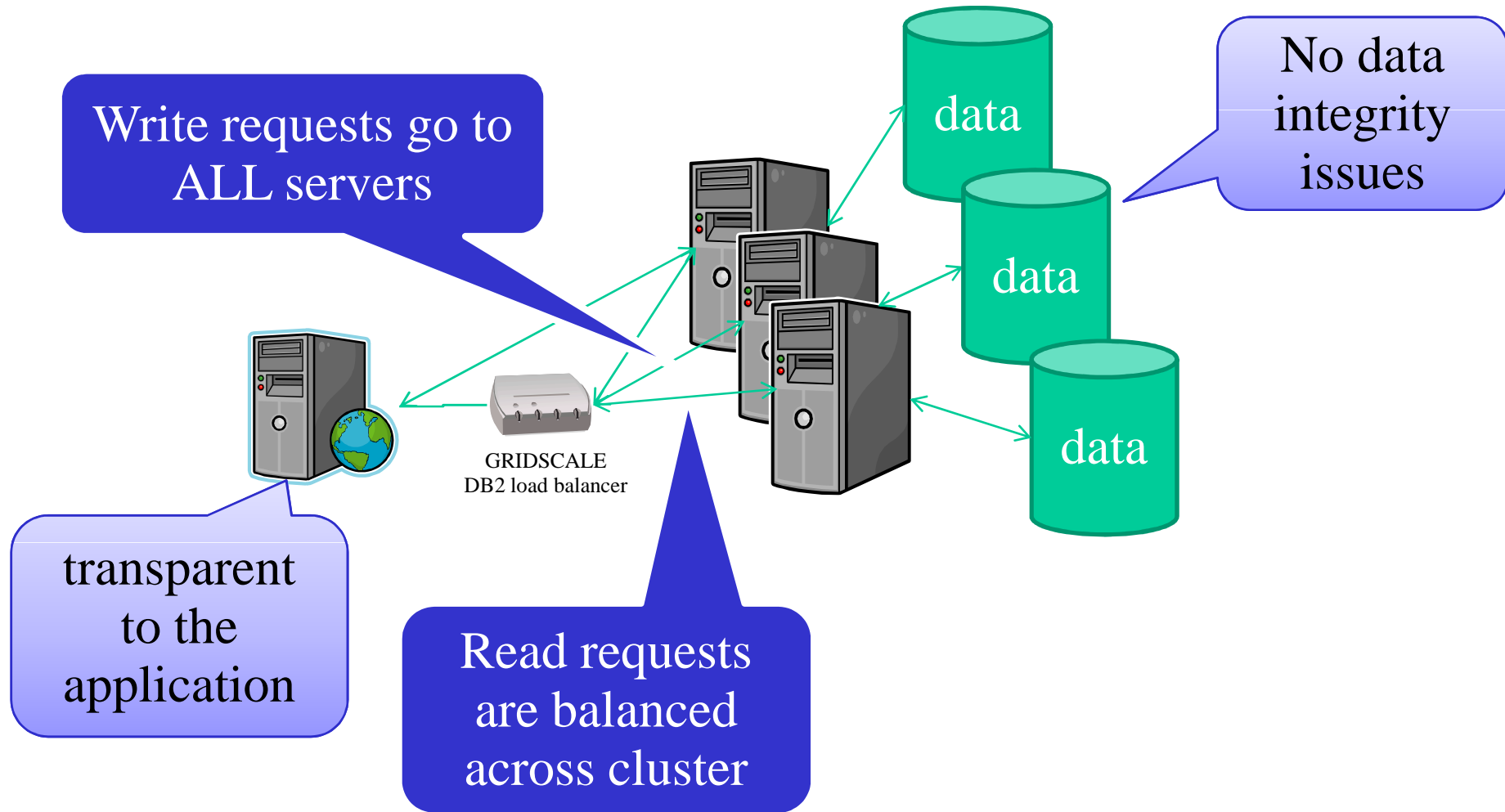
# No Oracle RAC in Cloud



# Master/Slave Replication in Cloud



# DB2 and GRIDSCALE Load Balancing in Cloud



# DB2 Data Partitioning Feature in the Cloud

- Database Sharding:
    - Shared Nothing database architecture
    - Data is partitioned across a number of DB2 nodes with each node managing only its own portion of data
  - Parallel Processing: DB2 DPF splits complex queries and parallelizes their execution across DB2 nodes in a DPF cluster. Results are recombined and returned to the calling application
-

# DB2 Data Partitioning Feature in the Cloud

- Shared Nothing architecture fits well in to the Cloud provisioning model
  - Capacity can be scaled up/down by adding/removing independent nodes to the cluster
  - Execution of complex queries automatically parallelized across a cluster
  - Partitioning and parallelization are transparent to an application
  - Designed to process complex queries over very large data sets
-

# DB2 High Availability Disaster Recovery in Cloud

Application sends a read



GRIDSCALE Servers



Database

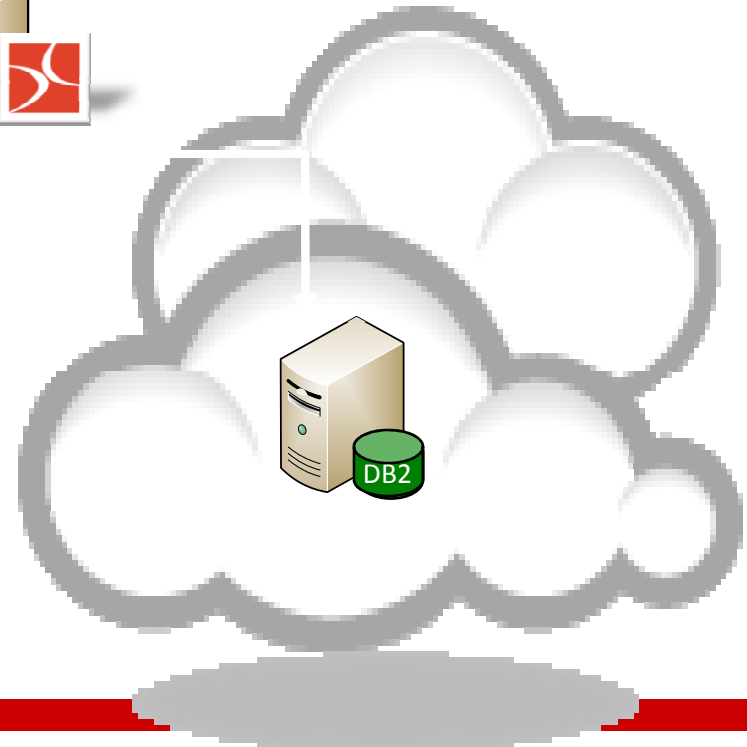


# DB2 High Availability Disaster Recovery in Cloud

Database server fails; read is automatically re-routed

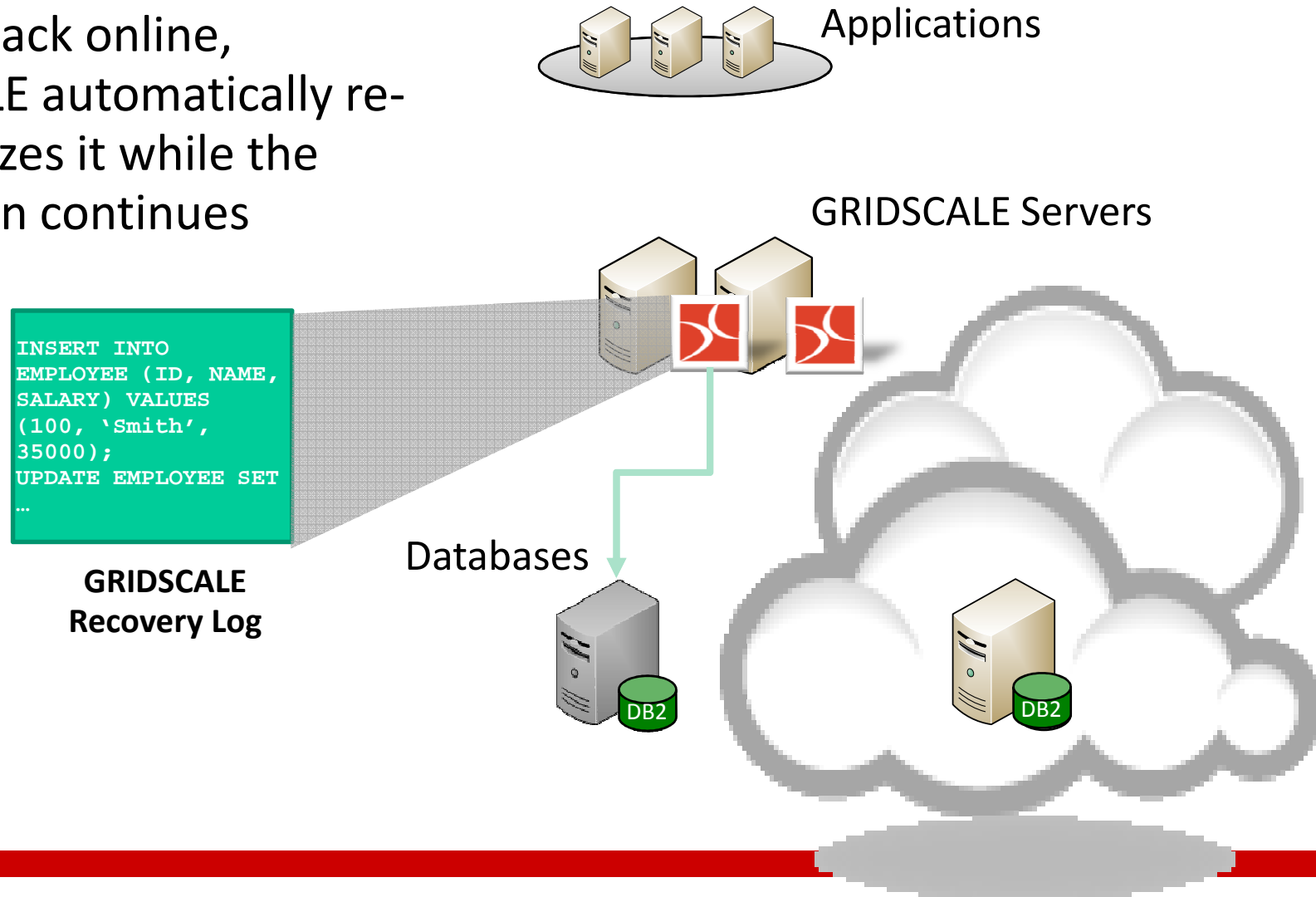


Database



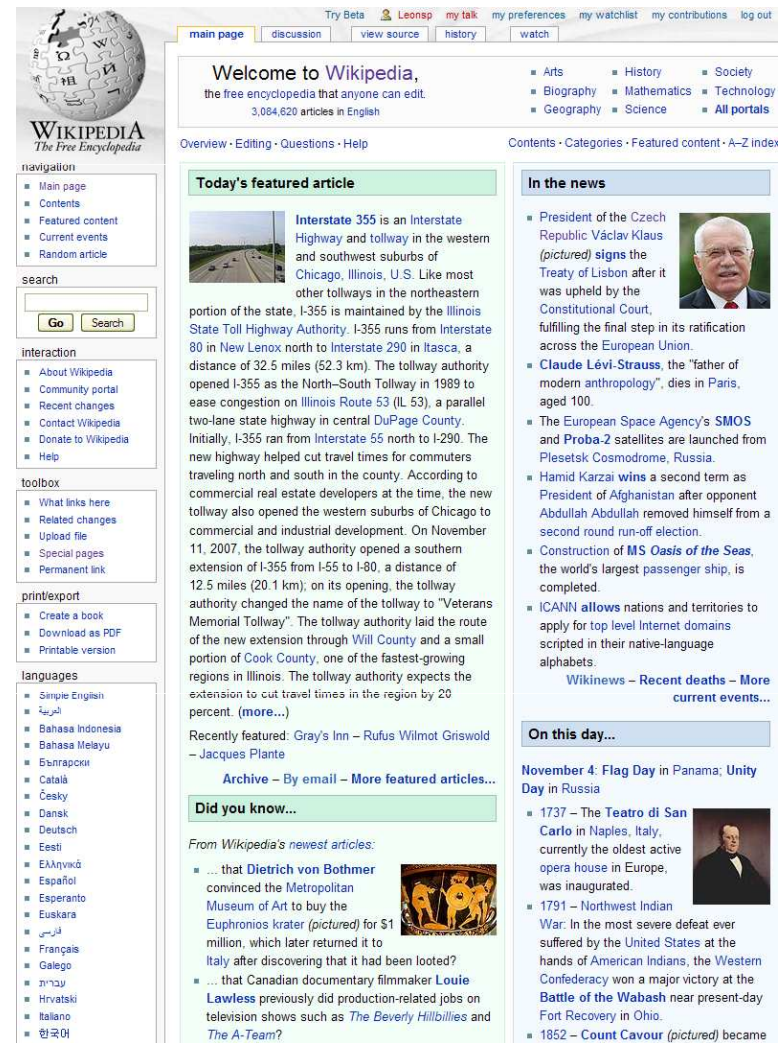
# DB2 High Availability Disaster Recovery in Cloud

When the database server is brought back online, GRIDSCALE automatically re-synchronizes it while the application continues



# What is MediaWiki?

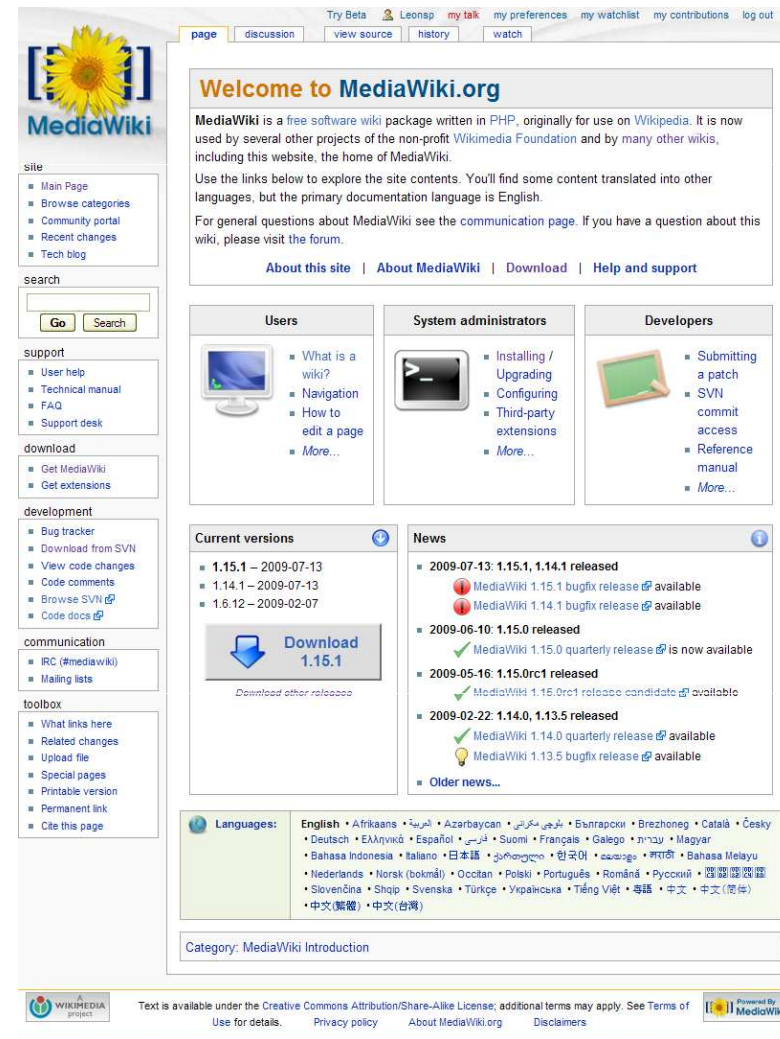
- Powers Wikipedia
  - since 2002
- Countless other websites:
  - Ubuntu Help
  - Folding@Home
  - FileZilla Documentation
  - Mozilla Developer Center
  - Super Mario Wiki
  - etc.



The image shows a screenshot of the English Wikipedia homepage. At the top, it says "Welcome to Wikipedia, the free encyclopedia that anyone can edit." and "3,084,820 articles in English". The page is divided into several sections: "Today's featured article" (Interstate 355), "In the news" (President of the Czech Republic Václav Klaus signs the Treaty of Lisbon), "Recently featured" (Gray's Inn - Rufus Wilmot Griswold), "Did you know..." (Dietrich von Bothmer), and "From Wikipedia's newest articles" (Louie Lawless). The page also includes a search box, navigation links, and a list of languages.

# What is MediaWiki?

- Written in PHP
- Traditionally runs on MySQL
- Ported to IBM DB2 among others
- Open source – community contributions welcome



The screenshot shows the MediaWiki.org homepage. At the top, there's a navigation bar with links like 'page', 'discussion', 'view source', 'history', and 'watch'. Below this is a 'Welcome to MediaWiki.org' banner. The main content area is divided into several sections: 'Users', 'System administrators', and 'Developers', each with a list of links. There's also a 'Current versions' section with a 'Download 1.15.1' button, and a 'News' section with a list of recent releases. At the bottom, there's a 'Languages' section with a list of supported languages and a footer with the Creative Commons license information.

# Why database abstraction?

- MySQL

  - `mysql_query(); mysql_affected_rows();`

- SQL Server

  - `mssql_query(); mssql_rows_affected();`

- IBM DB2

  - `db2_exec(); db2_num_rows();`

- Incompatible PHP APIs for every database

  - Everything a function in global namespace



# Why database abstraction?

- MySQL

```
SELECT * FROM table LIMIT 10;
```

- SQL Server

```
SELECT TOP 10 * FROM table;
```

- IBM DB2

```
SELECT * FROM table  
FETCH FIRST 10 ROWS ONLY;
```

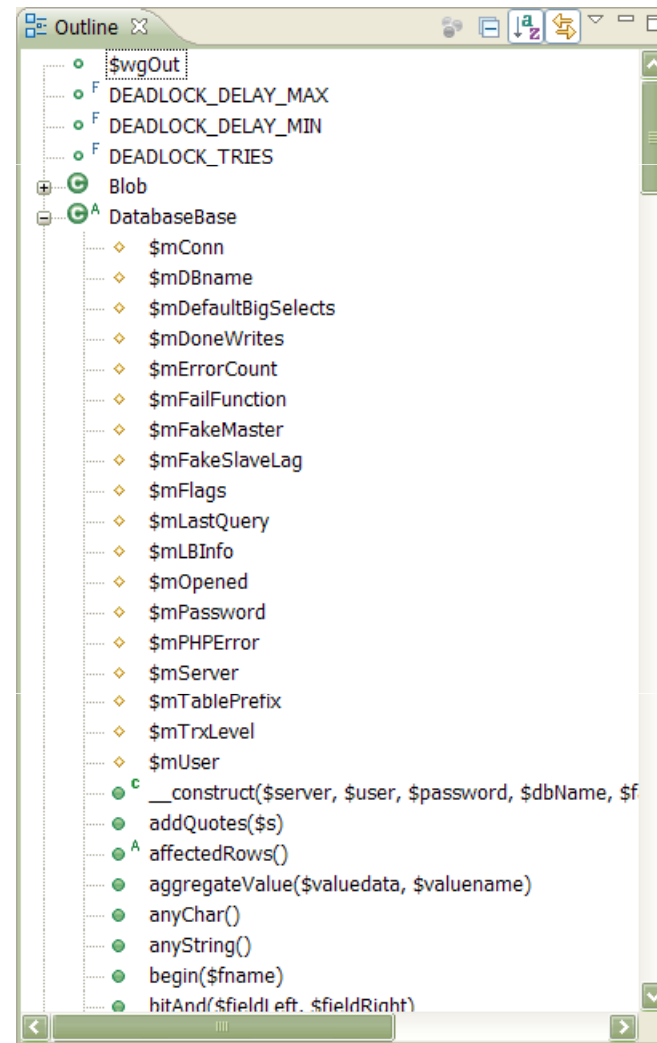
- Incompatible SQL syntax

- No single standard that just works

---

# MediaWiki Solution

- One abstract class that defines the database API
  - DatabaseBase
- Child classes that handle specific databases
  - DatabaseMySQL
  - DatabaseIBM\_DB2
  - etc.



# Resources

- Leons Petrazickis
    - <http://lpetr.org/blog/>
    - [leonsp@ca.ibm.com](mailto:leonsp@ca.ibm.com)
  - DB2 in the Cloud
    - [db2cloud@ca.ibm.com](mailto:db2cloud@ca.ibm.com)
  - Amazon Web Services
    - <http://aws.amazon.com/>
  - RightScale
    - <http://www.rightscale.com/>
  - IBM Cloud
    - <http://www-949.ibm.com/cloud/developer/>
-