

# Dolphins vs Elephants

A comparison of MySQL and PostgreSQL

# Major Differences

## MySQL

- Different plugin storage engines
- Single method of replication
- “More” commercial, controlled by single entity
- Minimal full text searching abilities, slow in comparison
- No GIS support
- Smaller amount of data types, no ability to create your own
- Began focused on speed, now adding features/stability
- Loosely follows SQL standard

## PostgreSQL

- Single all purpose storage engine
- Different plugin replication engines
- BSD licensed and core development group from several different companies
- Powerful full text searching
- Robust GIS and Geospatial support
- Rich data types and ability to create your own easily
- Began focused on features and stability, last several years have been devoted more to performance
- Only deviates from SQL standard when necessary

# Data Type Differences

## MySQL

- Normal Integer, Floating point, and arbitrary precision types
- Normal date/time support, without timezone support
- Typical string, text, and blob data types. Largest text type limited by size of server/client message buffer
- Ummm... that's it...

## PostgreSQL

- Same, with the addition of a money type which automatically formats input/output based on locale.
- Supports timezones
- Typical string, text, and blob data types. However, storage is only limited by available disk space.
- Also has built in support for Points, Lines, Segements, Boxes, Paths, Polygons, Circles, Ipv4/IPv6 addresses, CIDR, bit strings, UUIDs, XML, Arrays, and defining your own composite types

# Performance Differences

## MySQL

- MyISAM is very fast, but not reliable. Basically a step up from using SQLite. Need to use InnoDB for safety, which is much slower.
- Performance levels off at 5 simultaneous users and begins to drop after 20-30
- Very few tunable parameters
- Lacks many performance niceties such as partial indexes and tablespaces

## PostgreSQL

- No decision needed about storage engine. In between MyISM and InnoDB on performance in general.
- Performance actually increases the more simultaneous users there are in most cases. Performance doesn't level off until 75+ users
- Ability to tune database server in a variety of ways
- Scales vertically better than MySQL

# People using PostgreSQL

- Yahoo
- Skype
- FeedLounge
- .info and .org TLD databases
- EveryBlock.com
- LJW and Sunflower

# In the Real World...

When FeedLounge switched their site from MySQL to PostgreSQL they saw

- MySQL -> InnoDB increased DB size by 10x
- At time of move MySQL on disk was 34GBs, after move to PostgreSQL it was 9.6GBs
- Queries ran 3x faster and DB system used 1/3<sup>rd</sup> the memory
- Effectively a 10x improvement in performance

# In the Real World...

Yahoo uses PostgreSQL for their web analytics application

- Wrote their own columnar storage engine, with which they scaled to over 1 Petabyte. Estimates they will be at 10 Petabytes by end of 2009
- 24 billion events per day
- When asked why they didn't go with a commercial DB, response was “They don't scale well beyond 100 Terabytes.”

Questions?