How to distribute databases

HI1030 Database Technology Short Paper, Period 2, Fall 2016

Johan Lipecki, [lipecki@kth.se](mailto:lipecki@kth.se)

Viggo Lundén, [vlunden@kth.se](mailto:vlunden@kth.se)

December 2016

# Abstract

Sammanfattningen i en kortare rapport kan med fördel formuleras i fyra meningar:

* Första meningen presenterar problemet
* Andra meningen klargör varför detta är ett problem
* Tredje meningen presenterar resultatet
* Fjärde meningen presenterar slutsatsen

• Glöm inte att nyckelorden ska ingå i Abstract

Keywords

Ange ett antal nyckelord

# Introduction

The corners of the world are closer than ever, and so is its shared data. If you want to know the weather at Cape Horn, there will be no problem finding out almost immediately. And that is of course not because you are the only one querying one of Chile’s southernmost databases, but because the information is stored closer to you than the 14 800 km between you and Tierra del Fuego. It is because data is stored in databases all around the globe, and because those databases are part of a distributed system. Below, we will cover the basics in the setup of distributed databases, look at what considerations need to be taken by the database administrator and juxtapose some aspects of distributed relational and distributed non-relational databases.

# Distributing a Database

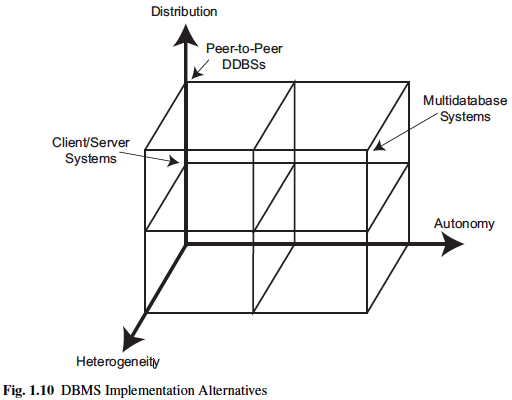
## Reasons for distribution

Distributing a database may be part of the optimization of a growing organization, the starting point of an information-heavy business model or perhaps a business startup’s approach to future scalability. Whatever the outset may be, there are many benefits of distributing a database – and for a global organization even more so.

## Local processing of local queries

One of the optimization reasons of distributing a database concerns the geographical displacement of users or applications. To query a database some physical distance away naturally leads to latency and for a dispersed organization that latency is exacerbated by every user attempting to connect to the same database. Further, the information needs of one office may not be the same as another and so distributing the database on the grounds of distance and disparate needs leads to at least three benefits: speedy retrieval, possible table tailoring and reduction in database payload.

## The Database Distribution System

* + 1. Architecture of the database[1]

## Database transparency

The ideal is that the user is unaware of an underlying database structure and only interacts with the views of interest – without having to specify any locations for its parts. If this objective of unawareness is met, the database is completely transparent. However, this is no simple accomplishment. But how do we know if a database is transparent? Let us answer that by restating the question: *do the users have to know that the database is distributed?* If so, the database is not completely transparent and, within the domain of database design, transparency refers to two considerations of distribution: *fragmentation and replication*.

## Fragmentation

The fragmentation of tables refers to the separation of tables into subsets. Those subsets may be divided either vertically, horizontally or both and need to rely on statistical data to supply the database with the required improvements.

## Vertical Fragmentation

## Horizontal Fragmentation

## Matching Fragments and Statistics

## Replication

# Database Distribution Steps

We do not live in Utopia, but let us pretend for a moment that we know exactly everything about the needs of our organization and that we can have an efficient and completely transparent database design – what would the setup process be?

# Analysis

# Conclusions

# References

[1] Ponniah, Paulraj. "Distributed Database Systems." *Database Design and Development*. Hoboken, NJ, USA: John Wiley & Sons, 2005. 551-97. Web.

[2] Özsu, M. Tamer., and Patrick. Valduriez. *Principles of Distributed Database Systems*. 3rd ed. New York: Springer New York, 2011. Studies in Mechanobiology, Tissue Engineering and Biomaterials. Web.