Location Matters: Limitaions of Global-Scale Datacenters

Yahel Ben-David, Shaddi Hasan, Paul Pearce

Summary

Cloud Computing becomes more and more popular. This technology enables cost reduction by reducing service costs through big datacenters, bulk purchases of needed hardware or electrical power, large-scale cooling and centralized management. Cloud providers tend to build big, centralized datacenters, but these systems have some deficiencies. Disaster tolerance, response time and regulatory constraints are the key issues.
Because of the advantages, the authors mention that building large datacenters, which are still profitable, are always the option cloud operators would choose. So nano datacenters are pushed to the background. Another promising idea is using telecommunication carriers. These are ideally positioned and can offer end-to-end quality guarantees.

Why not Nano Datacenters?

Volume pricing can enable the possibility to design cost effective specialized hardware, which makes a difference in saving power or cooling. But these options are only for big datacenter operators.
In addition, the cooling and power infrastructure is more efficient at big scales and bulk buyer can often get lower per-unit rates or incentives.
Another reason against nano datacenters is the fact that developing applications for one big datacenter is much easier and less failure-prone as developing for more than one datacenter. Furthermore, communication between multiple datacenters is slower and more costly.
To provide satisfactory results in an appropriate time, massive computing systems and ever-increasing storage is needed. Big datacenters allow this.

Cloud response times

The authors tested the cloud response time with four of Amazon's public cloud datacenters. They used HTTP response time and the lowest response times occur, when the datacenters and PlanetLab nodes (used for testing) are geographically close together. Response times can be also affected by routing failures, suboptimal routing or slow Border Gateway Protocols. The tests with single datacenters showed that a single datacenter cannot provide low response times to multiple end points. So more datacenters were added and the results are further reduced times.

Legal Concerns and Constraints

The "number one killer" for using public cloud computing is the geographic location (jurisdiction) of data.
Privacy laws govern how data is stored and who may access that data. There are many different laws. Transborder data laws govern where data can be stored. Moving data between different jurisdictions is called transborder data flow and can be difficult or illegal. This depends on content type or country of origin. This intimidated a cloud provider so that he builds redundant datacenters in each needed jurisdiction. Some jurisdictions allow transborder data crossing, when the other jurisdiction provides at least equivalent levels of protection. The amount and variety of these laws hamper the extension of cloud computing.

The European Union Data Protection Directive (EUDPD)

Holding personal data from a EU citizen in another country than the EU is forbidden by the EUDPD, except they can provide appropriate protection. Switzerland, Canada, Argentina and the Isle of Man can provide the required protection. This is a very limited area, but in special cases data can also be stored in the United States. In this case the United States Safe Harbor rules play an important role.

US Safe Harbor Frameworks

With this law companies in the US can process or store EU data. If a company wants to be certified it has to apply and agree to the requirements. But there are also limitations like the fact that the company has to be under the jurisdiction of the Federal Trade Commission. The inconsequent legislative facts and the complicated laws hamper the idea of global datacenters.

Conflicting Legislation: The USA PATRIOT ACT

Obtaining personal data of foreign companies and persons is allowed by this Act, when the data is stored in the US. Therefore, EU companies, which store data in the US, should reckon that the US may read their data.
If an US company stores data in the EU, then the Act gives the US also the right to collect private data.
This leads to mistrust, insecurity and does not support the idea of a global-scale datacenter but rather to distributed datacenters.

Impact on Application Innovation

Flexibility over capacity is provided by current cloud systems, but in future there maybe could be provided flexibility over location. This could be an advantage for software-defined networking and the outsourcing of network management. They can move the management to central places and cities and benefit from the skilled network administrators there.

Discussion

The actual cloud model “one datacenter fits all” is not right. Based on this system, some smaller businesses change their storing systems and use clouds but big companies still hesitate.
The findings show that telecom carriers’ locations are ideal for end-to-end-connections and connect many users. The low distance to these users reduces response times and the spread ensures that carriers are in many jurisdictions. One disadvantage is that this system may not be as efficient as the current big datacenters, but latency sensitive applications can benefit from the cloud system and also the network management outsourcing.

Related work (Eher Aufzählung von Autoren, die ähnliches gemacht haben)