Discussing “The Art of Scalability” Article

Thinhinane Ihadadene & Maya Kassis
Master 2 CCN
2019 - 2020
Chapter 22. Introduction to the AKF Scale Cube

Chapter 23. Splitting Applications for Scale

Book: The Art Of Scalability : Scalable Web Architecture, Processes, and Organizations for the Modern Enterprise

Authors: Martin L. Abbott and Michael T. Fisher
01
INTRODUCTION TO
SCALABILITY
Scalability, Scaling

**Scalability is**

The ability of a system (business or technology) to handle increased requests without altering response time.

**Scaling is**

The increase or decrease in the capacity of a system by changing the number of allocated processes.
Scalability, Scaling

Scale up or vertical scaling

8 CPU, 16 GB RAM
4 CPU, 8 GB RAM
2 CPU, 4 GB RAM

Scale out or horizontal scaling

2 CPU, 4 GB RAM
2 * (2 CPU, 4 GB RAM)
4 * (2 CPU, 4 GB RAM)

Scaling up vs Scaling out, adapted from: https://dev.to/wmahathre/horizontal-and-vertical-scaling-1lid, 9/20/2019
02
The AKF Scale Cube
The AKF Scale cube is a model for defining microservices and scaling technology products, invented in 2007 by AKF partners consulting company.

This cube has a 3-dimensional axis represent a unique method of scaling products, processes, and teams.
AKF scale cube

X-axis
Represents the cloning of services or data such that work can easily be distributed across instances with absolutely no bias.

Y-axis
Represents the separation of work responsibility for an action, also known as service or resource oriented split.

Z-axis
Represents the separation of work by customers or requestors.

Monolithic System

Infinite Scale

(0,0,0)
03
Splitting Applications For Scale
Scaling Applications: X-AXIS
Cloning/Replications

- Clone services or duplicate databases
- Use a load balancer

Pros:
- Easy and fast to implement
- Perfect for read only databases

Cons:
- Cost of replication
- Doesn't address caching
Scaling Applications: Y - AXIS
Split Different Things

- Separates into services
- Separates into resources

Pros:
- Increases hit cache
- Fault isolation
- More secured

Cons:
- Difficult and time consuming
Scaling Applications: Z-AXIS
Split Similar Things

- Separates by region, customers
- Separates into resources

Pros:
- Increases hit cache
- Fault isolation
- More secured

Cons:
- Difficult and time consuming
Is it sufficient to scale using one axis?

- Coupled with the needs of an organization
- They might be all used and mixed

AKF Scale Cube - [https://akfpartners.com/growth-blog/scale-cube](https://akfpartners.com/growth-blog/scale-cube), 9/21/2019
04
Practical Use of AKF Scale Cube
Airlines reservation system developed by IBM, consists of the below:

1- Original/Destination duals and inventories.

2- Global distributed system GDS, connect providers with travel agencies.

3- Real time dynamic pricing system, based on varied number of flight or search characteristics.
Conclusion
• AKF Scale Cube can be used to scale either organizations or applications.
• The AKF Scale Cube shows how we can scale but doesn’t say exactly when or which axis should be applied.
• It’s up to its users to define their needs and decide which strategies should be followed.
• Whenever you need to scale think AKF Scale Cube.
Any Questions?
Books

Webography
- https://gerardnico.com/code/design/scalability, 9/20/2019
- https://linuxacademy.com/blog/cloud/scalability-cloud-computing/, 9/20/2019
Credits

- iconfinder.com [used for icons]
- slidesgo.com [for templates]