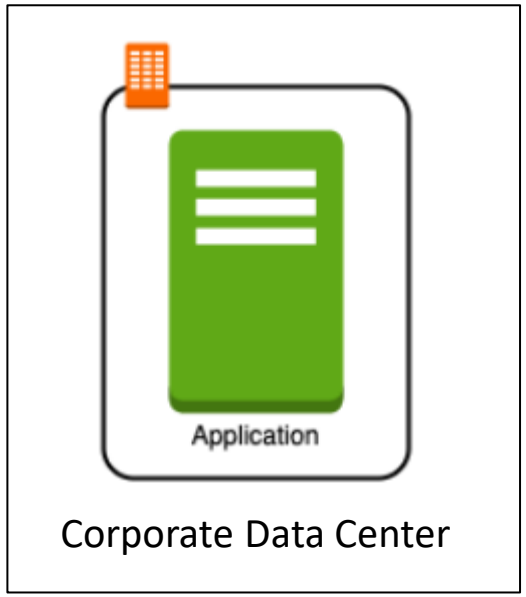
Regions and Availability Zones

Why Regions and Zones?

Imagine that your application is deployed in a data center in the US
What would be the challenges?

- Slow access for users from other parts of the world (high latency)
- What if the data center crashes? Your application goes down (low availability)



Multiple Data Centers

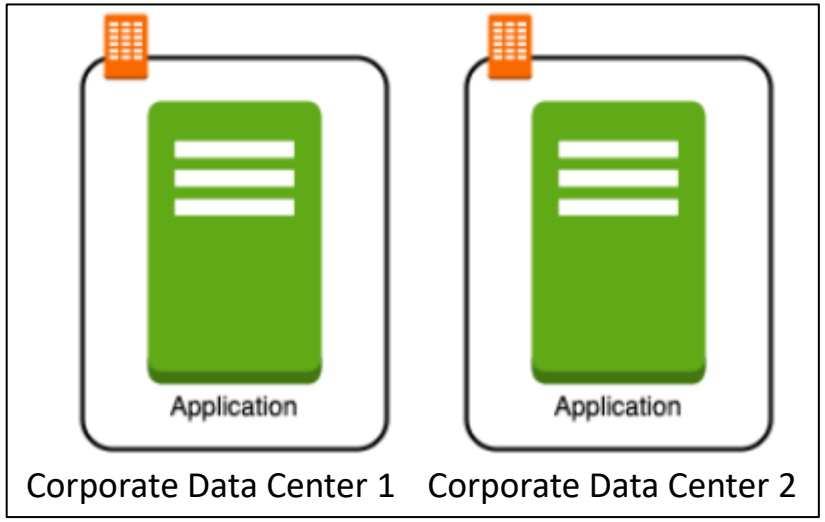
Let's add in one more data center in the US

- Slow access for users from other parts of the world
- What if one data center crashes?

(Your application is still available from the other data center)

- What if entire region of London is unavailable?

(Your application goes down)



US

Multiple Regions

What would be the challenges?

- Slow access for users from other parts of the world

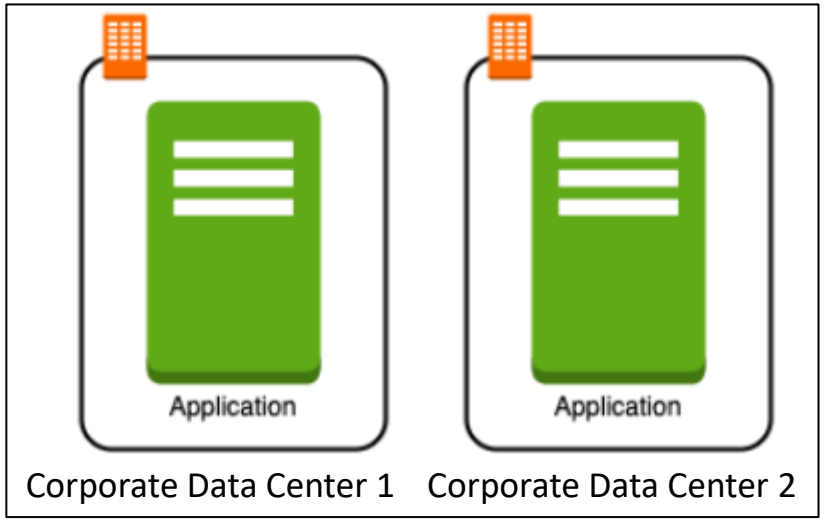
(You can solve this by adding deployments for your applications in other regions)

- What if one data center crashes?

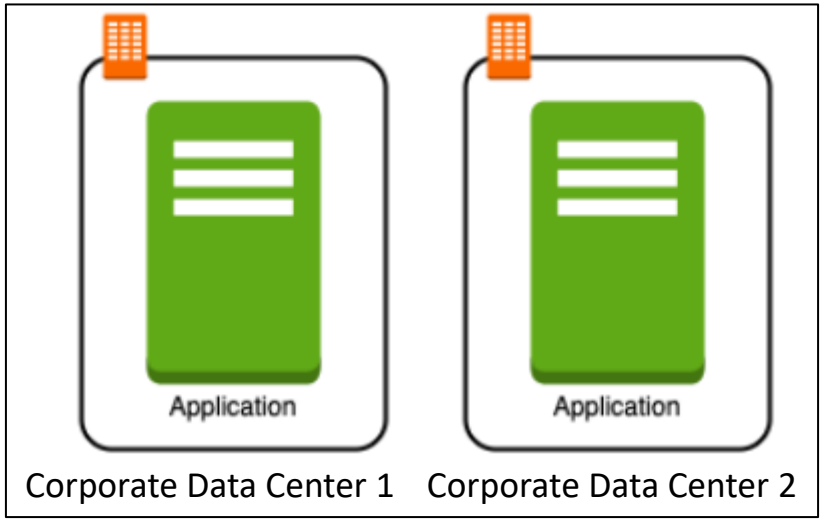
(Your application is still live from the other data centers)

- What if entire region of London is unavailable?

(Your application is served from Hyderabad)



US



Hyderabad

Regions

- Imagine setting up your own data centers in different regions around the world
- AWS provides multiple regions around the world (expanding every year)

https://aws.amazon.com/about-aws/global-infrastructure/regions_az/

Regions - Benefits

- High Availability
- Low Latency
- Adhere to government regulations

Choose the right region(s) based on:

- Where are your users located?
- Where is your data located?
- Regulatory and security compliance needs

AWS Services can be:

- Regional (OR)
- Global

Availability Zones

- Isolated locations in a Region
- Each AWS Region has at least two Availability Zones
- Increase availability of applications in the same region

https://gist.github.com/karstenmueller/98381c897178a260be8c08f98ffe2c3e

See you in the AWS Management Console!