CLOUD LEADERSHIP FORUM
Produced by IDC & IDG Enterprise
STRATEGIES FOR THE DYNAMIC ENTERPRISE

An Inside Look at Amazon.com's Own Cloud Strategy

Jerry J. Hunter
VP of Infrastructure
Amazon.com

OUT OF SCOPE

Amazon.com Website

Supply Chain Management

Corporate Applications

Financial Systems

Email & Calendar

Payments Systems

Laptops/Desktops & Telecom

HR Systems

3rd Party Seller Technologies

Developer Tools

Knowledge Management Tools

Employee Tools & Corporate Systems

Customer Service Center Software

Cost of IT

Corporate Applications KPI

$$\frac{\text{Cost of Corporate IT}}{\text{\# of Users}} = \text{Cost Per User}$$

How the Cloud Reduces Cost

Economies of scale

On-demand capacity

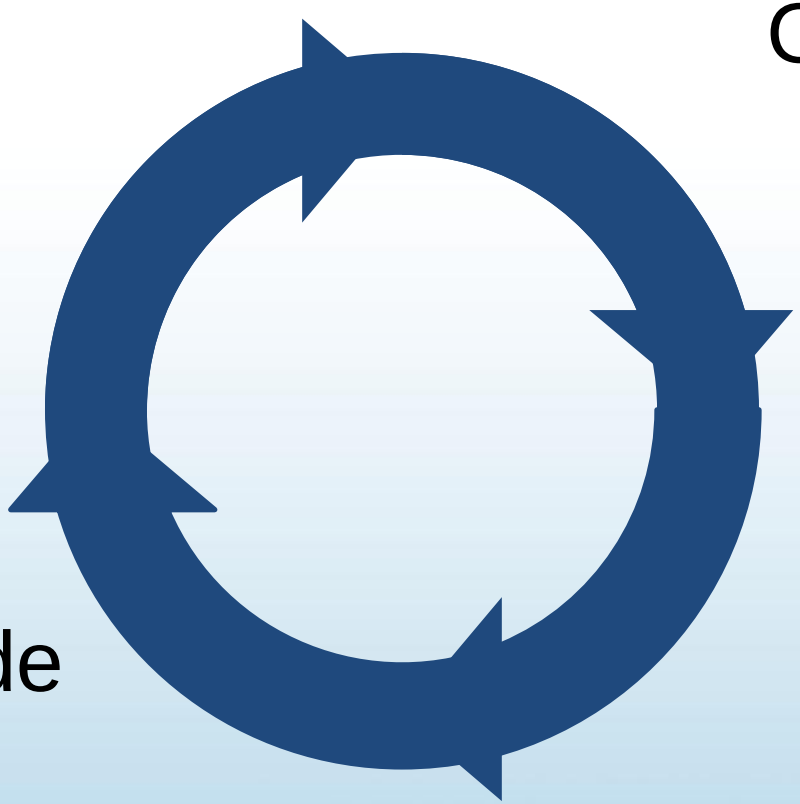
Reduced operational cost

Other Benefits >> Cost Savings


FU-012

Act

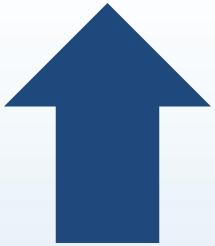
Observe



Decide

Orient

http://www.flickr.com/photos/jphilipson/2323690960/



Speed of Execution

Old Infrastructure = Speed Bumps

Hardware must be carefully chosen

Provisioning cycles are long

Everything must be built from scratch

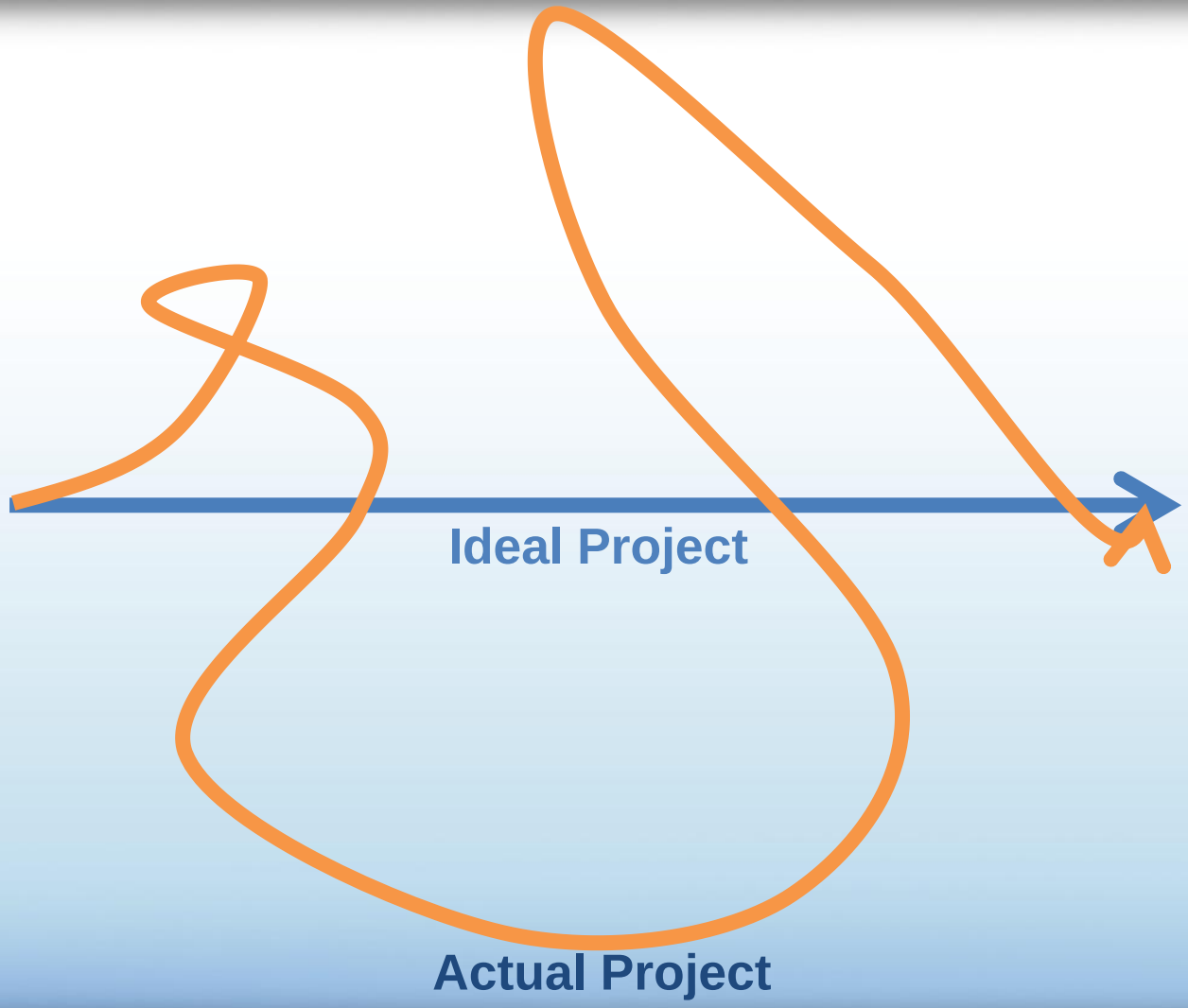
AWS Removes Speed Bumps

Switch between instance types

Provisioning is an API call

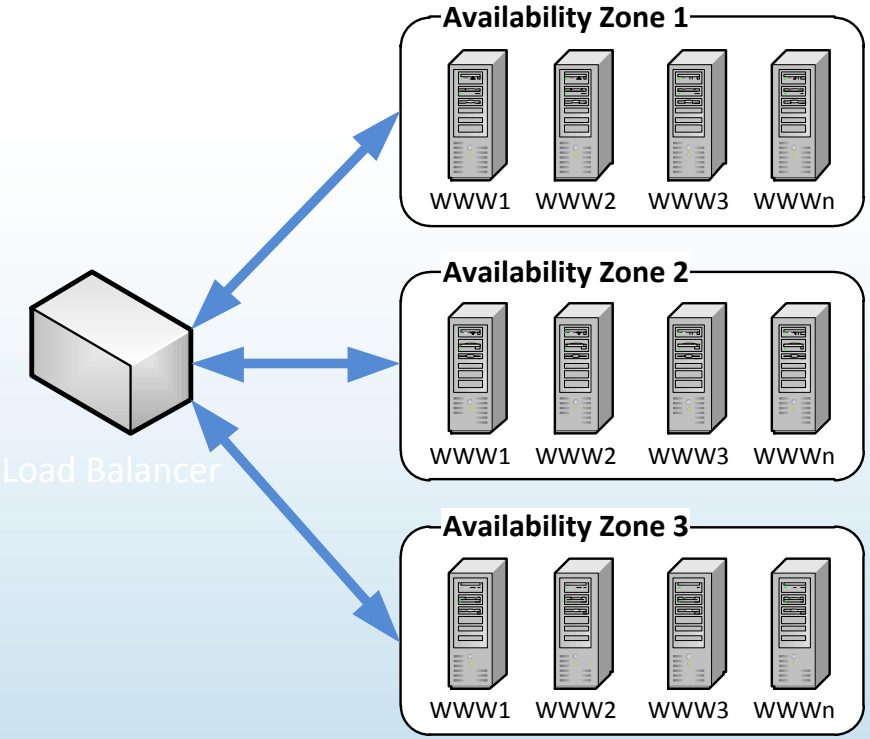
Prebuilt components

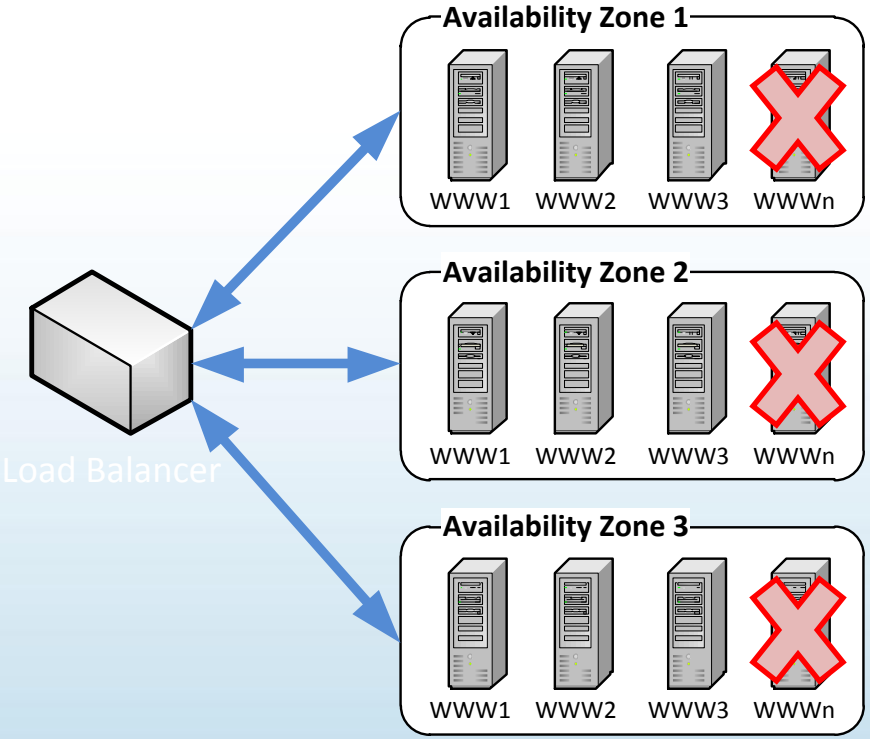

Ideal Project

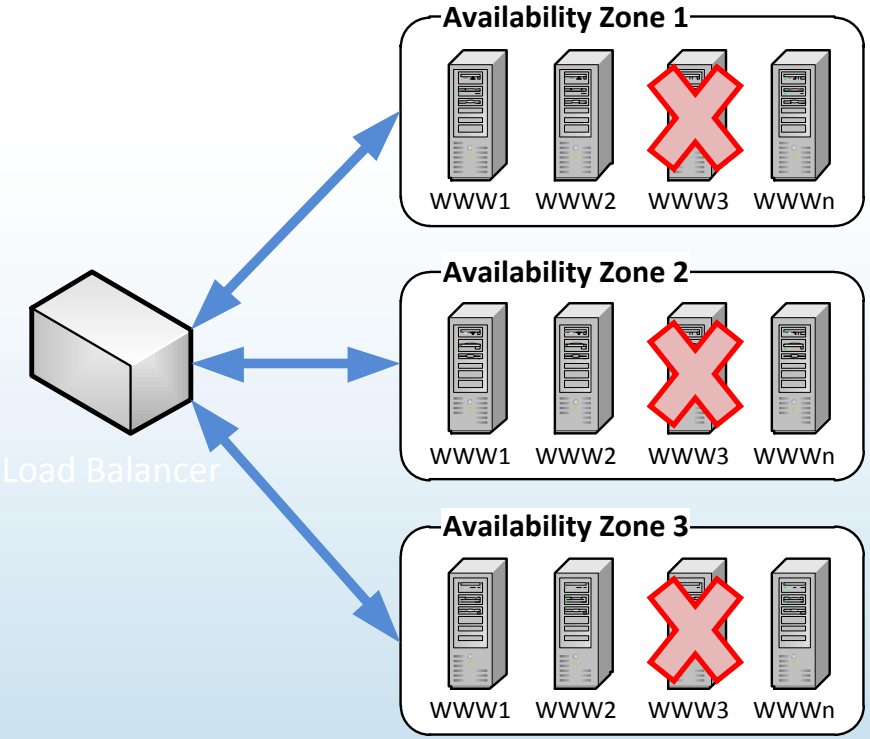

Ideal Project
Actual Project

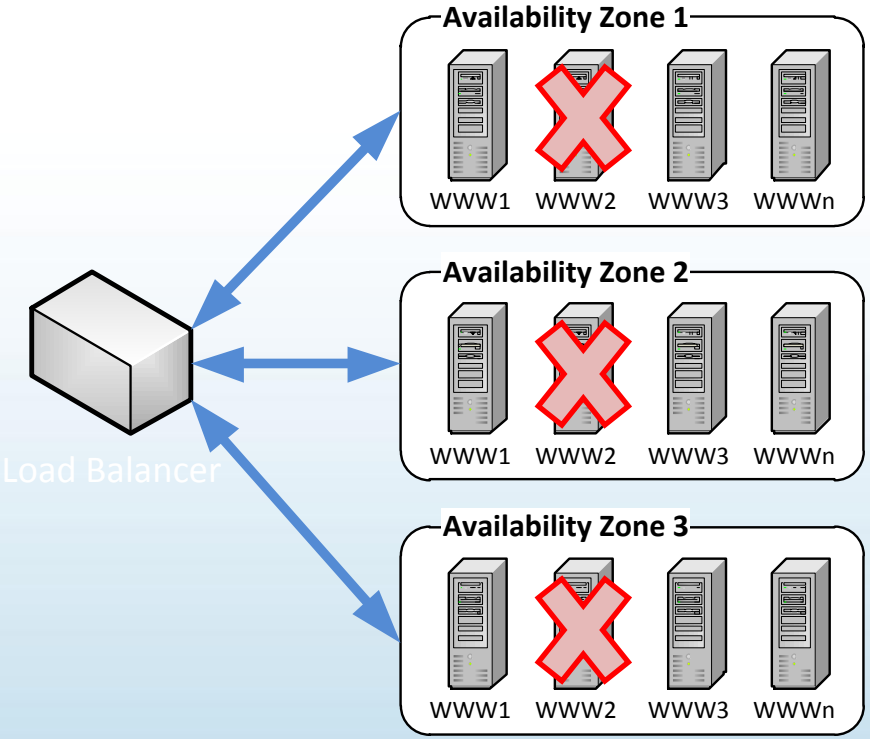
Iteration not Inspiration

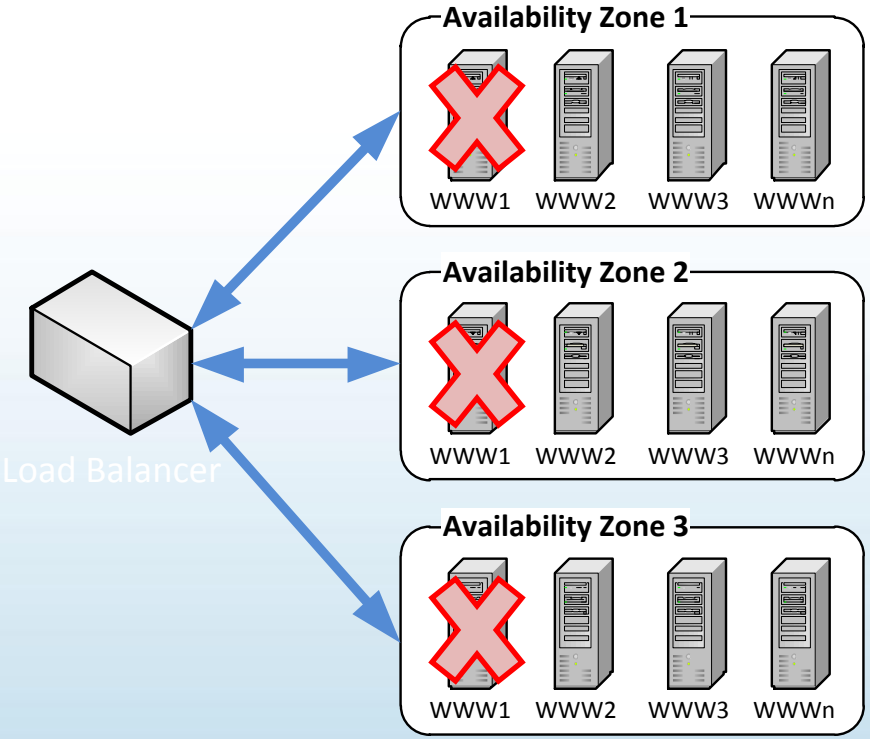
Case Study: Deployment at Amazon


Availability Zone 1
WWW1
WWW2
WWW3
WWWn
Availability Zone 2
WWW1
WWW2
WWW3
WWWn
Load Balancer
Availability Zone 3
WWW1
WWW2
WWW3
WWWn


Availability Zone 1
WWW1
WWW2
WWW3
WWWn
Availability Zone 2
WWW1
WWW2
WWW3
WWWn
Availability Zone 3
WWW1
WWW2
WWW3
WWWn
Load Balancer


Availability Zone 1
WWW1
WWW2
WWW3
WWWn
Availability Zone 2
WWW1
WWW2
WWW3
WWWn
Availability Zone 3
WWW1
WWW2
WWW3
WWWn
Load Balancer


Availability Zone 1
WWW1
WWW2
WWW3
WWWn
Availability Zone 2
WWW1
WWW2
WWW3
WWWn
Load Balancer
Availability Zone 3
WWW1
WWW2
WWW3
WWWn

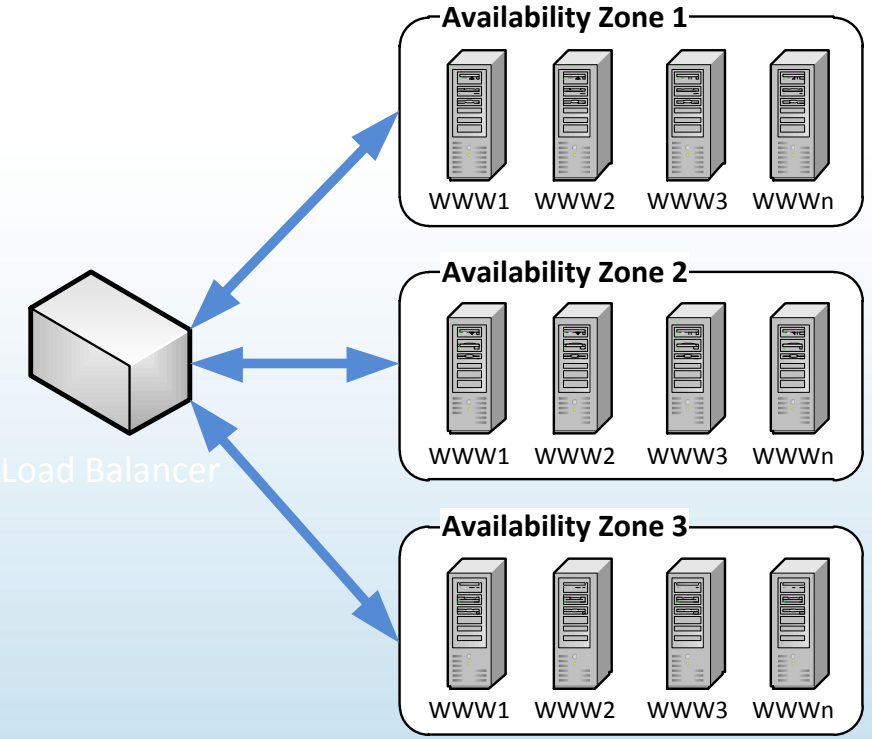

Availability Zone 1
WWW1
WWW2
WWW3
WWWn
Availability Zone 2
WWW1
WWW2
WWW3
WWWn
Availability Zone 3
WWW1
WWW2
WWW3
WWWn
Load Balancer

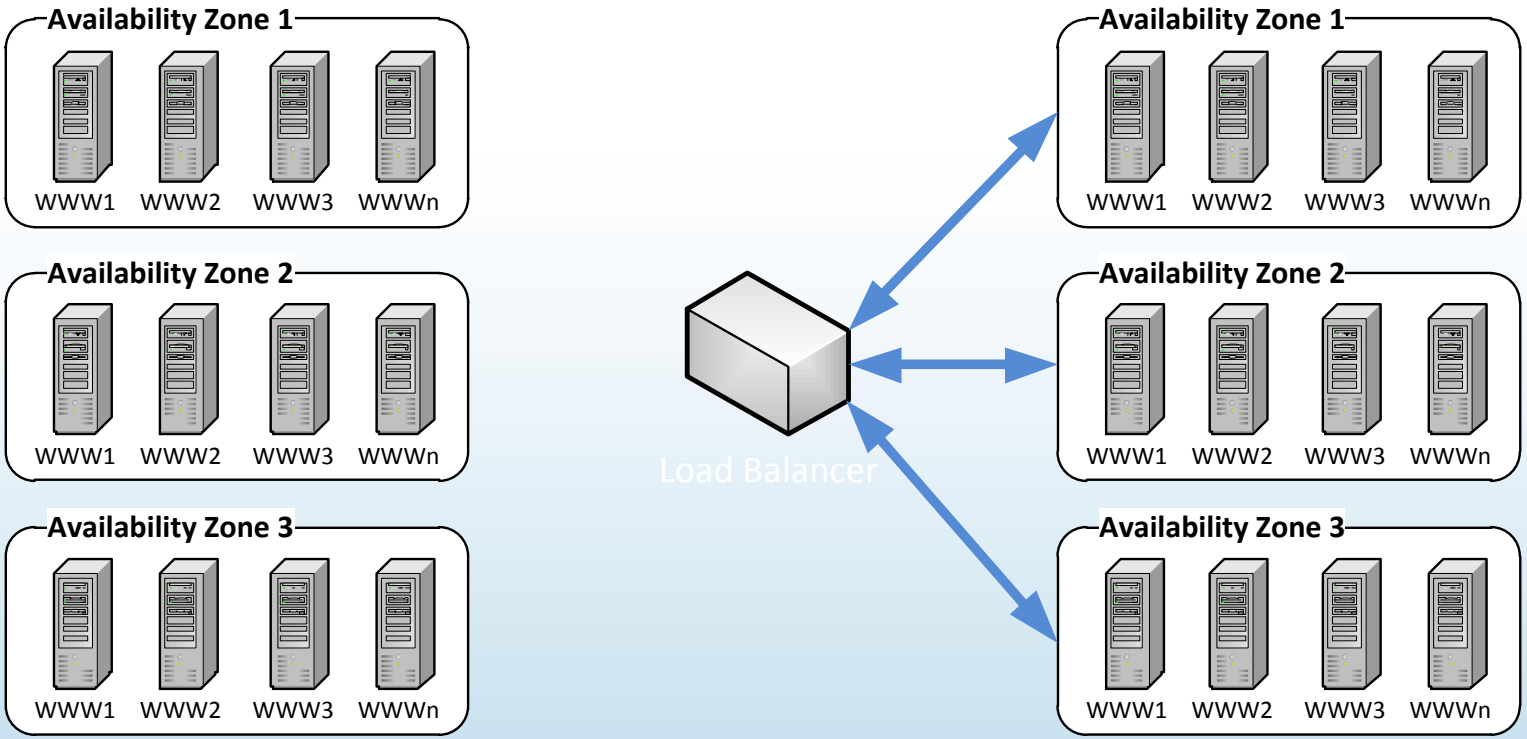
Traditional Deployment Problems

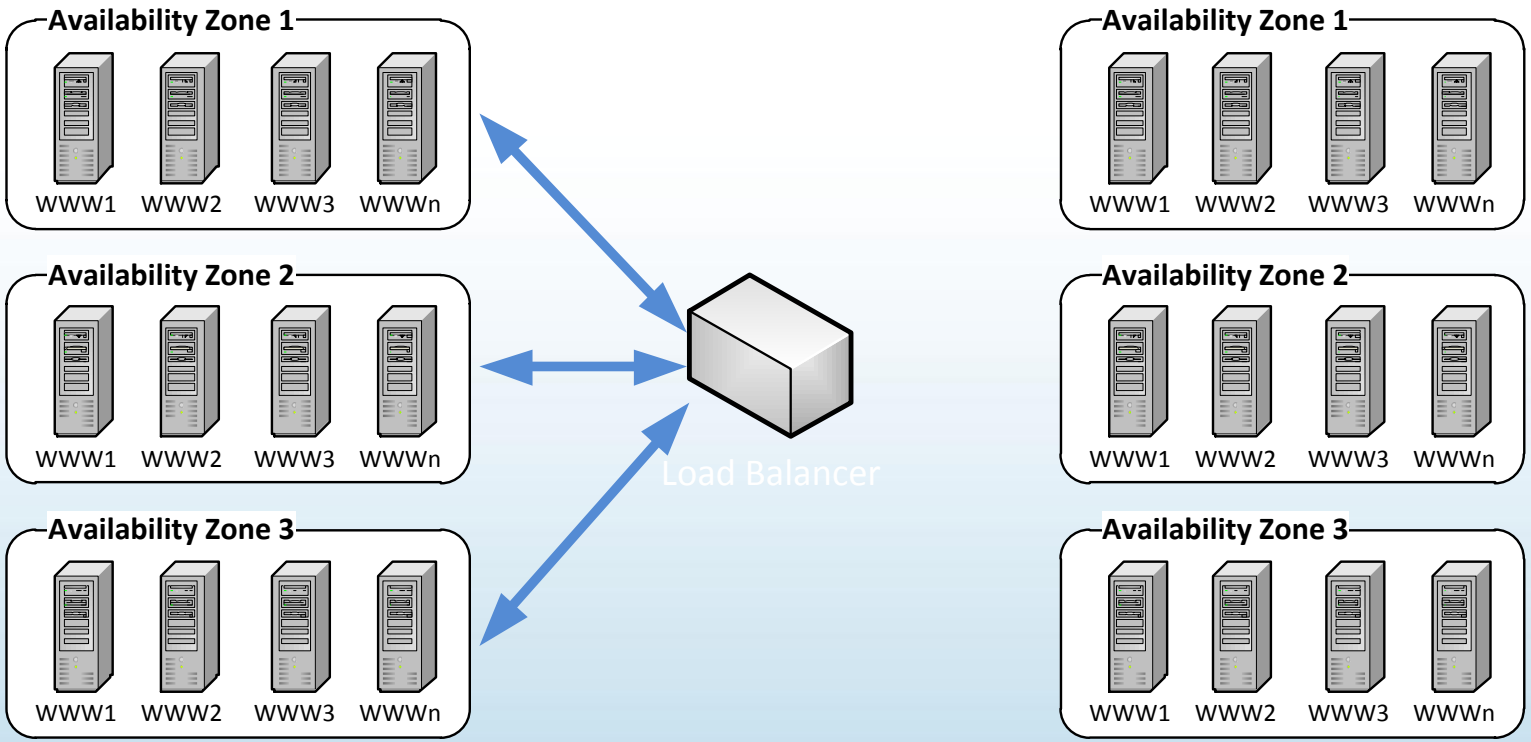
It’s a slow, serial process

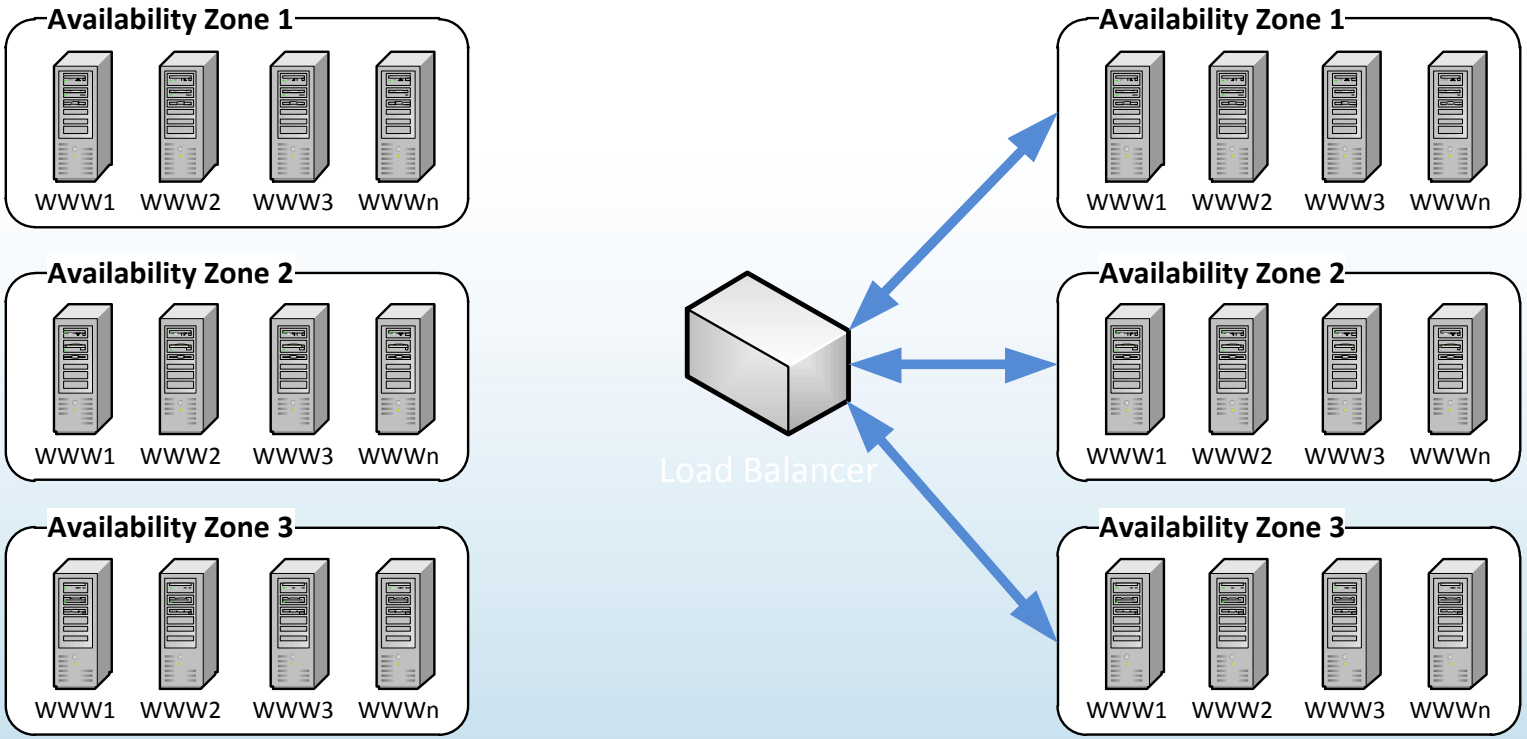
Inconsistent application state

Failures require high judgment decisions


Availability Zone 1
WWW1
WWW2
WWW3
WWWn
Availability Zone 2
WWW1
WWW2
WWW3
WWWn
Load Balancer
Availability Zone 3
WWW1
WWW2
WWW3
WWWn


Availability Zone 1
WWW1
WWW2
WWW3
WWWn
Availability Zone 2
WWW1
WWW2
WWW3
WWWn
Availability Zone 3
WWW1
WWW2
WWW3
WWWn
Load Balancer
Availability Zone 1
WWW1
WWW2
WWW3
WWWn
Availability Zone 2
WWW1
WWW2
WWW3
WWWn
Availability Zone 3
WWW1
WWW2
WWW3
WWWn


Availability Zone 1
WWW1
WWW2
WWW3
WWWn
Availability Zone 2
WWW1
WWW2
WWW3
WWWn
Availability Zone 3
WWW1
WWW2
WWW3
WWWn
Load Balancer
Availability Zone 1
WWW1
WWW2
WWW3
WWWn
Availability Zone 2
WWW1
WWW2
WWW3
WWWn
Availability Zone 3
WWW1
WWW2
WWW3
WWWn


Availability Zone 1
WWW1
WWW2
WWW3
WWWn
Availability Zone 2
WWW1
WWW2
WWW3
WWWn
Availability Zone 3
WWW1
WWW2
WWW3
WWWn
Load Balancer
Availability Zone 1
WWW1
WWW2
WWW3
WWWn
Availability Zone 2
WWW1
WWW2
WWW3
WWWn
Availability Zone 3
WWW1
WWW2
WWW3
WWWn

Amazon May Deployment Stats

(production hosts & environments only)

11.6 seconds

Mean time between deployments (typical weekday)

1,079

Max # of deployments in a single hour

10,000

Mean # of hosts simultaneously receiving a deployment

30,000

Max # of hosts simultaneously receiving a deployment

Safe Deployment at Amazon

75% reduction in outages triggered by deployments since 2007

90% reduction in outage minutes

~0.001% of deployments trigger an outage

⬇ Cost of Mistakes

New Metrics for CIOs

Mean Time to Traffic

date host received first traffic – date host was ordered

Mean Time to Deploy

date code was deployed – date code was written

Mean Time to Rollback

time was rolled back – time code was deployed

Thank You!