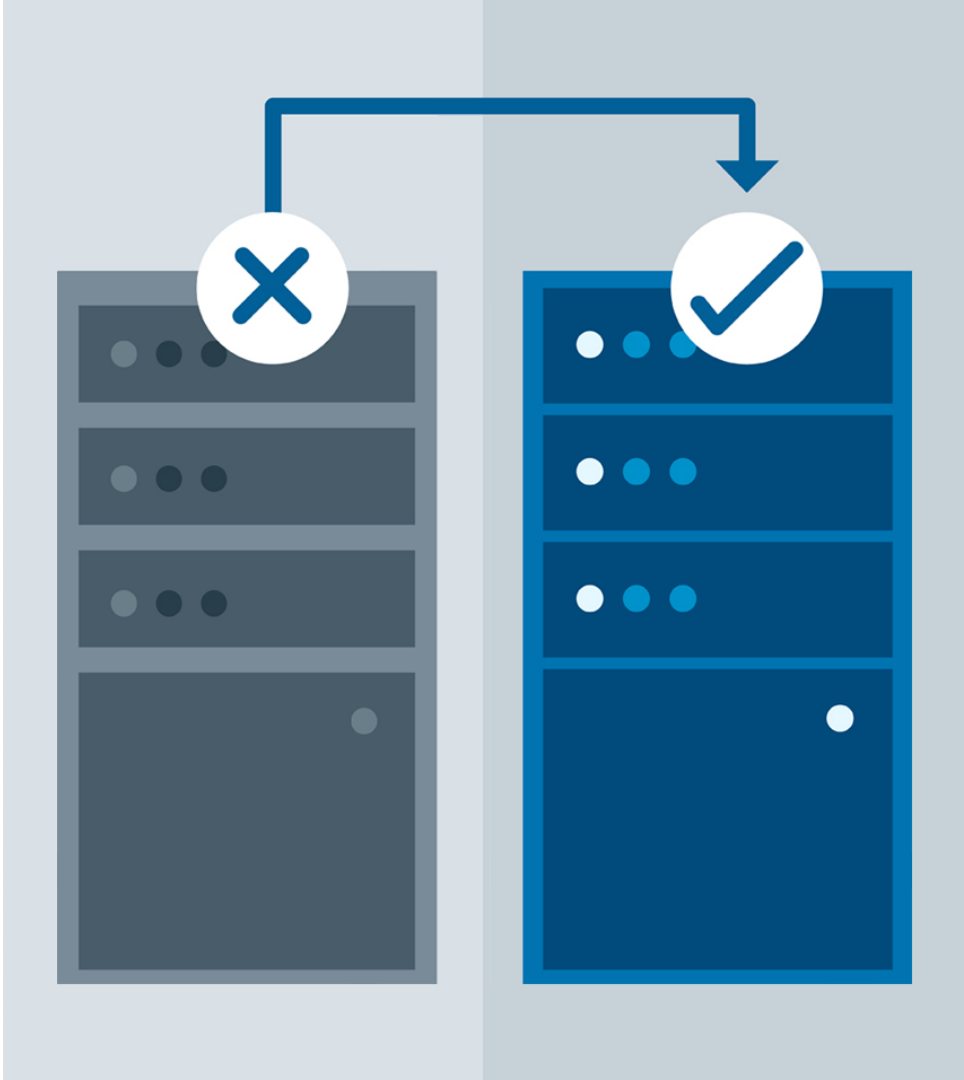
Application High Availability for Operational Continuity



Dasun Hegoda

Technical Lead at ICTA | Blogger | Speaker | Lecturer | Consultant

Overview

- Operational & Business Continuity
- High-Availability Explained
- Decision Criteria for High Availability
- Application High Availability
- Database High Availability
- Big Picture - All Together
- Demo
- Q&A

Operational & Business Continuity

Business Continuity

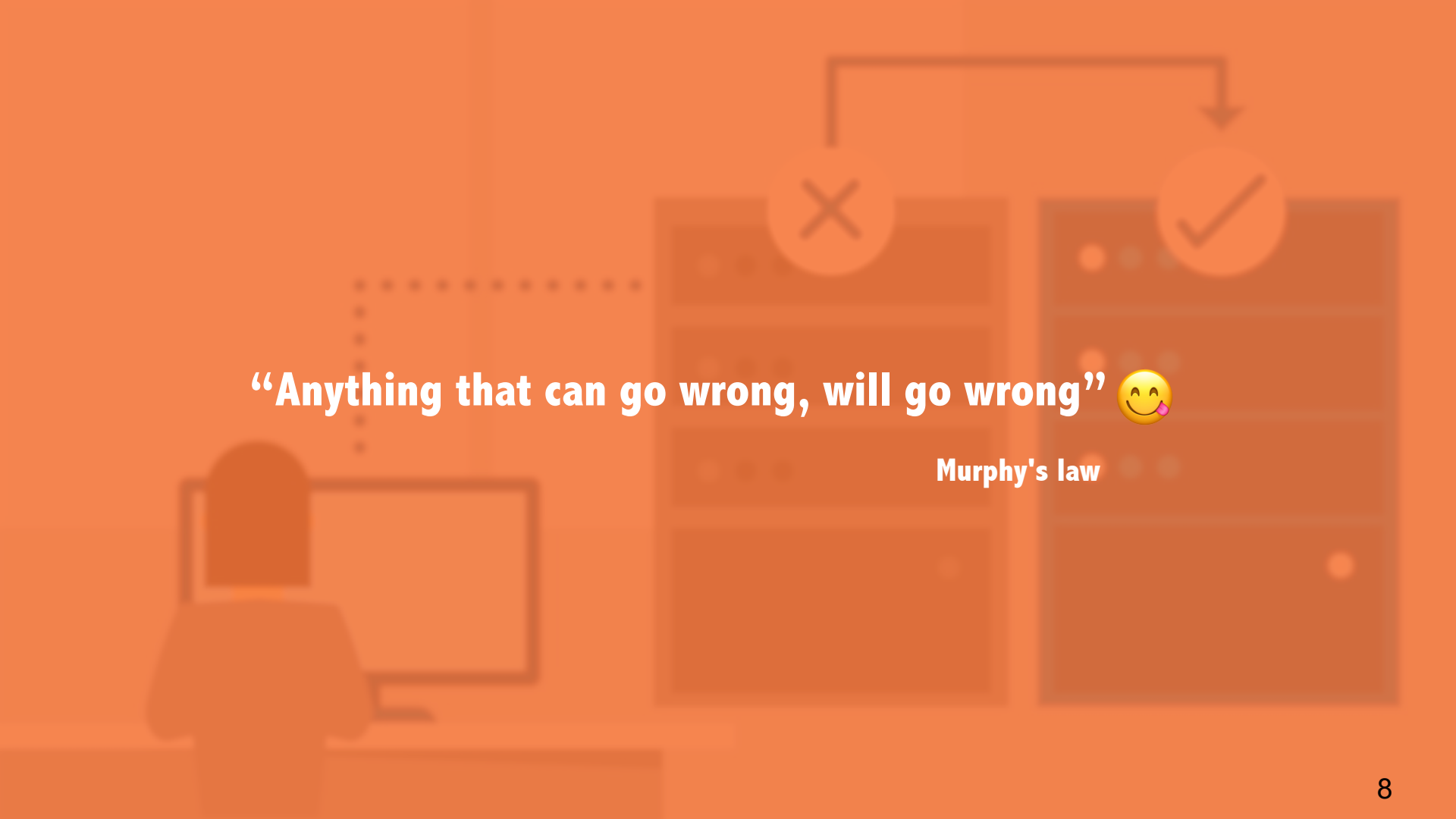
Business continuity encompasses planning and preparation to ensure that an organization can continue to operate in case of serious incidents or disasters and is able to recover to an operational state within a reasonably short period.

Operational continuity refers to the ability of a system to continue working despite damages, losses or critical events.

Business Continuity

Critical Success Factor

High Availability Explained


“Anything that can go wrong, will go wrong” 😋
Murphy's law

500: Error 

High Availability Explained

High Availability is in the eye of the beholder

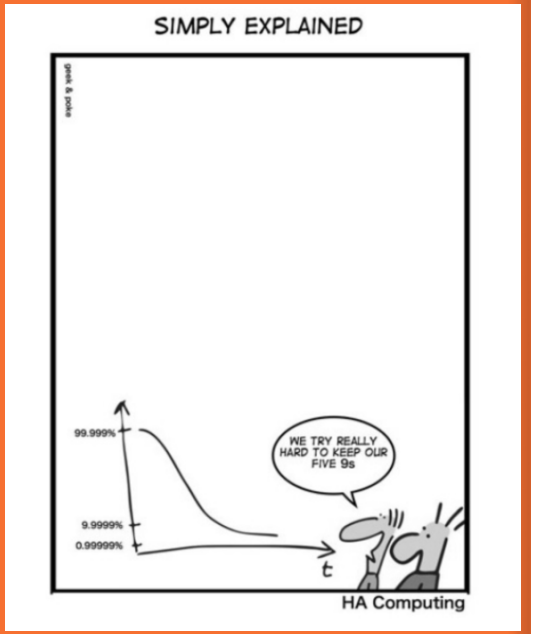
- CEO: we don't loose sales
- Sales: we can extend our offer basing on HA level
- Accounts managers: we don't upset our customers (that often)
- Developers: we can be proud – our services are working ;)
- System engineers: we can sleep well (and fsck, we love to!)
- Technical support: no calls? Back to WoW

Availability

Availability is the percentage of time that a system operates during its intended duty cycle.

$$\text{Availability} = \frac{\text{Uptime}}{(\text{Uptime} + \text{Downtime})} * 100\%$$

So How Many 9's?



Availability %	Downtime per year	Downtime per month	Downtime per week	Downtime per day
90% ("one nine")	**36.5 days**	**72 hours**	**16.8 hours**	**2.4 hours**
95% ("one and a half nines")	18.25 days	36 hours	8.4 hours	1.2 hours
97%	10.96 days	21.6 hours	5.04 hours	43.2 minutes
98%	7.30 days	14.4 hours	3.36 hours	28.8 minutes
99% ("two nines")	**3.65 days**	**7.20 hours**	**1.68 hours**	**14.4 minutes**
99.5% ("two and a half nines")	1.83 days	3.60 hours	50.4 minutes	7.2 minutes
99.8%	17.52 hours	86.23 minutes	20.16 minutes	2.88 minutes
99.9% ("three nines")	**8.76 hours**	**43.8 minutes**	**10.1 minutes**	**1.44 minutes**
99.95% ("three and a half nines")	4.38 hours	21.56 minutes	5.04 minutes	43.2 seconds
99.99% ("four nines")	**52.56 minutes**	**4.38 minutes**	**1.01 minutes**	**8.64 seconds**
99.995% ("four and a half nines")	26.28 minutes	2.16 minutes	30.24 seconds	4.32 seconds
99.999% ("five nines")	**5.26 minutes**	**25.9 seconds**	**6.05 seconds**	**864.3 milliseconds**
99.9999% ("six nines")	**31.5 seconds**	**2.59 seconds**	**604.8 milliseconds**	**86.4 milliseconds**
99.99999% ("seven nines")	**3.15 seconds**	**262.97 milliseconds**	**60.48 milliseconds**	**8.64 milliseconds**
99.999999% ("eight nines")	**315.569 milliseconds**	**26.297 milliseconds**	**6.048 milliseconds**	**0.864 milliseconds**
99.9999999% ("nine nines")	**31.5569 milliseconds**	**2.6297 milliseconds**	**0.6048 milliseconds**	**0.0864 milliseconds**

Factors That Determine System Availability

Reliability of the individual components that comprise the system.

Ex : hardware, operating system, database and the application itself.

Time it takes for the application to be restored once a failure has occurred.

Design for High Availability

Decision Criteria for High Availability

Decision Criteria for High Availability

Is the information system a revenue generator?

Are there alternate methods to conduct business while the system is repaired?

What is the value of the lost revenue if the information system is not available?

Does the information system downtime affect the employee productivity?

Is the information system mission critical?

Is the loss of reputation and trust due to poor availability critical to the enterprise?

Decision Criteria for High Availability

Does the non-functioning of the information system result in lost customers?

RPO: Recovery Point Objective; how much data can we loose?

RTO: Recovery Time Objective; how long does it take to recover?

SLA: Service Level Agreement

OLA: Operational Level Agreement; within organization; help us keeping provided SLAs

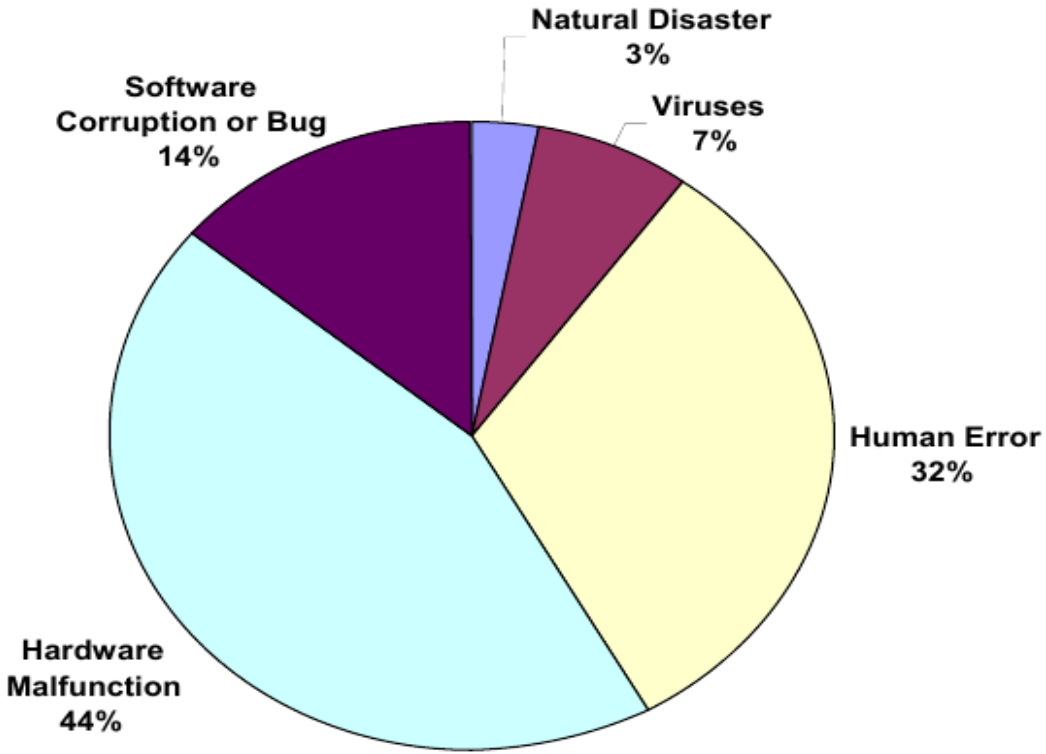
Designing For Downtime

1. Downtime are results from a system failure.

2. Downtime due to scheduled outages.

It is up to the **system designer** to **understand the business need** and **design** the system to allow for **planned** downtime

Causes of Downtime



Causes of Downtime

Lack of best practice in,

- change control
- monitoring of the relevant components
- operations
- avoidance of network failures
- avoidance of internal application failures
- avoidance of external services that fail
- physical environment
- network redundancy
- physical location
- infrastructure redundancy

General Recommendations

- Spend money, not blindly
- Examine the system history for failure patterns
- Remove single points of failure
- Establish Service Level Agreements (SLA)
- Plan ahead for outages and disasters
- Keep it simple
- Plan for system maintenance
- Conduct planning meetings for maintenance
- Maintain separate environment for test and lab
- Choose mature software
- Make the application stateless
- Microservices architecture with docker

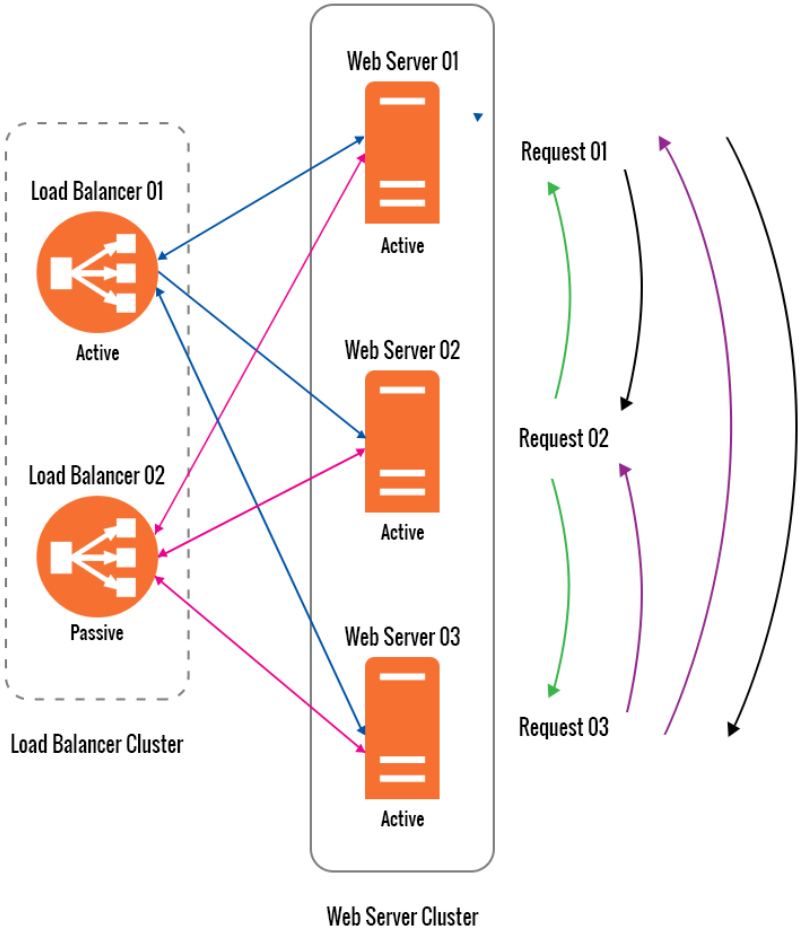
Application High Availability

Sticky and Non-Sticky sessions

- Sticky : only single session object will be there.
- Non-sticky session : session object for each server node

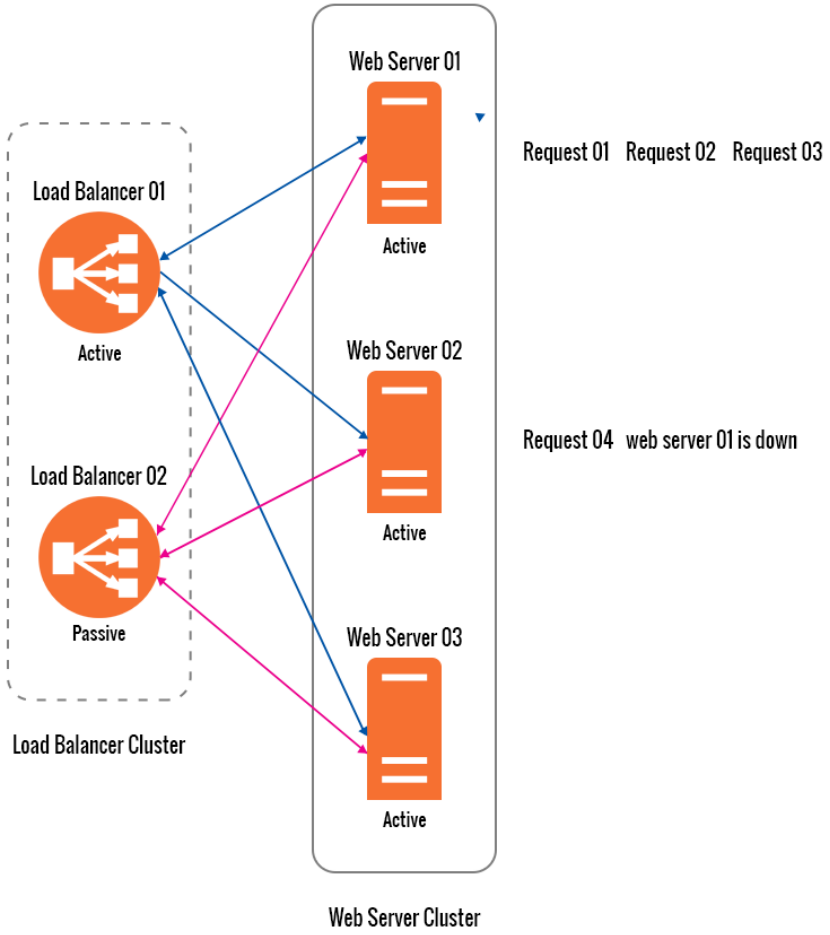
Application High Availability

Session replication without sticky session



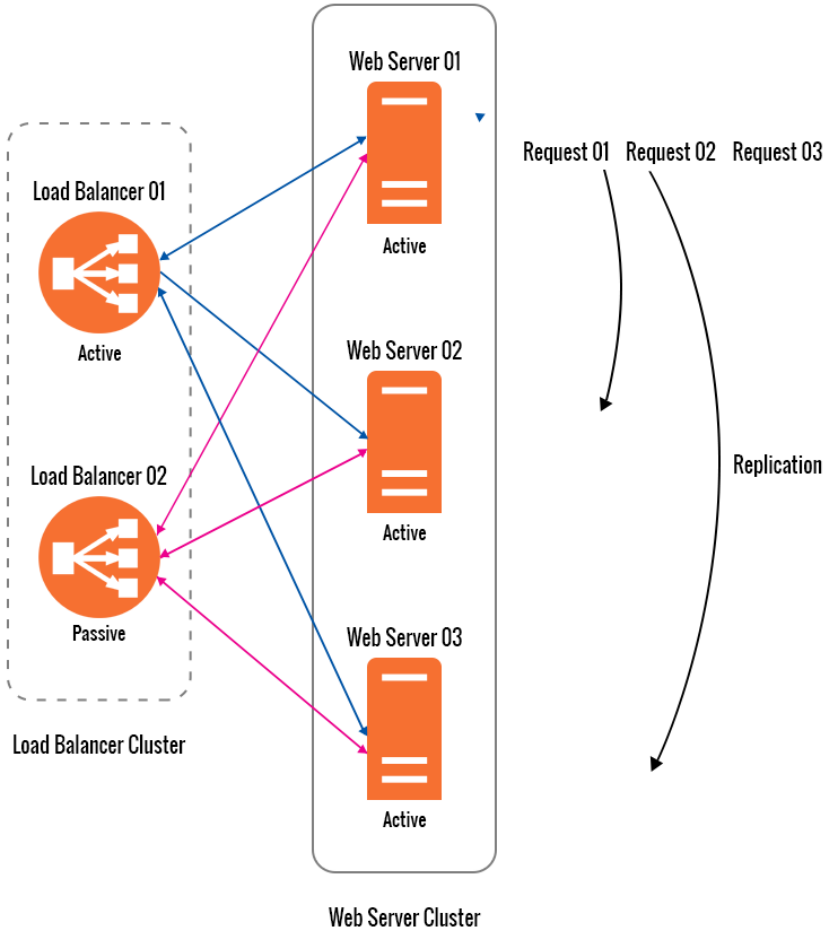
Application High Availability

Sticky session without session replication

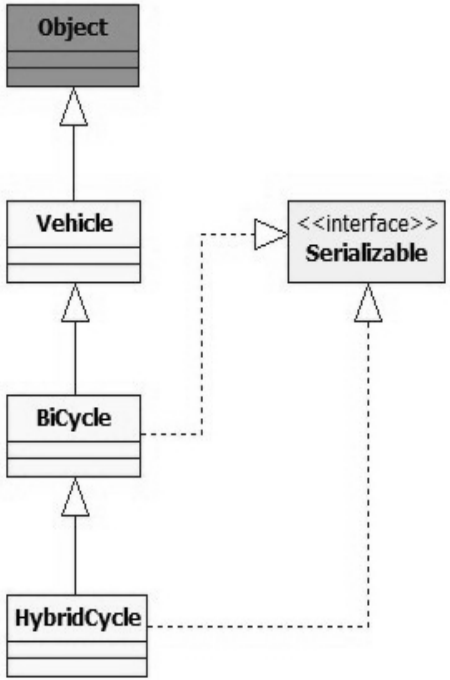


Application High Availability

Session replication with sticky session



Object Serialization



What is Java Serialization?

- Implement the *java.io.Serializable interface*
- Object Stream → Transport Through Network → Rebuilt Object

Database High Availability

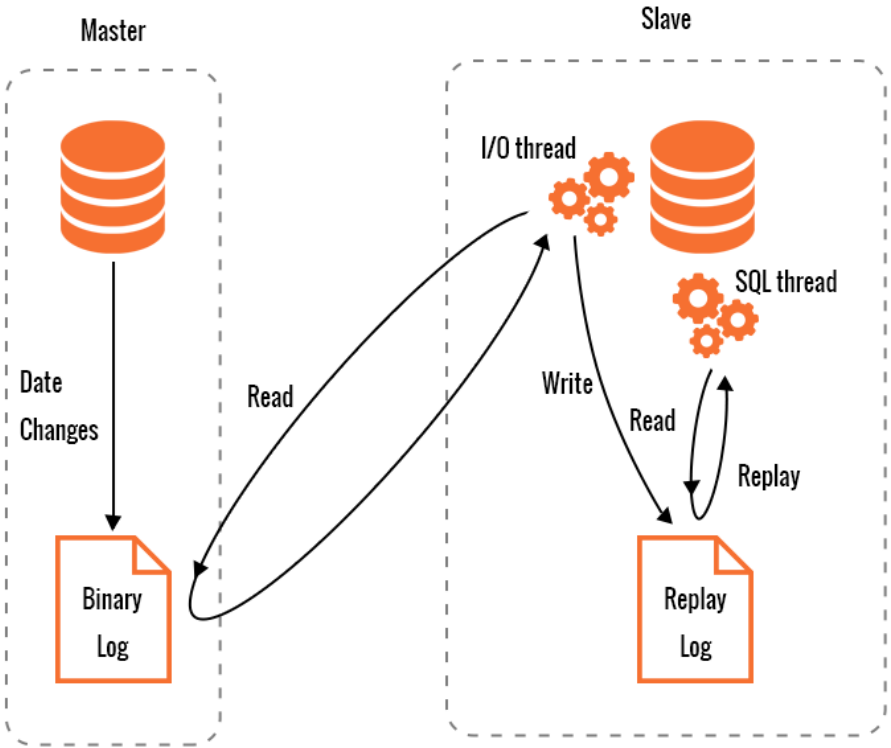
Database High Availability

Many of today's enterprise applications rely on commercial databases, therefore it is appropriate to review the current "best practice" regarding their design and configuration.

- Replication (Master-Master, Master-Slave)
- Database Clustering

Database High Availability

Database replication



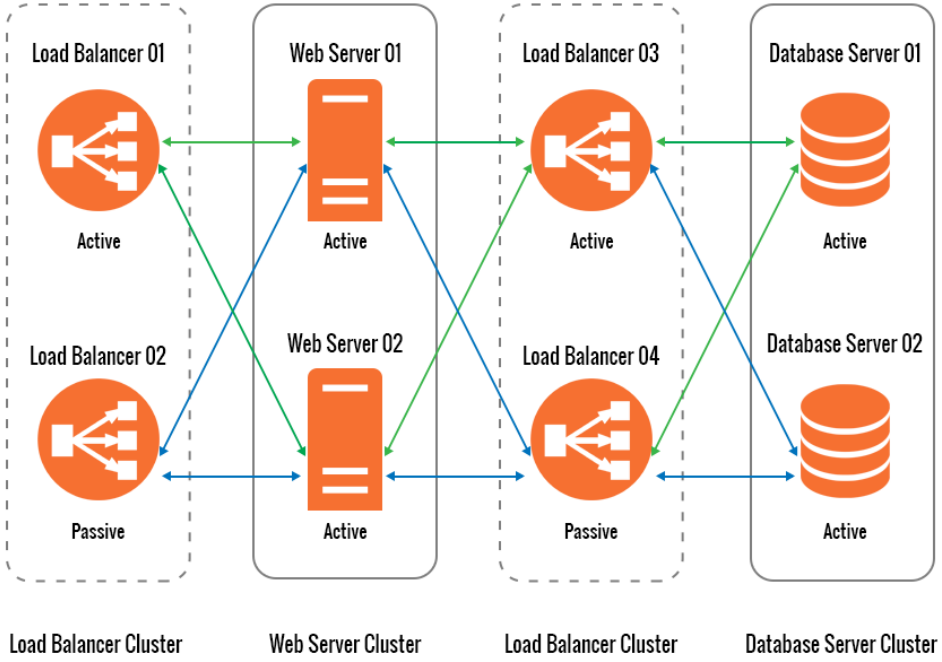
Software Load Balancers

- HAProxy
- Apache mod-proxy
- Apache mod-jk
- BalanceNG
- Percona
- NGINX
- Linux Virtual Server (LVS)
- List goes on.....

Big Picture – All Together

2 Node High Availability Cluster Deployment

All together in one



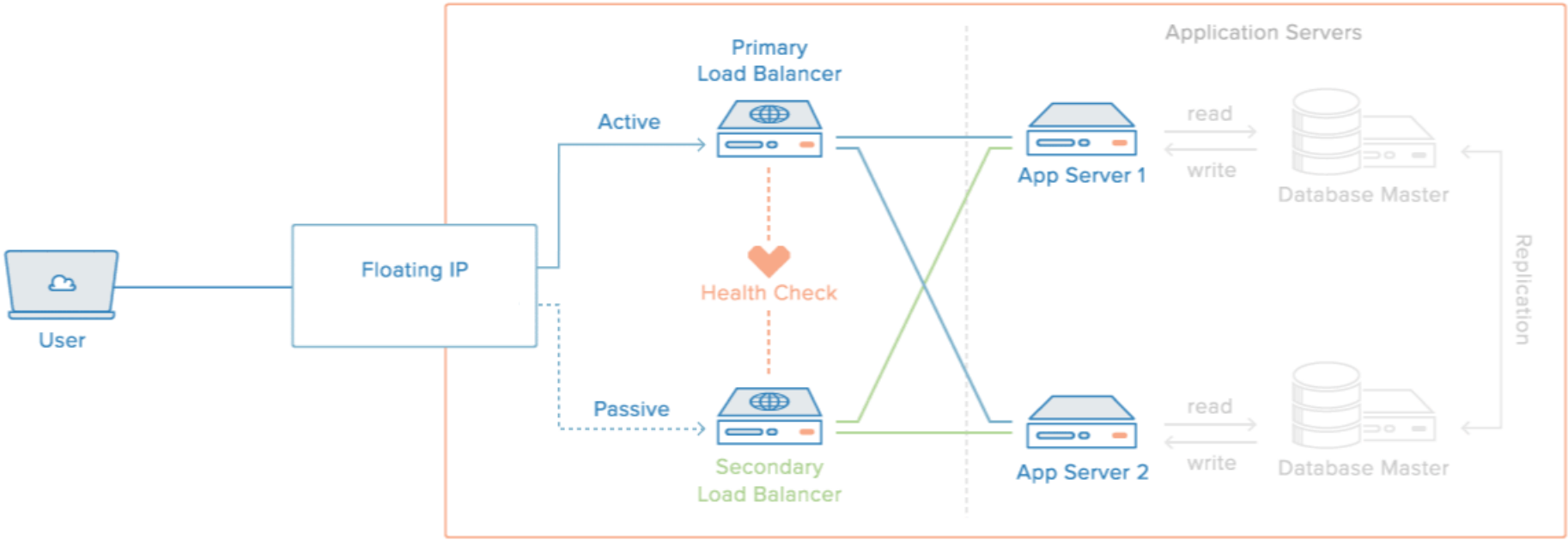
200: OK 

Demo

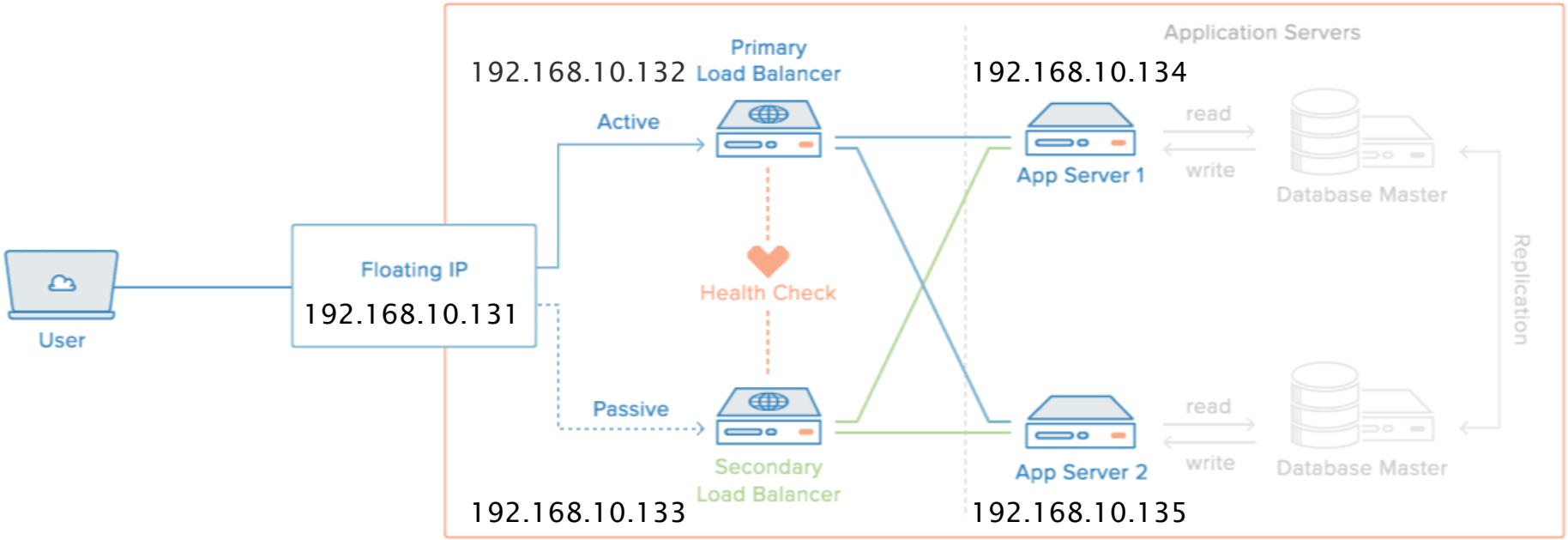
Step by Step Guides

You don’t have to take

dasunhegoda.com



1. Active/Passive Cluster is healthy
2. Primary node fails
3. Floating IP is assigned to Secondary node



1. Active/Passive Cluster is healthy
2. Primary node fails
3. Floating IP is assigned to Secondary node

Q&A

Thank you very much for your time