

Disaster Recovery

Architectural Strategies


Service Outage

Werner Vogels: “Everything fails all the time”

https://www.slideshare.net/AmazonWebServices/high-availability-websites-part-one/12-Everything_fails_all_the_time

365online.com failure

Tue March 2nd 2017 about 8PM GMT

Bank of Ireland 

365online

Need help using this site? [GET HELP](#)


Error

We are unable to process your request at this time, please try again later.

Alternatively, if you would like to access your accounts via 365 phone, please call;

- Republic of Ireland: 0818 365 365 or 01 404 4000
- Great Britain / Northern Ireland: 0345 7 365 555
- Outside these locations: + 353 1 404 4000

Opening Hours: 8am to Midnight, Monday to Friday or 9am to 6pm on Saturdays, Sundays, Bank and Public Holidays.

Try Again 

365online.com failure

- Bank of Ireland have not issued any public information on this outage
- Nothing on their FB page, in response to customer comments: <https://www.facebook.com/BankofIreland/posts/1271336836296733>



AWS S3 Failure

S3 Service Disruption - a fascinating read. One of the more interesting passages:

...

“ we have not completely restarted the index subsystem or the placement subsystem in our larger regions for many years”

...

<https://aws.amazon.com/message/41926/>

Lessons Learned

- Although we cannot say what caused the 365online outage we can say that the S3 outage, caused by human error, took longer to restore as AWS hadn't tested the restore process for many years.
- Test your backup and restore strategies on a regular basis - do they meet your SLAs?

What was the recovery time for the S3 outage?

Any comments on this as an acceptable timeframe?

Disaster Recovery

There are 4 Basic Architectural Approaches to Disaster Recovery

- Backup and Restore
- Pilot Light
- Fully Working Low-Capacity Standby
- Multi-Site Active-Active

Or possibly a mix and match of all 4

RTO & RPO

RTO: Recovery Time Objective. How long can you wait for your system to come back up if it fails?

RPO: To what point in time will you accept lost data?

Generally, the lower RTO & RPO goes, the higher the cost to the business.



Why Systems Fail

And will it happen to you?

Distributed Architecture

Most architectures have some form of distribution

- [aside - do you use a distributed architecture?]
- Failure in one component can cause a cascade across the architecture

Why?

- Because we typically use components we have no control over
- And if these fail, then our code fails too
- Worse, it doesn't even need failure, just resource saturation

S3 Outage Cascade

...” Other AWS services in the US-EAST-1 Region that rely on S3 for storage, including the S3 console, Amazon Elastic Compute Cloud (EC2) new instance launches, Amazon Elastic Block Store (EBS) volumes (when data was needed from a S3 snapshot), and AWS Lambda were also impacted while the S3 APIs were unavailable. ” ...

Synchronous Comms

Software is still written using synchronous communications even though there is sometimes a WAN in between communicating components

Distributed architectures can be robust when designed properly and unstable when not designed properly

An important architectural concern is **“how do we handle the failure of a component we can’t control?”**

=> We incorporate failure planning into our architectures

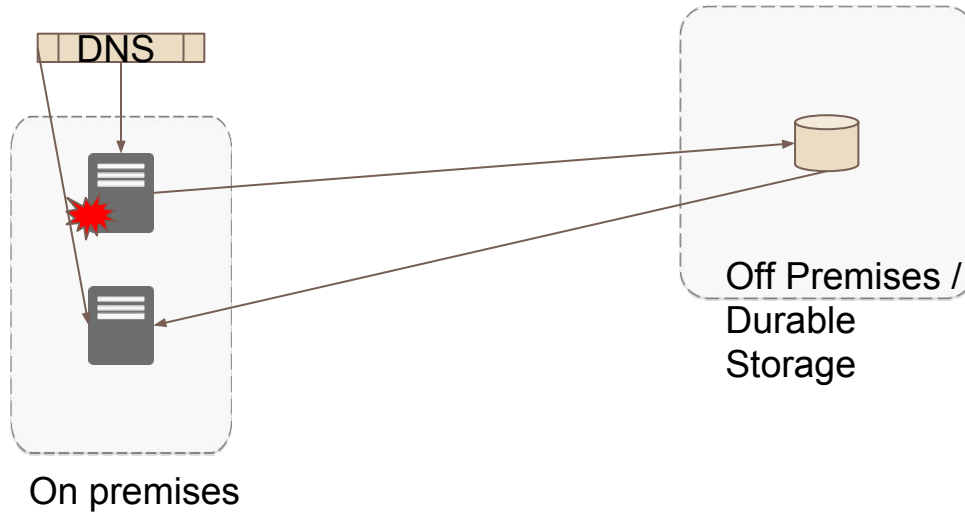


DR: Backup and Restore

Least complex, least expensive

Backup and Restore

This is the most basic DR strategy. Data periodically copied off the on-prem server to a remote storage facility



Backup and Restore

Once the failure is detected:

- snapshots data is copied back from the remote storage
- new service is provisioned
- data is loaded into the database
- DNS server is used to point to the new instance IP
- customer service is restored

Backup and Restore

Advantages

- Simple to get started
- Extremely cost effective (mostly backup storage)

Preparation Phase

- Take backups of current systems
- Store backups in Durable Storage
- Know how to restore system from backups.
- Know how to switch to new system.
- Know how to configure the deployment.

RTO & RPO

Objectives

RTO: as long as it takes to bring up infrastructure and restore system from backups.

RPO: time since last backup.

Q: Is this in the SLAs?

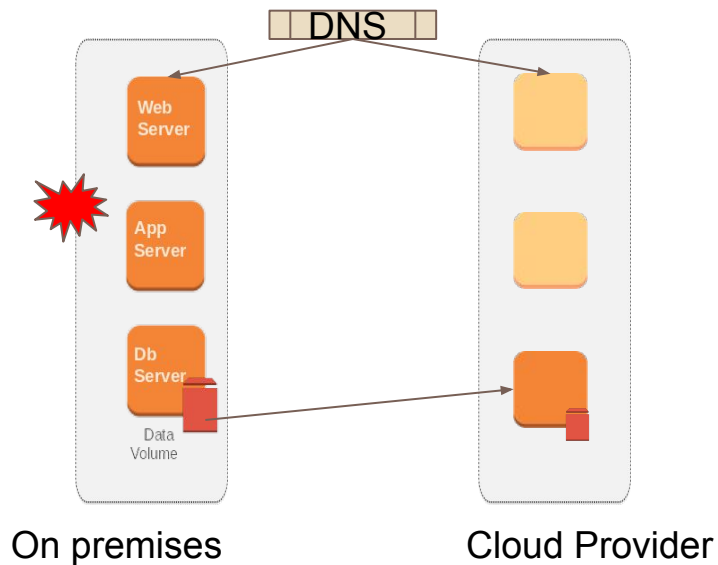


DR: Pilot Light

Low Cost, Medium Complexity

Pilot Light

Run a small instance DB on separate infrastructure



Pilot Light

- Smaller DB instance is used for replication / mirroring
- The database can be resized to PROD size in the event of outage
- The Web & App layers can be started in a few minutes

Advantages

- Very cost-effective (fewer 24/7 resources)

Pilot Light - Preparation

Preparation Phase

- Enable replication of all critical data to durable cloud storage
- Prepare all required resources for automatic start
- EG, in AWS: AMIs, Network Settings, Elastic Load Balancing, etc.
- Reserved Instances

Pilot Light DR & Objectives

In case of disaster

- Automatically bring up resources around the replicated core data set.
- Scale the system as needed to handle current production traffic.
- Switch over to the new system.
- Adjust DNS records to point to cloud provider.

Objectives

RTO: as long as it takes to detect need for DR and automatically scale up replacement system.

RPO: depends on replication type.

Q: Is this in the SLAs?

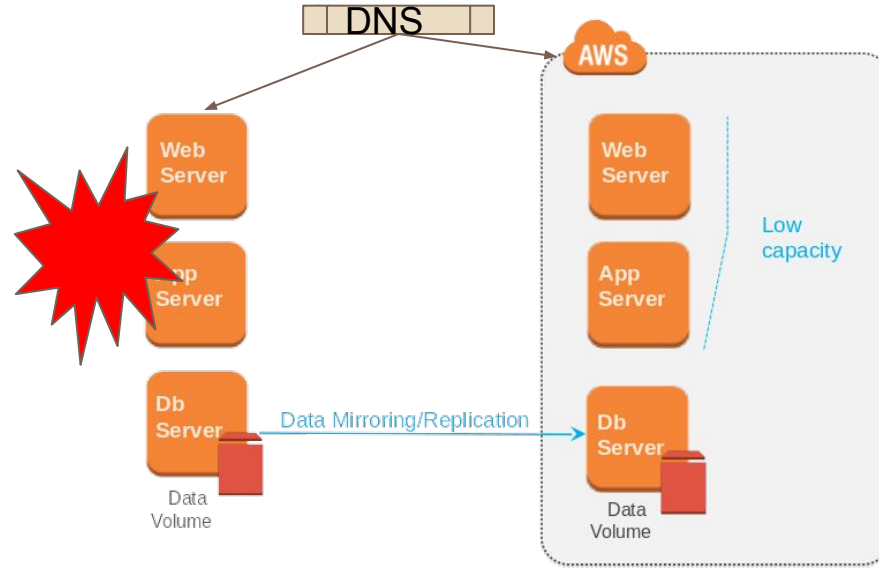


DR: Fully Working Low-Capacity Standby

Medium Cost, Medium Complexity

Fully Working Low-Capacity Standby

Using AWS



Fully Working Low-Capacity Standby

Advantages

- Can take some production traffic at any time
- Cost savings (IT footprint smaller than full DR)

Preparation

- Similar to Pilot Light
- All necessary components running 24/7, but not scaled for production traffic
- Best practice: continuous testing
- “Trickle” a statistical subset of production traffic to DR site

Fully Working Low-Capacity Standby

In case of disaster

- Immediately fail over most critical production load
- Adjust DNS records to point to AWS
- (Auto) Scale the system further to handle all production load

RTO/RPO Objectives

Objectives

- RTO: for critical load: as long as it takes to fail over; for all other load, as long as it takes to scale further
- RPO: depends on replication type

Q: Is this in the SLAs?

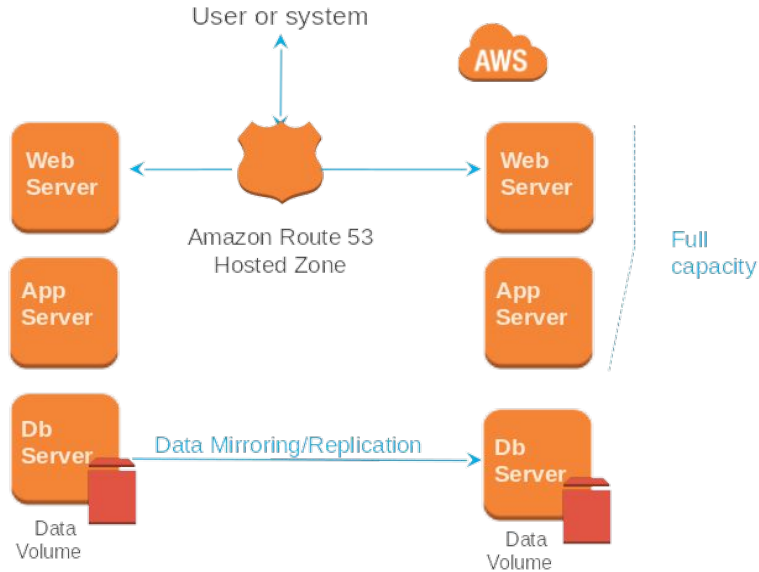


DR: Multi-Site Active-Active

Complex / Expensive

Multi-Site Active-Active

[Using AWS DNS]



Multi-Site Active-Active

Advantages

- At any moment can take all production load

Preparation

- Similar to Low-Capacity Standby
- Fully scaling in/out with production load
- In Case of Disaster
- Immediately fail over all production load

RTO/RPO Objectives

Objectives

RTO: as long as it takes to fail over

RPO: depends on replication type

Q: Is this in the SLAs?



DR: Best Practices

Cloud Provider or all in-house?

DR Best Practices

Start simple and work your way up

- Backups in AWS as a first step
- Incrementally improve RTO/RPO as a continuous effort

Check for any software licensing issues

Exercise your DR Solution

- Practice "Game Day" exercises
- Ensure backups, snapshots, AMIs, etc. are working
- Monitor your monitoring system

Cloud Provider Selection

Select a Cloud Provider or Vendor that has

- Various building blocks available
- Fine control over cost vs. RTO/RPO tradeoffs
- Ability to easily and effectively test your DR plan
- Availability of multiple locations worldwide
- Managed desktops available
- Possibly managed desktop

AWS is one possible solution, Azure is another



AWS DR Case Study

Unilever

Unilever DR with AWS



Amazon S3

Backup data, snapshots, product and recipe media files stored durably and made available to all regions



Amazon EBS

EBS Snapshot Copy used to backup SQL Server on Amazon EC2 data volumes and restore them to the disaster recovery region when necessary



Amazon EC2

Created ~400 AMIs of Windows and Linux instances to provide flexible deployment across environments

In DR region, all instances except master SQL Server are kept on standby, activated when needed

Unilever DR with AWS

Unilever

“We designed a disaster recovery solution to protect our content management system, content deployment architecture, and many GOLD-classified web properties—and to give the business confidence in the AWS Cloud.

Sreenivas Yalamanchili
Global Technical Manager, Unilever Digital marketing

