science.nasa.gov
Drupal Cloud Architecture

Chris Shenton

@shentonfreude
chris@koansys.com

2014-03-18
Drupal NoVA Meetup

science.nasa.gov, ciencia.nasa.gov

science.nasa.gov

2nd busiest site at NASA

NASA is a popular target for attacks

Implementation History

2008: Plone CMS: single box at NASA HQ

2010: Django + FeinCMS: 6 VMs, now at AWS

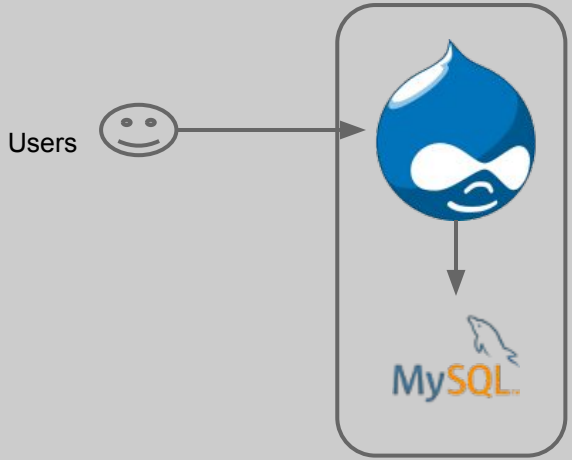
2014: Drupal: AWS, leverage cloud services

Architecture Needs to Support...

- High Volume
- Scalable
- Secure
- oh, yeah: Cost-Effective

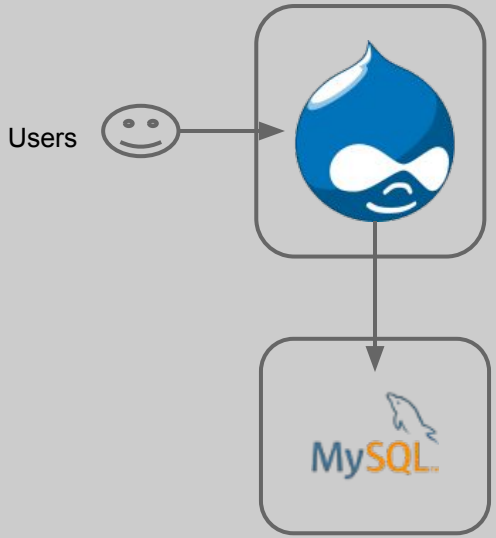
1. Simple: One Giant SPOF

Running on a single machine: any failure of any component takes down the entire site



2. Better Scale, 2 x SPOF

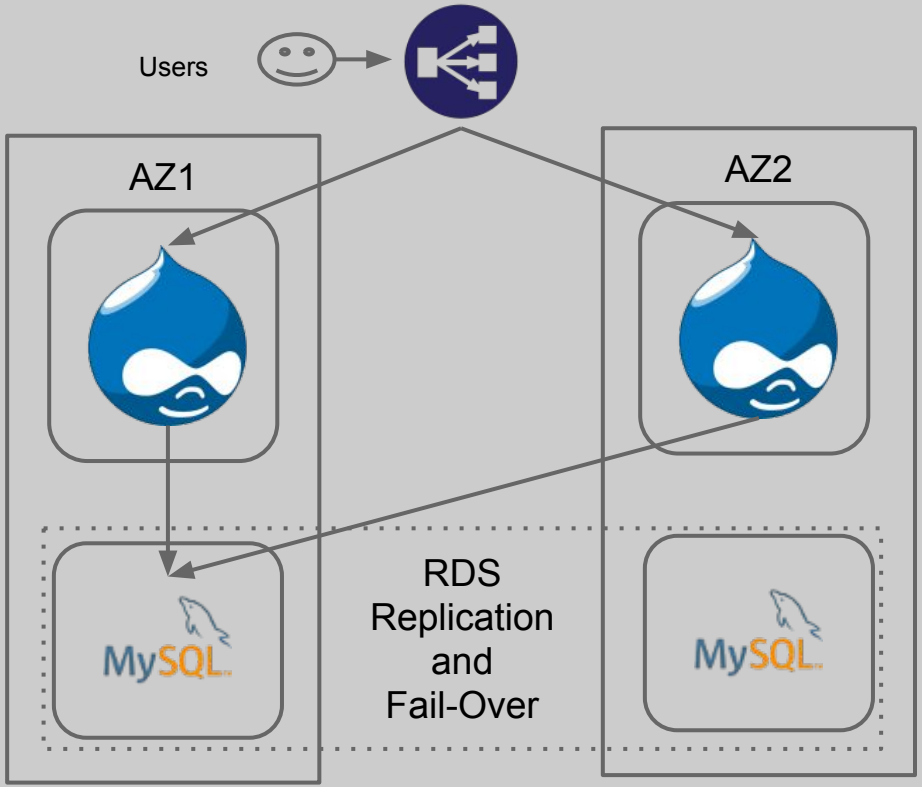
Running on two machines: any failure of any component still takes down the entire site



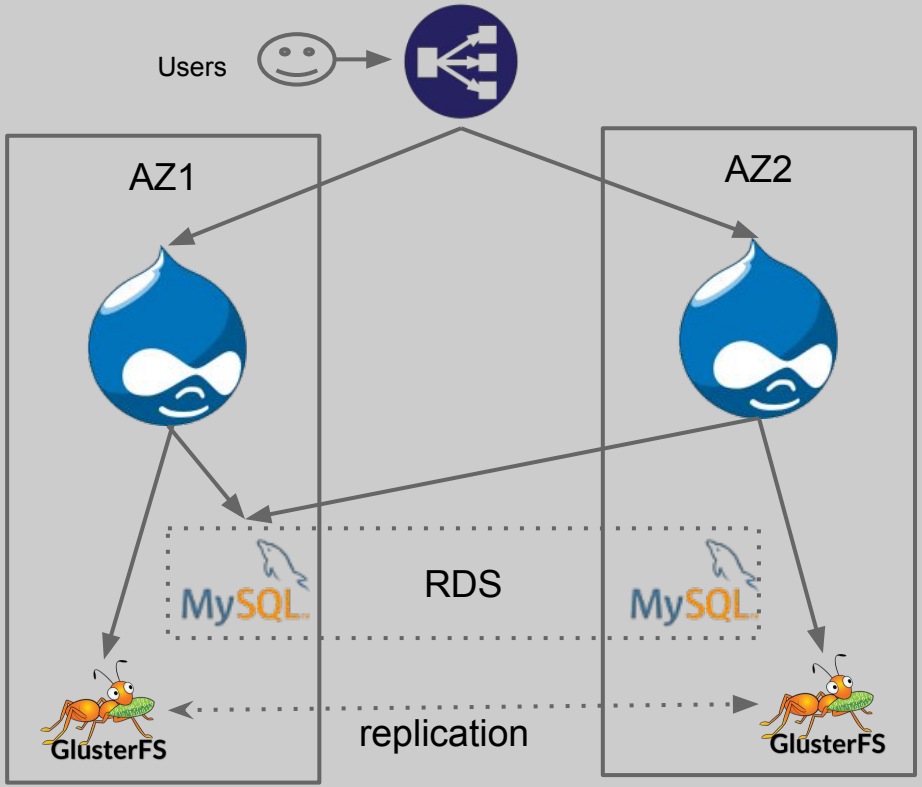
3. Fault Tolerance

Each AZ is physically-distinct Amazon datacenter; they're connected by high-speed networks.

This is better, but won't work as-is because /files is local to each Drupal instance



4. Distributed /files with Gluster

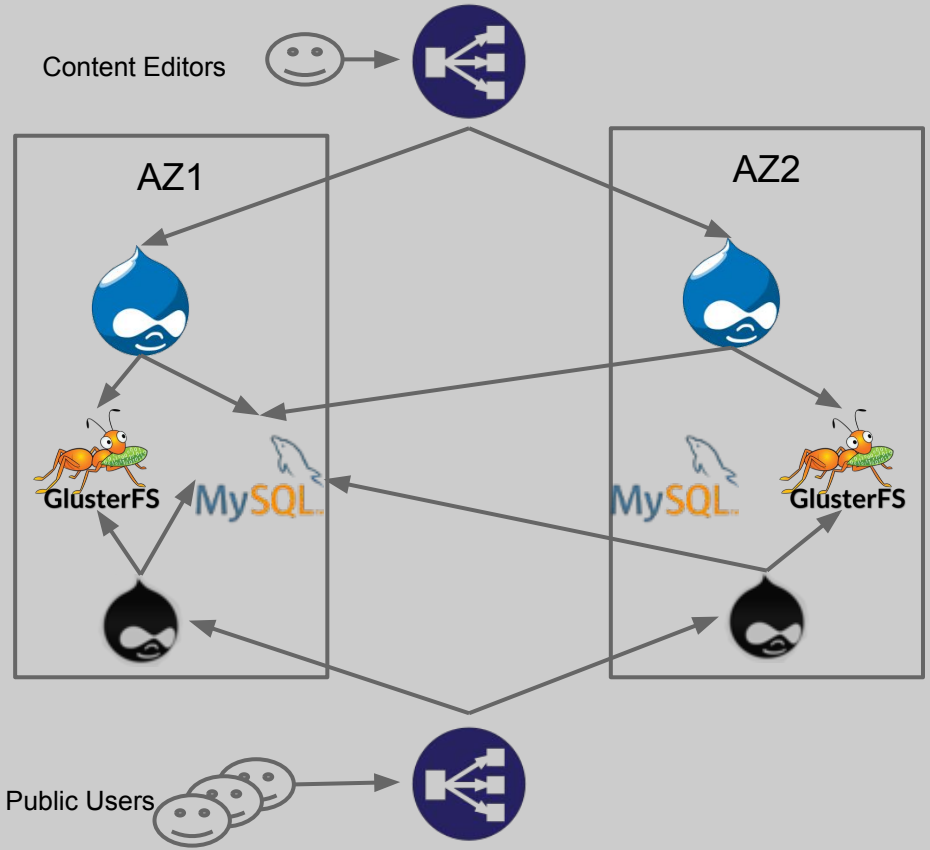


GlusterFS replicates /files so both Drupals have same media, just like the replicated DB

The server box outlines have been removed for clarity; each service runs on its own virtual box.

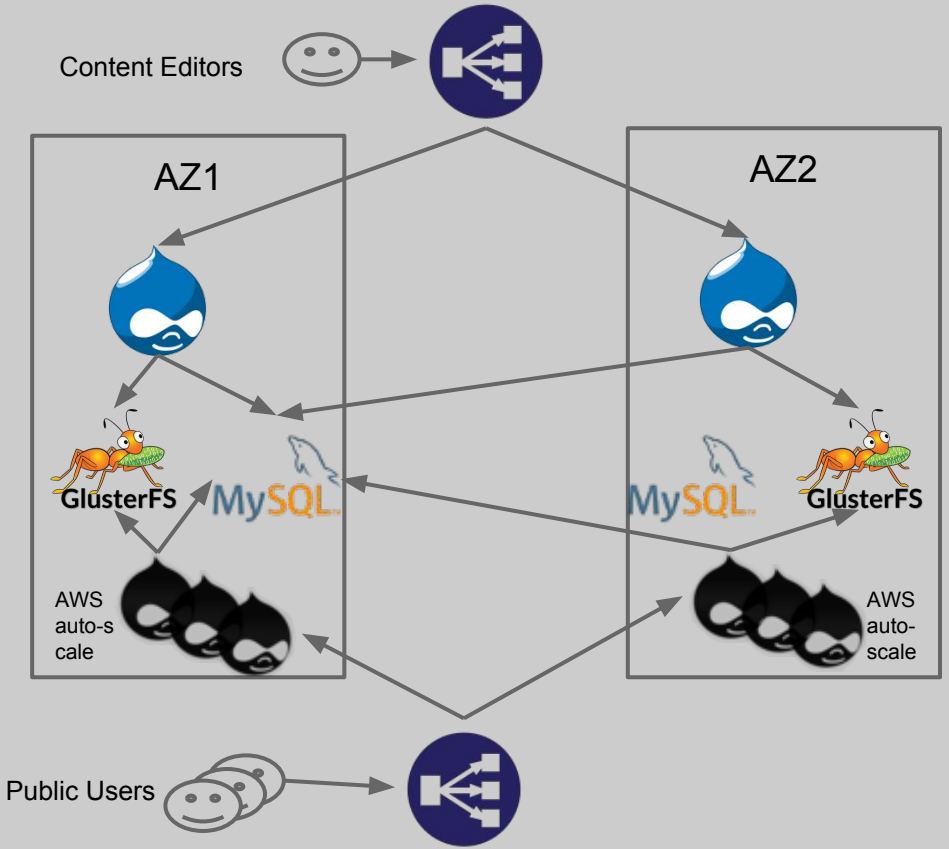
5. Improve Security: Read-Only Public

Public-facing Drupals mount the Gluster filesystem and access the database through read-only connections. Even if the boxes are compromised the site data cannot be changed.



The replication lines have been removed for clarity.

6. Auto-Scale

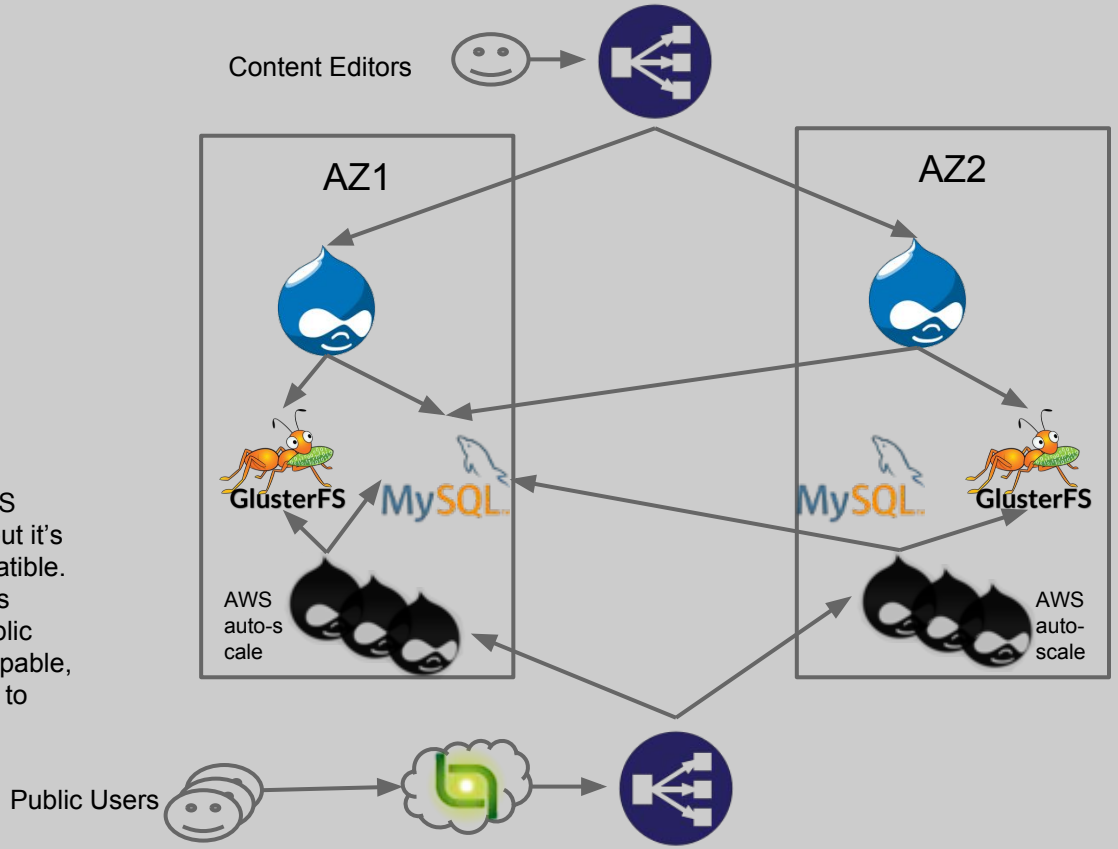


Auto-scaling lets us deploy on small-ish boxes, but respond to spikes in load by spinning up new instances; they boot from “baked” Amazon Machine Images. This saves money, rather than deploying on the biggest hardware you’ll ever need to handle a crush.

Not shown: a writable Drupal for POSTs, e.g.: forms

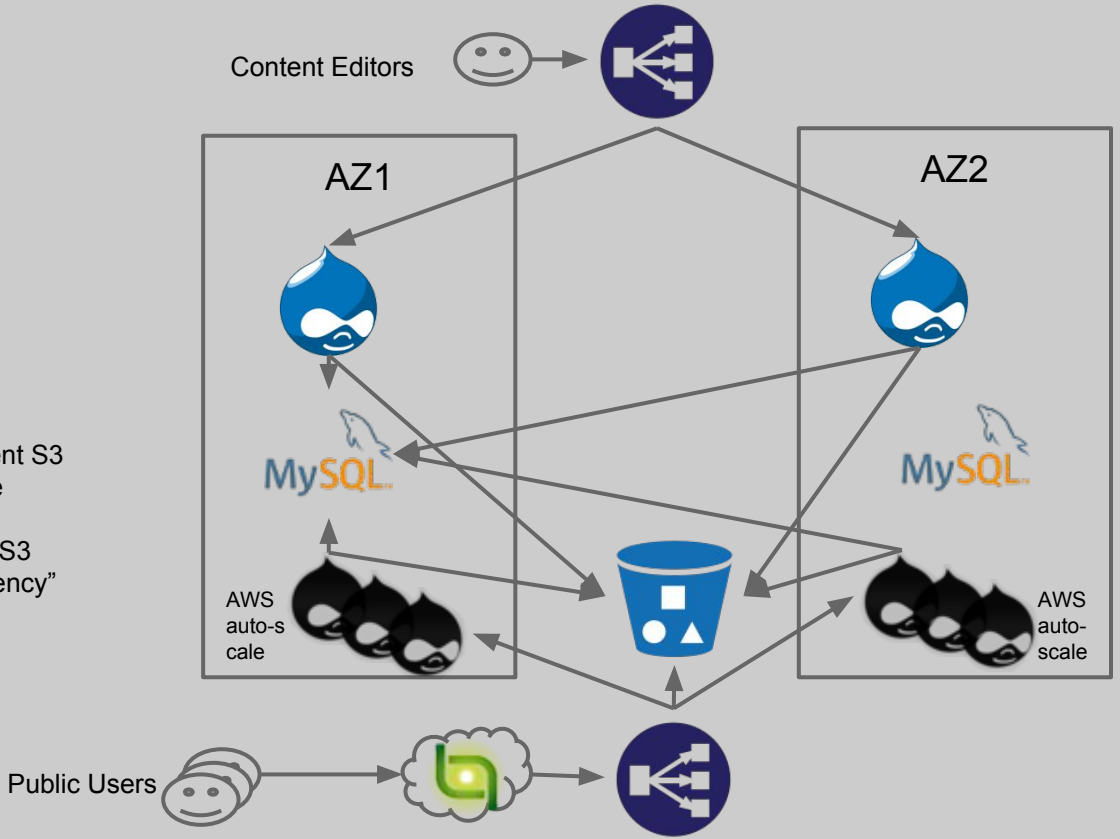
7. CDN

science.nasa.gov currently uses AWS CloudFront CDN but it's not yet IPv6 compatible. Federal regulations require all new public sites to be IPv6 capable, so we're switching to LimeLight.



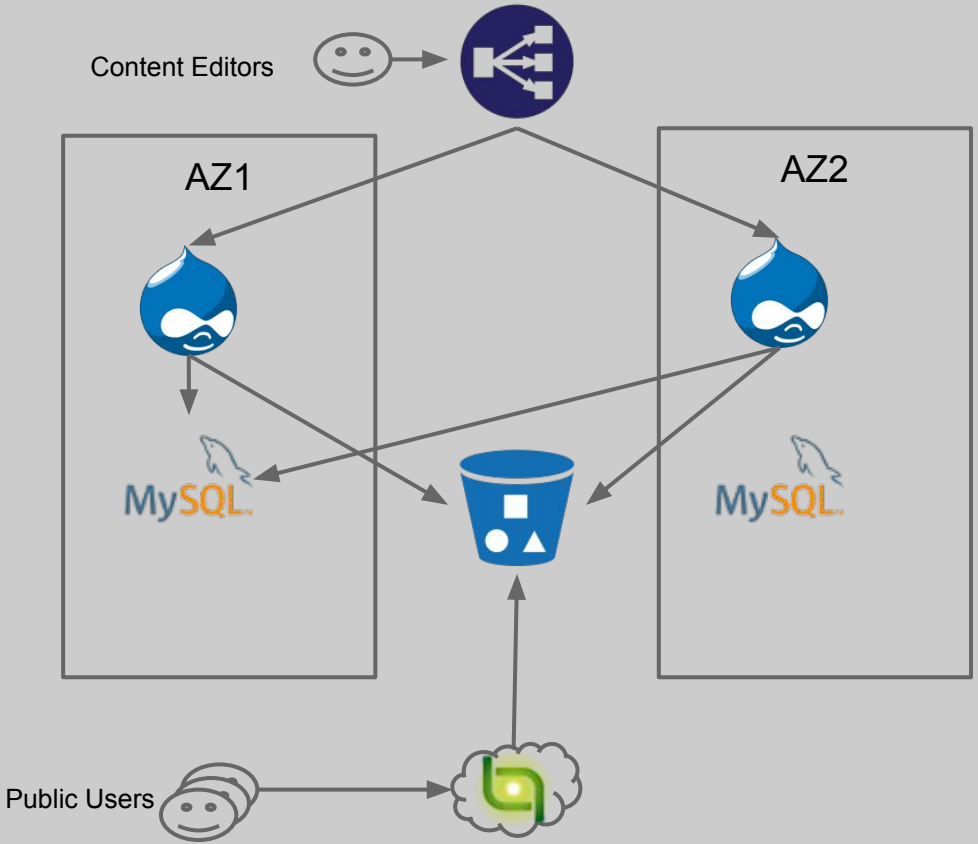
8. Extra-Credit: /files on S3?

Drupal vs. S3:
there are 3 different S3 modules but none seem ready for production. AWS S3 “eventual consistency” is problematic



9. Extra-Credit: Publish Pages to S3?

Drupal would have to publish media and Page content to S3 on every change. Publishing pages that depend on these could be very difficult. Pathological case: footer -- all pages would have to be re-pushed. Slow to push to S3.



Wanna get involved?

We’re looking to add some Drupal chops to our team.