Multiple Generator Failures Caused Amazon Outage

MANAGE > UPTIME

Multiple Generator Failures Caused Amazon Outage

Amazon Web Services says that the failure of multiple generators in a single data center caused last Friday night's power outage, which led to downtime for Netflix, Instagram and many other popular web sites.

Rich Miller | Jul 03, 2012

Amazon Web Services said last week's service outage for its EC2 cloud computing service was caused by generator failures at a single facility. (Photo by BCP via Flickr.

Amazon Web Services says that the repeated failure of multiple generators in a single data center caused last Friday night's power outage, which led to downtime for Netflix, Instagram and many other popular web sites. The generators in this facility failed to operate properly during two utility outages over a short period Friday evening when the site lost utility power, depleting the emergency power in the uninterruptible power supply (UPS) systems.

Amazon said the data center outage affected a small percentage of its operations, but was exacerbated by problems with systems that allow customers to spread workloads across multiple data centers. The company apologized for the outage and outlined the steps it will take to address the problems and prevent a recurrence.

The generator failures in the June 29 incident came just two weeks after a June 14 outage that was caused by a series of problems with generators and electrical switching equipment.

Just 7 Percent of Instances Affected

The incident affected only one availability zone within its US-East-1 region, and that only 7 percent of instances were offline. It did not identify the location of the data center, but said it was one of 10 facilities serving the US-East-1 region.

When the UPS units ran out of power at 8:04 p.m., the data center was left without power. Shortly afterward, Amazon staffers were able to manually start the generators, and power was restored at 8:24 p.m. Although the servers lost power for only 20 minutes, recovery took much longer. "The vast majority of these instances came back online between 11:15pm PDT and just after midnight," Amazon said in its incident report.

Amazon said a bug in its Elastic Load Balancing (ELB) system that prevented customers from quickly shifting workloads to other availability zones. This had the affect of magnifying the impact of the outage, as customers that normally use more than one availability zone to improve their reliability (such as Netflix) were unable to shift capacity.

Amazon: We Tested & Maintained The Generators

Amazon said the generators and electrical switching equipment that failed were all the same brand and all installed in late 2010 and early 2011, and had been tested regularly and rigorously maintained. "The generators and electrical equipment in this datacenter are less than two years old, maintained by manufacturer representatives to manufacturer standards, and tested weekly. In addition, these generators operated flawlessly, once brought online Friday night, for just over 30 hours until utility power was restored to this datacenter. The equipment will be repaired, recertified by the manufacturer, and retested at full load onsite or it will be replaced entirely."

In the meantime, Amazon said it would adjust several settings in the process that switches the electrical load to the generators, making it easier to transfer power in the event the generators start slowly or experience uneven power quality as they come online. The company will also have additional staff available to start the generators manually if needed.

Amazon also addressed why the power outage was so widely felt, even though it apparently affected just 7 percent of virtual machine instances in the US-East-1 region.

Though the resources in this datacenter … represent a single-digit percentage of the total resources in the US East-1 Region, there was significant impact to many customers. The impact manifested in two forms. The first was the unavailability of instances and volumes running in the affected datacenter. This kind of impact was limited to the affected Availability Zone. Other Availability Zones in the US East-1 Region continued functioning normally. The second form of impact was degradation of service "control planes" which allow customers to take action and create, remove, or change resources across the Region. While control planes aren't required for the ongoing use of resources, they are particularly useful in outages where customers are trying to react to the loss of resources in one Availability Zone by moving to another.

Load Balancing Bug Limited Workload Shifts

The incident report provides extensive details on the outage's impact on control planes for its EC2 compute service, Elastic Block Storage (EBS) services and Relational Database Service (RDS). Of particular interest is Amazon's explanation of the issues affecting its Elastic Load Balancing (ELB) service.

The ELB service is important because it is widely used to improve customer reliability, allowing them to shift capacity between different availability zones, an important strategy in preserving uptime when a single data center experiences problems. Here's a key excerpt from Amazon's incident report regarding the issues with ELB on the June 29 outage.

During the disruption this past Friday night, the control plane (which encompasses calls to add a new ELB, scale an ELB, add EC2 instances to an ELB, and remove traffic from ELBs) began performing traffic shifts to account for the loss of load balancers in the affected Availability Zone. As the power and systems returned, a large number of ELBs came up in a state which triggered a bug we hadn't seen before. The bug caused the ELB control plane to attempt to scale these ELBs to larger ELB instance sizes. This resulted in a sudden flood of requests which began to backlog the control plane. At the same time, customers began launching new EC2 instances to replace capacity lost in the impacted Availability Zone, requesting the instances be added to existing load balancers in the other zones. These requests further increased the ELB control plane backlog. Because the ELB control plane currently manages requests for the US East-1 Region through a shared queue, it fell increasingly behind in processing these requests; and pretty soon, these requests started taking a very long time to complete.

While direct impact was limited to those ELBs which had failed in the power-affected datacenter and hadn't yet had their traffic shifted, the ELB service's inability to quickly process new requests delayed recovery for many customers who were replacing lost EC2 capacity by launching new instances in other Availability Zones. For multi-Availability Zone ELBs, if a client attempted to connect to an ELB in a healthy Availability Zone, it succeeded. If a client attempted to connect to an ELB in the impacted Availability Zone and didn't retry using one of the alternate IP addresses returned, it would fail to connect until the backlogged traffic shift occurred and it issued a new DNS query. As mentioned, many modern web browsers perform multiple attempts when given multiple IP addresses; but many clients, especially game consoles and other consumer electronics, only use one IP address returned from the DNS query.

Hide comments

Comment *

Join the discussion...

PublishLog in or register to comment

SAVE

Leave this field blank

Related

How the Pandemic Has Forced a New Normal in the Data Center
Jun 26, 2020

Veeam Updates Backup Products to Unleash the Data
Jun 19, 2020

T-Mobile Blames Network Failure on a Fiber-Circuit Supplier
Jun 17, 2020

COVID-19: Five Recommendations for Telco IT Leaders
Jun 11, 2020
Load More

Data Center Knowledge.

About Us
Contact Us
Advertise
Sitemap
Ad Choices
CCPA: Do not sell my personal info
Privacy Policy
Terms of Service
Cookie Policy

Follow us:

in

Informa tech