关于云计算可用性的定性与定量研究

(A Qualitative and Quantitative Study on Availability of Cloud Computing)

(第七部分)

陈怀临, 弯曲评论创办人
北极光创投投资顾问, 云基地中云网技术顾问
Email: huailin@gmail.com

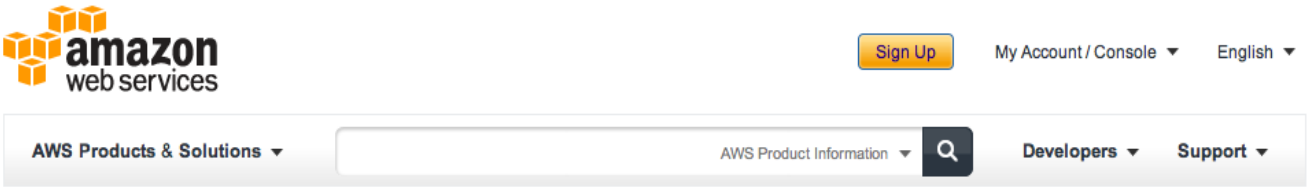
4 案例研究--亚马逊AWS

4.4 Amazon AWS服务宕机调查 (2011)

2011年, Amazon主站点和AWS比较明显的服务宕机事故大约为3次. 可计算的总共受到的影响时间大致为144个小时. 其中EBS子系统的事故影响很严重.

23. April 21, 2011

2011年4月21日, AWS遭受到了有史以来最大的一次宕机事故. 东部的一个数据中心的EC2和RDS服务突然全面无法提供服务.导致大批客户在线服务停顿, 例如,著名的foursquare等.
事故原因: EC2
事故恢复:4天
事故解释: EBS子系统崩溃导致EC2服务全面失效.
相关URL: Awazon的官方解释 http://aws.amazon.com/message/65648/



Summary of the Amazon EC2 and Amazon RDS Service Disruption in the US East Region

April 29, 2011

Now that we have fully restored functionality to all affected services, we would like to share more details with our customers about the events that occurred with the Amazon Elastic Compute Cloud ("EC2") last week, our efforts to restore the services, and what we are doing to prevent this sort of issue from happening again. We are very aware that many of our customers were significantly impacted by this event, and as with any significant service issue, our intention is to share the details of what happened and how we will improve the service for our customers.

The issues affecting EC2 customers last week primarily involved a subset of the Amazon Elastic Block Store ("EBS") volumes in a single Availability Zone within the US East Region that became unable to service read and write operations. In this document, we will refer to these as "stuck" volumes. This caused instances trying to use these affected volumes to also get "stuck" when they attempted to read or write to them. In order to restore these volumes and stabilize the EBS cluster in that Availability Zone, we disabled all control APIs (e.g. Create Volume, Attach Volume, Detach Volume, and Create Snapshot) for EBS in the affected Availability Zone for much of the duration of the event. For two periods during the first day of the issue, the degraded EBS cluster affected the EBS APIs and caused high error rates and latencies for EBS calls to these APIs across the entire US East Region. As with any complicated operational issue, this one was caused by several root causes interacting with one another and therefore gives us many opportunities to protect the service against any similar event reoccurring.

注:Amazon IaaS EBS子系统在这次事故之后, 又频频出事. EBS为了可靠性加大了系统的复杂性，结果是Amazon IaaS里最不可靠的元素。教训极大.

24. August 06, 2011

AWS在欧洲爱尔兰的数据中心遭遇雷电. 变压器起火从而数据中心的电力系统崩溃.
事故原因: EC2, EBS
事故恢复:48小时
事故解释: 电力失效.
相关URL:
http://www.datacenterknowledge.com/archives/2011/08/07/lightning-in-dublin-knocks-amazon-microsoft-data-centers-offline/

Summary: Amazon EU datacenter in Dublin outage due to lightning

Number Affected: Unknown
Organization: Amazon
- Affected Service: EC2

Reported Date: 2011-08-07
Occurred Date: 2011-08-07

注: 8月11日, 爱尔兰当地的电力公司声称事故原因不是变压器的问题.

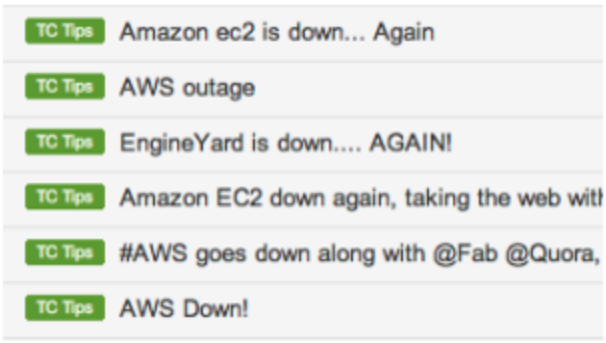
25. August 08, 2011

AWS在美国东部北弗吉尼亚的数据中心发生宕机事件. 影响了许多著名的互联网公司的正常服务,其中包括Reddit, Quora, Netflix 和 FourSquare等.
事故原因: EC2, EBS
事故恢复:25分钟
事故解释: 网络故障
相关URL: http://techcrunch.com/2011/08/08/amazon-ec2-outage/

Down Goes The Internet... Again. Amazon EC2 Outage Takes Down Foursquare, Instagram, Quora, Reddit, Etc

MG SIEGLER

Monday, August 8th, 2011 **104 Comments**



Are you trying to use the web right now? Just stop. It's largely broken.

As indicated by about 20 tips in the last few minutes and pretty much all of Twitter, Amazon's EC2 service appears to be down. That means services like Reddit, Heroku, Foursquare, Instagram, Fab, Quora, Turntable.fm, Netflix and many, many others are down.

If this sounds familiar, it's because **it just happened this past April**. So far, it looks like the outage has been **going on for about 30 minutes**.

Update: It looks like the outage may be **isolated to EAST-1**, so not all of EC2. Still, all of the companies above and hundreds of others are clearly affected right now.

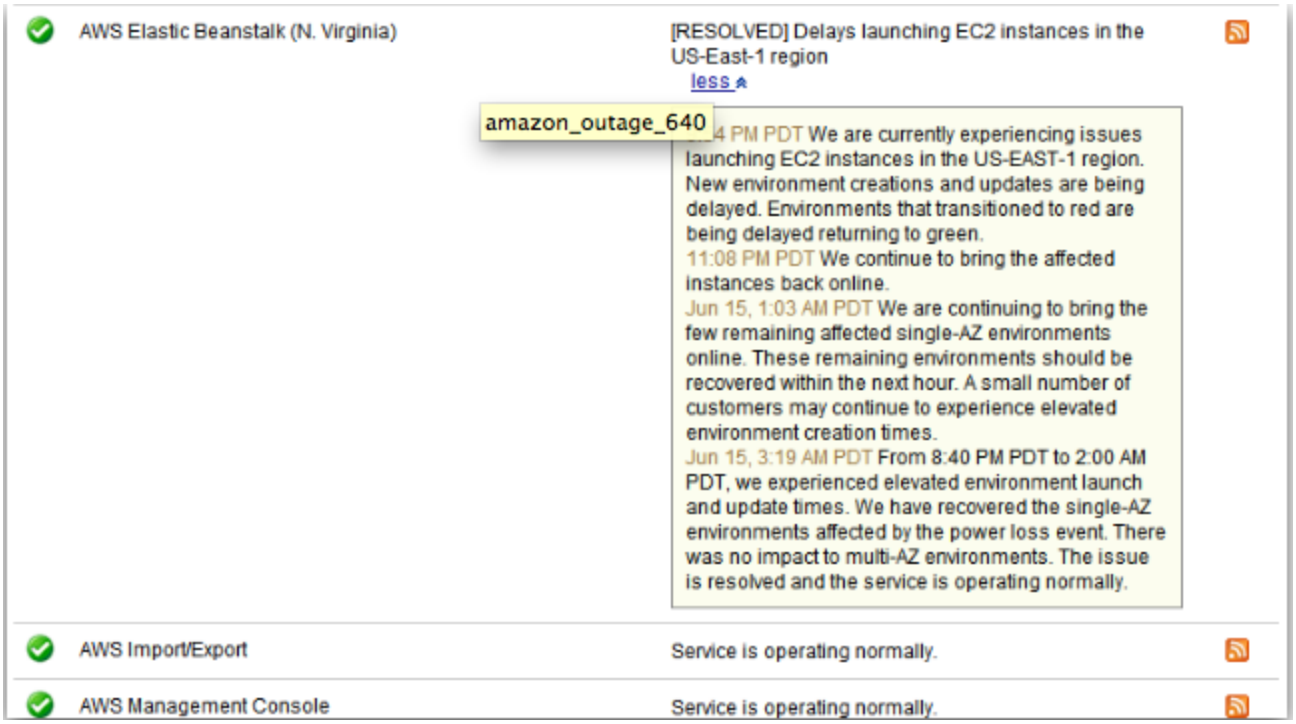
Update 2: And after roughly 40 minutes of downtime, the Internet appears to be coming back online. Amazon's **status site** confirms that it's being resolved.

4.5 Amazon AWS服务宕机调查 (2012)

2012年, Amazon主站点和AWS比较明显的服务宕机事故大约为4次. 可计算的总共受到的影响时间大致为41个小时. 除了再次发生了EBS子系统崩溃之外, 还第一次发生了ELB负载均衡系统的宕机事件.

26. June 15, 2012

AWS在美国东部北弗吉尼亚的数据中心发生宕机事件.
事故原因: EC2, EBS
事故恢复:6小时
事故解释: 电源故障
相关URL: http://www.wired.com/insights/2012/06/amazon-outage-pilot-error/



27. June 29, 2012

AWS在美国东部北弗吉尼亚的数据中心再次发生宕机事件. NetFlix, Instagram和Pinterest的服务都受到影响.
事故原因: EC2
事故恢复:6.75小时

事故解释：雷电
相关URL:
http://www.forbes.com/sites/anthonykosner/2012/06/30/amazon-cloud-goes-down-friday-night-taking-netflix-instagram-and-pinterest-with-it/

Friday Night, Taking Netflix, Instagram And Pinterest With It

As of 11:21 PM EST Amazon's Elastic Compute Cloud in North Virginia went down, due to severe thunder storms in the area. The Washington Post reports torrential rains, "scary winds," lightning and massive power outages in the D.C. area.

Amazon的官方解释为：http://aws.amazon.com/message/67457/

Summary of the AWS Service Event in the US East Region

July 2, 2012

We'd like to share more about the service disruption which occurred last Friday night, June 29th, in one of our Availability Zones in the US East-1 Region. The event was triggered during a large scale electrical storm which swept through the Northern Virginia area. We regret the problems experienced by customers affected by the disruption and, in addition to giving more detail, also wanted to provide information on actions we'll be taking to mitigate these issues in the future.

Our US East-1 Region consists of more than 10 datacenters structured into multiple Availability Zones. These Availability Zones are in distinct physical locations and are engineered to isolate failure from each other. Last Friday, due to weather warnings of the approaching storm, all change activity in the US East-1 Region had been cancelled and extra personnel had been called into the datacenters for the evening.

On Friday night, as the storm progressed, several US East-1 datacenters in Availability Zones which would remain unaffected by events that evening saw utility power fluctuations. Backup systems in those datacenters responded as designed, resulting in no loss of power or customer impact. At 7:24pm PDT, a large voltage spike was experienced by the electrical switching equipment in two of the US East-1 datacenters supporting a single Availability Zone. All utility electrical switches in both datacenters initiated transfer to generator power. In one of the datacenters, the transfer completed without incident. In the other, the generators started successfully, but each generator independently failed to provide stable voltage as they were brought into service. As a result, the generators did not pick up the load and servers operated without interruption during this period on the Uninterruptable Power Supply ("UPS") units. Shortly thereafter, utility power was restored and our datacenter personnel transferred the datacenter back to utility power. The utility power in the Region failed a second time at 7:57pm PDT. Again, all rooms of this one facility failed to successfully transfer to generator power while all of our other datacenters in the Region continued to operate without customer impact.

28. Oct 22, 2012

圣诞夜, AWS在美国东部北弗吉尼亚的数据中心再次发生宕机事件.
事故原因: EC2, EBS
事故恢复:4.25小时
事故解释: EBS子系统崩溃
相关URL:
http://techcrunch.com/2012/10/27/amazon-web-services-outage-caused-by-memory-leak-and-failure-in-monitoring-alarm/
AWS的官方解释: https://aws.amazon.com/message/680342/

Amazon Web Services Outage Caused By Memory Leak And Failure In Monitoring Alarm

ALEX WILLIAMS

Saturday, October 27th, 2012 **27 Comments**

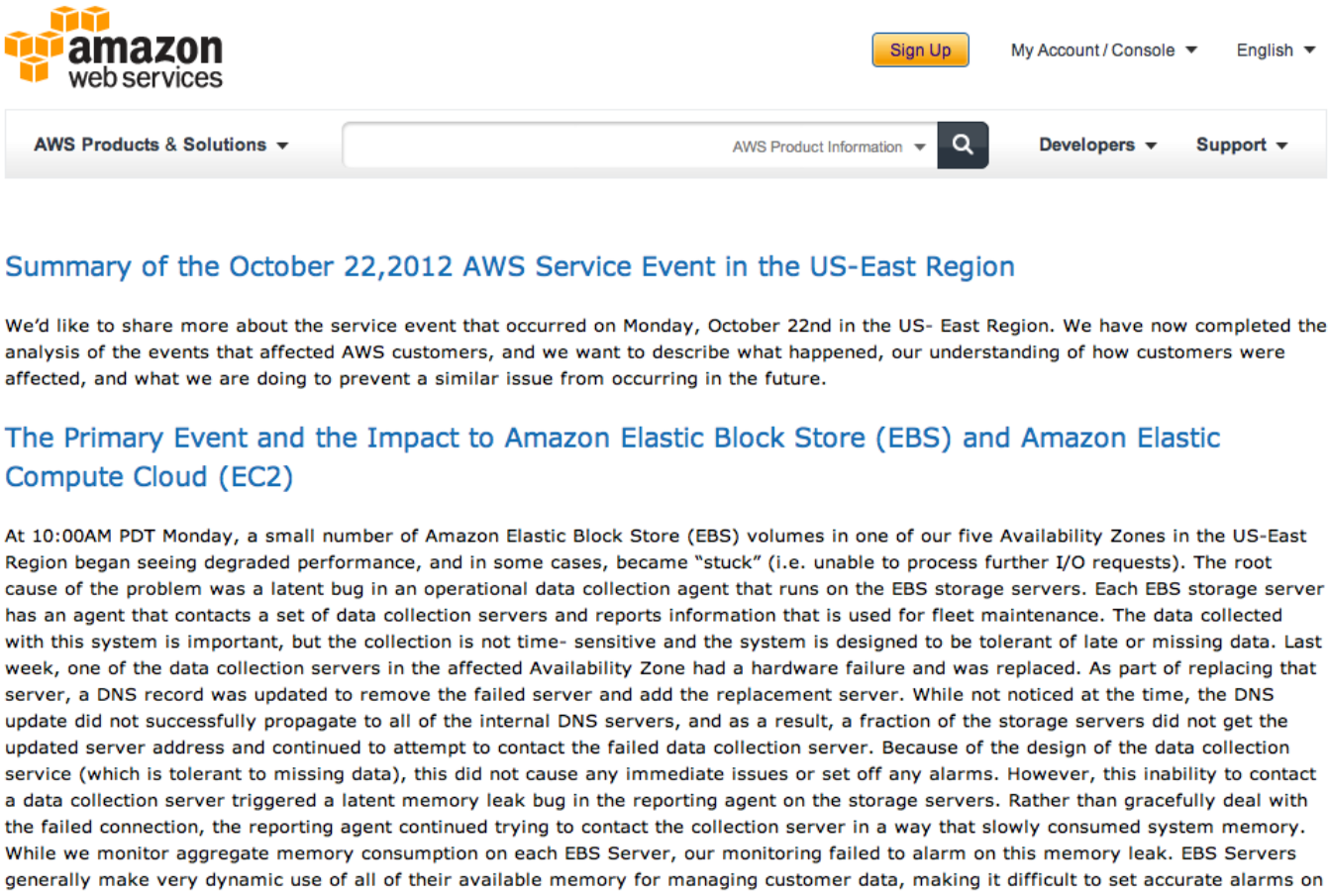
A memory leak and a failed monitoring system caused the **Amazon Web Services outage on Monday** that took out Reddit and other major services.

According to a **post** Friday night, AWS explained that the problem arose after a simple replacement of a data collection server. After installation, the server did not propagate its DNS address correctly and so a fraction of servers did not get the message. Those servers kept trying to reach the server, which led to a memory leak that then went out of control due to the failure of an internal monitoring alarm. Eventually the system ground to a virtual stop and millions of customers felt the pain.

AWS:

> *By Monday morning, the rate of memory loss became quite high and consumed enough memory on the affected storage servers that they were unable to keep up with normal request handling processes.*

The failure in its North Virginia region eventually interrupted Reddit, Foursquare, Minecraft, Heroku, GitHub, imgur, Pocket, HipChat, Coursera and a number of others.



29. Dec 24, 2012

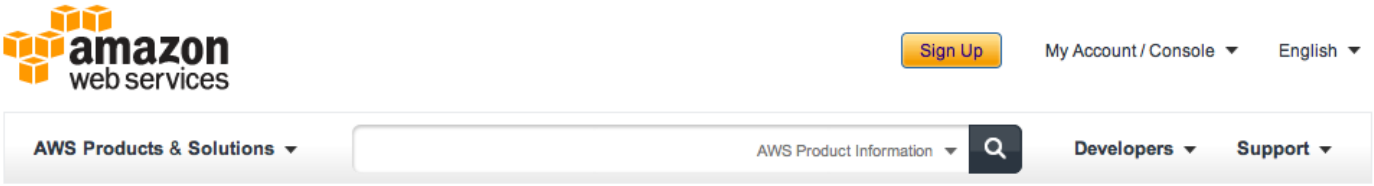
AWS在美国东部北弗吉尼亚的数据中心再次发生宕机事件.
事故原因: ELB
事故恢复:24小时
事故解释: ELB子系统
相关URL:
http://techcrunch.com/2012/12/24/netflix-crippled-on-christmas-eve-by-aws-outages/
AWS的官方解释: http://aws.amazon.com/message/680587/

Updated: Netflix Crippled On Christmas Eve By AWS Outages

CATHERINE SHU

Monday, December 24th, 2012 **234 Comments**

'Twas the night before Christmas, when all through the house, not a creature was stirring, not even those in the lighthearted holiday fare you were planning to watch on Netflix, which has been down since 1pm PST for some customers. Netflix confirmed the outage with a tweet on its official channel, though Netflix Cloud Architect Adrian Cockcroft said on Twitter that the service is still working on some devices.



Summary of the December 24, 2012 Amazon ELB Service Event in the US-East Region

We would like to share more details with our customers about the event that occurred with the Amazon Elastic Load Balancing Service ("ELB") earlier this week in the US-East Region. While the service disruption only affected applications using the ELB service (and only a fraction of the ELB load balancers were affected), the impacted load balancers saw significant impact for a prolonged period of time.

The service disruption began at 12:24 PM PST on December 24th when a portion of the ELB state data was logically deleted. This data is used and maintained by the ELB control plane to manage the configuration of the ELB load balancers in the region (for example tracking all the backend hosts to which traffic should be routed by each load balancer). The data was deleted by a maintenance process that was inadvertently run against the production ELB state data. This process was run by one of a very small number of developers who have access to this production environment. Unfortunately, the developer did not realize the mistake at the time. After this data was deleted, the ELB control plane began experiencing high latency and error rates for API calls to manage ELB load balancers. In this initial part of the service disruption, there was no impact to the request handling functionality of running ELB load balancers because the missing ELB state data was not integral to the basic operation of running load balancers.