

東京リージョン (AP-NORTHEAST-1) で発生した Amazon EC2 と Amazon EBS の事象概要

English | 中文 (简体) | 한국어

2019年8月28日(日本時間)更新:

最初の事象概要で言及した通り、今回のイベントは、東京リージョンの1つのアベイラビリティゾーン(AZ)の一部に影響を与えました。この影響は当該 AZ の Amazon EC2 および Amazon EBS のリソースに対するものですが、基盤としている EC2 インスタンスが影響を受けた場合には、当該 AZ の他のサービス(RDS、Redshift、ElastiCache および Workspaces 等)にも影響がありました。お客様と今回のイベントの調査をさらに進めたところ、個別のケースのいくつかで、複数のアベイラビリティゾーンで稼働していたお客様のアプリケーションにも、予期せぬ影響(例えば、Application Load Balancer を AWS Web Application Firewall やスティッキーセッションと組み合わせてご利用しているお客様の一部で、想定されるより高い割合でリクエストが Internal Server Error を返す)があったことを AWS では確認しております。AWS では、個別の問題についての詳細な情報を、影響を受けたお客様に直接、共有を行う予定です。

2019年8月25日(日本時間)初版:

日本時間 2019年8月23日 12:36 より、東京リージョン (AP-NORTHEAST-1) の単一のアベイラビリティゾーンで、オーバーヒートにより一定の割合の EC2 サーバの停止が発生しました。この結果、当該アベイラビリティゾーンの EC2 インスタンスへの影響及び EBS ボリュームのパフォーマンスの劣化が発生しました。このオーバーヒートは、影響を受けたアベイラビリティゾーンの一部の冗長化された空調設備の管理システム障害が原因です。日本時間 15:21 に冷却装置は復旧し、室温が通常状態に戻り始めました。室温が通常状態に戻ったことで、影響を受けたインスタンスの電源が回復しました。日本時間 18:30 までに影響を受けた EC2 インスタンスと EBS ボリュームの大部分は回復しました。少数の EC2 インスタンスと EBS ボリュームは、電源の喪失と過大な熱量の影響を受けたハードウェアホスト上で動作していました。これらのインスタンスとボリュームの復旧には時間がかかり、一部につきましては基盤のハードウェアの障害によりリタイアが必要でした。

EC2 インスタンスと EBS ボリュームへの影響に加えて、EC2 RunInstances API にも影響が発生しました。日本時間 2019年8月23日 13:21 に、影響の発生したアベイラビリティゾーンでの EC2 インスタンスの起動、および冪等性トークン (複数のインスタンスを起動させる危険なく、インスタンスの起動をリトライする機能) を使用して RunInstances API を東京リージョンで実行した場合に、エラー率の上昇が発生しました。その他の EC2 API や冪等性トークンを使用しない EC2 インスタンスの起動は正常に動作しておりました。この問題は、冪等性トークンを使用する Auto Scaling グループからのインスタンスの新規起動も阻害しました。日本時間 14:51 に、エンジニアは、冪等性トークンと Auto Scaling グループの問題を解決しました。影響の発生したアベイラビリティゾーンでの、新しい EC2 インスタンスの起動についてのエラー率向上は、EC2 コントロールプレーンサブシステムのリストアが完了した日本時間 16:05 まで継続しました。影響の発生した EBS ボリュームのスナップショットの作成も、この期間にエラー率の上昇が発生しました。

今回の事象はデータセンターで使用されるいくつかの冷却システムの制御と最適化に使用されるデータセンター制御システムの障害によって引き起こされました。制御システムは、高可用性を実現するために複数のホストで実行されます。この制御システムには、ファン、冷却装置、温度センサーなどのサードパーティ製との通信を可能にするサードパーティ製のコードが含まれていて、直接または組み込みプログラマブルロジックコントローラ (PLC) を介して通信し、実際のデバイスと通信します。今回の事象発生の直前に、データセンター制御システムは、制御ホスト群から制御ホストの1つを外す処理を行っていました。このようなフェイルオーバーの間、新しい制御ホストがデータセンターの最新状況を保持する為に、制御システムは、他の制御システムおよび制御するデータセンター機器 (データセンター全体の冷却装置および温度センサーなど) と情報を交換する必要があります。サードパーティ製の制御システムにおけるロジックのバグにより、この情報交換が制御システムとデータセンターのデバイス間で過度に発生し、最終的には制御システムが応答しなくなりました。AWS のデータセンターは、データセンターの制御システムに障害が発生した場合、制御システムの機能が回復するまで冷却システムが最大冷却モードになるよう設計されています。これはデータセンターのほとんどで正常に機能していましたが、データセンターのごく一部で冷却システムがこの安全な冷却構成に正しく移行できず停止しました。追加の安全策として、AWS のデータセンターオペレータは、データセンター制御システムを迂回し冷却システムを「バージ」モードにすることで故障に際しての熱風を素早く排出する能力を持っています。運用チームは、影響のあるデータセンターのエリアで冷却システムを「バージ」モードにしようとしたましたが、これも失敗しました。この時点で、データセンターの影響を受けるエリアの温度が上昇し始め、サーバーの温度が許容限度を超え、サーバーの電源が停止し始めました。データセンター制御システムが利用できなかったため、データセンターオペレータはデータセンターと冷却システムの健全性と状態に対する可視性が最小限に限定されていました。この状況を改善されるためにはオペレータは影響を受けるすべての機器を手動で調査してリセットし、最大冷却モードにする必要がありました。これらの対応時に一部の空調ユニットを制御する PLC も応答しないことが見つかりました。これらの PLC はリセットする必要があります、またこの障害によりフェォルトの冷却モードと「バージ」モードが正常に動作していないことが確認できました。これらのコントローラがリセットされると、影響のあったデータセンターのエリアで冷却が行われ室温が低下し始めました。

現在も、サードパーティのベンダーと協力し、制御システムと応答がなくなった PLC の両面を引き起こしたバグ、並びにバグによる影響の詳細な調査を進めております。並行して、事象の再発を防ぐため、バグを引き起こした制御システムのフェイルオーバーモードを無効にしました。また、もし万が一事象が再現しても対策が取れるよう、オペレータにこの度の事象の検知方法と復旧方法のトレーニングを実施しました。これにより、もし同様の事象が発生しても、お客様への影響が発生する前に、システムのリセットが可能になっております。その他にも、「バージ」モードが PLC を完全にバイパスできるよう、空調ユニットを制御する方法を変更するよう作業を進めております。これは、最新のデータセンターで使用している方法で、PLC が応答がなかった際でも「バージ」モードが機能するよう、空調ユニットにすることのための方法です。

この度の事象発生時、異なるアベイラビリティゾーンの EC2 インスタンスや EBS ボリュームへの影響はございませんでした。複数のアベイラビリティゾーンでアプリケーションを稼働させていたお客様は、事象発生中も可用性を確保できている状況でした。アプリケーションで最大の可用性を必要とされるお客様には、この複数のアベイラビリティゾーンのアーキテクチャに則ってアプリケーションを稼働させることを引き続き推奨します (お客様にとって高可用性に課題を生じ得る全てのアプリケーションのコンポーネントは、この耐障害性を実現する方法の下で稼働させることを強く推奨します)。

この度の事象により、ご迷惑をおかけしましたことを心よりお詫び申し上げます。AWS サービスがお客様ビジネスに極めて重要である点を理解しております。AWS は完全ではないオペレーションには決して満足することはありません。この度の事象から学ぶために出来得る全てのことを行い、AWS サービスの向上に努めてまいります。

Update August 28, 2019 JST:

As we mentioned in our initial summary, this event impacted a small portion of a single Availability Zone (“AZ”) in our Tokyo Region. The impact was to the Amazon EC2 and Amazon EBS resources in that AZ, though some other services (such as RDS, Redshift, ElastiCache, and Workspaces) would have seen some impact in that AZ if their underlying EC2 instances were affected. As we have further investigated this event with our customers, we have discovered a few isolated cases where customers’ applications running across multiple Availability Zones saw unexpected impact (i.e. some customers using Application Load Balancer in combination with AWS Web Application Firewall or sticky sessions, saw a higher than expected percent of requests return an Internal Server Error). We are sharing additional details on these isolated issues directly with impacted customers.

Summary of the Amazon EC2 and Amazon EBS Service Event in the Tokyo (AP-NORTHEAST-1) Region

We'd like to give you some additional information about the service disruption that occurred in the Tokyo (AP-NORTHEAST-1) Region on August 23, 2019. Beginning at 12:36 PM JST, a small percentage of EC2 servers in a single Availability Zone in the Tokyo (AP-NORTHEAST-1) Region shut down due to overheating. This resulted in impaired EC2 instances and degraded EBS volume performance for some resources in the affected area of the Availability Zone. The overheating was due to a control system failure that caused multiple, redundant cooling systems to fail in parts of the affected Availability Zone. The affected cooling systems were restored at 3:21 PM JST and temperatures in the affected areas began to return to normal. As temperatures returned to normal, power was restored to the affected instances. By 6:30 PM JST, the vast majority of affected instances and volumes had recovered. A small number of instances and volumes were hosted on hardware which was adversely affected by the loss of power and excessive heat. It took longer to recover these instances and volumes and some needed to be retired as a result of failures to the underlying hardware.

In addition to the impact to affected instances and EBS volumes, there was some impact to the EC2 RunInstances API. At 1:21 PM JST, attempts to launch new EC2 instances targeting the impacted Availability Zone and attempts to use the “idempotency token” (a feature which allows customers to retry run instance commands without risking multiple resulting instance launches) with the RunInstances API in the region began to experience error rates. Other EC2 APIs and launches that did not include an “idempotency token,” continued to operate normally. This issue also prevented new launches from Auto Scaling which depends on the “idempotency token”. At 2:51 PM JST, engineers resolved the issue affecting the “idempotency token” and Auto Scaling. Launches of new EC2 instances in the affected Availability Zone continued to fail until 4:05 PM JST, when the EC2 control plane subsystem had been restored in the impacted Availability Zone. Attempts to create new snapshots for affected EBS volumes, also experienced increased error rates during the event.

This event was caused by a failure of our datacenter control system, which is used to control and optimize the various cooling systems used in our datacenters. The control system runs on multiple hosts for high availability. This control system contains third-party code which allows it to communicate with third-party devices such as fans, chillers, and temperature sensors. It communicates either directly or through embedded Programmable Logic Controllers (PLC) which in turn communicate with the actual devices. Just prior to the event, the datacenter control system was in the process of failing away from one of the control hosts. During this kind of failover, the control system has to exchange information with other control systems and the datacenter equipment it controls (e.g., the cooling equipment and temperature sensors throughout the datacenter) to ensure that the new control host has the most up-to-date information about the state of the datacenter. Due to a bug in the third-party control system logic, this exchange resulted in excessive interactions between the control system and the devices in the datacenter which ultimately resulted in the control system becoming unresponsive. Our datacenters are designed such that if the datacenter control system fails, the cooling systems go into maximum cooling mode until the control system functionality is restored. While this worked correctly in most of the datacenter, in a small portion of the datacenter, the cooling system did not correctly transition to this safe cooling configuration and instead shut down. As an added safeguard, our datacenter operators have the ability to bypass the datacenter control systems and put our cooling system in “purge” mode to quickly exhaust hot air in the event of a malfunction. The team attempted to activate purge in the affected areas of the datacenter, but this also failed. At this point, temperatures began to rise in the affected part of the datacenter and servers began to power off when they became too hot. Because the datacenter control system was unavailable, the operations team had minimum visibility into the health and state of the datacenter cooling systems. To recover, the team had to manually investigate and reset all of the affected pieces of equipment and put them into a maximum cooling configuration. During this process, it was discovered that the PLCs controlling some of the air handling units were also unresponsive. These controllers needed to be reset. It was the failure of these PLC controllers which prevented the default cooling and “purge” mode from correctly working. After these controllers were reset, cooling was restored to the affected area of the datacenter and temperatures began to decrease.

We are still working with our third-party vendors to understand the bug, and subsequent interactions, that caused both the control system and the impacted PLCs to become unresponsive. In the interim, we have disabled the failover mode that triggered this bug on our control systems to ensure we do not have a recurrence of this issue. We have also trained our local operations teams to quickly identify and remediate this situation if it were to recur, and we are confident that we could reset the system before seeing any customer impact if a similar situation was to occur for any reason. Finally, we are working to modify the way that we control the impacted air handling units to ensure that “purge mode” is able to bypass the PLC controllers completely. This is an approach we have begun using in our newest datacenter designs and will make us even more confident that “purge mode” will work even if PLCs become unresponsive.

During this event, EC2 instances and EBS volumes in other Availability Zones in the region were not affected. Customers that were running their applications thoroughly across multiple Availability Zones were able to maintain availability throughout the event. For customers that need the highest availability for their applications, we continue to recommend running applications with this multiple Availability Zone architecture; any application component that can create availability issues for customers should run in this fault tolerant way.

We apologize for any inconvenience this event may have caused. We know how critical our services are to our customers’ businesses. We are never satisfied with operational performance that is anything less than perfect, and we will do everything we can to learn from this event and drive improvement across our services.

2019 年 8 月 28 日 (中国标准时间) 更新:

在我们一开始的总结中提到，这次事件影响了东京区中一个可用区 (“AZ”) 里的一小部分。被影响到的有 Amazon EC2 及 EBS 资源，一些使用到受损 EC2 资源作为底层硬件的其它服务 (比如 RDS、Redshift、ElastiCache 以及 Workspaces) 也受到了影响。在我们进一步和客户调查这个事件的时候，我们也看到了一些少数的例子，客户在用多个可用区的时候也受到了一定的影响 (比如一些客户在同时使用 Application Load Balancer，以及 AWS Web Application Firewall 或者粘性会话时，遇到了比预期比例还高的内部服务器错误)。我们会直接和被这些少数的受影响客户直接分享更多详细的信息。

关于 Amazon EC2 以及 Amazon EBS 在东京区域 (AP-NORTHEAST-1) 的服务事件的说明

针对在东京区域 (AP-NORTHEAST-1) 的服务中断事件，我们在这里提供更多信息。从 2019 年 8 月 23 日 11:36 AM CST (中国标准时间) 开始，一小部分的 EC2 服务器在东京 (AP-NORTHEAST-1) 区域中单一可用区 (Availability Zone) 由于服务器过热造成停机。这导致在该可用区中受到影响的 EC2 实例与 EBS 卷效能降低。造成服务器过热的原因是控制系统故障，造成受影响的可用区的部分冷却系统失效。受到影响的冷却系统已经在 2:21 PM CST (中国标准时间) 修复，服务器温度也恢复到正常状态。在温度恢复正常后，EC2 实例的电源供应也已恢复。在 5:30 PM CST (中国标准时间)，大部分受影响的 EC2 实例与 EBS 卷都恢复正常工作，但仍有一小部分的实例与卷因为过热与断电暂时无法修复，因为底层硬件的故障，其中有些实例与卷需要更多的时间进行修复。

除了 EC2 实例与 EBS 卷受到影响外，在 12:21 PM CST (中国标准时间) EC2 RunInstances API 也受到了影响。在受影响的可用区中，尝试启动新的 EC2 实例和尝试使用 RunInstances API 的 “idempotency token” 功能 (一个允许用户启动新的实例时重试而不会产生多余的实例的功能) 时，也有发生错误。其他没有调用 “idempotency token” 的 API 则可正常运转。这个事件也导致透过 “idempotency token” 使用 Auto Scaling 时，无法启动新实例。后台团队已经在 1:51 PM CST (中国标准时间) 修复了 “idempotency token” 与 Auto Scaling 相关的问题。并且于 3:05 PM CST (中国标准时间) 在受影响的可用区中，修复了 EC2 控制面板的子系统，开启新实例的功能已经可以正常工作。但在本事件中受到影响的服务所建立的新快照 (Snapshot) 依旧有一定的错误率。

本件事件是由于数据中心负责控制和优化冷却的控制系统故障所造成，这个控制系统在多个主机都有部署以实现高可用性，本控制系统中包含了允许与风扇、冷却器和温度传感器等硬件组件相互传递信号的数据三方程序，该程序可以直接或透过 Programmable Logic Controllers (PLC) 来与实际的硬件组件沟通。在这件事件发生时，数据中心的控制系统正在为了其中一台失效的控制主机进行备份处理。在备份处理中，控制系统要彼此互相交换信号 (例如：冷却装置与温度传感器交换信号) 以保持最新的信息。由于该第三方程序中的一个错误，导致控制系统与组件过度的进行信息交换而造成控制系统无法回应。我们的数据中心被设计成一旦控制系统发生错误，冷却系统就会自动进入最冷的模式，直到控制系统恢复正常为止，这样的设计对于我们大部分的数据中心都是有效的，但有一小部分的数据中心，由于冷却系统无法正确进入安全降温模式，而造成系统故障。我们的数据中心加入了安全防护设计，在控制系统故障时，可以略过控制系统，直接进入净空模式将数据中心中的热空气迅速排出，但控制中心的团队在启动净空模式时发生了故障，所以数据中心中的温度才会持续攀升，而服务器在到达温度上限后也开始自动关机了。由于数据中心的控制系统故障，运维团队无法得知数据中心冷却系统的即时信息，在进行故障排除时，团队必须要对所有组件进行逐一的人工检查，才能让控制到达安全最冷模式，在这故障排除的过程中，发现控制空调组件的 PLC 控制系统无法回应，控制器需要进行重置，是 PLC 控制器的错误造成了预设的冷却模式与净空模式无法正确动作，在 PLC 控制器被重置之后，在该区域数据中心的冷却系统就可以正常工作了，而数据中心的过高的温度也开始慢慢降低。

我们仍在与第三方供应商合作以了解导致控制系统和受影响的 PLC 无响应的错误和后续交互。在此期间，我们已经禁用在我们的控制系统上触发此错误的故障转移模式，以确保我们不会再次出现此问题。我们还培训了我们的本地运营团队，以便在发生这种情况时快速识别和修复这种情况，并且我们相信，如果再次发生类似情况，无论什么原因，我们可以在客户受影响之前重置系统。最后，我们正在努力修改我们控制受影响的空气处理单元的方式，以确保“清除模式”能够完全绕过 PLC 控制器。这是我们在最新的数据中心设计中开始使用的一种方法，即使 PLC 无响应，我们也会更加确信“清除模式”将起作用。

在这次事件中，EC2 实例以及 EBS 存储在另一区域的其它的可用于没有受到影响。同时在多个可用区上充分执行他们的应用程序的客户，在这次的事件中依然可以维持服务可用。对于需要绝对对客户的服务，我们持续建议您使用高可用性的架构设计。任何与应用程序相关的元件都应该采用这种容错设计。

针对这次事件造成的不便，我们感到十分抱歉。我们理解对于我们的用户的业务来说，稳定的服务有多么至关重要。我们也从未满足於我们所提供的响应速度以及服务品质。对于这次事件，我们将会尽一切所能从此次事件中学习并且继续不断改善我们的服务。

2019年 8月 28日 업데이트:

초기 요약 글에서 설명된 바와 같이, 이번 이벤트로 도쿄 리전 내 단일 가용 영역(AZ: Availability Zone)의 작은 부분이 영향을 받았습니다. 이는 해당 가용 영역의 Amazon EC2와 Amazon EBS 리소스에 영향을 미쳤습니다. 나란히, 기존 EC2 인스턴스 운영을 받은 경우, 해당 가용 영역의 다른 서비스(예: RDS, Redshift, ElastiCache, Workspaces)에 다소 영향이 있었을 것입니다. 고객과 함께 이번 이벤트를 자세히 조사한 결과, 여러 가용 영역에서 실행되는 고객의 애플리케이션(이후) 예상치 못한 영향을 받은 몇 가지 도메인(예: AWS 웹 애플리케이션 방화벽 또는 고정 세션과 함께 애플리케이션 로드 밸런서를 사용하는 일부 고객에서 예상보다 높은 비율로 내부 서버 오류가 발생)를 발견했습니다. AWS는 이러한 예상외인 문제에 대한 추가 세부 사항을 영향을 받은 고객과 직접 공유하고 있습니다.

도쿄(AP-NORTHEAST-1) 리전에서 발생한 아마존 EC2 및 아마존 EBS 서비스 이벤트 요약

2019년 8월 23일 도쿄(AP-NORTHEAST-1) 리전에서 발생한 서비스 장애와 관련해 추가 정보를 전달드립니다. 일본표준시간 오후 12시 36분부터 도쿄(AP-NORTHEAST-1) 리전 내 한 가용영역(AZ: Availability Zone)의 EC2 서버 중 일부가 과열로 인해 가동중단 되었습니다. 그 결과 EC2 인스턴스가 손상되고, 해당 가용영역의 영향을 받는 일부 리소스의 EBS 볼륨 성능이 저하되었습니다. 이번 과열은 제어 시스템 장애로 발생한 것으로, 지나치게 발생한 가용영역 일부에서 다수의 서버 장애로 이어졌습니다. 해당 냉각 시스템은 일본표준시간 기준 오후 3시 21분에 복구되었으며, 해당 영역의 온도도 정상을 회복하기 시작했습니다. 온도가 정상화되면서, 영향을 받은 인스턴스에 전력 공급도 복구되었습니다. 일본표준시간 기준 오후 6시 30분까지 영향을 받은 인스턴스와 볼륨 대다수가 복구되었습니다. 전력 중단과 과열로 인해 영향을 받은 하드웨어에 호스팅되어 있던 인스턴스와 볼륨은 소수였습니다. 해당 인스턴스와 볼륨을 복구하는 데 보다 오랜 시간이 걸렸으며, 일부는 기존 하드웨어의 장애로 인해 사용을 중단해야만 했습니다.

해당 인스턴스와 EBS 볼륨에 영향을 미친 것 외에도, EC2 RunInstances API에도 일부 영향이 있었습니다. 일본표준시 오후 1시 21분, 영향을 받은 가용영역 내 새로운 EC2 인스턴스 개시 시도와 리전의 RunInstances API로 “idempotency token”(고객이 여러 개의 인스턴스를 시작하는 위험 부담 없이 run instance 명령어를 제시하게 해주는 기능)을 사용하려는 시도가 오류를 보이기 시작했습니다. “idempotency token”을 포함하지 않은 다른 EC2 API와 개시는 정상적으로 작동했습니다. 이 이유는 “idempotency token”에 의존하는 오토 스케일링(Auto Scaling)의 새로운 시작을 방해했습니다. 일본표준시 기준 오후 2시 51분 엔지니어들은 “idempotency token”과 오토 스케일링을 해결했습니다. 영향을 받은 가용영역 내 새로운 EC2 인스턴스 시작은 오후 4시 5분 해당 가용영역에서 EC2 컨트롤 플레인 서비스시스템을 복구할 때까지 장애를 일으켰습니다. 영향을 받은 EBS 볼륨에 새로운 스냅샷을 생성하고자 한 시도 역시 해당 이벤트 동안 오류율이 증가했습니다.

이번 이벤트는 데이터센터에서 사용하는 다양한 냉각 시스템을 제어하고 최적화에 사용하는 데이터센터 컨트롤 시스템의 장애로 발생하게 되었습니다. 해당 컨트롤 시스템은 고가용성 확보를 위해 다수의 호스트 상에서 운영됩니다. 그리고, 컨트롤 시스템은 팬, 냉각장치, 온도 센서 등 제3자 장치와 커뮤니케이션을 할 수 있도록 제3자 코드를 가지고 있습니다. 컨트롤 시스템은 직접적으로 또는 실제 장치와 커뮤니케이션을 하는 임베디드 PLC(Programmable Logic Controller)를 통해 커뮤니케이션을 합니다. 이번 이벤트가 발생하기 바로 전, 해당 데이터센터 컨트롤 시스템을 호스트 중 하나에서 장애를 일으키고 있었습니다. 이러한 종류의 페일오버(failover)가 발생하게 되면, 해당 컨트롤 시스템은 새로운 컨트롤 호스트가 데이터센터 상태에 대한 가장 최신 정보를 갖도록 다른 컨트롤 시스템 및 컨트롤러는 데이터센터 장비에: 데이터센터 전반에 설치된 냉각 장비와 온도 센서)와 정보를 교환해야만 합니다. 제3자 컨트롤 시스템 로직 내 버그로 인해 예상치 못한 정보 교환은 데이터센터의 컨트롤 시스템과 장치 간 과도한 인터랙션이 발생하게 되었고, 이로 인해 컨트롤 시스템이 무응답 상태가 되었습니다. AWS의 데이터센터는 데이터센터 컨트롤 시스템이 장애를 일으키게 되면, 컨트롤 시스템 기능이 복구될 때까지 냉각 시스템이 최대 냉각 모드도 작동하도록 설계되었습니다. 이러한 설계가 대부분 데이터센터에서는 제대로 작동했지만, 일부에서 냉각 시스템이 안전한 냉각 설정으로 전환하지 못하고 가동 중단되었습니다. 추가 안전장치로, 데이터센터 운영자들은 오작동 상황에서 데이터센터 컨트롤 시스템을 우회해서 냉각 시스템을 “퍼지(purge)” 모드로 두고 재빨리 뜨거운 공기를 빼낼 수 있습니다. 운영팀에서는 문제가 발생한 데이터센터 구역에 퍼지 가열장치를 시도했지만 실패했습니다. 이 때 해당 부분의 데이터센터 온도가 오르기 시작했고, 지나치게 과열된 공간이 차단되었습니다. 데이터센터 컨트롤 시스템을 사용할 수 없었기 때문에, 운영팀에서는 데이터센터 냉각 시스템 상태에 대해 최소한의 가시성만 확보할 수 있었습니다. 복구를 위해 운영팀에서는 데이터센터 장비 전체를 조사하고 재설정해 최대 냉각 설정 상태로 놓아야 했습니다. 이 과정에서 일부 공기조화장치를 제어하는 PLC 역시 무응답 상태라는 것을 알게 되었습니다. 그래서 PLC 역시 재설정이 필요했습니다. 이 PLC 컨트롤러가 장애를 일으켜 디폴트 냉각 및 “퍼지” 모드가 작동을 하지 못했습니다. 해당 컨트롤러를 재설정할 때에는 데이터센터 이벤트 발생 구역의 냉각이 복구되었으며, 온도가 떨어지기 시작했습니다.

AWS는 아직 제3자 벤더들과 함께 컨트롤 시스템과 해당 PLC의 무응답을 일으킨 버그와 추가 인터랙션을 파악하는 중에 있습니다. 이번 방편으로 해당 이벤트가 재발하는 것을 방지하고자 컨트롤 시스템의 버그를 일일히 패치하며 모니터링과 상태로 만들었습니다. AWS는 만약 재발하게 될 경우 상황을 재빨리 파악하고, 해결할 수 있도록 현재 운영팀도 교육했습니다. 이번 사유로 비슷한 상황이 발생하게 되면 고객 영향이 발생하기 전에 시스템을 재설정할 수 있을 것으로 확신하고 있습니다. 마지막으로, “퍼지 모드”가 PLC 컨트롤러를 완전히 우회할 수 있도록 문제를 일으킨 공기조화장치 제어 방식을 변경 작업을 하고 있습니다. 이 방식은 최신 데이터센터 설계에 적용하기 시작한 것으로, PLC 무응답 상태에도 “퍼지 모드”가 보다 확실히 작동하도록 할 것입니다.

해당 이벤트 동안 리전의 다른 가용영역에 있는 EC2 인스턴스와 EBS 볼륨은 영향을 받지 않았습니다. 여러 가용영역에서 애플리케이션을 실행한 고객은 이벤트 기간 동안 가용성을 유지할 수 있었습니다. 애플리케이션이 가장 높은 가용성이 필요한 고객에게는 다중 가용영역이 추가되거나 애플리케이션을 실행할 것을 권장합니다. 고객에게 가용성 문제를 일으킬 수 있는 모든 애플리케이션 구성 요소는 이러한 내구성이 높은(fault tolerant) 방식으로 실행되어야 합니다.

了解有关 AWS 的信息

AWS 资源

AWS 上的开发人员

帮助

早9:00-晚8:00 (法定假期除外)
在线咨询 AWS 云计算专家

什么是 AWS?

入门

开发人员中心

联系我们

AWS 职位

AWS 咨询热线: 1010 0866

什么是云计算?

培训和认证

软件开发工具包与工具

AWS 职位

提交支持 TT

创建 AWS 账户

什么是开发运营?

AWS 解决方案组合

运行于 AWS 上的 .NET

提交支持 TT

提交支持 TT

什么是数据湖?

架构中心

运行于 AWS 上的 Python

运行于 AWS 上的 Python

知识中心

AWS Support 概览

什么是容器?

产品和技术常见问题

运行于 AWS 上的 Java

运行于 AWS 上的 PHP

运行于 AWS 上的 PHP

运行于 AWS 上的 JavaScript

AWS 云安全性

分析报告

运行于 AWS 上的 PHP

运行于 AWS 上的 JavaScript

运行于 AWS 上的 JavaScript

最新资讯

AWS 合作伙伴网络

运行于 AWS 上的 JavaScript

运行于 AWS 上的 JavaScript

运行于 AWS 上的 JavaScript

博客

新闻稿

Amazon 是一个倡导机会均等的雇主: 反对少数族裔、妇女、残疾人、退伍军人、性别认同和性取向歧视。