Alibaba Cloud Auto Scaling

ユーザーガイド

Document Version20190628

目次

1 使用上の注意1
1.1 ECS インスタンスのライフサイクル1
1.2 スケーリンググループのステータス2
1.3 スケーリングアクティビティのプロセス2
1.4 スケーリングアクティビティのステータス4
1.5 スケーリングアクティビティの失敗時に実行されるインスタンスのロールバック5
1.6 異常な ESC インスタンスの削除5
1.7 通知機能6
1.8 強制操作6
1.9 数量制限
1.10 考慮事項10
1.11 操作手順11
1.12 ワークフロー12
2 スケーリング設定15
2.1 スケーリング設定の作成15
3 Auto Scaling の実装16
3.1 スケジュール済みタスク16
3.1.1 スケジュール済みタスクの作成16
3.2 アラームタスク19
3.2.1 アラームタスクの表示、変更、削除19
4 Auto Scaling の管理21
4.1 スケーリングルールの変更、照会、削除21
4.2 削除ポリシー21
4.3 スケーリンググループの削除22
5 手動スケーリング24
5.1 ECS インスタンスの削除24
6 ECS インスタンスリストの照会25
7 スケーリングアクティビティの表示

1 使用上の注意

1.1 ECS インスタンスのライフサイクル

ここでは、ECS インスタンスのライフサイクル管理について説明します。

スケーリンググループに追加された ECS インスタンスには、自動作成されたインスタンスと手動 で追加したインスタンスの 2 種類があります。

自動作成された ECS インスタンス

自動作成された ECS インスタンスとは、スケーリング設定とルールに従って自動的に作成された インスタンスを指します。

このタイプの ECS インスタンスのライフサイクルは、Auto Scaling で管理されます。 スケール アップ時には ECS インスタンスが作成され、スケールダウン時には ECS インスタンスが停止お よびリリースされます。

手動で追加した ECS インスタンス

手動で追加した ECS インスタンスとは、手動でスケーリンググループに関連付けられたインスタ ンスを指します。

このタイプの ECS インスタンスのライフサイクルは、Auto Scaling で管理されません。 この ECS インスタンスが手動またはスケールダウンアクティビティによってスケーリンググループか ら削除された場合、インスタンスは Auto Scaling によって停止およびリリースされることはあ りません。

インスタンスのステータス

ライフサイクルの間、ECS インスタンスは、次に示すステータスのいずれかになります。 追加 中:

- ECS インスタンスはスケーリンググループに追加中です。具体的には、Auto Scaling により インスタンスが作成されて Server Load Balancer インスタンスに追加されているか、また は RDS のアクセスホワイトリストに追加されているところです。
- ・サービス中: ECS インスタンスはスケーリンググループに追加され、正常に機能しています。
- ・削除中: ECS インスタンスはスケーリンググループから削除中です。

インスタンスのヘルスステータス

ECS インスタンスは、次のいずれかのヘルスステータスになります。

- ・正常
- ・異常

ECS インスタンスのステータスが実行中(Running)でないとき、異常とみなされます。 異常 な ECS インスタンスは、Auto Scaling によってスケーリンググループから自動的に削除されま す。

- ・異常なインスタンスが自動作成されたものに限り、Auto Scaling によって自動的に停止およ びリリースされます。
- ・手動で追加したインスタンスは、停止およびリリースされません。

1.2 スケーリンググループのステータス

ここでは、スケーリンググループのステータスについて説明します。

スケーリンググループには、Active、Inactive、Deletingの3つのステータスがあります。

ステータス	API インジケーター
作成中	Inactive
作成済み	Inactive
有効化中	Inactive
実行中	Active
無効化中	Inactive
停止	Inactive
削除中	Deleting

1.3 スケーリングアクティビティのプロセス

ここでは、スケーリングアクティビティのプロセスを説明します。

スケーリングアクティビティのライフサイクルは、スケーリンググループのヘルスステータスと 境界条件の確認から始まり、クールダウン時間の有効化で終わります。

自動スケーリング

スケールアップ

- 1. スケーリンググループのヘルスステータスと境界条件を確認します。
- 2. アクティビティ ID の割り当て後、スケーリングアクティビティを実行します。
- 3. ECS インスタンスを作成します。
- 4. 合計容量を変更します。

- 5. 作成した ECS インスタンスに IP アドレスを割り当てます。
- 6. RDS のアクセスホワイトリストに ECS インスタンスを追加します。
- 7. ECS インスタンスを起動します。
- ECS インスタンスを Server Load Balancer にアタッチし、重みを 0 に設定します。 60 秒 間待機した後、重みを 50 に設定します。
- 9. スケーリングアクティビティが完了し、クールダウン時間を有効化します。

スケールダウン

- 1. スケーリンググループのヘルスステータスと境界条件を確認します。
- 2. アクティビティ ID の割り当て後、スケーリングアクティビティを実行します。
- Server Load Balancer が、ECS インスタンスへのトラフィックの転送を停止します。 60 秒 待機した後、Server Load Balancer から ECS インスタンスを削除します。
- 4. ECS インスタンスを無効化します。
- 5. RDS のアクセスホワイトリストから ECS インスタンスを削除します。
- 6. ECS インスタンスをリリースします。
- 7. 合計容量を変更します。
- 8. スケーリングアクティビティを完了し、クールダウン時間を有効化します。

既存 ECS インスタンスの手動での追加または削除

既存 ECS インスタンスの手動での追加

- 1. スケーリンググループのヘルスステータスと境界条件を確認します。また、ECS インスタンス のステータスとタイプを確認します。
- 2. アクティビティ ID を割り当て、スケーリングアクティビティを実行します。
- 3. ECS インスタンスを追加します。
- 4. 合計容量を変更します。
- 5. RDS のアクセスホワイトリストに ECS インスタンスを追加します。
- Server Load Balancer に ECS インスタンスをアタッチし、重みを 0 に設定します。 60 秒 間待機した後、重みを 50 に設定します。
- 7. スケーリングアクティビティを完了し、クールダウン時間を有効化します。

既存インスタンスの手動での削除

- 1. スケーリンググループのヘルスステータスと境界条件を確認します。
- 2. アクティビティ ID を割り当て、スケーリングアクティビティを実行します。
- 3. Server Load Balancer が、ECS インスタンスへのトラフィックの転送を停止します。
- 4. 60 秒待機した後、Server Load Balancer から ECS インスタンスを削除します。

- 5. RDS のアクセスホワイトリストから ECS インスタンスを削除します。
- 6. 合計容量を変更します。
- 7. スケーリンググループから ECS インスタンスを削除します。
- 8. スケーリングアクティビティを完了し、クールダウン時間を有効化します。

1.4 スケーリングアクティビティのステータス

- ここでは、Auto Scaling のスケーリングアクティビティのステータスについて説明します。 実行リクエストが拒否された場合、スケーリングアクティビティのステータスは、拒否になりま す。
- 実行中の場合、スケーリングアクティビティのステータスは、進行中になります。

スケーリングアクティビティが完了した後、ステータスは次の3つのいずれかになります。

・成功(Successful):スケーリングアクティビティが正常に実行され、スケーリングルールで指定した MaxSize 値または MinSize 値に従って、ECS インスタンスがスケーリンググループに追加、またはスケーリンググループから削除されました。

■ 注:

ECS インスタンスがスケーリンググループに正常に追加された場合は、ECS インスタンスの 作成が完了し、Server Load Balancer インスタンスと RDS のアクセスホワイトリストに追 加されています。 上記のいずれかの手順が失敗した場合、ECS インスタンスは「失敗」とみ なされます。

- ・警告(Warning): スケーリングアクティビティは実行されましたが、スケーリングルールで 指定した MaxSize 値または MinSize 値に従って、スケーリンググループに追加、またはス ケーリンググループから削除できなかった ECS インスタンスが、少なくとも1つあります。
- ・失敗(Failed):スケーリングアクティビティは失敗し、スケーリングルールで指定した MaxSize 値または MinSize 値に従って、スケーリンググループに追加、またはスケーリング グループから削除された ECS インスタンスは1つもありません。

例

スケーリングルールで、5 個の ECS インスタンスを追加すると定義します。 スケーリンググルー プの現在の合計容量は 3 個の ECS インスタンスで、MaxSize 値は 5 個の ECS インスタンスで す。 スケーリングルールが実行されると、MaxSize 値の指定に従い、ECS インスタンスが 2 個 追加されます。 このスケーリングアクティビティの完了後、ステータスは次の 3 つのいずれかに なります。

- ・成功: ECS インスタンスが2 個正常に作成され、Server Load Balancer インスタンスと RDS のアクセスホワイトリストに正しく追加されました。
- ・ 警告: 2 個の ECS インスタンスが正常に作成されましたが、Server Load Balancer インスタンスと RDS のアクセスホワイトリストに正しく追加されたのは 1 個だけです。もう 1 個は失敗したため、ロールバックされてリリースされます。
- ・ 失敗: ECS インスタンスが作成されませんでした。または、2 個の ECS インスタンスが正常 に作成されましたが、どちらも Server Load Balancer インスタンスと RDS のアクセスホワ イトリストに追加されませんでした。2 個のインスタンスは、ロールバックされてリリースさ れます。

1.5 スケーリングアクティビティの失敗時に実行されるインスタン スのロールバック

ここでは、スケーリングアクティビティの失敗時に実行されるインスタンスのロールバックにつ いて説明します。

スケーリングアクティビティがスケーリンググループに 1 つ以上の ECS インスタンスを追加でき なかった場合、失敗した ECS インスタンスはロールバックされます。 スケーリングアクティビ ティはロールバックされません。

たとえば、スケーリンググループに 20 個の ECS インスタンスがあり、そのうち 19 個のイン スタンスが Server Load Balancer インスタンスに追加された場合、追加できなかった 1 つの ECS インスタンスのみが自動的にリリースされます。

Auto Scaling では、Alibaba Cloud の Resource Access Management (RAM) サービスを使用し、ECS Open API を介して ECS インスタンスのリソースを調整します。 したがって、API 使用料が適用されます。

1.6 異常な ESC インスタンスの削除

本ページでは、異常な ECS インスタンスを削除する方法を紹介します。

ECS インスタンスがスケーリンググループに正常に追加されると、インスタンスのステータスは Auto Scaling サービスに定期的にスキャンされます。 Auto Scaling では、実行中 (Running)) ステータスではない ESC インスタンスがある場合、その ECS インスタンスは異常と見なされ スケーリンググループから削除されます。

- ・ 自動的に作成された ECS インスタンスは、すぐに削除され、リリースされます。
- ・手動で追加された ECS インスタンスは、すぐに削除されるが、停止およびリリースされることはありません。

MinSize 値による制限があっても、異常なECSインスタンスは削除されます。 削除されたことに よって、スケーリンググループ内の ECS インスタンスの数 (合計容量) MinSize 値を下回った場 合は、インスタンス数が MinSize 値に達するまで、ECS インスタンスがグループに自動的に追 加されます。

1.7 通知機能

ここでは、Auto Scaling の通知機能について説明します。

スケーリングアクティビティが、次のいずれかの条件を満たしたとき、ショートメッセージまた はメールが送信されます。

- スケジュール済みタスク、CloudMonitor アラームタスク、またはヘルスチェックによって
 スケーリングアクティビティがトリガーされた場合
- ・ ECS インスタンスが作成またはリリースされた場合

上記の条件が満たされると、対応するスケーリングアクティビティにショートメッセージまたは メールが送信されます。

1.8 強制操作

ここでは、強制操作について説明します。

Auto Scaling では、自動作成された ECS インスタンスを ECS コンソールから削除するなどの強 制操作をすることができます。 Auto Scaling では、ユーザーによる強制操作を次のように処理 します。

リソース	強制操作のタイプ	処理方法	
ECS	ECS コンソールまたは API を使用して、スケーリンググ ループから ECS インスタンス を削除する。	Auto Scaling は、ヘルス チェックを介して ECS インス タンスのステータスが異常か どうかを判断します。異常な 場合、スケーリンググループ からそのインスタンスを削除 します。この ECS インスタ ンスのイントラネット IP アド レスは、RDS のアクセスホワ イトリストからは自動的に削 除されません。スケーリング グループ内の ECS インスタン スの数 (合計容量) が MinSize 値を下回る場合は、MinSize 値に達するまで、ECS インス タンスがグループに自動的に 追加されます。	
ECS	ECS OpenAPI の許可を Auto Scaling から取り消す。	すべてのスケーリングアク ティビティのリクエストが拒 否されます。	
Server Load Balancer	Server Load Balancer コン ソールまたは API を使用し て、Server Load Balancer インスタンスから ECS インス タンスを強制的に削除する。	Auto Scaling では、このアク ションを自動的に検出しませ ん。また、このような例外を 処理しません。この ECS イ ンスタンスはスケーリンググ ループに保持されます。ただ し、このインスタンスが、ス ケールダウンアクティビティ の削除ポリシーに従って選択 された場合は、リリースされ ます。	

リソース	強制操作のタイプ	処理方法
Server Load Balancer	Server Load Balancer コン ソールまたは API を使用し て、Server Load Balancer インスタンスを削除するか、 インスタンスのヘルスチェッ ク機能を無効にする。	ECS インスタンスは、この Server Load Balancer イン スタンスに追加されているス ケーリンググループに追加さ れません。スケーリングタ スクは、このスケーリンググ ループから ECS インスタンス を削除するスケーリングルー ルをトリガーできます。ま た、ヘルスチェック機能によ り異常とみなされた ECS イン スタンスも削除できます。
Server Load Balancer	(料金滞納や障害により) Server Load Balancer イ ンスタンスが使用不可能にな る。	すべてのスケーリングアク ティビティが失敗します。ただ し、手動でトリガーした ECS インスタンスの削除アクティ ビティは除きます。
Server Load Balancer	Server Load Balancer API の許可を Auto Scaling から 取り消す。	Auto Scaling は、Server Load Balancer インスタンス に追加されているスケーリン ググループに対するすべての スケーリングアクティビティ のリクエストを拒否します。
RDS	RDS コンソールまたは API を 使用して、RDS ホワイトリス トから ECS インスタンスの IP アドレスを削除する。	Auto Scaling では、このアク ションを自動的に検出しませ ん。また、このような例外を 処理しません。この ECS イン スタンスは、スケーリンググ ループに保持されます。この インスタンスが、スケールダ ウンアクティビティの削除ポ リシーに従って選択された場 合は、リリースされます。

リソース	強制操作のタイプ	処理方法
RDS	RDS コンソールまたは API を 使用して RDS インスタンスを 強制的に削除する。	この RDS インスタンスを設 定したスケーリンググループ には、ECS インスタンスが追 加されなくなります。 ECS イ ンスタンスは、この RDS イ ンスタンスに設定されている スケーリンググループには追 加されません。 スケーリング タスクは、このスケーリング グループから ECS インスタ ンスを削除するスケーリング ルールをトリガーできます。 また、ヘルスチェック機能に より異常とみなされた ECS イ ンスタンスも削除できます。
RDS	(料金滞納や障害により) RDS インスタンスが使用不可能に なる。	すべてのスケーリングアク ティビティが失敗します。ただ し、手動でトリガーした ECS インスタンスの削除アクティ ビティは除きます。
RDS	RDS OpenAPI の許可を Auto Scaling から取り消す。	Auto Scaling は、RDS イン スタンスに追加されている スケーリンググループに対す るすべてのスケーリングアク ティビティのリクエストを拒 否します。

1.9 数量制限

ここでは、Auto Scaling に関する数量制限について説明します。

現在、Auto Scaling には、次の数量制限があります。

- 1つのリージョン内の1つのアカウントで、最大 50 個のスケーリンググループを作成できます。
 - 1つのスケーリンググループに最大10個のスケーリング設定を作成できます。
 - 1つのスケーリンググループに最大 50 個のスケーリングルールを作成できます。
 - 1つのスケーリンググループに最大6個のイベント通知を作成できます。
 - 1つのスケーリンググループに最大6個のライフサイクルフックを作成できます。
- ・最大 20 個のスケジュール済みタスクを作成できます。

・1 つのスケーリング設定に最大 10 個のインスタンスタイプを選択できます。

1.10 考慮事項

ここでは、Auto Scaling に関する考慮事項について説明します。

スケーリングルール

スケーリングルールの実行と計算の際、スケーリンググループの MaxSize 値と MinSize 値に基 づいて ECS インスタンスの数が自動的に調整されます。 たとえば、スケーリングルールで ECS インスタンスの数が 50 に設定されていても、スケーリンググループの MaxSize 値が 45 に設定 されている場合、45 の ECS インスタンスでスケーリングルールを計算し実行します。

スケーリングアクティビティ

- 1つのスケーリンググループで一度に実行できるスケーリングアクティビティは、1つだけです。
- スケーリングアクティビティは中断できません。たとえば、20の ECS インスタンスを追加するスケーリングアクティビティを実行中で、5つのインスタンスのみ作成されている場合、強制終了させることはできません。
- スケーリンググループに対して ECS インスタンスを追加または削除するスケーリングアク ティビティが失敗した場合、スケーリングアクティビティではなく ECS インスタンスの整合性 が維持されます。つまり、スケーリングアクティビティではなく、ECS インスタンスがロー ルバックされます。たとえば、スケーリンググループに 20 の ECS インスタンスが自動作成 されていて、19 の ECS インスタンスが Server Load Balancer インスタンスに追加された 場合、追加に失敗した ECS インスタンスのみがリリースされます。
- Auto Scaling は、Alibaba Cloud の Resource Access Management (RAM) サービスを 使用し、ECS API を介して ECS インスタンスを置き換えるため、ロールバック ECS インスタンスは引き続き課金されます。

クールダウン時間

- ・クールダウン時間中、CloudMonitor アラームタスクからのスケーリングアクティビティリ クエストのみが、スケーリンググループで拒否されます。他のタスク (手動で実行するスケー リングルールやスケジュール済みタスクなど)は、クールダウン時間が終了するまで待機せず に、即座にスケーリングアクティビティをトリガーできます。
- ・クールダウン時間は、最後の ECS インスタンスがスケーリングアクティビティによってス ケーリンググループに追加、またはスケーリンググループから削除された後に開始されます。

1.11 操作手順

ここでは、Auto Scaling ソリューションを作成するための手順を説明します。



- スケーリンググループを作成します (CreateScalingGroup)。 スケーリンググループ内の ECS インスタンスの最小値と最大値を設定し、関連付ける Server Load Balancer と RDS インスタンスを選択します。
- 2. スケーリング設定を作成します (CreateScalingConfiguration)。 イメージ ID やインス タンスタイプなど、Auto Scaling の ECS インスタンスの属性を設定します。
- 3. 手順2で作成したスケーリング設定を使用して、スケーリンググループを有効にします (EnableScalingGroup)。
- 4. スケーリングルールを作成します (CreateScalingRule)。 たとえば、N 個の ECS インスタンスを追加する、などです。
- 5. スケジュール済みタスクを作成します (CreateScheduledTask)。 たとえば、手順4で作成 したスケーリングルールを午前 12:00 にトリガーする、などです。
- アラームタスクを作成します (CloudMonitor API PutAlarmRule)。 たとえば、平均 (最大 または最小も可) CPU 使用率が 80 %以上の場合に ECS インスタンスを1つ追加する、など です。

1.12 **ワークフロー**



ここでは、Auto Scaling のワークフローを説明します。

スケーリンググループ、スケーリング設定、スケーリングルール、スケーリングトリガータスク の作成後、以下のプロセスが実行されます (この例では、ECS インスタンスを追加します)。

- スケジュール済みタスクでは、指定された時間にスケーリングルールの実行リクエストをトリ ガーします。
 - CloudMonitor タスクでは、スケーリンググループ内の ECS インスタンスのパフォーマン スをリアルタイムでモニタリングし、設定されたアラームルールに基づいて (たとえば、ス

- ケーリンググループ内の全 ECS インスタンスの平均 CPU 使用率が 60% を上回った場合など)、スケーリングルールの実行リクエストをトリガーします。
- ・スケジュール済みタスクでは、トリガー条件に従ってスケーリングアクティビティをトリガーします。
- カスタムタスクでは、モニタリングシステムとアラームルールに基づいて (たとえば、オンラインユーザー数やジョブキュー数など)、スケーリングルールの実行リクエストをトリガーします。
- ヘルスチェックタスクでは、スケーリンググループとグループの ECS インスタンスのヘル スステータスを定期的にチェックします。 ECS インスタンスに異常が見つかった場合 (実 行中ステータスでない場合)、グループから ECS インスタンスを削除するリクエストをト リガーします。
- スケーリングアクティビティは、ExcuteScalingRule インターフェイスを介してトリガーされます。このインターフェイスでは、一意の Alibaba Cloud リソース ID (ARI) により、実行するスケーリングルールを指定します。

カスタムタスクを実行する場合、ユーザーのプログラムで ExcuteScalingRule インターフェ イスを呼び出しておく必要があります。

- システムは、手順2で入力されたスケーリングルール ARI に基づいてスケーリングルール、 スケーリンググループ、スケーリング設定に関する情報を取得し、スケーリングアクティビ ティを作成します。
 - a. システムは、スケーリングルール ARI に基づいてスケーリングルールとスケーリンググ ループを照会し、追加する ECS インスタンスの数を計算します。また、Server Load Balancer と RDS インスタンスを設定します。
 - b. システムは、スケーリンググループに基づいてスケーリング設定を照会し、新しい ECS イ ンスタンスの作成時に使用するパラメーター (CPU、メモリ、帯域幅)を決定します。
 - c. システムは、追加する ECS インスタンス数、ECS インスタンスの設定、および Server Load Balancer と RDS インスタンスの設定に基づいて、スケーリングアクティビティを 作成します。
- 4. スケーリングアクティビティの実行中、 ECS インスタンスが作成され、Server Load Balancer と RDS インスタンスが設定されます。
 - a. インスタンスの設定に基づいて、指定した数の ECS インスタンスが作成されます。
 - b. 作成された ECS インスタンスのイントラネット IP アドレスが、指定した RDS インスタン スのホワイトリストに追加されます。また、作成された ECS は、指定した Server Load Balancer インスタンスにも追加されます。

 スケーリングアクティビティが完了すると、スケーリンググループのクールダウン機能が開始 されます。 クールダウン時間が経過するまで、そのスケーリンググループで新しいスケーリン グアクティビティを実行することはできません。

2 スケーリング設定

2.1 スケーリング設定の作成

ここでは、スケーリング設定の作成方法について説明します。

スケーリング設定の作成プロセスは ECS インスタンスの作成と似ていますが、リージョンなどの 一部の項目は設定できません。 コンソールの指示に従って操作してください。 ページの各設定 項目には、簡単な説明が付いています。 設定項目の意味と使用方法の詳細は、「ウィザードを使 用したインスタンスの作成」をご参照ください。

スケーリング設定を作成するには、次の手順に従います。

- 1. Auto Scaling コンソール にログインし、[操作] 列にある [管理] をクリックします。
- 2. [スケーリング設定] ページに進み、[スケーリング設定の作成] をクリックします。
- [基本設定] ページで、課金方法、インスタンス、イメージ、ストレージ、パブリックネット ワーク帯域幅、セキュリティグループを設定し、[次のステップ: システム構成] をクリックし ます。

三注:

Auto Scaling では、従量課金インスタンスとプリエンプティブルインスタンスのみを選択で きます。また、パブリックイメージ、カスタムイメージ、共有イメージのみを使用できます。

4. [システム構成] ページで、タブ (オプション)、ログイン認証情報、インスタンス名 (オプション)、および詳細オプション (オプション) を設定し、[確認] をクリックします。

三注:

詳細オプションでは、インスタンス RAM ロールとユーザー定義データを設定します。 詳細 オプションは、VPC 接続のスケーリンググループに対してのみ設定できます。

- 5. [確認] ページで、選択した設定項目を確認し、スケーリング設定名を入力して、[作成] をク リックします。
- (作成成功) ダイアログボックスで、[適応] をクリックします。 スケーリング設定を今すぐ使用 しない場合は、ダイアログボックスを閉じます。

作成が成功したら、スケーリング設定リストでスケーリング設定を表示および選択できます。

3 Auto Scaling の実装

3.1 スケジュール済みタスク

3.1.1 スケジュール済みタスクの作成

ここでは、スケジュール済みタスクの定義と作成方法について説明します。

スケジュール済みタスクは、入力パラメーターに従って、最大 20 個まで作成できます。

スケジュール済みタスクとは

スケジュール済みタスクとは、指定した時間に指定したスケーリングルールを実行するデフォル トのタスクを指します。 したがって、ビジネスニーズを満たし、コストを管理するために、コン ピューティングリソースが自動的にスケールアップまたはスケールダウンされます。 また、柔軟 なルールでビジネスの変化に対応できるように、スケジュール済みタスクの繰り返し周期を指定 することもできます。

注:

1つのアカウントで作成できるスケジュール済みタスクの数には、制限があります。「数量制 限」をご参照ください。

スケーリンググループでは、一度に1つのスケーリングアクティビティしか実行できないため、1 つのスケーリングルールが失敗した場合に自動的に再試行する機能により、スケジュール済みタ スクの結果を保証します。実行対象のスケジュール済みタスクが1分間に複数存在する場合、最 後に作成されたスケジュール済みタスクが実行されます。

手順

スケジュール済みタスクを作成するには、次の手順に従います。

- 1. Auto Scaling コンソールにログインします。
- 2. [自動トリガータスク] を選択して、[スケジュール済みタスク] ページに移動します。[スケ ジュール済みタスクの作成] をクリックします。

Scaling group manage	eduled task
 Auto-trigger task mana 	
Scheduled task Description Status Scaling group info Execution time Recurrence end time Retry expiration time	Operation
Alarm task schedule_task_d schedule task Stopped Scaling group:test_scaling_group Scaling rule:Scaling_rule 2016-10-14 22:48 每1天执行 2016-11-12 22:48 600秒 Enable Mo	Jify Delete
Help Total: 1 Itern(s). Per Page: 10 Itern(s) a d	
FAQs	

スケジュール済みタスクの作成] ダイアログボックスで、タスク名、実行時刻、スケーリング ルール、再試行有効期限 (オプション)、繰り返し周期 (オプション) を指定します。後で参照す るための説明を追加することもできます。 [送信] をクリックします。

el

门注:

スケジュール済みタスクの属性については、「スケジュール済みタスクの属性」をご参照く ださい。

スケジュール済みタスクの属性

属性名	説明	例
タスク名	名前は 2~40 文字で指定しな ければなりません。 小文字、 数字、または漢字で始める必 要があります。"."、","、"-" を 使用できます。	st-yk201808301442
説明	スケジュール済みタスクの目 的、機能などの情報を記述し ます。	PV は月の初めに大きいです。 3つのインスタンスを追加しま す。
実行時刻	スケジュール済みタスクをト リガーする時刻	00:00, September 2, 2018

属性名	説明	例
スケーリングルール	スケーリングルールの名前で す。タスクがトリガーされた ときに実行するスケーリング 操作を指定します。	add3
再試行有効期限	時間の範囲は、0 秒~21,600 秒 (6 時間) です。指定した時 刻にスケーリング操作の実行 に失敗した場合、スケジュー ル済みタスクは、再試行有効 期限まで繰り返し実行されま す。	600
繰り返し周期	スケジュール済みタスクを繰 り返し実行する周期です。 毎 日、毎週、毎月のいずれかで す。 異なる要件が必要な場合 は、Cron 式を使用します。	毎月 毎月 2 日から 3 日に実行され ます。
繰り返し終了時刻	スケジュール済みタスクの繰 り返し実行を停止する時刻で す。	00:00, September 30, 2018

Cron 式

Cron 式では、UTC + 0タイムゾーンが使用されます。 システム時間を中国の現地時間に変換す る際、8 時間を追加する必要があります。 また、Cron 式の初回実行時間は繰り返し終了時間よ り短くしなければなりません。そうしないと、スケジュール済みタスクは作成されません。

Cron 式は、スペースで区切られた文字列です。5から7つのフィールドに分けられます。 現 在、Auto Scaling のスケジュール済みタスクでは、分、時、日、月、週の5フィールドの Cron 式を使用できます。 以下の表に、値の範囲を示します。

フィールド	必須かどうか	有効値
分	はい	[0, 59]
時	はい	[0, 23]
Н	はい	[1, 31]
月	はい	[1, 12]
週	はい	[0, 7]; Sunday = 0 or 7

フィールドには複数の値を入力できます。

- ・複数の値を指定するには、カンマ(,)を使用します。たとえば、1,3,4,7,8と入力します。
- ・ 値の範囲は "-" を使用して指定します。 たとえば、1-6 と指定します。 1, 2, 3, 4, 5, 6 と指定 するのと同じ結果になります。
- ・任意の値を指定する場合は、アスタリスク (*) を使用します。 たとえば、時フィールドのアス タリスクは各時間を表し、結果は 0-23 と同じです。
- ・間隔を指定する場合は、スラッシュ (/)を使用します。たとえば、時フィールドに 0-23/2 と入力すると、2 時間ごとに実行されることを示します。スラッシュ (/) はアスタリスク (*) と一緒に使用できます。たとえば、時フィールドに*/3 と入力すると、3 時間ごとに実行されることを示します。

3.2 アラームタスク

3.2.1 アラームタスクの表示、変更、削除

ここでは、アラームタスクを表示、変更、削除する方法について説明します。

Ceeling group menage								
Scaling group manage							C Refresh	Create alarm task
 Auto-trigger task mana 								
Scheduled task	Alarm task	Status	Monitor resource	Statistical cycle	Scaling trigger rule	Number of trigger rules		Operation
Alarm task	alarm_task_demo	Normal	Scaling group:test_scaling_group	2Minutes	CPUUsageAverageContinual3Times>=70.0%	1	View Deta	ails Disable Delete More operations -
Help						Total: 1 item(s), Per Page: 1	0 item(s) «	< 1 > »
FAQs								

メトリックの詳細の表示

アラームタスクが正常に作成されると、アラームタスクリストにアラームタスクが表示されま す。

- 1. Auto Scaling コンソールにログインします。
- 2. [システムモニタリング] をクリックして、作成したシステムモニタリングのアラームタスクを 表示します。
- 3. [カスタムアラームタスク] をクリックして、作成したカスタムアラームタスクを表示します。
- アラームタスク名をクリックして、詳細ページに移動します。このページには、アラームタス クのメトリックデータが表示されます。

アラームタスクの変更

アラームタスクリストページでは、アラームタスクを変更できます。また、アラームタスクの詳 細ページに移動して、アラームルールを変更することもできます。 アラームタスクの変更は、アラームタスクの基本情報の変更と、アラームルールのトリガールー ルの変更の2つの部分に分けられます。

基本情報の変更には、タスク名、メトリック、統計期間、統計方法、繰り返し回数などの変更が 含まれます。 アラームタスクのメトリックを変更することは、異なるインデックスをモニタリ ングするということなので、メトリックを変更しないことを推奨します。 この場合、新しいイン デックスの新しいアラームタスクを作成することを推奨します。

アラームタスクの削除

[アラームタスク] ページの [操作] 列でアラームタスクを削除します。

4 Auto Scaling の管理

4.1 スケーリングルールの変更、照会、削除

ここでは、スケーリングルールの変更手順を説明します。

スケーリングルールの作成後、スケーリングルールを変更、照会、削除することができます。 変 更する前に、スケーリングルールの詳細を表示します。

√ test_scaling_	g			${\mathfrak Q}$ Refresh
Scaling rule				Create scaling rule
You can have up to 10 scaling rul	les in a single scaling group.			
Scaling rule	Adjusted type	Adjusted value	Cool-down time	Operation
Scaling_rule	Adjusted to	3台	600秒	View Details Execute Modify Delete
				Total: 1 item(s), Per Page: 10 item(s) a c 1 , s

- 1. [スケーリングルール] ページに移動し、[操作] 列で [詳細の表示] をクリックします。
- 2. 変更対象のスケーリングルールを確認してから、[変更] をクリックします。
- 3. [スケーリングルールの変更] ダイアログボックスで、必要に応じて属性を変更し、[スケーリ ングルールの変更] をクリックします。

注:

スケーリングルールの属性の詳細については、「スケーリングルールとは」をご参照くださ い。

4.2 削除ポリシー

ここでは、削除ポリシーについて説明します。

削除ポリシーには、デフォルトポリシーとカスタムポリシーの2種類があります。

デフォルトの削除ポリシー

このポリシーでは、最も古いスケーリング設定 (OldestScalingConfiguration) に従って作成さ れた ECS インスタンスに対して、最初にレベル 1 のインスタンススクリーニングが実行されま す。続いて、最も古い ECS インスタンス (OldInstance) に対して、レベル 2 のスクリーニング が実行されます。

このポリシーでは、スケーリンググループの最も古いスケーリング設定 (OldestScal ingConfiguration) に従って作成された ECS インスタンスが、最初に選択されます。続い

て、これらの ECS インスタンスから最も古い ECS インスタンス (OldestInstance) が選択さ れます。 最も古い ECS インスタンスが複数見つかった場合、そのうちの 1 つがランダムに選 択され、スケーリンググループから削除されます。

- ・手動で追加した ECS インスタンスは、どのスケーリング設定にも関連付けられていないため、最初に削除対象として選択されることはありません。
- スケーリング設定に関連付けられているすべての ECS インスタンスを削除しても、さらに多くのインスタンスをスケーリンググループから削除する必要がある場合、このポリシーでは手動で追加された時期が最も古いインスタンスが選択されます。

カスタムリリースポリシー

複数のポリシーを設定して、スケーリンググループから ECS インスタンスを連続的に選択および 削除できます。

- リリースポリシーの種類
- OldestInstance: このポリシーでは、作成された時期が最も古い ECS インスタンスが選択されます。レベル1スクリーニングでは、手動または自動で作成された時期が最も古い ECS インスタンスが選択されます。
- NewestInstance: このポリシーでは、直近に作成された ECS インスタンスが選択されます。
 レベル1スクリーニングでは、手動または自動で作成された時期が最も新しい ECS インスタンスが選択されます。
- OldestScalingConfiguration: このポリシーでは、最も古いスケーリング設定に従って作成 されたインスタンスが選択されます。手動で追加したインスタンスは無視されます。ただし、 スケーリング設定に関連付けられているすべての ECS インスタンスを削除しても、さらに多 くのインスタンスをスケーリンググループから削除する必要がある場合、このポリシーでは手 動で追加した ECS インスタンス (スケーリング設定に関連付けられていないインスタンス)が ランダムに選択されます。

4.3 スケーリンググループの削除

ここでは、スケーリンググループの削除手順について説明します。

不要になったスケーリンググループを削除できます。

注:

スケーリンググループを削除すると、そのグループのスケーリング設定とスケーリングルールも 削除されます。 スケーリンググループにステータスが実行中の ECS インスタンスが含まれる場 合、最初にその ECS インスタンスが停止されてから、手動で追加したすべてのインスタンスが 削除されます。次に、自動作成されたすべてのインスタンスがリリースされます。

- [スケーリンググループ] ページで、削除対象のスケーリンググループの横にある [操作] 列の
 [削除] をクリックします。
- 2. [スケーリンググループの削除] ダイアログボックスで、[確認] をクリックします。
- 3. [スケーリンググループ] ページで [更新] をクリックして、削除が完了したことを確認します。

5 手動スケーリング

5.1 ECS インスタンスの削除

指定したスケーリンググループから ECS インスタンスを削除できます。

自動的に作成された ECS インスタンスがスケーリンググループから削除されると、そのインスタンスは停止およびリリースされます。

手動で追加した ECS インスタンスがスケーリンググループから削除されても、そのインスタンス は停止もリリースもされません。

削除操作が正常に行なわれるには、次の条件を満たす必要があります。

・スケーリンググループがアクティブであること。

・スケーリンググループは、スケーリングアクティビティを実行していないこと。

スケーリンググループに対してスケーリングアクティビティが実行されていない場合、ECS イン スタンスの削除操作はクールダウン時間を待たずにすぐに実行されます。

正常に返された場合、Auto Scaling サービスはすぐにスケーリングアクティビティを実行 しますが、スケーリングアクティビティが正常に実行されるわけではありません。 返された ScalingActivityID を使用して、スケーリングアクティビティのステータスを確認してくださ い。

スケーリンググループ内の既存の ECS インスタンス数 (合計容量) から削除対象の ECS インスタンス数を引いた数が MinSize 値よりも小さい場合、操作は失敗します。

<	test_scaling_g						${\mathfrak C}$ Refresh
Basic info	Total number of instances:1						
ECS instance list		Active instances		Pending instances	Removed instances		
Scaling activity		1 item(s)		O item(s)	O _{item(s)}		
Scaling configuration							
Scaling rule							
	Automatically create Manually attach						Add existing instance
	CS name		Scaling configuration	Status	Health check status(All) 👻	Time added	Operation
	SS-sg-test_scaling_group-ecs-l-2ze7mrdp		test-config	In Service	Healthy	2016-10-13 22:44	Remove from scaling group and release
Ξ	Remove from scaling group and release						Total: 1 item(s), Per Page: 10 item(s)

例

6 ECS インスタンスリストの照会

ここでは、ECS インスタンスリストを照会する方法について説明します。

ステータスが実行中(Running)でない ECS インスタンスは異常とみなされます。 異常な ECS インスタンスは、スケーリンググループから自動的に削除されます。 自動作成された ECS インスタンスとは、スケーリング設定とルールに基づいて、Auto Scaling サービスによって作 成されたインスタンスです。 手動で追加した ECS インスタンスとは、Auto Scaling サービスに よって作成されたのではなく、スケーリンググループに手動で追加したインスタンスです。

例

以下に例を示します。

<	🚾 test_scaling_g							
Basic Info								
ECS instance list	total number of instances:1	Active instances	Pending instances	Removed instances				
Scaling activity		1 item(s)	O item(s)					
Scaling configuration								
Scaling rule								
	Automatically create Manually attach					Add existing instance		
	ECS name	Scaling configuration	Status He	ith check status(All) 👻	Time added	Operation		
	ESS-sg-test_scaling_group-ecs-I-2zebmpnh	test-config	In Service He	lthy	2016-10-13 22:27	Remove from scaling group and release		
Ξ	Remove from scaling group and release					Total: 1 item(s), Per Page: 10 item(s)		

例

7スケーリングアクティビティの表示

ここでは、スケーリングアクティビティの表示方法について説明します。

この操作では、過去 30 日間に実行されたスケーリングアクティビティの情報を照会します。

<u>~</u>					Heroduct News	${f C}$ Refresh
Scaling Activities						Table Chart
Scaling Activities	Total Instances (Updated)	Started At	Stopped At	Description	Status(All) 👻	Actions
au (101) a' 1010	1	21 February 2019, 17.29	21 February 2019, 17.29	Add "1" ECS ins	Successful	View Details
101100-00100	0	19 February 2019, 14.27	19 February 2019, 14.27	Remove "2" ECS	Successful	View Details
-	2	19 February 2019, 14.25	19 February 2019, 14.26	Add "2" ECS ins	Successful	View Details
-	0	19 February 2019, 14.23	19 February 2019, 14.24	Remove "2" ECS	Successful	View Details
101010-00	2	19 February 2019, 14.21	19 February 2019, 14.23	Add "2" ECS ins	Successful	View Details
-	0	19 February 2019, 14.05	19 February 2019, 14.08	Add "2" ECS ins	Failed	View Details
101010-008	0	19 February 2019. 14.02	19 February 2019. 14.03	Remove "1" ECS	Successful	View Details
Scaling Activity ID:asa-m5ei5om4p049gqjma3d5		Status:Successful				×
Started At:21 February 2019, 17.29		Stopped At:21 February 2019, 17.29				
Cause: A user re	equests to attach instance "	" from the specified scaling group.				
Details: new ECS	5 instances ' are attached.					
Status: "1" ECS instances are added						