1 プロダクト概要

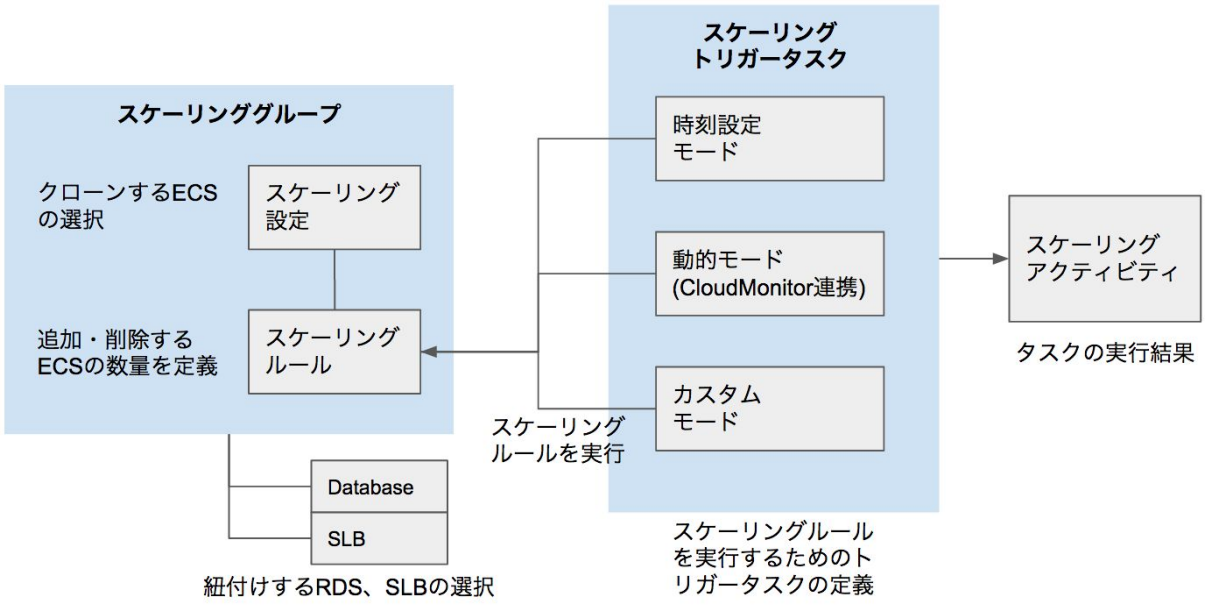
1-1 プロダクト概要

Auto Scaling は、ユーザーリクエストのボリュームに応じてECSコンピューティングリソースを自動的に調整するサービスです。急激なユーザーリクエストの増加に対しても、Auto Scaling によって自動的に ECS インスタンスを追加して対応できます。ユーザーリクエストが減少したときに、自動的に追加されたインスタンスが削除することもできます。

1-2 ワークフロー

図1はAuto Scalingで利用する概念の関連を表したものです。各概念の詳細な説明は本資料の「2 プロダクトの機能」に記述しています。

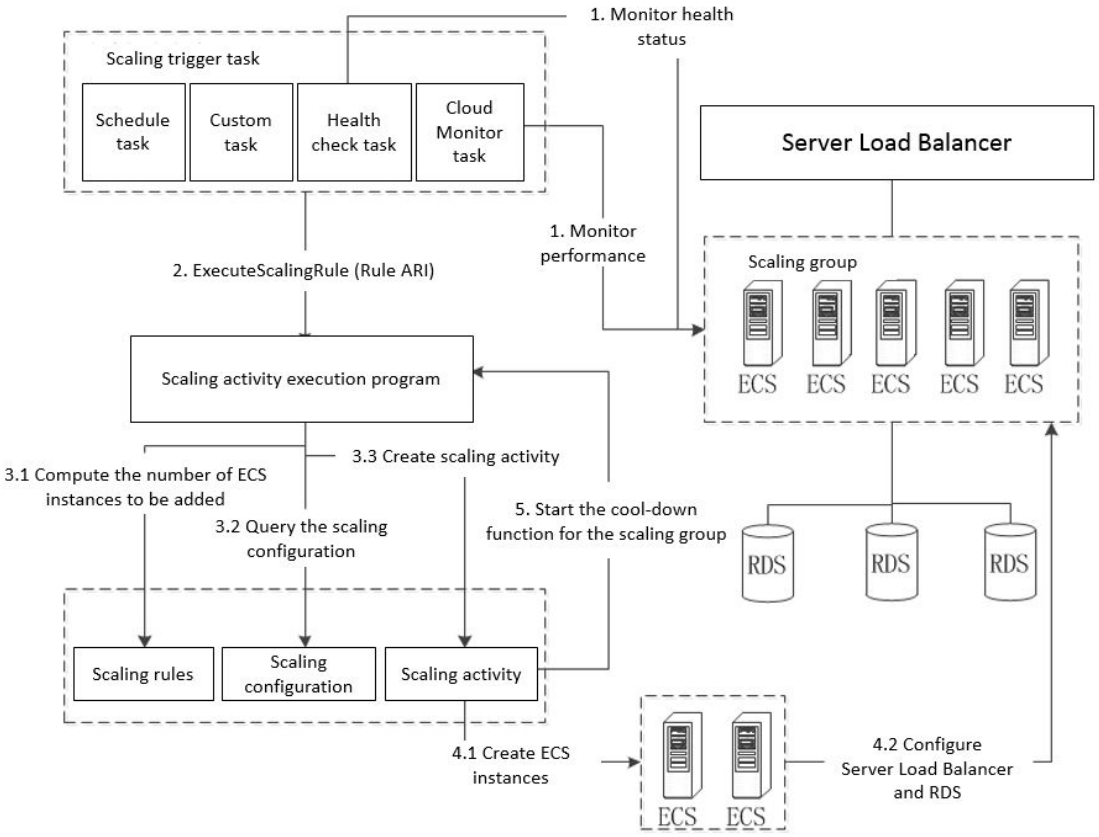
図1 Auto Scaling概念関連図



© 2017 SB Cloud Corp. © 2009-2017 Alibaba Cloud.

また、図2はAuto Scalingの処理のワークフローを示します。

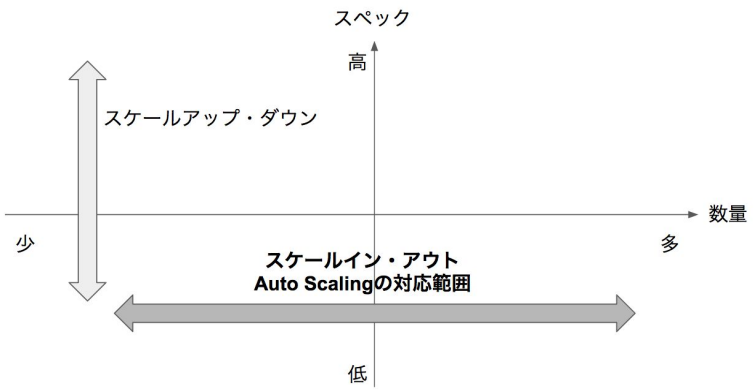
図2 ワークフロー



1-3 適用範囲

Auto ScalingはECSサーバのスケールイン・アウトに対応するサービスです。ECSのスケールアップ・ダウンを行いたい場合には別途ご対応が必要です。

図3 Auto Scalingの適用範囲



© 2017 SB Cloud Corp. © 2009-2017 Alibaba Cloud.

2 プロダクトの機能

2-1 スケーリンググループ

スケーリンググループは、同じアプリケーションが稼働する ECS インスタンスのグループのことです。スケーリンググループでは、稼働するECS インスタンスの最大数・最小数（スケール可能なECSの数は0～100台）や、関連する SLB インスタンス、RDS インスタンスを設定できます。

2-2 スケーリング設定

スケーリング設定では、Auto Scaling に使用される ECS インスタンスの設定を行います。具体的には、スケールするECSインスタンスの選択、属するセキュリティグループ、ネットワークの設定です。

2-3 スケーリングルール

スケーリングルールは、追加・削除するECSインスタンスの数量を設定することができます。これは、後述のスケーリングトリガータスクによって実行されるルールです。また、ルールに対するクールダウン時間の設定も可能です。１つのスケーリンググループに登録できるスケーリングルールは最大10個までです。

2-4 スケーリングアクティビティ

スケーリングルールが実行されると、スケーリングアクティビティに結果が保存されます。スケーリングアクティビティは、スケーリンググループの ECS インスタンスの数量の変化およびスケーリングの成否を確認することができます。なお、スケーリングアクティビティは過去30日分保存されます。

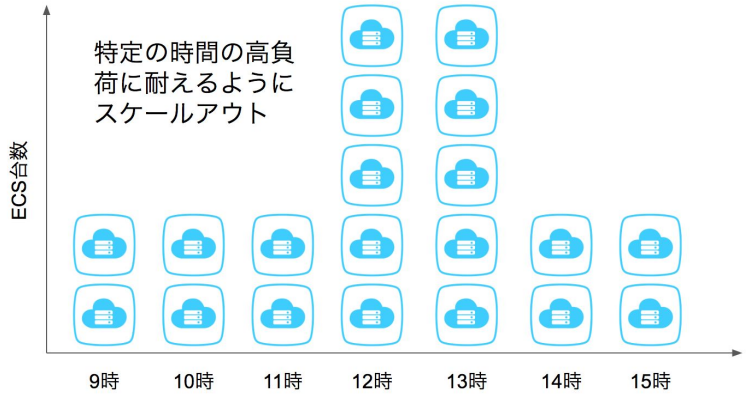
2-5 スケーリングトリガータスク

スケーリングトリガータスクは、スケジュール済みタスクや CloudMonitor アラームタスクなど、スケーリングルールを起動するための設定を行います。スケーリングルールを実行するにはいくつかの方式があります。

2-5-1 時刻設定モード

ECS インスタンスを定期的に追加または削除するスケジュール済みタスク (毎日午後 1 時など) を設定します。

© 2017 SB Cloud Corp. © 2009-2017 Alibaba Cloud.



2-5-2 動的モード

CloudMonitor パフォーマンス指標に基づいて ECS インスタンスを自動的に追加または削除することができます。パフォーマンス指標には、CPU、メモリ、平均システム負荷（ロードアベレージ）、送受信トラフィックが指定できます。ただし、利用するパフォーマンス指標によっては、事前にCloudMonitorのエージェントをインストールする必要があります。（表1参照）

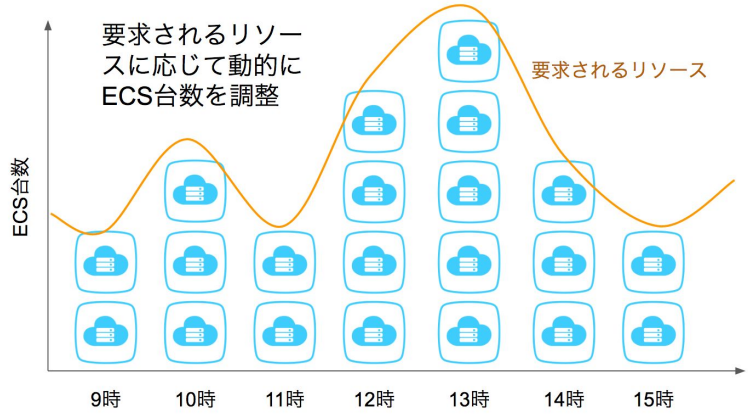


表1. パフォーマンス指標とエージェントインストール要否

	CPU	メモリ	平均システム負荷※	トラフィック使用状況
エージェントインストール要否	不要	必要	必要	不要

※Windows Serverの場合は、平均システム負荷指標が取得できないため、パフォーマンス指標に選択することはできません。

2-5-3 固定数量モード

属性を利用して、定型的なシナリオでのリアルタイム運用のために、一定の数の正常な ECS インスタンスを維持することができます。

© 2017 SB Cloud Corp. © 2009-2017 Alibaba Cloud.

2-5-4 カスタムモード

コンソール画面からあるいはAPI を利用して手動でのスケーリングを実行することできます。具体的に下記のような実行方法があります。

1. スケーリングルールを手動で実行する。
2. 既存の ECS インスタンスを手動で追加または削除する。
3. MinSize 属性または MaxSize 属性を手動で調整し、ECS インスタンスの数が MinSize 値と MaxSize 値の範囲に収まるようにAuto scalingを実行する。

2-5-5 ヘルシーモード

ECS インスタンスが Running ステータスでない場合、そのインスタンスは Auto scaling により自動的に削除またはリリースされます。
異常なインスタンスが検出された場合は、Auto Scaling によって新しいインスタンスに自動的に置き換えられるため、サービスに中断が生じません。

2-5-6 マルチモード

上記のモードを任意に組み合わせて使用できます。たとえば、ユーザーは毎日午後 1 時から午前 2 時までの間がピークタイムと予測された場合、時刻設定モードで 20 個の ECS インスタンスを定期的に作成します。予測したピークタイムよりも高い実需要があるかどうかを把握できない場合(たとえば、40 個の ECS インスタンスが 1 日に必要となるような場合) は、時刻設定モードに加え、動的モードも同時に設定しておけば、このような予想外の変化に対応することができます。

2-6 クールダウン時間

クールダウン時間は、スケーリンググループのスケーリングアクティビティが完了した後の、他のアクティビティが一切実行されない期間を指します。クールダウン時間はスケーリングルールで設定することが可能です。また、デフォルトのクールダウン時間は300秒で、設定可能なクールダウン時間は0～86400秒（24時間）です。

2-7 Auto ScalingとECSクォータの関係について

Alibabaアカウントはそれぞれのクラウドリソースの購入可能なクォータが制限されます。例えば、日本リージョンにおいて、デフォルトのAlibabaアカウントは従量課金のECSインスタンスを最大10インスタンス購入可能です。AutoScalingにより自動的に作成されたECSインスタンスもこのクォータで制限されます。そのため、自動スケーリングECSインスタンスを10台以上に作成したい場合は、この上限を引き上げる必要があります。リソース作成クォータの上限を引き上げたい場合には、サポートチケットを起票し依頼してください。

© 2017 SB Cloud Corp. © 2009-2017 Alibaba Cloud.

ご利用上の注意事項

この資料は、Alibaba Cloudの提供するクラウドサービスの機能について説明したもので、サービスのご利用を検討する際の参考となる技術的情報を提供するものです。
今後、本資料はクラウドサービスの機能追加・変更等に合わせて、予告なく変更される場合があります。閲覧された情報は最新のものではない場合がありますので、予めご了承下さい。

改版履歴

日付	版数	変更内容
2017/2/27	1.0	初版作成
2017/4/13	1.1	「2−6 AutoScalingとECSクォータの関係について」を追加
2017/9/1	1.2	スケーリングタスクの動的モードに、指定できるパフォーマンス指標取得のためのエージェントインストール要否を追加

© 2017 SB Cloud Corp. © 2009-2017 Alibaba Cloud.

本文書中に記載されている社名・商品名等は各社の商標または登録商標です。

© 2017 SB Cloud Corp. © 2009-2017 Alibaba Cloud.

SB Cloud