Redshift Serverless (Preview) ワークショップ

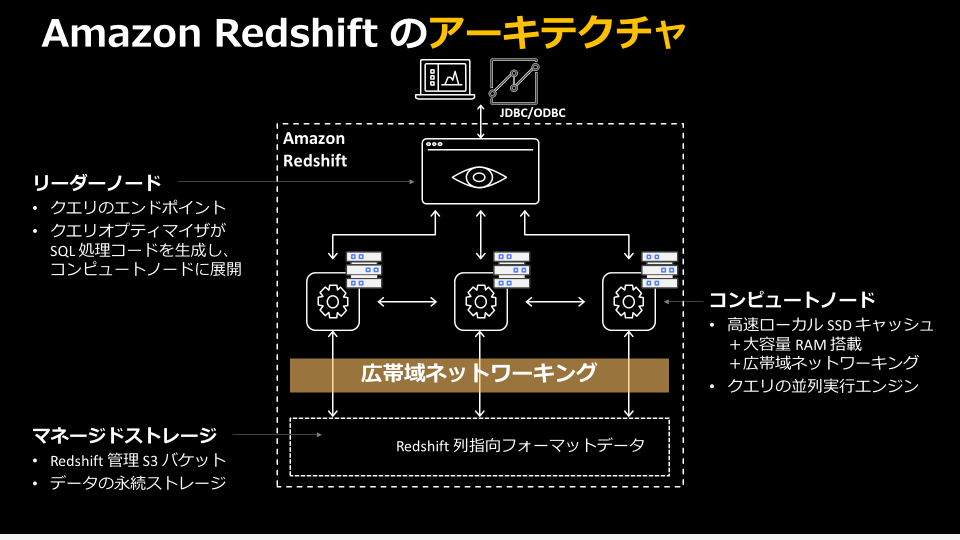
2022/04/29

シニアエバンジェリスト 亀田

はじめに：

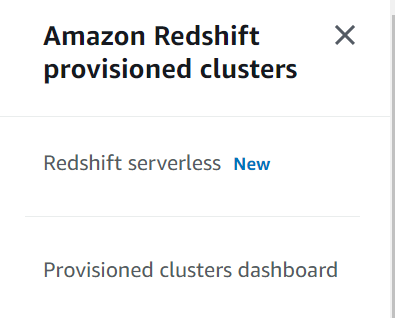
Amazon Redshift Serverlessは現在Preview中のサービであり、従来のクラスター型Redshiftと同様に、あらゆる規模で高いパフォーマンスで分析を非常に簡単に実行できる新機能です。使い方はデータを読み込んでクエリを開始するだけで、クラスターのセットアップや管理は不要です。データのクエリやロード中など、データウェアハウスの使用中は秒単位で課金されデータ、ウェアハウスがアイドル状態の場合は課金されないのが特徴です。

Redshift Serverless では、使用するコンピューティングとストレージに対して別途料金を支払います。コンピューティング性能は Redshift Processing Units (RPU) で測定され、ワークロードの料金は RPU 時間で 1 秒単位で請求されます。ストレージについては、Amazon Redshift が管理するストレージに保存されたデータと、スナップショットに使用したストレージに対して課金されます。これは、RA3 インスタンスを使用してプロビジョニングされたクラスターで支払うのと同じです。

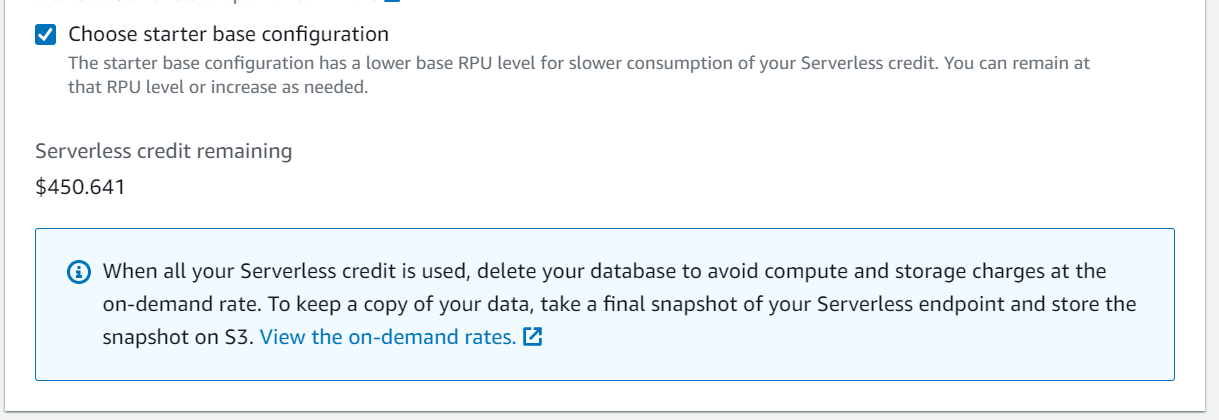


現在Redshift ServerlessはPreview中であり、仕様は変更になる可能性があります。

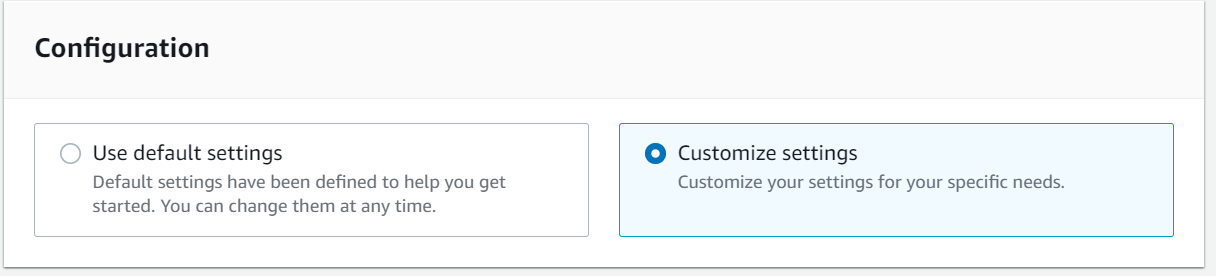
1. Redshiftマネージメントコンソールに移動します。
2. 左ペインの[Redshift serverless]をクリックします



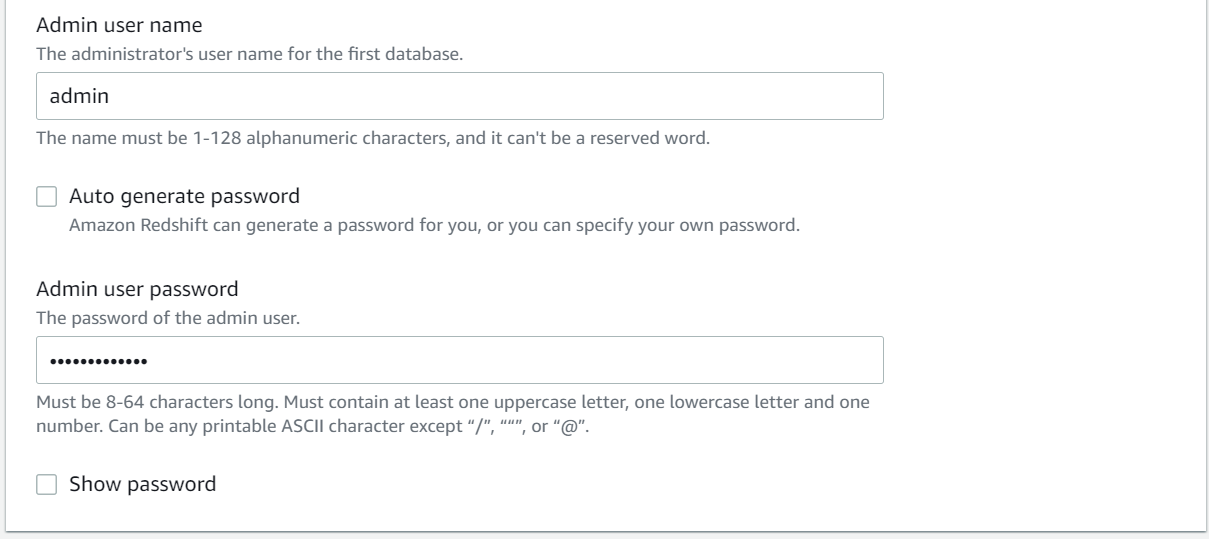
1. [Choose starter base configuration]をクリックします。以下の通り検証用500＄クレジットが付与されます。



1. クラスターを起動するためのNetworkは自動でdefault VPCが設定されています。現在Redshift ServerlessはPublicエンドポイントには対応しておらず、VPC経由でのエンドポイントのみが作成可能です。[Customize Settings]を選びます



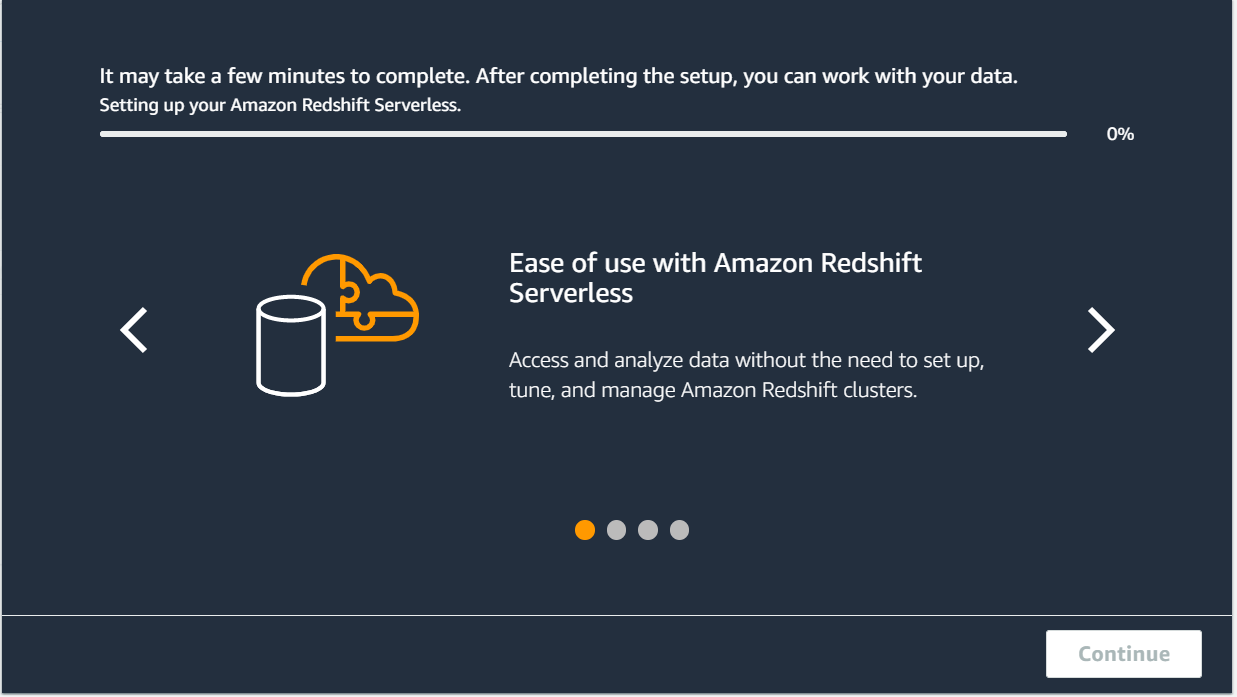
1. Adminのパスワードをセットします



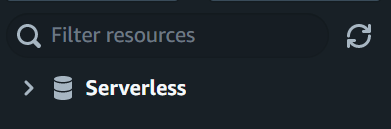
1. Redshift ServerlessはRedshiftと異なりIAMのクレデンシャルが不要で、ユーザー/パスワードのみが利用が可能です。このため、IAMを設定せず[Create serverless endpoint]をおします



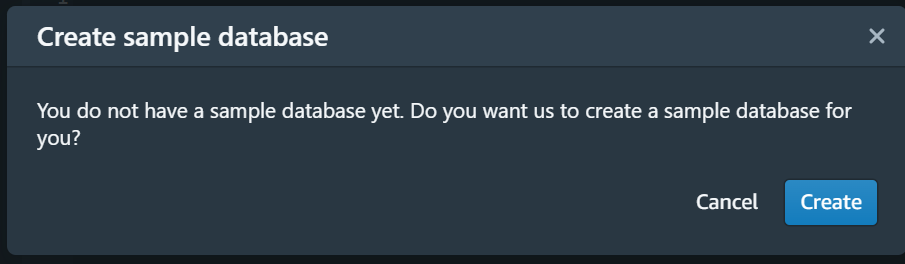
1. しばらくまつと[Continue]ボタンが押せるようになります



1. 起動したら、左ペインから[Query Editor]をクリックします
2. [Serverless]データベースにテストデータをロードします。[>Serverless]の矢印をクリックします

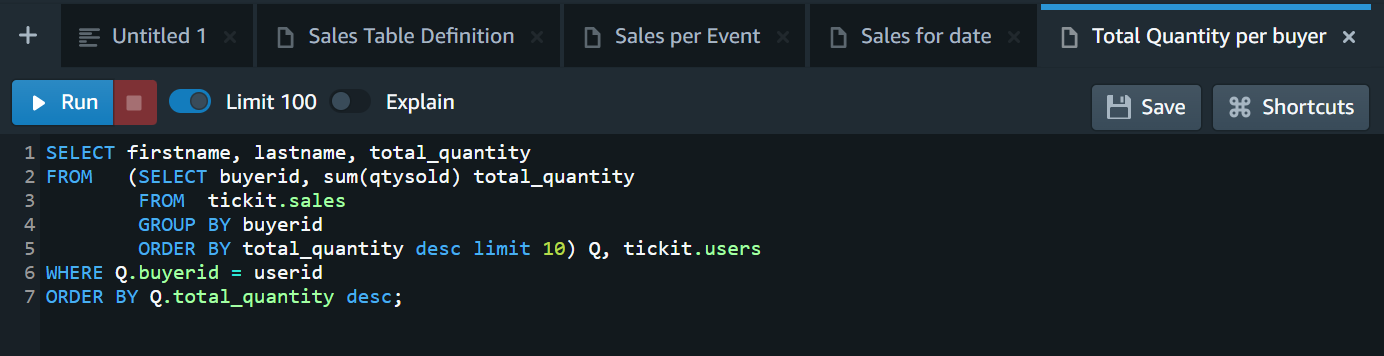


1. 順に、[> sample\_data\_dev]の矢印、[tickti]のフォルダをクリックするとテストデータがロードされ、サンプルクエリが表示されます。途中以下の確認ダイアログが表示されますので、[Create]をおします

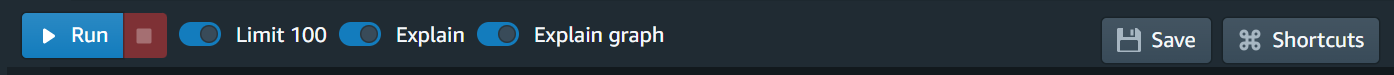


あらかじめ存在している[dev]データベースは、Redshift Serverlessの本体であり様々な設定データ、RPU消費履歴等が格納されています。

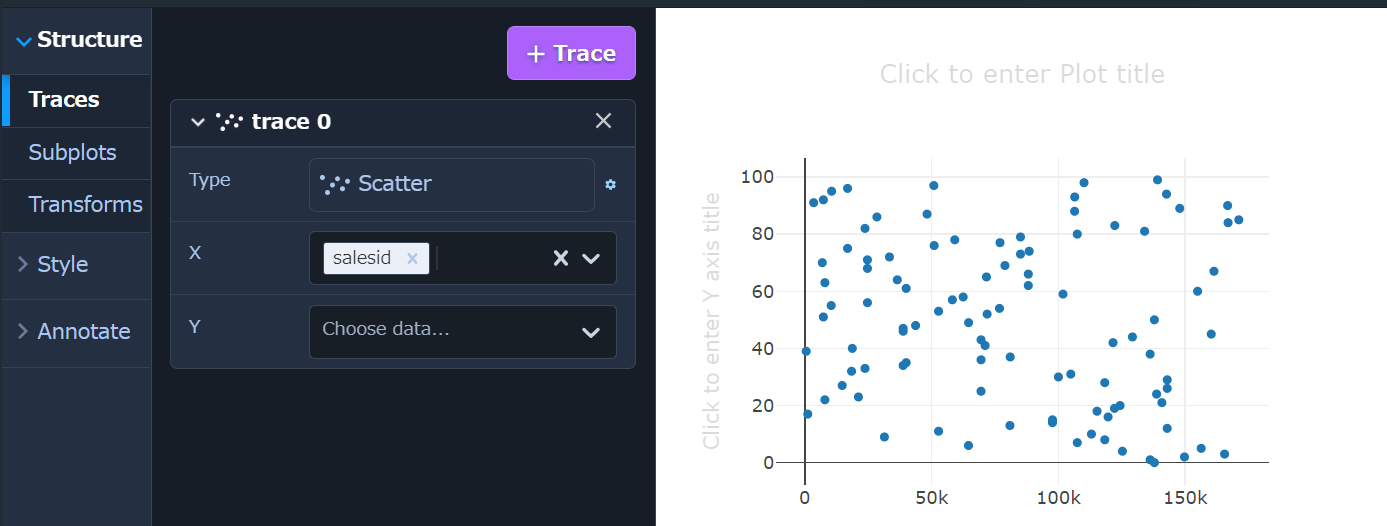
1. [Total Quantity per buyer]を[Run]ボタンをおし実行してみます



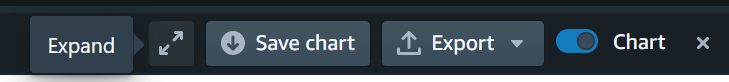
1. 実行には3桁msかかっていることが画面右下からわかります。
2. もう一度[Run]を押し実行すると、実行速度が2桁msになっていることがわかります。これは、Serverlessのコールドスタートではないことに注意してください。Redshiftはコンピュートノードに高速なSSDキャッシュを備えています。これによりよく呼び出されるデータがキャッシュ化され処理が高速化します
3. [Explain]トグルをオンにして再実行すると、実行計画が表示されクエリーチューニングのヒントなどが表示されます



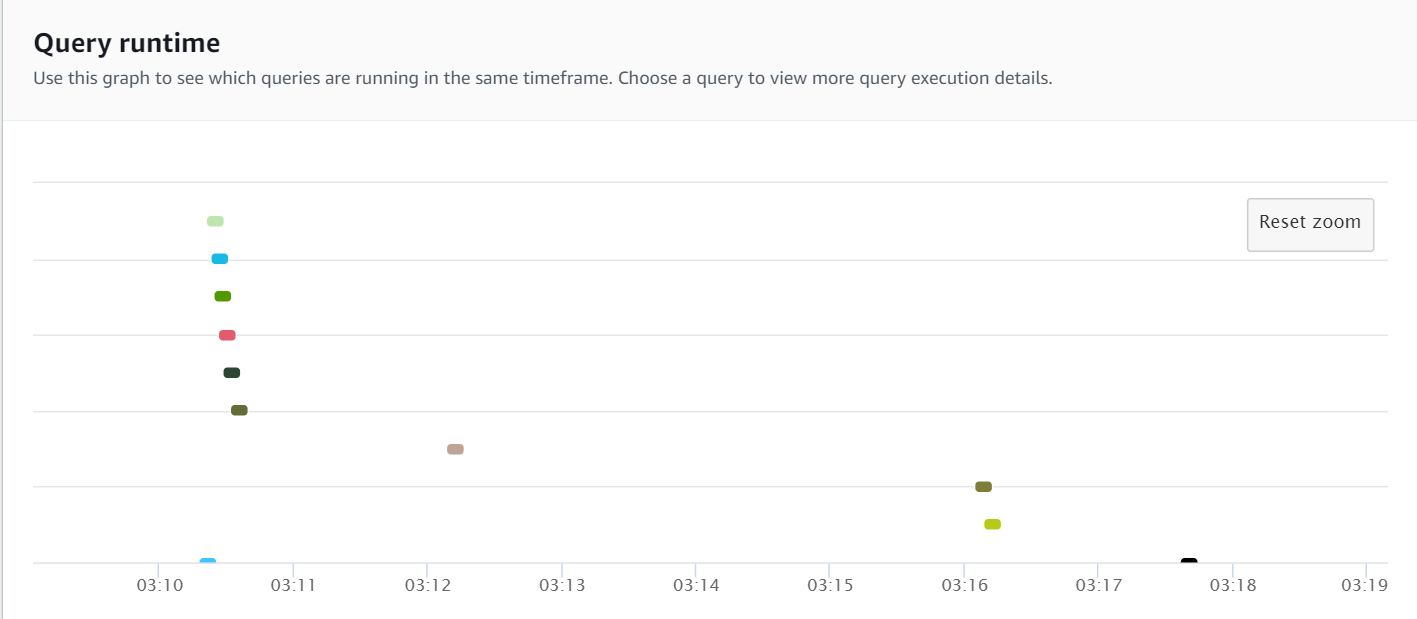
1. 次に[SELECT \* FROM tickit.sales;]を実行した後、[chart]トグルをオンにします
2. 以下のように様々なテンプレートからグラフを生成しデータ分布を視覚化することが出来るようになっています。

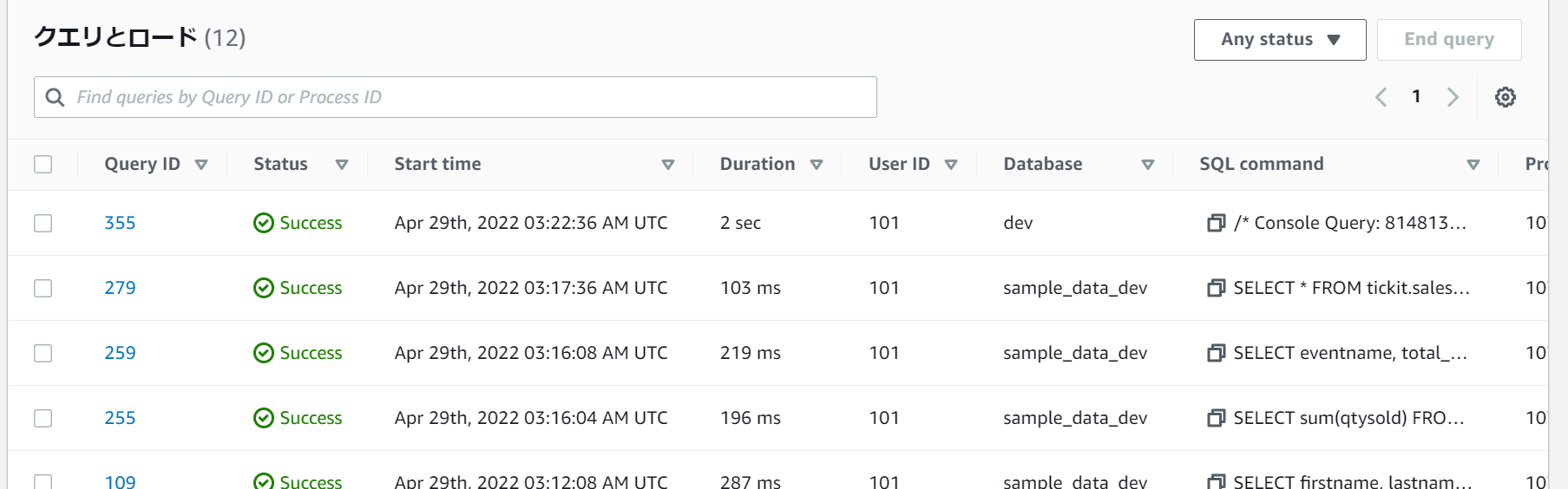


見づらい場合は[Expand]を押してみてください



1. マネージメントコンソールに戻り、左ペインから[Query and database monitoring]をクリックします
2. 以下の通りクエリの実行履歴が出力されています。





1. 実行されたクエリに付与されているQuery IDをクリックすると、実行されたクエリが管理可能な画面に移動します
2. クエリ一覧に対象Databaseが[dev]となっているものがあることに注目してください。これはRedshift Serverless の稼働状況などがdevデータベースで管理されていることを意味します。つまり、マネージメントコンソール上での操作自体も、Redshift ServerlessのCapacityを微小に消費することに留意してください。
3. では続いて、左ペインの [Resource Monitoring]をクリックしてください。[RPU capacity used]等何も表示されていないことがわかります。これは現時点でPreviewが抱える問題でありバグ報告済です。このデータが入っている[dev]のテーブルに対して以下のクエリを発行してみます。Query Editorに戻り以下を実行してください。

select

trunc(start\_time) "Day"

, sum(compute\_capacity) RPU

, sum(compute\_seconds) RPU\_seconds

,(sum(compute\_seconds/3600)\*0.5) "compute\_cost in $"

, sum(data\_storage)/1024/1024/1024 "storage in GB"

from

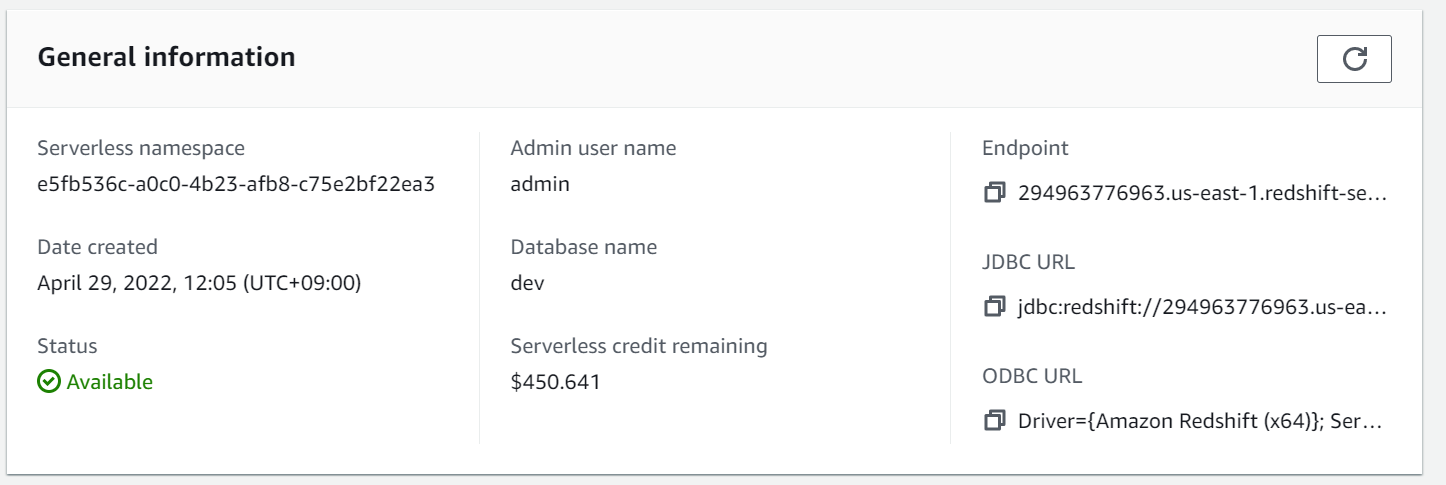
sys\_serverless\_usage

group by

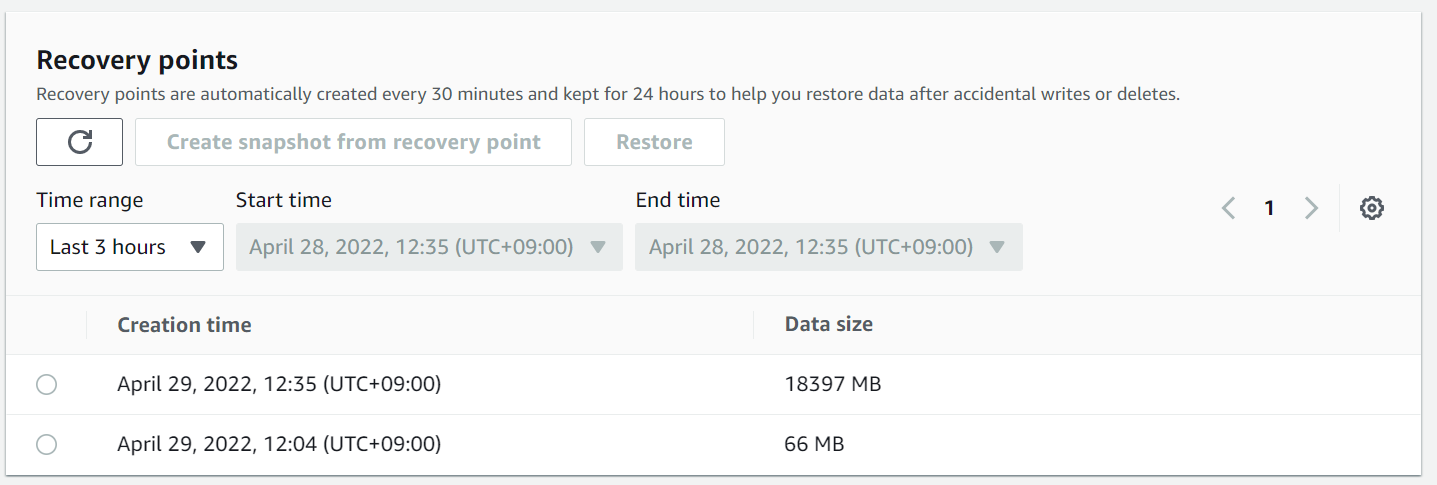
trunc(start\_time);

今までに消費したRPUの累計及び課金額がわかります。

1. マネージメントコンソールの[Serverless Configuration]をクリックします。[Endpoint]が表示されますので、コピーしておきます。



1. [Recovery Points]にいくつかエントリが存在しています。Redshift Serverlessでは自動で30分おきに取得し、24時間保存します。



1. [Limits]タブで[Base capacity in Redshift processing units (RPUs)]の[Edit]ボタンを押します。現在デフォルトとして設定されているRPUを調整することでデフォルトのパフォーマンスを変更させることができます。現時点の仕様では、このCapacityは自動で変動しないことに留意してください。（Amazon Aurora Serverlessと異なり、Lambdaと同じ形態です）
2. つぎに、Redshift ServerlessやQuery Editorのマネージメントコンソールを全て閉じた後、Cloud9のマネージメントコンソールを開きます。（Redshift 関連マネージメントコンソールやQuery Editorは、リロード時点でServerlessにクエリを実行してしまうためです）
3. [Create environment]をおします
4. 適当な名前を設定し、全てデフォルトのままCloud9を起動します。Cloud9はデフォルトでdefault VPC、つまりRedshift Serverlessと同じVPCに起動されます。
5. 以下のコマンドを実行します。RedshiftはPostgresql互換ですので、postgresqlライブラリから操作が可能です。

sudo yum install postgresql-server -y

1. 以下のコマンドを実行します

psql -h <endpoint> -U admin -d dev -p 5439

<endpoint>は先程コピーした値に置換します。その際、文字列後半の[:5439/dev]は取り除いて下さい。

以下のようになります、。

psql -h 294963776963.us-east-1.redshift-serverless.amazonaws.com -U admin -d dev -p 5439

1. パスワードが求められますので、入力をしてください。
2. 以下のSQLを実行すると値が表示されます。

select \* from sample\_data\_dev.tickit.sales limit 10;

1. Redshift Serverlessはセッションを3600秒保持します。このため　1時間以上経過したのち再度クエリーを実行すると以下のようにColdスタートが再現します。の挙動はPreview中に変更があることに留意してください。



おつかれさまでした！

削除は以下をおこなってください

・Redshift Serverlessクラスター

・Cloud9