

## AWS INNOVATE 2020 オンラインカンファレンス

### 「データ分析に最適なクエリエンジンの選定方法」のクイズおよび解答

AWS INNOVATE 2020のセッションの視聴およびアンケートにご記入頂きありがとうございます。  
本資料が「データ分析に最適なクエリエンジンの選定方法」で出題されたクイズの解答になります。

#### クイズ：

業務上マスターデータも含めすべての明細データ、履歴データを保管したいという要件があります。

- ・通常業務では直近1年分のデータを扱い、多くのユーザーがアクセスします。
- ・データ分析業務では約5年分のデータを利用したいとのリクエストがあります。
- ・履歴データは業務上約10年保管が必要で普段はアクセスしませんが、いつでも検索できるようにしておきたいとのリクエストがあります。

以上の要件があるときに、最適なクエリエンジンは何か？

それを判断する理由を考えてみてください。

解答：このクイズでは皆様に考察頂くために明確な要件を記載しておりません。解答では最終的に選択されたクエリエンジンの確認よりも、どのような判断でクエリエンジンを選択されたかの選択過程を重要視し解説しています。

最初に「マスターデータ」、「明細データ」、「履歴データ」の種類がある事に着目します。

“マスターデータ” は一般的に業務上の基礎情報として利用され変更される事が少ないデータです。

“明細データ” は業務システムで発生するデータで参照・更新・削除などのクエリが実行されます。

“履歴データ” は業務上保管しておかなければならないデータで件数が膨大で参照のみに利用されます。

次に説明を俯瞰して見ると、1年、5年、10年という記載があり、データボリュームが想定出来ます。

ここでは最大で10年分のデータがあり、業務に応じて1年分のデータ、5年分のデータを扱うことができます。

このシナリオではデータを扱うユースケースが3つあるので、それぞれ見ていきましょう。

通常業務で扱うアクティブなデータの範囲は1年。多くのユーザーがアクセスします。明細データを扱

うということから参照や更新、削除など様々な種類のクエリが実行されるものと想定できます。データも1年の範囲内で増え続けず常に一定のデータ量を保つと想定してみましよう。

このようなケースでは汎用的に利用可能なクエリエンジンが最適と判断し、Amazon RDSを選択しました。

例えば1年の範囲内でも取り扱うデータが増加していくようなケースでは、参照が中心か、更新が中心か、クエリの特徴を考えてみましょう。参照が中心の場合は他のクエリエンジンの選択も検討します。

次のデータ分析業務では大量のデータに対して分析や集計のクエリが”頻繁に”実行される特性があります。データファイルに対して直接クエリ可能なAmazon Athenaと迷うかもしれませんが、データ分析業務では利用ユーザーが多いことや、分析や集計のクエリでは一時的にデータを書き出す事もあるためAmazon Redshiftを選択しました。

最後の履歴データ。10年分の大量データを保管しているものの、データへのアクセス頻度は非常に少ない事がわかります。アクセス頻度の少ない大量のデータに対してクエリを実行したい場合、Amazon Redshiftに格納してはコストに見合わない可能性があるため、保管データに対して直接クエリ実行が可能なAmazon Athenaを選択しました。

これまで1つのユースケースに対して1つのクエリエンジンの選択肢を解説しましたが、例えば大量のデータを扱うケースにおいて、アクセス頻度の高いデータとあまりアクセスされないデータを結合するようなケースでは、Amazon Redshiftと、本セミナーでは紹介していませんが、Amazon Redshift Spectrum を組み合わせた利用を検討します。

クエリを実行するユーザーのスキルセットも一つの選択肢となります。データ量が多く参照頻度が高いケースにおいて、ユーザーにデータベーススキルがない場合は Amazon Athena や 直接グラフ表示が可能なQuickSightも選択肢となるでしょう。クエリエンジンを選択される際はなぜそれを選択したか、理由を残しておくことを推奨します。

どのような想定でクエリエンジンを選択したのか半年後や1年後にパフォーマンスや運用の課題が出た際にクエリエンジンを見直すための重要な情報となります。

この解答では出来るだけ多くの選択肢について解説しましたが、ここで重要なのはすべてのユースケースを1つのクエリエンジンで対応しなくとも良いということです。

AWSではそれぞれのクエリエンジンを組み合わせ、連携して利用する事が可能です。このセッションおよびクイズを通して皆様のクエリエンジン選択時の参考になれば幸いです。