

データベース 演習問題 (No.8) 解答例

1. リレーショナルデータベース管理システムはビッグデータの管理に向いていないと言われていたが、その理由についてビッグデータの特徴である 4V のうち Veracity (データの正確さ) を除く 3V をもとに説明しなさい。

解答例

Volume (データの量)

データの量が膨大であることに対応する方法の一つは、データベースを格納するサーバの処理性能を上げるとともにディスクを高速かつ大容量にすること (スケールアップ) であるが、これには物理的およびコスト的に限界がある。別の方法は、データベースを複数のサーバに分散 (スケールアウト) させることであるが、リレーショナルデータベースではデータベースの整合性を保証するため、更新処理ごとにデータベース全体を常時監視する必要があり、サーバ間の通信が多数発生し性能のボトルネックになる。

Velocity (データ増加の速度)

データ量が急速に増加することに対応するには、Volume と同様に、スケールアップでは物理的およびコスト的に限界がある。スケールアウトで対応するには、整合性の保証のためのサーバ間の通信の増加がボトルネックとなる。

Variety (データの種類)

データの種類の多様であり、かつ、あらかじめスキーマを決めておくことが困難なデータが多いため、新たな種類のデータが加わるたびに、リレーショナルデータベースではスキーマの修正が必要となる。

2. NoSQL データベースの特徴を、大量データの格納方式、データアクセスの方式、データベースのデータ構成の 3 つの観点から、それぞれリレーショナルデータベースと比較して説明しなさい。

解答例

大量データの格納方式:

リレーショナルデータベースでは、データベースのデータが大量となる場合は、1. でも説明したように、スケールアップで処理性能を向上させるしかない。

NoSQL データベースでは、厳密な整合性を保証するのではなく、結果整合性のみを保証するように変更することで、スケールアウトにより複数のサーバに負荷を分散させて並行に処理することができ、処理性能を向上させることができる。

データアクセスの方式:

リレーショナルデータベースでは、問合せのときにリレーションのどの属性についても検索条件で指定して均一にアクセスできる。また、データベースの更新について整合性が保証される。

キー・バリュー型や列指向などの NoSQL データベースでは、キーを検索条件に指定するアクセスは高速であるが、キー以外の項目を検索条件で指定する問合せは、その項目をキーとするような転置インデックスを別に作成しなければ効率よくアクセスできない。また、データベースの更新に対しては結果整合性しか保証されておらず、タイミングによっては整合性が保たれていないことが起こり得る。

データベースのデータ構成:

リレーショナルデータベースでは、スキーマを必ず定義する必要があり、スキーマに適合するデータしか格納できない。また、リレーションは必ず第一正規形である必要がある。

NoSQL データベースでは、スキーマを定義する必要はない (スキーマレス) ため、多様なデータ型のデータを格納することができ、第一正規形という制約もない。