

データベースの種類と使い方

速川 徹

機械学習ではたくさんの画像、文章、数値データ（目的によるが、数千件や、数千万件）などを学習し、学習モデルを生成します。このとき十分な量のデータをどのように保持するかはモデルの精度を大きく左右します。ビジネスの現場でAIを開発する際も、モデルの精度が上がらない原因はデータが少ないせいだったと言うことがよくあります。しかも、データ量をそろえるには、長い期間、データを集めて保存し続けなければなりません。

そこで本章では、データベースの基礎知識と、クラウド・データベースとして高いシェアを持つAWS（Amazon Web Services）でデータをためる方法を紹介합니다（図1）。

データを自動で保存したいと思ったときに考えるべきこと

● ストレージ容量

最近のSSDやハード・ディスクには数Tバイトものデータが入りますが、無限ではありません。例えば、監視カメラの画像が1枚100Kバイト程度として、これを10秒おきに保存していくとすると、1年で300Gバイトになります。監視カメラが10台あったら3Tバイトですね。パンクさせないためには、古いデータを消したり退避させたりすることが必要です。

● バックアップの有無

ストレージには故障がつきものです。故障でデータを失いたくないのであれば、複数のストレージへ同時にデータを保存するにしたり、定期的にバックアップを取ったりする処置が必要です。当然、予備のストレージも用意しなければなりません。

● ストレージの入るサーバ

SSDやハード・ディスクを稼働させるサーバも必要です。その上、データを保存するシステムを止めたくないのであれば、停電の対策で無停電電源装置（UPS）を用意したり、サーバの故障に備えて複数のサーバを用意したりする必要もあります。

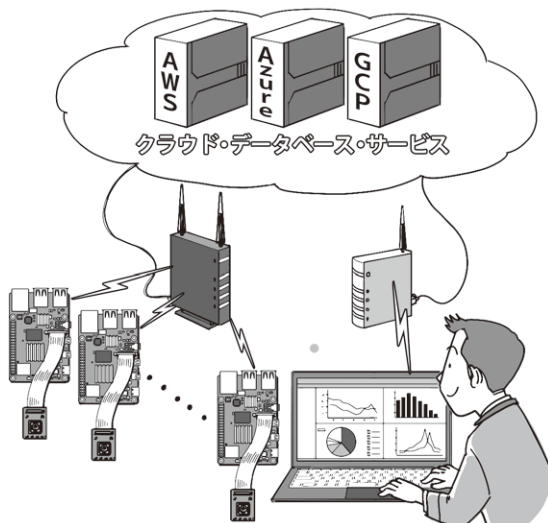


図1 クラウド・データベース・サービスを使ってデータ管理

● 情報セキュリティの対策レベル

人に見られたくないデータであれば、認証なしにはデータにアクセスできないようにしたり、そもそもPCのある場所に管理者以外は立ち入れないようにしたりするといった対処が必要です。

クラウド・データベースなら手軽に始められる

業務の種類によっては多少はデータを失ってしまっても構わないケースもあります。ですが、前述のうち1つでも用意するのが難しい項目があるのなら、クラウド・サービスを使うことが視野に入ってきます。

クラウド・サービスは、インターネット経由でサーバやストレージ、アプリケーションを借りられるサービスです。たいていは従量制で、1分当たりいくら、または1Mバイト当たりいくらというように、時間やデータ量に応じた課金がなされます。クラウドであれば、クラウド事業者がサーバを用意し、ストレージがパンクしそうになったら自動的に容量を拡張し、複数の拠点でコピーを作って故障に対処してくれます。料金面