

ビッグ データと分析について調べる

4 分

数年前、Tailwind Traders では、すべての配送車両に新しい GPS 追跡システムをロールアウトしました。この新しいシステムによって、リアルタイムの追跡データがプライマリ データセンターに提供されます。CTO は、傾向を見極めるため、あなたのチームに数年間の追跡データを調査してほしいと考えています。たとえば、重要な傾向として、休日の前後に配達が増え、追加のスタッフを雇う必要があることが判明したとしましょう。CTO は、あなたが記録した追跡データを詳細に分析することにより、変更が必要なタイミングを予測し、急増を適切に管理するために必要な措置を積極的に講じようとしています。

データは、あらゆる種類のフォームと形式で取得されます。ビッグ データとは "大量" のデータを指します。この Tailwind Traders のシナリオでは、データは GPS センサーから収集されます。これには、位置情報、気象システムからのデータ、および大量のデータを生成するその他の多くのソースが含まれます。これほどのデータ量になると、意味を理解し、意思決定の基準にすることはますます難しくなっています。量が非常に多いため、従来の形式の処理と分析は適切ではなくなりました。

このような大量のデータ セットに対応するために、オープン ソースのクラスター技術が徐々に開発されるようになってきています。Microsoft Azure により、ビッグ データと分析ソリューションを提供するさまざまな技術とサービスがサポートされています。たとえば、Azure Synapse Analytics、Azure HDInsight、Azure Databricks、Azure Data Lake Analytics などがあります。

Azure Synapse Analytics

Azure Synapse Analytics (旧称 Azure SQL Data Warehouse) は、エンタープライズ データ ウェアハウスとビッグ データ分析を組み合わせた無制限の分析サービスです。サーバーレスまたはプロビジョニングされたリソースを大規模に使用することにより、各自の条件に基づいてデータを照会できます。即時の BI および機械学習のニーズに合わせてデータの取り込み、管理、および提供を行うための統合されたエクスペリエンスが用意されています。





Azure HDInsight

Azure HDInsight は、フル マネージドの、エンタープライズ向けのオープン ソースのクラウド分析サービスです。これは、大量のデータをより簡単かつ迅速に処理して高いコスト効率を実現できるようにするクラウド サービスです。 広く普及しているオープンソース フレームワークを実行し、Apache Spark、Apache Hadoop、Apache Kafka、Apache HBase、Apache Storm、Machine Learning Services などの種類のクラスターを作成できます。 HDInsight を使用すると、抽出、変換、読み込み (ETL)、データ ウェアハウス、機械学習、IoT などのさまざまなシナリオもサポートされます。

Azure Databricks

Azure Databricks は、すべてのデータから分析情報を引き出し、人工知能ソリューションを構築するのに役立ちます。 Apache Spark 環境を数分で設定して自動スケーリングし、対話型ワークスペースで共有プロジェクトの共同作業を行うことができます。 Azure Databricks によって、Python、R、Java、SQL だけでなく、TensorFlow、PyTorch、scikit-learn などのデータサイエンス フレームワークおよびライブラリがサポートされます。





Azure Data Lake Analytics

Azure Data Lake Analytics は、ビッグ データを簡略化するオンデマンド分析ジョブ サービスです。ハードウェアのデプロイ、構成、チューニングを行う代わりに、クエリを作成してデ

ータを変換し、価値ある洞察を抽出します。この分析サービスでは、必要な性能をダイヤルで設定して、どのような規模のジョブでも即座に処理できます。実行中のジョブに対してのみ課金されるため、コスト効率に優れています。