

何故Hadoopが重要なのか

Hadoopは、ビッグデータアプリケーションにおいて幅広く使われている。これは、Hadoopの持つ柔軟性が、ビッグデータ環境における抽出、加工、ロード、(ETL: Extract, Transform, Load)を繰り返す動的性質をサポートしているためである。ストレージの近くに処理エンジンを置くHadoopの分散アーキテクチャーは、出力が直接ストレージに行くETLのようなバッチ処理によく適合している。HadoopのMapReduce機能は大量のデータ取り込み処理を小さなかげらに砕いて多数のノード(Map)に送り、次にそれを組み合わせて(Reduce)データウェアハウスにロードする最終のデータセットにする。