

バッチ処理の大幅な高速化を実現する「かんたんHadoopソリューション for バッチ処理 (Asakusa Framework & JP1)」

企業で扱うデータ量が爆発的に増加するなか、バッチ処理の長時間化をいかに解決するかが重要な課題となっています。

日立は大量データの並列分散処理技術「Hadoop」※1と、

バッチ処理プログラムを効率的に開発できる「Asakusa Framework」を組み合わせ

「かんたんHadoopソリューション for バッチ処理 (Asakusa Framework & JP1)」を開発。

バッチ処理の大幅な高速化により業務プロセスの短縮だけでなく、バッチ処理頻度を増やすことができるため、

経営判断などに必要なデータの高精度で迅速な把握と利活用を支援します。

※1 オープンソースソフトウェアコミュニティ「The Apache Software Foundation」にて開発・公開されている、大量データの効率的な分散処理を実現するソフトウェア基盤

Q. 処理するデータ量が増えたことで、バッチの処理スピードが追いつかない状況に悩んでいます。また、エンドユーザー部門から経営判断のため月次処理だったバッチを日次にして欲しいといった要望があります。システム投資を抑えながらバッチ処理時間を短縮したり、処理回数を増やすことができる方法はないでしょうか。

A. いま多くの企業では、事業の拡大やグローバル化にともない、処理すべきデータ量が急増しています。これにより、日々の売上

データの集計や会計処理などのバッチ処理も長時間化し、夜間バッチが翌日の業務開始時間までに終わらない「突き抜け」リスクが高まっているケースも見受けられます。これに対し、サーバ自体の性能を強化するスケールアップ型ではなく、サーバ数を増やすことで大量データを適正コストで高速に処理できるスケールアウト型の並列分散処理技術Hadoopを活用して、バッチ処理時間の短縮が可能です。

日立が提供を開始した「かんたんHadoopソリューション for バッチ処理 (Asakusa

Framework & JP1)」は、スケールアウトに適した高集積のエントリーブレードサーバ「HA8000-bd/BD10」(以下、HA8000-bd)に、Hadoopの商用ディストリビューション※2であるCloudera Enterprise/CDH (Cloudera's Distribution Including Apache Hadoop) (以下、CDH)と、ノーチラス・テクノロジーズ社から提供するHadoop上のバッチ処理に必要な開発・実行・運用環境を実装したAsakusa Framework、統合システム運用管理「JP1」をプレインストールし、導入しやすい初期設定済みのシステムとしてお届けする業界初※3のソリューションです(図1)。

※2 複数のソフトウェアをパッケージ化して提供するもの
※3 「Asakusa Framework」をサーバにプレインストールするソリューションとして(日立調べ、2013年4月17日現在)

Q. 「かんたんHadoopソリューション for バッチ処理」の特長を教えてください。

A. 本ソリューションの適用により、お客さまはHadoopなど関連ソフトウェアのインストール作業や、設定に必要な専門知識を詳しく習得することなく、バッチ処理高速化に向けたシステム構築、立ち上げに必要な期間を短縮することが可能となります。導入時は長時間化している特定バッチ処理のみを移行

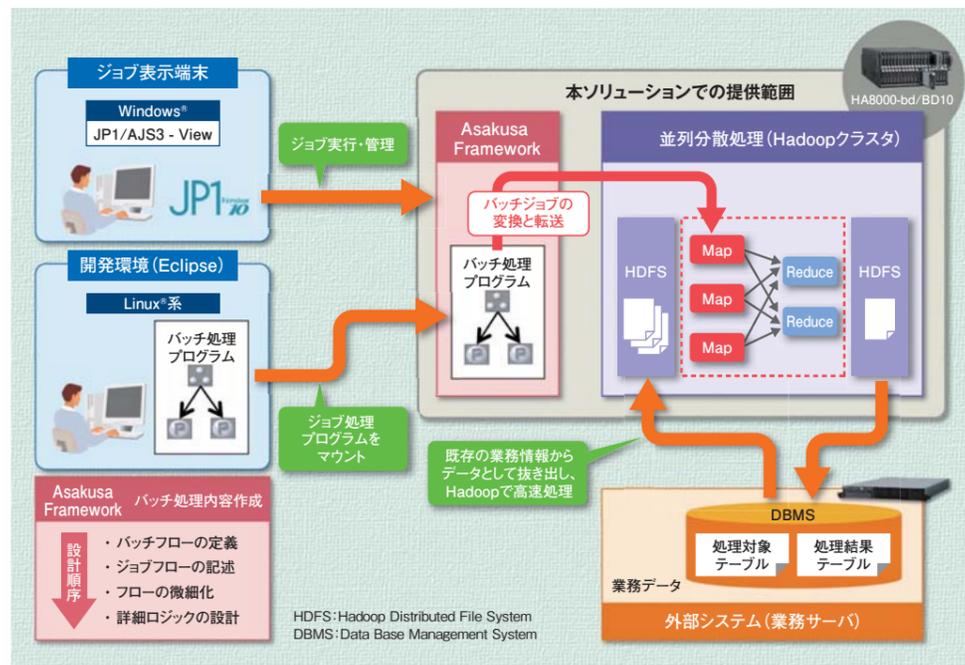


図1 「かんたんHadoopソリューション for バッチ処理」システム全体構成

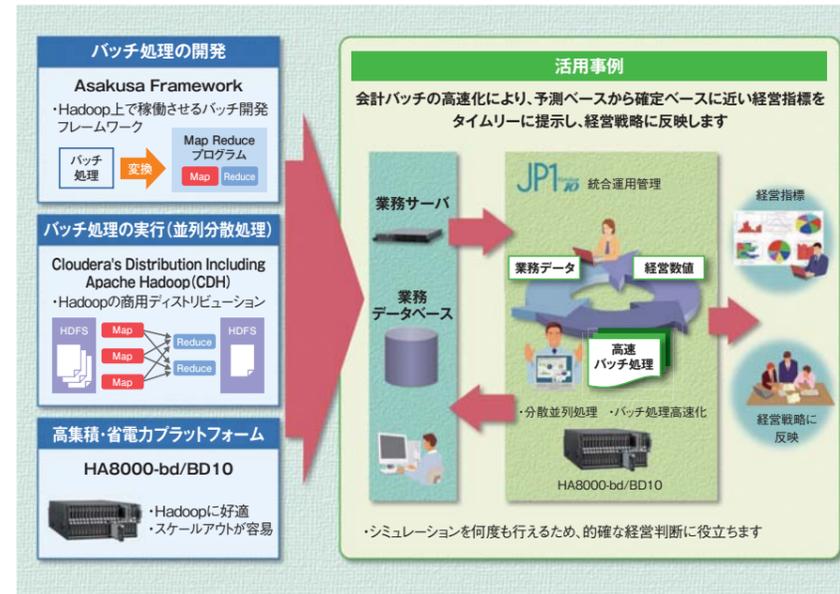


図2 HadoopとAsakusa Frameworkの組み合わせで、バッチ処理高速化を実現

し、徐々にHadoopの適用を広げたり、さらなるバッチ処理の短時間化にはHA8000-bdのサーバブレードを増設するなど、コストを調整しながら効果的にバッチ処理の高速化を実現できます(図2)。

Q. 「かんたんHadoopソリューション for バッチ処理」を導入すると、どのような効果が期待できますか。

A. HadoopによってI/Oボトルネックを分散し、並列実行することでバッチ処理時間を大幅に短縮できます。日立の検証では同一のバッチ処理について、RDBMS※4を実行基盤にした8コアCPU搭載のPCサーバ1台と、Hadoopを実行基盤に4コアCPUを搭載したHA8000-bd 5台と2コアCPUを搭載したHA8000-bd 1台での実行時間を測定・比較した結果、これまで2時間15分かかっていた処理時間を12分33秒に短縮(約11倍の実行速度)できることを確認しました※5。処理速度をさらに高速化したい場合、データ量が増大した場合なども、HA8000-bdをスケールアウトすることで柔軟に対応できます。

その結果、例えば従来なら月次・週次で

実行していた処理を日次でも実行しやすくなります。これにより、財務会計データなら予測ベースではなく確定ベースに近い経営指標をタイムリーに提示できたり、SCM※6分野なら受発注バッチ処理の高速化でサプライチェーン自体のスピードを上げ、商品・部材などの納期短縮によりビジネスチャンスを拡大するなど、お客さまの経営戦略にも大きな効果が期待できます。

※4 Relational Data Base Management System
※5 本ソリューションによる高速化の度合いは、対象とするバッチ処理によって異なり、本結果は11倍の高速化を保証するものではありません
※6 Supply Chain Management

Q. Hadoopシステムの導入検討や、業務バッチの再構築などを任せられる専任者がいません。それでも導入することは可能でしょうか。

A. 本ソリューションでは、Asakusa Framework導入時の簡易診断や実機検証、本番導入時におけるバッチ設計/開発支援まで、また実行基盤となるCDHについて基盤構築や運用、さらにJP1の統合管理設計や連携ソリューションも含め、日立の専門チームがトータルに支援します。お客さまのご要望に合わせたソリューションサービスや

オプションサービスを豊富にラインアップ※7し、

※7 日立製作所ならびに日立ソリューションズが提供

Q. Hadoopによる分散処理システムの運用は、システムを構成するサーバ数が多いためスペース、電力などデータセンターの運用を圧迫しかねないと聞いています。構築後の運用・監視業務やバッチ処理開発の負担が大きくなる心配はないでしょうか。

A. 本ソリューションでは、サーバ台数が多くなるHadoop環境でも省スペース、省電力な運用ができるHA8000-bdと、統合システム運用管理として実績豊富なJP1が、お客さまの運用・監視業務の効率化を強力にサポートします。例えばJP1により既存の業務サーバから抜き出したデータをHadoop上の複数サーバで高速分散処理し、結果を再び業務サーバに返す一連のフローを、これまでの通常業務と一元化された画面から運用・監視できるため、管理者の負担を増やす心配がありません。

また、Asakusa Frameworkで開発したバッチ処理プログラムを、「JP1/AJS3」※8のジョブ定義に自動生成する「JP1ジョブ定義生成サービス(オプション)」※9も提供するため、Hadoopを含むバッチジョブの定義を効率化できます。

※8 JP1 Automatic Job Management System 3
※9 株式会社日立ソリューションズが提供

以上にご紹介した「かんたんHadoopソリューション for バッチ処理 (Asakusa Framework & JP1)」は、2012年5月から提供しているアクセスログ解析ソリューション「かんたんHadoopソリューション for ログ解析 (QlikView & JP1)」に続く、かんたんHadoopソリューションの第2弾となるものです。今後も日立は、Hadoopに関する豊富な実績とノウハウを生かした付加価値の高いHadoopのビジネス活用ソリューションを、積極的に開発・提供していきます。

お問い合わせ先

HCAセンタ
☎0120-2580-12
利用時間 9:00~12:00、13:00~17:00(土・日・祝日を除く)

情報提供サイト

<http://www.hitachi.co.jp/ha8000-bd/>

※Apache Hadoopは、Apache Software Foundationの米国およびその他の国における登録商標または商標です。
※Asakusa Frameworkは、株式会社ノーチラス・テクノロジーズの登録商標です。

※Clouderaは、Cloudera, Inc.の登録商標です。
※その他記載の会社名、製品名は、それぞれの会社の商標もしくは登録商標です。