ビッグデータ技術を活用した
バッチ処理の高速化
ーバッチ処理時間１／１０も夢じゃない
～Hadoop 適用ノウハウ教えます～ー

アブストラクト

1. 研究の背景

近年、企業で取り扱うデータ量が日々増大することにより、夜間に行うバッチ処理が翌朝のオンライン開始時刻までに終了せず、バッチ処理だけでなく通常業務にも支障が出始めている、といった事例を散見する。その一方で、オンライン時間の延長が求められており、バッチ処理を高速化し処理時間を短縮することが急務となってきている。

今までとは一線を画する発想と技術が必要であり、その解決手段のひとつとしてビッグデータ技術に白羽の矢が立った。

しかし、一言にビッグデータ技術と言っても様々な特性をもつ技術が存在するため、利用する際の明確な指標や指針が見当たらない。このため、ビッグデータ技術の導入効果や、システム投資の際の費用対効果が判別できないため、システム部門及び経営トップでも導入判断を下すことができず、結果として適用には至らず悪循環に陥っている。

2. 研究のアプローチ

まず、ビッグデータ技術適用の実態を調査するために、適用事例の収集を行った。

数ある適用事例の中でも Hadoop に焦点をあて、高速化に寄与する要因の分析を行った。

分析結果を基に、既存バッチへの Hadoop 適用ガイドラインをまとめ、メンバーにて机上検証を行いガイドラインのブラシュアップを行った。そして、分科会メンバー各社において第 3 者アンケートを行い、ガイドラインの有用性を検証した。

3. 研究内容/研究成果

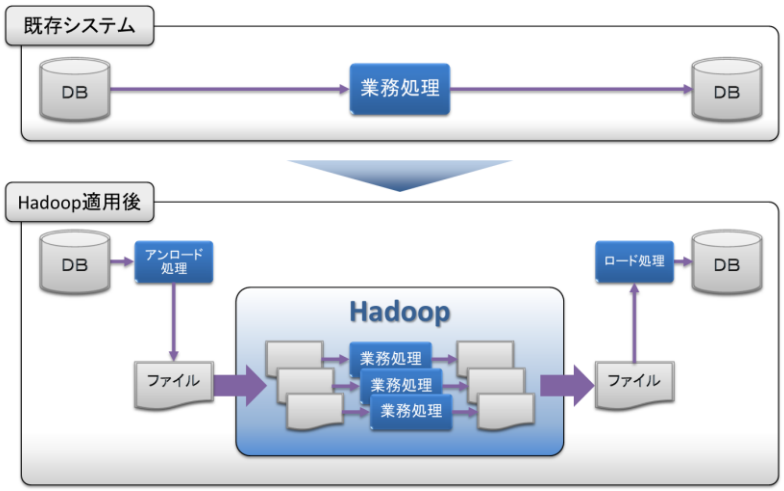
既存バッチ処理への Hadoop 適用事例収集と、Hadoop アーキテクチャーの両面を分析することで、Hadoop の適用可否判断を行うための要素を抽出した。

図表　1　Hadoop 適用可否判断

	共通点		理由
1	Hadoop適用箇所	バッチ処理全体をHadoopに置き換えるのではなく、一部のジョブステップに限定してHadoopを適用している。	Hadoop上でアプリケーションを実行するには、MapReduceアプリケーションを開発する必要がある。そのためバッチ処理全体をHadoop上で処理するのではなく、MapReduceで処理すべきジョブステップのみにHadoopを適用する。
2	処理内容	Hadoopを適用したジョブステップの処理内容は、データの結合やソートなどが多い。	複数のキーそれぞれに対して結合を行うような処理は、キー毎に並列実行すること高速化が可能。 ソートはMapReduce処理内で行うShuffle&Sortが利用できる。
3	データ	処理対象となるデータサイズは、分散配置の効果が表れてくる1GB以上である。	Hadoopでは、入力データを分散させる（並列化）オーバヘッドが存在する。目安として、入力データサイズが1GB以下の場合、分散させるオーバヘッドの割合が高くなり、高速化が見込めない。

さらに、当分科会メンバーが持ち寄った「遅いバッチ処理」に対して、適用が可能か机上検証を行った。机上検証を行うことで、既存バッチ処理へ Hadoop を適用するための考慮すべきポイントが見えた。

図表　2　既存システムへの Hadoop 適用箇所と実行環境イメージ

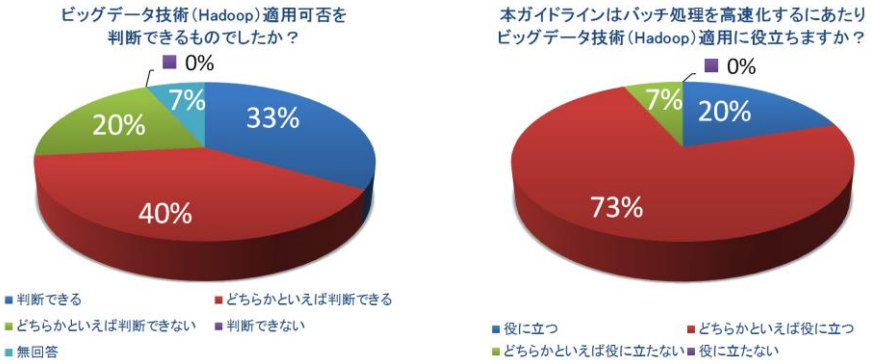


これらの結果を基に、Hadoop 適用可否判断だけではなく、Hadoop アーキテクチャーを考慮した設計を行うためのノウハウを盛り込んだガイドラインを作成した。

ガイドラインを作成後、第 3 者によるアンケートを実施した。

アンケートを行うことで、ガイドラインの有用性を調査すると共に、ガイドラインへの要望を拾い上げる事ができた。アンケートによるガイドラインの評価結果は以下のとおりである。

図表　3　ガイドライン評価結果



4.　研究の総括と提言

今回の研究で Hadoop アーキテクチャーを調査して Hadoop とバッチ処理の相性の良さを確認することができた。また、Hadoop 適用事例より明らかにした既存バッチ処理への適用ポイントは、今後、バッチ処理に対して Hadoop を適用する際の重要なポイントとなる。これらをまとめた Hadoop 適用ガイドラインは、そのアンケート結果からも有用性を確認できた。

ただし、今回の研究を通じて Hadoop に対する認知度はまだまだ低く、バッチ処理の高速化の手段としてビッグデータ技術を検討している人が少ないということもわかった。そのような現状において、今回の研究成果であるガイドラインは、Hadoop に対する入門書、バッチ処理に対してビッグデータ技術を適用するための一つの指針として有用であるといえる。

今後、Hadoop がバッチ処理の高速化手段として浸透してきた際には、本研究がその礎の一部となっていることを期待したい。