



5 Steps To

# OFFLOAD

YOUR DATA WAREHOUSE

WITH HADOOP

A QUICK STARTGUIDE

*Free Up Your Data Warehouse...  
and Budget*

**syncsort**

# 目次

## 序論:

ETL の概要と経緯

## *TheHadoop* の出番

### ステップ1:

事業目標の理解と定義

### ステップ2:

**Hadoop** に適切な接続を

### ステップ3:

ETL/ELT ワークロードのトップ 20% を特定

### ステップ4:

**MapReduce** と同等の処理に置換

### ステップ5:

自社向け **Hadoop ETL** 環境設定

## 結論

# ETL の概説と経緯

古くから「知識は力なり」と言われています。しかし、「データ」が「情報」を生成し、「情報」が「知識」を生成する場合、それは本当に企業にとって「力」となるのでしょうか？企業が「データ」に焦点を置く度合いが増えている理由は、より戦略的なビジネスの意思決定を可能にする手段として考えているからです。では、一体費用は幾らかかるのでしょうか？

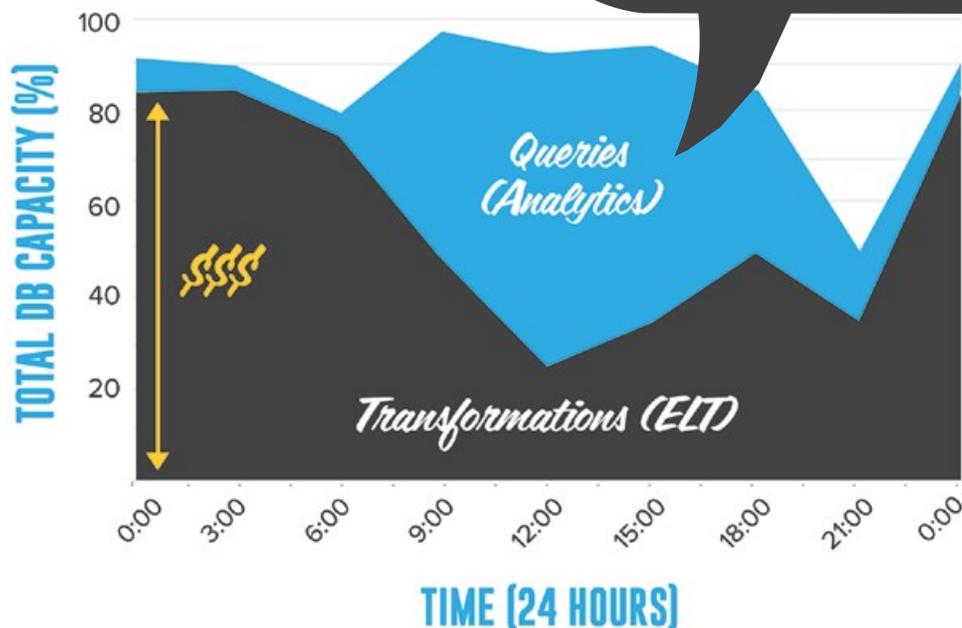
今日データの世界では、「何兆ギガバイト」もの膨大なデータが存在します。

これらを最大限に活かす鍵を握るのは、データ統合、より正確に言えば、ETL（抽出、変換／加工、書き出し）です。

ETL とは、ソース・システムからデータを取りだし、利用できる形に変換し、高度な分析及びレポート生成を行なう為のターゲットシステムへ書き出されるプロセスです。企業は何年にも渡り 3V と呼ばれる、大容量（high-Volume）、高速（high-Velocity）、そして様々なデータ資産（high-Variety of data assets）のビックデータに追いつく為に、従来の ETL アーキテクチャの拡張に苦労しています。IT システムはデータ増加のペースについていけず、その役目は ETL 処理からデータウェアハウスに移ることを余儀なくされ、ETL から ELT（抽出し、データウェアハウスに書き出し、データウェアハウス・エンジンで変換／加工する）というトレンドを生んでしまいました。しかし、これは費用がかかり非効率的なアプローチである事が判明しました。

データ統合により増加したデータはデータベース能力の 80% まで消費してしまい、結果持続不可能な支出、継続的なチューニングや保守の工数、またクエリパフォーマンスの悪さにつながりました。

## THE TRUE COST OF ELT



# 70%

全てのデータウェアハウスの70%は、  
パフォーマンスと能力の制約を持つ

(出典：ガートナー)

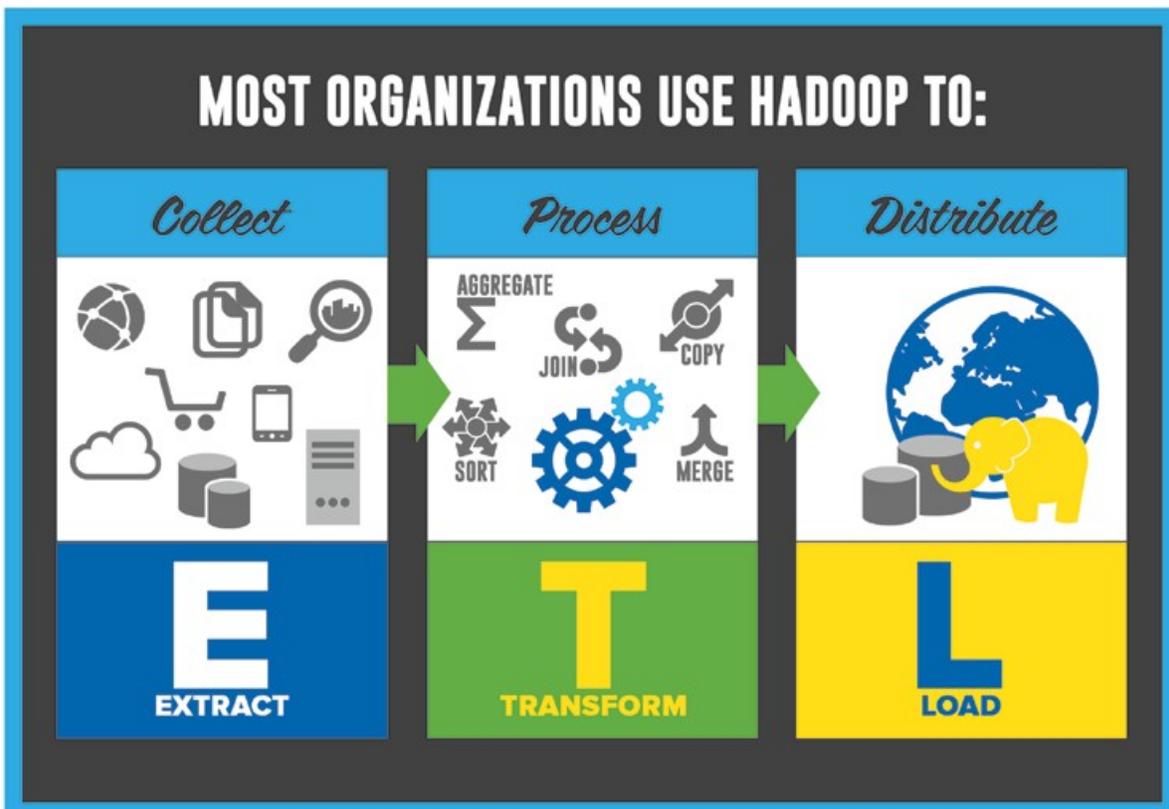
多くの企業がデータ統合ツールを使用する際直面する一番の課題は、TCOです。

一方、データ量は増え続けています。

事実、IDCの調査では企業が抱えるデータ量は2年で2倍に増えているという報告があります。

そういった終わりのないデータ爆発の観点から、企業は今後ますます増大するデータの収集、処理及び配布に備えて、Hadoopを検討し始めます。

企業はETLをHadoopへ移行させる事によって、エンタープライズ・データウェアハウス(EDW)から重い処理をオフロード(負荷軽減)し、迅速な分析とエンドユーザの照会の為にデータベース処理にかかる費用を劇的に削減し、データベース能力を解放します。



# Hadoop の出番

Hadoop はビッグデータ管理のための新しいオペレーティング・システムとなりました。その無限の拡張性により、開発者は破格の費用でビッグデータ・アプリケーションを生成する事が出来るようになりました。

複数の情報ソースからわかる価格(データ単価)は、ハイエンド・データウェアハウスがテラバイト辺り\$20,000 から \$100,000 かかるのに対して、Hadoop ではデータのテラバイト辺り\$250~\$2,000 です。

多くの企業は Hadoop の効率とパワーを生かした HadoopETL の取り組みに着手していますが、Hadoop はまだ完成された ETL ソリューションではありません。

Hadoop は強力で事実上無制限な水平方向のスケーラビリティを提供していますが、ユーザがエンタープライズ ETL で必要とする全ての機能を提供していません。

ETL から Hadoop への移行は、複雑な手作業でのコーディングが必要なこともあり、Hadoop 導入が遅れ、結果を出したい企業がなかなか思う様にいかないといった弱点がありました。

そこで、ETL ソリューション全ての利点を得つつ、Hadoop でのハイパフォーマンス ETL とエンタープライズ・データウェアハウス (EDW) の解放を両立する方法があります。

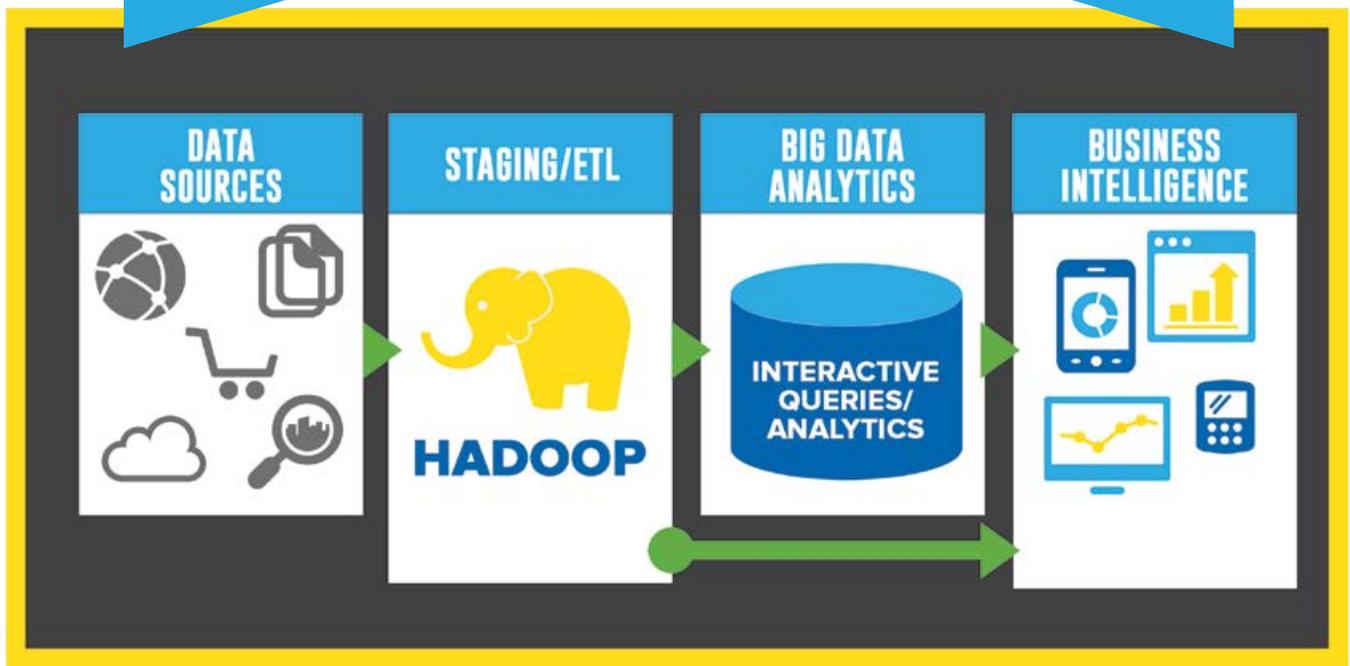


Syncsort Hadoop ETL ソリューションは、Hadoop とエンタープライズ ETL 間の溝を埋め、Hadoop をより頑丈で機能豊富な ETL ソリューションにします。

Syncsort のハイパフォーマンス ETL ソフトウェアは、従来の ETL ツールの使いやすさを損なう事なく、MapReduce の利点を最大化します。

Syncsort Hadoop ETL ソリューションは、Hadoop 内でネイティブに処理をする唯一のアーキテクチャを持っているため、企業が Hadoop の可能性を最大限に発揮する事ができます。

# ビッグデータ分析のための 新しいアーキテクチャ



このガイドでは、課題を克服してデータ統合の取り組みを加速するエンタープライズ・データウェアハウス（EDW）の Hadoop オフロード（負荷軽減）を始めるのに役立つ5つのステップを紹介しています。

どんなツールを選択したとしても、これらのステップはハイパフォーマンス分析と高速なユーザ・クエリを提供し、柔軟にエンタープライズ・データウェアハウス（EDW）開始に対応することができます。

## 事業目標の理解と定義

多くの企業は、既存のデータ統合基盤の構築にかなりの時間とお金を費やしてきました。しかし、従来のアーキテクチャを使用して情報量を増やすというのは色々な課題に直面します。

- **費用：**

重いデータ変換をエンタープライズ・データウェアハウス（EDW）に依存してしまうと持続不可能な費用がかかります。データベース・ワークロードの40%～80%使用するELT処理の為に、年間の追加容量に最高30万ドルまで費やす企業は珍しくありません。

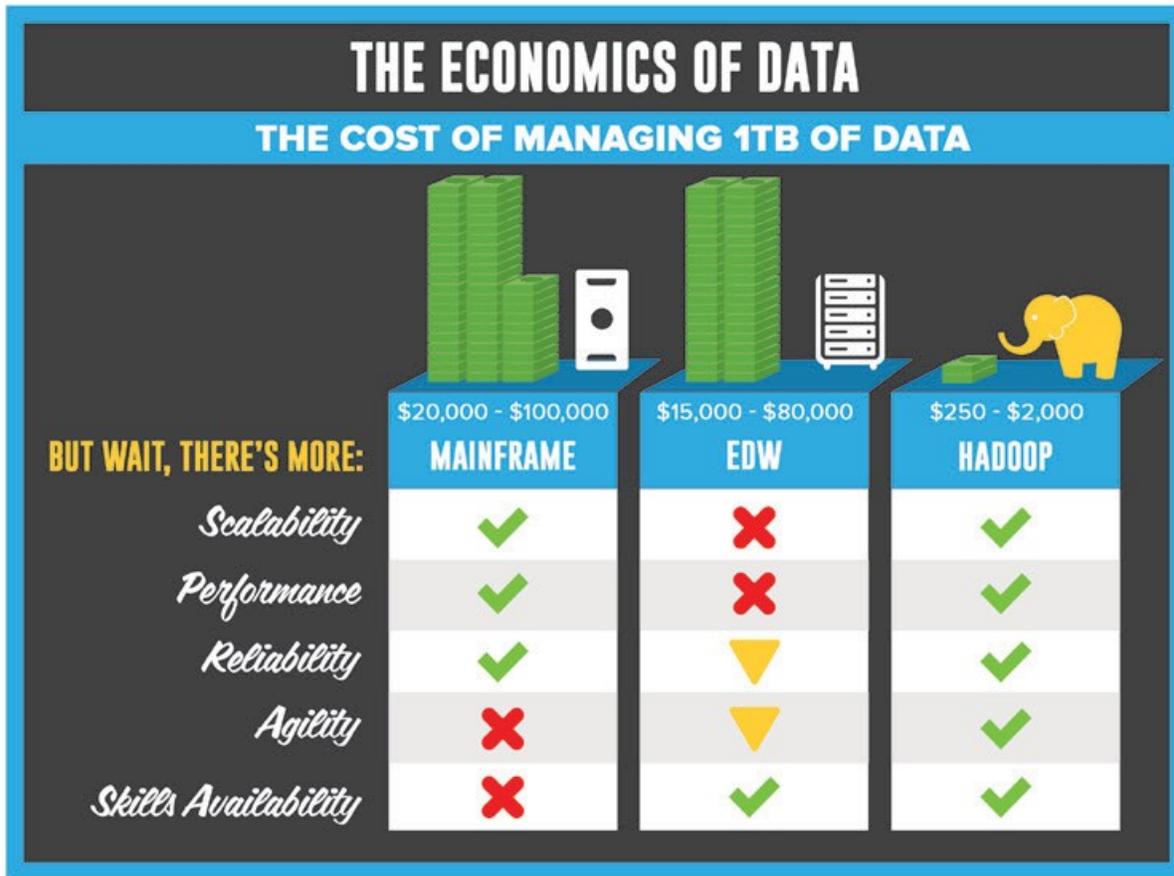
- **データ量：**

最適なビジネスの意思決定を行い、新しいビジネスチャンスを発見する為に、増加するデータ量とデータソースへ接続し管理するのは、従来の技術と手動での対応ではもはや不可能です。

- **サービス品質保証（SLA）：**

既存のインフラ環境とサービス品質保証（SLA）が適応しない場合、レポート生成に数時間から数週間も要するボトルネックを増加させる傾向にあり、ユーザのストレスも増加する傾向にあります。

コスト削減、より多くのデータ活用、またサービス品質保証（SLA）に適合する等どの目的に対しても多くの企業は別のアプローチを探す事を余儀なくされています。



Hadoop 内にエンタープライズ・データウェアハウス（EDW）の、重い ETL ワークロードを移行させる事により、企業は下記の重要な利益の享受を迅速に実現化する事ができます。

- バッチウィンドウの短縮化
- より最新で、関連性の高いデータ
- より速い洞察への到達
- より高速なデータベース・ユーザのクエリパフォーマンス

Hadoop は ETL ツールとしてデザインされているのではなく、あくまでオペレーティング・システムであるという事を踏まえておく事が重要です。正しいツールを使う事でビッグデータの能力を可能にしますが、Hadoop に高性能な ETL ソリューションで行う全ての処理は期待できません。

第一に、熟練した Hadoop のプログラマを見つける事自体が困難です。MapReduce 中の ETL ジョブを Pig、Hive、Sqoop で作らなくてはなりませんが、多くの企業は社内にそういった人材を置いていません。

第二に、直観的なインターフェース、メタデータ、再利用性、及び生産性の向上と接続性といった“エンタープライズ ETL”機能を諦めざるをえないということです。

そして、複雑なコーディングへの回帰は、長期の開発サイクルへの回帰、並びに保守及び企業が解決しようと試みている問題に再び直面することを意味します。

第三に、間違った ETL アプローチは余計なオーバーヘッドを生み、パフォーマンスに影響を与えます。

Hadoop は水平方向の拡張性、多くのノード間でデータ分散処理に優れていますが、各ノードの効率を最大化する為の ETL プロセスの最適化を行わなくてはなりません。

## 最善手法

ビジネス目標に対して下記の機能を確認する：

- 既存 ETL 開発者のスキルセットを活用できコストを抑える為の直観的なインターフェースを備えた **GUI**。
- 再利用、生産性を高める為に **メタデータは必要です**。  
これにより IT のバックログを削減し、サービス品質保証 (SLA) に合ったより速い開発を実現します。
- Hadoop の垂直方向の拡張性を最大化する為の自動 **最適化機能**。  
数少ないノードと短い時間でより多くのデータを処理し、コストを削減。
- **ノーコーディング** によって長期に渡る開発期間、非効率なコードによるノードあたりの性能の低下、リソース不足が原因で起こる高価なハードウェアコスト等、従来のデータ統合の欠点を Hadoop に持ち込まない。
- **Hadoop を ETL に使う、または Hadoop と他の ETL ツールを使うことに比べ、** サービス品質保証 (SLA)、費用、拡張性、そして接続の要件に合うか確認する「ストレステスト」と実現性の確認に注力することができる。
- 重い ETL ワークロードを簡単に移行できる **クラウドソリューション**。  
簡単に Hadoop クラスタを得る為、一部の企業は Amazon Elastic MapReduce (EMR)としてクラウドのアプローチを検討するかもしれませんが、クラウドの生産性と迅速な結果を出す為に適切なツールを得る事が必要です。

# Syncsort が如何にして役立つか

**Syncsort Hadoop ETL ソリューションは、企業が ETL の取り組みに Hadoop を使用する事でビジネス目標の達成を助ける為に設計されています。**

- ノードごとにより速いパフォーマンスと効率性を提供する、MapReduce ネイティブの処理。
- コード生成ではない： MapReduce は Hadoop に欠かすことができない全てのノード上で稼働する Syncsort ランタイムを呼び出す。
- 「コーディングなし」アプローチ：複雑な Java や Pig、HiveQL コード等を強力で使いやすいグラフィカルな開発環境で置き換える。
- Amazon クラウドの大規模な拡張性を活用しながら、生産性を加速させる。AmazonEMR 上の HadoopETL の為の最初で唯一の「ETL-as-a-Service」としてのソリューション。

## Hadoop に適切な接続を

ビジネス目標の定義ができれば、次のステップでは必要とする全てのソースとの接続を確保します。そうすれば、Hadoop は企業内でサイロにならずに済みます。ビッグデータは、RDB、ファイル、CRM システム、ソーシャルメディアといったデータソースとターゲットの大量のリストを指します。

企業は、レガシーとメインフレームといった、他のソースを考慮する必要があります。

世界中のトランザクション開発のアプリケーション中、最低70%がメインフレームで実行されているので、Hadoop を使用してより深い分析を妥当な費用で実施する事が重要です。

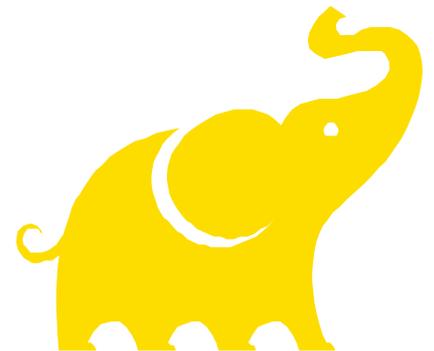
Hadoop はメインフレームにネイティブサポートを提供しておらず、面倒な手動での作業になります。

***Hadoop の分散ファイルシステム (HDFS) でロード時間を短縮する事は、毎日数十億というデータを取り扱う企業にとって重要な事です。***

アプリケーションとビジネスの需要が高まるにつれ、ロード時間を削減する必要性がさらに高まります。

もちろんエンタープライズ・データウェアハウス (EDW) への合理的な接続は必須です。しかし、多くの異種ツールや手動コーディングを含む接続のアプローチは毎日何らかの変更があり、IT チームはかなりの時間と労力を費やす必要があるので洞察に至る時間が妨げられます。従って、単一のツールでコーディングなしで全てのソースに効率的に接続できるかが鍵となります。

また、ソート、クレンジングフィルターへの事前処理機能、そしてそれを圧縮する事でさらに性能を強化し、ストレージスペースを節約します。



# 最善手法

ビッグデータから価値を引き出すためには広範囲に渡るデータ接続が必要です。

より簡単により速くデータ接続を行えば、データ及び Hadoop 投資から派生するであろうたくさんの価値を生み出します。

- リレーショナル、クラウド、ファイルそしてメインフレーム・ソースを含んだ様々な **ツール選択** は、簡単に Hadoop へインポート、また Hadoop からエクスポートを行う。
- 最大効率を導く **事前処理** データを実行するアプローチを特定する。
- **将来の保証** としてパートナーとオープン API やオープンソースを役立てる。

## Syncsort が如何にして役立つか

Syncsort Hadoop ETL ソリューションは、Hadoop ETL 展開のための重要な接続機能を含みます。

- 単一ツールで、RDB、アプライアンス、ファイル、JSON、XML、クラウド、さらにはメインフレームを含む全てのデータソースに接続。
- コーディングなしで Hadoop を全てのデータへ接続。
- パフォーマンスとストレージ節約の為 Hadoop にローディングする事前データ処理。
- Hadoop のメインフレームデータを読み、変換、配布するユニークな機能
- ニーズの進化に伴う新しいデータ・ソース及びターゲットに対して接続の機会を提供する データ接続 API

## ETL/ELT ワークロードのトップ 20%を特定

開始する準備が整った所で、次はどの様に優先順位を付け、最も高い効果が出せるかを策定・実施します。

長年、多くの企業はエンタープライズ・データウェアハウス（EDW）を使用したデータ統合にかかるコストと処理制限に頭を痛めてきました。一番の最善手方はステージング領域を構築する事ですが、データウェアハウス環境ではステージング領域は時間とお金と労力を吸い込む「秘密領域」になります。従って、安価なストレージ、高い信頼性、大規模な拡張性を提供する Hadoop は全てのデータの為に理想的なステージング領域を作ります。

Hadoop 内で変換とバッチ処理を行う事は従来のデータ統合の重大な欠点を補うことを意味します。

一方で、Hadoop の価値を証明するためには、最初に目標とするデータ変換を特定するのが鍵となります。

Hadoop の推進と経営層のサポートを得るために必要なのです。

*通常 ETL/ELT ワークロードの上位 20%にハードウェア、チューニング、保守費用に 80%を消費しており、これらを劇的に削減する事ができます。*

これらの変化を目標とすれば、最高の ROI（投資収益率）を短時間で達成します。

この 20%に着目すると、HadoopETL の成果を最適化し、浮いた資金をより戦略的に活用する事ができるようになります。

# 最善手法

Hadoop への移行の為に変換の上位 20%を特定する場合、下記の課題が考えられます。

- 比較的速い処理経過時間。
- 変更データキャプチャ (CDC)、SCD(Slowly Changing Dimensions)、かき集め機能(raking functions)、発揮性テーブル、複数のマージ・結合、カーソルとユニオンを含めた非常に複雑なスクリプト。
- WEB のログやクリックストリーム分析の様なファイルと半構造化データ。
- CPU、メモリ、ストレージを含めたリソース影響度の高いもの。
- 不安定またはエラーの発生しやすいコード。

## Syncsort が如何にして役立つか

Syncsort ソリューションを使用すれば、Hadoop による ETL で下記を可能にします。

- 自動生成されたフローチャートを利用した複雑な SQL スクリプトの分析と理解。
- 何千行ものコードをいくつかのグラフィカル Syncsort HadoopETL ジョブに移行。
- 総合的なログとデバック機能を介した、データエラーの迅速な特定及びトラブルシューティング。

## MapReduce と同等の処理に置換

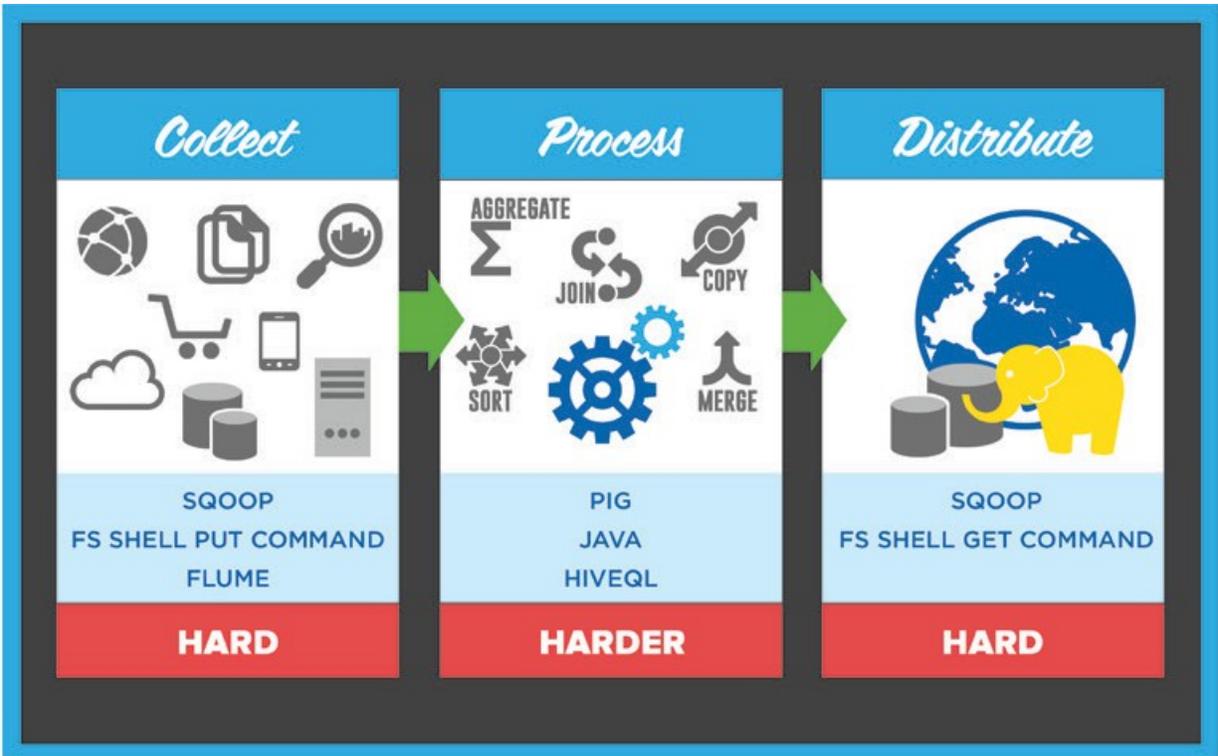
ETL ワークロードを特定したら、エンタープライズ・データウェアハウス (EDW) で全てのデータを統合し、Hadoop 内でそれらを複製する方法を理解しなければいけません。

限られたメタデータ及びドキュメントと共に作成された Oracle PL/SQL の様な独自の SQL の場合は特に困難です。SQL ベースの複雑な ETL ジョブを Hadoop ユティリティを使用して生成する事は可能ですが、全て個々のプロジェクトなので特別な新しいスキルを必要とします。開発者は、MapReduce ETL ジョブを開発する為に Java、HiveQL、Pig に精通している必要があります。

これらのツールは生産性を低下させるハンドコーディングだけではなく、多くの場合 Hadoop と MapReduce の深い理解、特にユーザー定義関数 (UDF) の指定が必要となります。

また、スクリプトの中には追加のオーバーヘッドを発生することがあり、専門の開発者によって書かれた場合でも最適なパフォーマンスを達成するためには幾度もの試行錯誤が必要となります。

# 今日の Hadoop 内 ETL



Hadoop とエンタープライズ ETL 間の溝に起因してソート、ジョイン、集計、及び接続の様なタスクは、複雑な手動のコーディングを必要とし、従来のデータ統合の非効率性と高い費用を要すると言う同じ轍を踏むことになります。

今日 ETL で幅広く使用されている変更データキャプチャ (CDC) 処理の様な、より洗練されたデータフローを Hadoop に構築する事は困難です。データセットは通常非常に大きく HDFS のデータノードに分散されています。レコードが変更を明確にする為には同じ場所に配置する必要があります。また許容可能なパフォーマンスにするためにはハンドコーディングとチューニング (簡単に数百行を越える自家製コードになる) が必要となります。

# 最善手法

エンタープライズ・データウェアハウス（EDW）のオフロード（負荷軽減）処理を開始する為の近道：

- エンタープライズ・データウェアハウス（EDW）で発生する複雑な変換及び処理を、**分析、理解、文書化する事から始めます。**
- 一般的な ETL 処理は、**ポイントアンドクリックインターフェースを備えたレバレッジツールで最初に移行します。**
- **手動コーディングを必要とするツールは避けます。**
- 事前に構築されたデータ統合機能とグラフィカルなユーザーインターフェースが、**既存のプログラミングスキルで活用できることを確認してください。**

## Syncsort が如何にして役立つか

**Syncsort Hadoop ETL ソリューションは、下記により Hadoop ETL ジョブの作成／維持を容易にする高度なデータ統合ツールです。**

- Windows ベースの GUI 上でローカルに開発及びテストし、Hadoop に展開できる。
- Mapper と Reducer の両方で総合的な組み込み変換 MapReduce ジョブがすぐに利用可能。
- 大幅にラーニング曲線を上回り、Hadoop における一般的な ETL タスクの迅速な開発を実現するライブラリー群（CDC、集計、結合など）。
- 内蔵されたメタデータ機能で再利用性、影響分析、データ系統を増加させる。

## 自社向け Hadoop ETL 環境設定

ビジネスを成功に導く為に、ETL 処理を Hadoop へ移行して強固な基盤を築く必要があります。

Hadoop への移行を阻害するのは、期限と ETL 処理のパフォーマンス（サービス品質保証（SLA））が出ないことに他なりません。

また、企業はより迅速かつ低コストで、従来のエンタープライズ・データ統合ツールで提供される機能とサービスを期待しています。

しかし、エンタープライズ・レベルでの Hadoop 導入は無償ではなく、大量のハードウェアと IT 生産コストが ROI を悪化させる可能性があります。

Hadoop は一般的なハードウェアを活用しますが、多数のノードを扱う場合ハードウェアの費用がかさみます。

HiveQL、Pig、Java、MapReduce といったプログラムは高額であり、生産性と折り合いをつける必要があります。

多くのデータ統合ツールは Hadoop の周辺で作動します、Hive、Pig もしくは Java を生成しますがオーバーヘッド層が増えるのでパフォーマンスが低下します。Hadoop に統合されコード生成を回避している ETL ソリューションは、パフォーマンスへの影響やロードの障害なく展開と保守が容易に行えます。エンタープライズ・データウェアハウス（EDW）でのデータ変換を行う課題のひとつはメタデータの欠如で、分析に影響を与え、ジョブの追跡や再利用が不可能になります。所定の何千ものコードが含まれている BTEQ スクリプトは、各ジョブ毎に手動で再コーディングして維持する必要があります。

また、サービス品質保証（SLA）を実現し、管理と再利用を簡素化するためには Hadoop ETL の一部としてメタデータ機能は不可欠です。

情報は企業にとって最も貴重な資産の一つで、ビッグデータには更なる責任が伴います。

従って、Hadoop 内でもエンタープライズレベルのデータセキュリティを維持する機能が重要ですが、Hadoop でデータ統合処理を実現するツールは機能がまだ制限されています。

ETL ツールには GUI と接続を提供しているツールがありますが、企業の標準に統合する事が困難な独自のセキュリティモデルを提供している場合があります。既存のセキュリティを使用したセキュリティ要件に沿ったサポートが不可欠です。

# 最善手法

今後の Hadoop への取り組みに役立つ ETL の為に、Hadoop を使用してエンタープライズ・データウェアハウス (EDW) をオフロード (負荷軽減) する上でのサービス品質保証 (SLA) を確認します。それを行うには下記が必要です。

- Hadoop の具体的な種類と **相互作用** するソリューションの違い、そしてそれらが生成する **コード量** の理解。
- 削減と効率を提供する **オープンソースの利点** を補完したアプローチの特定。
- パフォーマンスのサービス品質保証 (SLA) を満たす **ネイティブ Hadoop 統合ツール** を検討、不要なオーバーヘッドを回避。
- 開発とデータシステムのトラッキングや再利用を可能にする為、 **メタデータ・リポジトリ** を提供する解決策を模索。
- **セキュリティが損なわれていないことを確認**。 実行可能なアプローチを制御し、すべてのデータを保護するために、既存のインフラストラクチャを活用しなければなりません。
- ETLHadoop 環境を展開、モニター及び管理する **拡張性の有るアプローチ** を提供するツールを検討。



# Syncsort はどの様に役立ちますか？

**SYNCSORT の Hadoop ETL ソリューションは、企業の HadoopETL への対応を提供します。**

- Hadoop 内でネイティブに実行：ランタイムエンジンは Hadoop フレームワークと統合され全てのノード上で実行。
- Apache、Cloudera、Hortonworks、MapR、PivotalHD と Amazon EMR を含む全ての主要な Hadoop ディストリビューションとの密な統合。
- 全ての Hadoop クラスタで展開する SyncsortHadoopETL ソリューションのワンクリック展開とアップグレード。Cloudera Manager とシームレスな統合。
- MapReduce ETL ジョブのモニタリングを容易にするために Hadoop Jobtracker と完全に統合。
- Hadoop クラスタのより優れたパフォーマンスと効率の為、既存の HiveQL 及び MapReduce ジョブをシームレスに最適化し既存の Hadoop クラスタへの差し込み。
- リソース及びタスクを基準とした自動セルフオプティマイザ機能により、より速く、持続可能なパフォーマンスと効率を提供するスマート ETL オプティマイザーを備える。
- データをセキュリティを維持するために、市場をリードする LDAP や Kerberos の様な一般的認証プロトコルをサポート。

# 結論

企業は、コストを削減し、より高速な分析とエンドユーザークエリのデータベース容量を解放する為に、エンタープライズ・データウェアハウス（EDW）から Hadoop ETL へ移行します。しかし、Hadoop は完全な ETL ソリューションではなく、ビッグデータのオペレーション・システムという主旨で使用されています。最適な結果を得る為には、エンタープライズ ETL ソリューションに結合されなければなりません。

Syncsort の Hadoop ETL ソリューションを選ぶにしても他のソリューションを選ぶにしても、これら 5 つのステップは、エンタープライズ・データウェアハウス（EDW）から Hadoop へのデータオフロード（負荷軽減）に役立ち、大きな課題を幾つか解決し、データ統合を加速させ、今後の Hadoop に対する新たな取り組みを確実に成功へと導く様最善手方を活用する様にします。

## ステップ 1:

ビジネス目標を理解し、定義します。

ビジネスの目標がコスト削減にしる、より多くのデータ活用にしる、もしくはサービス品質保証（SLA）に適應する事にしる、エンタープライズ・データウェアハウス（EDW）から Hadoop への移行が ETL ワークロードに適合するかどうかを確認する事が重要です。

ETL のパフォーマンスを最適化するために Hadoop ネイティブで、既存の開発者のスキルセットを活用してエンタープライズ ETL 機能を提供するアプローチは、あなたの目標を達成するのに役立ちます。

## ステップ 2:

Hadoop の為に正しい接続を行います。Hadoop が自社のサイロにならないことを確認します。

ビッグデータから値を抽出する為には、大規模なデータ接続が必要で、Hadoop がこれを実現する為には大きな手作業を必要とします。

コーディングすることなく、すべてのデータ・ソースおよびターゲットへの接続を提供するツールは、あなたのデータからより速く、あなたの Hadoop への投資からより多くの価値を得ます。

## ステップ 3:

ETL/ ELT のワークロードの上位 20%を特定します。

通常、データ変換の上位 20%が、処理機能、高額なコスト、そしてハードウェア、チューニング、保守といった IT にかかる工数の 80%を消費しています。

## ステップ4:

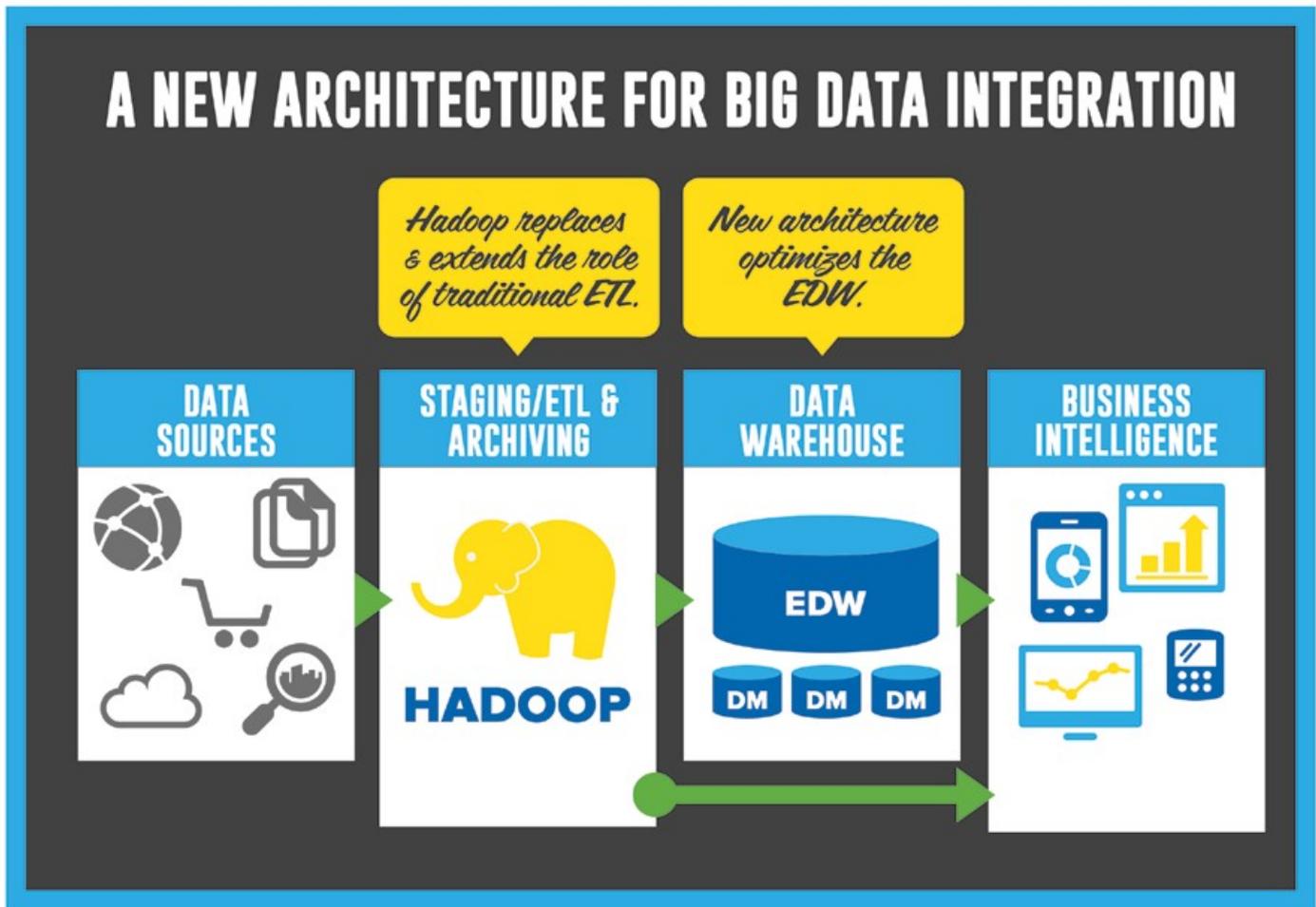
MapReduce 中の等価変換を再生成します。データ変換は Hadoop 中ですぐに複雑になるので、Java、HiveQL、Hadoop と MapReduce の深い理解とかなりの手動コーディングとチューニングが必要となります。変換、共通 ETL タスク、メタデータ機能全てを内蔵した GUI ベースのエンタープライズ ETL のアプローチは Hadoop ETL ジョブ生成と維持を容易にします。

## ステップ5:

自社の Hadoop ETL の環境設定を行なってください。

ETL 処理の為にサービス品質保証 (SLA) のが充分でない場合は、必然的に Hadoop 導入の障壁となります。

Hadoop の中で大規模な展開、監視および管理、およびデータのセキュリティを容易にする機能は、今後の企業での取り組みのための道を開きます。これらの機能は、ネイティブの Hadoop との統合を含む、組み込みパフォーマンスの最適化、および既存のセキュリティ・インフラストラクチャを活用する能力となります。



## about us

Syncsort provides data-intensive organizations across the big data continuum with a smarter way to collect and process the ever-expanding data avalanche. With thousands of deployments across all major platforms, including mainframe, Syncsort helps customers around the world to overcome the architectural limits of today's ETL and Hadoop environments, empowering their organizations to drive better business outcomes in less time, with less resources and lower TCO. For more information visit [www.syncsort.com](http://www.syncsort.com).

***syncsort***

like this? share it!

