

情報爆発時代がもたらす 新たなビジネスへのアプローチ

2011/07/06

株式会社 日立製作所
情報・通信システム社
ソフトウェア事業部 大量データ処理ビジネス推進室

山口 俊朗

「お客様システムへ、Hadoopなど新しい
大量データ処理技術を活用した
新ビジネスの創出に向けて、
お客様とともに考え、ご提案します」

Contents

1. 情報爆発時代の到来
2. 大量データのビジネス活用
3. 日立の取り組み
4. まとめ

1

情報爆発時代の到来

1-1. 情報爆発時代の到来

- 人やモノの行動や振る舞いを示すデータを活用した新サービスへの期待。
- 社会基盤の『機械』から生成されるデータが飛躍的に増大。
- 企業内では、各種業務ログやメール、Webのログなどが増大。

情報システムの増大するデータ

業務ログ

Web/メール

オフィス文書

ブログ

機械から生成されるデータ

モバイル端末利用

ICカード利用

センサデータ

交通や気象データ



大量データの活用が他社差別化や競争力の源泉！

課題1: データ量の爆発

- センシング機器が増え、秒・分単位で自動的にデータ取得が発生
- 従来型RDBでは、性能が出ない、運用管理コストが高くなるなどの課題が生じる



課題2: リアルタイムなデータ活用

- 証券のアルゴリズム取引や渋滞予測など、リアルタイム性が必要なデータ活用が増加
- 従来型RDBのように、一旦ディスクに蓄積してから検索するような手段では、処理が追いつかない



1-3. 技術トレンド

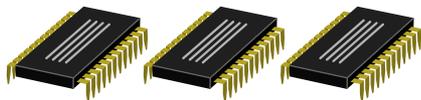
大容量メモリの活用した
高速処理
(数GB～数百GB)

大容量ディスクを並べた
ディスク/IOの分散
(数TB～数PB)

安価なサーバを並べた
並列分散処理
(数万円～20万円/台)

サーバ
メモリ

インメモリデータ処理



サーバ

ディスク/IOの並列化

ディスク



ディスク



ディスク



安価な
サーバ

安価な
サーバ

安価な
サーバ

処理の並列化

メモリ

ディスク



メモリ

ディスク



メモリ

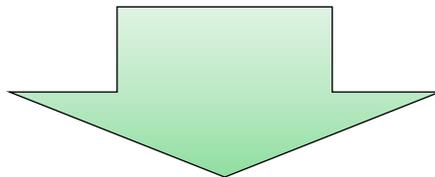
ディスク



技術の進歩により、大量データの高速処理が可能となる

対応1: 急激な負荷変動やデータ量増加に対応

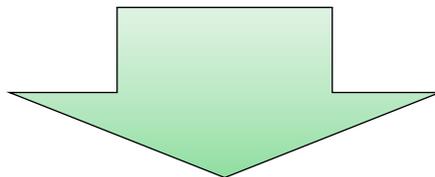
【従来】データ管理サーバの台数をピーク時に合わせて設計



【今後】トラフィック量に応じて、自由にサーバを増減できる

対応2: リアルタイム処理による予兆検知

【従来】1日前以上の過去データを使って、次の行動を決定



【今後】過去データに加えて、いま流れている現在データを使って、今後起き得るイベントの予兆を検知できる

1-5. 何ができるようになるのか？

① データの価値創造

- 当初は意識していなかったデータの価値が、後になって生まれてくることがある。
- 「とりあえず残しておく」ということも将来に向けた有益な戦略となり得る。

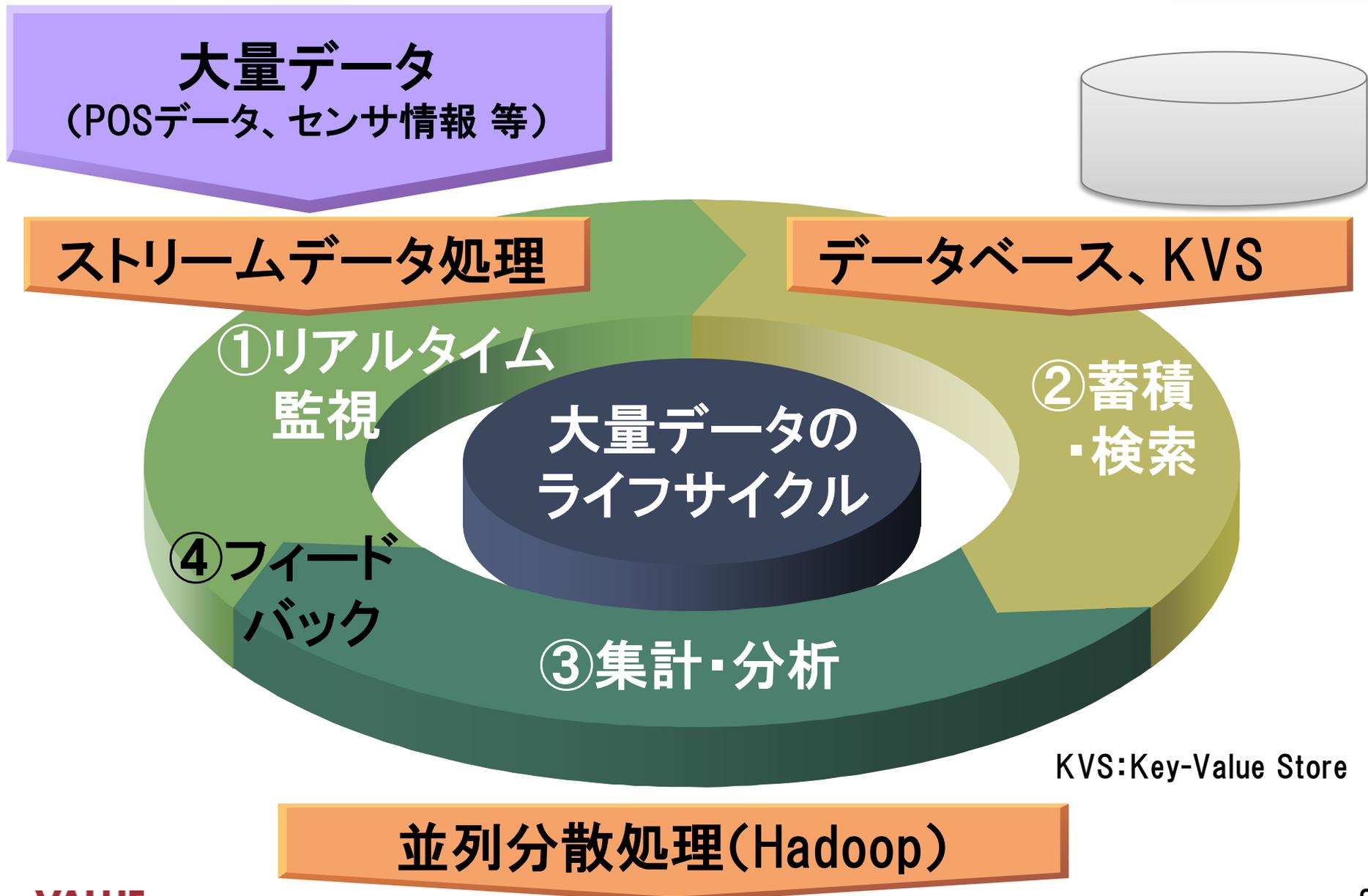
② 全件蓄積が将来の布石に

- (今は)特定の目的に沿った分析を効率よく行えるように、正規化してからデータを保存しているが、それによって失われてしまう情報は少なくない。あえて“生”のままの状態データをアーカイブしておくことにより、その蓄積が将来における自社の強みや付加価値になり得る。

③ 新たなビジネスの創生

- まさに情報を握る者こそが、次の時代の勝者！
- 業種の垣根を越えて相互に共有し、循環させることが可能となったとき、情報は知識に変わり、そこから新たなビジネスやサービスが創出されていく。

1-6. 大量データ処理の全体像

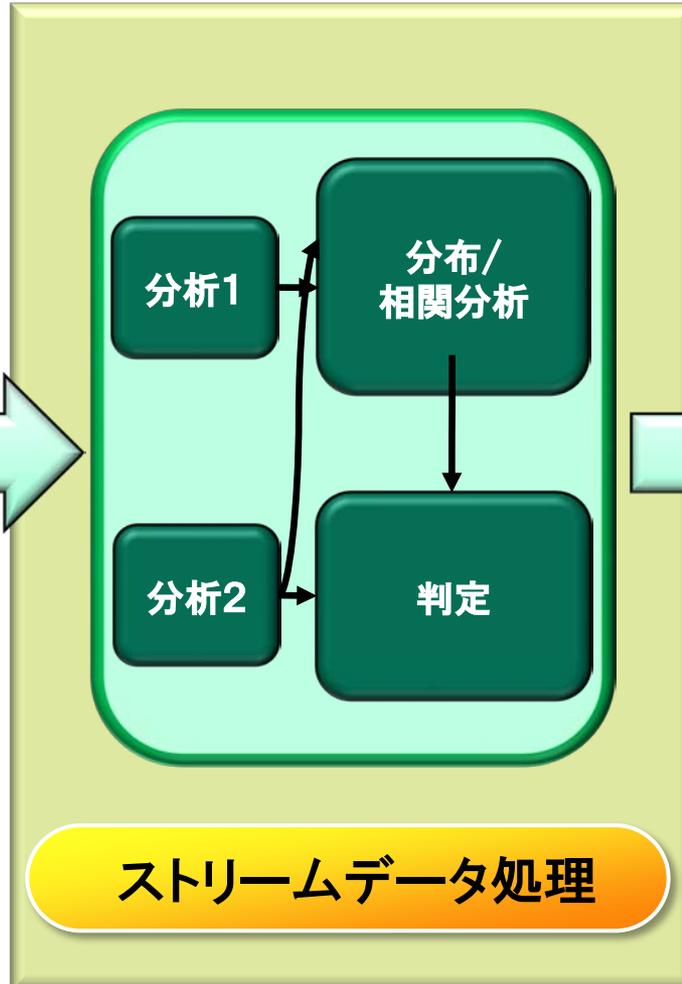


1-7. ストリームデータ処理

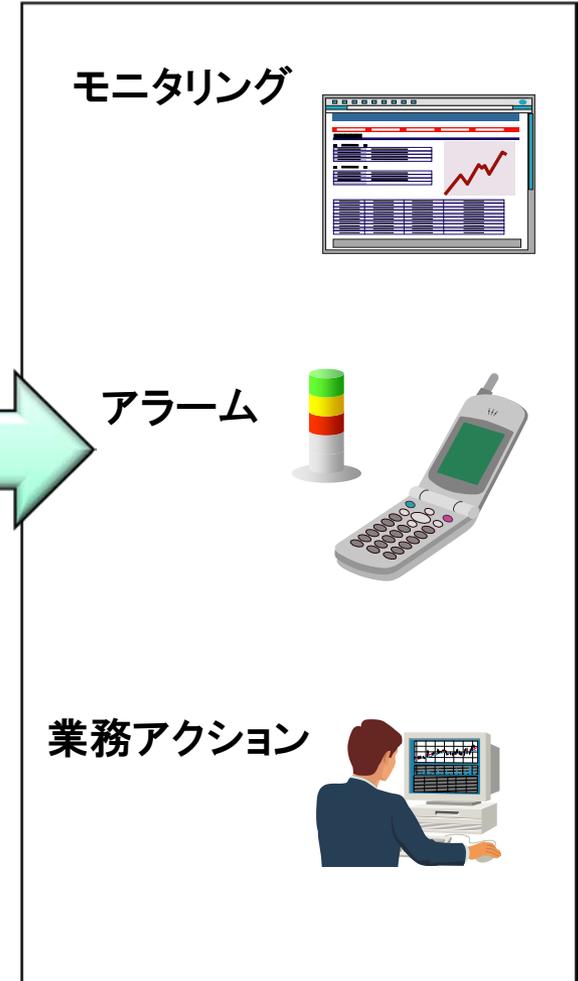
① 大量な実世界データ



② データ発生と同時に分析



③ 分析結果の活用



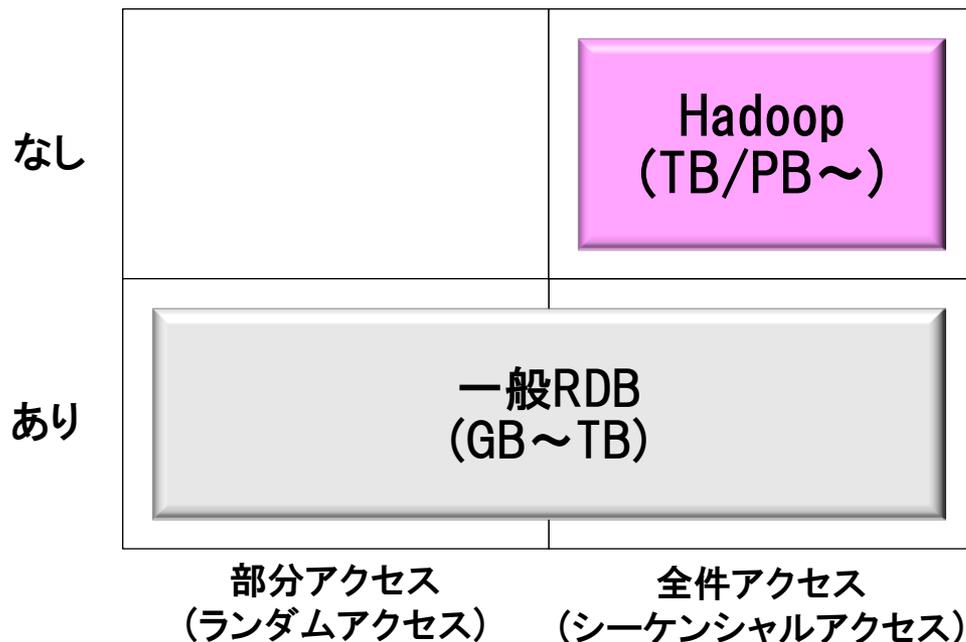
● Hadoopは以下の特性を持つシステムに適する

- 大量データを扱うシステム(TBオーダなど)
- 全件アクセスが必要な処理が多いシステム
- 将来、データ量の増大が予想されるシステム

バッチ処理が得意

● トランザクション制御機能はない。運用で区切られたデータを全件アクセスをして傾向をつかむ処理への適用が多い

トランザクション制御



データアクセスパターン

業種	代表的な適用分野
産業流通	販売ログデータを用いたマーケット分析
通信	通信ログデータを用いた分析、リコメンド生成
金融	顧客・取引データを用いたリスク分析
電力	センサデータを用いた予防保守
Web事業者	アクセスログを用いたリコメンド生成

2

大量データのビジネス活用

株価指数算出システム

- 電文受信(株値の変化)ごとに一部の指数のみをリアルタイムに算出。
- 8,000件/秒の株価情報もミリ秒レベルで処理、相場変化の反映は1秒 → 10ミリ秒以下。



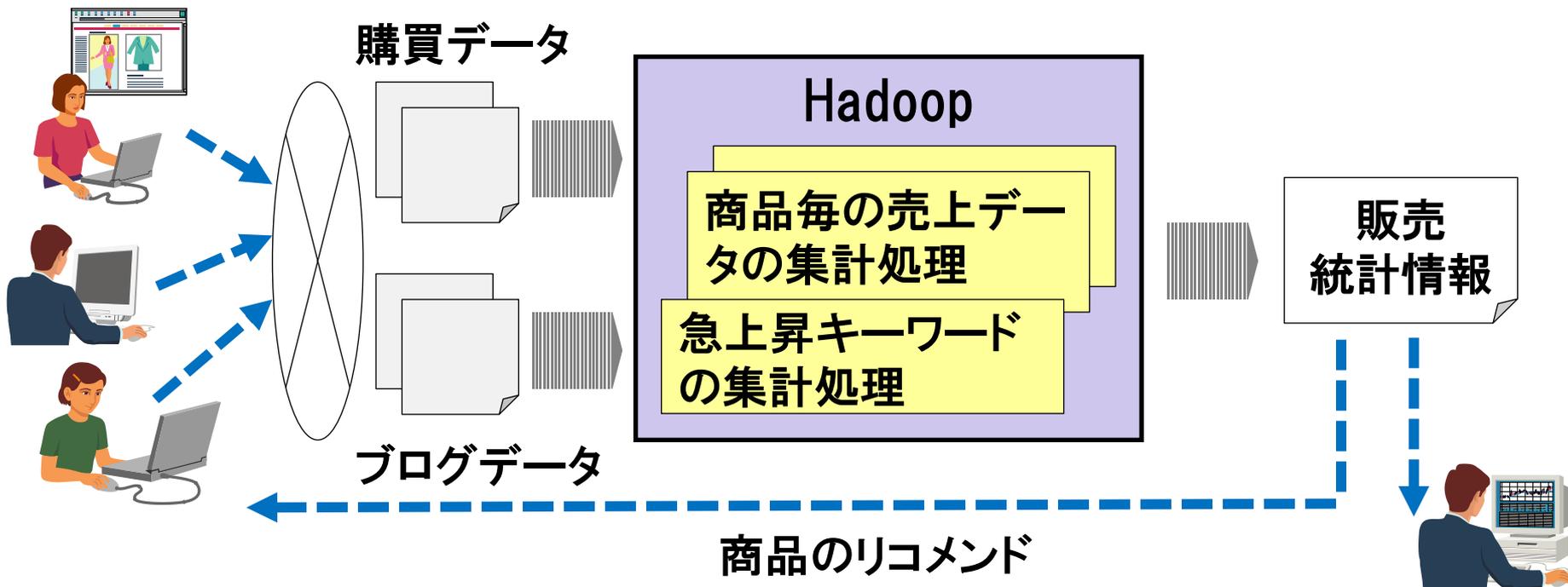
交通状況モニタリング

- 自動車の車両位置情報を分析して、車の速度や走行方向を算出。
- 速度や走行方向をもとに、渋滞を検出。



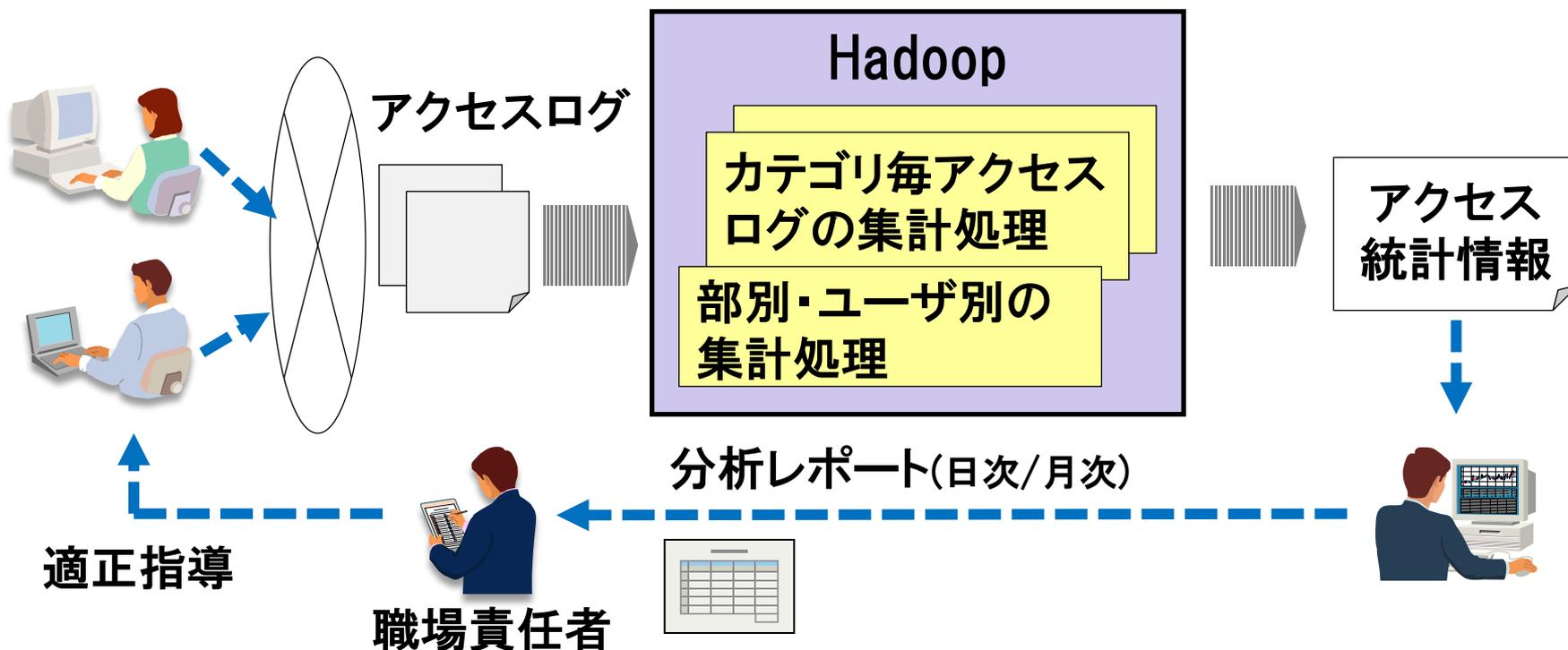
アクセスログ解析による商品リコメンド

- ECサイトの購買データやブログのテキストデータを解析して、人気のある商品をレコメンドし、購買行動に結び付ける。
- 集計処理時間を大幅に短縮できる。また開発容易性により、レコメンド内容を試行錯誤しやすい。



アクセスログ解析によるコンプライアンスチェック

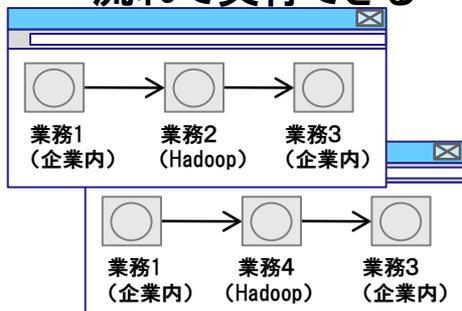
- 従業員がインターネットサイトを閲覧したWebアクセスログを解析して、コンプライアンスチェックを行なう。
- 集計処理時間を大幅に短縮(数時間⇒数分／数十秒など)。



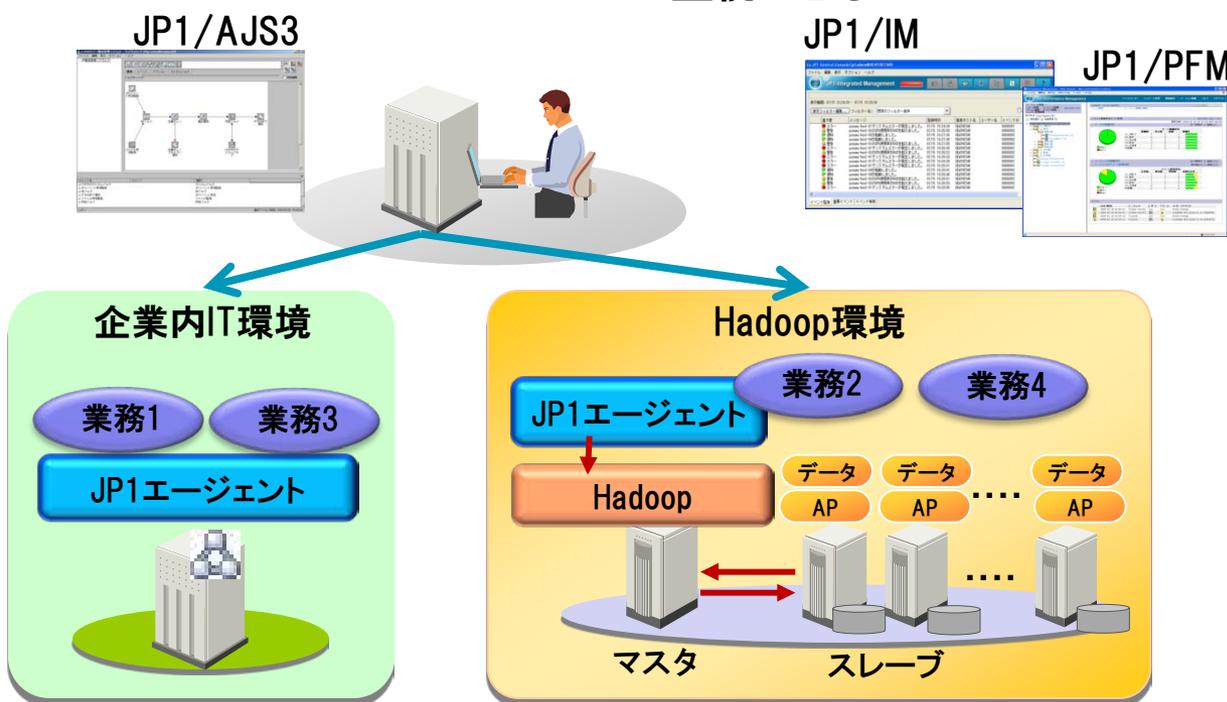
Hadoopの運用・監視もJP1で一元化

- Hadoop環境にも対応。シームレスなジョブ連携や一元的な監視で、運用を効率化します。

通常の業務とHadoopの業務を一連の流れで実行できる



通常環境もHadoopも1つの画面で監視できる



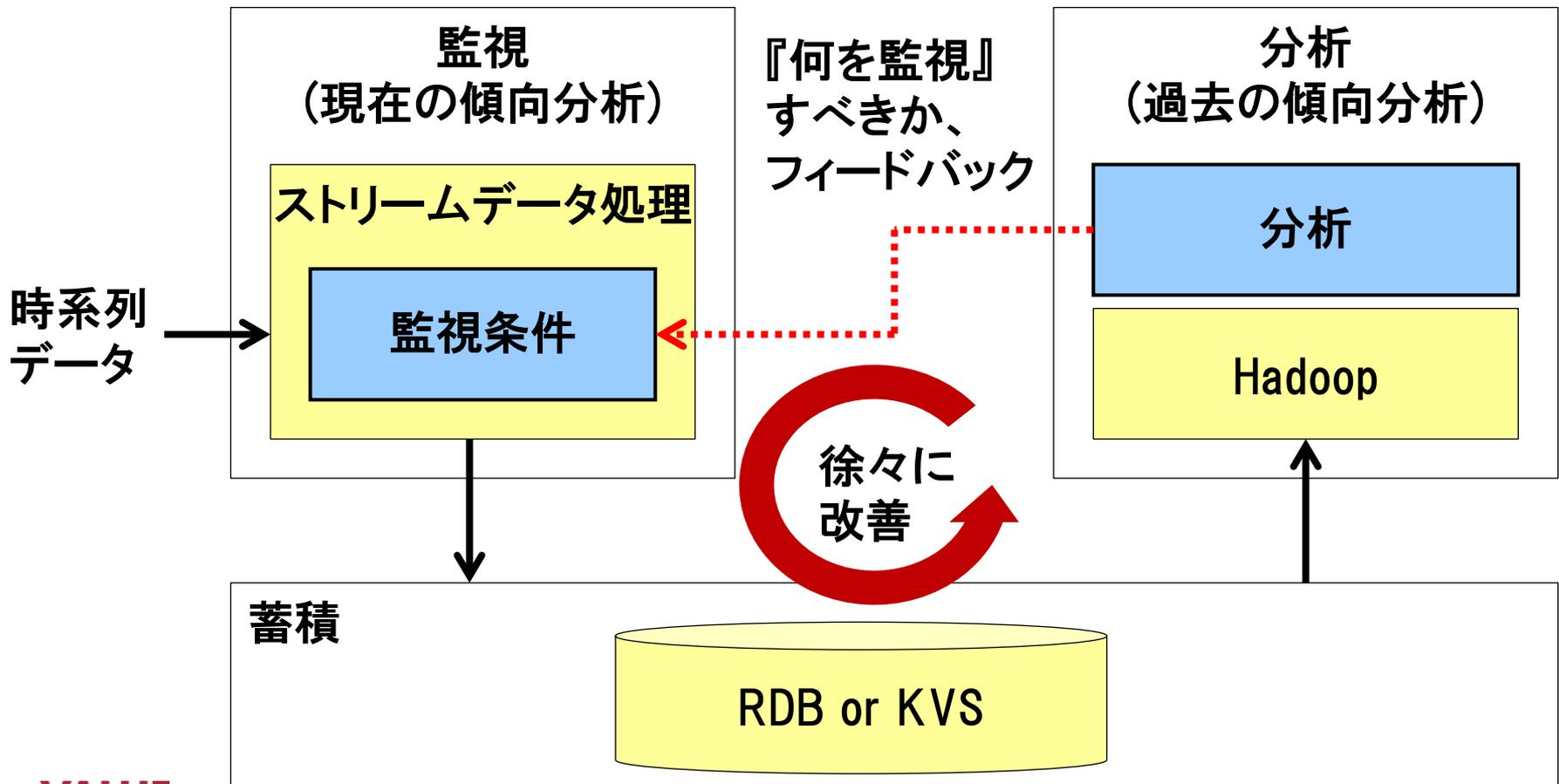
JP1/AJS3 : JP1/Automatic Job Management System 3

JP1/IM : JP1/Integrated Management

JP1/PFM : JP1/Performance Management

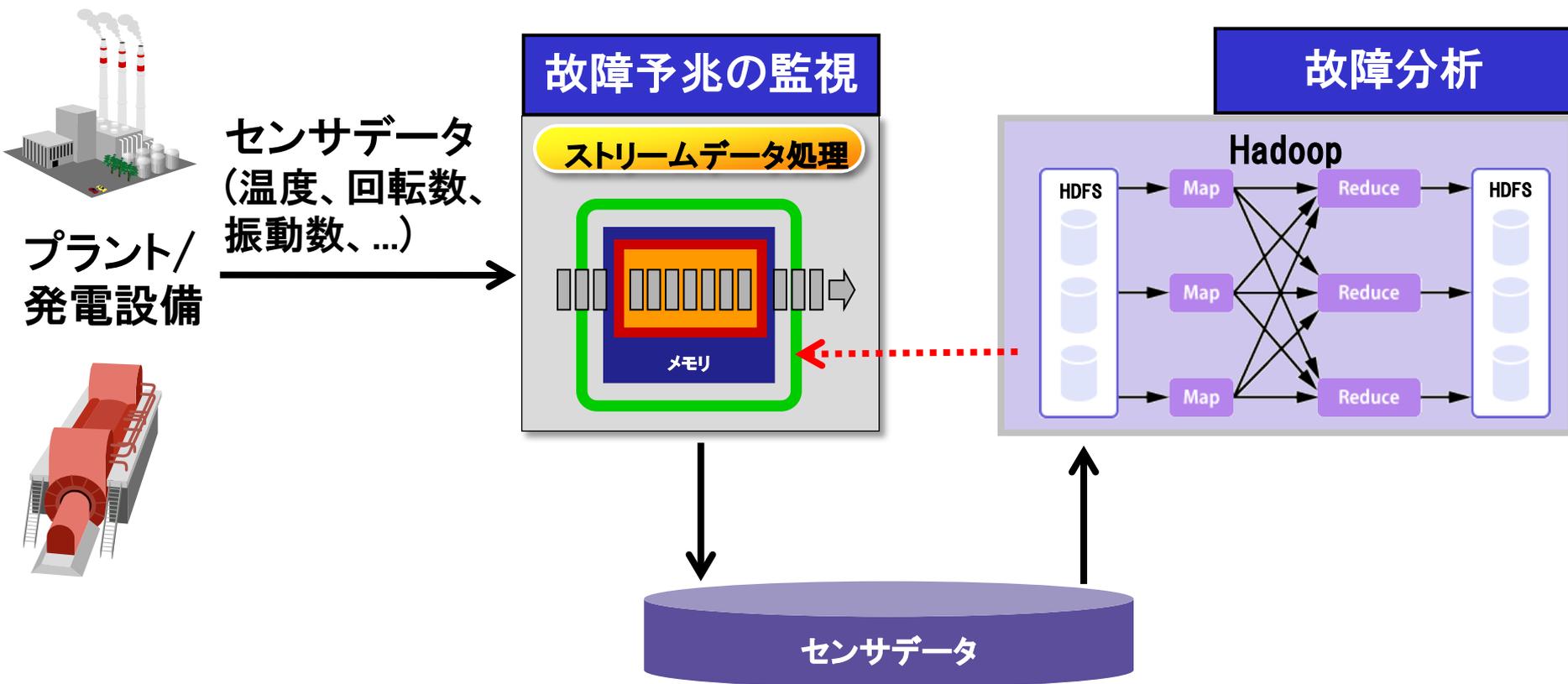
「今」を監視、「過去」を分析、監視へのフィードバック

- ストリームデータ処理で、『何を監視』するのが重要。
- Hadoopで過去の傾向を分析し、その結果をフィードバック。



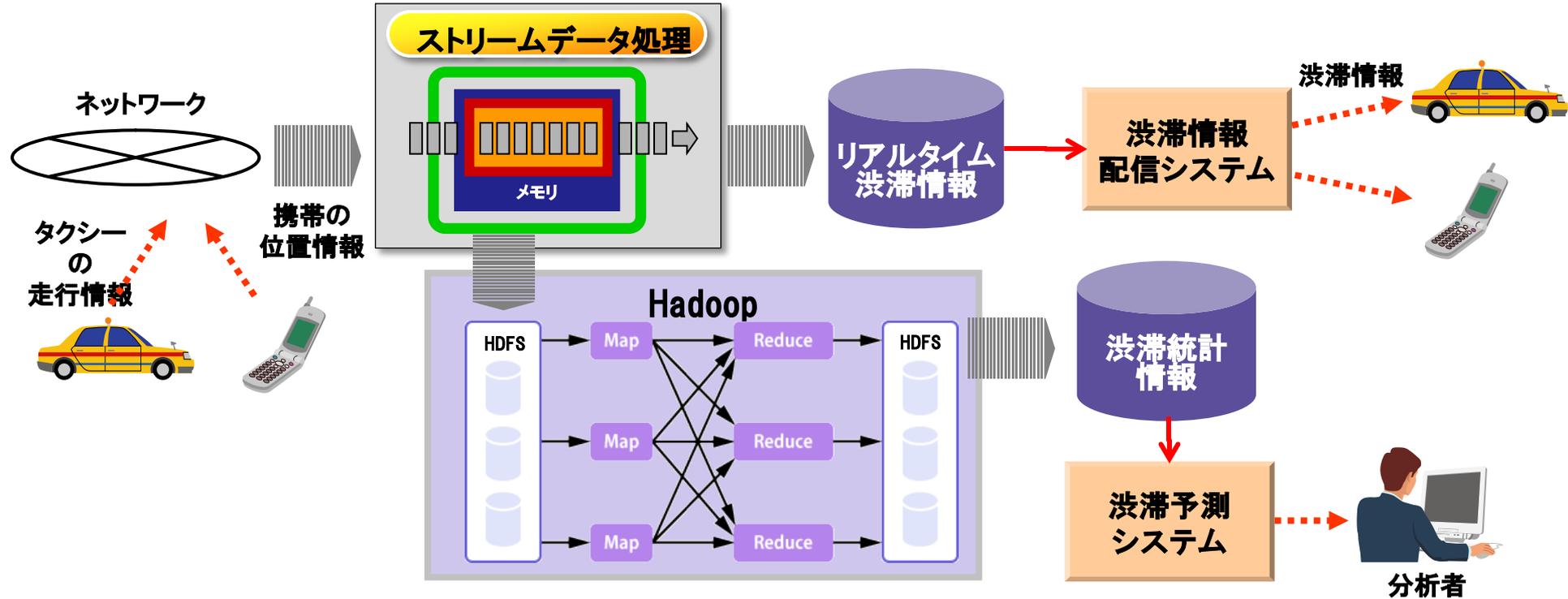
プラントの予防保全

- センサから発生する稼働情報を監視し、異常予兆を検知。
- プラント故障による仕損費および稼働率保証リスクを低減。



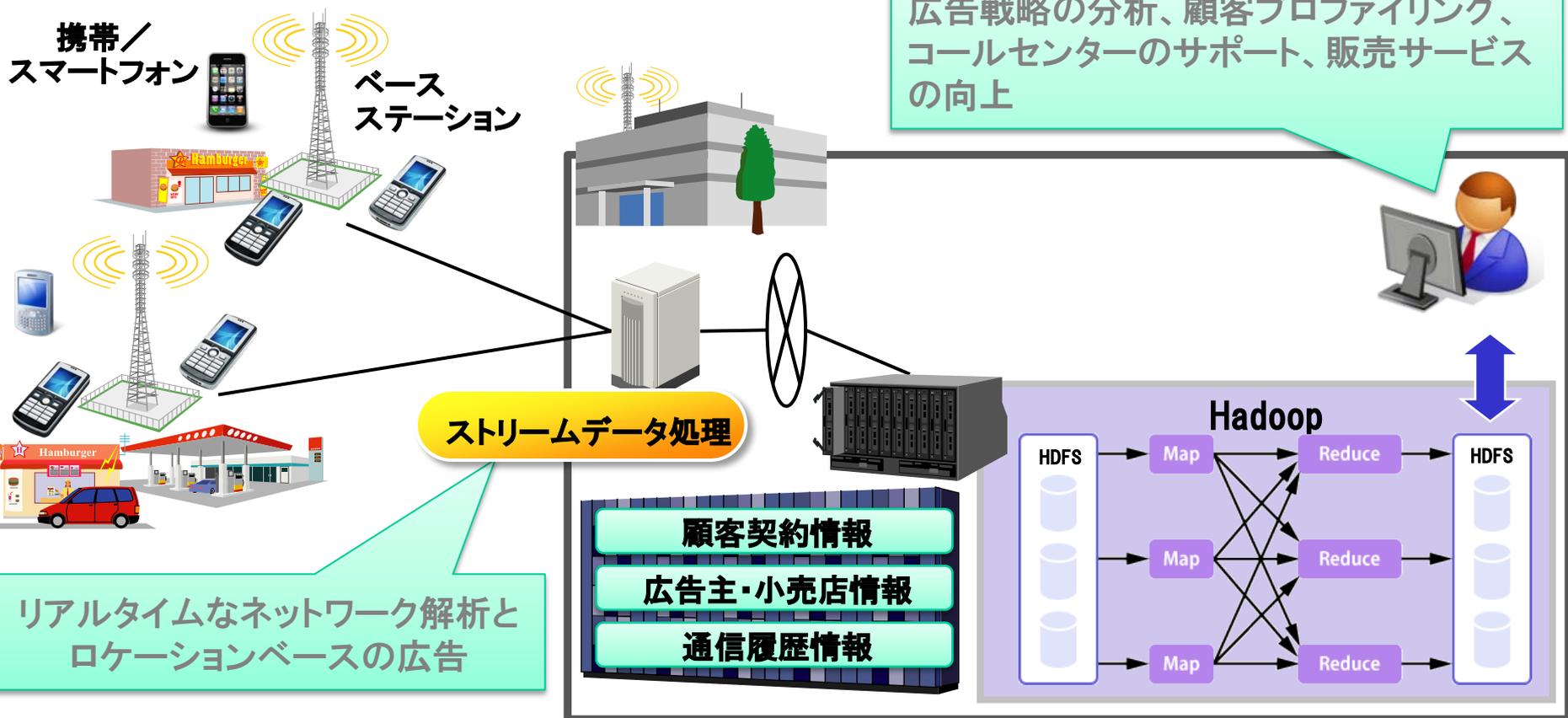
GPSデータを用いた渋滞予測

- リアルタイムに渋滞情報の即時配信。
- 蓄積したGPSデータから、統計的な渋滞発生 の情報を取得して、渋滞予測システムを実現。



位置情報を活用した広告配信

- 格納した各種データを検索して消費者行動のモデルを分析。
- 分析結果をもとにストリームデータ処理でリアルタイムに広告を送信。



新技術の適用に対するお客様の期待や不安

- ① データがたくさん溜まっているが、データをうまく活用できていない。
- ② 新しい技術がたくさんあるが、どの技術を採用すれば良いかわからない。
- ③ 試してみたいが、実証実験をおこなうための場所や環境がない。

お客様のニーズ

大量に蓄積されたログを
分析したいのだけど...

データ量が増大してバッチ
時間を短くしたいのだけど...

もっとリアルタイムに異常を
検知したいのだけど...



新技術・製品

Hadoop?

KVS?

ストリーム?

クラウド?

新技術の適用にはシステム構築に入る前の試行評価が重要。
⇒お客様のシステム企画・検証フェーズにおいて、大量データ処理に
関するコンサルティングや実証実験の場の提供が必要。

2-11. 課題解決に向けたアプローチの流れ

●お客さまの課題



データの
活用方法が
わからない

分析モデル
は判っている
が、適切な
処理方式が
わからない

処理方式は
決まっており、
実機評価をし
たいが、環境
がない

大量データの
処理技術につ
いて学びたい

●お客さまの工程

システム
企画

実証
実験



必要なアセスメント

活用方法の提言、
ビジネスへの貢献を
コンサルティング

処理方式の提案
「検証環境の場」の提供

「検証環境の場」
の提供

技術研修の
実施

適切なシステム形態へ

(注)分析モデル:分析対象データ、分析方法、分析結果活用のモデルを示す。

3

日立の取り組み

ストリームデータ処理

uCosminexus Stream Data Platform

SQL : Structured Query Language
CQL : Continuous Query Language

- ✓インメモリ処理と差分計算処理による超高速処理
- ✓SQLライクなスクリプト言語(CQL)で分析シナリオを記述可能な開発容易性

並列分散処理

uCosminexus Grid Processing Server (グリッドバッチ)

- ✓データを分割配置し、バッチ処理の並列実行により、バッチ時間の短縮とバッチ終了時間の遅延防止
- ✓基幹系にも対応出来る高い可用性、障害の局所化
- ✓COBOL資産等を有効活用可能(プログラミング言語非依存)

データベース

東大と日立が共同で 超高速データベースエンジンを研究開発
従来型のデータベースエンジン比で約800倍の処理性能をめざす
(<http://www.hitachi.co.jp/New/cnews/month/2011/06/0602a.html>)

3-2. 大量データ処理をトータルでご支援

システム企画・実証実験～システム構築・本番稼働まで

大量データ
分散処理
アセスメント
サービス

分析計画 策定
コンサルティングサービス

分析方式 検証支援サービス

NEW

技術研修サービス

基盤導入 検証支援サービス

システム
インテグレーション

大量データ処理
基盤構築サービス

NEW

エンタープライズグリッド
ソリューション(金融向け)

パブリッククラウド

大量データ分散処理
プラットフォーム
提供サービス

NEW

Hadoop、uGPS、uSDP

PaaS

従来型
(オンプレミス型)
システム向け
製品・サービス

HA8000-bd/BD10用
ユーザ環境設定サービス

NEW

Hadoop MapReduce
サポートサービス

日立オープンミドルウェア製品
(JP1等)

日立サーバ製品、
日立ストレージ製品、サービス他

(注)uGPS:uCosminexus Grid Processing Server、uSDP:uCosminexus Stream Data Platform、PaaS: Platform as a Service

お客様のシステム企画や検証に必要な環境、技術支援を提供

大量データ分散処理アセスメントサービス	概要
分析計画 策定コンサルティングサービス	<ul style="list-style-type: none">✓お客様の新事業・業務革新に寄与する、大量データ処理システムの計画立案を支援します。✓日立の豊富なノウハウを用い、分析モデルを提案します。
技術研修サービス	<ul style="list-style-type: none">✓Hadoopや日立オープンミドルウェアによる大量データの分散処理技術を座学・マシン演習で研修します。
分析方式 検証支援サービス	<ul style="list-style-type: none">✓日立の分析処理に関するノウハウを使って、適切な基盤や処理方式の決定に導きます。✓環境構築済みのクラウド環境により、すぐに検証に着手できます。
基盤導入 検証支援サービス	<ul style="list-style-type: none">✓環境構築済みのクラウド環境により、すぐに検証に着手できます。✓お客様が開発・移行・性能評価などを行われる際の技術支援により、効率よく検証を推進できます。

3-4. 大量データ処理基盤構築サービス

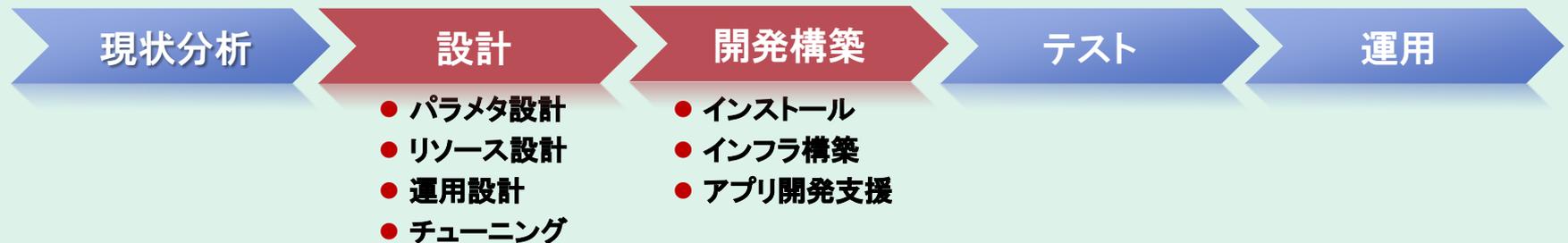


Hadoopや日立オープンミドルを使用してシステム構築をしたいが、システム環境や運用の設計ノウハウがなく不安。構築方法がわからないため調査コストがかかってしまう。

ポイント

- 専門技術者がお客様に代わって、環境設計/構築します。
- パラメタ設計書/環境構築手順書をご提供します。

■Hadoop、日立オープンミドルウェア(※1)を導入する際の開発/実行環境パラメタ設計やチューニング、環境構築作業、アプリケーション開発支援を実施します。



■お客さまから構築したいシステムの構成をヒアリングし、各種設定パラメタの設計、インストール及び動作環境構築を行います。

■お客さまの業務/ユースケースをヒアリングし、アプリケーション開発や運用設計、チューニングのご支援を致します。

(※1)対象製品については、個別にお問い合わせください

3-5. 大量データ分散処理プラットフォーム提供サービス



- ・ 大量データ分析をしたいが、機器の準備が必要。
- ・ 大量データ分析をするための機器設置スペースがない。

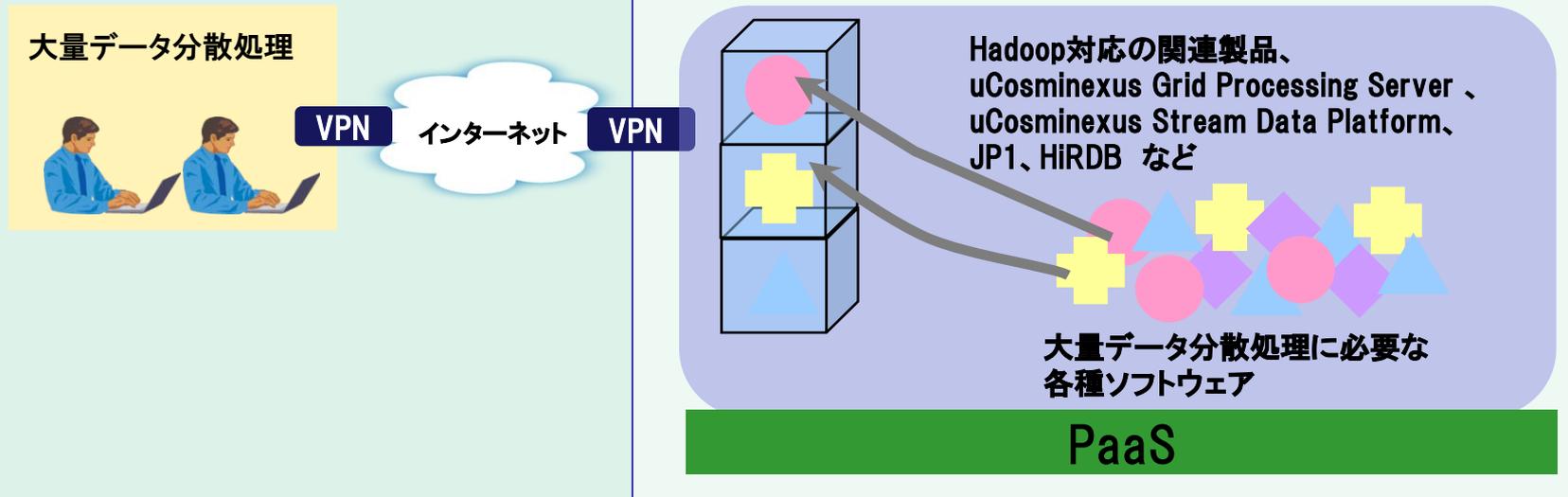
ポイント

- 大量データ分散処理のためのクラウド環境を提供しますので、機器の準備が容易になります。
- セキュリティの高い、環境配慮型データセンタからご提供します。

大量データ分散処理の実行に必要なミドルウェアを月額サービスで利用可能

<Harmonious Cloud センタ>

大量データ分散処理プラットフォーム提供サービス





Hadoopを適用しようとして検討中だが、並列実行になるためサーバが大量に必要になり、導入費用や、短期間で立ち上げられるかが心配だ。

ポイント

- 多数のノード(*1)に対する導入時の環境設定作業を、日立の専門技術者が当社出荷前に実施するサービスです。
- ブレード毎の個別設定も可能です。

HA8000-bd/ BD10 ユーザ環境設定 サービス



マスターコピーサービス

お客さまよりインストール媒体と手順書をお預かりして、新規に出荷するサーバブレードにコピーいたします。

個別設定サービス

各サーバブレードに対するOS上のネットワーク設定や、ベースユニットへの管理ネームラベルの貼り付け作業を代行いたします。

LAN配線代行サービス

ベースユニットに内蔵のLANスイッチと、ラック内統合ハブ間のLAN配線作業を代行いたします。

サービスの 効果

- インストール・設定作業を日立が代行することにより、お客さまのご負担を軽減します。
- 製品ご導入後、業務運用開始までの期間を短縮できます。

(*1) ノード: Hadoopプログラム(JobTracker, TaskTracker, NameNode, DataNode)が実行されるOS環境の単位



オープンソースなので、障害発生時などに問い合わせ先がなく不安だ。

ポイント

- Hadoopに関する一般的なQ&Aから障害発生時の調査など、コミュニティ情報調査等によりお客さまの問題解決を支援します。

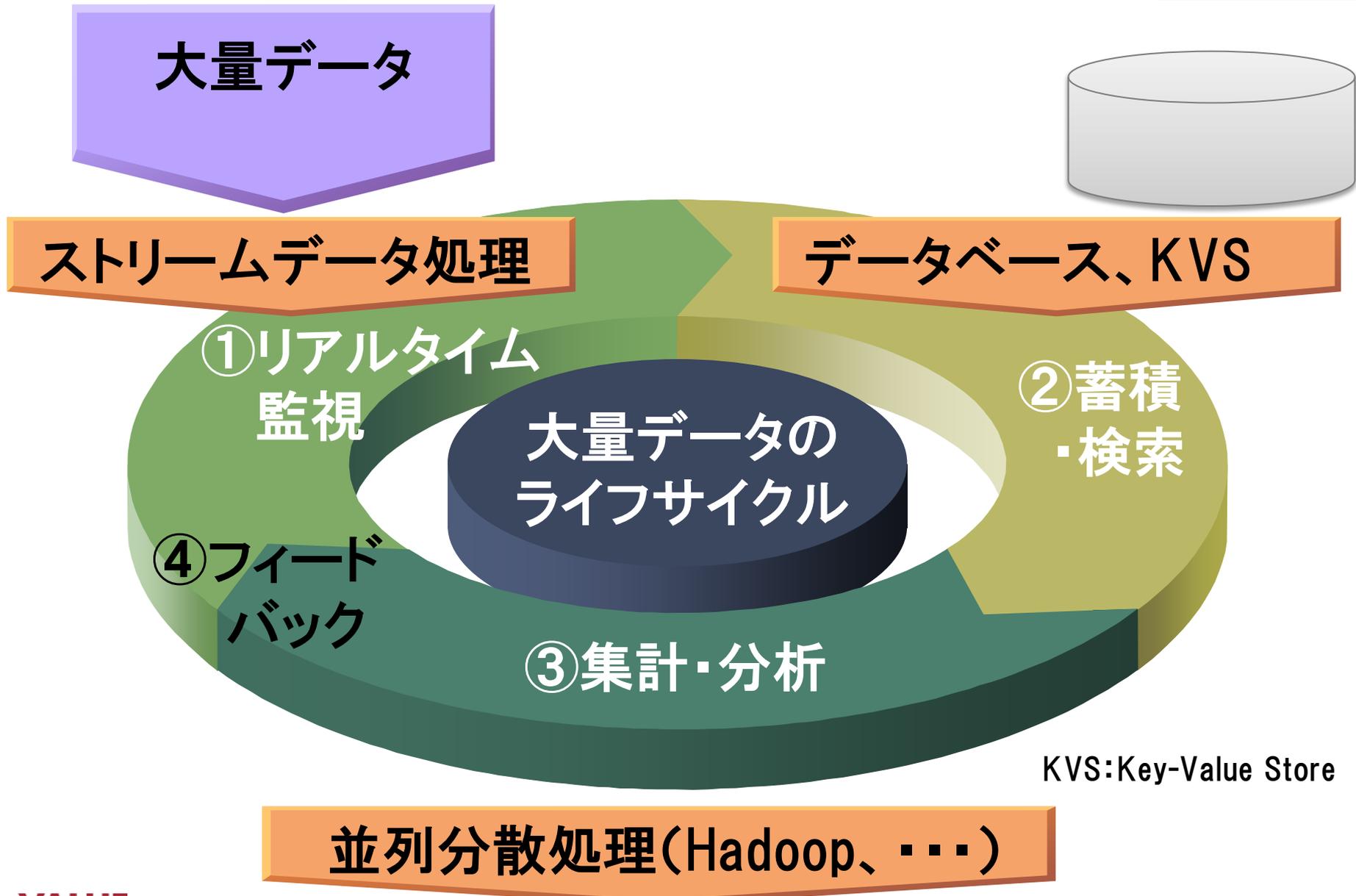
サービス内容

項目	内容	
サービス対象ソフトウェア	Hadoop Common Hadoop MapReduce Hadoop Distributed File System	
サービス内容	問題解決の支援	対象ソフトウェアの機能や使い方についての一般的なQ&Aや、問題発生時の回避策の調査や確認を実施します。
	情報の提供	FAQやOSSコミュニティからのリリース情報などを、お客さまへご提供いたします。
サービス前提環境	Red Hat Enterprise Linux 5以降 Java 1.6.x	

4

まとめ

4-1. 大量データのライフサイクル



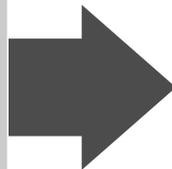
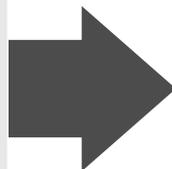
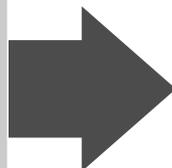


課題

保有の大量データを
新事業・業務改革に活用で
きないだろうか。

初めての技術なので、
不安だ。
検証できる環境がない。

大量データ活用を、
どのようなシステムで
どのように実現すればよい
かわからない。



アセスメントのメリット



大量データの活用方法がわかる

ポイント
その1

日立の豊富なノウハウを用い、
大量データの分析・活用戦略を検討し、
新事業・業務改革に寄与。

すぐに着手・短期間で検証ができる

ポイント
その2

初期設定済みの検証環境を用意して、
専門技術者の支援を受けながら
効率よく検証を実施。

適切なシステム形態の立案ができる

ポイント
その3

新技術を検証・評価をすることにより、
適切なシステム形態を立案。

協創のパートナーとして
お客さまの次のビジネスイノベーションに貢献

uVALUE



ITの技術力、システム構築力

協創に挑むパートナーシップ、協創する力

自らのIT経営改革を通じて培った経験、ノウハウ

他社商品名、商標等の引用に関する表示

- Javaは、Oracle Corporation 及びその子会社、関連会社の米国 及びその他の国における登録商標 または商標です。
- Red Hatは、米国およびその他の国でRed Hat, Inc. の登録商標もしくは商標です。
- Linuxは、Linus Torvalds氏の日本およびその他の国における登録商標または商標です。
- Googleは、Google Incの登録商標です。
- インテル, Intel, およびIntel Core は、アメリカ合衆国およびその他の国におけるIntel Corporationの商標です。

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。

■ 大量データ分散処理Webサイト

http://www.hitachi.co.jp/soft/big_data/

■ お問い合わせ

Eメール: big_data@itg.hitachi.co.jp

■ uCosminexus Grid Processing Server

<http://www.hitachi.co.jp/soft/cosminexus/gps/index.html>

■ uCosminexus Stream Data Platform

<http://www.hitachi.co.jp/soft/cosminexus/sdp/index.html>

uVALUE

HITACHI
Inspire the Next 