TOSHIBA
Leading Innovation

ビッグデータによる価値創造を実現する データ収集・蓄積・分析クラウドサービス

“簡単！賢く！データを活かす！” 東芝データレイクサービスの取り組みのご紹介

株式会社 東芝 インダストリアルICTソリューション社

商品統括部　プラットフォームソリューション商品技術部
栗田 雅芳

© 2015 Toshiba Corporation

ビッグデータという言葉が使われ始めた当初はバズワードと揶揄されることもありましたが、今ではさまざまな活用事例が登場しています。また、データウェアハウスのような従来の方法では多くの課題がみつかり、最近ではデータレイクという概念が注目されています。

本セッションでは、東芝のビッグデータの活用事例と、ビッグデータの活用の阻害要因を解決し、データの収集、蓄積から分析までトータルに対応する東芝のデータレイクサービスの取り組みをご紹介します。

アジェンダ

- 最近のビッグデータを取り巻く状況
- 東芝のビッグデータ活用事例
- 東芝データレイクの取り組み

© 2015 Toshiba Corporation

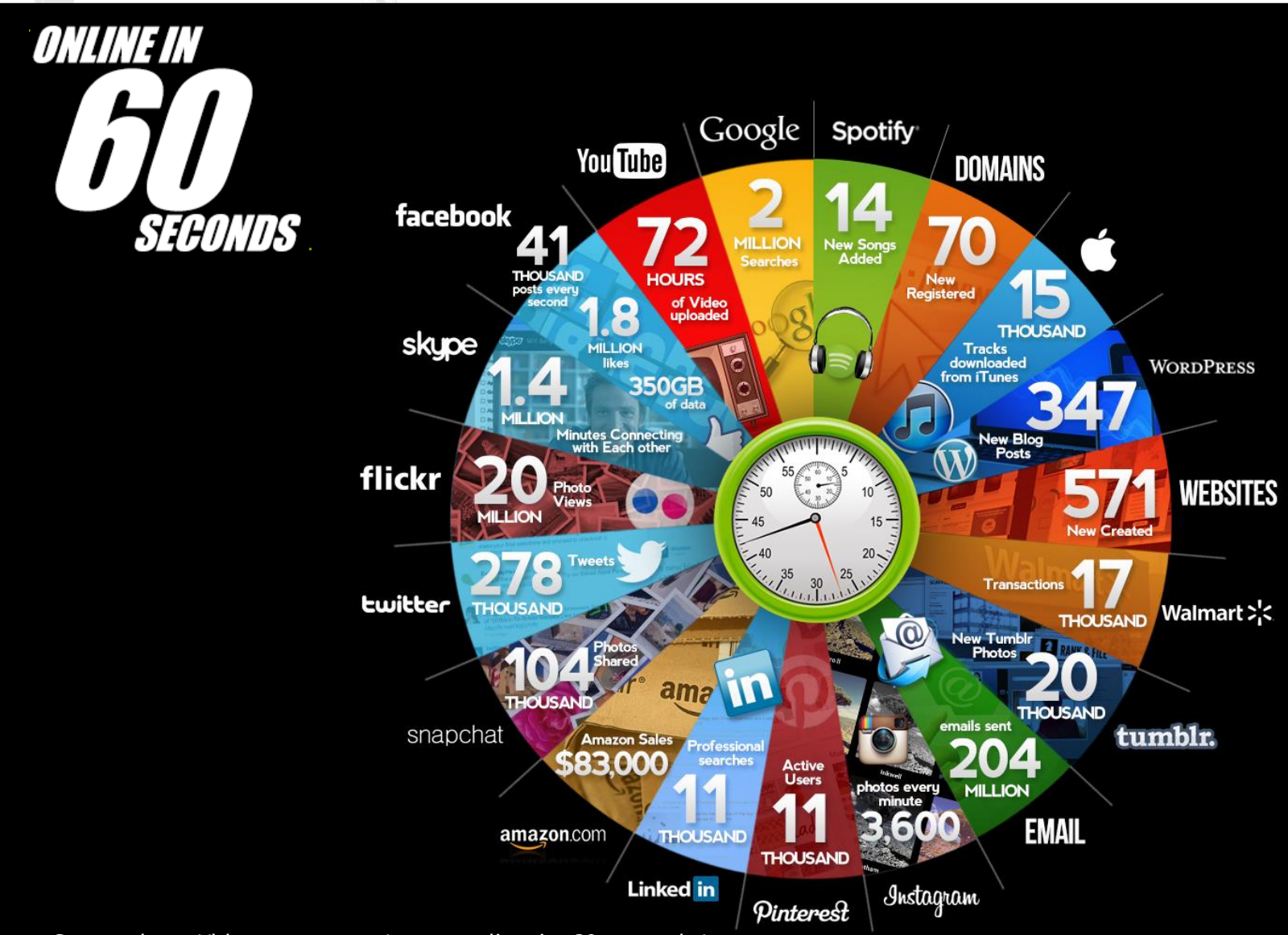
最近のビッグデータを取り巻く状況

Source: NBC News

ビッグデータ　×　IoT & IoE

Source: Deloitte

ビッグデータ　×　リアルタイム



Source: http://blog.qmee.com/qmee-online-in-60-seconds/

ビッグデータ × IoT & IoE × リアルタイム

リアル社会での人間活動と自然・社会環境を
データとして把握・融合・分析してリアルタイムに活用

チャンスやリスクいち早くつかみ、新しい価値創造

東芝のビッグデータ活用事例


TOSHIBA
TOSHIBA
Oui
ABC
Cooking Studio
JTB
JTBトラベランド
4Fにて営業中
knt!
営20:00まで営業
www.knt-t.co.jp

スマートビル　~省エネと快適な環境の共存~

© 2015 Toshiba Corporation

川崎スマートコミュニティセンターでの実証

省エネと執務者の快適性を両立したオフィス環境を実現
ビル全体でのCO2削減量 54% ※1

快適性はPMV(※2) を活用

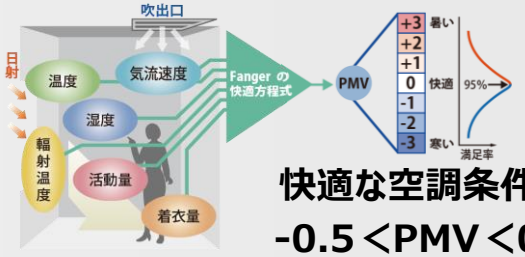
※1: 東京都一般事務所ビル平均値(2005年)基準（東芝試算） ※2: 快適性指数 PMV(Predicted Mean Vote)：ISO7730にて規定

国内初

モデルベース最適空調制御

空調システムをエネルギーモデル化。
執務者の快適性を維持しながら、
最も省エネとなるように温度/湿度
をリアルタイム制御。

従来制御に比べ
省エネ率7%向上



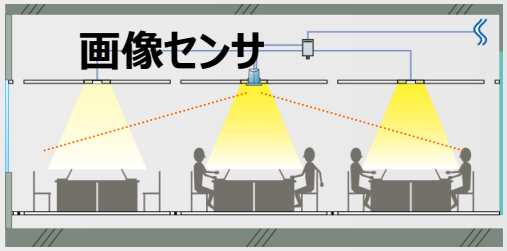
快適な空調条件：
-0.5＜PMV＜0.5

国内初

画像センサ応用照明制御

画像センサにより執務者の在/不在
をリアルタイム検知、照明１灯ごとに
調光制御。

画像センサにより
省エネ率11%向上

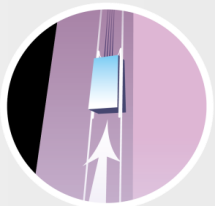


国内初

エレベータ混雑階優先制御

混雑しているエレベータホールにかごを
リアルタイム優先割付け、待ち時間を
短縮。

画像センサにより
待ち時間20%短縮



ELV優先割り付け

©

快適・効率的な住空間（ホーム・ソリューション）

省エネ×スマート家電×ライフログ活用で快適な住空間を実現

省エネ

見える化・わかる化・できる化

平均 約10%節電

デマンドレスポンス制御

ピーク時省エネ率 20%削減

×

スマート家電

家電の統合コントロール

家電コンシェルジュ

・省エネ

・見守り

・運転状態、故障予知

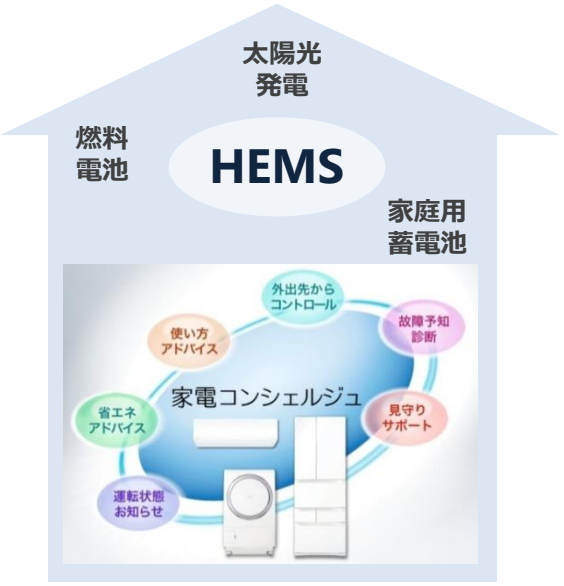
×

ライフログ活用サービス

生活パターンに沿った家電自動運転

生活スタイルに合ったライフサポート

（買い物支援、家事代行等）



パターン（ルール）の導出

タイムリーなサービスの提供

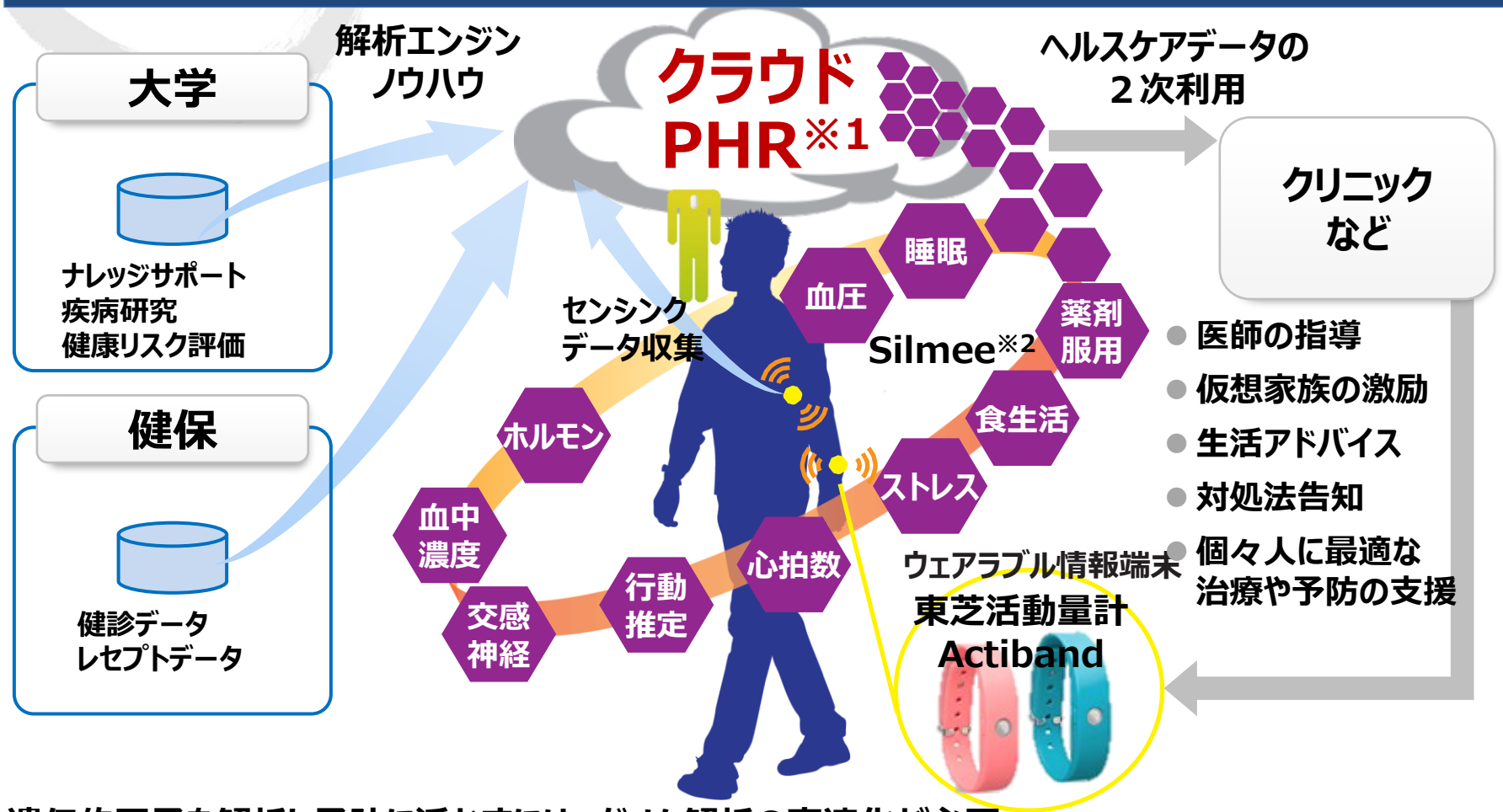
自炊好きの家庭に
タイムリーに食材の紹介

外食好きの家庭に
タイムリーにレストランの紹介

© 2015 Toshiba Corporation

ヘルスケア

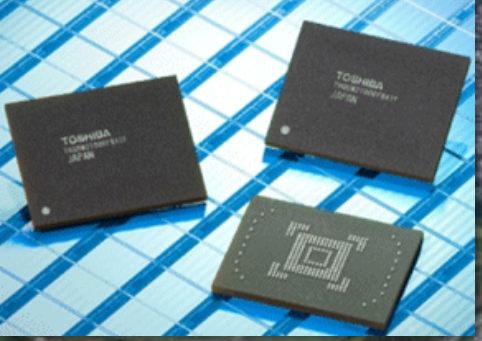
膨大な医療情報＆センシング情報の解析で健康な生活を実現



■遺伝的因子を解析し予防に活かすには、ゲノム解析の高速化が必要

- 1億2千万人のゲノム解析（フルシーケンス）には約2万年以上の処理時間が必要
- 1億2千万人のデータ蓄積では120エクサバイトの領域が必要

※1: PHR（Personal Health Record ）
個人の健康記録
※2: Silmee　生体情報をセンシングし、端末に
無線でデータを送り出すセンサ・モジュール


TOSHIBA
TOSHIBA
JAPAN

東芝半導体工場のビッグデータ活用

1日16億件のビッグデータで
生産能力最大化 と コスト最小化　を追求

製造装置の最適組合せ

装置稼働率の向上

歩留りの改善

コントロールセンター

工場内の”渋滞“を解消

搬送効率の向上

データ活用

データ分析

16億件／日
収集するデータの件数

データ活用

半導体製造装置

ビッグデータ

データ収集

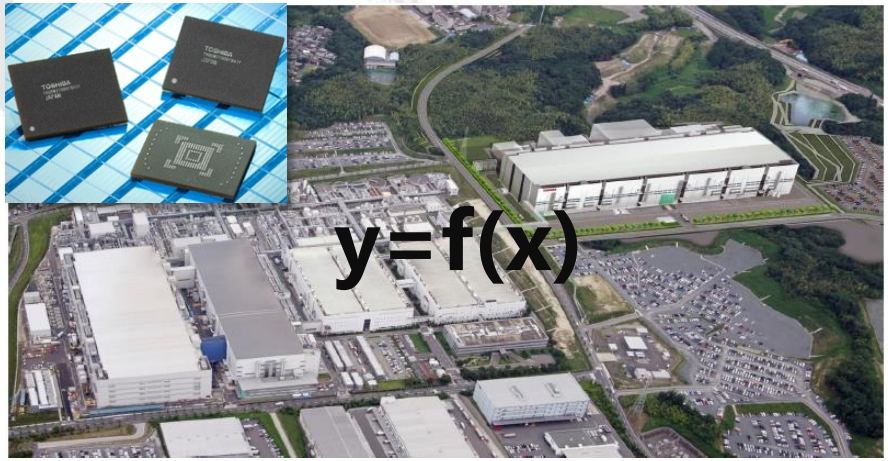
自動搬送システム

50品種/2万工程
生産する製品の種類と
総工程数

200機種/4000台
制御対象の製造装置

0人
ウェハーの
運搬作業者

東芝半導体工場の歩留り・生産効率向上



約4000台 製造装置

ロット投入 → 製造工程 → 製造工程 → 検査工程 → 製造工程 → 製造工程 → 検査工程 → 最終テスト → ロット払出

ペタバイト級 データ

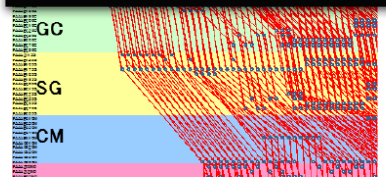
加工条件 … 検査工程データ … 加工時のセンサデータ … 最終検査データ

数兆通り 組合せ

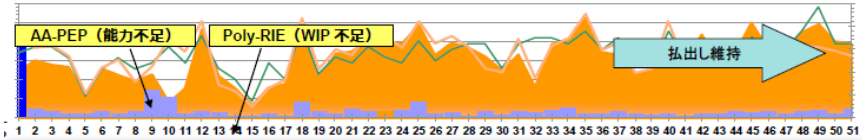
相関分析

データが取得可能な装置等から全てデータを取得し分析＆シュミレーションで活用

ライン平準化による生産効率向上



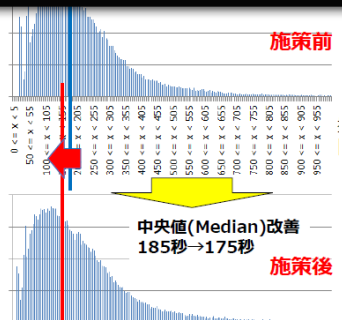
装置情報、払出計画、工程間TAT、装置間リンク制約等をもとに条件展開最適化エンジンをもとにライン平準化を図り、生産効率を最大化



不良パターンの活用により歩留り改善

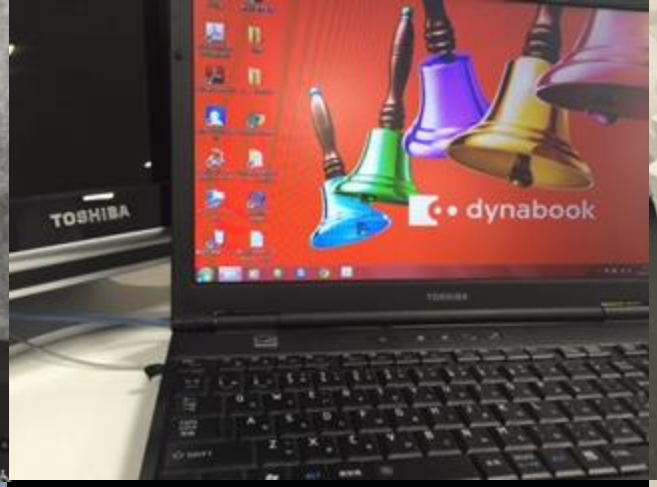
最終検査で不良となるパターンを学習
製品×装置毎に最適値をシミュレーションしリアルタイムに利用製造装置に反映

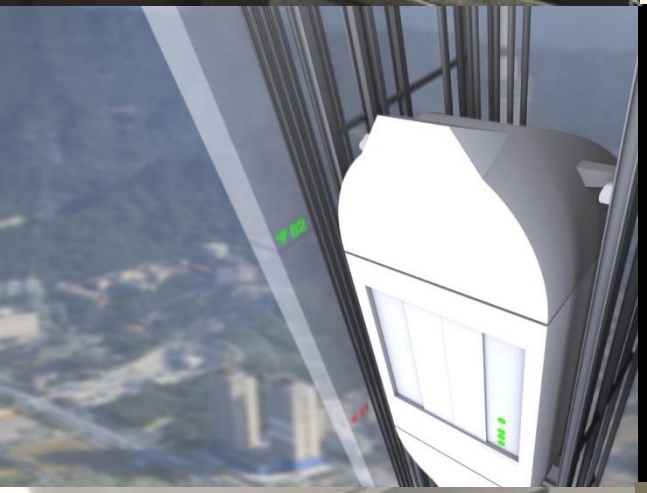
搬送効率化等による稼働率の向上



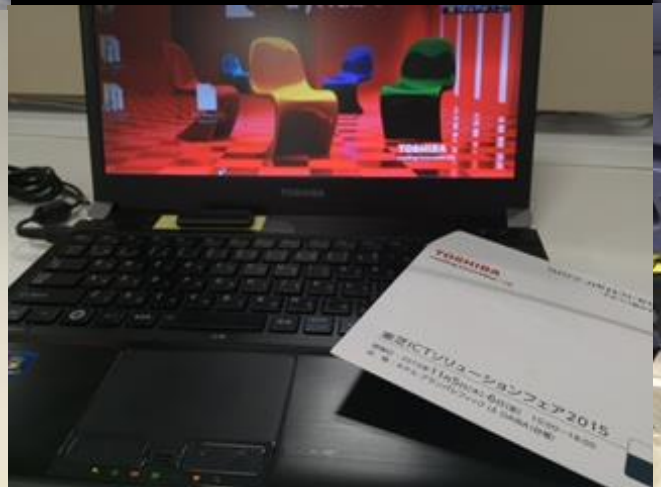
自動搬送や自動着工などのルールをリアルタイムに変更し、製品の流し方を制御
分析結果をもとに、稼働率低下の要因となるものがあれば、即座にシステムの設定を自動で変更


4
4


TOSHIBA
dynabook




東芝ＰＣヘルスモニタ
TOSHIBA
東芝ＰＣヘルスモニタ
電源監視
３Ｄセンサ
熱監視
消費電力
CPU温度
バッテリ残容量
ファン回転速度
良好
バッテリ充電能力
デバイス温度
設定
注意と許諾
閉じる

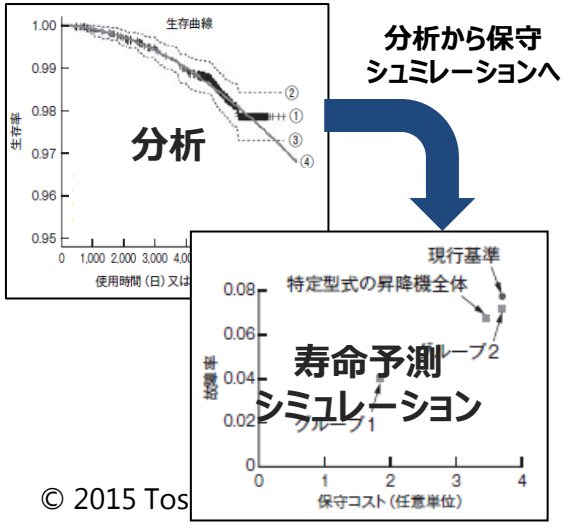


東芝製品の故障予防・ダウンタイム極小化

様々な製品の故障予兆・寿命分析アルゴリズムを開発
停止時間の極小化、不具合の未然防止、メンテナンスの最適化

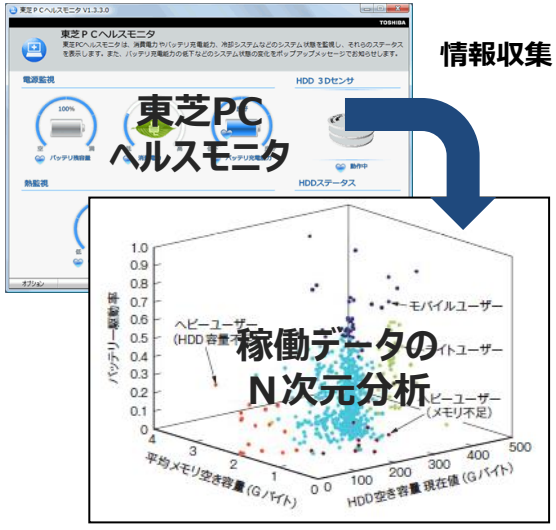
寿命予測
故障実績に加えて、製品の使われ方まで考慮し、部品の寿命を予測

十数年間分の保守履歴データを利用して、データマイニング技術、統計解析のモデル化技術、及びシュミレーション技術を組み合わせた保守計画シミュレータを開発し、およそ2,000種類のエレベーター部品の寿命モデルを構築。



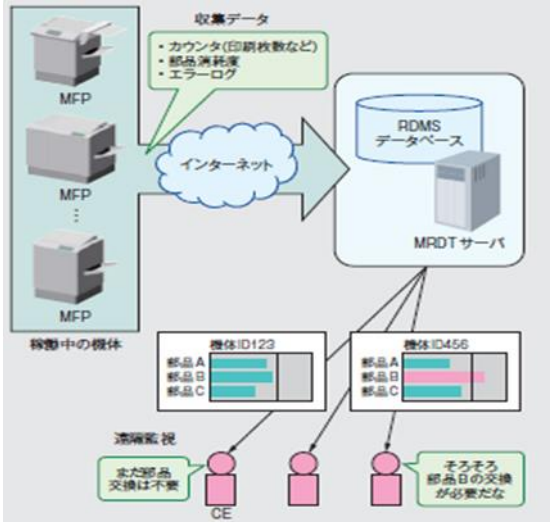
故障予兆分析
約200万台以上のPCのデータを利用した故障予兆診断で、確実なデータ保全を実現

バッテリー劣化や冷却性能など多岐にわたって稼働データの収集・分析を行い、故障予兆検知モデルを構築。故障予兆されたHDDに対して適切な対応を打つことで、ノートPCのデータ保全をより確実に実施。



故障停止時間
遠隔診断と予兆分析でメンテナンスをリコメンドし、故障による停止時間を極小化

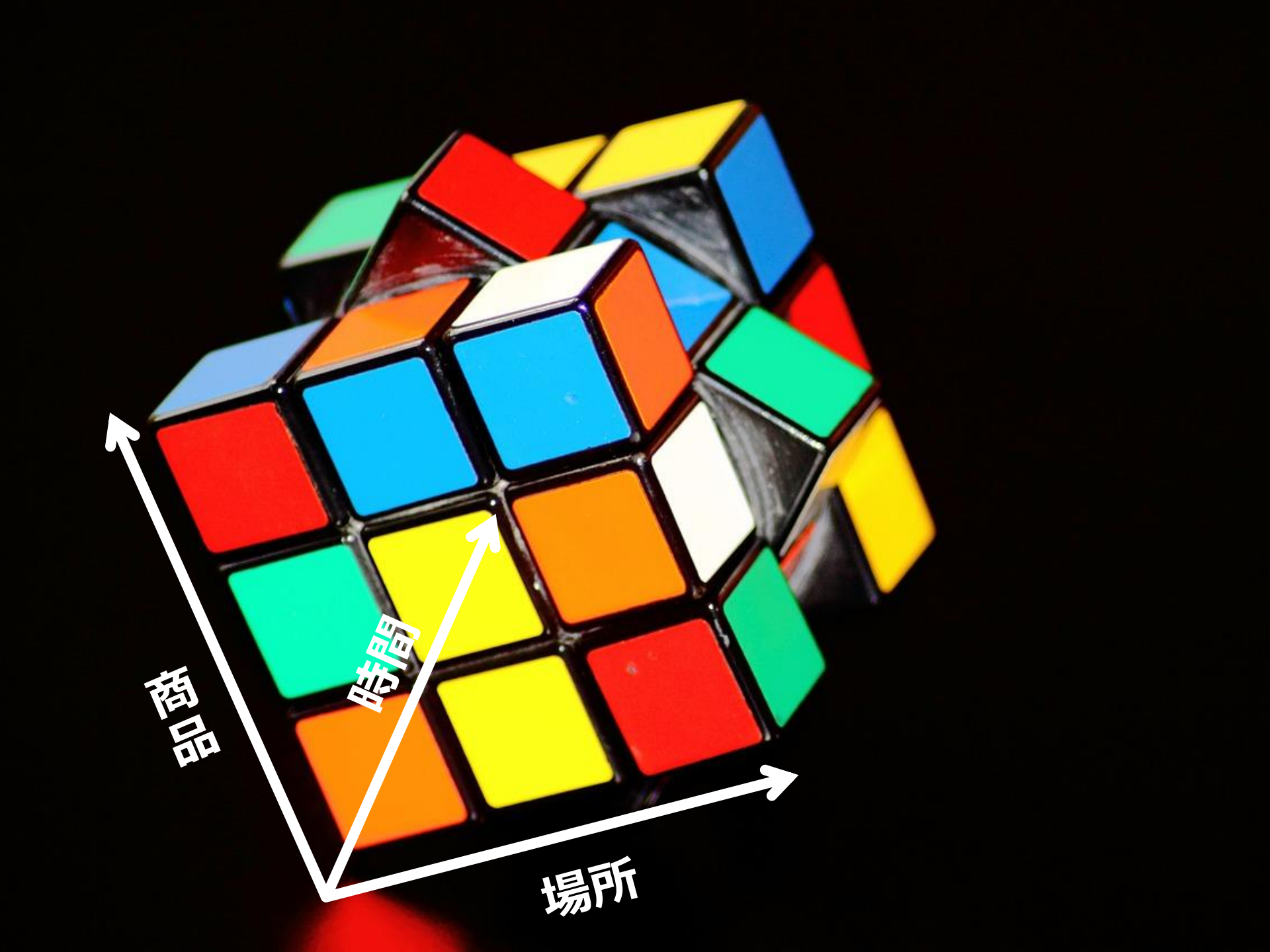
MFPの遠隔管理システムから収集されるデータを利用し、利用状況や部品の消耗度合い、エラーなどの履歴から直近の障害発生を高精度に予測する技術を開発し、経験・勘に頼らない適切なタイミングでのメンテナンスを実現。



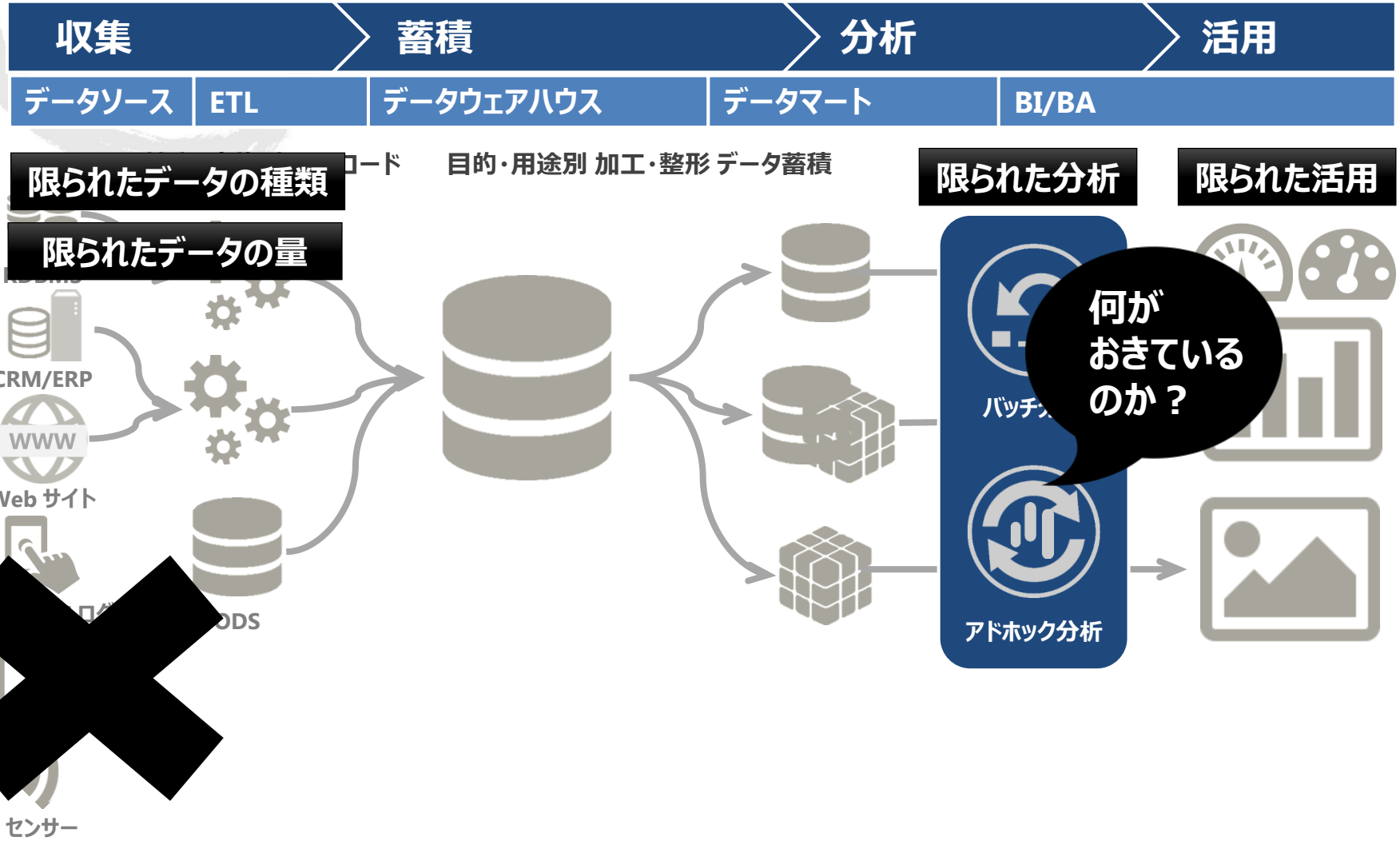
© 2015 Tos

ビッグデータ活用のプロセス

東芝データレイクの取り組み


商品
時間
場所

典型的な データウェアハウス　アーキテクチャ



①データの格納時点で目的・用途に応じて分類や体系化して倉庫(データウェアハウス)に保管
②必要に応じて市場(データマート)に運び込む

©

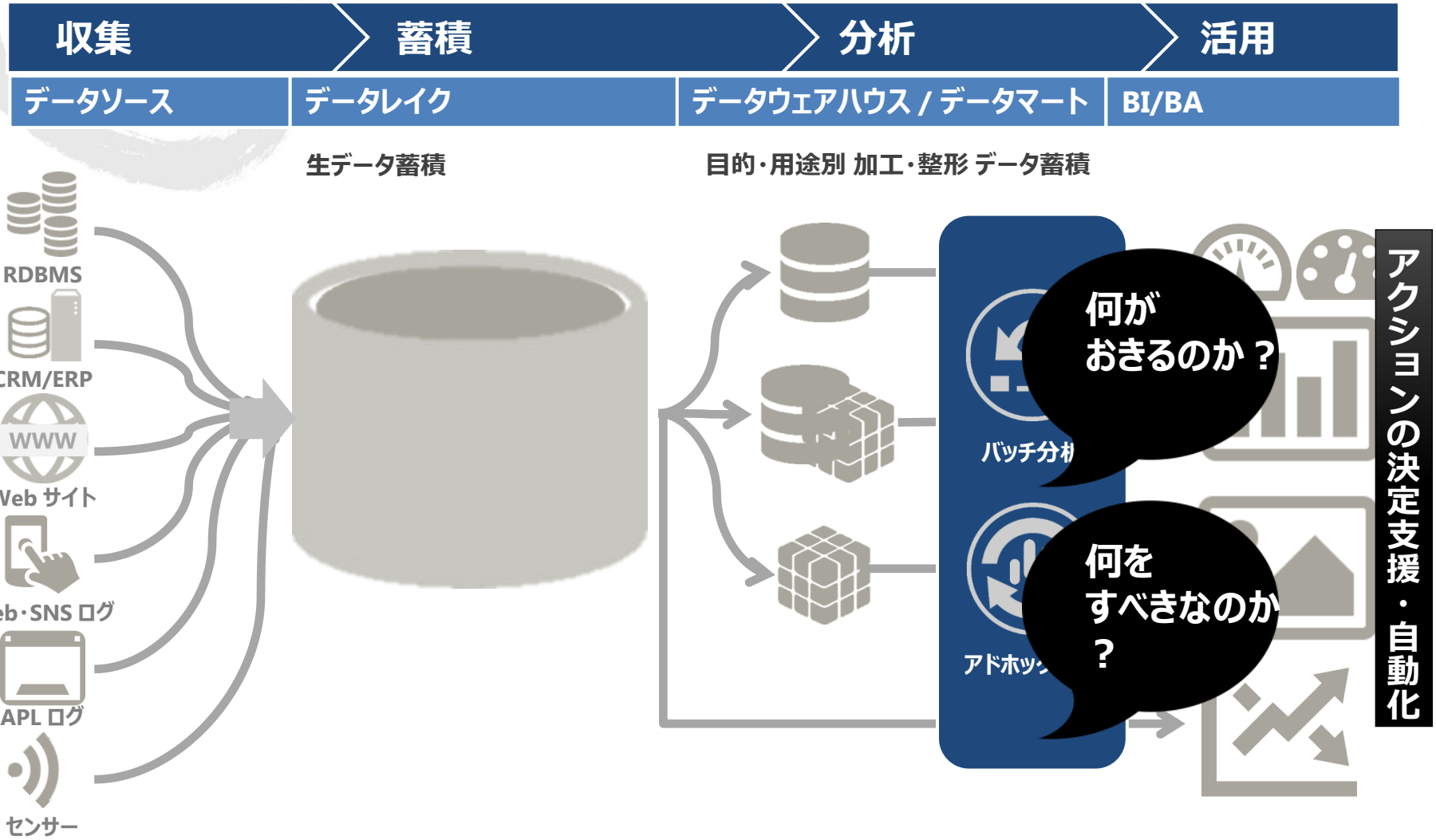
データレイク

ビッグデータの保管に適した（比較的）安価なHW上に構築された大規模な簡単にアクセス可能なデータリポジトリ
米Pentaho社のCTO、ジェームズ・ディクソンが2010/10最初に提唱（出典：Wektionaryより）

すべてのデータ (生データ)
事前処理未 (スキーマレス)
利用者による直接的データアクセス
データ取得が困難
参照時にスキーマ定義が必要

整理されたデータ (多次元モデル化データ)
事前処理済 (スキーマフル)
利用者による間接的データアクセス
データ取得が容易
新しい要素の追加が困難

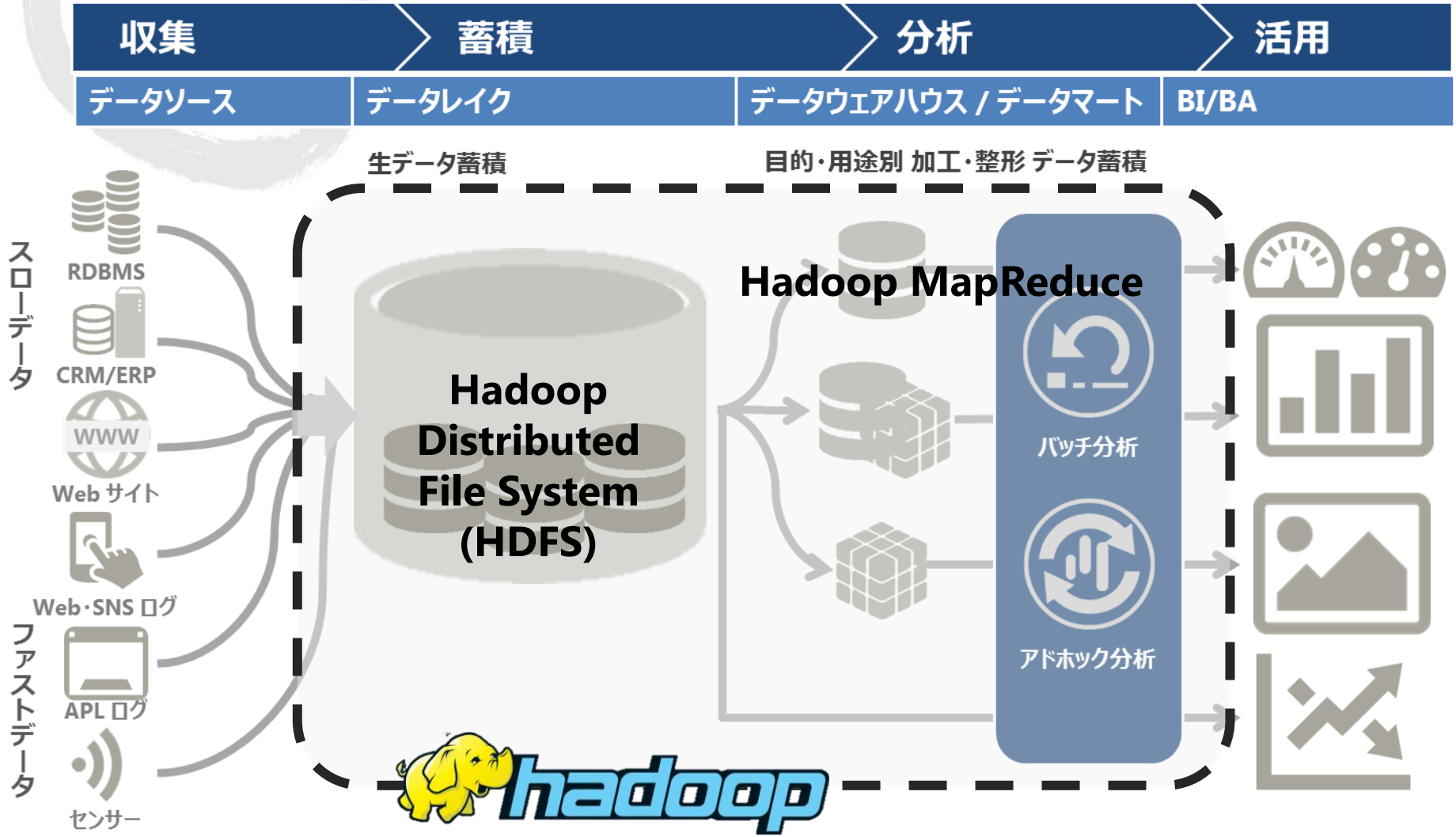
一般的な データレイク アーキテクチャ



①湧水、雨水…(生データ)を湖(データレイク)に貯めておく。②用途に応じて蒸溜・加工して
③必要に応じて倉庫(データウェアハウス)に移して、④必要に応じて市場(データマート)に運び込む

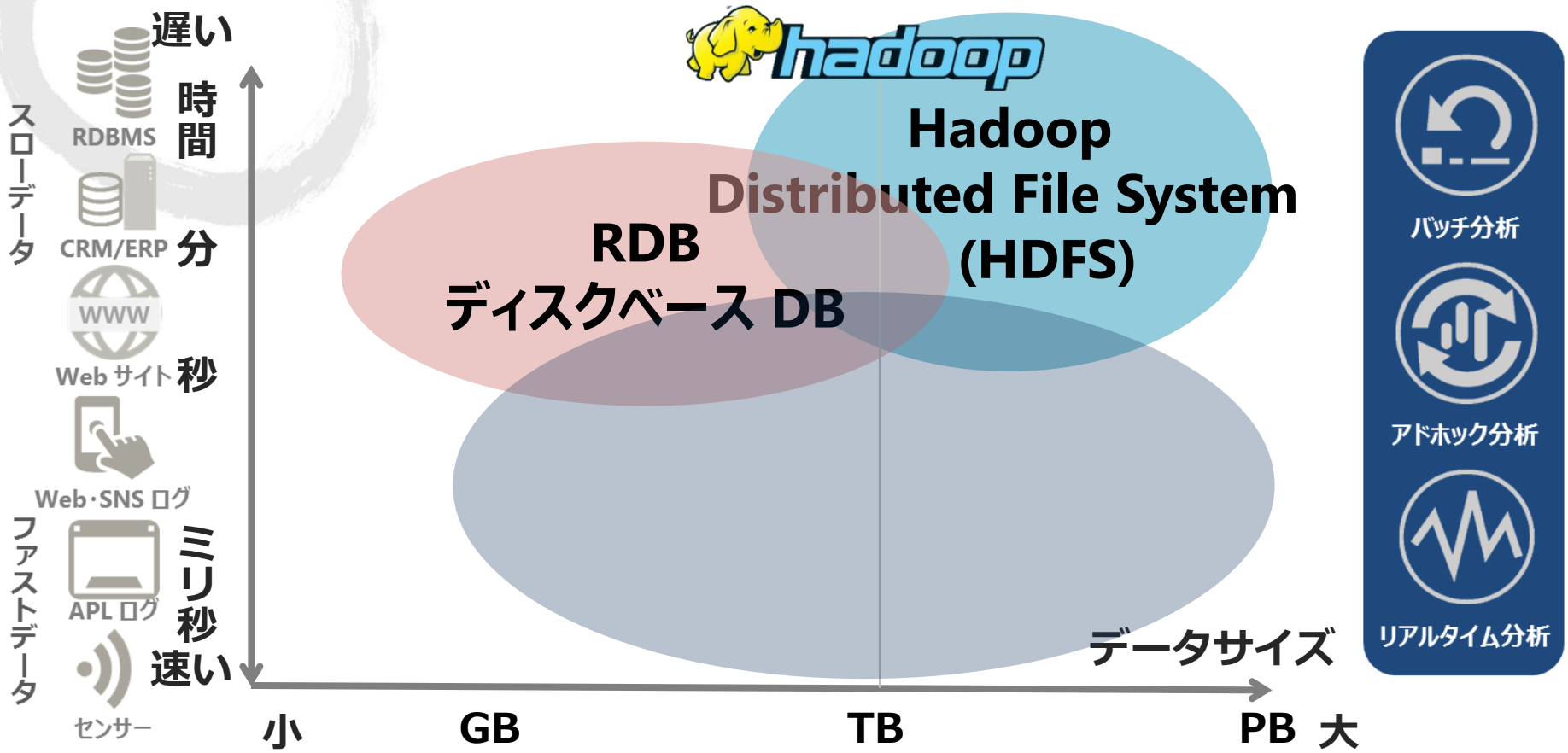
©

一般的な データレイク アーキテクチャ



© 2015 Toshiba Corporation

Hadoop Distributed File System (HDFS)



ファイルサーバの視点

- 一般的なファイルシステムのインターフェイスをもっていない。もしくは足りない
- 小さいサイズを扱うには適していない
- 単一障害をなんとかしなければならない

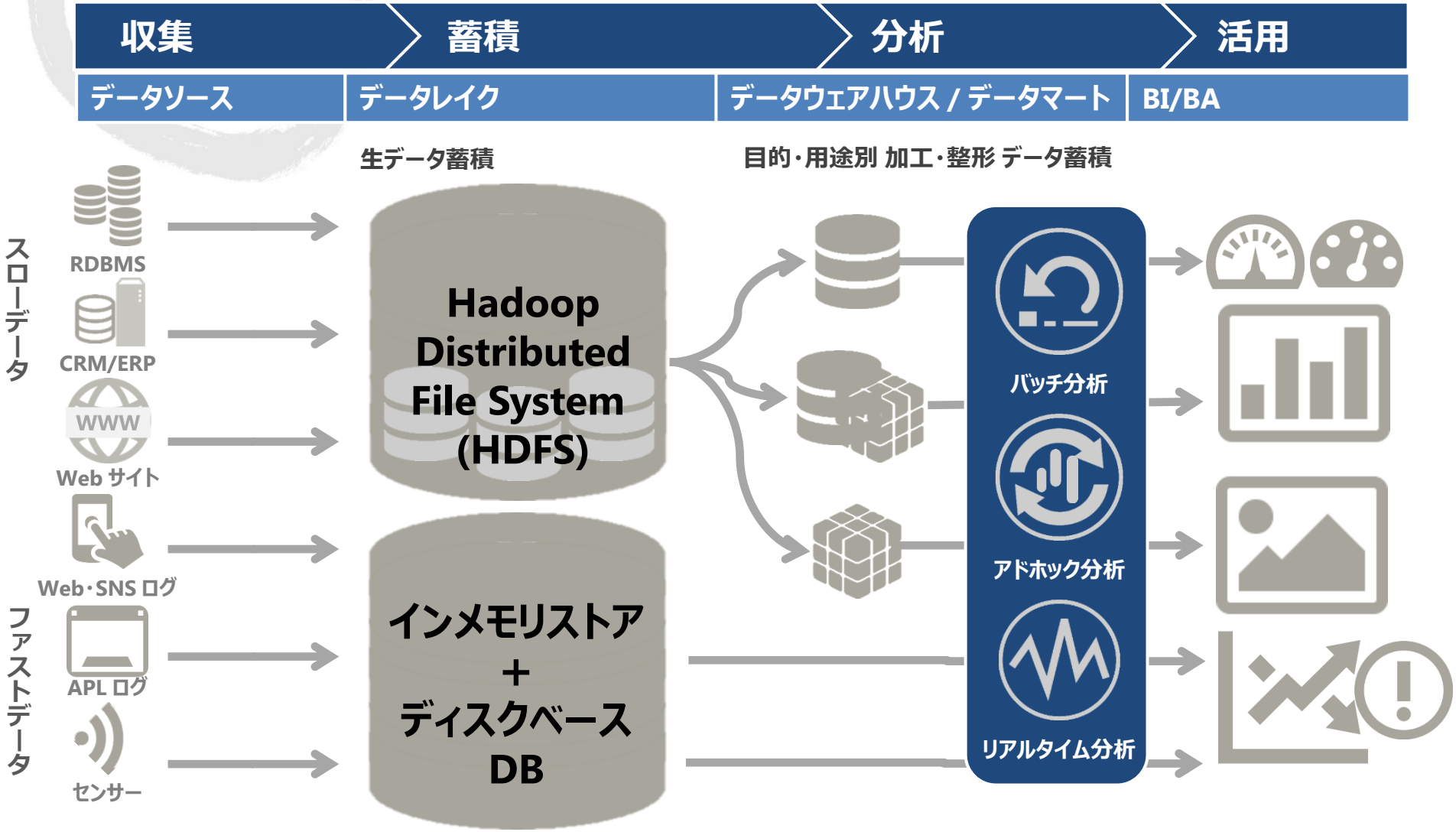
オンライン処理の視点

- 素早くレスポンスを返すのは苦手。
  他との組合せが必要
- トランザクションを扱う機能が欠けている。
  エラーが起きた時にどうするのか？
  RDBとの組合せが必要である
- 同時アクセス時の制御が難しい

リレーショナルデータベースの視点

- 書き込みは 1 回限りで読み取りは何度もできるという、write-once-read-many モデルでデータのピンポイントで削除したり更新したりできない
- フルスキャンのみ。検索を高速化するためのインデックス機能はない

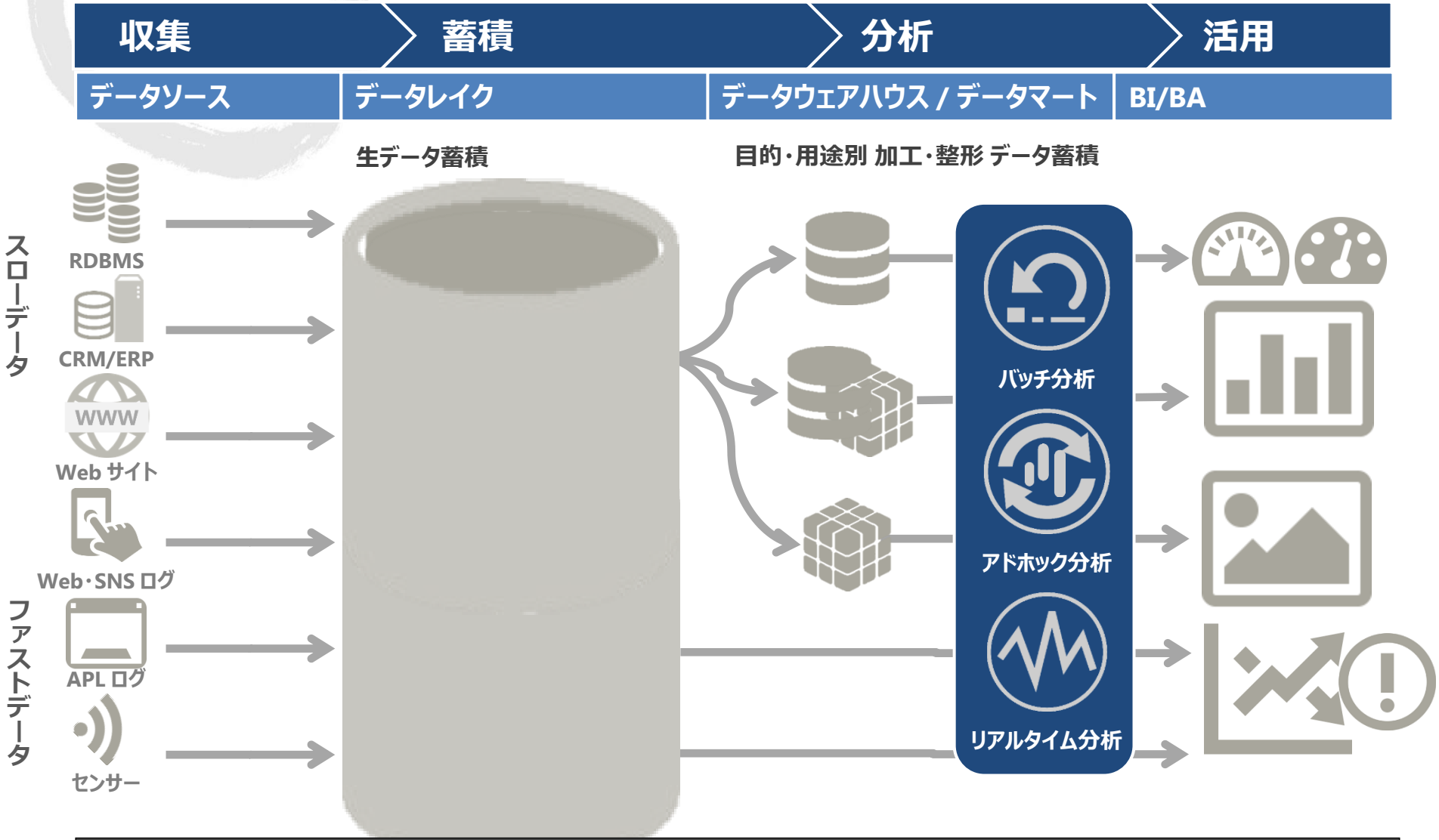
リアルタイム処理とバッチ処理の両立



Hadoopでは不十分なので、他の組み合わせが必要・・・一元化できない

©

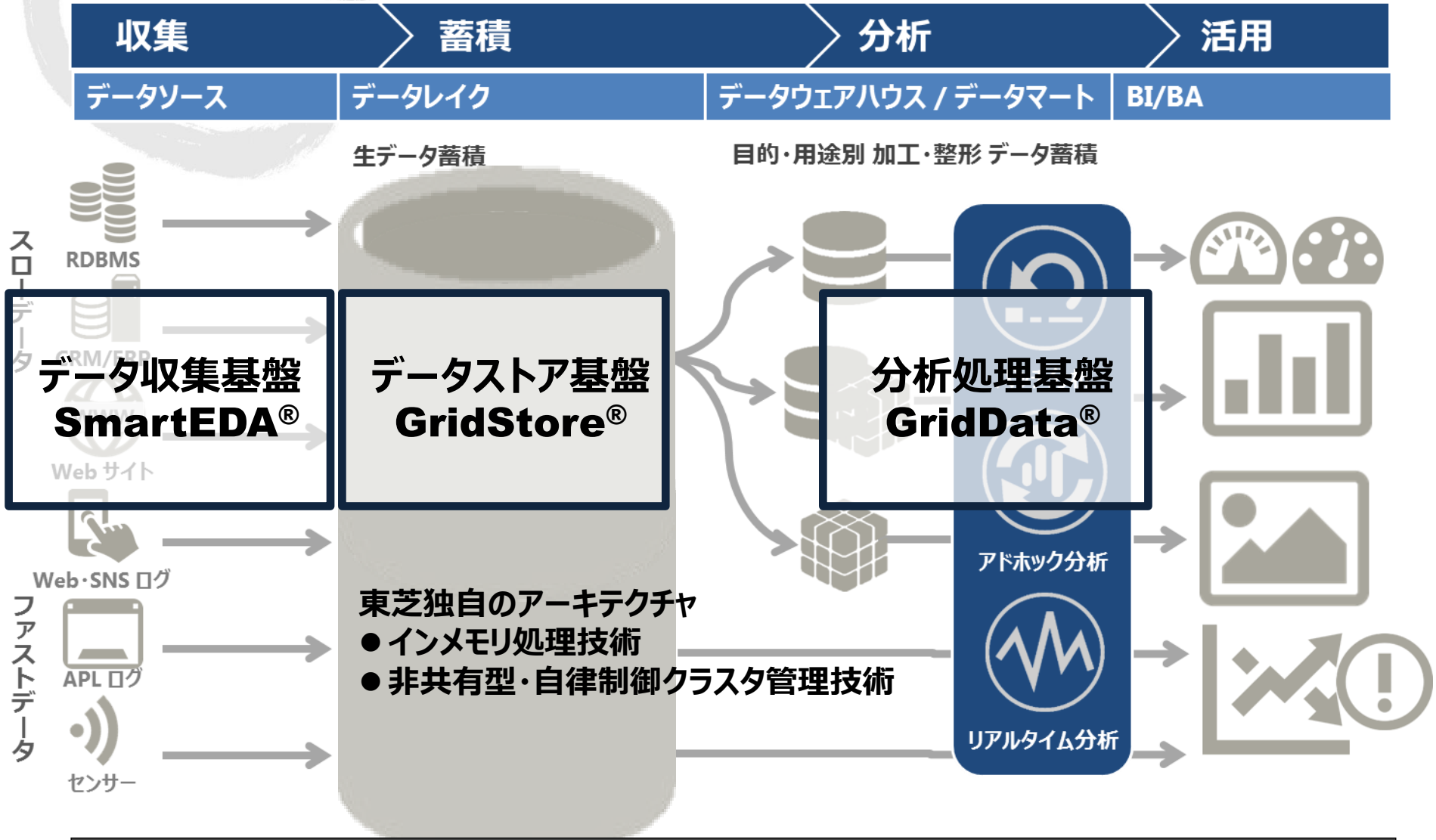
ファストデータとスローデータの一元化



一元化するには・・・高速性＆大容量性の両立させたものが必要

©

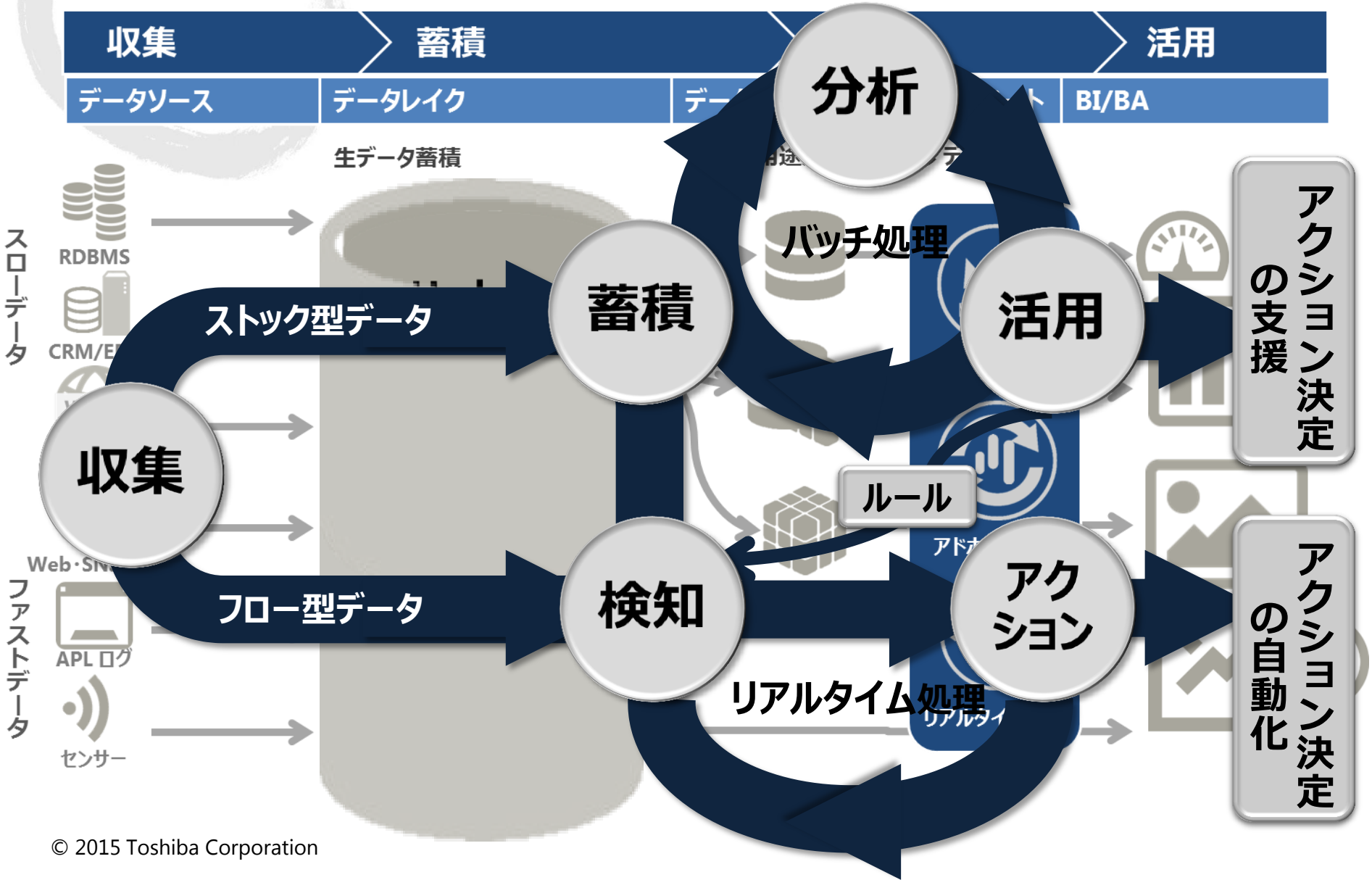
東芝データレイク アーキテクチャ



メモリがもつ高速性とディスクのもつ大容量性を両立

©

東芝データレイク アーキテクチャ

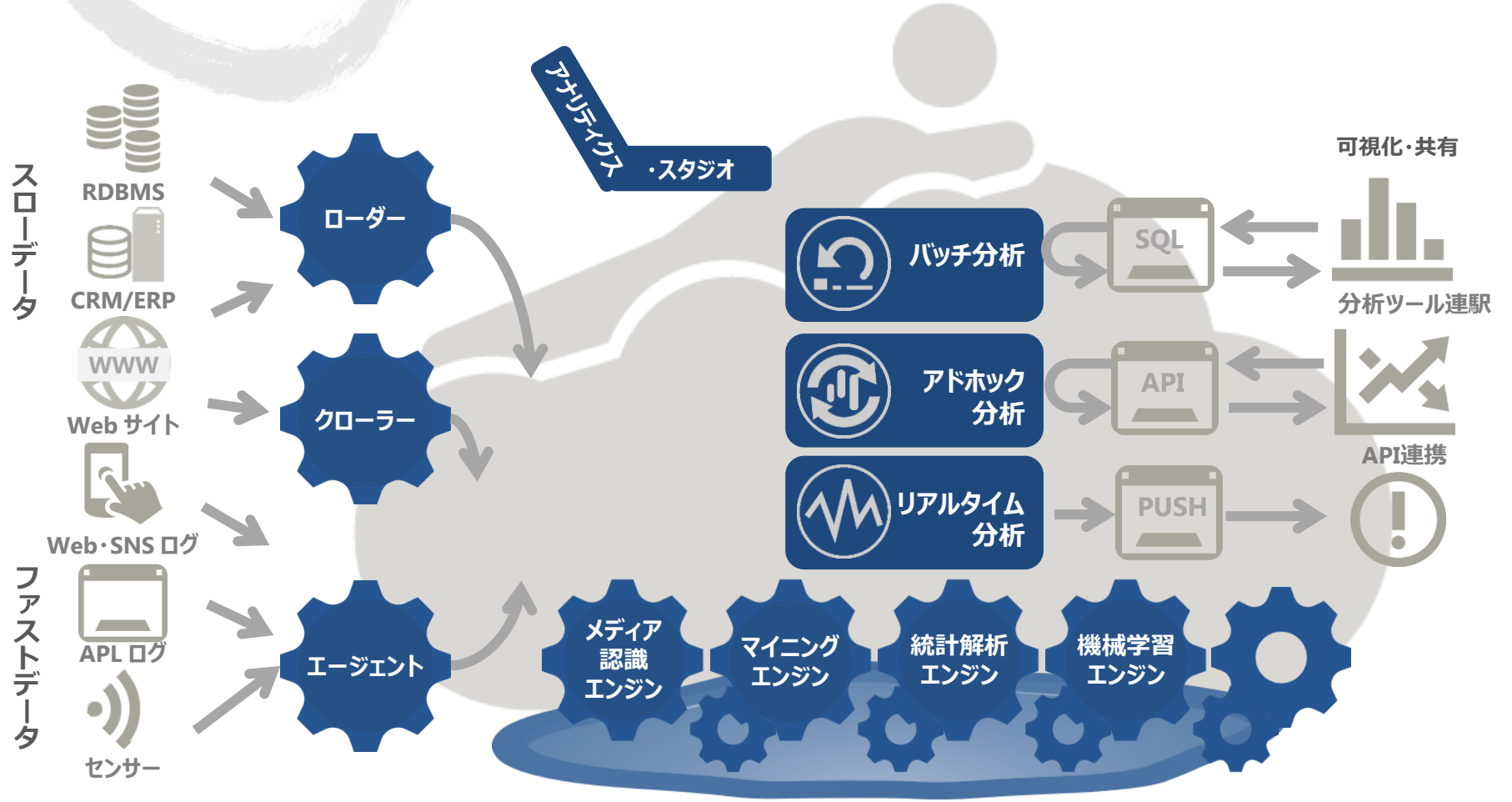


© 2015 Toshiba Corporation


東芝データレイクサービス

東芝データレイクサービス

“簡単!”・“賢く!” ビッグデータ活かす収集・蓄積・分析サービス



© 2

東芝データレイクサービスの特長

“簡単!”・“賢く!” ビッグデータ活かす収集・蓄積・分析サービス

POINT 1 初期コストを抑えて、スタート効果を先取り

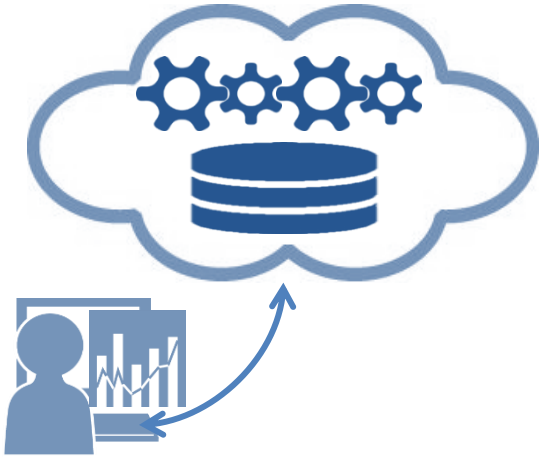
フルマネージドのクラウドサービスなので、ワンクリックするだけで、必要なインフラの環境設定作業などの準備に手を煩わすことなく、速やかにかつ初期コストも抑えてスタートし効果を先取り可能。また、インフラなどのメンテナンスも不要

POINT 2 実証済みの分析作業のテンプレートを用意

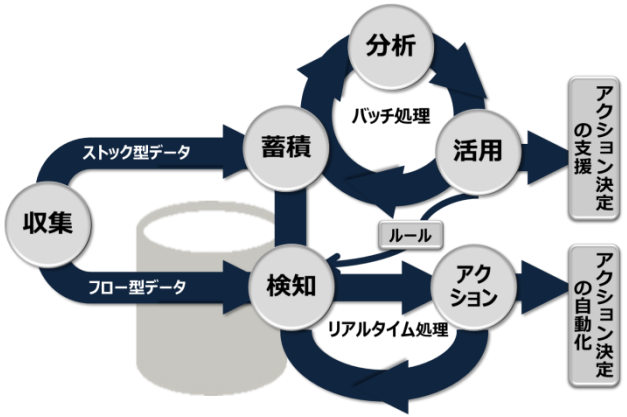
アナリテックス・スタジオ（作業環境）のギャラリーから実証済みの分析作業のテンプレートからパターンを選ぶだけで、自動的にビッグデータの収集・蓄積・分析の一連の作業をスルーして効率よく実行

POINT 3 導いたルールでアクション決定・自動化の仕組み

東芝独自の各種分析アルゴリズムやストア型とフロー型のデータ処理の組合わせの仕組みにより、「原因把握・情勢判断」だけではなく、導出したルールを用いた「アクション決定の支援・自動化」まで、高度なデータ活用を容易に実現



実証済み分析作業テンプレート



収集

さまざまなデータソースと取得手段に対応したコレクターで簡単!
エッジ（データソース）側での「集めないビッグデータ®」で賢く!

POINT 1 さまざまなデータの高度な取得・収集を簡単に実現

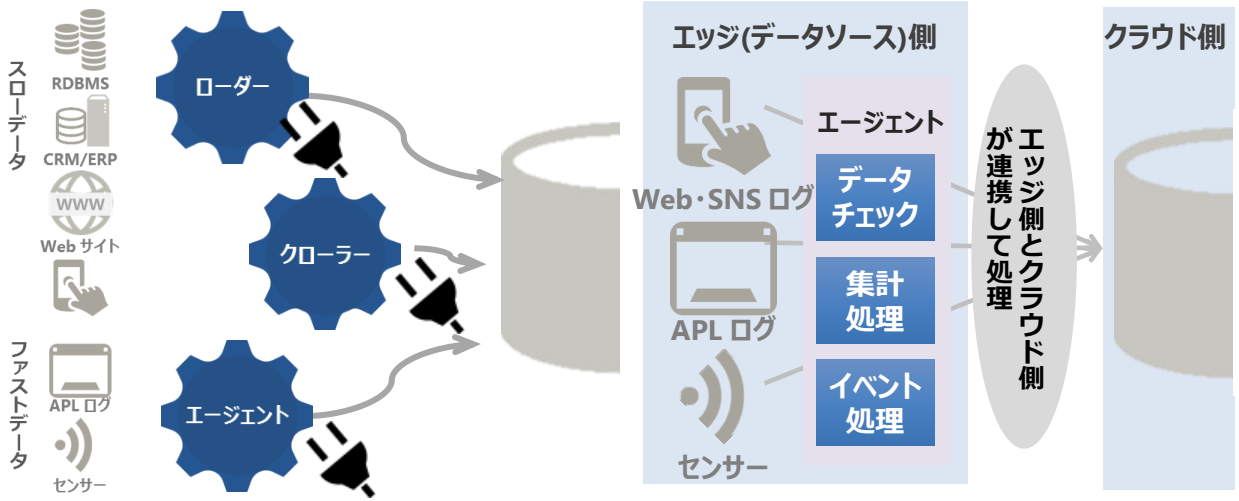
用途に応じたコレクター(ローダー/クローラー/エージェント)とデータの抽出・変換・加工や認識、意味づけなどの前処理ができるプラグインを用意しており、高度なデータ収集が容易に可能

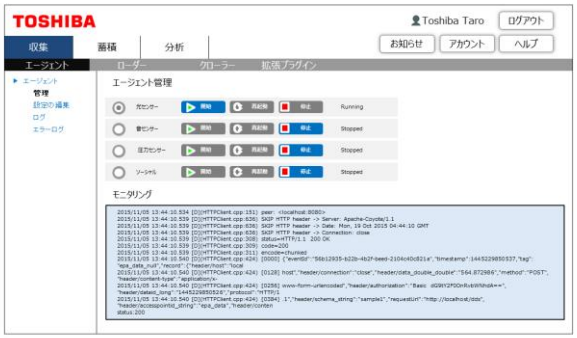
POINT 2 「集めないビッグデータ®」で、リアルタイムに収集

アプリログ、センサーデータなどの膨大な非構造化ビッグデータをエッジ(データソース)側のエージェントの「集めないビッグデータ®」テクノロジーで、ネットワークトラフィックを抑えつつ、リアルタイムに収集

POINT 3 各種ビジネスデータやWebデータを高速収集

- RDBMSやCRM/ERPなどの構造化データを各種ローダーで高速ロード
- ウェブ上の文書や画像など周期的に取得し、自動的に認識・加工しながら収集するクローラーを用意





蓄積

時々刻々に発生する膨大なビッグデータ
“賢く!” 高速に登録・更新・検索＆ “簡単に!”長期に安定保管

POINT 1 リアルタイム性重視 遅延なく登録・更新・検索

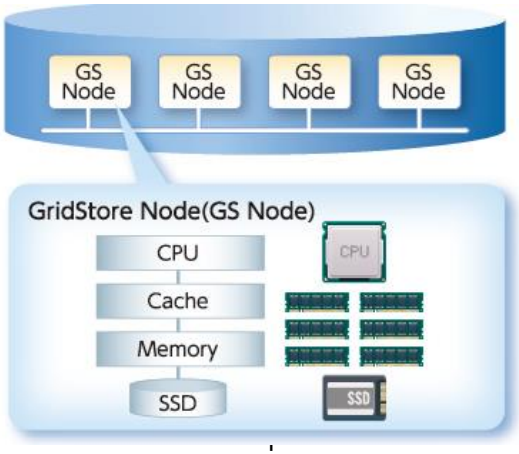
- 東芝独自のメモリが主・ディスクが従のアプローチとCPU処理オーバヘッド排除を追求したインメモリ処理技術で高速レスポンスや高スループットを実現
- SQLを用いた高速なアドホック分析も可能

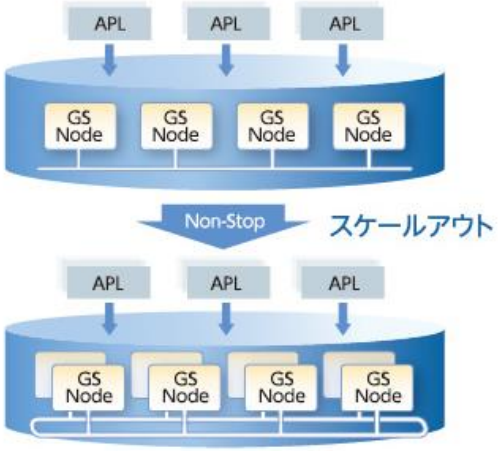
POINT 2 容量や性能に応じた 拡張・縮退

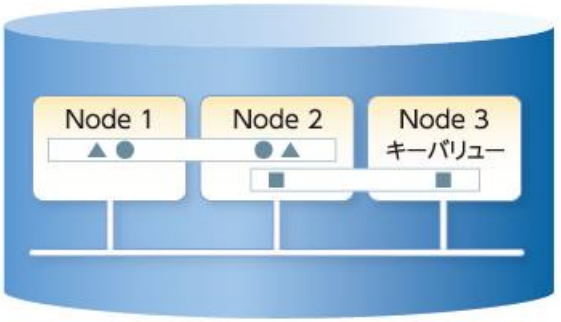
- 構造化・非構造化を問わず、あらゆるデータを１つに集約して長期間保存・蓄積
- 非共有型クラスタ技術により、将来予想ができない容量追加や性能強化はノードの追加のみで容易に可能

POINT 3 24時間を通して 安定運用

- 明示的なマスターノードを必要としない東芝独自の自律制御クラスタ管理技術により、容量と性能の拡張が無停止で可能
- 自動データ同期複製・自律的マスター決定により、障害時には素早く処理を引き継ぎサービスを継続





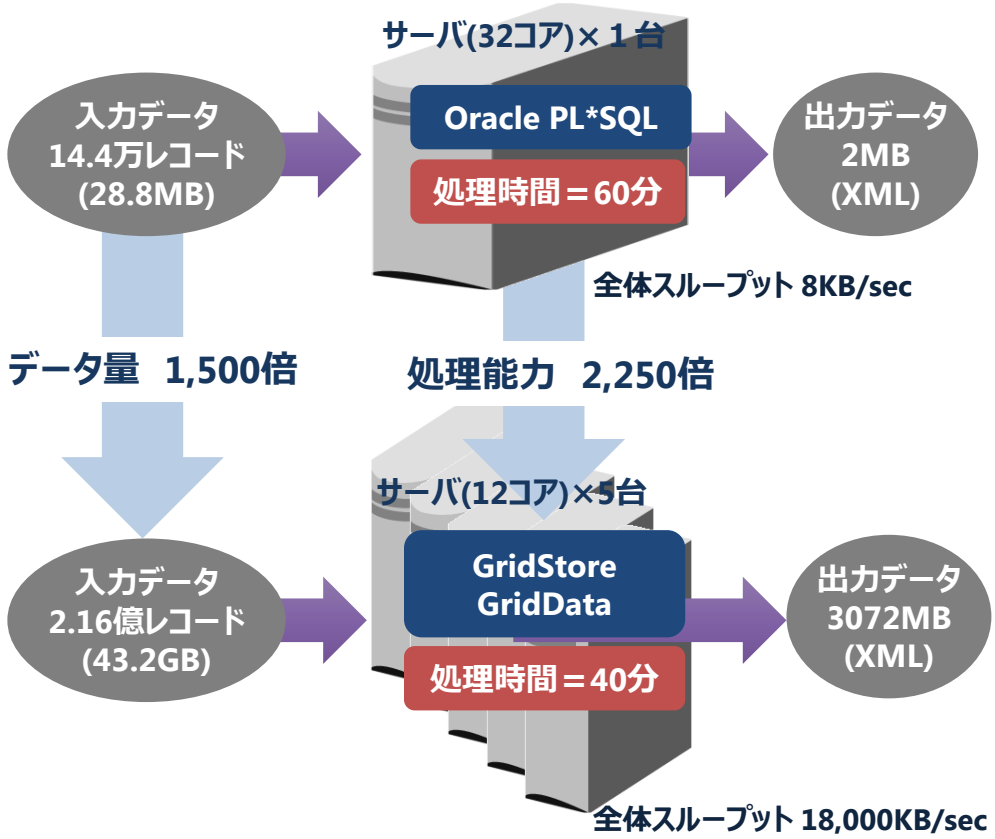


データ(図:▲●■)を自動的に複数ノードに配置

某社様　業務で扱うデータ処理の高速化

東芝のビッグデータ技術を適用し、1,500倍のデータ量を従来比の2,250倍の処理能力で対応

電力小売り事業者に対し、電力送配電網を提供し、契約ユーザの利用量に応じた料金を請求するシステム
電力の自由化に伴い、多数の電力小売り事業者が参入し、契約数の増加（3,000契約→450万契約）による
データ量の爆発的増加へビッグデータ技術を適用し対応



©

ニュースリリース

2015年9月9日

ビッグデータテクノロジーの普及・促進に向けた業界団体「Open Data Platform initiative」に加入
～加盟各社との連携強化を目的に大容量スケールアウト型データベース「GridStore」のソースコードを公開～

マークの付いたリンクは、別ウィンドウで開きます。

当社は、ビッグデータ分野の大手企業が設立を進めている業界団体「Open Data Platform initiative(以下、ODP)」に、設立メンバーとして加入します。ODPは、ビッグデータを効率的に処理・管理するオープンソースソフトウェア「Apache Hadoop」とメンバー各社のビッグデータ関連ソフトウェアの相互接続性を高めることで、ビッグデータテクノロジーの普及・促進に取り組む業界団体です。

ODPに加入することで、当社のビッグデータ関連ソフトウェアと他社ソフトウェアおよび「Apache Hadoop」との相互接続性を高め、加盟各社との連携を強化します。これにより、ビッグデータの利活用を行うユーザーは、相互接続性の高い開発環境で、効率的にシステムを構築することができるようになります。

また、ODPへの加入を機に、当社は、ビッグデータテクノロジーの普及促進を加速するため、自社開発のビッグデータ向け大容量スケールアウト型データベース「GridStore」のソースコードを来年２月から公開します。

「GridStore」は、当社が社会インフラシステム向けの大規模・高速データ管理用として開発したデータベース管理システムです。従来ハードディスクに配置していたデータを、メインメモリに配置することで、高速なデータベース処理を実現しました。また、少量のデータを取り扱う際には少数のサーバで処理することで初期投資を抑え、データの増加に応じ、サーバを増やすことで大量のデータ処理に対応するスケールアウトが可能です。

今回、「GridStore」のソースコードを開示することで、ユーザーは、変化のスピードが著しいソフトウェアやシステムへのニーズに自身で対応できるようになります。また、「GridStore」オープンソースコミュニティ注の活動によって「GridStore」の機能や利便性がさらに高まります。

当社は、オープンソースソフトウェアをコアにビッグデータテクノロジーの普及を促進するODPと連携し、さまざまなシーンでのビッグデータ利活用に貢献していきます。

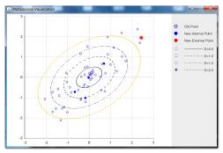

SELECT CATEGORY
-CULTURE
-ECONOMIC
-FINANCE
-BUSINESS
-MEDIA
-PEOPLE
-CREATIVE
-TUTORIALS
-INVESTMENT
-NETWORKING
MEDIA
-VIDEO
-MUSIC
-FILMS
-SEARCH
-CONTACTS
-MESSAGES
-SHOW BUSINESS
-NETWORK
-MUSIC
-CINEMA
-BUSINESS/FINANCE
-WORLD NEWS
DISTRIBUTION OF NEW GOODS IN SECONDARY MARKETS
NETWORK SEARCH
-PEOPLE
-FORUMS
-MAIL
-SHOP
-BUY
-SALE
180 EUROPE
210 USA
240 ASIA
240 AFRICA
222 CHINA
190 INDIA
170 UK
222 CANADA
240 RUSSIA
WORLD
PROJECTED SALES GROWTH D
COMMON CEREALS AS A PERCENTAGE
LOADING 100%
BUSINESS
NETWORK SEARCH
TOTAL SALES BY REGION
TOTAL SALES BY CATEGORY
11%
5%
14%
8%
11%
5%
26%
JAN
FEB
MAR
APR
MAY
JUN
Silver
Platinum
Gold
Diamonds
Oil
Gas
Electric power
Medicine
Textile
IT
Electronics
Media
TOTAL SALES BY REGION
REGION SALES
West € 1 236 345,0
South € 1 896 354,0
North € 2 569 345,0
East € 1 893 543,0
Total € 7 595 587,0
SALES BY CATEGORY
SALES
Copper 85
Steel 89
Gold 74
Silver 50
Platinum 27
SALES BY CATEGORY
SALES
West 930
South 854
North 652
East 450
Other 223
1010110110101010110

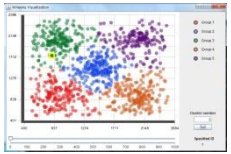
分析

実証済み分析シナリオのパターンとインタラクティブ環境で簡単!
さまざまな(東芝独自含む)分析アルゴリズムのエンジンで賢く!

POINT 1 実証済みの分析パターンで、専門知識なくても…

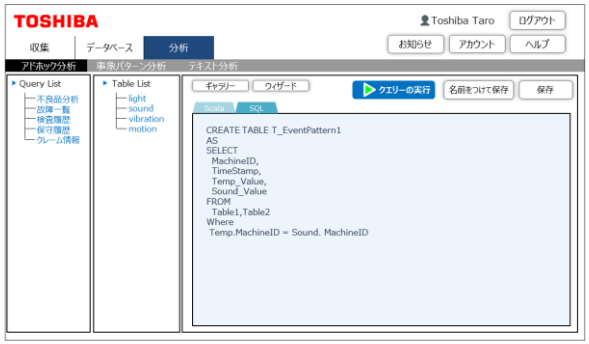
- 専門知識がなくても、用意された実証済みの分析パターンを選択するだけで、スタート可能
- バッチ分析（定型分析)、アドホック分析（非定型分析)、リアルタイム分析と分析結果の可視化やアクションの仕組みを用意

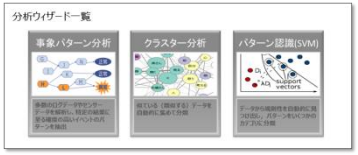




POINT 2 泥臭い作業をインタラクティブな環境でサポート

- 面倒な試行錯誤などの泥臭い分析作業をインタラクティブな環境でSQLやScalaを実行することで、スマートに対応
- 分析シナリオの分析パターンを作成することで、分析作業の自動化が簡単に実現





POINT 3 東芝独自のアルゴリズムを分析エンジンとして搭載

- マイニング(SparkSQL) 、統計解析(SparkR)や機械学習(MLlib)などさまざまな分析アルゴリズムを搭載したエンジンを用意
- 東芝独自のアルゴリズムの分析シナリオのパターンで、より高度なルールを容易に発見

事象パターン分析(時系列分析)エンジン

過去のさまざまな事象(イベント)から特定の結果に至る確率の高いパターン (ルール)を事前に掴んで、最適なアクションを実現

POINT 1 大量の事象からある事象に至るまでのパターンを抽出

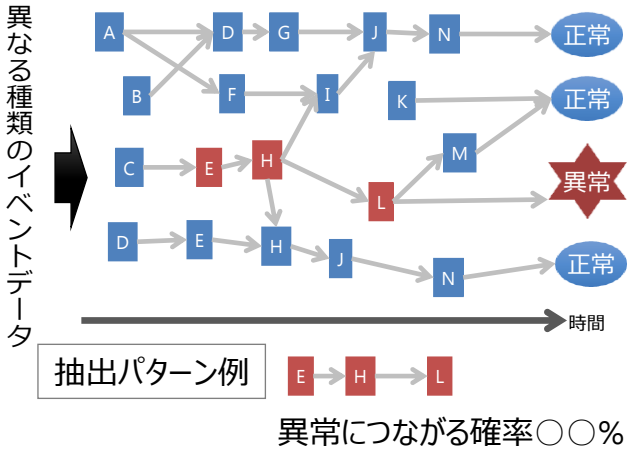
- 東芝独自の人手では発見できない大量の過去の事象(イベント)から、ある特定の事象の結果に至る確率の高いパターン(ルール)を抽出する技術
- 「異常予測」、「異常内容推定」、「原因推定」などで活用

POINT 2 順序だけではなく、時間間隔も高速に予測可能

- 事象の順序関係だけではなく、事象間の時間間隔も含んだ精度の高いパターンを抽出
- 事象を指定し、そこか遡って頻出するパターンを抽出する方法なので分析する対象が限定でき高速なパターン抽出が可能

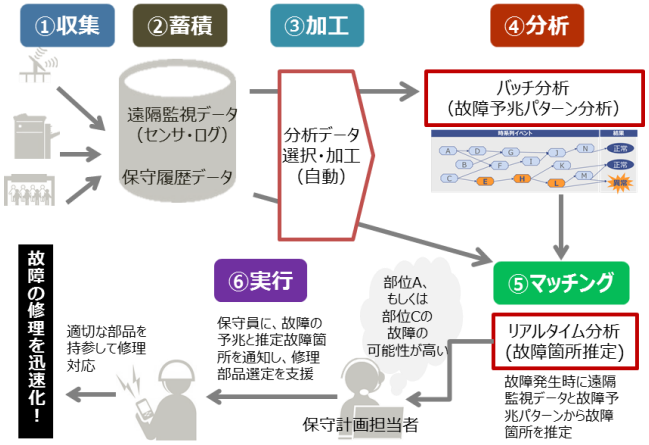
POINT 3 リアルタイム分析と組合せでいち早く対処可能

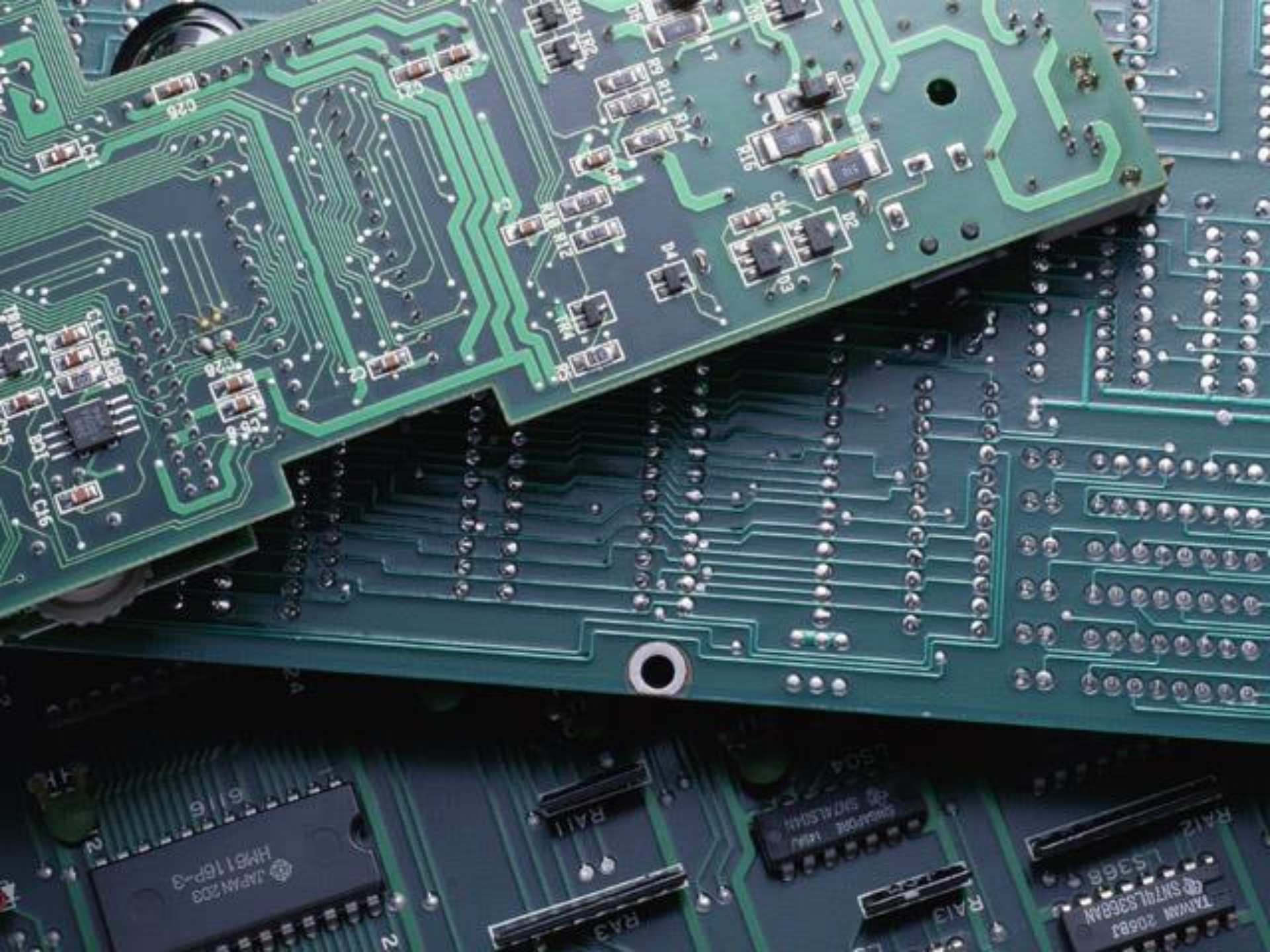
- 導いたルールをリアルタイム分析に適用し、ある特定の事象 (故障など)が発生する前の段階で対処するなどが可能
- すなわち。今の状況からチャンスやリスクを掴み、いち早く対処することが可能



予防保全 保守効率化	故障を未然に防ぎダウンタイム削減 保守項目の優先順位付け
「特定の結果の発生パターン」と「実データ」とのマッチングによる「異常予測」での活用	
保守迅速化	保守部品を事前に用意
「特定の結果の発生パターン」と異常発生時データのマッチングによる「異常内容推定」での活用	
品質改善	パターンから故障原因を特定
「特定の結果の発生パターン」の深掘りによる「原因推定」での活用	

故障箇所の推定による保守部品調達の迅速化





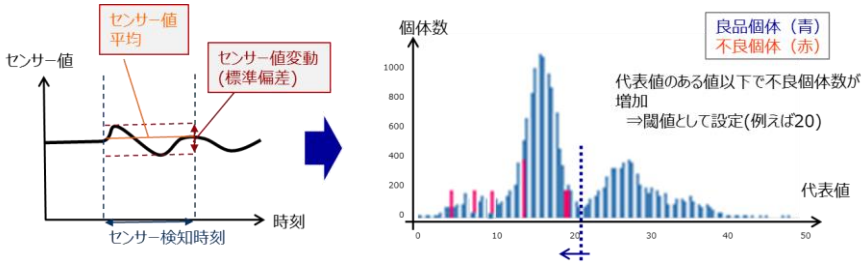
某社様 製造ラインの不具合パターン検出

東芝独自分析手法を適用し、イベント抽出/パターン分析
「匠」でも気づかなかった不具合パターンを発見

❶ 装置データを収集し、製品・時系列で紐付け

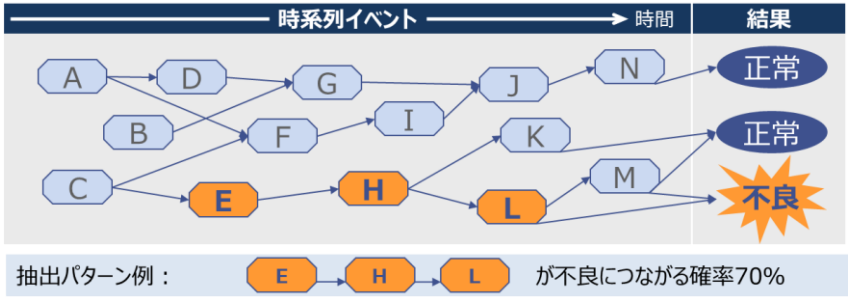
❷東芝独自手法で相関関係からイベント抽出

部品個体別装置センサー値の代表値(平均値、標準偏差)を算出し、不良の種類ごとに、統計上有意な相関のある、センサー代表値の閾値を算出。その閾値を利用して、イベント化を実施



❸東芝独自パターン分析で不具合パターン抽出

多数のログデータやセンサーデータを解析し、特定の結果に至る確度の高いパターンを抽出を実施



❹見える化により新たな気づきをサポート

製造現場の「匠」も気づかなかった不具合の因果関係(パターン)を発見

テキストマイニング(文書分類)エンジン

さまざまな形式(SNSデータ/会話的テキスト/論文…)の膨大なテキストデータを利用目的に応じた手法で自動的に分類

POINT 1 隠れた話題・関心を把握(教師なし話題分類)

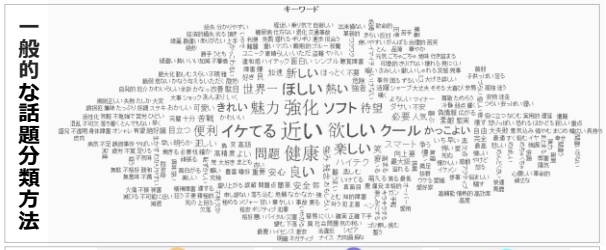
- キーワードの出現回数ではなく統計情報を用いて似た意味の単語の集まりとして話題を抽出し、確率的に近い意味を持つ概念として分類(辞書・キーワード不要)
- 1万件/分/台の高速処理で出現回数が少ない隠れた話題を発見

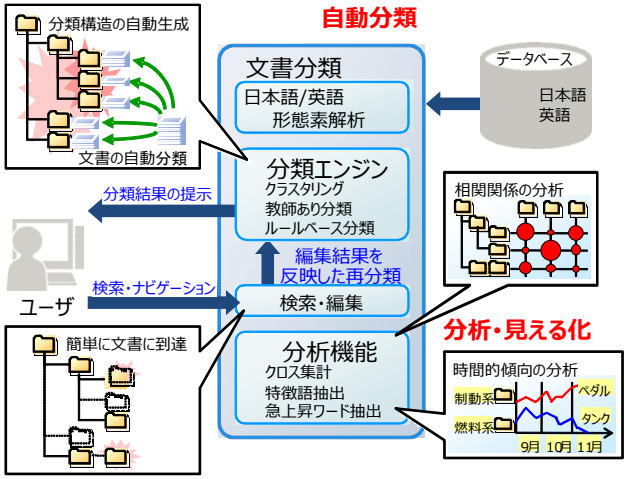
POINT 2 階層分類 (タクソノミ)(教師なし階層分類)

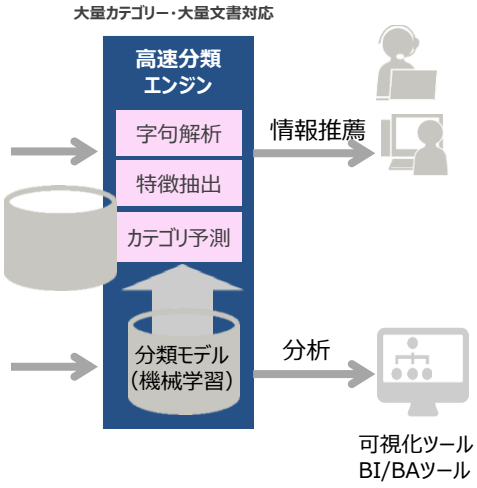
- 東芝が長年培ってきた自然言語意味解析技術を用いて、記述内容に応じて自動的に分類カテゴリを生成し、文書を分類
- 大量文書を5万件/分/台の高速処理階層的に体系化し、情報アクセスの効率化を向上

POINT 3 高速分類(教師あり高速分類)

- 類似語を自動認識することで分類数を大幅に削減した高速分類
- 単語辞書を用いないため、言語や業務分野を問わず利用可能
- 東日本大震災時のピーク時の流量(50万ツィート/分)に遅延なく分類





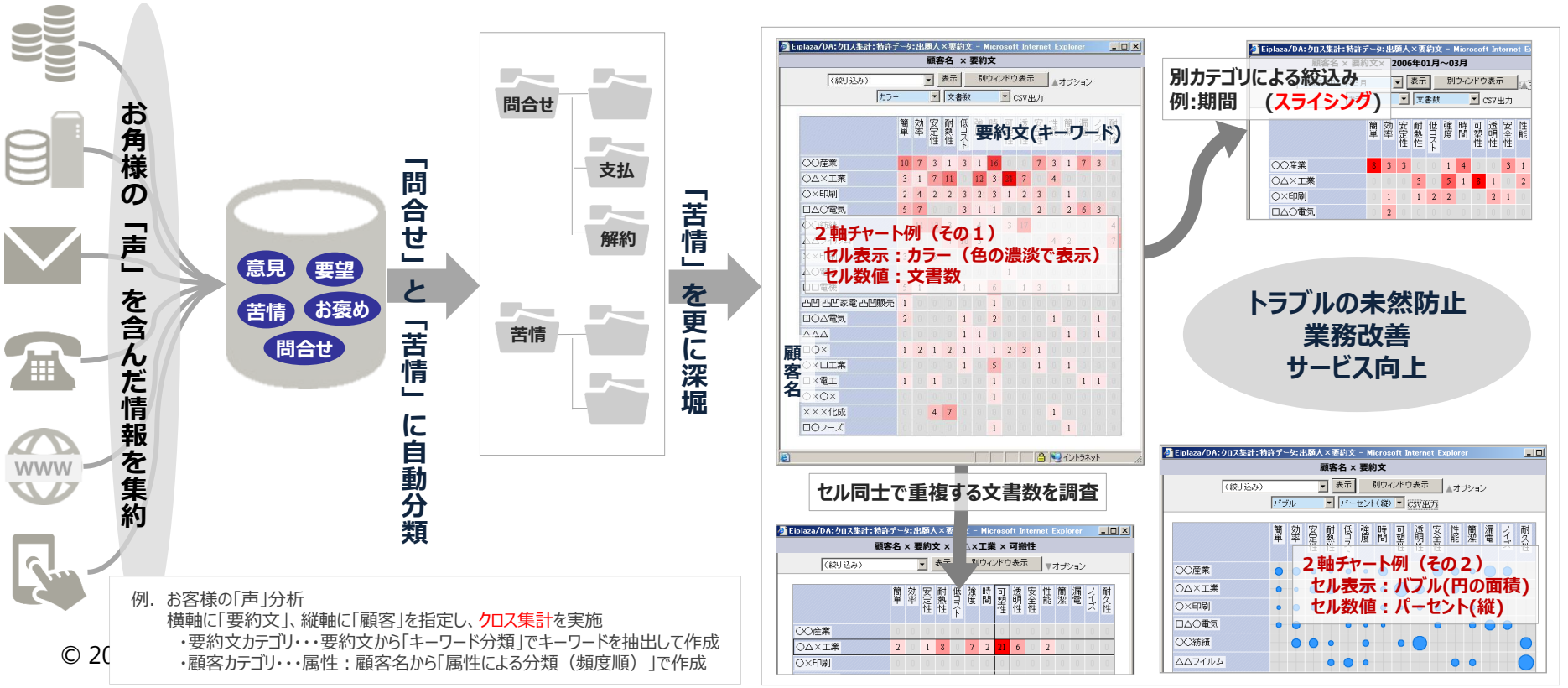



Service

某社様 お客様の声の自動分類

さまざまなお客様の声（メール・電話・Web・SNS…）を自動分類・分析し、顧客満足度の向上を実現

メールや、コンタクトセンターの中に溜まったお客様からの問い合わせやWebやSNSなど、膨大なお客様の声を自動的に分析することで、業務改善につながるキーワードを発見
お客様の「苦情」と発見したキーワードなどを軸として、2軸でマッピングし、顧客の声を分析

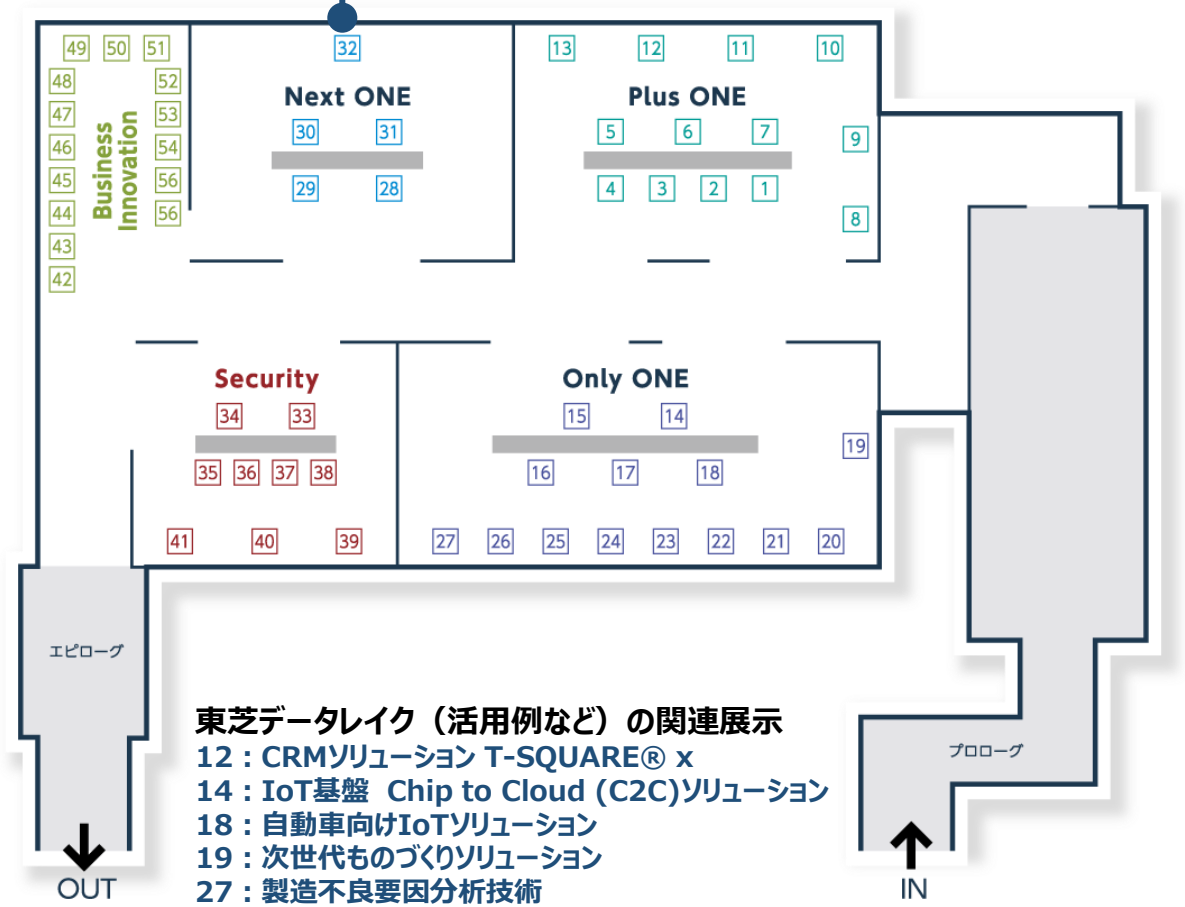


例．お客様の「声」分析
横軸に「要約文」、縦軸に「顧客」を指定し、クロス集計を実施
・要約文カテゴリ・・・要約文から「キーワード分類」でキーワードを抽出して作成
・顧客カテゴリ・・・属性：顧客名から「属性による分類（頻度順）」で作成

© 20


ビッグデータ
×
IoT & IoE
×
リアルタイム
東芝データレイクサービスで
“簡単！”・”賢く！ ”

東芝データレイクサービスは、こちらで展示しております。
是非お立ち寄りください。



東芝データレイク（活用例など）の関連展示

12：CRMソリューション T-SQUARE® x
14：IoT基盤　Chip to Cloud (C2C)ソリューション
18：自動車向けIoTソリューション
19：次世代ものづくりソリューション
27：製造不良要因分析技術
30：ビッグデータ・大規模メディアを高速照会するデータベース技術

TOSHIBA

Leading Innovation