

TOSHIBA

Leading Innovation >>>

データの価値を引き出す 統合ビッグデータソリューション

～センシングデータやソーシャルメディアをスピーディーにビジネスで活用～

東芝ソリューション株式会社

商品・技術統括部 プラットフォーム商品技術部

栗田 雅芳

東芝クラウド&ソリューションフェア 2014

TOSHIBA Cloud & SOLUTION FAIR 2014

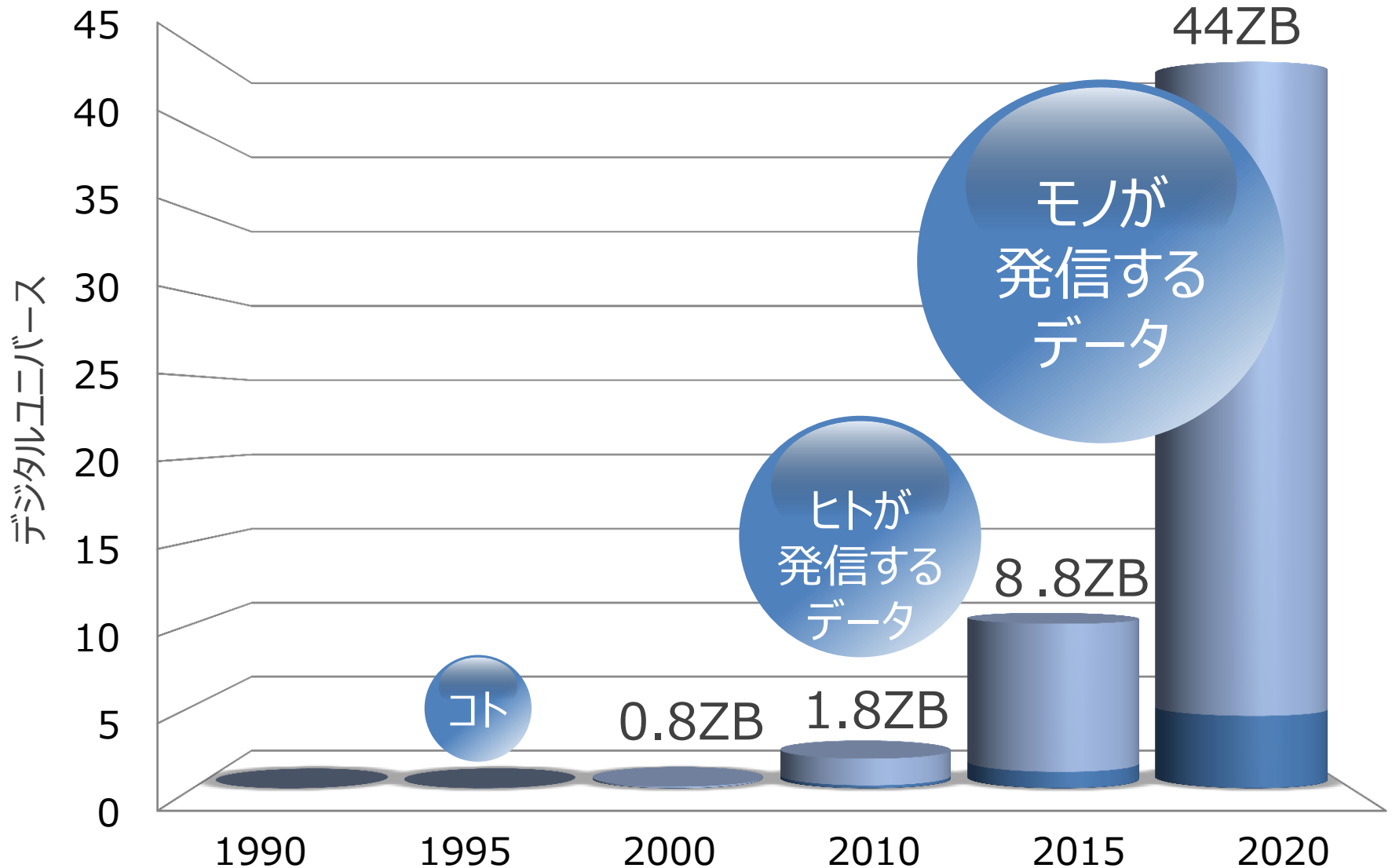


アジェンダ

- **ビッグデータの価値を引き出すには…**
- **弊社の統合ビッグデータソリューション**
- **まとめ**

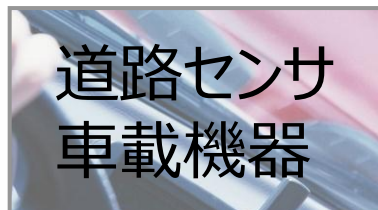
ビッグデータの価値を 引き出すためには…

地球上で生成されるデータの量の推移

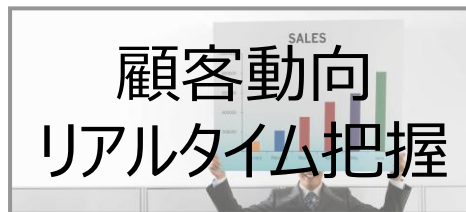
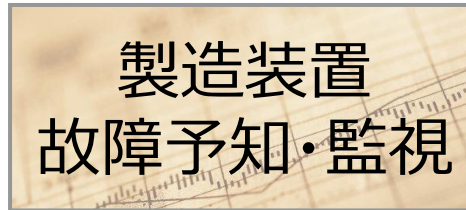


ビッグデータがもたらす新たな価値

実世界データ



ナウキャスト



実現できる世界



スマート家電 家電“コンシェルジュ”サービス

スマート家電×省エネ×ライフログ活用で快適な住空間を

便利

冷蔵庫内をカメラで
チェックして、
重複買いを防ぎます。



快適

室温が高温になると
メールでお知らせ。
外出先からお部屋の
温度がわかります。



省エネ

洗濯機の乾燥フィルターの
詰まりをメールでお知らせ。
お掃除をお願いして
省エネにつなげます。



安心

故障する前に
機器の異常をメールでお知らせ。
万一、故障の場合はサービスセンター
(東芝テクノネットワーク株式会社)に
連絡もできます。



パターンの導出

タイムリーなサービスの提供

今までのデータ利活用

診断的アナリティクス
記述的アナリティクス
勘・経験・度胸

コスト

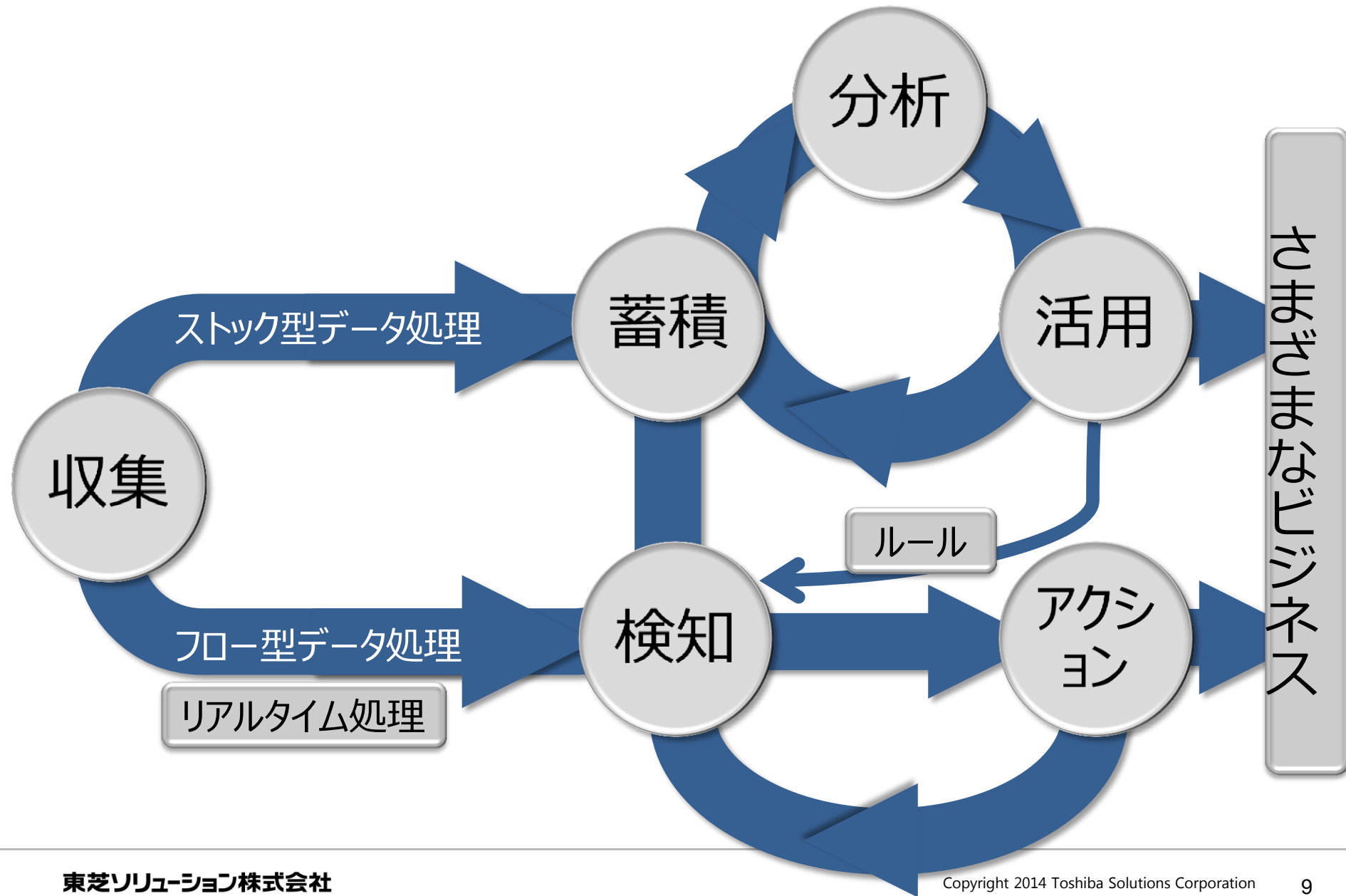
わかる

しる

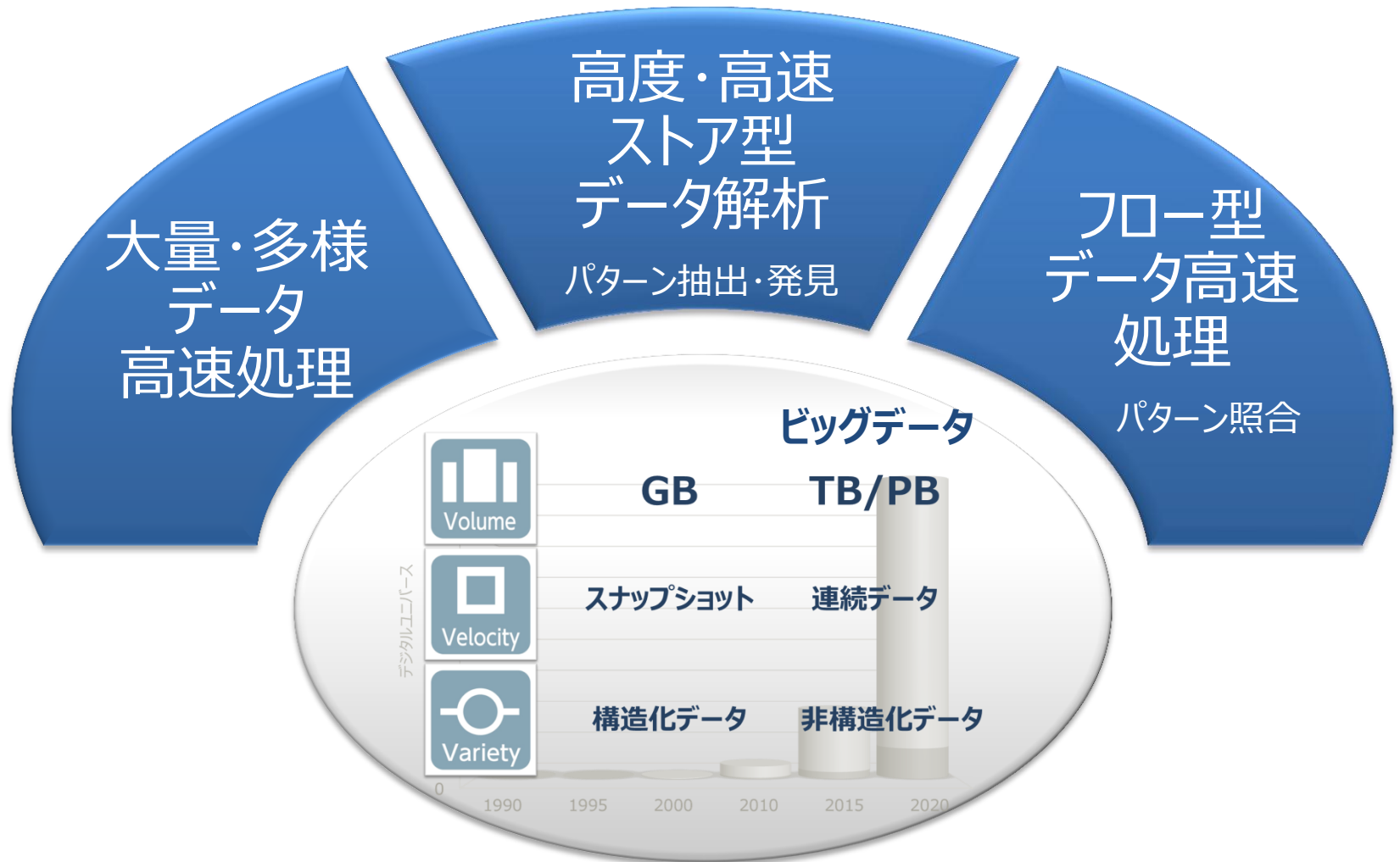
みる



ビッグデータの活用プロセス

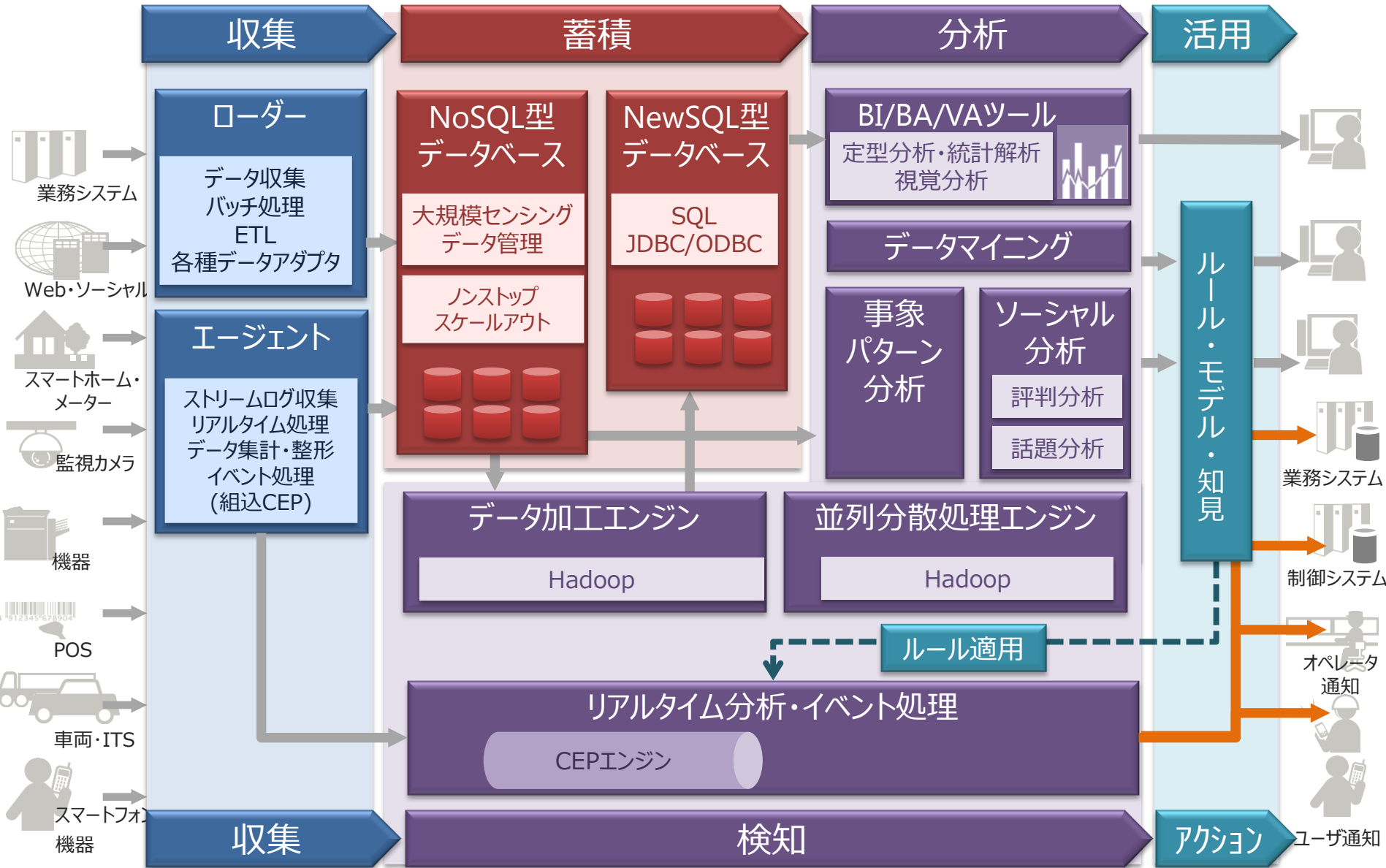


必要なビッグデータ技術

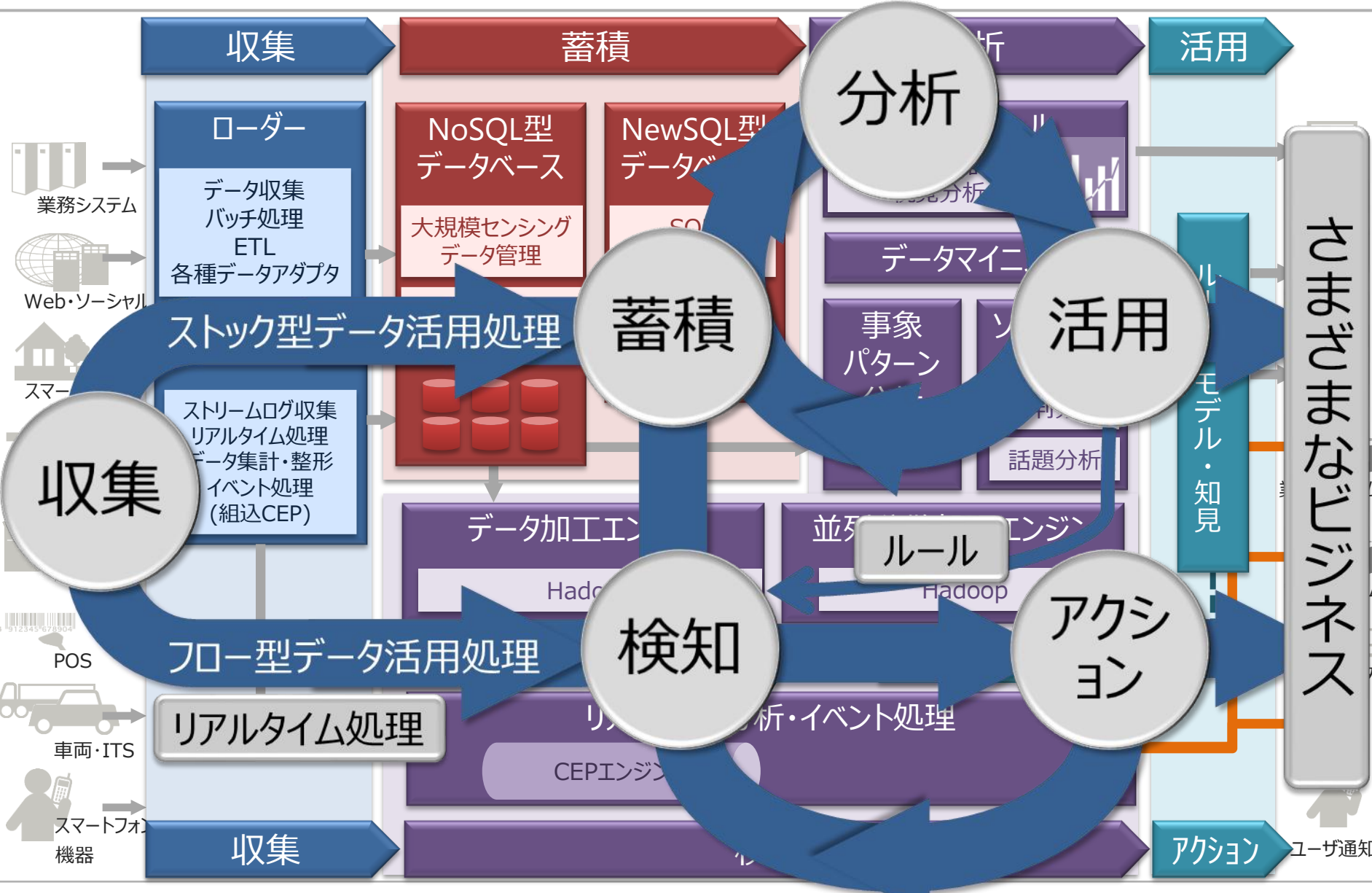


統合ビッグデータソリューション

統合ビッグデータソリューション



統合ビッグデータソリューション



収集
SmartEDA®

ビッグデータでの「収集」の要件

ビジネスデータ

- ☑ 従来型のデータ
- ☑ 定型、スキーマがある
- ☑ 意味づけがされている
- ☑ そのまま利用可能

ソーシャルデータ

- ☑ 主にテキスト
- ☑ 画像、動画、音声など
- ☑ 認識処理が必要

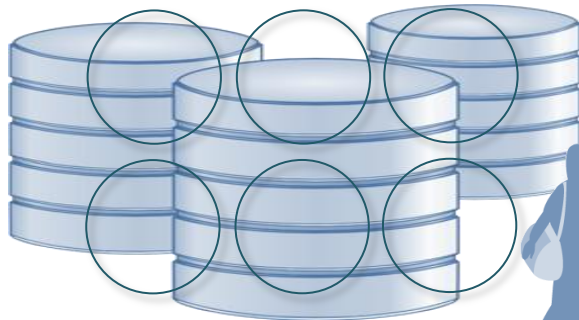
センシングデータ

- ☑ 主に数値データ
- ☑ 時系列に多数並ぶ
- ☑ 値の判定が必要
- ☑ 傾向、パターンの認識が必要

コト

ヒト

モノ



ビッグデータでの「収集」の要件

ビジネスデータ

- ☑ 従来型のデータ
- ☑ 定型、スキーマがある
- ☑ 意味づけがされている
- ☑ そのまま利用可能

ソーシャルデータ

- ☑ 主にテキスト
- ☑ 画像、動画、音声など
- ☑ 認識処理が必要

センシングデータ

- ☑ 主に数値データ
- ☑ 時系列に多数並ぶ
- ☑ 値の判定が必要
- ☑ 傾向、パターンの認識が必要

- 異常は一瞬（見逃してはいけない）
 - 瞬間値ではなく、時系列の波形が意味を持つ場合もある
- 全件サーバーに送信するとネットワークの負担が大

- データの頻度が多い（1秒に100件など）
- デバイスの数が多い（数千デバイス、数万デバイス）
- ほとんどが正常データ（ためる価値が低い）

集めないビッグデータ®

エンドポイント側にエージェントを配置して一次処理のアプローチが必要

SmartEDA

ローダー

- ☑ ストア型データソース：業務データ、RDBMS / CRM / ERP
- ☑ バルク処理（ブロック単位でのデータロード）

エージェント

- ☑ フロー型データソース：センシングデータ / Webログ / アプリログ
- ☑ 「集めないビッグデータ®」を実現するエンドポイント処理
 - ☑ データの集計・整形：ミリ秒単位の集計/分析に適した粒度への変換
 - ☑ イベント処理：端末側での異常検知
 - ☑ ネットワーク側の負担軽減：正常データ定期送信/異常データ即時送信

蓄積
GridStore®

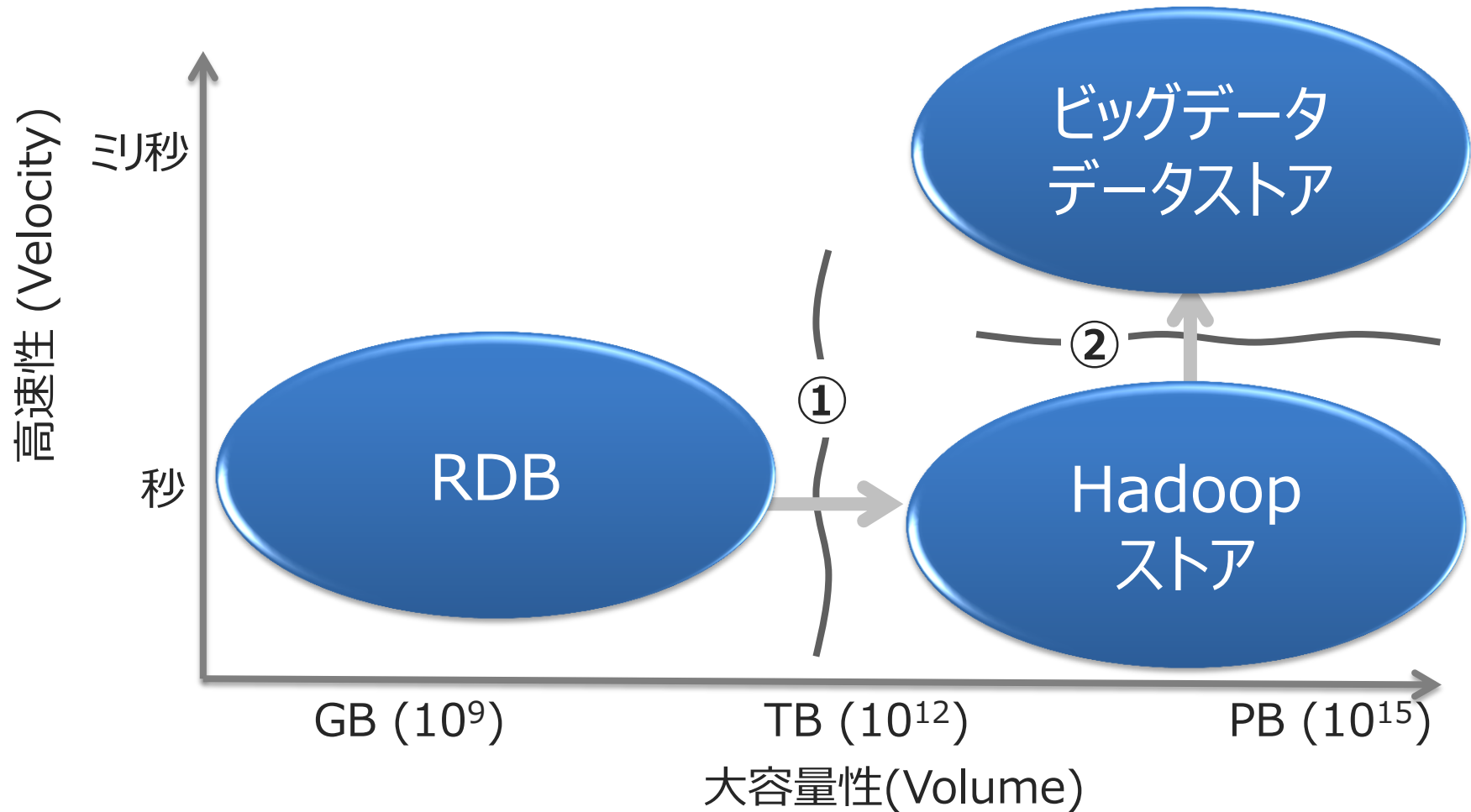
ビッグデータを支える「蓄積」の要件

「将来、データがどれだけ集まるかわからないこと」



スケールアウトによる大容量化と高速化

既存技術では越えられない壁



① RDBでは越えられない大容量化の壁

② Hadoopストアでは越えられない高速化の壁

GridStore

高スケーラビリティ

- ☑ 弊社独自の自律的データ再配置とノンストップスケールアウト

高パフォーマンス

- ☑ インメモリ指向メモリ管理によるオーバヘッド最小化

高可用性

- ☑ 弊社独自のSPOFなしハイブリットクラスタと高速レプリケーション

非構造化データ

- ☑ 時系列コンテナ、データ圧縮機能、データ期限解放機能
- ☑ 2D・3Dデータ型、空間データ索引機能

SQLサポート (ODBC/JDBC)

GridStore

高スケーラビリティ

- ☑ 弊社独自の自律的データ再配置とノンストップスケールアウト

高パフォーマンス

- ☑ インメモリ指向メモリ管理によるオーバヘッド最小化

高可用性

- ☑ 弊社独自のSPOFなしハイブリットクラスタと高速レプリケーション

非構造化データ

- ☑ 時系列コンテナ、データ圧縮機能、データ期限解放機能
- ☑ 2D・3Dデータ型、空間データ索引機能

SQLサポート (ODBC/JDBC)

事例：エネルギー情報管理システム

家庭・マンション・ビル等施設の各種メーターデータの見える化を低コストで実現

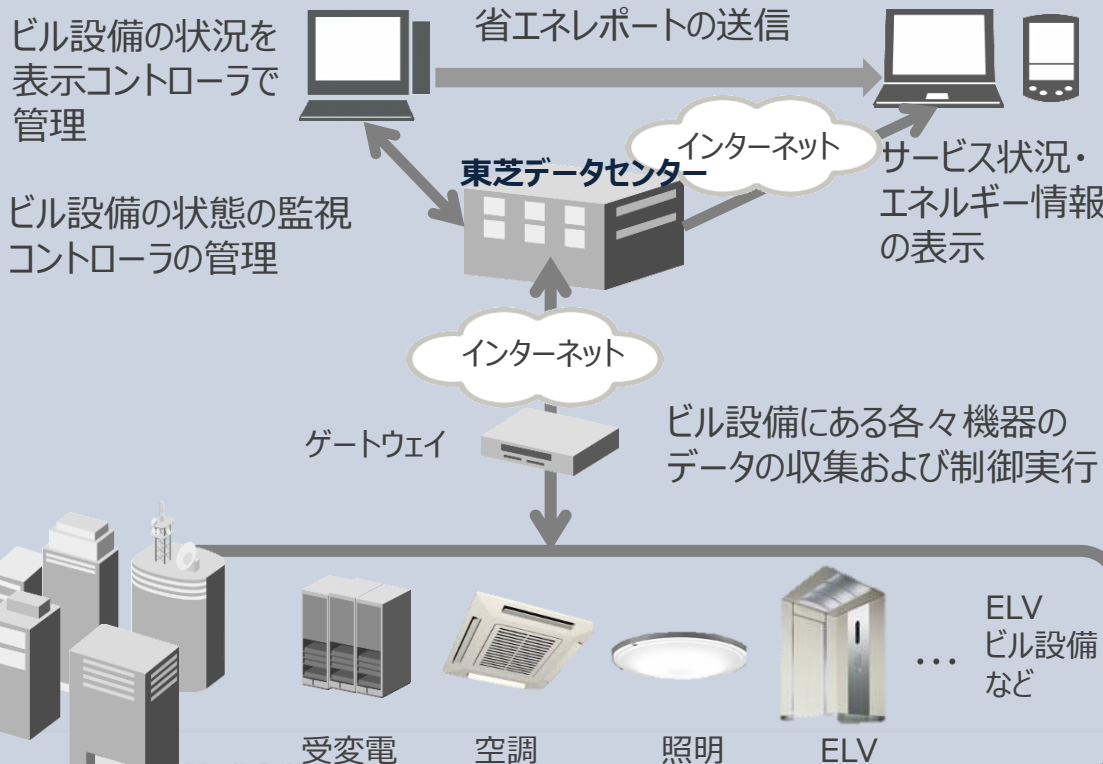
機能概要

- 施設の各種メーターデータの収集・保存
- 施設ごとのエネルギー情報の表示
- 複数の施設を跨いだエネルギー情報の表示

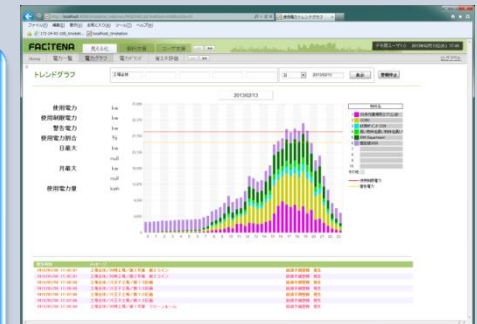
導入効果

- 大量メーターデータの低コストの蓄積・管理の実現
- 多数の施設のデータの一括管理と施設増設時の安定処理の継続

システム概念図



見える化画面イメージ



電力消費量

設備の状況

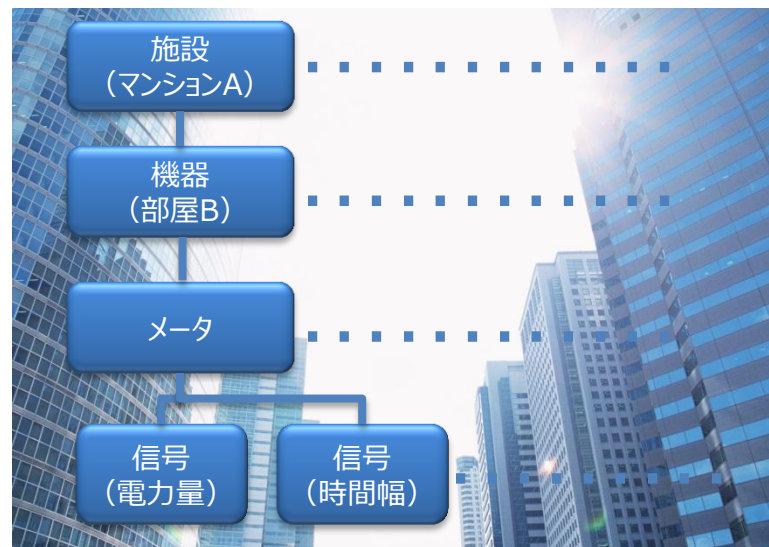
工場全体	電力	ガス	水	CO2
1,400	206	299	542	

設備の状況

エネルギー情報管理システムでの実証データ

管理する規模

項目	個数
施設数	10,000
メータ数	500,000
設備数/施設	1
メータ数/設備	50
測定項目数/メータ	2 (電力量、時間幅)
短周期値収集間隔	1分
積上の種別	最大、最小、合計
積上区間の種別	時、日、月、年



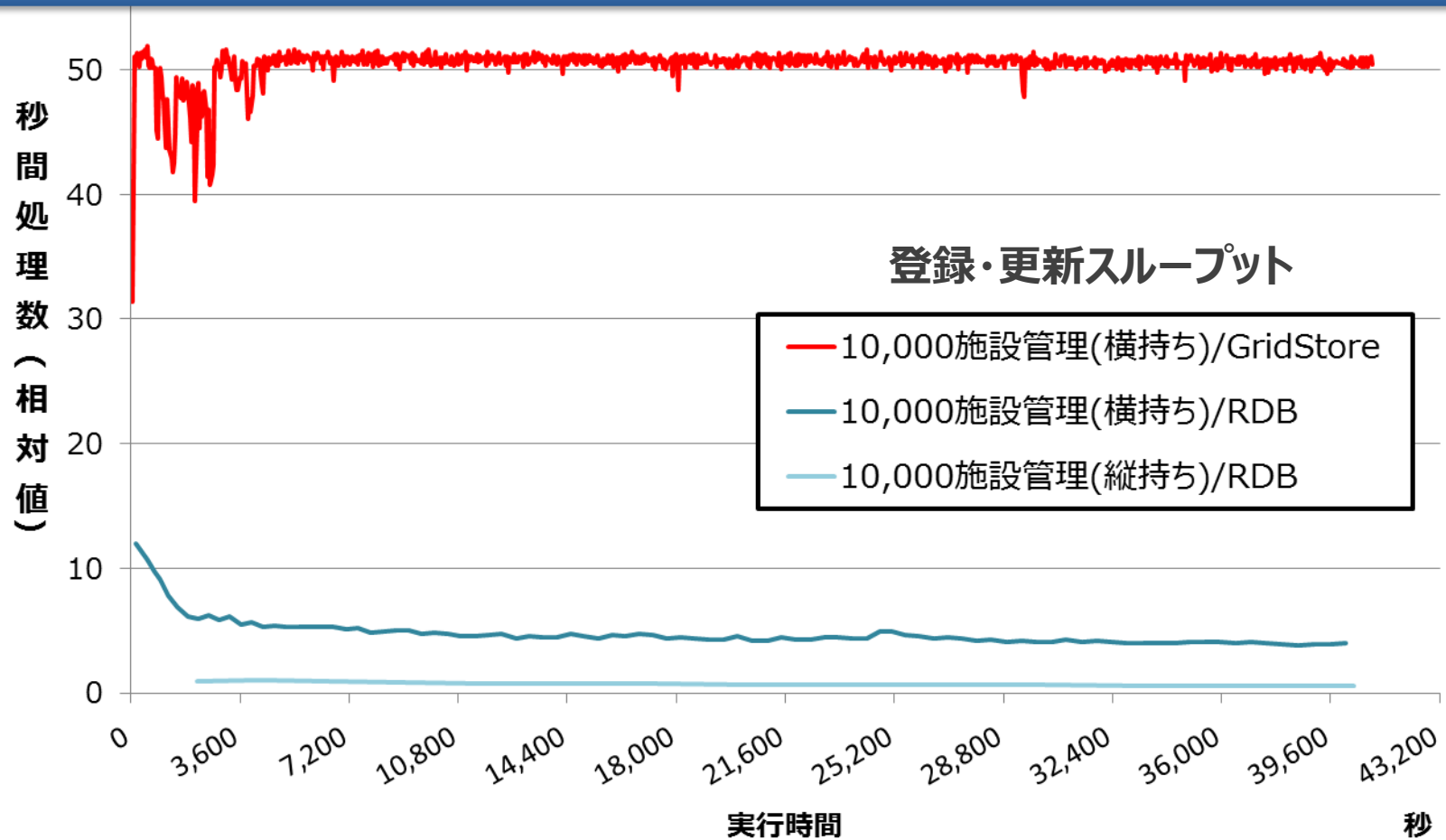
実証データ

時刻	施設名	機器名	メータ名	信号名	信号値
2014/2/20 13:00:00.1234	施設 1	機器 1	メータ1	信号1 (電力量)	0.01
2014/2/20 13:00:00.1234	施設 1	機器 1	メータ1	信号2 (時間幅)	60.0
2014/2/20 13:00:00.1234	施設 1	機器 1	メータ2	信号1 (電力量)	100
2014/2/20 13:00:00.2345	施設 1	機器 1	メータ2	信号2 (時間幅)	59.5
2014/2/20 13:00:00.2345	施設 1	機器 1	メータ3	信号1 (電力量)	250
2014/2/20 13:00:00.2345	施設 1	機器 2	メータ3	信号2 (時間幅)	60.5
2014/2/20 13:01:00.0123	施設 1	機器 2	メータ1	信号1 (電力量)	0.01
2014/2/20 13:01:00.0123	施設 1	機器 2	メータ1	信号2 (時間幅)	60.2
2014/2/20 13:01:00.0123	施設 1	機器 2	メータ2	信号1 (電力量)	1000
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮

信号値
(短周期値)

高い基礎性能の実現

RDBに対して10倍以上のスループットを達成！！

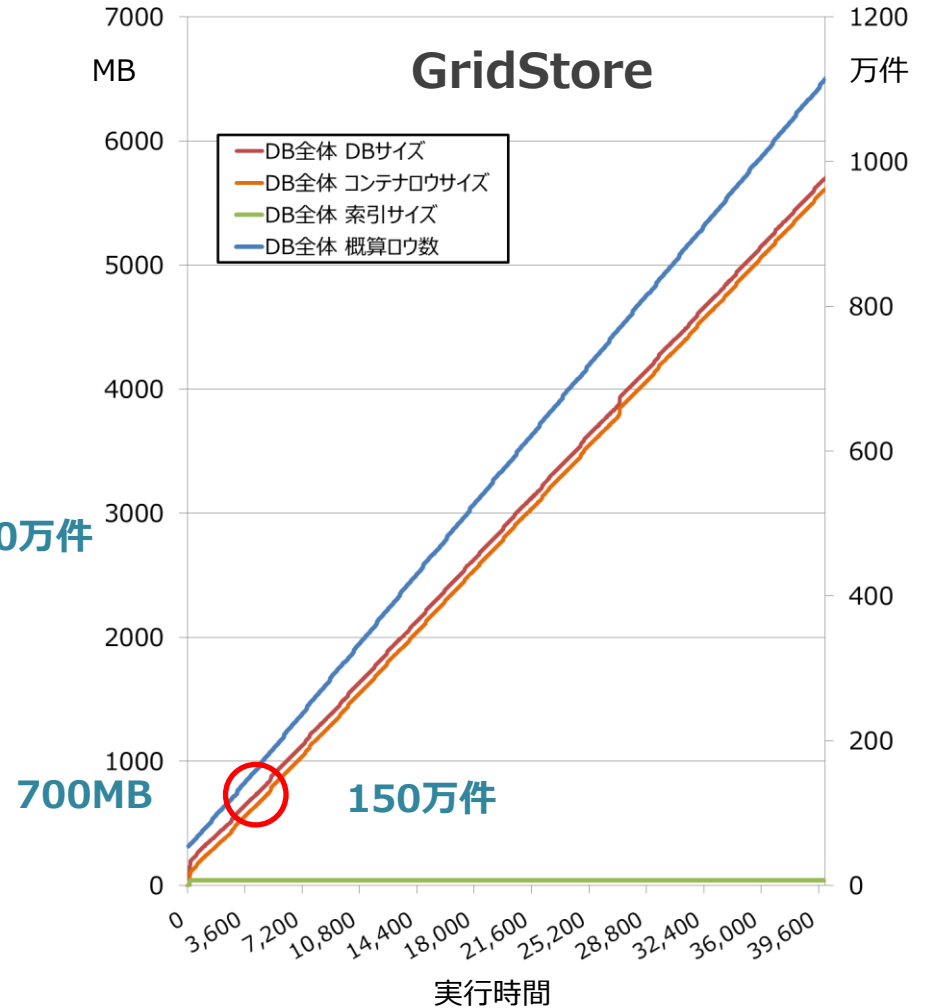
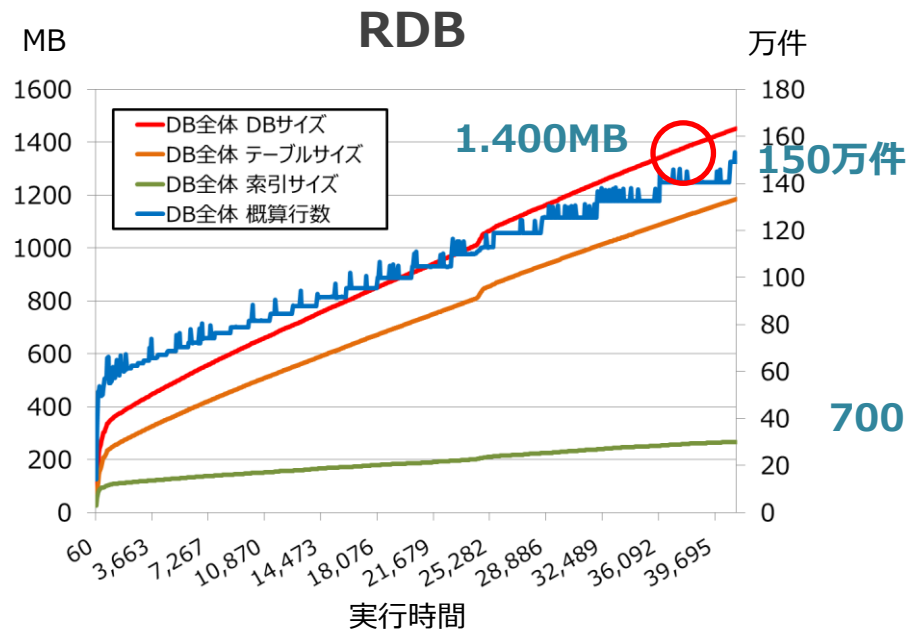


※RDBの10,000施設処理ケースでの最大値を1とした場合の相対値
横持ち：1時刻で機器毎のメータをまとめて（50メータで）1レコード作成
縦持ち：1時刻で1メータごとに1レコード作成

効率的なデータ格納

DBサイズがRDBの50%程度で効率的にデータ格納

	行数・件数	DBサイズ
RDB	150万	約1.400MB
GridStore	150万	約700MB



分析 GridData®

ビッグデータの「分析」の要件

	記述的/診断的 アナリティクス	予測的/処方的 アナリティクス	
対象データの 範囲	数日～数年 サンプリング	数日～数年 全数	直近 全数
分析目的	全体感の把握	パターンの抽出	パターンとの照合
分析タイミング (頻度)	バッチ/アドホック (逐次)	バッチ/アドホック (逐次)	リアルタイム (常時実行)
データ処理	ストック型	ストック型	フロー型
データの蓄積	ストレージ	ストレージ	メモリ→破棄
活用例	<ul style="list-style-type: none"> • 店舗別売上の前年同月比 • 商品Aと商品Bの同時購買傾向 	<ul style="list-style-type: none"> • 不正値・異常値のリアルタイムな検知 • 在庫連動タイムセール通知 	

GridData

分析エンジン・ツール

- ☑ 事象パターン分析：モノ・ヒトの特定の事象に至るパターンの抽出
- ☑ ソーシャルメディア分析：ヒトが発信する“生の声”で、隠れた事実の発見

BI/BAツール連携

- ☑ ビッグデータ加工・取り込みによる各種BI（定型分析）/BA（統計分析）ツールによる分析

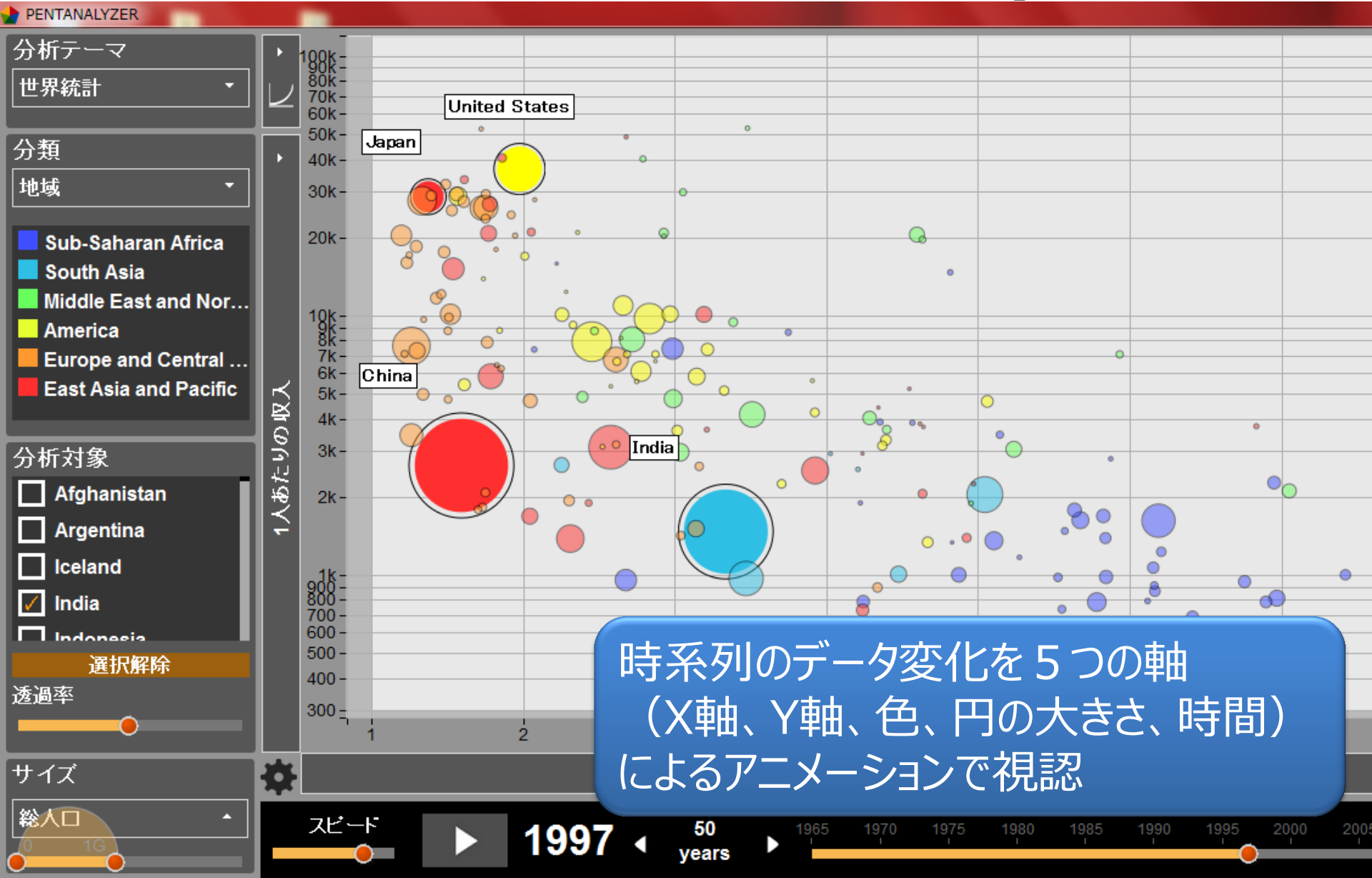
VA（視覚分析）

- ☑ ビッグデータから「きづき」を引き出す視覚化による分析

Hadoopソリューション

- ☑ HDFSを經由せず高速なMapReduce処理を実現した並列分散処理エンジンとエコシステム

視覚分析 (VA: Visual Analytics)



検知 SmartEDA®

ビッグデータを支える「検知」の要件

刻々と発生するイベントから直近のリアルをつかむ

現在

過去



■ 予兆検知



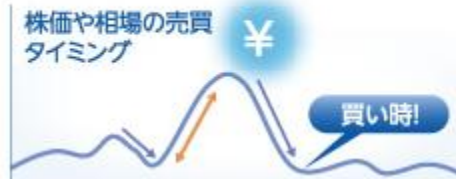
■ 3シグマ手法



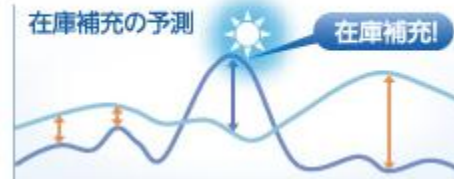
■ 組み合わせ検知



■ 変化パターン検知



■ 相関関係



■ キーワードの出現ランキング



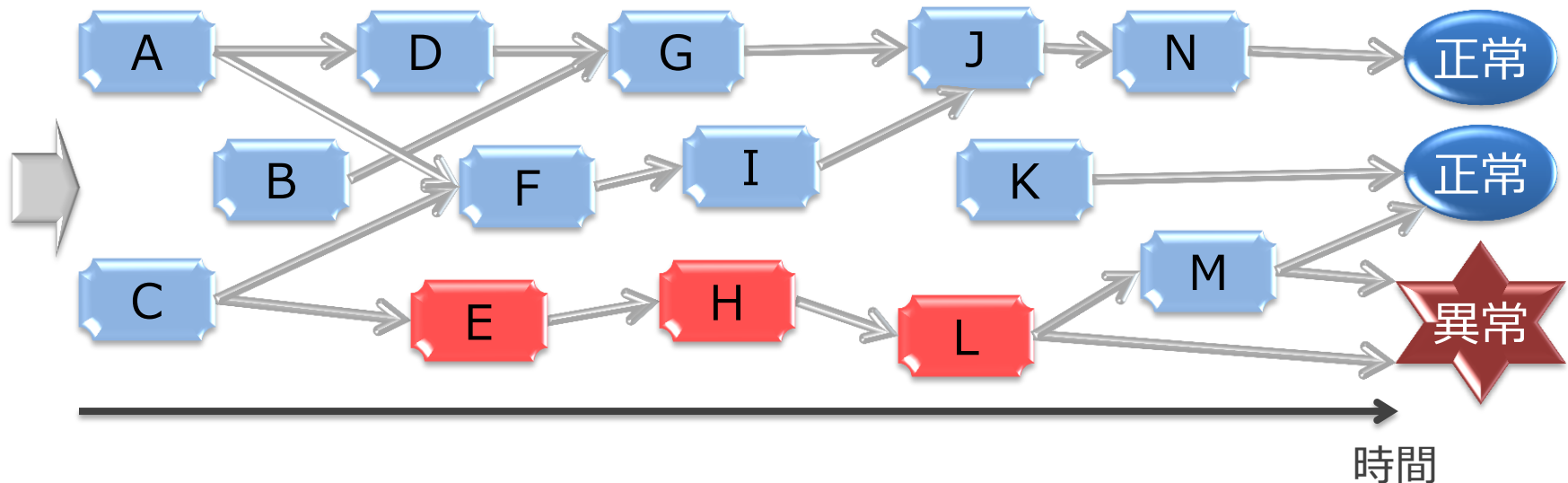
事象パターン分析 GridData[®] / Analytics

事象パターン分析

さまざまな事象データから特定の事象（出来事）の発生につながる確率の高いパターンを自動的に発見

ある出来事（異常）に至る予兆のパターンを、過去データから見つけ出す

異なる種類の
時系列データ



抽出パターン例：**E** → **H** → **L** が異常につながる確率82%

「故障」に至る予兆パターンの抽出

時刻	稼働ログ		センサーデータ		
	ユーザ操作/ コマンド	警告/ エラー	位置ずれ センサー	振動 センサー	温度 センサー
13711710350460			なし	なし	低
13711710661870	操作1		微小	微小	低
13711710661871			微小	微小	低
13711710850460	操作2		小	微小	低
13711710850780			小	微小	低
13711710851090	コマンド1		微小	微小	低
13711710851250			微小	微小	中
13711710853900		警告A	中	中	中
13711710853902	操作1		中	大	高
13711710854530			大		
13711710858750	コマンド3		大		
13711710867810		エラーA	大		
13711710870460			大		

機器1

時刻	センサーデータ				
	ユーザ操作/ コマンド	警告/ エラー	位置ずれ センサー	振動 センサー	温度 センサー
13711710661871	コマンド5	警告B	中	なし	低
13711710661872			中	大	高
13711710661873	操作2		なし	大	高
13711710661874			なし	中	高
13711710848120	コマンド2		なし	中	高
13711710851090			なし	中	低
13711710851250	コマンド1		なし	中	低
13711710851251			なし	小	低
13711710853900	コマンド6		なし	小	低
13711710854530			大	小	低
13711710858750	コマンド5		大	大	高
13711710867810			大	大	高
13711710870460		エラーA	大	大	高

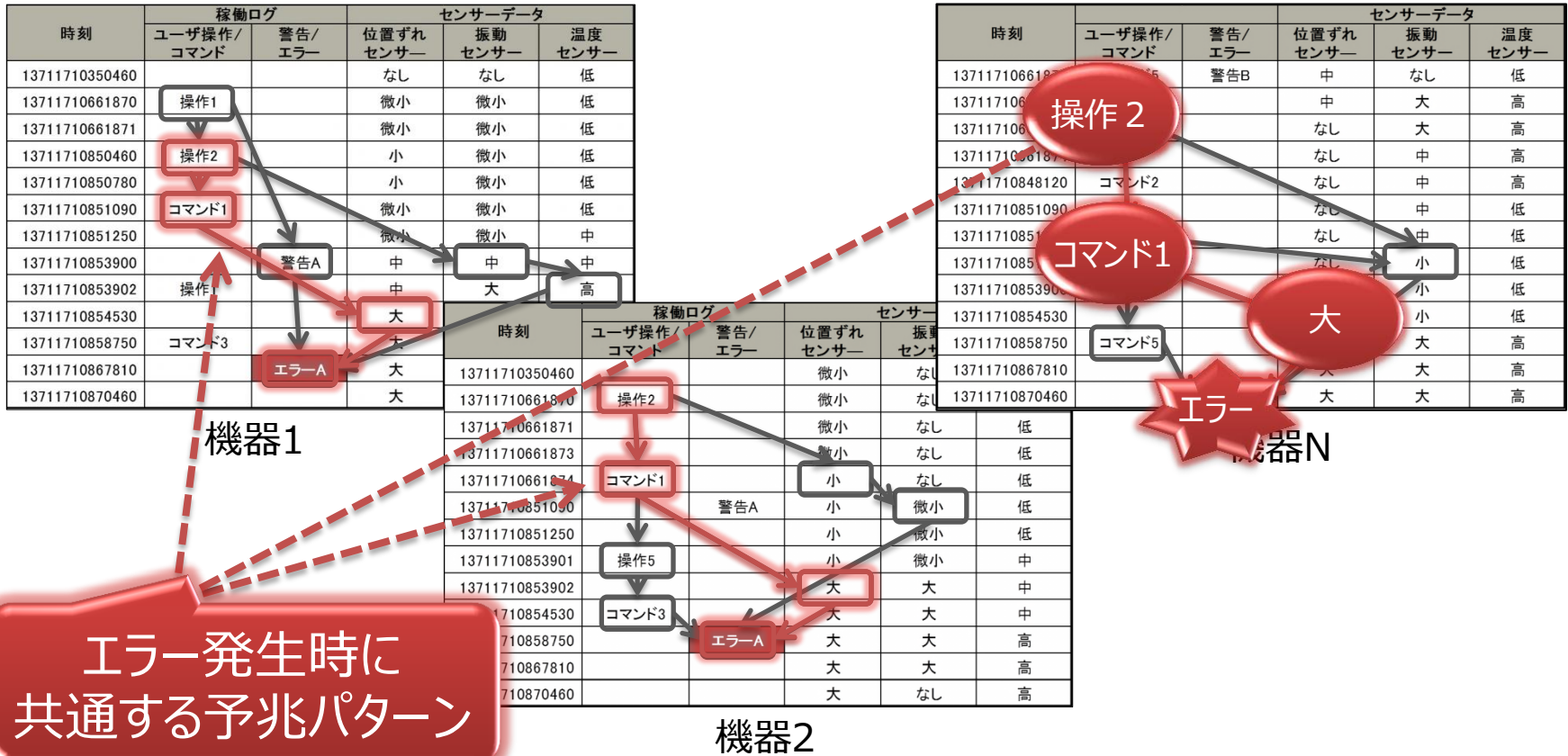
機器N

時刻	稼働ログ		センサー	
	ユーザ操作/ コマンド	警告/ エラー	位置ずれ センサー	振動 センサー
13711710350460			微小	なし
13711710661870	操作2		微小	なし
13711710661871			微小	なし
13711710661873			微小	なし
13711710661874	コマンド1		小	なし
13711710851090		警告A	小	微小
13711710851250			小	微小
13711710853901	操作5		小	微小
13711710853902			大	大
13711710854530	コマンド3		大	大
13711710858750		エラーA	大	大
13711710867810			大	大
13711710870460			大	なし

機器2

何が故障（エラーA）の予兆か分からない

「故障」に至る予兆パターンの抽出

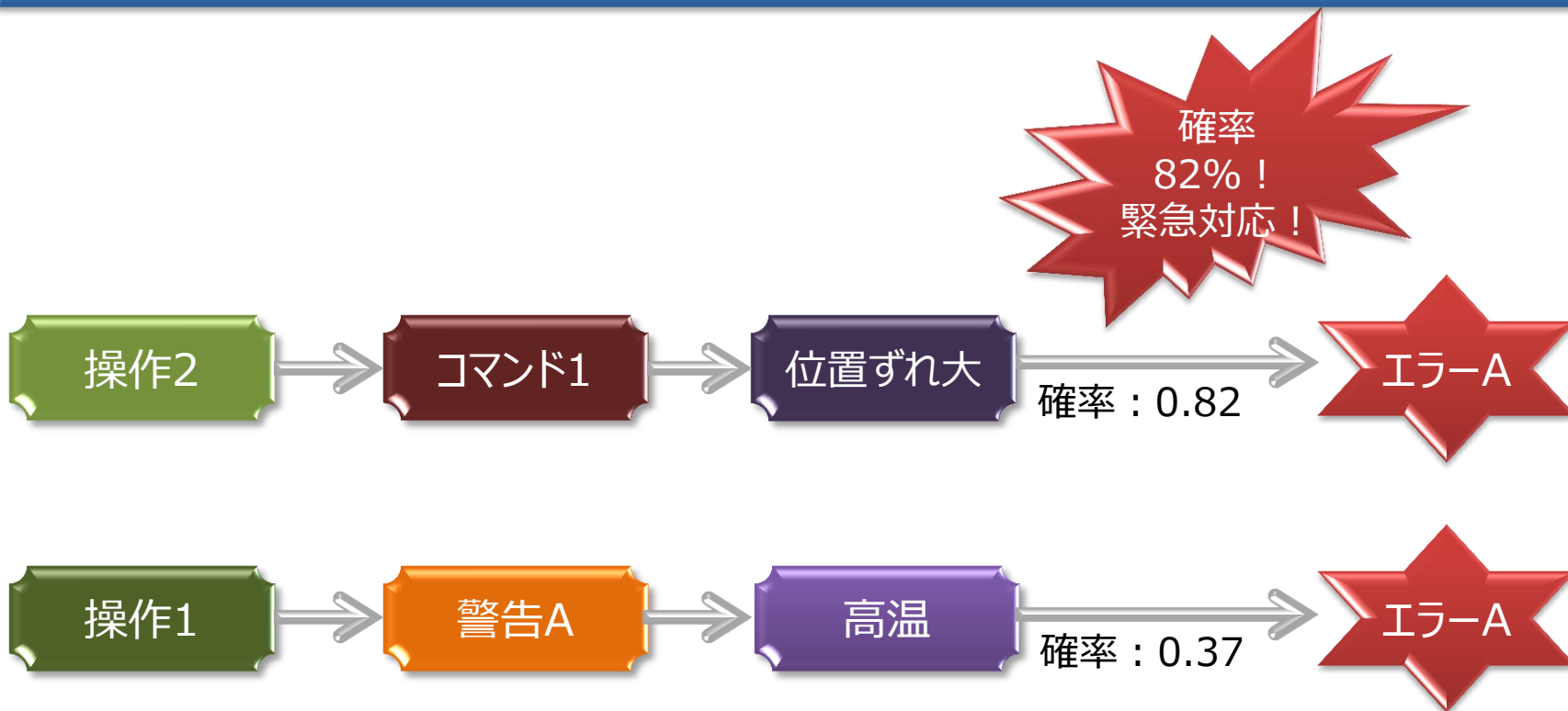


現象のパターン（予兆パターン）を導き出すことができる

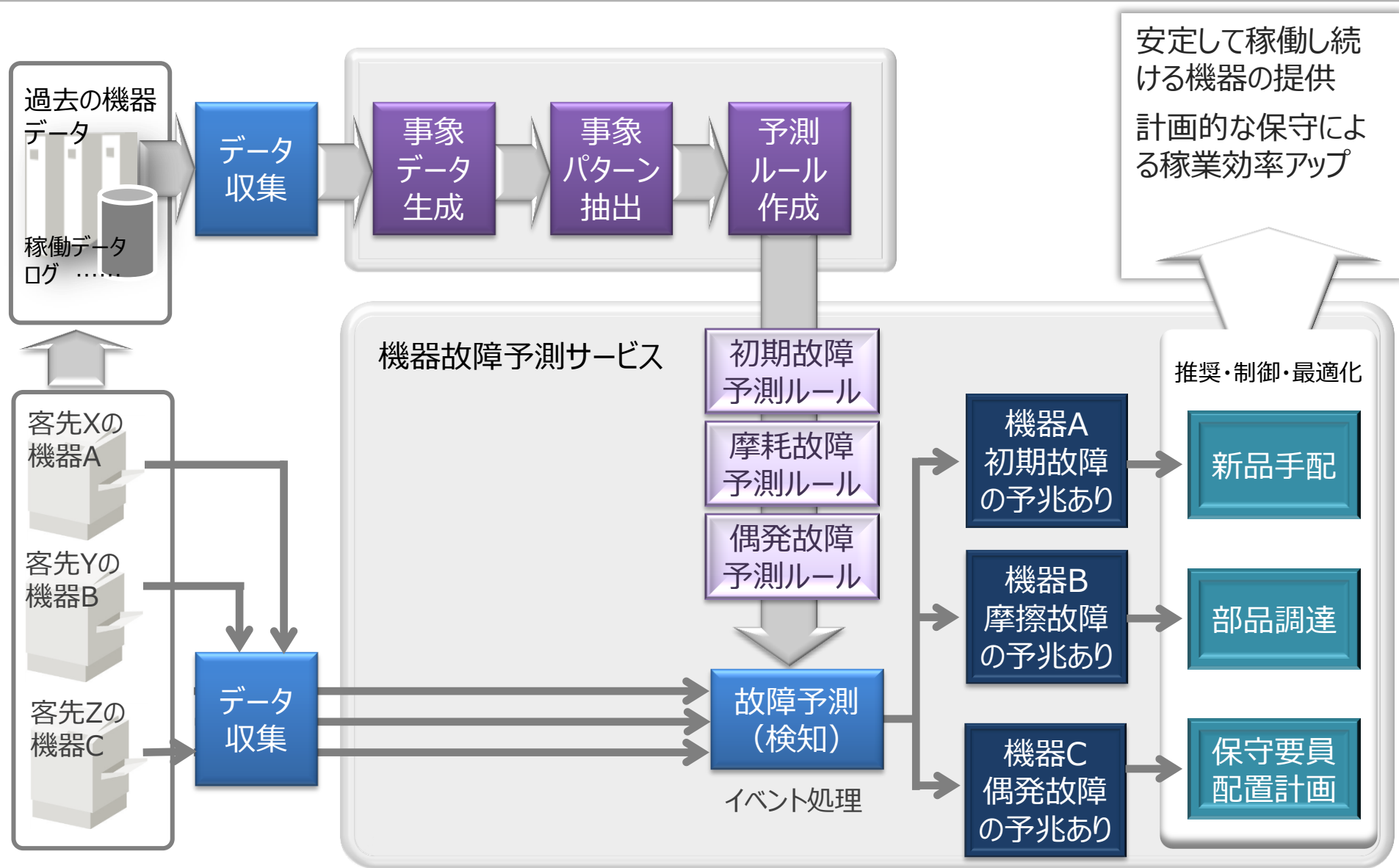
故障予測への適用

特定の事象（故障）に至る確率を算出

確率の高いパターンを検知ルールに採用



機器の故障予測への適用事例



特長：事象の発生時期まで予測

類似技術：
時系列アソシエーション分析

事象の順序関係のみ抽出可能

× いつ事象が発生するか予測ができない

事象パターン分析

事象間の時間間隔も含めたパターンも抽出可能

◎ いつ事象が発生するか予測ができる



特長：高速な事象パターン抽出

類似技術： 時系列アソシエーション分析

全ての現象系列を分析し、頻出するパターンを抽出する方法



予測に関係がない大量の現象系列パターンが抽出されてしまい、分析に時間がかかる

時刻	稼働ログ		センサーデータ		
	ユーザ操作/コマンド	警告/エラー	位置ずれセンサー	振動センサー	温度センサー
13711710350460			なし	なし	低
13711710661870	操作1		微小	微小	低
13711710661871			微小	微小	低
13711710850460	操作2		小	微小	低
13711710850780			小	微小	低
13711710851090	コマンド1		微小	微小	低
13711710851250			微小	微小	中
13711710853900		警告A	中	中	中
13711710853902	操作1		中	大	高
13711710854530			大	大	高
13711710858750	コマンド3		大	大	高
13711710867810		エラーA	大	大	高
13711710870460			大	なし	高

色々な事象に至る系列が出てきてしまう

事象パターン分析

事象を指定し、そこから遡って、頻出するパターンを抽出する方法



分析する対象を限定することができ、高速なパターン抽出が可能

時刻	稼働ログ		センサーデータ		
	ユーザ操作/コマンド	警告/エラー	位置ずれセンサー	振動センサー	温度センサー
13711710350460			なし	なし	低
13711710661870	操作1		微小	微小	低
13711710661871			微小	微小	低
13711710850460	操作2		小	微小	低
13711710850780			小	微小	低
13711710851090	コマンド1		微小	微小	低
13711710851250			微小	微小	中
13711710853900		警告A	中	中	中
13711710853902	操作1		中	大	高
13711710854530			大	大	高
13711710858750	コマンド3		大	大	高
13711710867810		エラーA	大	大	高
13711710870460			大	なし	高

「エラーA」に至る現象系列のみ抽出



事象データ
作成

事象パターン
抽出

生データ

事象化済
データ

パターン
(ルール)

イベント監視モニタ

データ監視モニタ

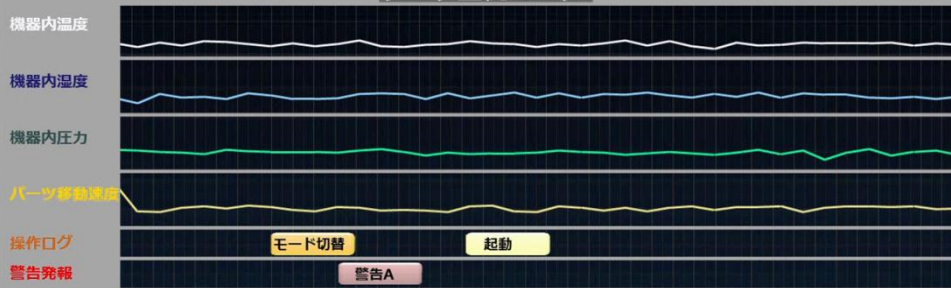
イベント検知モニタ

イベント検知モニタ



データ監視モニタ

データ監視モニタ



ソーシャルメディア分析 GridData[®] / Analytics

ソーシャルメディア分析

種類	機能概要
時系列分析	形態素解析結果（単語に分割）からキーワードの集計などの時系列での分析
ターゲット分析	指定した検索ワードを含む内容の発信者の属性別（年代別/性別）や地域別の分析
影響力分析	拡散規模など影響力の度合いの分析
評判分析 （ポジ・ネガ分析）	発信内容をポジティブ・ネガティブの観点での評価・カウントの分析
話題分析 （共起語分析）	頻出する共起語などによる話題の抽出・分析
重要表現分析	重要な発言（事故、法令違反、健康被害など）を自動抽出・分析

特長：ヒトの生の声から思い・関心の把握

ヒトが発信する「生の声」を収集・高精度に分析

- ☑ 長年にわたり、東芝の機械翻訳などで培った自然言語意味解析技術で高精度に分析
- ☑ 業種・業界の専門用語辞書（約30種類）で、対象分野に合わせた適切な解析

ヒトの「思い」を把握

- ☑ 発言内容や文書全体から好評・不評の判定

隠れた情報からヒトの「関心」を把握

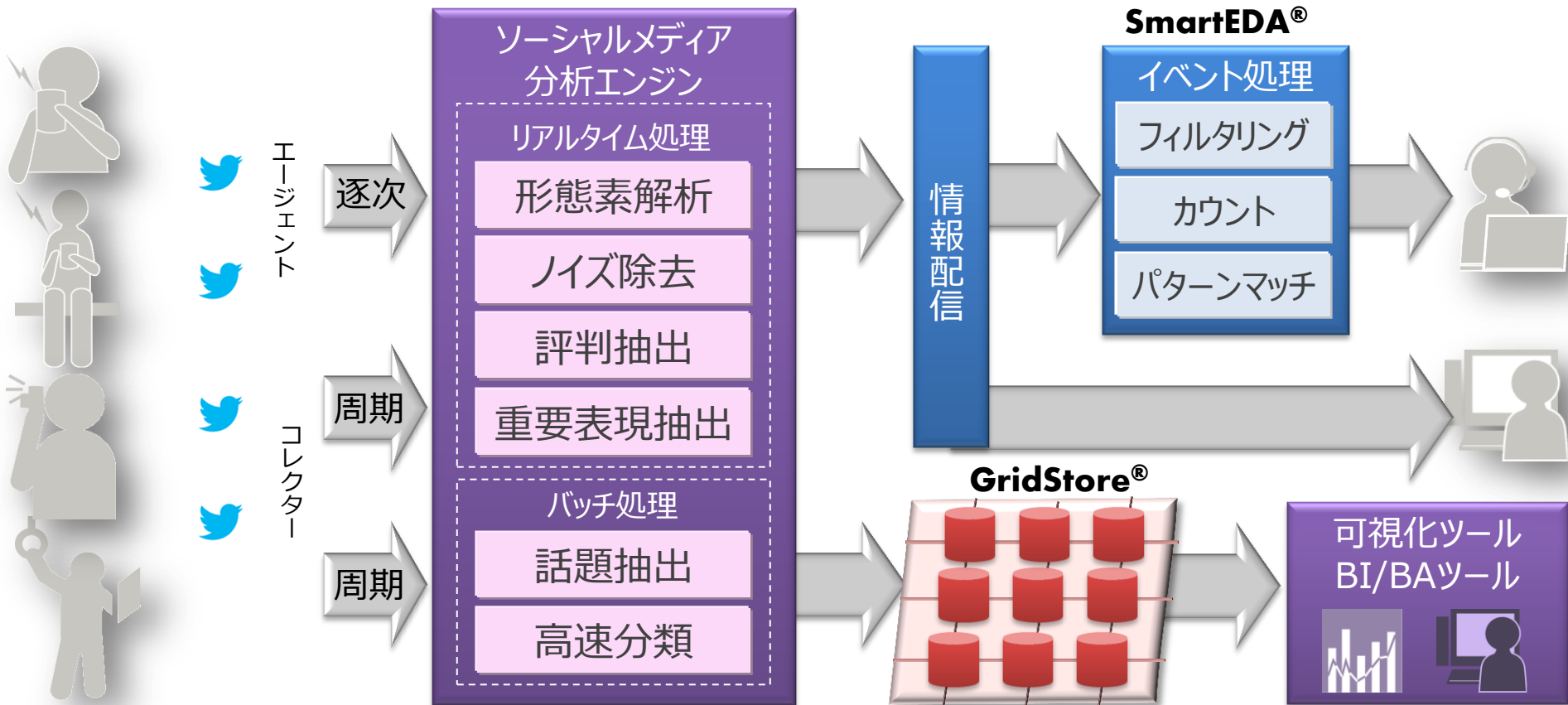
- ☑ キーワードを用いることなく話題を自動で抽出・分類
- ☑ 想定していない世の中の隠れた話題を見つけ出すこと可能

重要表現抽出により素早く検知

- ☑ 重要な発言（事故、健康被害など）を抽出し、イベント処理と組み合わせることで、苦情への早期対応やリスクの回避が可能

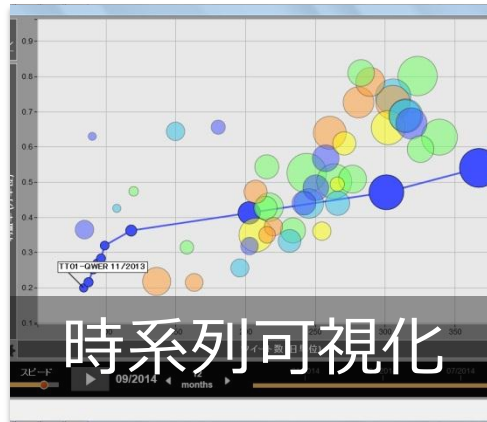
特長：リアルタイム処理で素早く活用

ヒトの生の声をリアルタイムに把握し、
タイムリーに効果的な対応を可能とする



ビッグデータ利活用フレームワーク

ビッグデータのパワーをスピーディーにビジネスで活用！



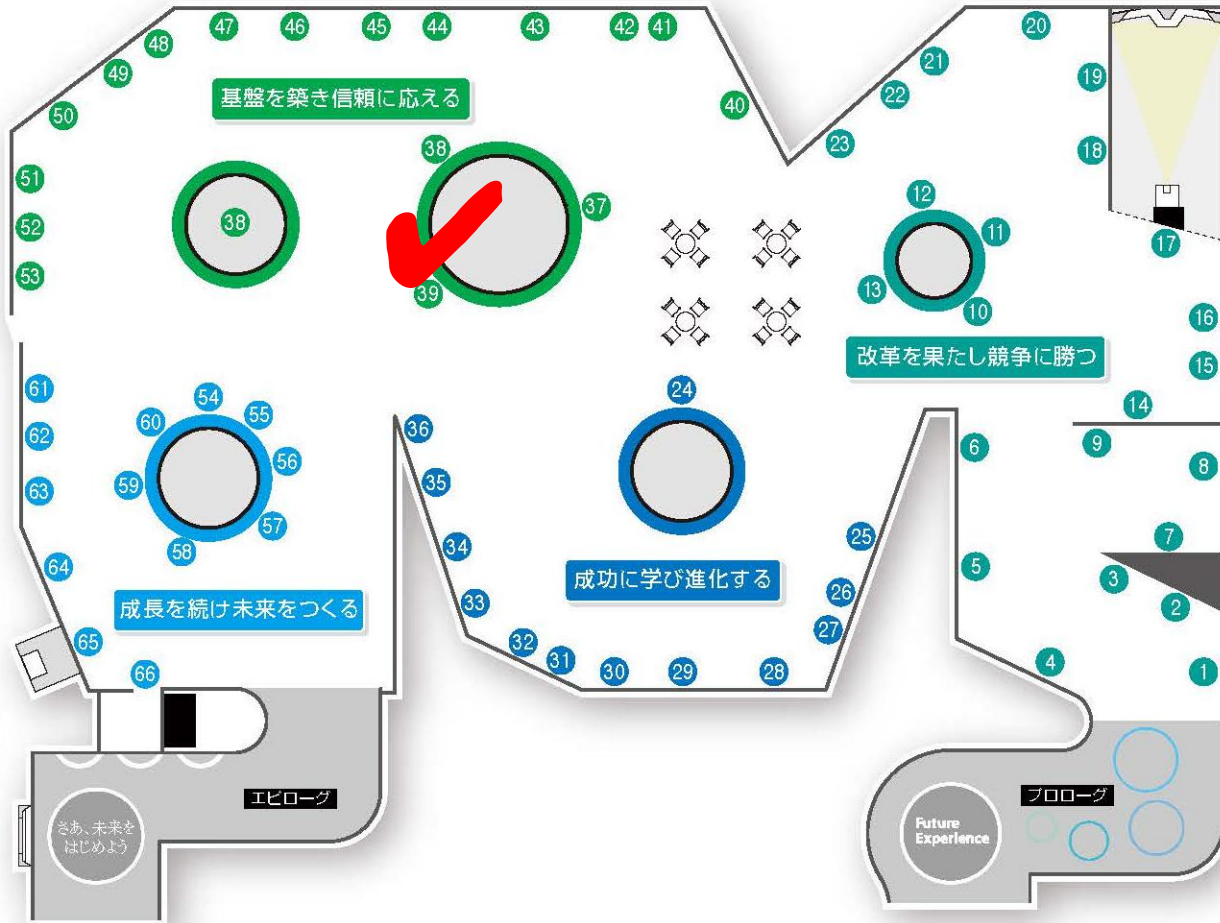
まとめ

まとめ

- ☑ ヒト・モノが発信するデータから一歩先を読んで手を打つことで新たな価値を創出
- ☑ 「みる・しる・わかる」と「勘・経験・度胸」の判断から「わかる・みぬく・いかす」と「ルール」による判断
- ☑ データの蓄積・分析・活用（ルール化）とルールによる検知とアクションのサイクル化が重要

データの収集・蓄積・分析・検知、そしてビジネスでの活用まで包括的なソリューションの提供

展示のご案内



ご静聴ありがとうございました



安心、安全、快適な社会。

Human Smart Community

by lifenology - the technology life requires

TOSHIBA

Leading Innovation >>>