

7月11日(木) A-3: 16:35~17:20

**実業とITの融合で、
今そこにある課題に答える！**

日立のビッグデータ活用基盤

株式会社 日立製作所 情報・通信システム社
ソフトウェア開発本部 DB設計部
主任技師 石川 太一

**Human Dreams.
Make IT Real.**



はじめに

ビッグデータを活用して、どんな価値を生み出したいですか？

真実を伝えられないサマリデータ、
収集できず切り捨てられた詳細データ、
膨大に溜まったまま活用できない履歴データ・・・。
ごみの山か、宝の山か？

データが発生し、消滅するまでのライフサイクルに注目すると、
価値を生み出せる瞬間が、そして今必要な技術が見えてきます。

今、目の前にある課題を素早く解決し、
新たな価値を創出するために。

実業とITを手掛ける、日立のビッグデータ活用基盤を、
ご紹介します。

ビッグデータは、お題目なのか？

「**ビッグデータ**」という言葉だけが
大きく取り上げられますが、

そもそもベンダがモノを売りこむために作った、
「**お題目**」なのでは？

「ビッグデータ」は、お題目か？

いいえ。これは、
正しい判断をするための、解決手段の一つなのです。

そもそも経営方針や、営業戦略、開発合議など、
なにを根拠にGo/NoGoを判断していますでしょうか？

実はまだ「**KKD**」に頼る、という場面も少なくない。

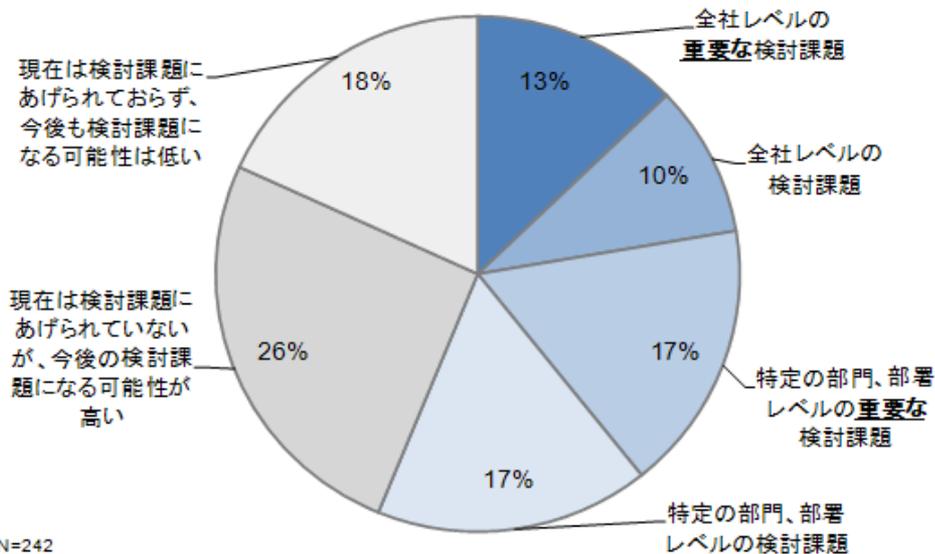
一方で、
権威ある人の「勘、経験、度胸」に頼るマネジメントでは、
そろそろ限界、と感じていないでしょうか？



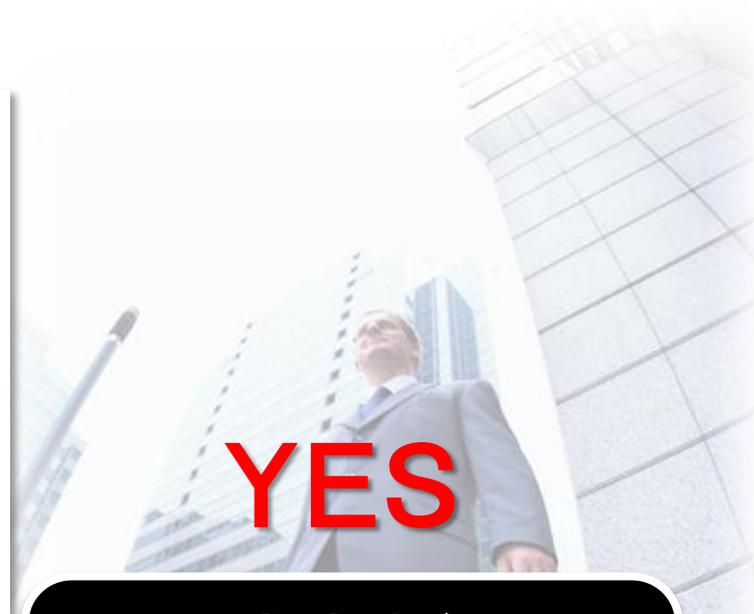
そこで質問です。

貴社ではビッグデータ活用が、
組織的な検討課題にあげられていますか？

貴社ではビッグデータ活用が、 組織的な検討課題にあげられていますか？



出典:2012年12月25日 株式会社野村総合研究所
ビッグデータの利活用に関するアンケート調査
<http://www.nri.co.jp/news/2012/121225.html>



約60%が
ビッグデータ活用を
自社の検討課題と認識

判断材料が欲しい！

つまり、もっとデータが欲しい。

An aerial photograph of a city skyline at dusk or dawn. The sky is filled with soft, grey clouds, and the city below is illuminated by a low sun, creating a hazy, golden glow. The text "最善の応えは、データの中にある。" is overlaid in the center of the image. The word "データ" is highlighted in yellow, while the rest of the text is white with a subtle drop shadow.

最善の応えは、データの中にある。

しかし・・・。

- ・分布や傾向だけじゃわからなかった不具合の原因を、
詳細データを追うことで確認したい。
- ・仮説と検証を繰り返して、
最も効果の高い組み合わせを見つけたい。

本当は、これをやりたかったのに、
できなかった。

それはなぜか？

真実を伝えられないサマリデータ、
収集できず切り捨てられた詳細データ、
膨大に溜まったまま活用できない履歴データ...

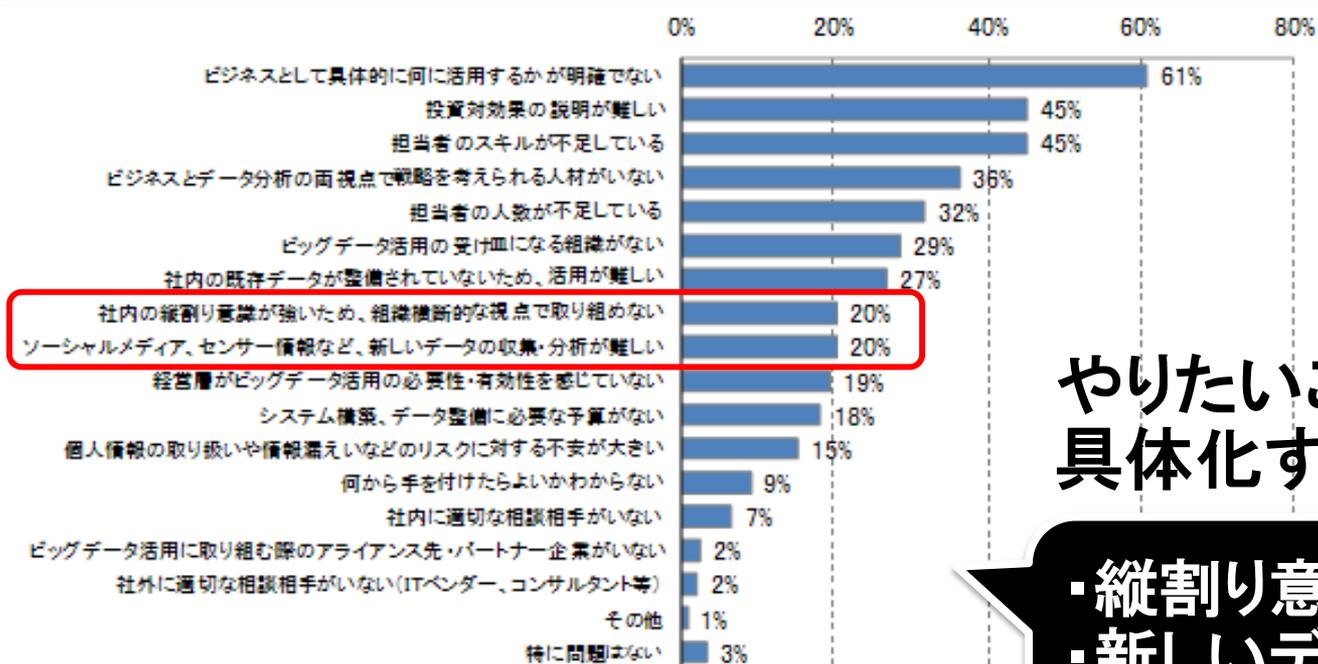
データはそこにある。
でも、上手く活用できていない。



そこで質問です。

今後、貴社でビッグデータの活用を進めていく場合、
どのようなことが問題・課題となりそうですか？

今後、貴社でビッグデータの活用を進めていく場合、 どのようなことが問題・課題となりそうですか？



やりたいことが
具体化すると…。

- ・縦割り意識の壁
- ・新しいデータの収集
や分析が難しい

出典:2012年12月25日 株式会社野村総合研究所
ビッグデータの利活用に関するアンケート調査
<http://www.nri.co.jp/news/2012/121225.html>

縦割り組織の壁？

新しいデータの収集と、分析が難しい？

いったい、どういうことなのでしょうか？

An aerial photograph of a lush green field with a path leading towards a line of trees in the distance. The text is centered in the middle of the image.

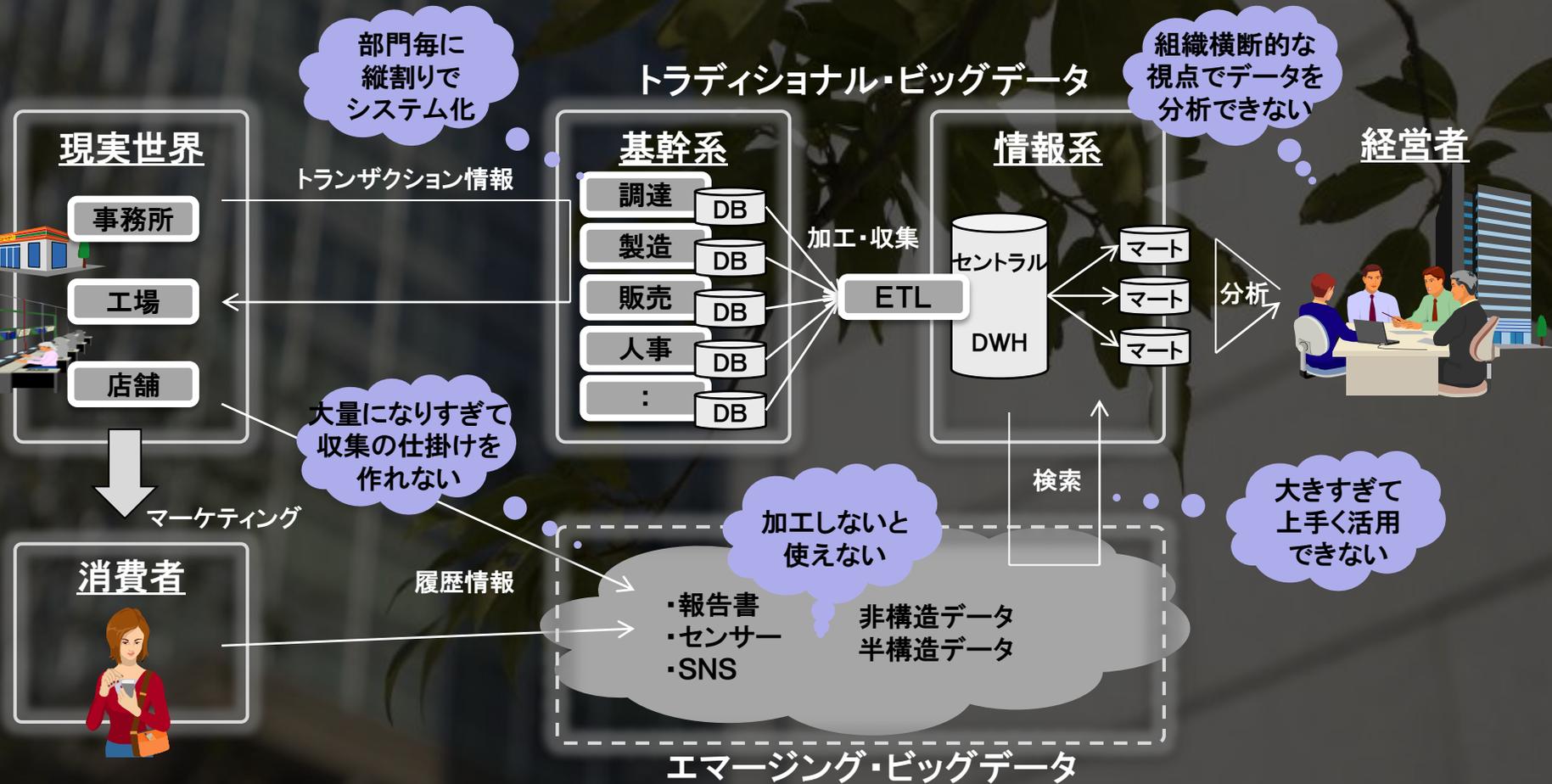
多様なプロセスで構成される
ビッグデータのライフサイクル

実は、

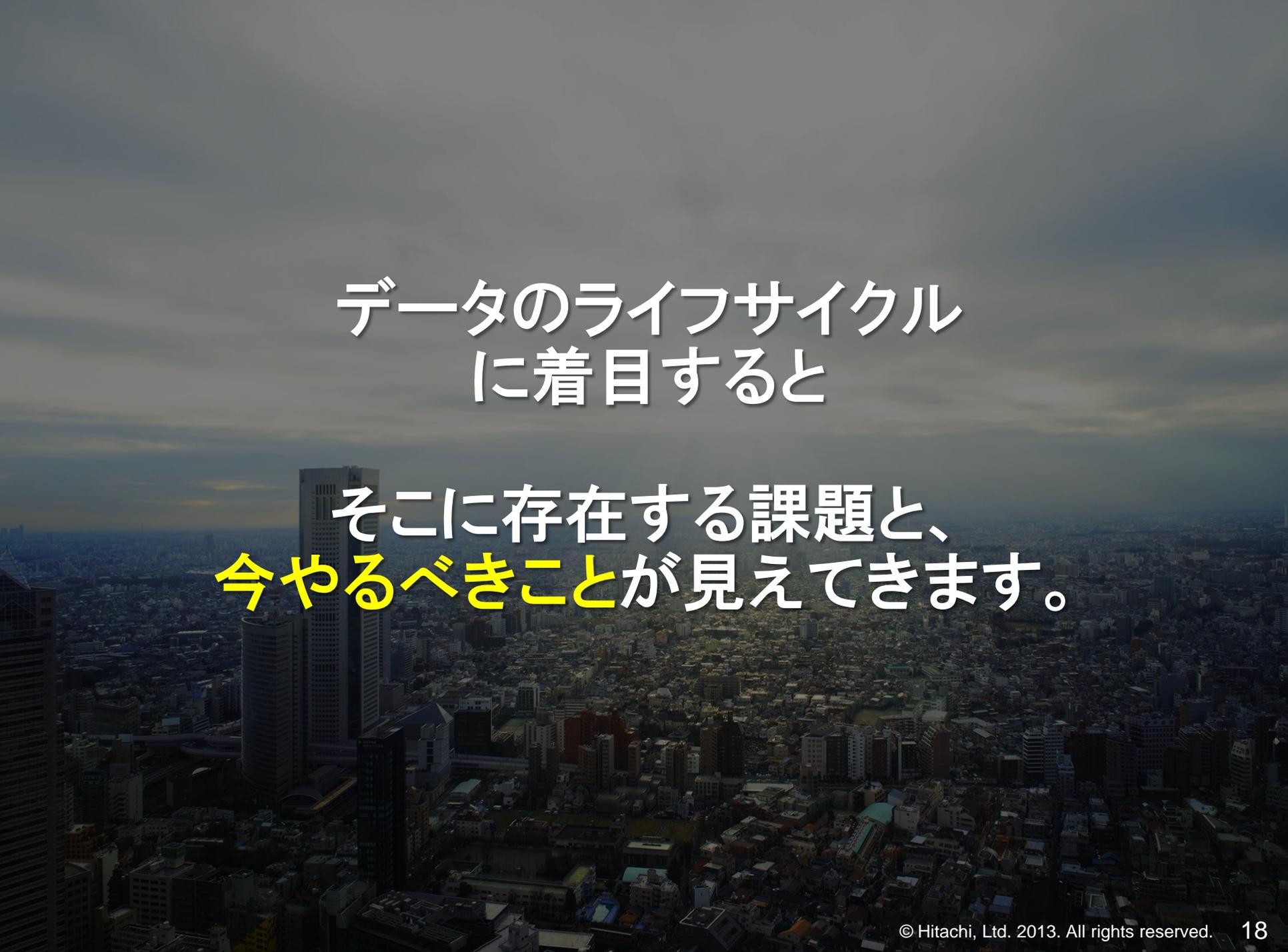
「データのライフサイクル」

に着目すると、
その問題の本質が、解決策が見えてきます。

縦割り組織とは、ということなのか？



新しいデータを活用できないのは、なぜか？

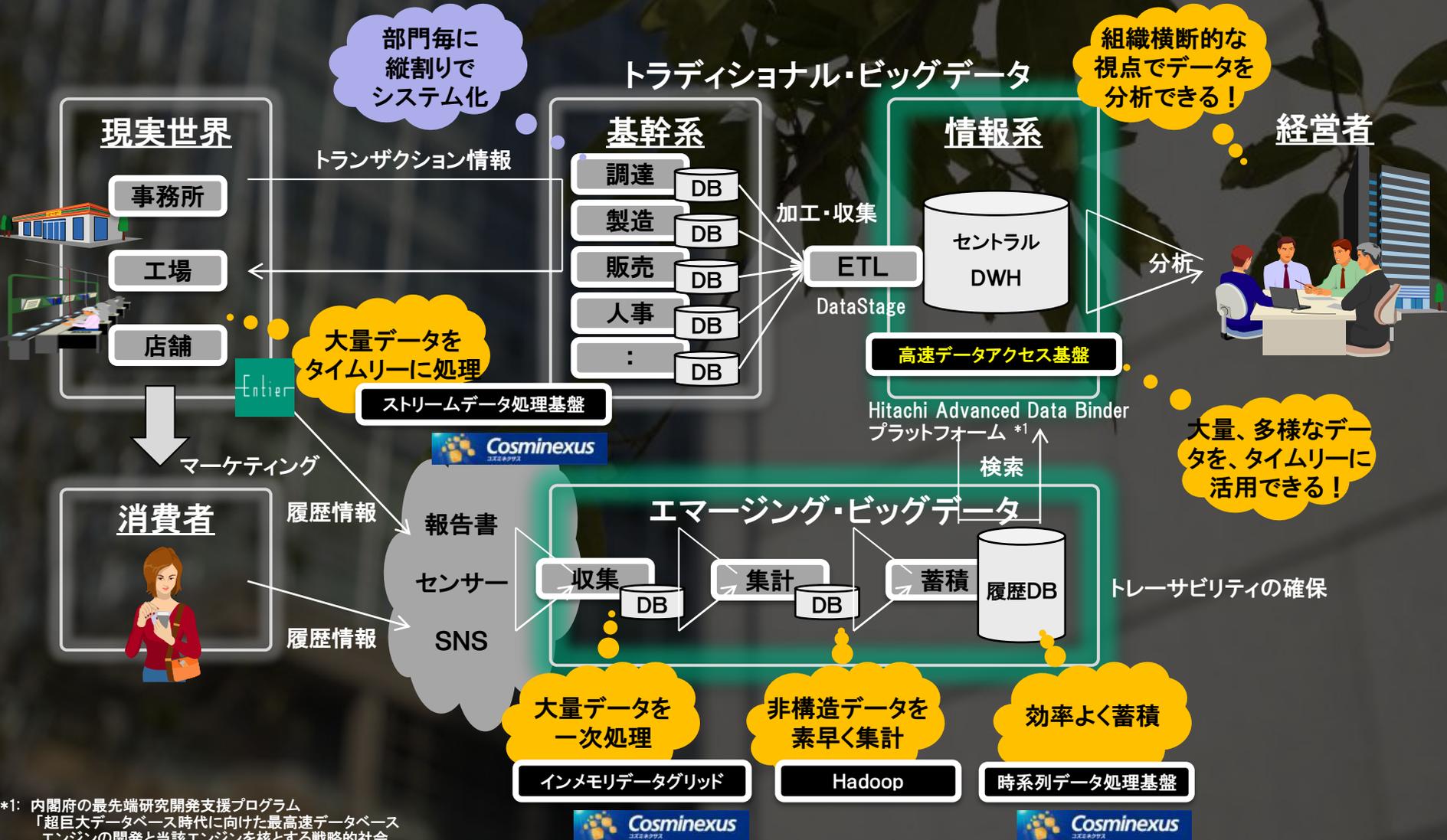


データのライフサイクル に着目すると

そこに存在する課題と、
今やるべきことが見えてきます。

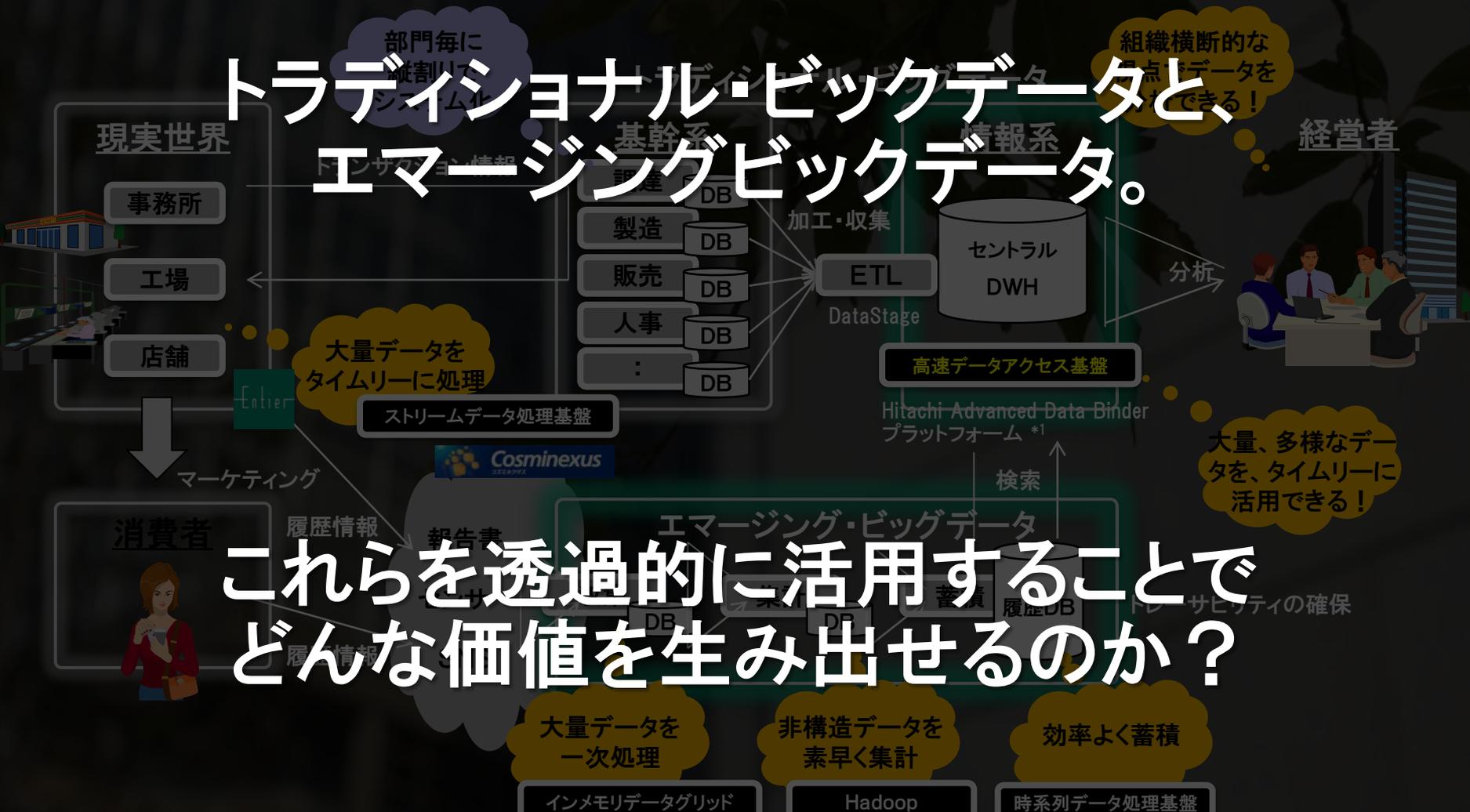
では、何をやるべきなのでしょう？
どんな方法で？

ひとつの方法では、ビッグデータを操れない。 データライフサイクルに応じて、適切な手段を使い分ける。



*1: 内閣府の最先端研究開発支援プログラム
「超巨大データベース時代に向けた最高速データベース
エンジンの開発と当該エンジンを核とする戦略的
社会サービスの実証・評価」(中心研究者:喜連川 東大
教授/国立情報学研究所所長)の成果を利用。

トランザショナル・ビッグデータと、 エマージングビッグデータ。



これらを透過的に活用することで
どんな価値を生み出せるのか？

- 大量データを一次処理 → インメモリデータグリッド
- 非構造データを素早く集計 → Hadoop
- 効率よく蓄積 → 時系列データ処理基盤

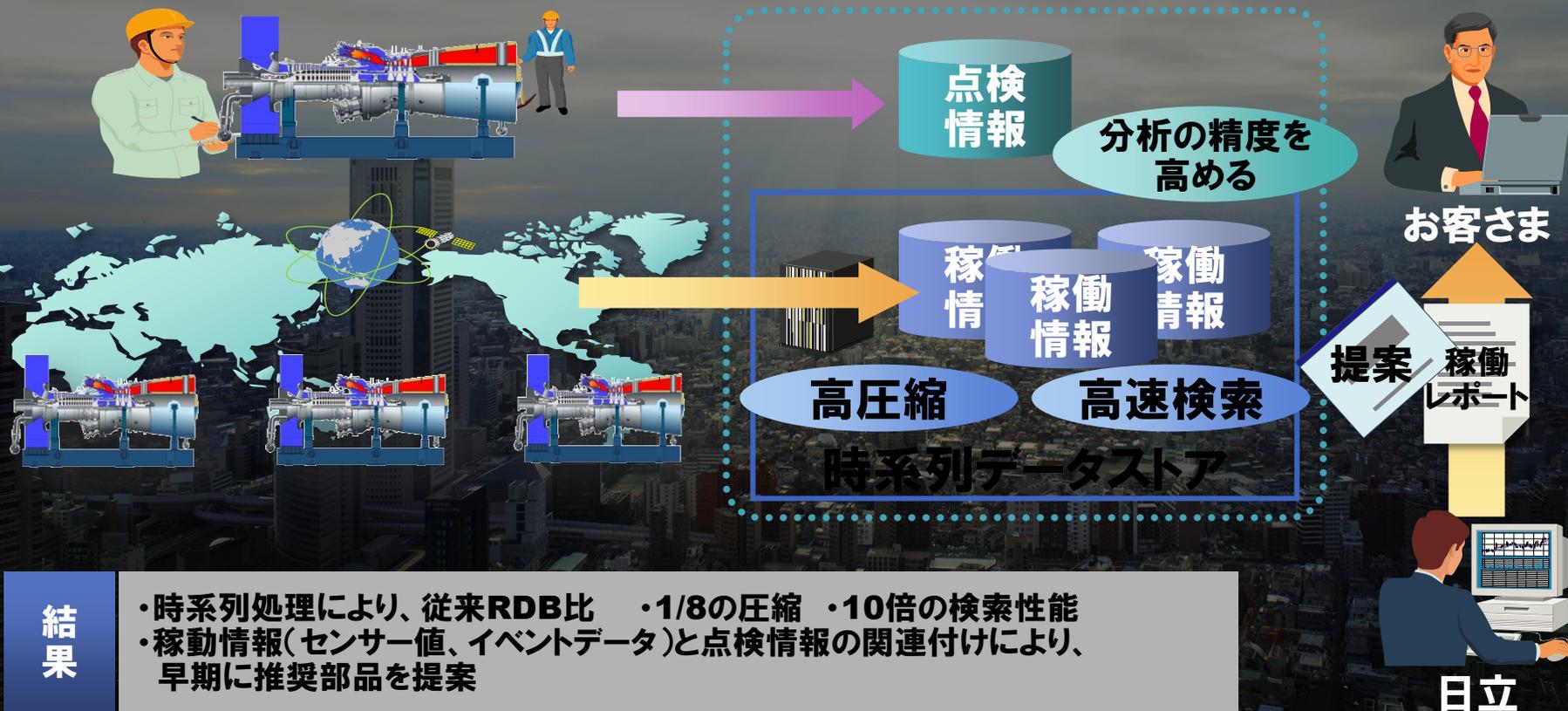
*1: 内閣府の最先端研究開発支援プログラム
「超巨大データベース時代に向けた最高速データベース
エンジンの開発と当該エンジンを核とする戦略的の社会
サービスの実証・評価」(中心研究者:喜連川 東大
教授/国立情報学研究所所長)の成果を利用。

実業とITの日立だから生み出せる、価値

例えば、ガスタービン保全業務でのビックデータ活用例

早く精度の高い情報をお客さまに伝えるために

- 素早く分析するために、長期にわたる稼働情報を効率よく管理
- 多角的に分析するために、さまざまなデータを活用



結果

- ・時系列処理により、従来RDB比 ・1/8の圧縮 ・10倍の検索性能
- ・稼働情報(センサー値、イベントデータ)と点検情報の関連付けにより、早期に推奨部品を提案

例えば小売業でのビックデータ活用例

多様なデータを活用

多量のデータを活用

店舗販売

CRM

EC,
ネットスーパー

コンタクトセンタ
活用

SNS活用

◆ 単品販売情報
(POS売上明細)

◆ 階層
・会社
・店
・ライン
・クラス
・棚
・単品

◆ 時系列
・年
・期
・四半期
・月
・週
・日
・曜日
・時間帯



品揃え改善
対面サービス強化

◆ 単品販売情報
◆ 顧客情報
・会員種別
・年代
・性別
・居住エリア
・ロイヤリティ
・家族、収入
など



One to One
マーケティング

◆ 単品販売情報
◆ 利用チャネル
◆ Access Log
サイト訪問履歴
商品参照履歴



Omni Channel
アプローチ

◆ 問合せ、
会話内容
◆ クレーム



顧客接点
強化

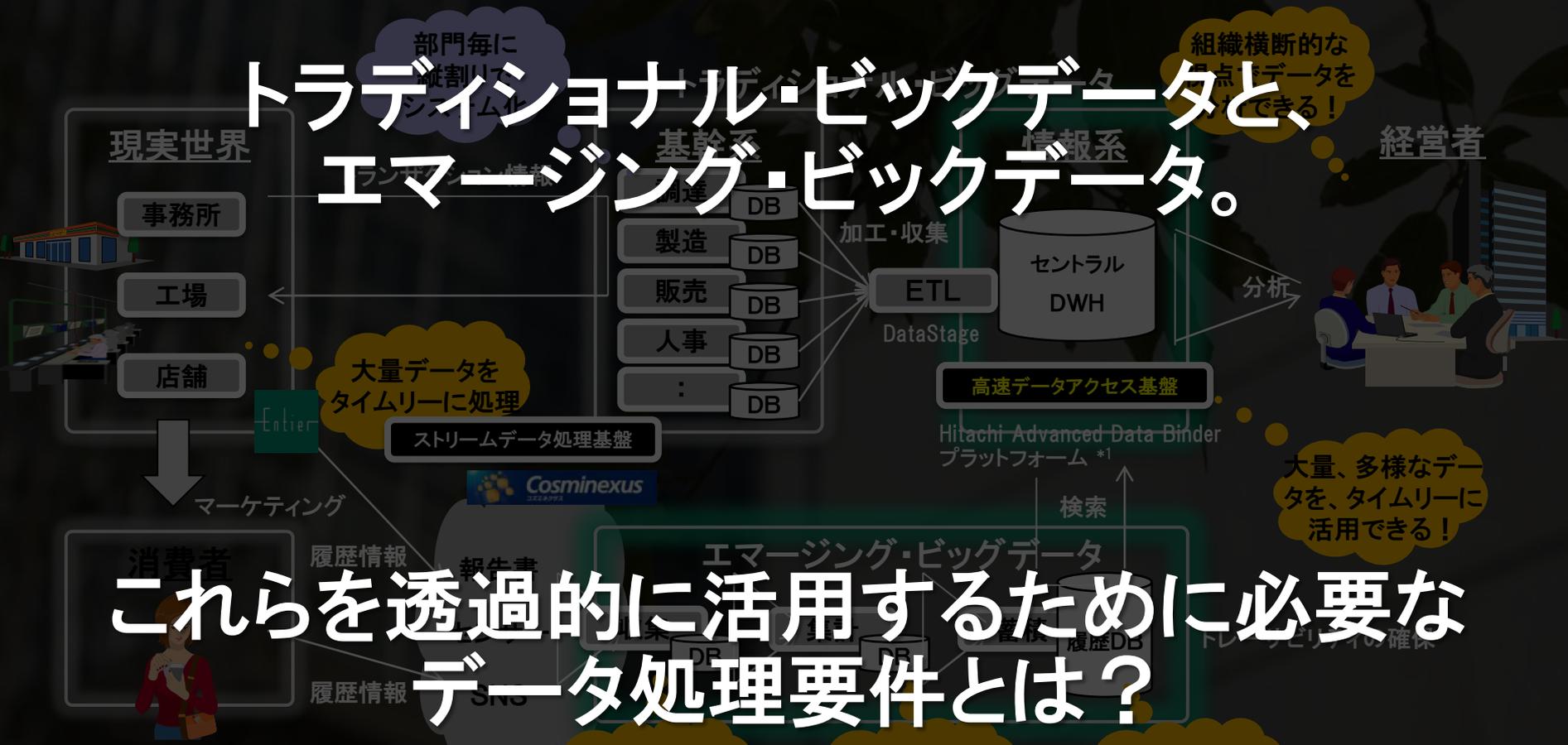
◆ つぶやき、
コメント、
情報発信



原因究明
未来予測

傾向の把握 ⇒ 事実確認 ⇒ 原因究明 ⇒ 未来予測へ

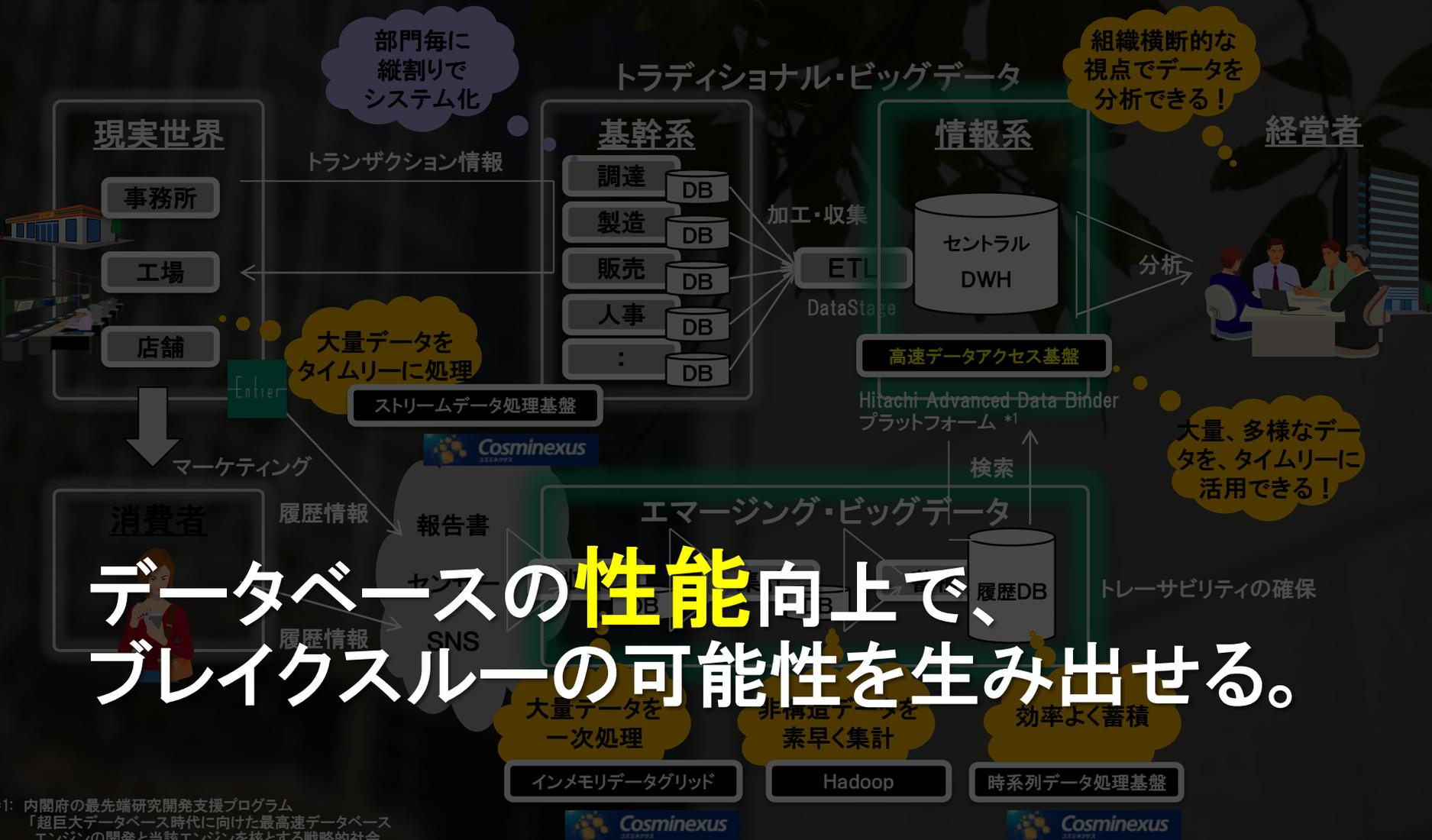
トランザショナル・ビッグデータと、 エマージング・ビッグデータ。



これらを透過的に活用するために必要な データ処理要件とは？

*1: 内閣府の最先端研究開発支援プログラム
「超巨大データベース時代に向けた最高速
データベースエンジンの開発と当該エンジンを
核とする戦略的サービスの実証・評価」
(中心研究者: 東大喜連川教授)の成果を利用。

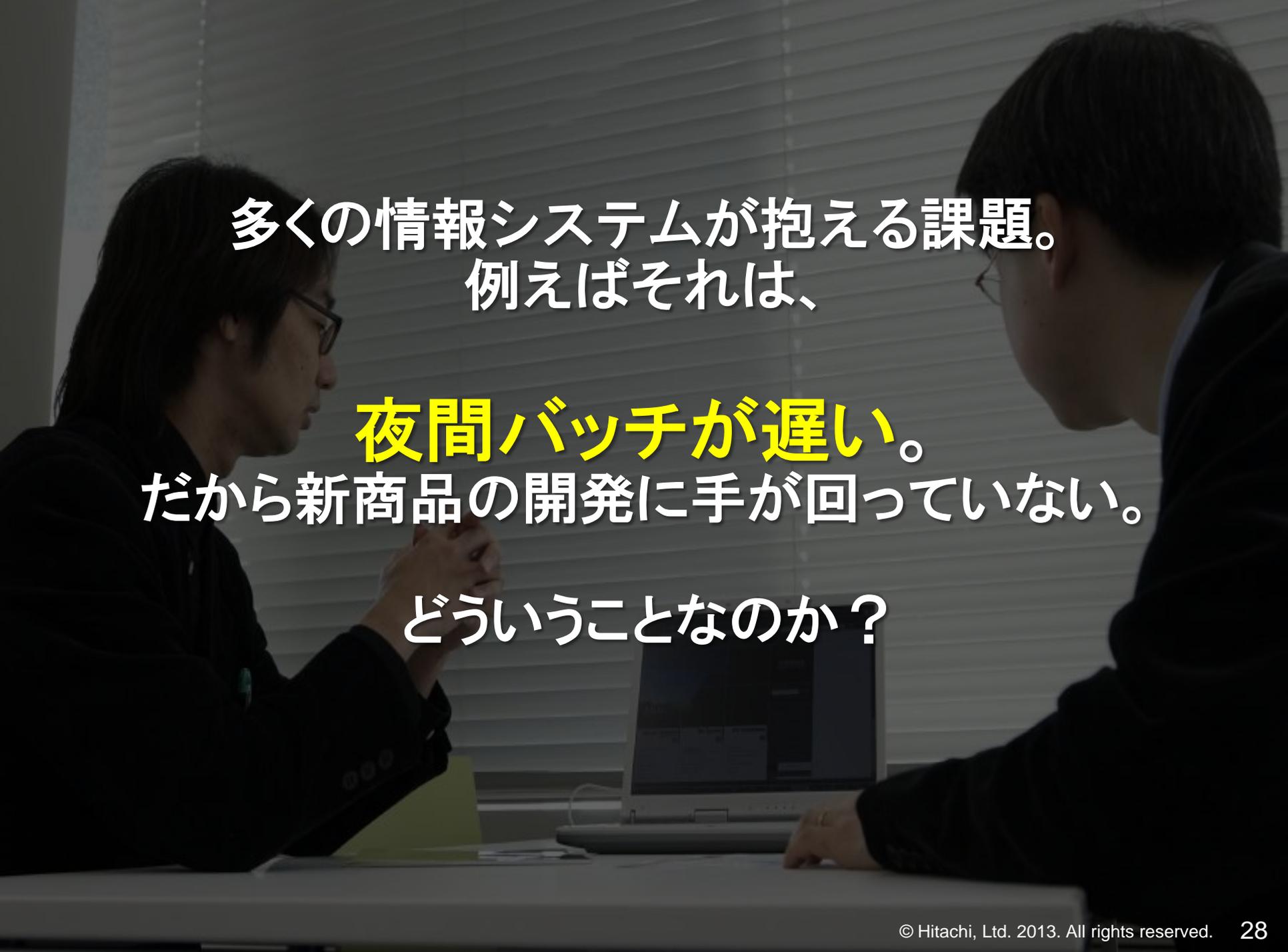
性能



データベースの性能向上で、
ブレイクスルーの可能性を生み出せる。

*1: 内閣府の最先端研究開発支援プログラム
「超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的サービスの実証・評価」(中心研究者:喜連川 東大教授/国立情報学研究所所長)の成果を利用。

100倍速くなれば、情報の価値が変わってくる



多くの情報システムが抱える課題。
例えばそれは、

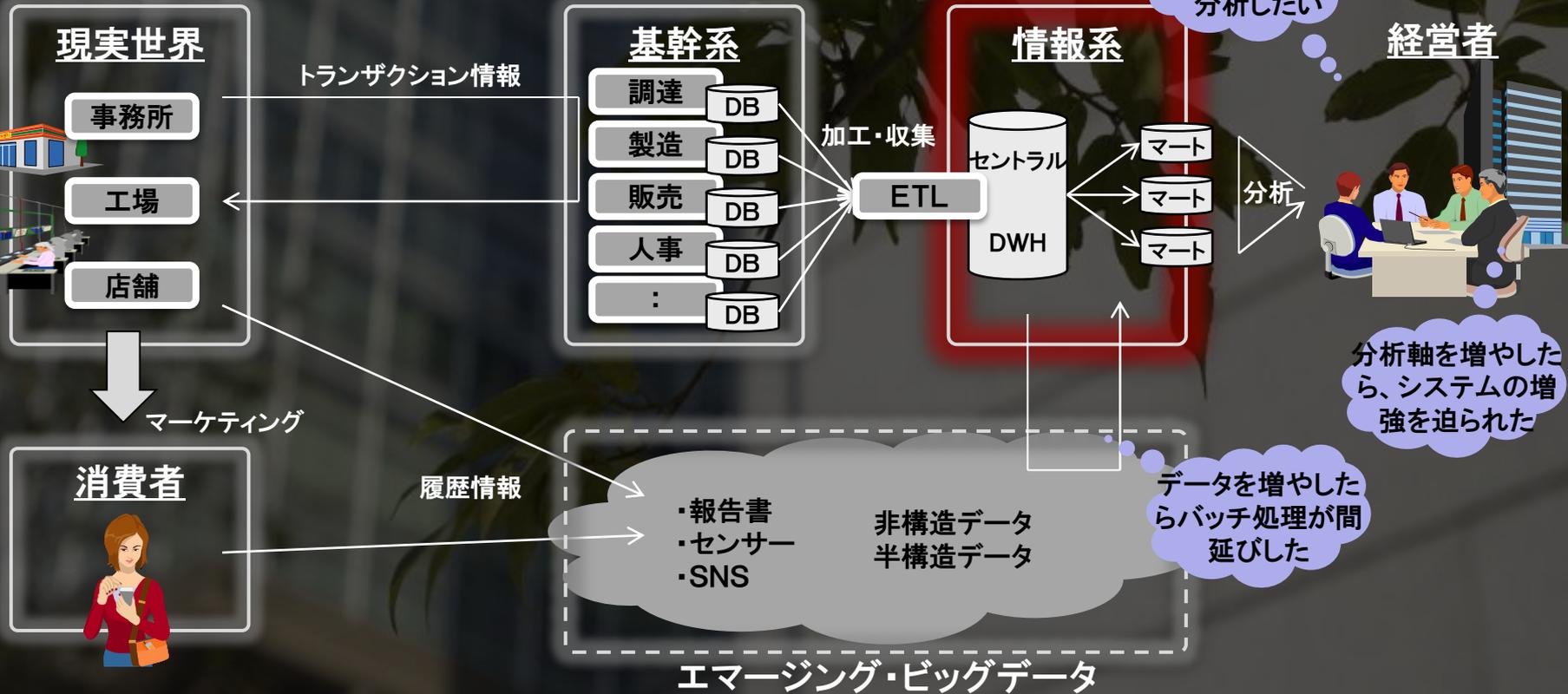
夜間バッチが遅い。
だから新商品の開発に手が回っていない。

どういうことなのか？

分析シナリオに従ったデータマート(集計DB)を 夜間バッチで作成している。

トラディショナル・ビッグデータ

組織横断的な
視点でデータを
分析したい



分析軸を増やしたら、
システムの増強を迫られた

データを増やしたら
バッチ処理が間延びした

もしデータ処理が、100倍速くなったら？

100倍速くなったら、何が変わるのか？

例えばPOSデータの分析だと、
性能上の制約によって1ヶ月だった分析対象の期間を、
10年間に**拡大**できる！

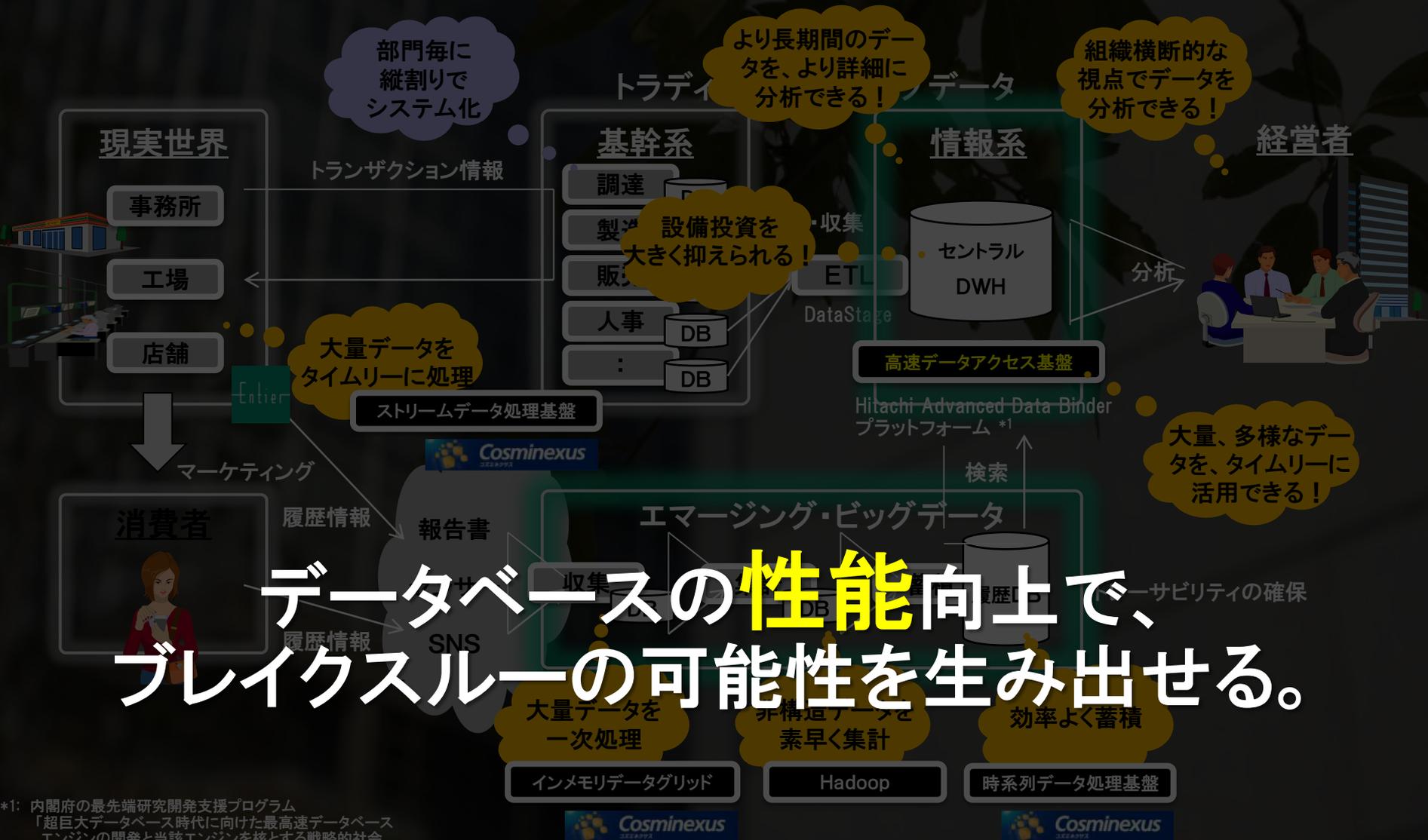
分析単位を
1日単位から1時間単位に**詳細化**できる！

100倍速くなったら、何が嬉しいのか？

そもそも分析処理を高速化するために、
夜間の実行していた
トータル100本のバッチ集計処理が、
全部要らなくなる！

そのバッチを実行するための
サーバやストレージ、
そもそものバッチシステムの構築と保守コストが、
まるまる不要になる！

もし100倍速だったら——。



部門毎に縦割りでシステム化

より長期間のデータを、より詳細に分析できる！

組織横断的な視点でデータを分析できる！

設備投資を大きく抑えられる！

大量データをタイムリーに処理

大量、多様なデータを、タイムリーに活用できる！

データベースの性能向上で、ブレイクスルーの可能性を生み出せる。

*1: 内閣府の最先端研究開発支援プログラム
「超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的サービスの実証・評価」(中心研究者:喜連川 東大教授/国立情報学研究所所長)の成果を利用。

自社従来比100倍の性能に挑む
超高速データベースエンジン
Hitachi Advanced Data Binder

昨年のリリース直後から、注目を集めています。



■大賞 [日立製作所]
高速データアクセス基盤
Hitachi Advanced Data Binder プラットフォーム

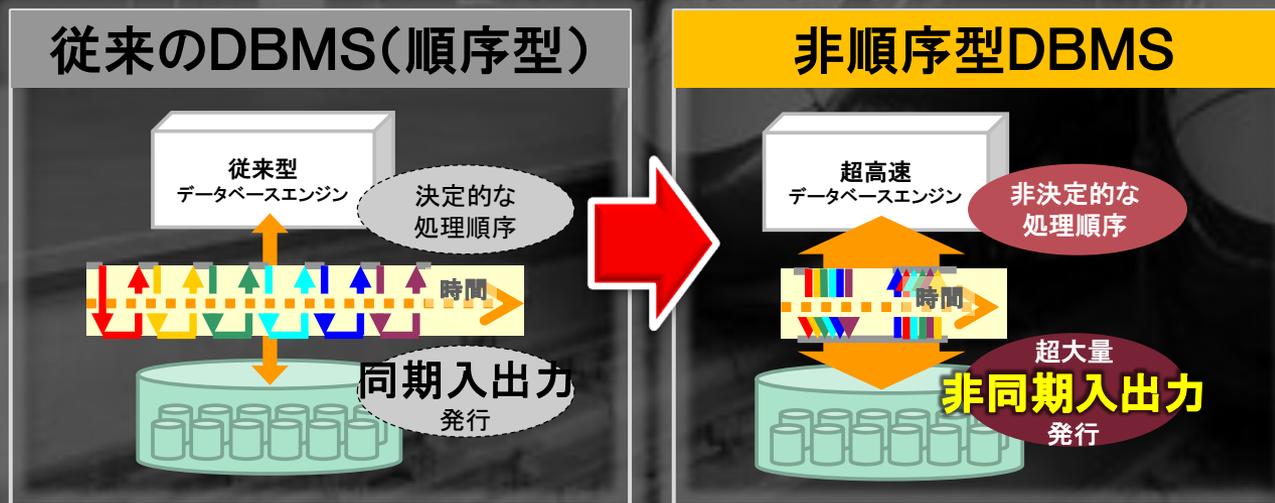
■優秀賞
[日本マイクロソフト] Windows Server 2012
[富士通] STYLISTIC Q702/F
[ジャストシステム] UnitBase
[NTTデータ] フルオープン仮想化基盤構築ソリューション

従来の常識を超えた、RDBエンジン

非順序型実行は、通常のノード「並列」や「分散」とはメカニズムが違う。
ある意味、順序性を無視して処理の多重度を高める技術。

辻褄は、最後に合わせられる。

非順序型実行原理*1



*1: 喜連川 東大教授/国立情報学研究所所長・合田 東大特任准教授が考案した原理。

*2: 内閣府の最先端研究開発支援プログラム「超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的サービスの実証・評価」(中心研究者:喜連川 東大教授/国立情報学研究所所長)の成果を利用。

「非順序型実行原理」

- 喜連川 東大教授／国立情報学研究所所長・合田 東大特任准教授が考案した原理。
- 内閣府の最先端研究開発支援プログラム*
「超巨大データベース時代に向けた最高速データベースエンジンの開発と、当該エンジンを核とする戦略的社会サービスの実証・評価」
(中心研究者:喜連川 東大教授／国立情報学研究所所長)の成果。

*最先端研究開発支援プログラム(FIRSTプログラム)の詳細については下記URLをご参照ください。
URL:<http://first-pg.jp/about-us/about-30.html>

従来のデータベースは、基本的にデータ格納順に
ストレージへのI/O処理をおこないます。
従ってI/O待ち(無駄)が避けられない。

「NO」が「100」の行を
検索したい！

NO	VALUE
100	A1
100	B1
200	C1
300	D1
100	E1
100	E1
200	D1

データ格納順にI/O処理

非順序実行方式でない場合

100 A1

待ち

100 B1

待ち

100 E1

待ち

検索結果

100	A1
100	B1
100	E1

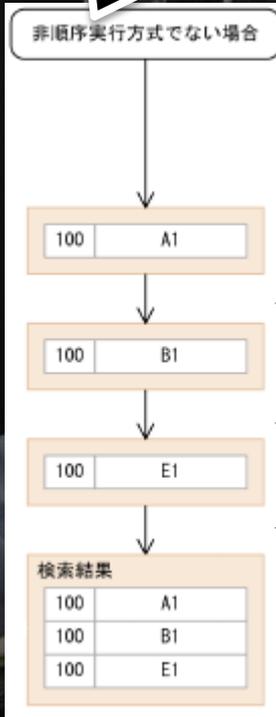
100	E1
100	B1
100	A1

非順序実行原理を採用した Hitachi Advanced Data Binderは、I/Oを待たずに、次々に処理を行うことができます。

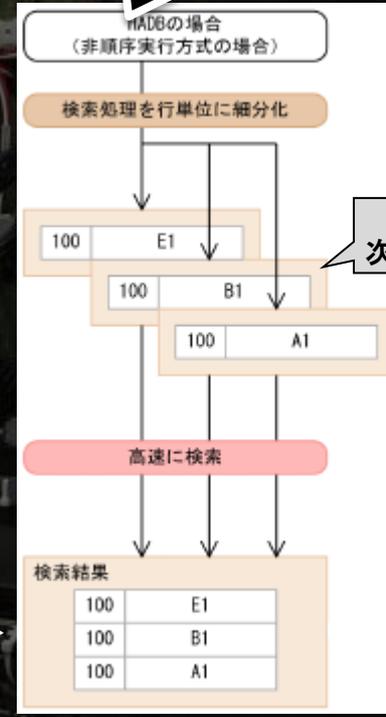
「NO」が「100」の行を検索したい！

NO	VALUE
100	A1
100	B1
200	C1
300	D1
100	E1
100	E1
200	D1

データ格納順にI/O処理



行単位にタスクを割り当て 並列にI/O処理



I/Oを待たずに 次の行を取りに行く。

だから毎回、結果の順序が異なる。順序を保障したいなら、ORDER BYでソートすれば良い。

処理の順番を無視できるから、メリットがある。

●従来のDBMS(順序型)

- ❑サーバ側のCPUリソースを十分に使いきれなかった。
- ❑だからサーバをスケールアウトさせて並列度を高めていた。
- ❑結果としてサーバ台数が増え、コストが増大してしまう。

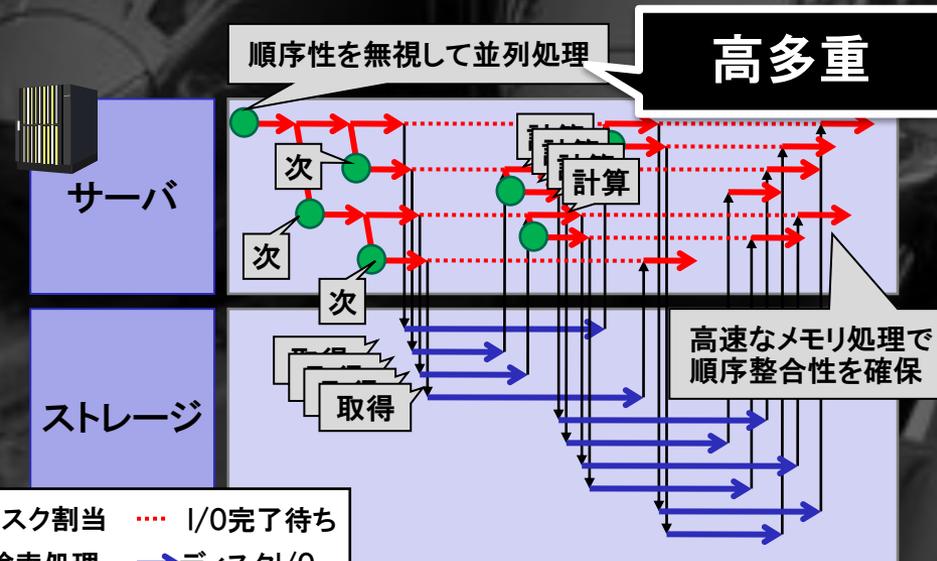
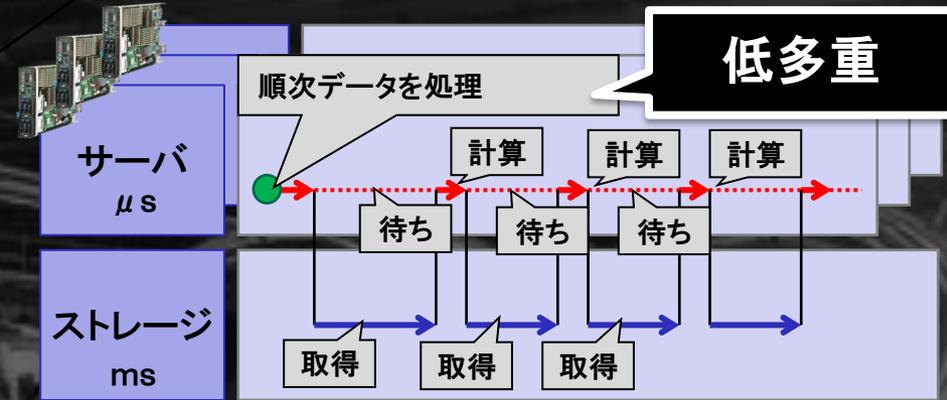
※従来方式では、SQLパラレル処理やパーティショニングを駆使して処理多重度を稼いでいた。

●HADB(非順序型)

- ❑サーバ側のマルチコアCPUを使い切るので効率的。
- ❑結果としてサーバ台数も減り、コストを低減できる。

SQLクエリ単位の処理多重度の差異

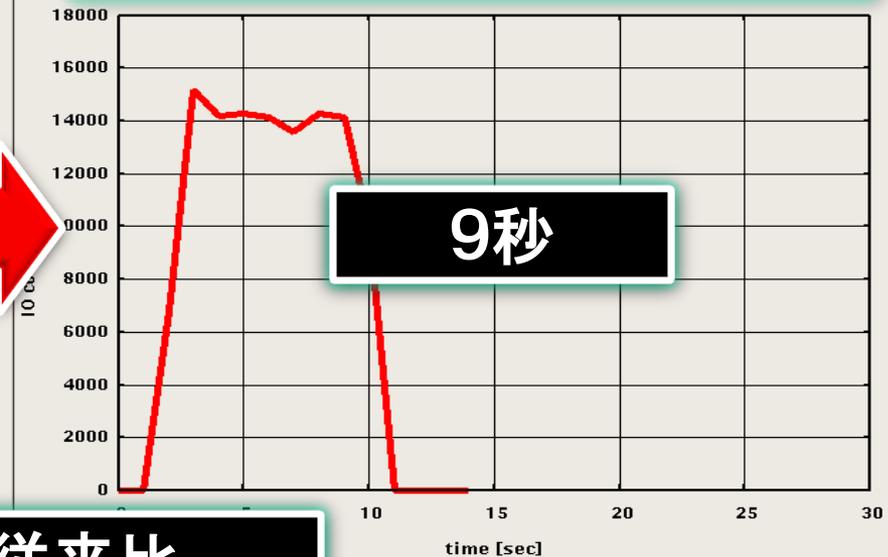
サーバ台数を増設(スケールアウト)して多重度を稼ぐ方式



非順序型、Hitachi Advanced Data Binderの威力

順序型 従来自社DBMS

非順序型 Hitachi Advanced Data Binder
プラットフォーム



自社従来比
約100倍の高速化



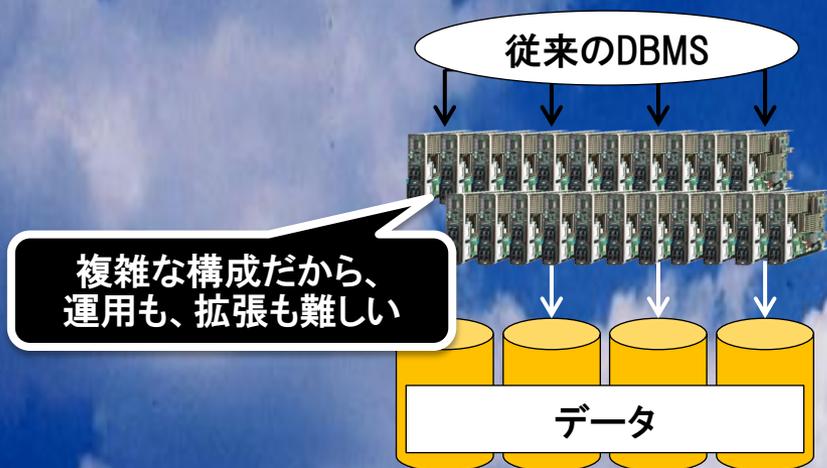
非順序型、Hitachi Advanced Data Binderの威力

デモンストレーションでご確認ください。

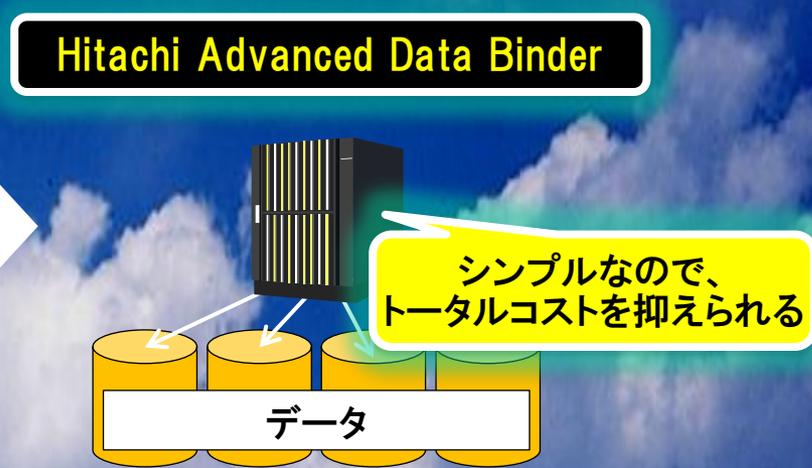
構造がシンプルだから、メリットが大きい

- シンプルな構成なので、運用時、拡張時のコストメリットあり
非順序型実行原理により、シンプルなシングルサーバ構成でありながら、サーバのマルチコアプロセッサ*およびストレージシステムの利用効率を最大限に高めることで、処理性能の大幅な向上を実現。

スケールアウト型
(複雑な構成)

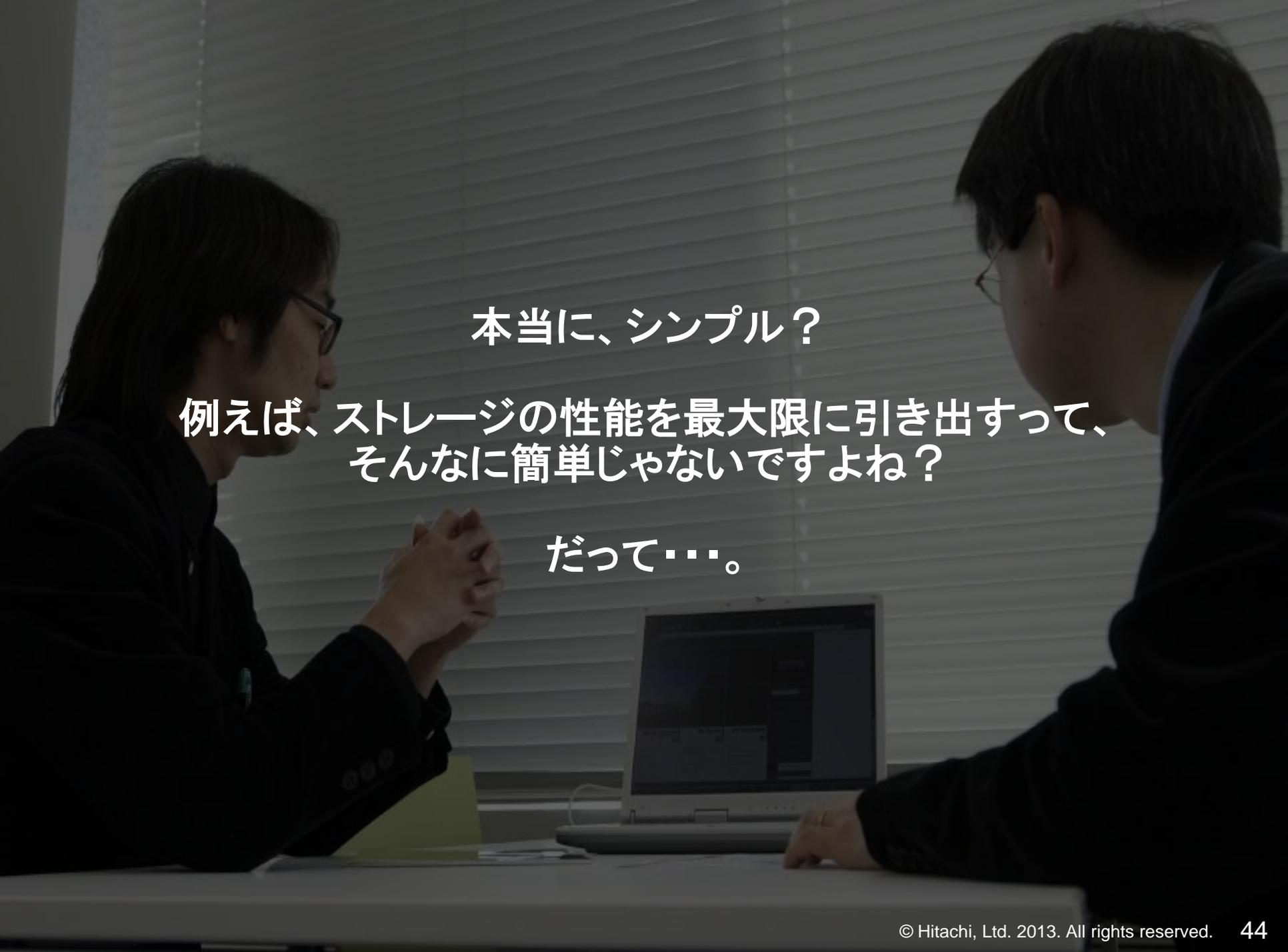


スケールアップ型
(シンプルな構成)



*: 多数のコアを集積したプロセッサ。

自製の日立だから作れる、ビッグデータ活用基盤

A dimly lit office scene with two men in business attire. The man on the left has his hands clasped and is looking towards the man on the right. A laptop is open on the desk between them. The background consists of horizontal blinds.

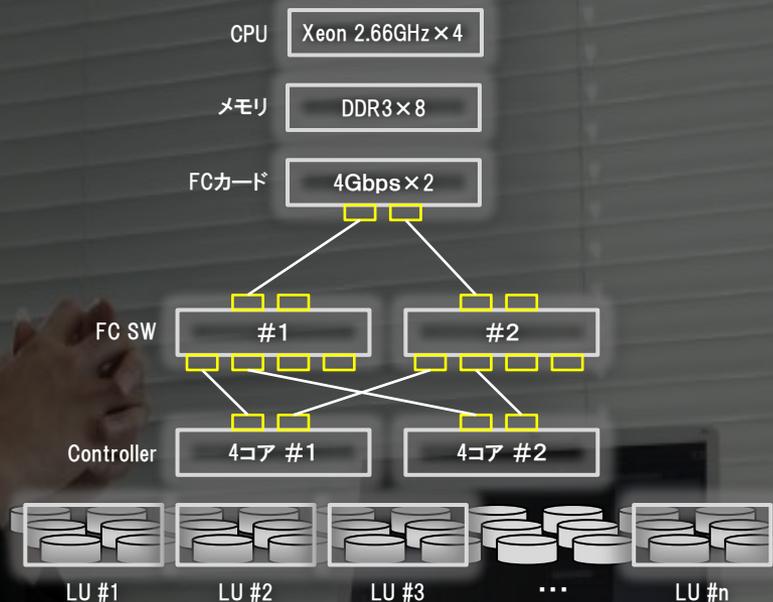
本当に、シンプル？

例えば、ストレージの性能を最大限に引き出すって、
そんなに簡単じゃないですよね？

だって…。

良くSEと、こんな会話しませんか・・・。

「あ・・・。なぜ想定した性能が出ないかという・・・、ストレージは、1RAIDグループ1LUで設定していたのですね。おそらく、こうしてしまうと1RAIDグループ1LUに対してストレージのCPUコアは1個しか割り当てないので、多重性能が上がらないものと思います。今回のストレージ装置はコントローラ上にコアが4個なので1RAIDグループ4LU以上にはしておいたほうがよかったですね・・・。」



本当にシンプル？ 結局、使うのが難しいのでは？

大丈夫

データベースも、ストレージも
どちらも同じ組織で、全て自製している日立ですから。

データ処理の特性やサイズに応じて 最適な基盤の組み合わせをご提案できます。

日立データベースミドルウェア

HiRDB

XDM/SD

XDM/RD

グリッドバッチ

ストリームデータ処理基盤

インメモリデータグリッド

高速データアクセス基盤



BS2000

日立アドバンスサーバ



TS20



RS440



BS320



BS500

日立ストレージ

Hitachi Unified Storage VM



Hitachi Virtual Storage Platform



● お客様の業務サイズに応じた推奨構成を、「ベストプラクティスモデル」としてご提供します。

✓サイジング済サーバ/ストレージセットだから安心、簡単。

Hitachi Advanced Data Binder プラットフォーム



Hitachi Advanced Data Binder プラットフォーム

超高速データベースエンジン

日立ラックサーバ

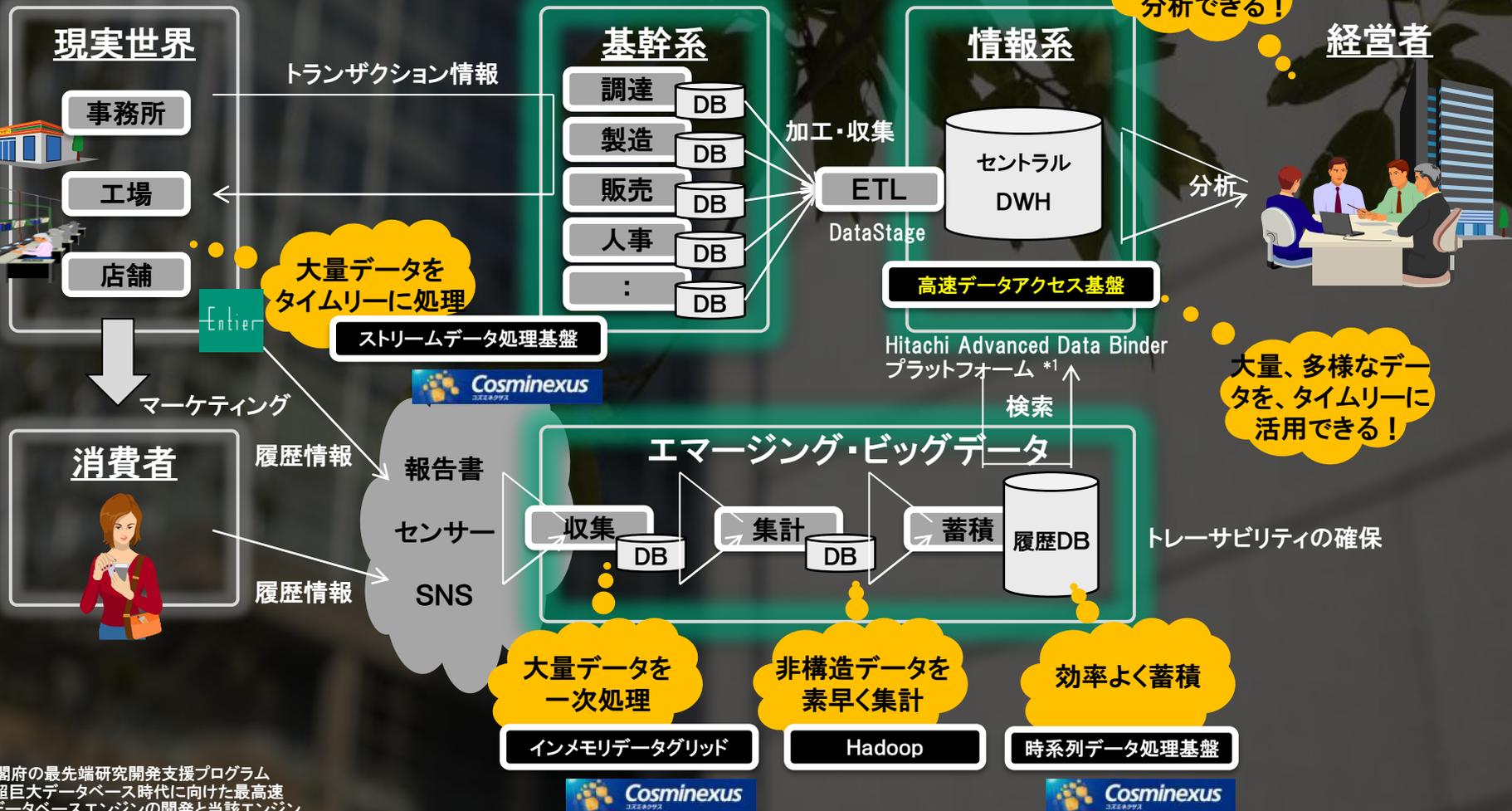
日立ストレージ

つまり…、
ビッグデータを活用して価値を生み出すには、
どんな基盤を選べば良いのでしょうか？

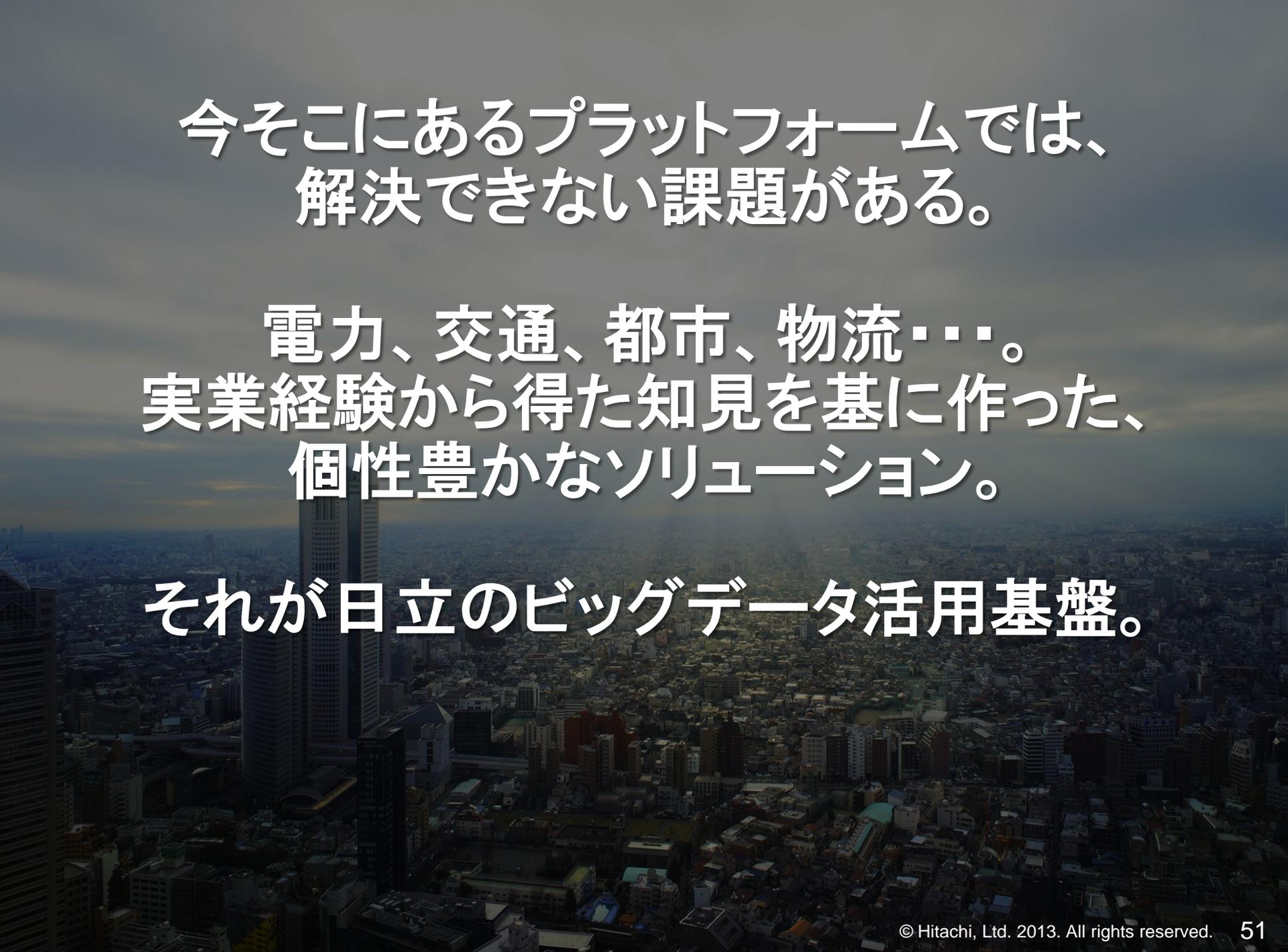
データライフサイクルに応じて、適切な**基盤**を使い分ける。 日立はお客様に最適な選択肢を、ご提案します。

トラディショナル・ビッグデータ

組織横断的な
視点でデータを
分析できる！



*1: 内閣府の最先端研究開発支援プログラム
「超巨大データベース時代に向けた最高速
データベースエンジンの開発と当該エンジ
ンを核とする戦略的サービスの実証・評価」
(中心研究者: 東大喜連川教授)の成果を利用。



今そこにあるプラットフォームでは、
解決できない課題がある。

電力、交通、都市、物流・・・。
実業経験から得た知見を基に作った、
個性豊かなソリューション。

それが日立のビッグデータ活用基盤。

全ては、お客様のスペシャルな夢を、
現実に変えるために。

**Human Dreams.
Make IT Real.**

これからも日立は、
お客様のスペシャルな夢を叶える
選択肢をご提案し続けます。

- ・ 本資料に記載の会社名，製品名は，それぞれの会社の商標もしくは登録商標です。