

ビッグデータ概説 ダイジェスト

NECラーニング テクノロジー研修事業部

横垣 裕史

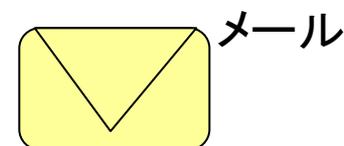
アジェンダ

- **ビッグデータとは**
- **ビッグデータが注目される背景**
- **ビッグデータの活用**
- **ビッグデータの活用事例**
- **ビッグデータ活用の目的**
- **ビッグデータ活用方法の検討**
- **データ分析実現方法の検討**
- **システム構成の検討**
- **ビッグデータ活用のシステム構成**
- **データの生成、転送、蓄積、整理、分析**

ビッグデータとは

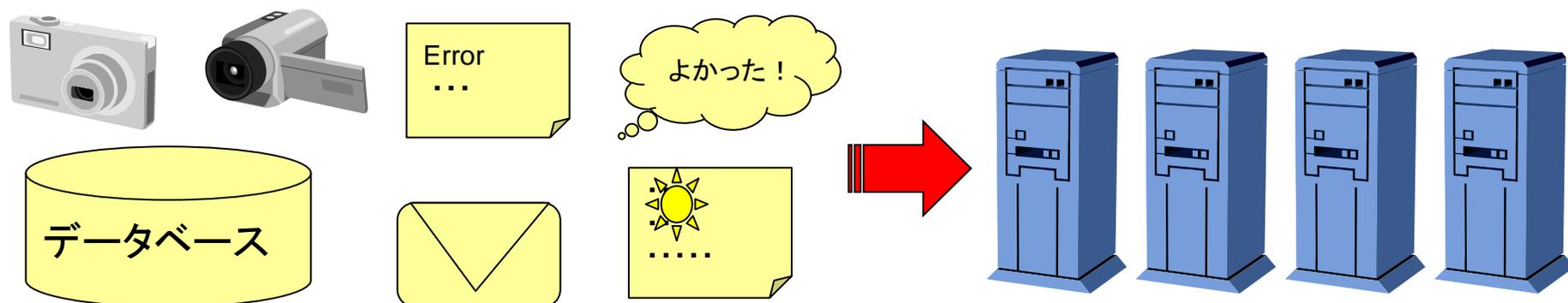
データのサイズ、量、種類、発生頻度が拡大

- データサイズの拡大
- データ量の増大
- データの種類(形式)の増加
- 発生頻度の増加



ビッグデータが注目される背景

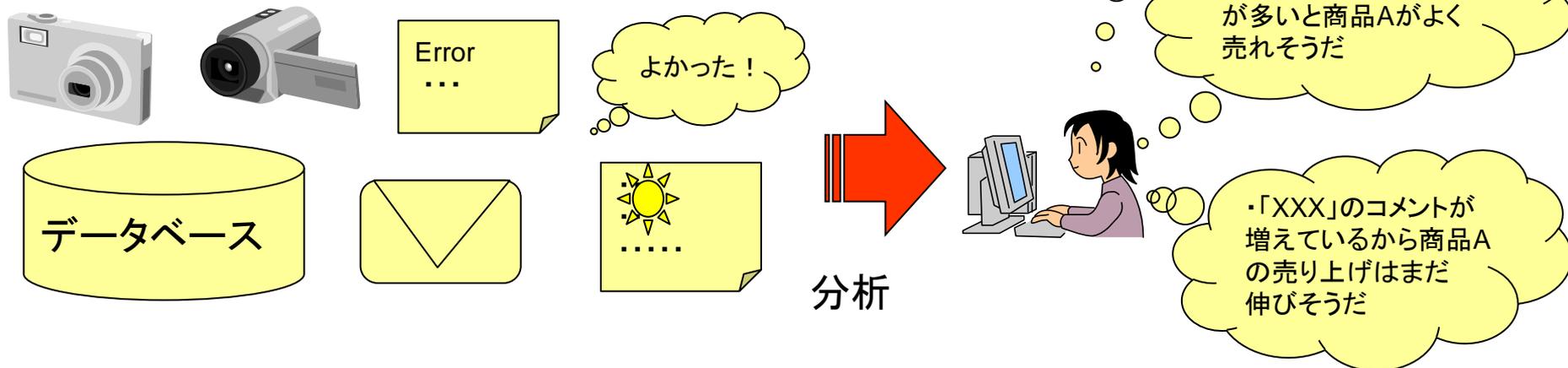
- 基幹業務システム化の成熟
- ビジネス成長実現への期待
- 分散処理技術の発展
- クラウド技術の普及



ビッグデータの活用

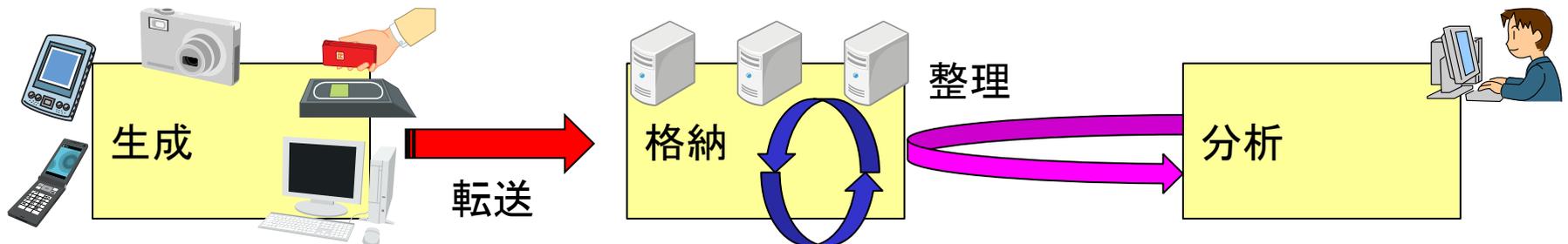
ビッグデータを分析し、新しい知見を導出

- 現状や実績の把握 (何が起きたか)
- 原因の検討 (なぜ起きたか)
- 未来の予測 (何が起きるか)



ビッグデータの活用事例

事例	目的	収集データ	データ分析
ウェザーニュース	高精度の天気予報、ゲリラ雷雨予測により、航海、農業、電力の最適化を実現	<ul style="list-style-type: none"> 観測器からの観測データ 有料会員からの5感による観測データ 	観測データと天気との相関を分析
NTTドコモ 「モバイル空間統計」	人口分布、人口構成、移動人口の空間統計を、まちづくり、防災計画に活用	<ul style="list-style-type: none"> 携帯電話の位置情報 性別、年齢などの利用者の属性情報 	位置情報、利用者属性情報を集計
VISA	不正利用パターンを作成し、リアルタイムにカードの不正利用を検知	<ul style="list-style-type: none"> クレジットカードの利用・取引状況 	不正利用パターンの学習 不正利用のリアルタイム検知
カブドットコム証券	株価動向の分析精度を向上し、新しい投資情報サービスを提供	<ul style="list-style-type: none"> ツイッターなどのソーシャルメディア上のデータ 株価動向 	ソーシャルデータと株価動向との相関を分析
エスエス製薬 「カゼミル」	ツイッター、天気予報により「風邪話題度」を集計・予測し、風邪の流行に注意を促す	<ul style="list-style-type: none"> ツイッター上の風邪情報 各地の天気予報 	「風邪話題度」と天気予報との相関



ビッグデータ活用の目的

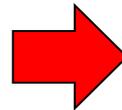
業務や生活を改善するために、ビッグデータ活用により新しい知見を導出する

■ 例

- ・ 電力供給の無駄をなくすために、電力の使用状況を把握、予測する
- ・ 顧客に適切な商品をリコメンドするために、購買履歴の把握、動線分析をする
- ・ システム異常を早期発見するために、システムログを機械学習する



・費用の無駄が多い
・売上が上がらない
etc...



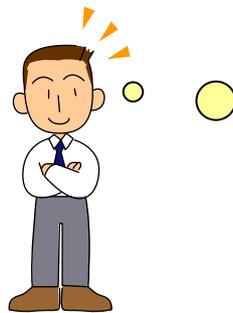
・費用の無駄の原因が分かり、
無駄を削減できた！！
etc...

ビッグデータ活用方法の検討

業務改善を実現するためのビッグデータの活用方法を検討

■求められるスキル

- 対象業務知識 ⇒ 問題・課題の発見
- データ分析知識 ⇒ データ分析で可能なことを把握
- 問題解決力 ⇒ 問題解決方法を提案



・ネットショップの売上を上げるには、
検索エンジン対策をすべきだ！
・売上に関係の深い要因を見つけ、
改善案を考えたほうがいい

データ分析実現方法の検討

問題解決に必要なデータ分析を実現する方法を検討

■求められるスキル

➤ データ分析手法の知識

⇒ 使用するデータ分析手法の決定

⇒ データ分析に必要なデータの設計



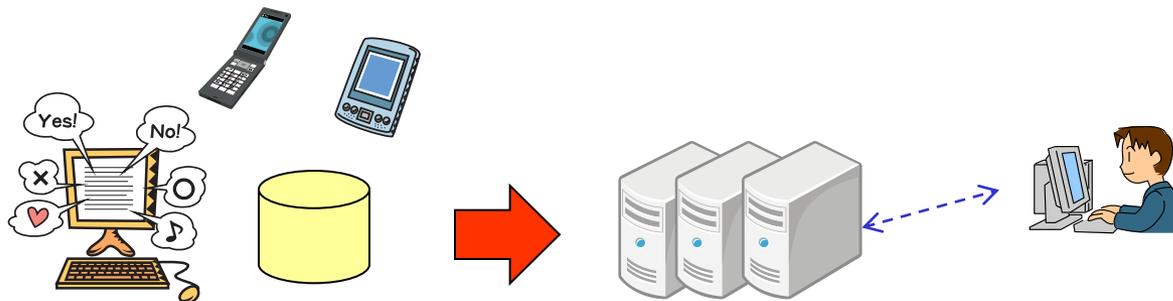
- ・売上に関係の深い要因を見つけるには、相関分析がよさそうですね。
- ・相関分析するには、売上データとそれに関わる要因（日付、クリックストリーム、つぶやき数など）が必要です。

システム構成の検討

ビッグデータを活用するシステム構成を検討

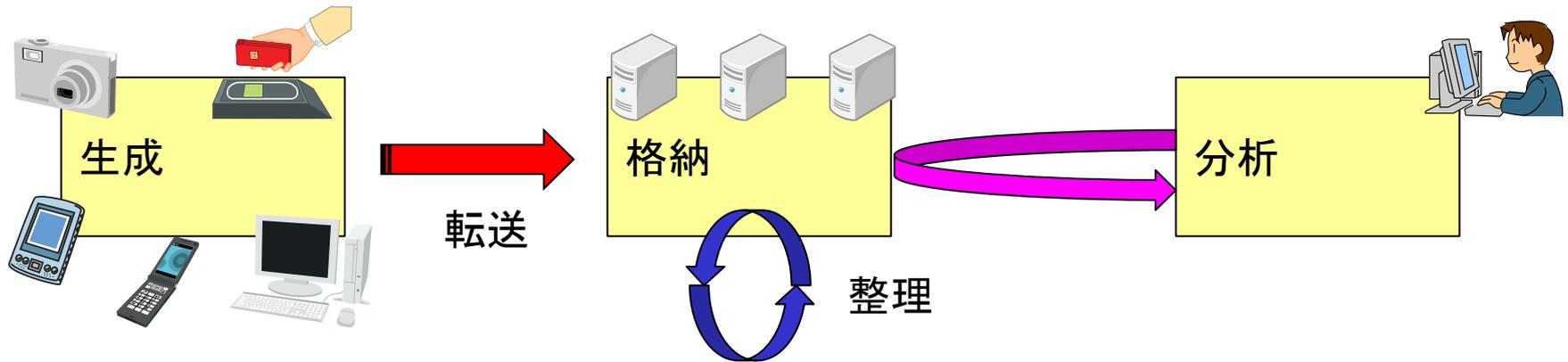
■求められるスキル

- システム設計スキル ⇒ システム構成の決定
- システム構築スキル ⇒ システムの実現
- システム運用スキル ⇒ システムの運用・保守

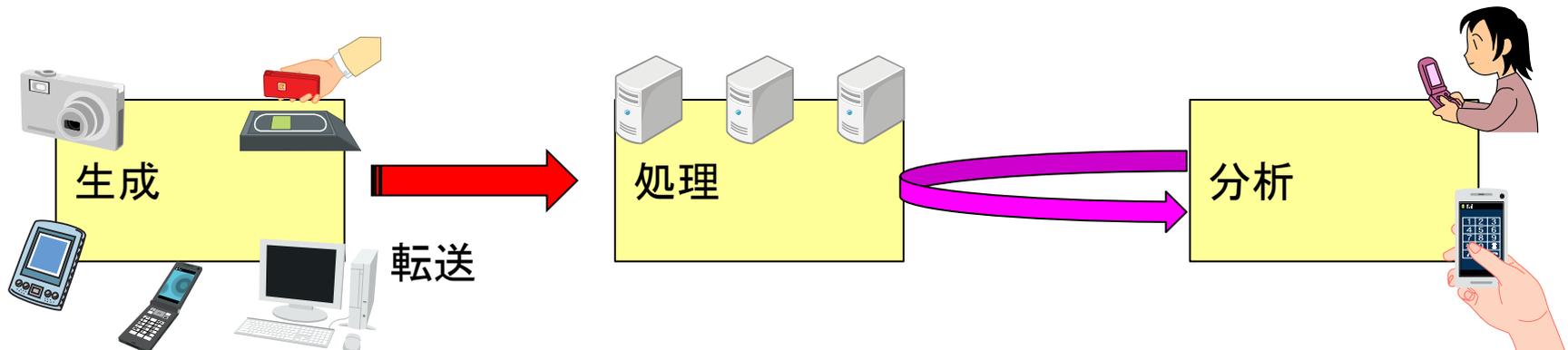


ビッグデータ活用のシステム構成

蓄積型のビッグデータ活用



リアルタイム型のビッグデータ活用

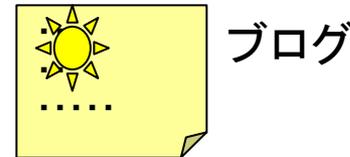
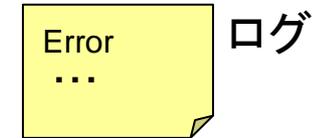
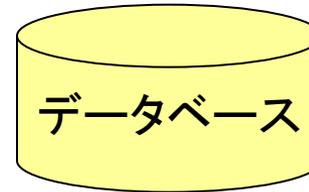


データの生成

ビッグデータ活用に必要なデータを生成

■ 既存データの利用

- ・システムログ
- ・基幹業務システムデータ
- ・SNS(Twitter、ブログ)



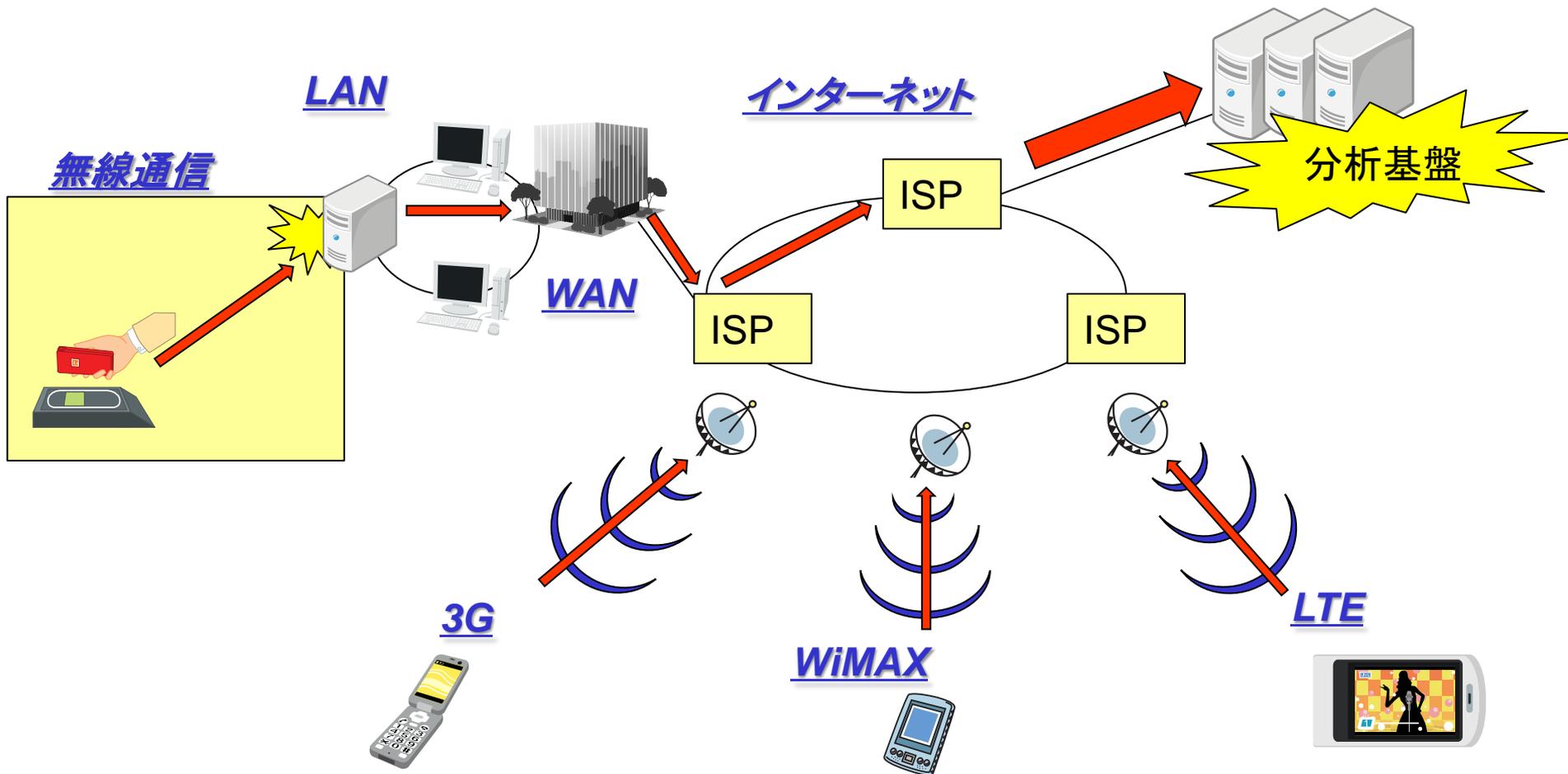
■ 新規データの生成

- ・データ生成デバイス(RFID、GPS、センサ、カメラなど)の導入
- ・情報収集アプリケーションの導入



データの転送

生成したデータを分析基盤に転送



データの蓄積

分析に必要なデータを蓄積

- リレーショナルデータベース
- NoSQLデータベース
- 多次元データベース
- Hadoop

ID	NAME	DEPT_ID
1	村野	10
2	鈴木	20
3	吉田	NULL

DEPT_ID	NAME
10	営業
20	経理

Key		
1	Key	
2	100	
3	101	
	102	

	東京	名古屋	大阪
1月	45	7	7
2月	12	48	13
3月	11	7	9

2012/05/17 16:30 error 89
2012/05/17 16:31 access 10.1.1.1



データの整理

分析ができる形式にデータを整理

■ データクレンジング

- ・コード変換、表記変換、不正データの発見と修正、名寄せ

■ マスターデータ管理(MDM)

- ・異なるシステムで使用するマスターデータを統合

■ 非構造データから構造データへの変換

- ・テキストマイニング
- ・画像認識
- ・音声認識

データの分析

データを分析し、価値ある知見を導出

■ データ分析ツール

- ・データ検索・レポートツール
- ・OLAPツール
- ・データマイニングツール
- ・ポータルツール

■ システム構成

- ・クライアントサーバー型
- ・Webアプリケーション型
- ・クラウド型