

ビッグデータとセキュリティ

2013年2月1日

株式会社ラック
システム事業本部SSU
上野 津代志

1. ビッグデータ事例
2. ビッグデータ技術とは
3. ビッグデータのセキュリティ

1. ビッグデータ事例

検索エンジン (入力：検索キーワード、HTML)

入力補助機能



他のキーワード候補 検索結果の表示順



Web系企業では様々なシーンで要素技術としてビッグデータ技術が使われている。

たべみる (入力：検索キーワード)



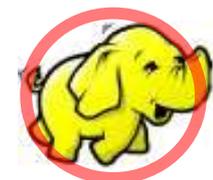
日本全国の「食べたい」がみえる

たべみる



食の検索データサービス

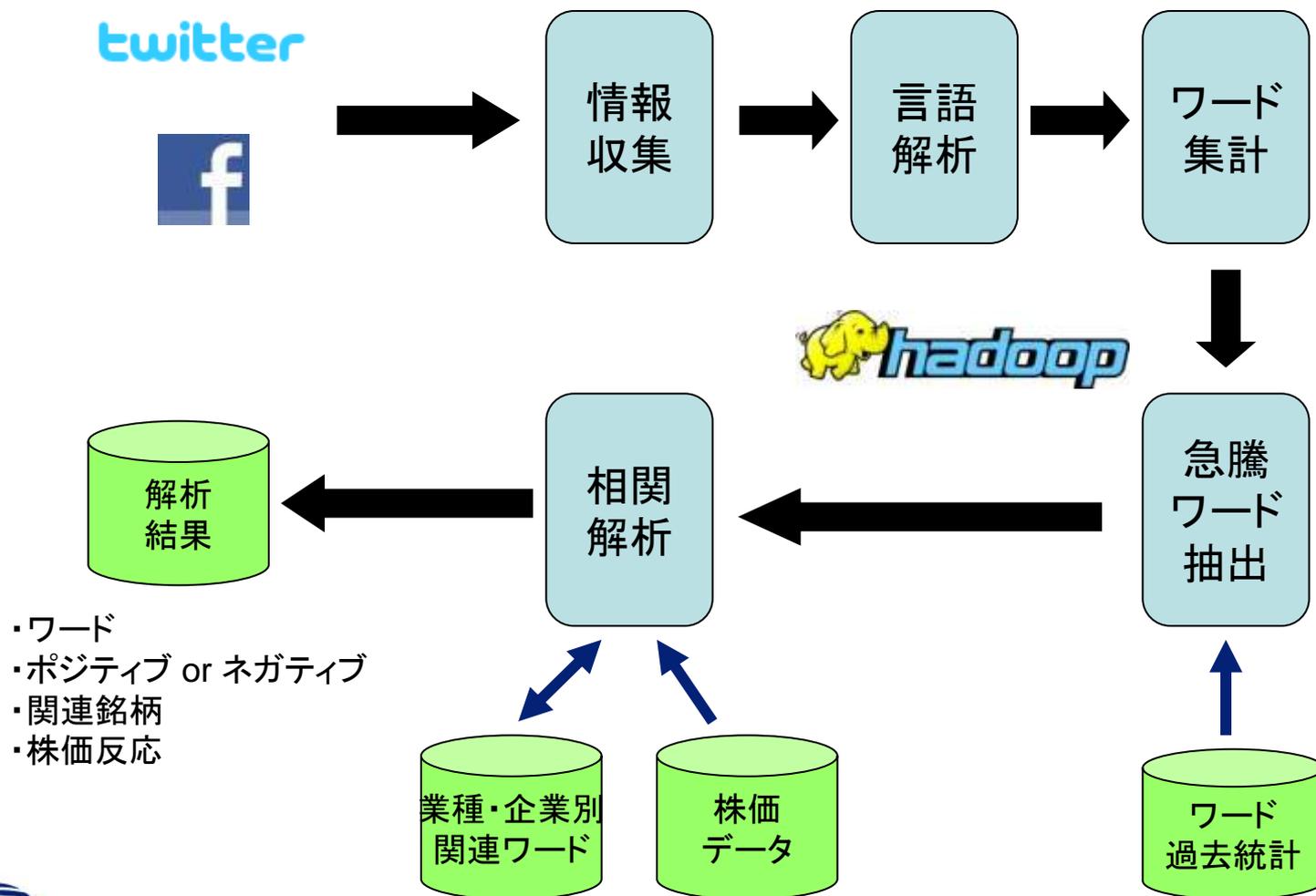
利用料：180万円/年



RDBで推定7000時間が
Hadoopで30時間

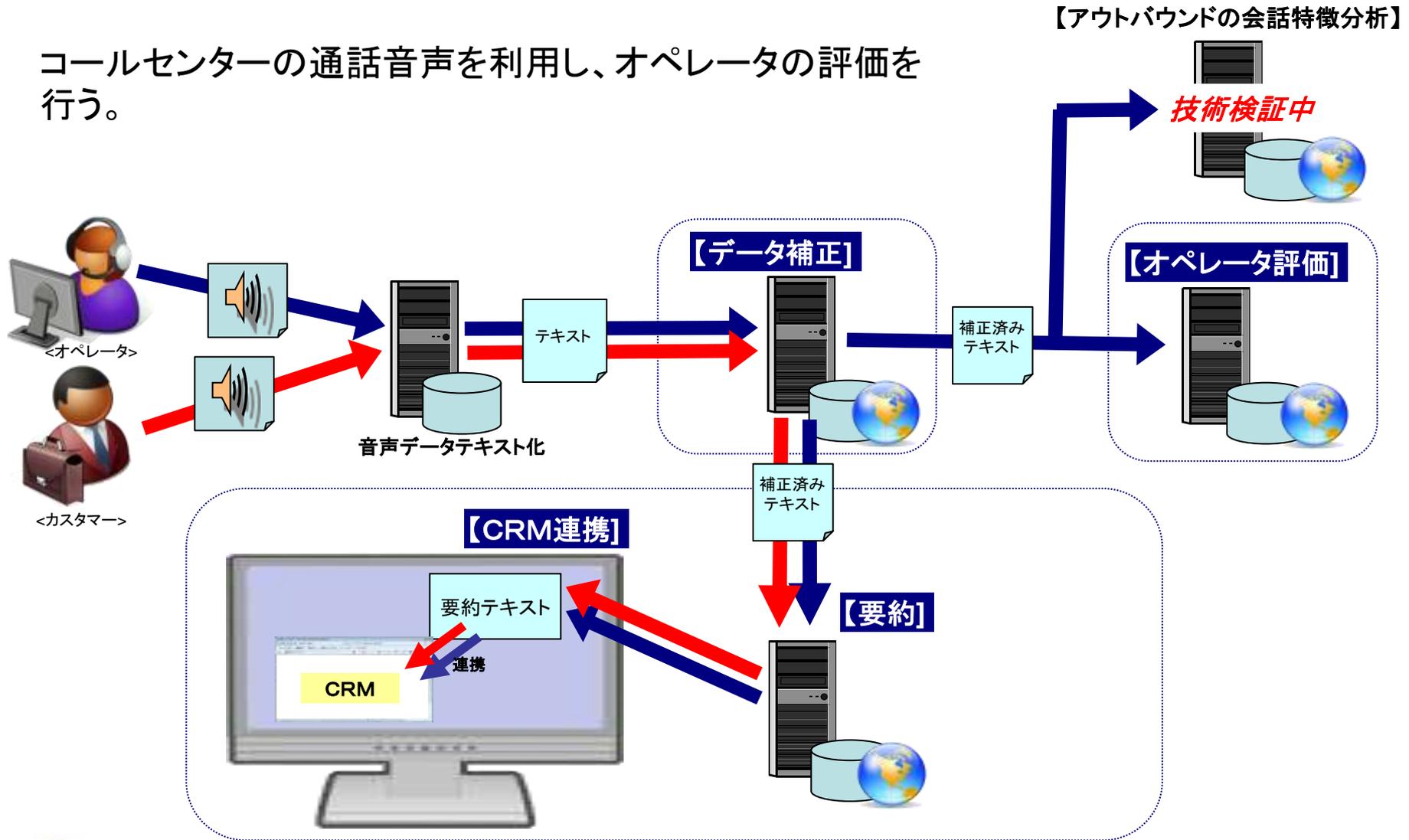
株銘柄関連情報分析(入力:ソーシャルネットワークデータ)

ソーシャルメディアデータから急騰ワード分析、センチメント分析を行い、株式銘柄の株価との相関を分析。



コールセンターオペレータ評価(入力: 音声データ)

コールセンターの通話音声を利用し、オペレータの評価を行う。



サービス解約予測 (入力：顧客行動ログ)

ケーブルテレビ
契約者



毎時サービス利用状況収集

センターA



センターB



サービス利用状況
統計処理

RDBMS

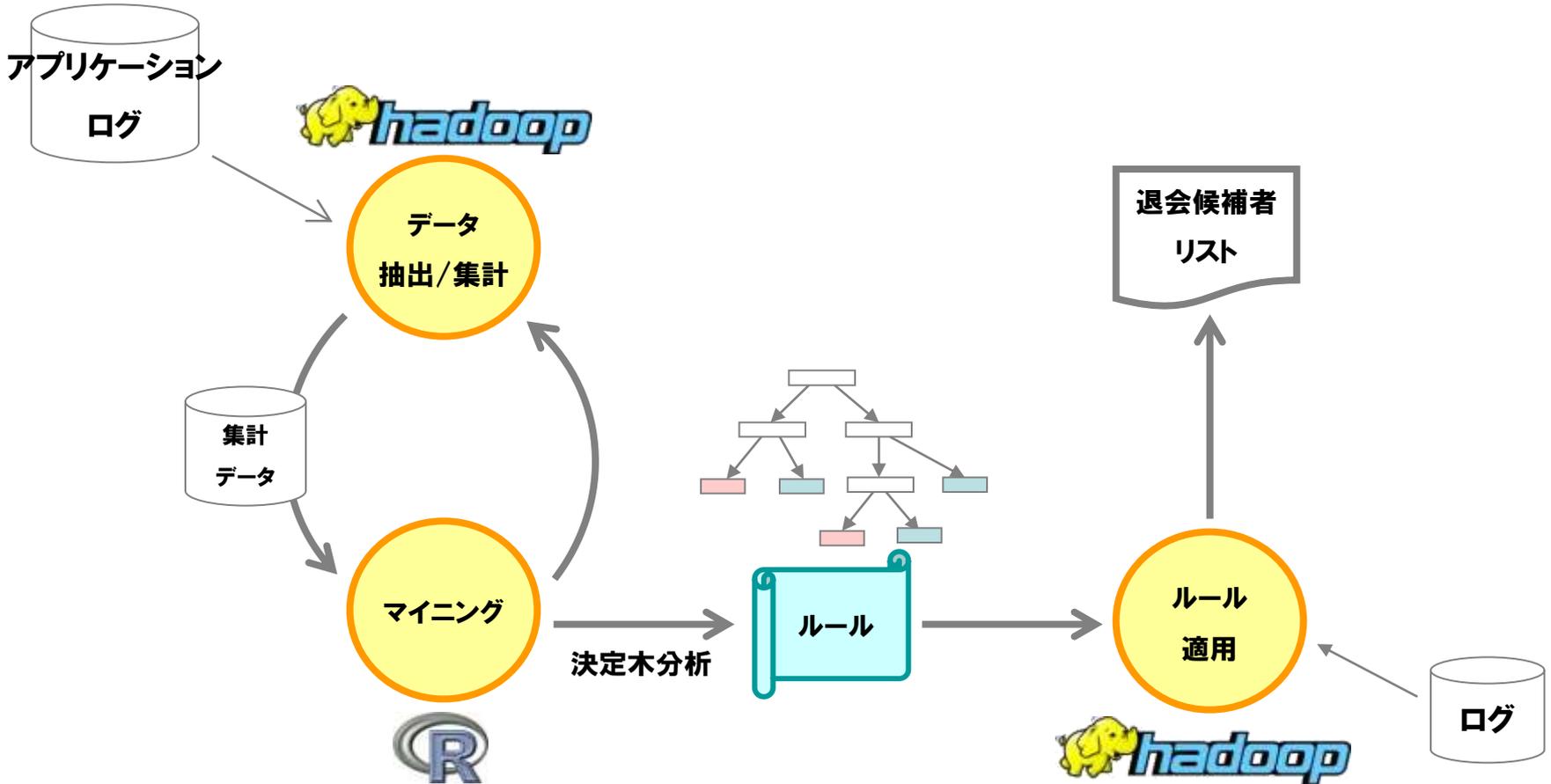
帳票出力



ターゲティング広告
サービス解約予測

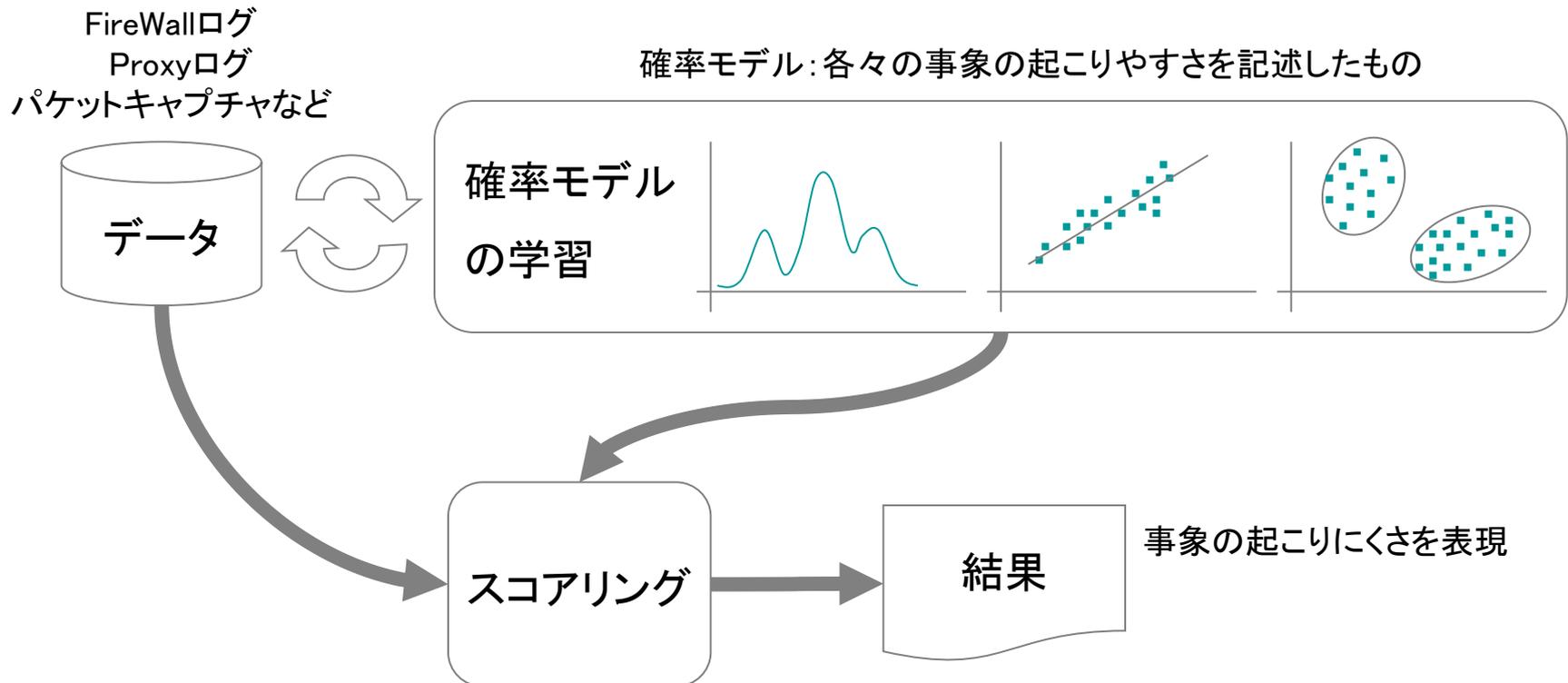
顧客離反分析（入力：Webアプリケーションログ、顧客マスター）

顧客の属性と行動履歴から離反確率の高い「顧客」「行動」を予測、ルールを発見し、予測結果から適切なオファーをすることによって離反を抑止する。



Anomaly Detection(入力：各種サーバログ)

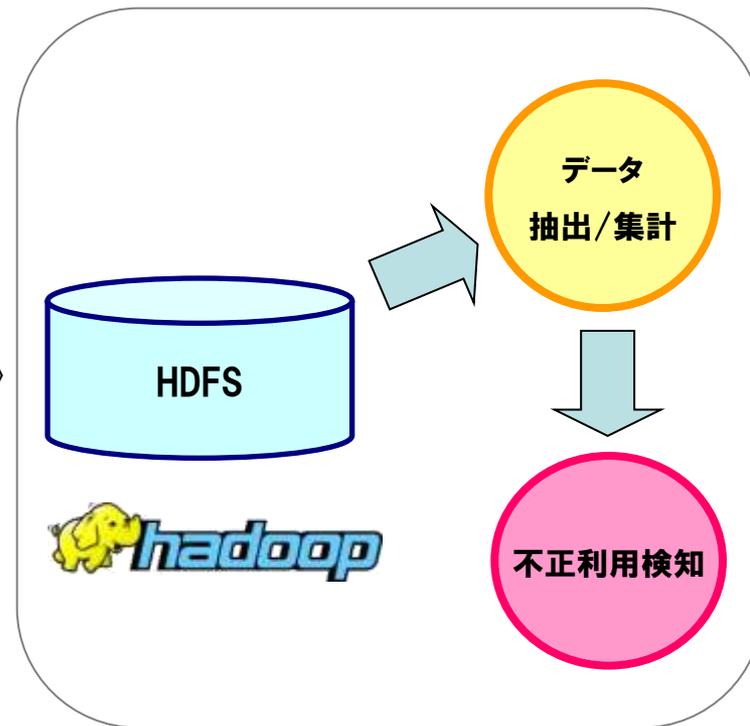
統計的手法を用いて異常検知を行う。具体的には入力データをもとに確率モデルを学習し、確率モデルをもとに異常度合いをスコアリングするという方法をとる。



リスク分析 (入力：口座情報・顧客情報・取引/決済データ)

カード不正利用やマネーロンダリングなどのリスク管理の例

銀行：勘定系システム等



東京



アラート
&
レポートイング

大阪

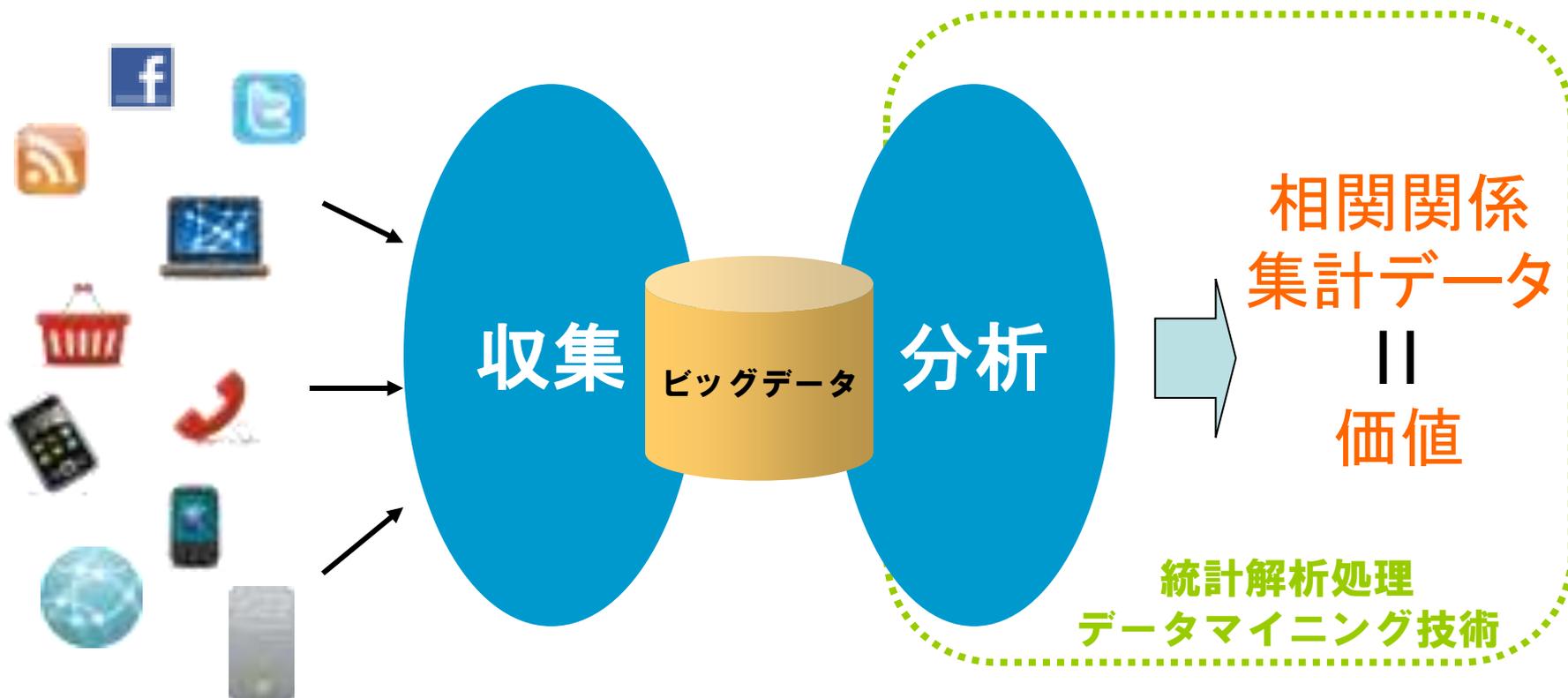


カード：基幹系システム等

2. ビッグデータ技術とは

ビッグデータ処理とは

多種多様で膨大なデータを収集、分析し有用な意味や洞察を導き出す技術



サンプリングではなく全量データ
多種多様なデータの組み合わせ

Volume
Variety
Velocity

ビッグデータ処理基盤

大量データを高速に処理する経済的なシステム



ビッグデータ

※ビッグデータの目安は
3桁ギガバイト

バッチ処理機能

MapReduce
(並列分散処理)

例)  **hadoop**

オンラインデータベース機能

NoSQLデータベース

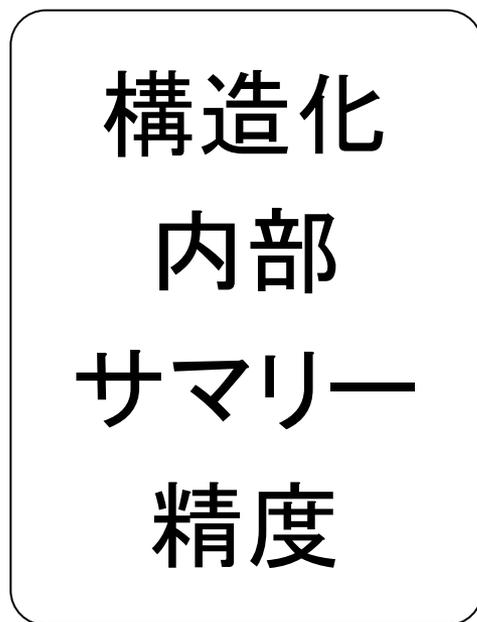
例)  **Cassandra** 

分散処理技術をベースにしたシステム

3. ビッグデータのセキュリティ

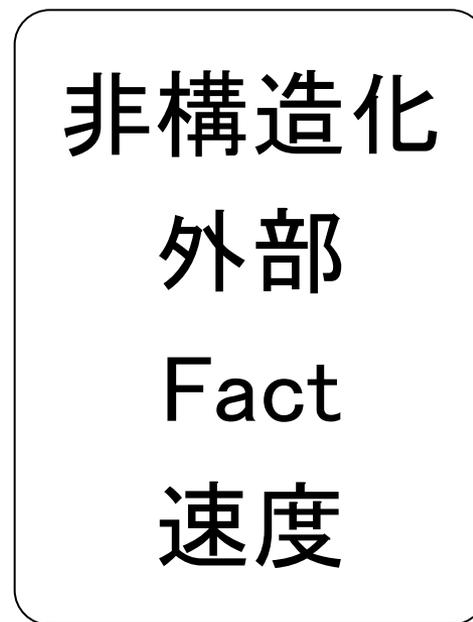
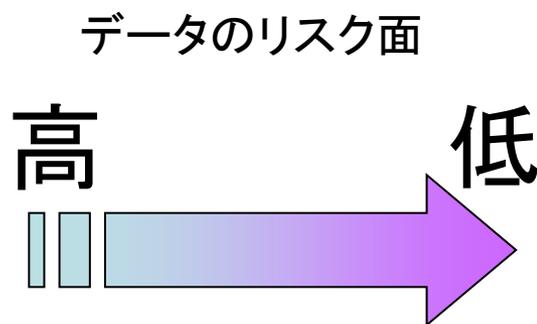
ビッグデータシステムが扱うデータ

非ビッグデータ？



例) 顧客データ 売上データ など

ビッグデータ？



例) 動画 音声 センサー GPS など

ビッグデータシステムが扱うデータ

【事例データ】

構造化

内部

サマリー

精度

HTML

検索ワード

SNSテキスト

顧客行動ログ

Webアプリケーションログ

顧客マスター

各種サーバログ

口座情報

取引/決済データ

音声データ

非構造化

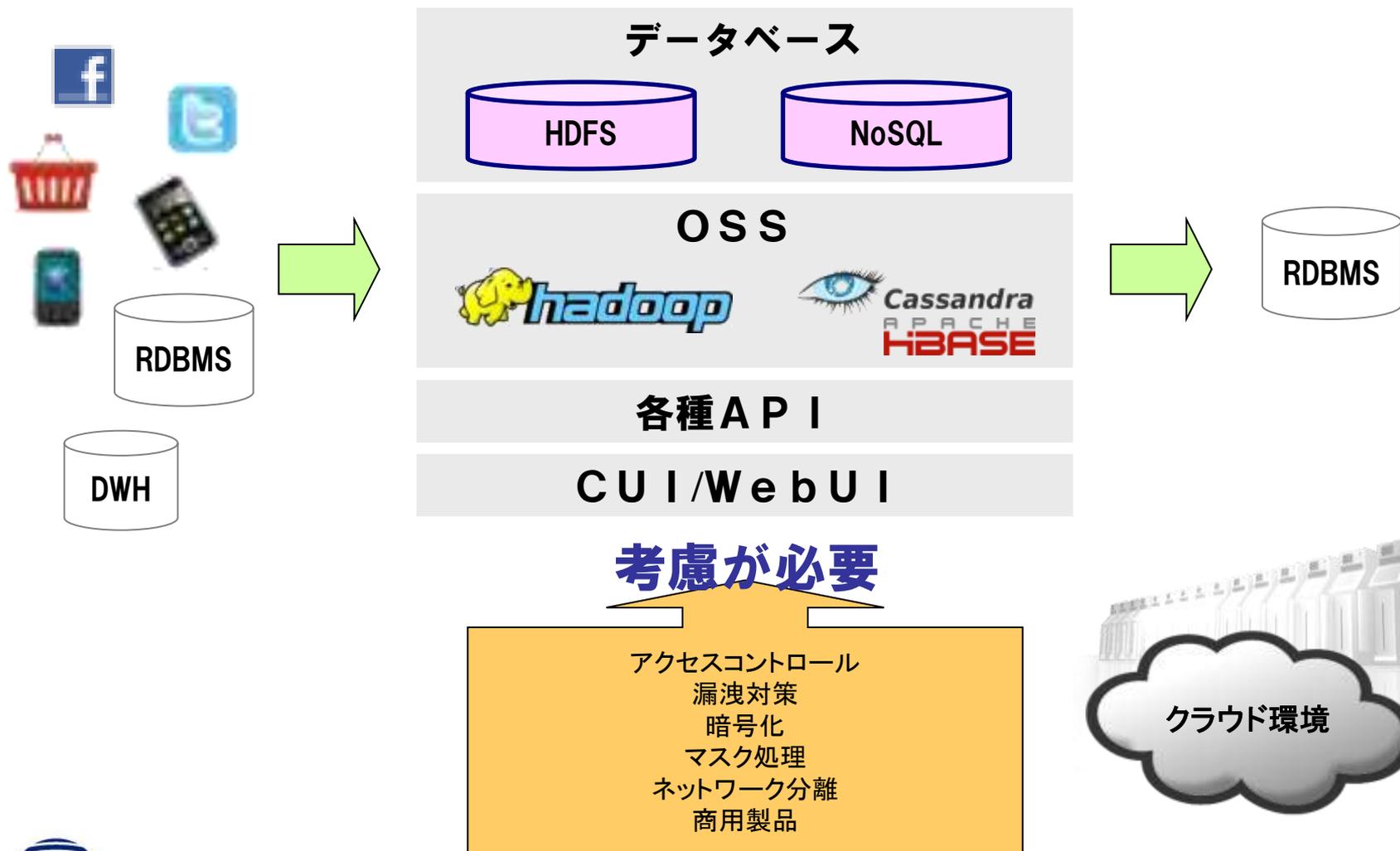
外部

Fact

速度

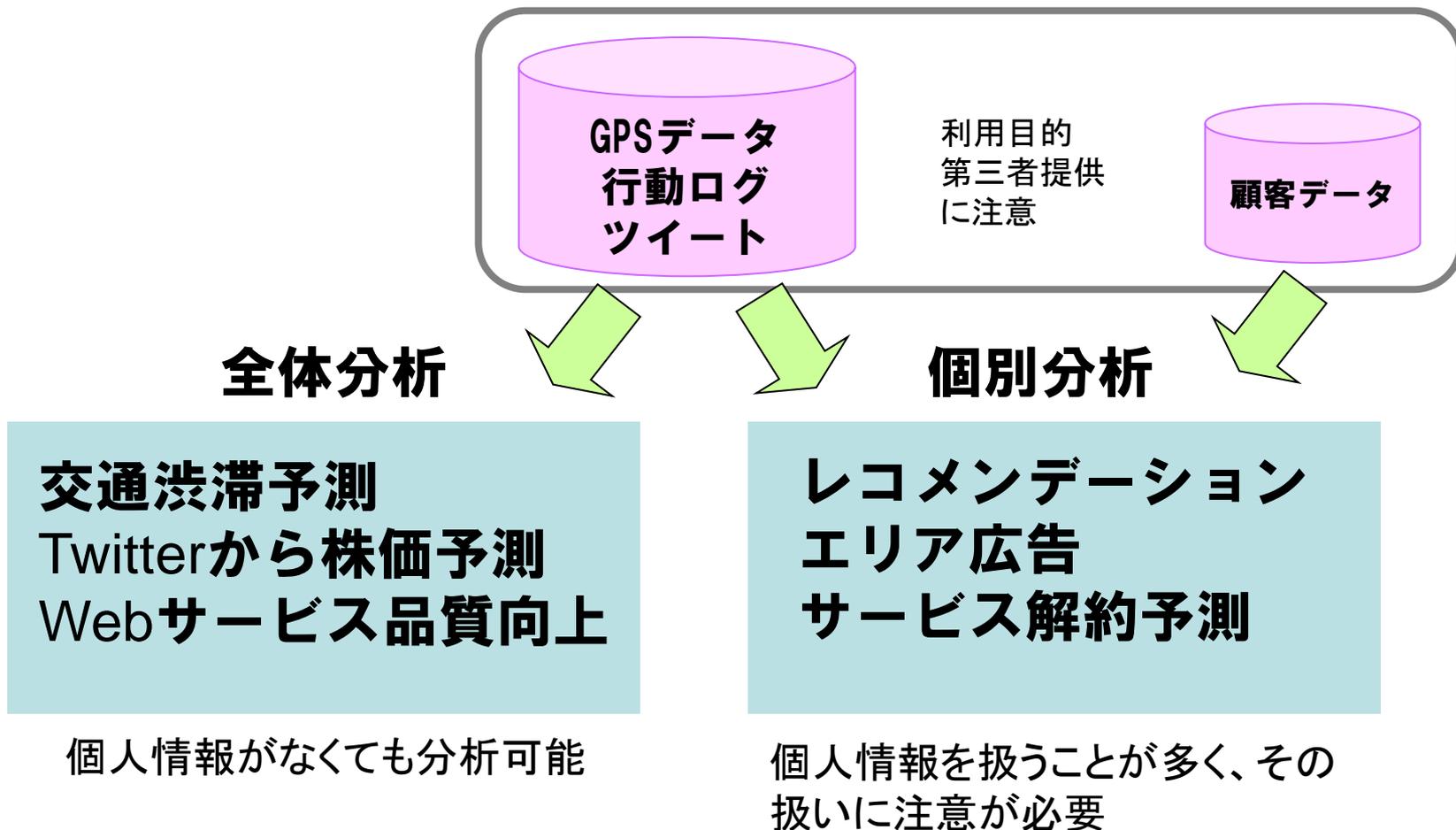
ビッグデータ処理の構成要素とセキュリティ上の課題

ビッグデータシステムの道具立てはセキュリティに関して未成熟



個人情報保護、プライバシーに関する問題

個人から取得したデータを利用する場合は、個人情報保護の観点やプライバシー問題にも考慮が必要



参考) 総務省：利用者視点を踏まえたICTサービスに係る諸問題に関する研究会

まとめ

- **ビッグデータ処理で扱うデータはセキュリティリスクが低いわけではない。**
- **ビッグデータ基盤技術はセキュリティに関して未成熟なため、セキュリティに関する考慮が必要。**
- **ビッグデータ処理で扱うデータおよび分析結果の利用には個人情報保護の観点やプライバシーに関する考慮も必要。**

Thank you. Any Questions ?