

クラウドとビッグデータ

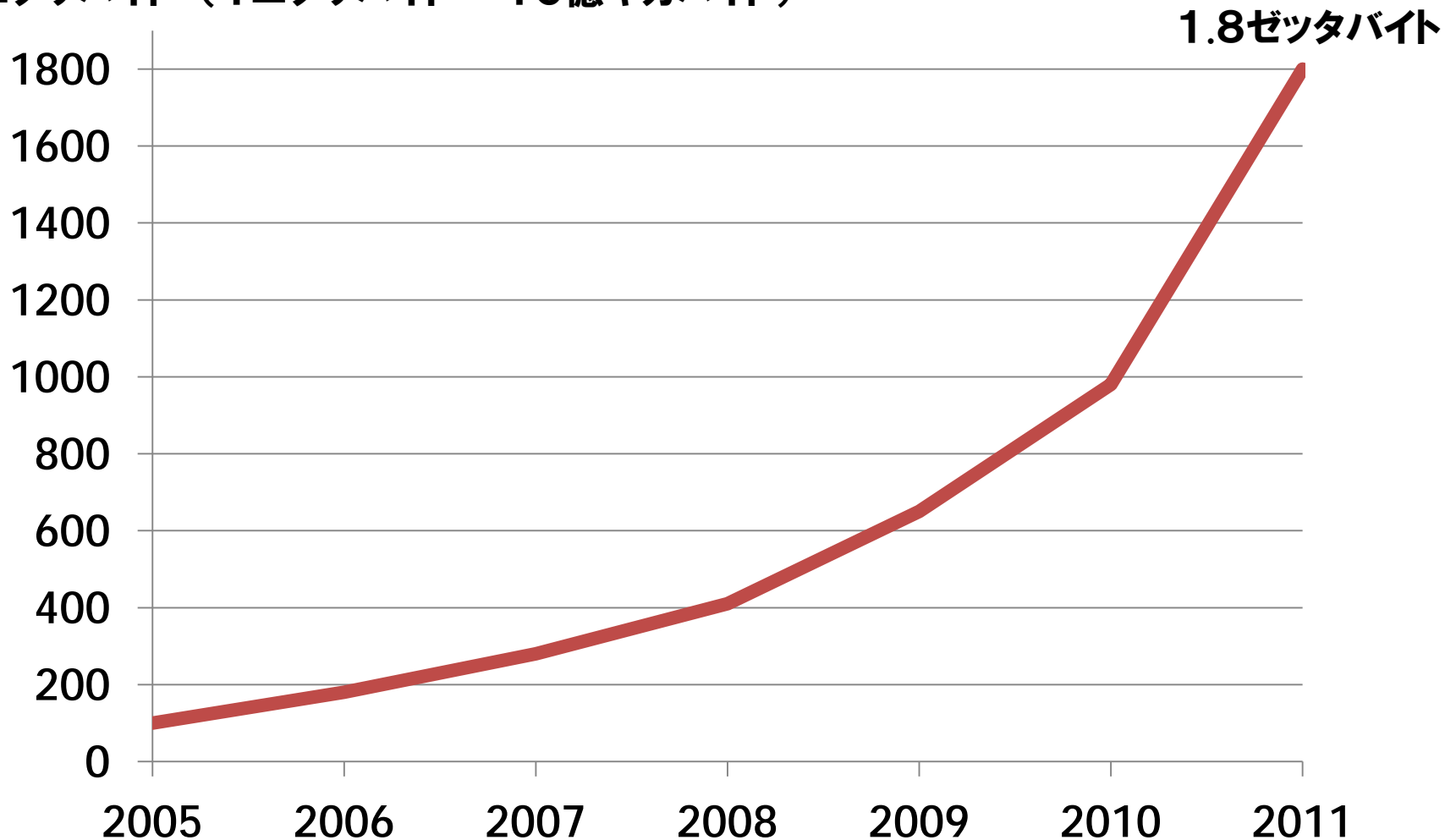
スマートファクトリーオートメーションを
実現するICT技術

NTTアドバンステクノロジー株式会社

遠藤公啓

世界で生成されるデータ量

エクサバイト（1エクサバイト = 10億ギガバイト）



IDC, "The Expanding Digital Universe" に基づいて作成

ビッグデータは量だけではない

■ 大量

- 数百テラバイト以上

■ 非構造

- テキストメッセージ、音声、画像、動画、センサデータ、ログ、…

■ リアルタイム・高頻度

- Twitterのつぶやき、スマートフォン、販売履歴ログ、センサ、RFID、…

データ量の恐怖

- 1台の機械から毎秒1Kバイトのデータ生成
- 1工場に1000台の機械・装置
- 1年間
- 100工場

3,153,600,000,000,000bytes
=3.15Pbytes

なぜ今ビッグデータなのか

■今までも

- データマイニング
- データウェアハウス
- ビジネスインテリジェンス

■しかし今は

- (本当の)ビッグデータが生成されるようになった
- ビッグデータを処理する技術が確立されてきた
- ビッグデータは宝の山であることがわかってきた

スマートグリッドとビッグデータ

- 家庭や企業の電力消費量をリアルタイムでモニタリングし、発電所や送電網を細かく調整する
- 数百万から数千万のスマートメーターから電力消費量が送信されてくる
- ビッグデータの処理が不可欠

ビッグデータの活用例

- **イタリアの電力会社 ENEL**
 - **スマートグリッドの例**
 - **契約家庭2700万世帯にスマートメーターを設置**
 - **電力使用量を15分ごとに送信**
 - **時間帯ごとに異なる料金体系を適用できるようになった**
 - **検針員の削減と発電所の効率的な運用により年間7億5000万ドルのコスト削減**

日経ビジネスオンライン、「イタリアは世界一の“スマートグリッド先進国”」
<http://business.nikkeibp.co.jp/article/world/20091120/210286/>

ビッグデータを扱うには

■現在のサーバのスペック

- ハードディスクの最大容量 3テラバイト
- CPUの最大速度 3.8GHz

■1台では蓄積・処理できない

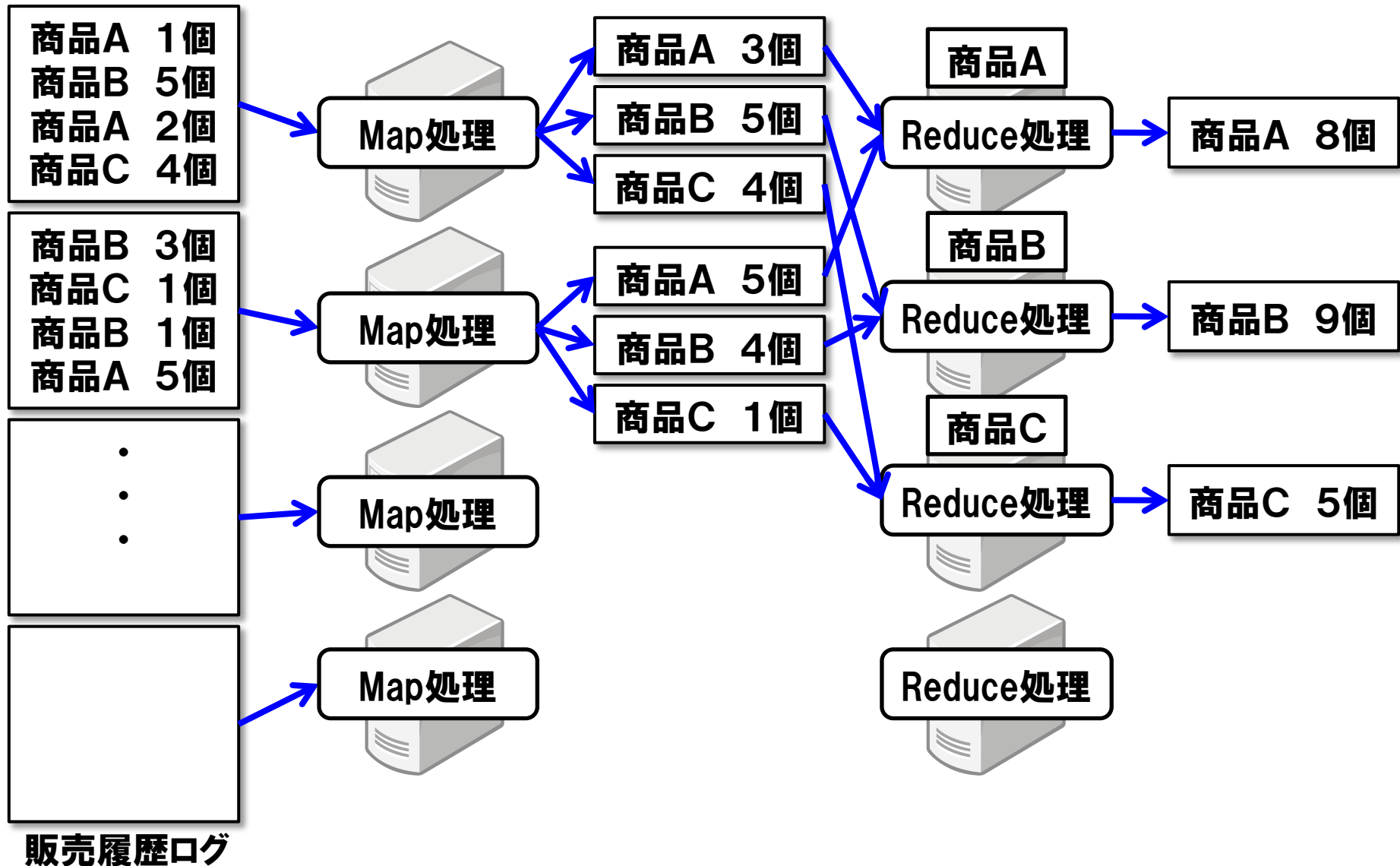
■安価なサーバを大量に用意し分散蓄積、分散処理するというアプローチ

- Googleはサーバを90万台以上保有

MapReduce

- **大規模分散処理アルゴリズム**
- **Googleが開発し、2004年に論文を発表**
- **Map処理とReduce処理**

MapReduceの仕組み



Hadoop

- 読み方は「はどぅ〜ぶ」
- MapReduceの実装
- オープンソースソフトウェア



MapReduce/Hadoopの何がよいのか

- スケーラビリティが高い
 - 数千ノードでも動作し、性能がほぼ線形に向上
- 障害に強い
 - ノードの障害を検知したら別ノードに処理を割り当て
- 並列処理の面倒なところをしてくれる
 - データの分割、ノードの割り当て、データの集約
 - Map処理とReduce処理を作るだけでよい

Hadoopの利用事例

■ VISA

- 不正なカード利用を発見するため、クレジットカードの取引情報を分析
- 2年分730億件(36テラバイト)の取引情報を分析
- 従来1ヶ月かかっていた
- Hadoop導入により13分

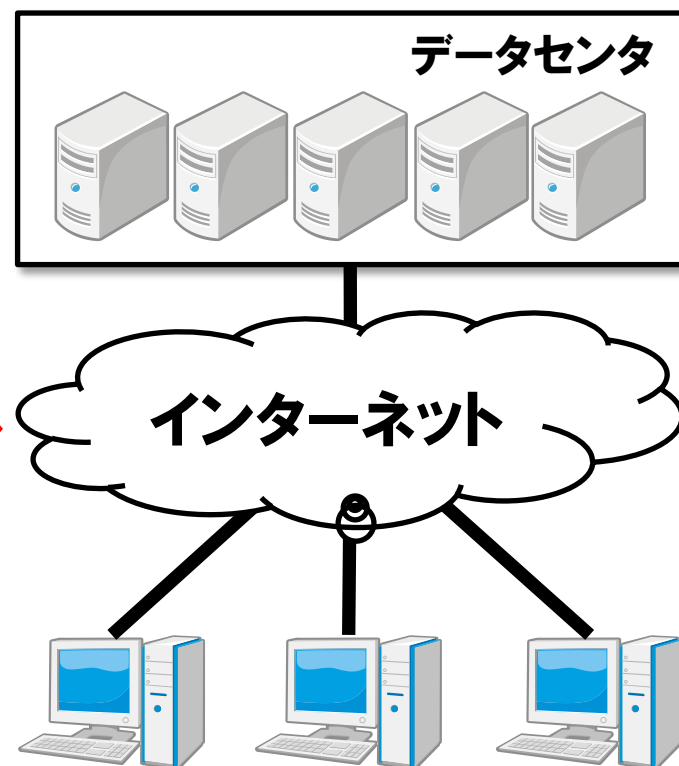
Hadoop World 2009, "Large Scale Transaction Analysis"

<http://www.slideshare.net/cloudera/hw09-large-scale-transaction-analysis>

クラウドコンピューティング

- Google 元CEO エリック・シュミット氏が提唱
- クラウド=雲
- インターネットの向こう側に
あるコンピュータを必要な
量を必要な時だけ利用する

これが雲 →



クラウドの種類

ソフトウェアを利用したい

- **SaaS型 (Software as a Service)**
 - ソフトウェアをレンタルする

自分の好きなソフトウェアを実行したい

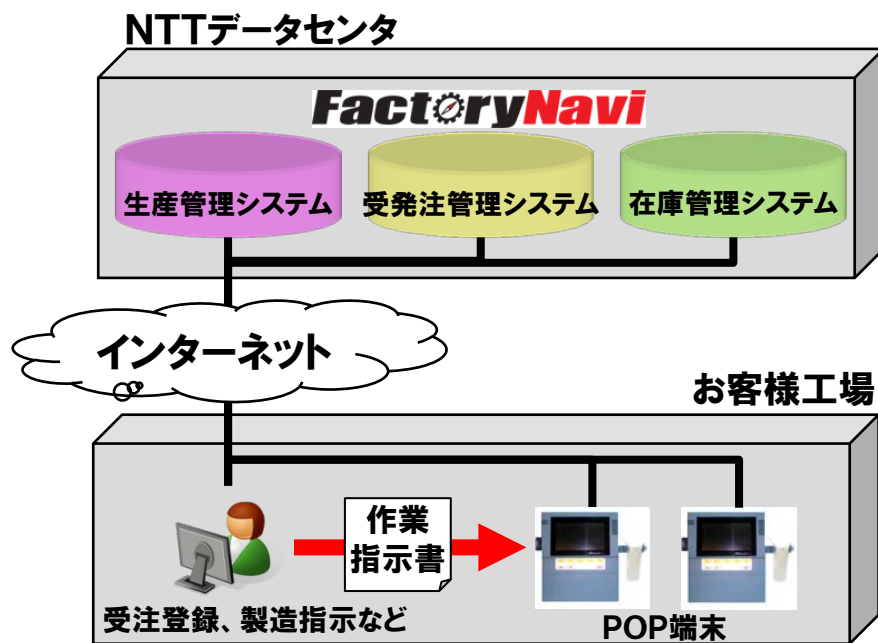
- **PaaS型 (Platform as a Service)**
 - ミドルウェアをレンタルする
- **IaaS型 (Infrastructure as a Service)**
 - 実行基盤をレンタルする

SaaS型クラウドの例

工場経営改善ナビゲーションシステム

FactoryNavi

- クラウドで提供する中小製造業向け基幹業務アプリケーション
- 月額利用料金26,040円(税込)から



FactoryNavi

検索

IaaS型クラウドの利用事例

■ ニューヨークタイムズ

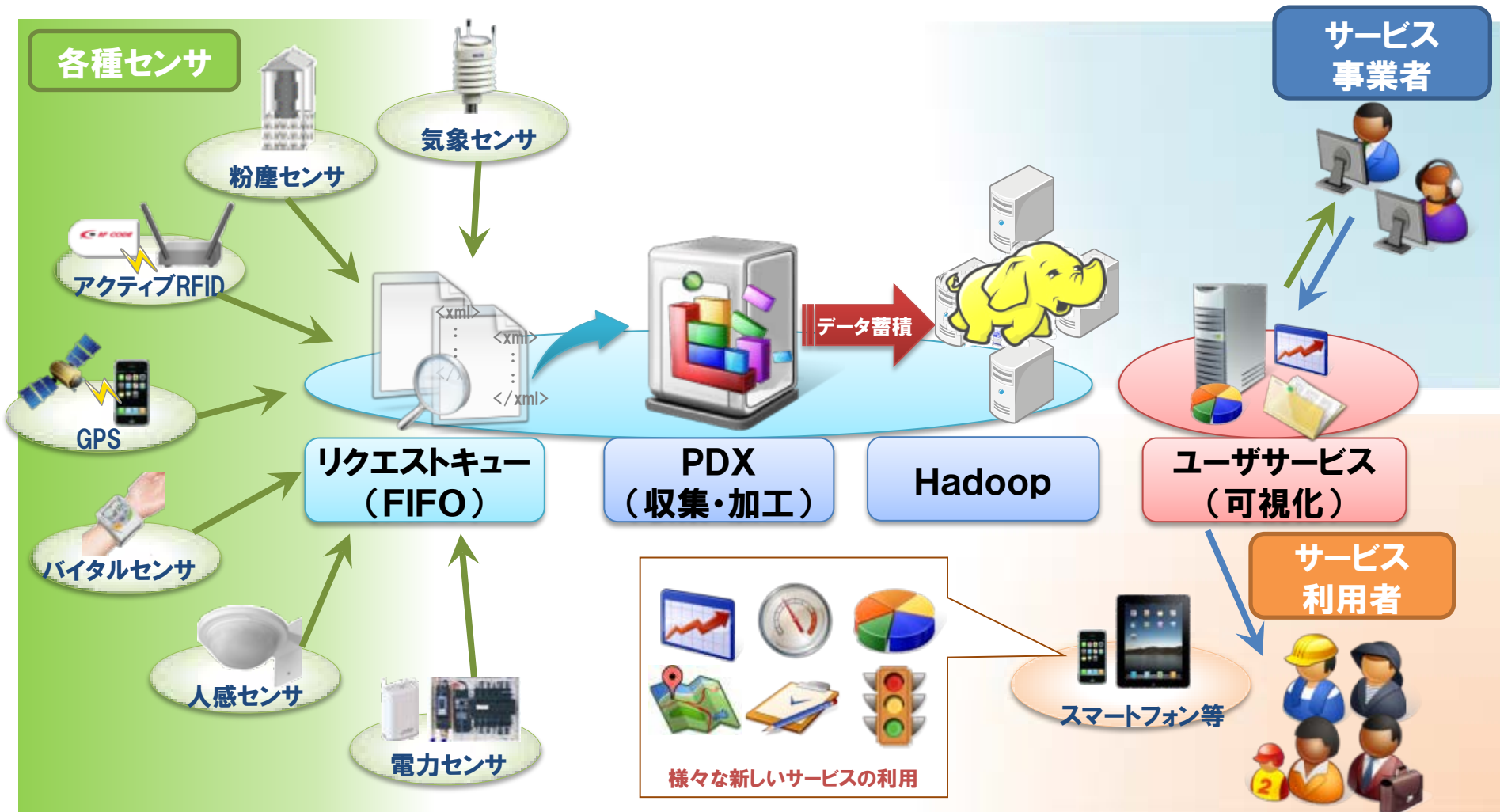
- 1851年から1980年までの過去記事の1100万枚の画像(4テラバイト)のPDF化
- Amazon EC2を利用
- 100インスタンスを使用してHadoopで処理し、24時間で完了
- かった費用はわずか240ドル

The New York Times, "Self-Service, Prorated Supercomputing Fun!"
<http://open.blogs.nytimes.com/2007/11/01/self-service-prorated-super-computing-fun/>

NTTアドバンステクノロジーのビッグデータへの取組み

超大量計測データ分散処理基盤

- 電力管理、防災、環境、見守り、医療、...



スマートファクトリーオートメーション

サプライサイド
情報

デマンドサイド
情報

クラウド

ビッグデータ処理



スマートグリッド



スマートコミュニティ



サプライチェーン

稼働情報

エネルギー
消費情報

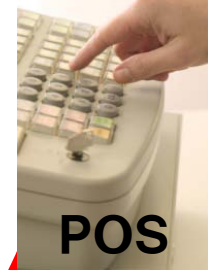
製造計画
生産指示
運転調整



スマートファクトリー

Twitter

Facebook



POS

販売履歴
ログ

おわりに

- ビッグデータは既にそこにある
- ビッグデータを扱う技術が確立されてきた
- これからはビッグデータの活用の段階

- スマートファクトリーオートメーションの実現には
ビッグデータ処理が不可欠
- そしてその鍵はクラウドの活用にある