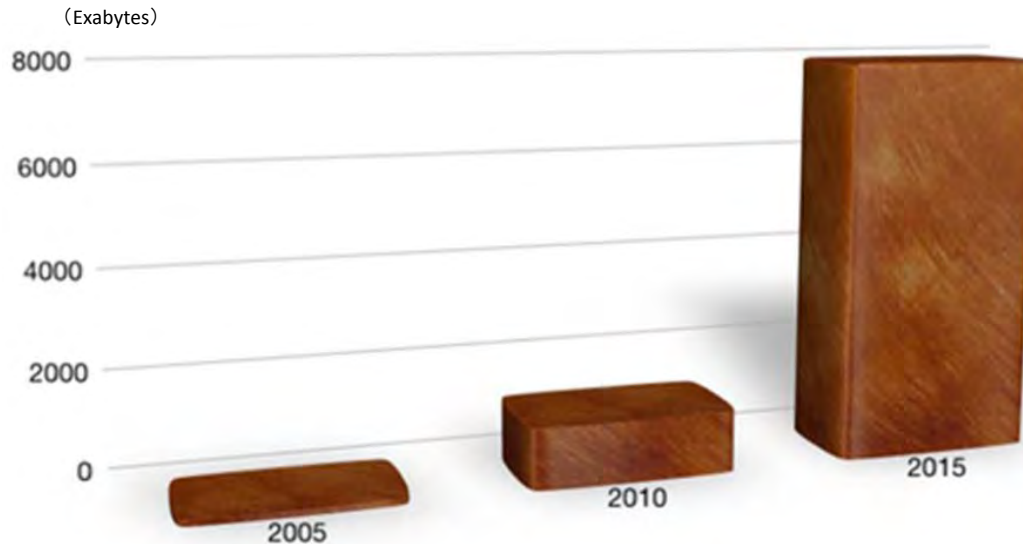


- 国際的なデジタルデータの量は、2011年の約2ゼタバイト(2兆ギガバイト = 2千エクサバイト)から約4倍増加し、2016年には約8ゼタバイトへ拡大する見込み。
- ビッグデータの活用により、例えば、米国ヘルスケアで年間3千億ドル、EU公共セクターで年間2.5千億ユーロ、位置情報データの活用により年間6千億ドルの消費者価値創出等が期待。

## 10年間のデジタルデータの成長



【出典：IDC「2011 Digital Universe Study: Extracting Value from Chaos」(平成23年6月)】

## いわゆる「ビッグデータ」の定量的価値(例)

50億台の携帯電話が使用(2010年)

300億のコンテンツが毎月Facebook上で共有

IT費用の5%増加で、年間40%増のデータ創出

米国のヘルスケアでは年間3000億ドルの価値創出が期待(スペインの年間ヘルスケアコストの2倍)

EUの公共セクターでは年間2500億ユーロの価値創出が期待(ギリシアのGDPを超える)

個人の位置情報データを活用することで年間6000億ドルの消費者価値創出が期待

小売の営業利益に60%改善の見込み

【出典：McKinsey Global Institute「Big data: The next frontier for innovation, competition, and productivity」(平成23年5月)】

## NoSQL/Not only SQL

- ▶ 表形式によるリレーショナルデータベース管理システム(RDBMS)とは異なる設計によって実装されたデータベースシステム
- ▶ RDBMSが定型データの処理を必要とする業務システムでの利用に適しているのに対し、NoSQLはセンサーやソーシャルメディア等の非定型データを含む多様なデータを大量にデータベース化するために利用

## Hadoop

- ▶ 米国NPOのApacheソフトウェア財団のプロジェクトで開発が進められている、大規模データの効率的な分散処理等のためのオープンソースソフトウェアフレームワーク
- ▶ 複数のサーバを通じた並列処理により、柔軟かつ継続的な大規模データの高速処理が可能



## クラウドサービス

- ▶ 利用者が必要なコンピュータ資源を「必要な時に、必要な量だけ」利用でき、拡張性、可用性、俊敏性や経済性等の特徴を有するサービス
- ▶ クラウドサービスの利用により、多種多量のデータの蓄積や計算処理のために必要となる多数のマシンについて、自前で用意する必要がなく、低コストで同様の環境の構築が可能
- ▶ また、例えば、「Hadoop」稼働環境を提供するサービスを利用すれば、マシンの調達に加え、ソフトウェアのインストール等の設定作業の省略も可能

## DWH (Data Ware House)

- ▶ 定型データ・非定型データを問わず、大量データの蓄積を目的とするデータベースの総称
- ▶ これらの大量データを高速に処理する方法により、
  - ①従来のRDBMSとは異なる設計技術によるNoSQLデータベースと、
  - ②標準的なRDBMSとハードウェアレベルでの高速化技術を組み合わせたDWHアプライアンスの2つに大別

## CEP (Complex Event Processing)

- ▶ データをディスクに格納せずに、書込速度がディスクに比べて高速なメモリ上で逐次的に処理することにより、必要な情報をリアルタイムに抽出する技術
- ▶ ディスクにデータを蓄積して分析する手法と比べ、短時間で処理が可能であるため、クレジットカードの不正利用や防犯カメラ映像の異常検知等、短期間での対応が必要となる場合に利用
- ▶ 予め利用者が定義するリアルタイム処理の内容には、単一のデータ属性の閾値による判別に加え、複数の属性を組み合わせた処理の設定も可能

## MDM (Master Data Management)

- ▶ 業務を遂行する際の基本情報である顧客情報等のマスターデータを管理するためのシステム
- ▶ 多種多様な情報システムに重複・散在し、かつ、多量に生成等されるマスターデータについて、常時の最新状態への更新やシステム全体の整合性の確保等が可能

## PPDM (Privacy Preserving Data Mining)

- ▶ プライバシーを保護した上で、大規模なデータから特徴や規則性等を抽出する技術
- ▶ 匿名化や秘密計算等により、個々のデータを暗号化したままデータマイニングを実施することで、個人情報流出等のリスクを回避した上で、データの解析等が可能
- ▶ 例えば、どのデータについても同じものがk件以上存在するようにデータの粒度や曖昧さを制御するk-匿名化技術等が研究

## 秘密計算

- ▶ 入力データや演算ロジックを暗号化したままで任意の計算を可能にする技術
- ▶ 複数のコンピュータにデータの断片を送り、断片の部分計算を繰り返し行うことで、データを秘匿したまま統計などの各種計算を行うことが可能
- ▶ PPDMでの利用の他に、民間企業、公共機関、教育現場などにおけるプログラムの不正解析防止、知的財産の侵害防止、情報漏えい防止などの多様な分野への応用が期待