

# ビッグデータの登場



左の画像はNHKが東北の震災時にツイートのデータを加工分析したものです。

2016年2月 上田ヤマト作成



# ビッグデータのメニュー

0. コンピューターの発達
1. スマートフォンの出現
2. インターネットの拡大
3. データの量と質
4. ビッグデータとは何か
5. ビッグデータの処理、分析
6. 企業への活用
7. 今後の展望



0. コンピューターの発達

# ビルゲイツ

1955年シアトルで裕福な家庭に生まれる

高校でCOBOL言語で給与計算ソフト作成

ハーバード大学でBASIC4言語作成  
マイクロソフト社創設

1980年IBMの為にPC-DOSを開発

1990年WINDOWSを発表、世界一位となる

現在57歳、世界一の金持ち、慈善活動を行う。



0. コンピューターの発達

# スティーブ・ジョブズ

1955年シリア人の男性とアメリカ人の母を大学院生とのあいだにサンフランシスコに生まれた。

1972年養父先から大学に入学するが座禪に心酔

1976年APPLE Iをウオズと作成、販売

1977年APPLE IIを発売

1987年Macintosh発売

1985年APPLE追放

2000年APPLE社社長に復帰

2003年膵臓癌発症

2007年IPHONE発売

2010年IPHONE 4発売

2011年10月4日IPHONE 4S発表

翌日死去



0. コンピューターの発達

# ウィンドウズ

- 1992年 — — — — Windows 3.1
- 1995年 — — — — Windows 95
- 1998年 — — — — Windows 98
- 2000年 — — — — Windows Millennium, 2000
- 2002年 — — — — Windows XP
- 2007年 — — — — Windows Vista
- 2009年 — — — — Windows 7
- 2012年 — — — — Windows 8
- 2016年 — — — — Windows 10

## 1. スマートフォンの出現

# APPLE関連

- 2001年1月 ITUNES発表
- 10月 IPOD 発表
- 11月 IPOD 日本発表
- 2003年 WINDOWS ITUNES 発表
- 2005年 ipod shuffle
- ipod Mini
- 2007年 Iphone 発表
- 2010年 4月 Ipad米国発表
- 5月 Ipad日本発表
- 2011年 11月 iphone4S発売 (日本も同じ)
- 2012年 9月 iphone5発売 (日本も同じ)
- 2013年 9月 iphone5S発売

## 1. スマートフォンの出現

# I p h o n e で 可 能 な こ と

インターネット

メール

写 真

Y o u t u b e

M a p

P a p e r N e w s

文 庫

電 話

ゲーム

ワ-フ口

表 計 算

パワ-ポイント

音 楽

T w i t t e r



# 1. スマートフォンの出現

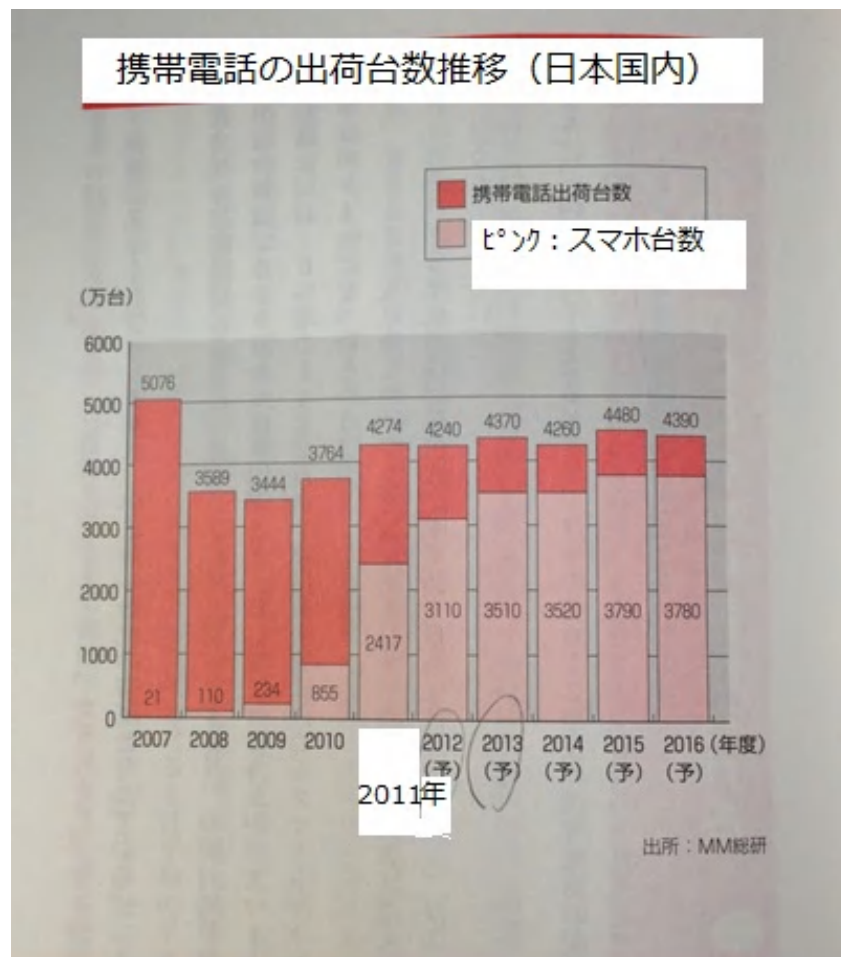
# IPAD - 雑誌





## 1. スマートフォンの出税

# モバイルデバイスの普及



○データが急速に増えた「**第一の原因**」は、インターネットに接続する端末がふえているから。2011年に世界で40億台が2015年には150億台に達すると見られる。

○サーバー、外付デスク、PC、タブレット、スマートフォン、ゲーム機、テレビ、デジカメ、家電等

○中でもスマートフォンの広がり、メールの送受信、ツイッター、フェイスブックへの写真や動画の投稿、GPS連動のアプリの利用、つまり個人が情報を持って移動し、多くの情報を発信するという使い方が広がっている。

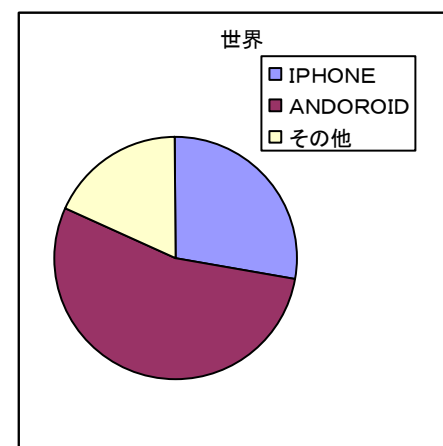
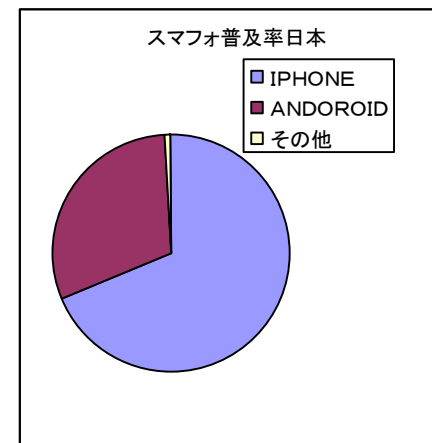
○携帯電話は2011年までは半々だったが、2012年以後スマートフォンが急速に増え、そのうち40%がツイッター等を使うと見られる。

## 1. スマートフォンの出現

# スマートフォンとは

- スマートフォンとは高機能電話でこれまでの携帯電話にできなかった事が出来るようになった。スマートフォンには大きく分けて3種があります。2007年にアメリカのAPPLE社にて初めてIPHONEを発売しました。

- | 普及率         | 日本    | 世界    |
|-------------|-------|-------|
| 1) IPHONE   | 68.7% | 27.9% |
| 2) ANDOROID | 30.5% | 53.7% |
| 3) その他      | 0.8%  | 18.4% |
- (世界ではアジアがandroid多く、EUにIphone多い-2013年の調査)



## 1. スマートフォンの出現

# ますます進むスマートフォン

1. 世界の携帯電話にスマホが占める割合: 2011年3割(4億7千万台)、2015年6割(12億台)
2. スマートフォンとは携帯電話+パソコンのようなもので@外出先でも調査、注文、電話が出来る。
3. パソコンは自宅か事務所で使う: 点の作業、スマホは移動中でも使える: 面で利用出来る。

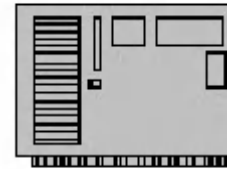


# 第3期のコンピュータ時代

○第1期のコンピュータ：**汎用機時代**  
第2期は**分散PC時代**、

○第3期は**クラウド、スマートフォン、SNS、ビッグデータ**からなっている。  
クラウドにより投資額を抑えて、短期間に膨大なビッグデータ収集と分析を可能とする。

第一段階  
汎用機



第二段階  
PCインター  
ネット



第三段階

クラウド  
モバイル  
ソーシャル  
ビッグデータ



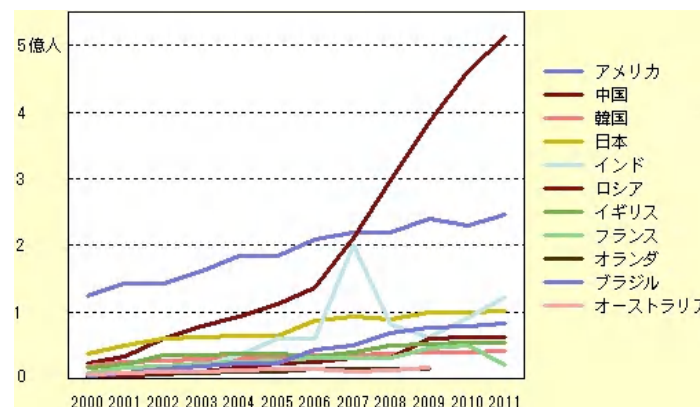
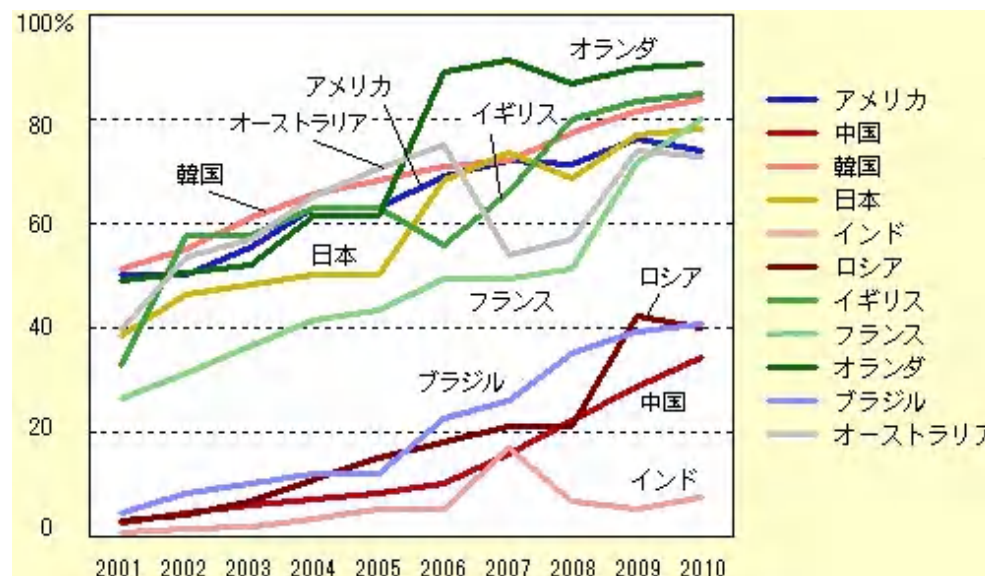
## 2.インターネットの拡大

# インターネットの普及

2013年インターネット普及率は1位アイスランド：96.55%、6位オランダ93.96% 16位日本86.25%

中国は普及率は89位：45.8%だが人口別では5億人を超え、2位アメリカ2.5億、3位インド1.2億、4位日本1億人です。

2007年スマホ発売以後急速にインターネット人口が爆発的に増加





## 2. インターネットの拡大

# SNS（ソーシャルネットワークサービス）

- **社会的ネットワークをインターネット上で構築するサービス）**  
**社会的ネットワークとは価値、構想、提案、金銭、友人、親類、嫌悪、取引、病氣、航空等1つ以上の関係で結ばれた関係から成る社会的構造である。**

**1. 日本最大会員数：MIXI、2010年日本で2000万人加入**

**2. 最近話題は：Twitter 2006年開始：2009年日本で320万人（世界1億人）、2012年世界4億6500万人、2015年日本2100万人**

**3. 世界最大会員数：FACEBOOK 2010年全世界5億人（日本ではまだ少ない200万人）、2012年世界で10億人、2015年日本で1050万人**

**4. LINE：2012年7月発表、2013年8月世界で2億3000万人（一時間に6万3000人増加）、2015年世界で4億人、日本で5000万人**

## 2.インターネットの拡大

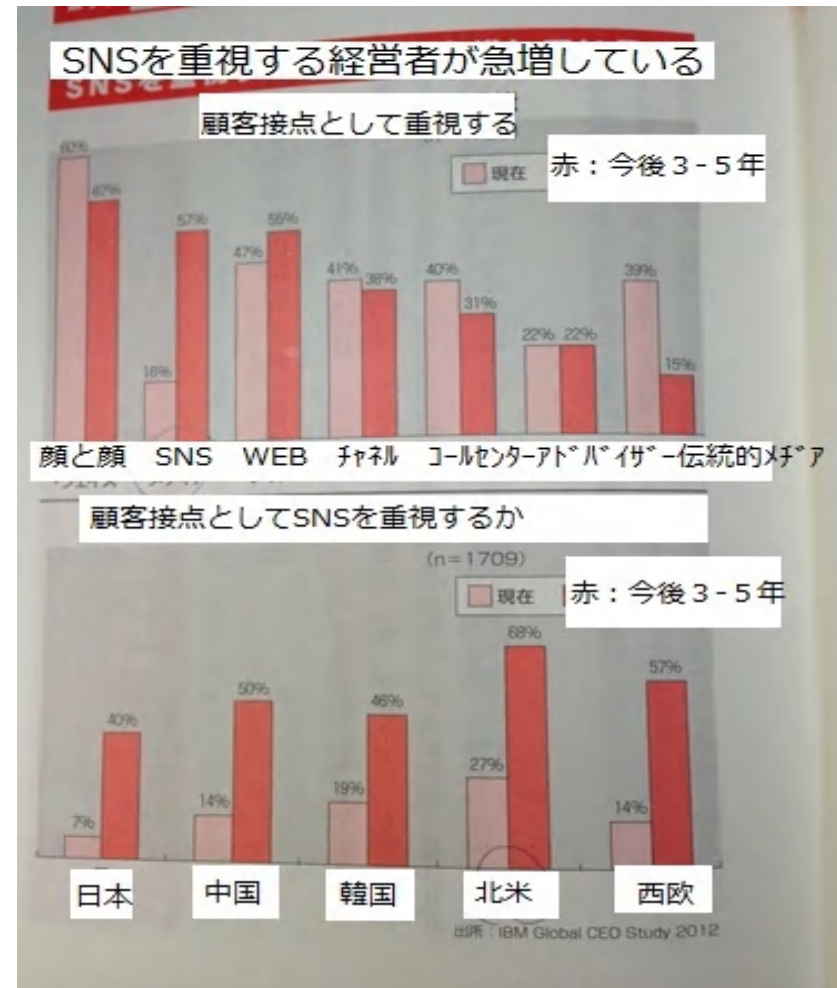
# 世界のツブヤキが収集される

○データ増加の第2の原因はソーシャルメディアの拡大だ。現代利用者は世界で45億人となっている。

○ツイッターは5億人、フェイスブックは9億5千万の人に達している。

○日本でもソーシャルメディア利用者は2012年5月で5千万人になっている。

○企業への調査で「顧客との接点を持つために重視するもの」2番目にソーシャルメディアとなっている。



## 2. インターネットの拡大

# インターネットへのデータの蓄積

1. SNSの台頭、ツイッターは世界で毎日4億件、facebookは毎日25億件(500TB)の文字や写真が蓄積され続けている。

2. クラウドの出現でパソコンだけでなくインターネットに保存。

3. モバイル、ソーシャル、クラウドの情報技術の発展で生活や仕事の仕方が変化してきた。

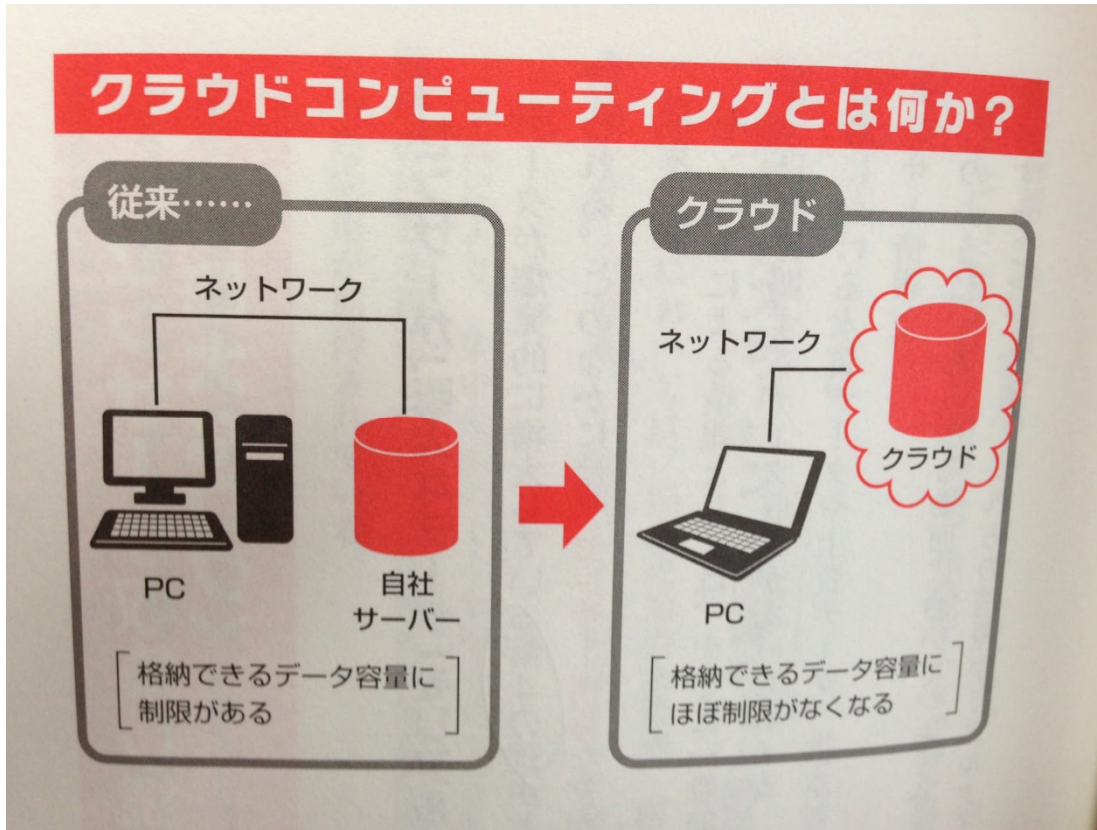
4. 行動履歴、なまの声、文書、写真データが爆発的に増え、ビッグデータと呼ばれる様になった。

5. 自動販売機が在庫データを離れたコンピュータに自動的に送信する。



### 3. データの量と質

# クラウド時代の到来



- データ増加の3つめはクラウドコンピュータの登場だ。
- クラウドは莫大に大きなサーバーをネットワークを通じて利用し利用料金を払う。
- どんどん増えるメールや添付資料、写真等を保存しても限度があるがクラウドなら可能になった。

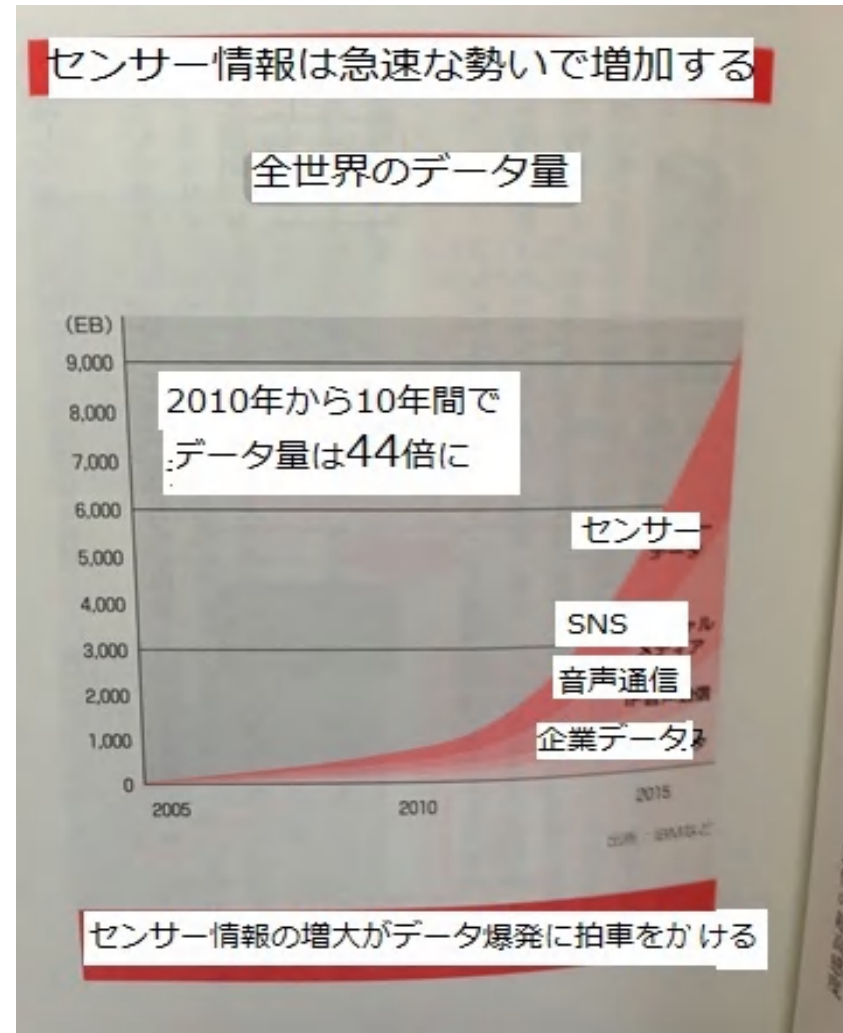
### 3. データの量と質

# センサー情報

○センサーの小型化、軽量化、低消費電力、低価格化により活用が進み、センサー情報が拡大している。

○気象データや自動車や人の交通情報、農業の土質データ、洪水の水量データが増えている。

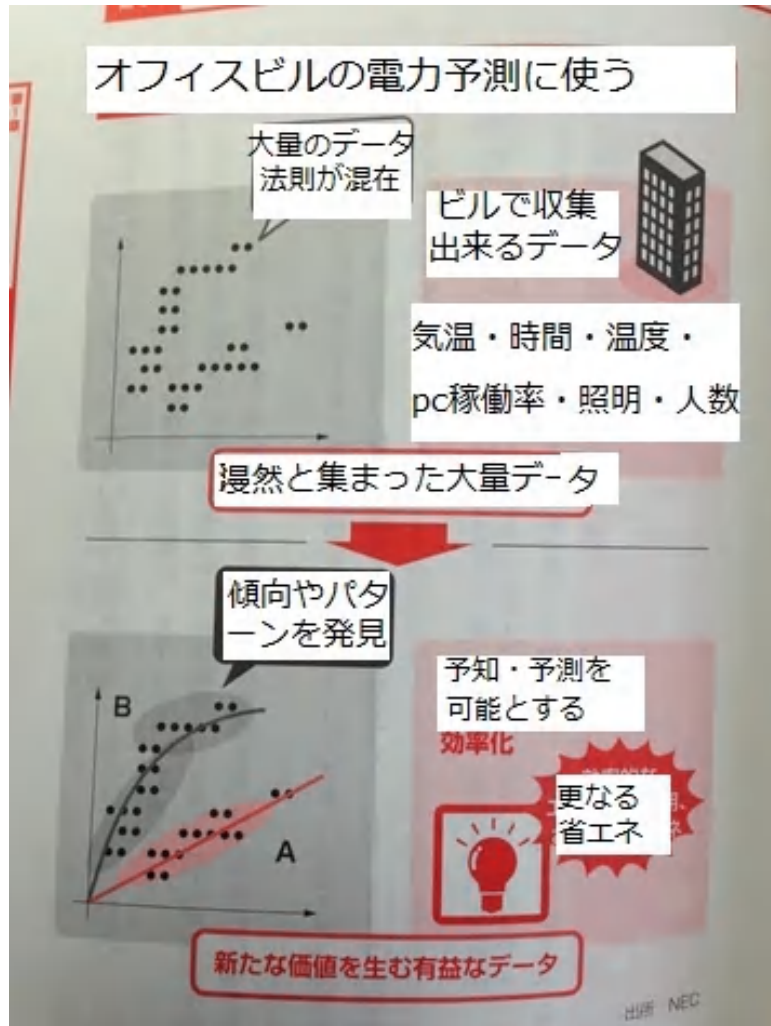
○センサー情報は数値だけでなく、画像データを含んでいる。





### 3.データの量と質

# 高感度振動センサー



○高感度振動センサーは微弱な振動から建物の劣化監視、水道管の漏水検知、人体の血管の流れを検知できる。

○オフィスではビルに設置したセンサーから気温、湿度、PC稼働率、勤務時間帯、社員数、照明の数、天気情報等により1週間後の電力使用料を予測する。

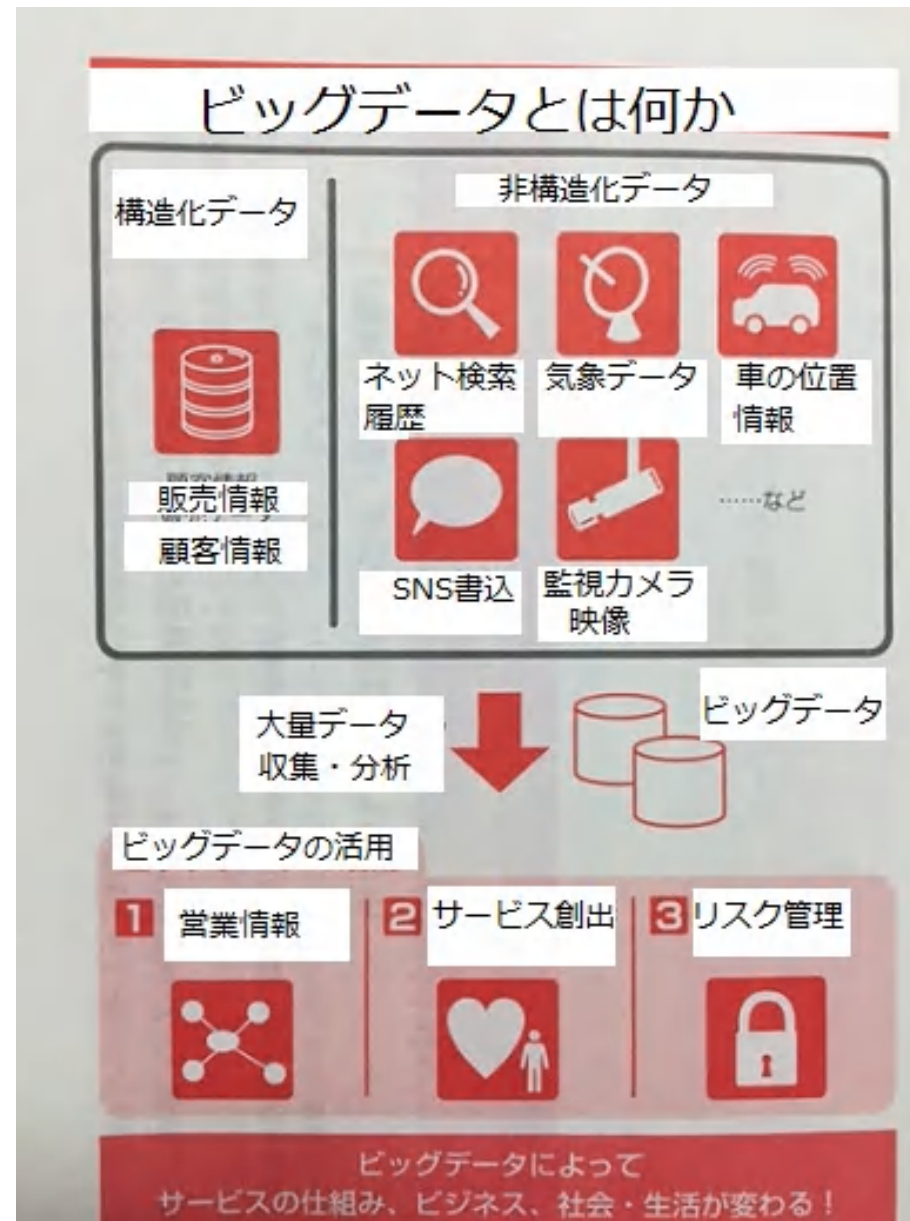
## 4.ビッグデータとは何か

○ビッグデータとは大量のデータを蓄積、分析することで、これまで見えなかった発見を導きサービスに生かす事を言う。(非構造化データ)

○さらに分析結果から事前に対策を打てるため、サービスや仕事の仕方が変わることになる。

(構造化データ)

○ビッグデータによって「サービスの仕組み、ビジネス、社会・生活が変わる



#### 4.ビッグデータとは何か

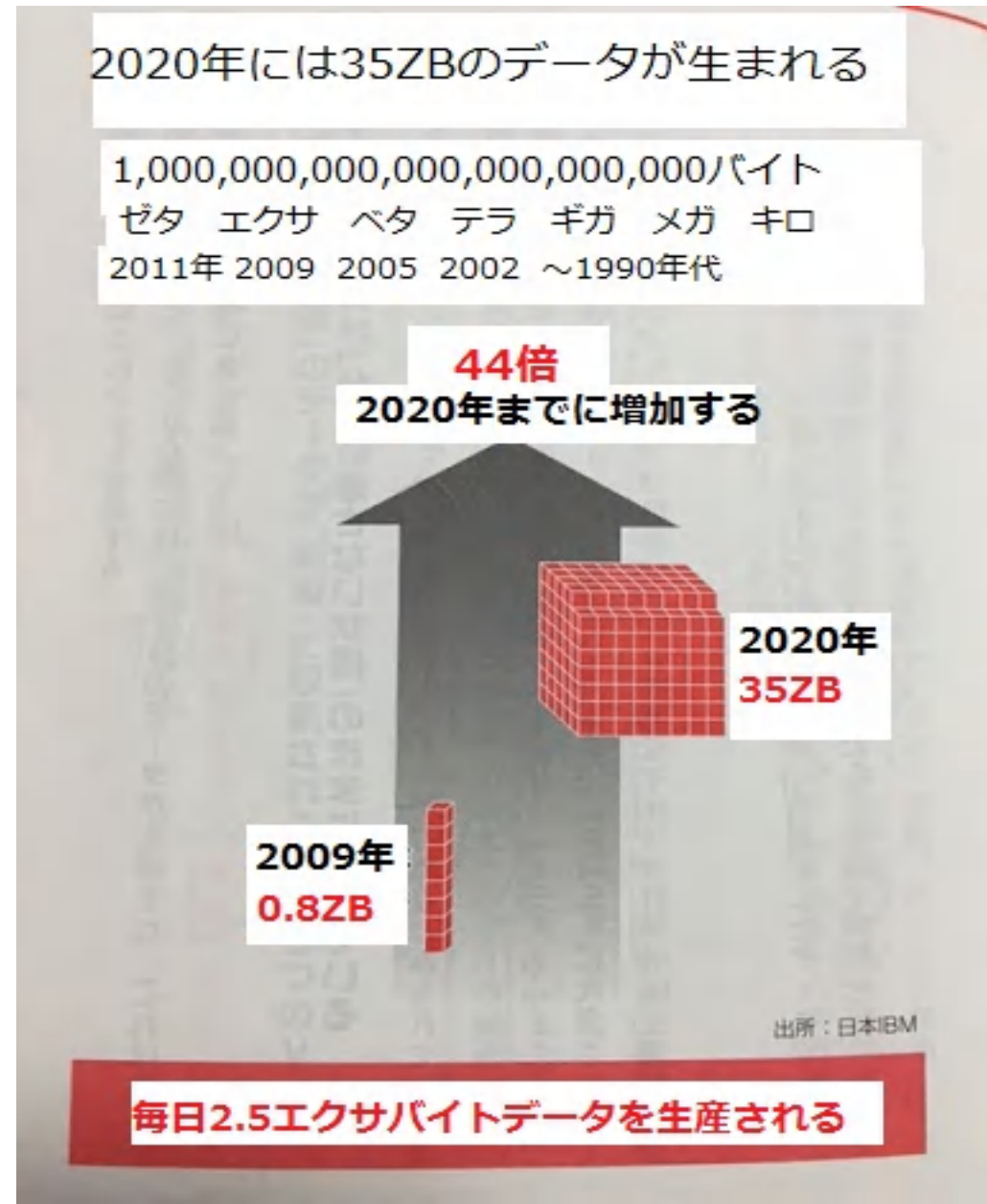
# 巨大な情報が宝の山

- コールセンターに届かない「ユーザーの声」を拾う
- ツイッターやフェイスブックから消費者の不満を収集する
- 商品棚の前でどんな行動をしたか（ショーケーストラッカー）を収集する（手に取ってもどしたか、悩んでから買ったか等）
- 情報システム部門からマーケット、営業部門へのIT投資

#### 4.ビッグデータとは何か

# これが「ビッグデータ」だ

- 毎日世界で毎日ブルーディスク10億枚分 (2.5EB=25GB×10億枚) の情報が生まれている。
- 2009年には0.8ZBのデータ量だが (90%が2年間で生まれた) 2020年には35ZBとなる。



#### 4.ビッグデータとは何か

# データの大爆発が進んでいる

○2011年に生成されたデータ量は1.8ZBと言う。

○1.8ZBとは2時間の動画で2千億本（1人24時間見続けて4700万年かかる）

○32GBのIPADに保存するには575億個（積上げると富士山の25倍の高さ）

○米国の全国民が2万6976年のあいだ毎分3ツイートするのと同じデータ量

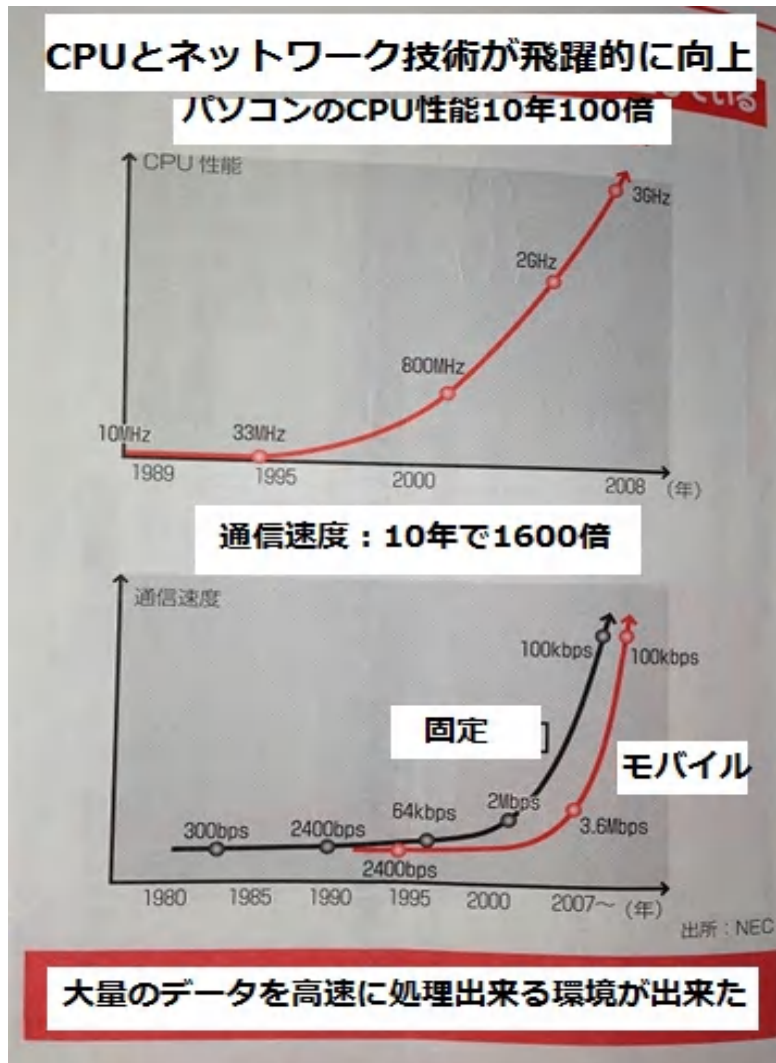
○ツイッターは毎日7TB、フェイスブックは10TB、グーグルは24PB、2013年インターネット量は667EB（人の一生分を記録すると3TB）





#### 4.ビッグデータとは何か

# なぜ今ビッグデータなのか



○巨大なデータを処理する技術が追いついて来た。

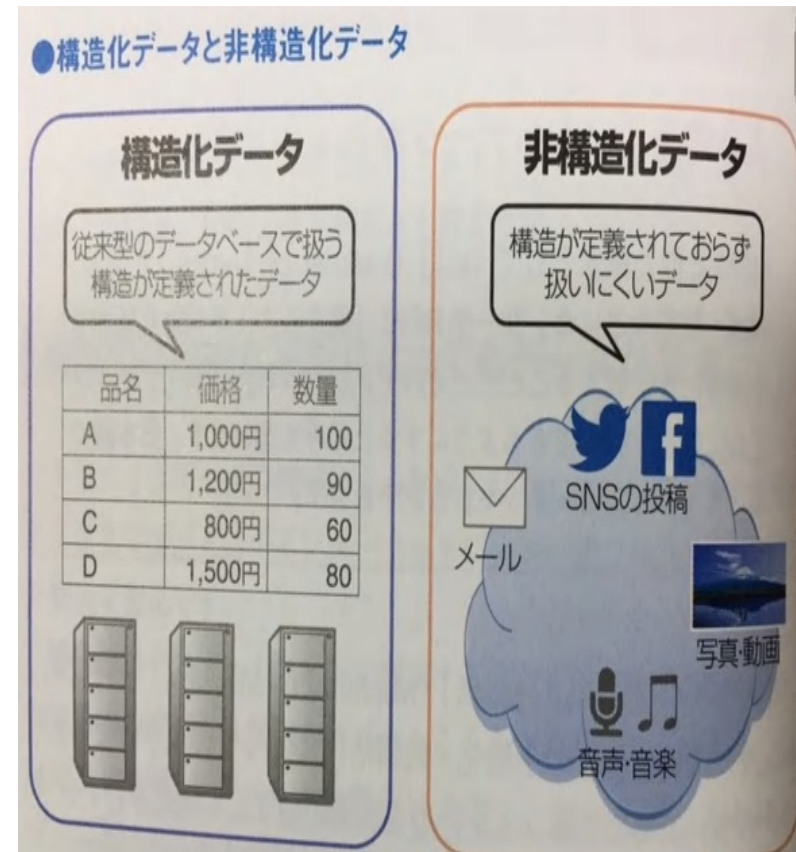
○CPUの性能は10年で100倍、データを伝送するネットワークはワイヤレス化と共に高速化され、通信速度は10年で1600倍になっている。

○ハードディスクより高速なSSDの登場、情報発信のスマートフォンやデータ収集のセンサーが進化した。

#### 4.ビッグデータとは何か

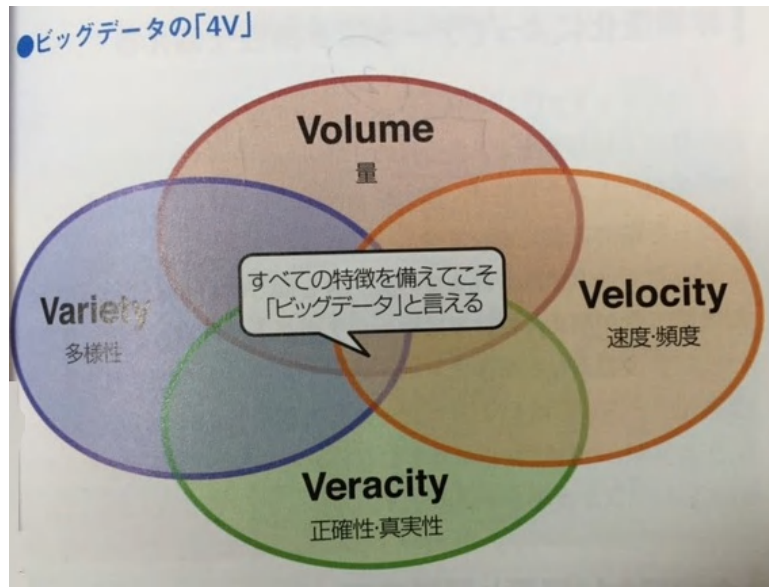
# 構造化データと非構造化データ

1. 従来のPOSでは商品名、価格、数量の集計がされた:ラージデータとか大規模データと言い構造化データと言う。
2. ツイッターやFACEBOOKの投稿は文字や写真や動画、音声、が混じったもので非構造化データと言いそれをビッグデータと言う。
3. 従来は蓄積しても価値がなかった非構造化データが様々な手法で構造化され活用する動きとなった。



## 4.ビッグデータとは何か

# ビッグデータの定義



## 4V+ビジョン

データの背景、  
なにを導きたいか、  
継続的に活用できるか

■IBMではビッグデータを4Vで定義している。

「データ量」**多い-Volume**が明確な定義無い

「多様性」**非構造化-Variety**データ、テキスト、写真、動画、音声等

「**速度**」**Velocity**JR-ICOCA、は改札口を通る度に大変な速度でデータが蓄積される。2015年8月で1100万枚売れている。SUIKAは4300万枚

「**正確さ**」**Veracity**多少不正確なデータが混じっても全体の量が大きいので影響が少ない、

「**洞察力**」4Vがあるだけではダメでそこに「**ビジョン**」がないと活用できない

## 5.ビッグデータの処理、分析

# インターネットの普及と比例して医療が発達

1.分析結果や学者の研究論文が昔は5年ほど後しか伝わらなかったがインターネットのお蔭で翌日皆が利用できる時代になった。

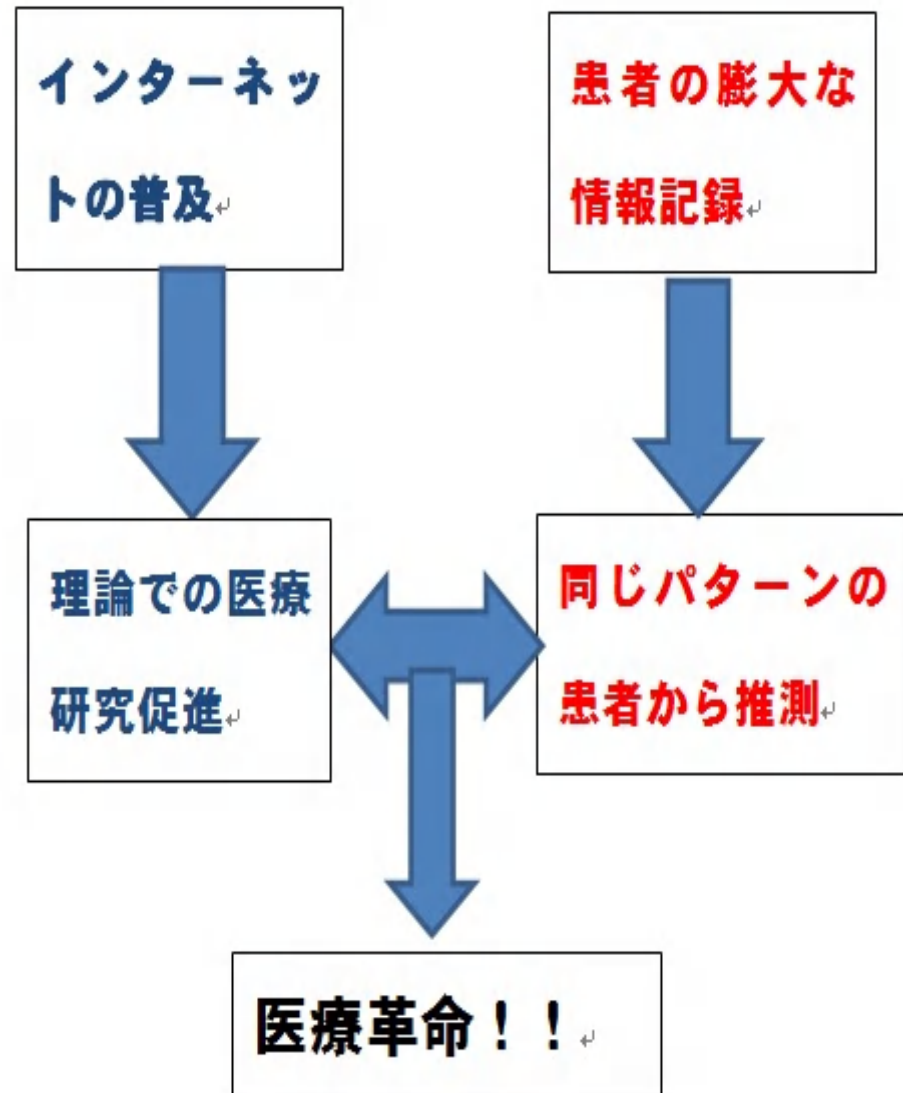
2.病院では患者のデータの記録が進みその結果患者達の膨大なデータが蓄積されてきた。

3.この2つが結びついて

4..2015年に医療テクノロジーの大革命が起きた

5.2015年をスタートで年5ヶ月ずつ伸びて行く

6.30年後には平均寿命100歳になると予想



## 5.ビッグデータの処理、分析

# ビッグデータ活用には



「技術」・・・ 列指向DB  
メモリ高速  
Hadoop

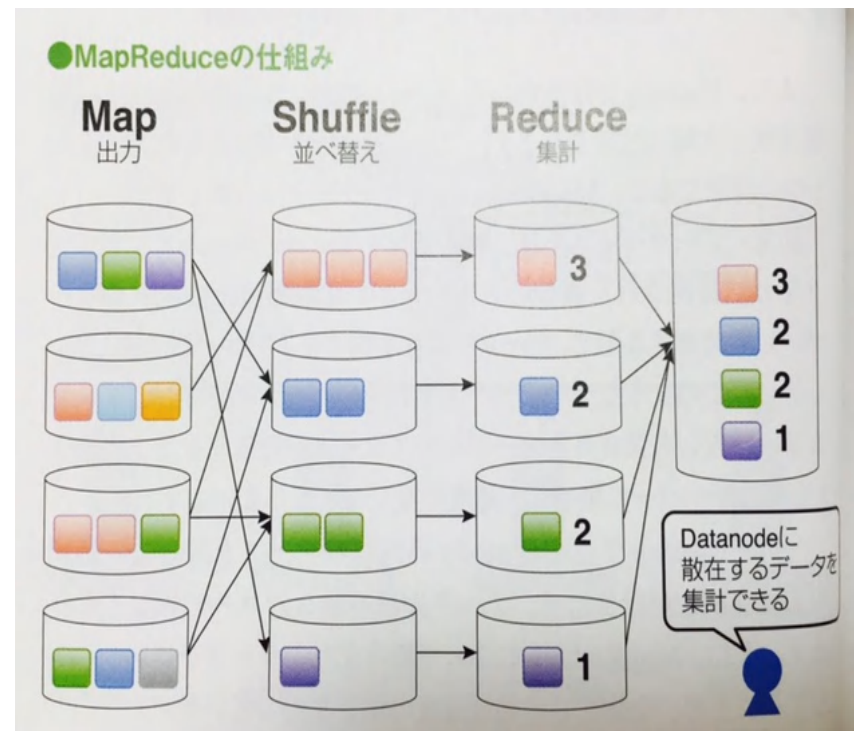
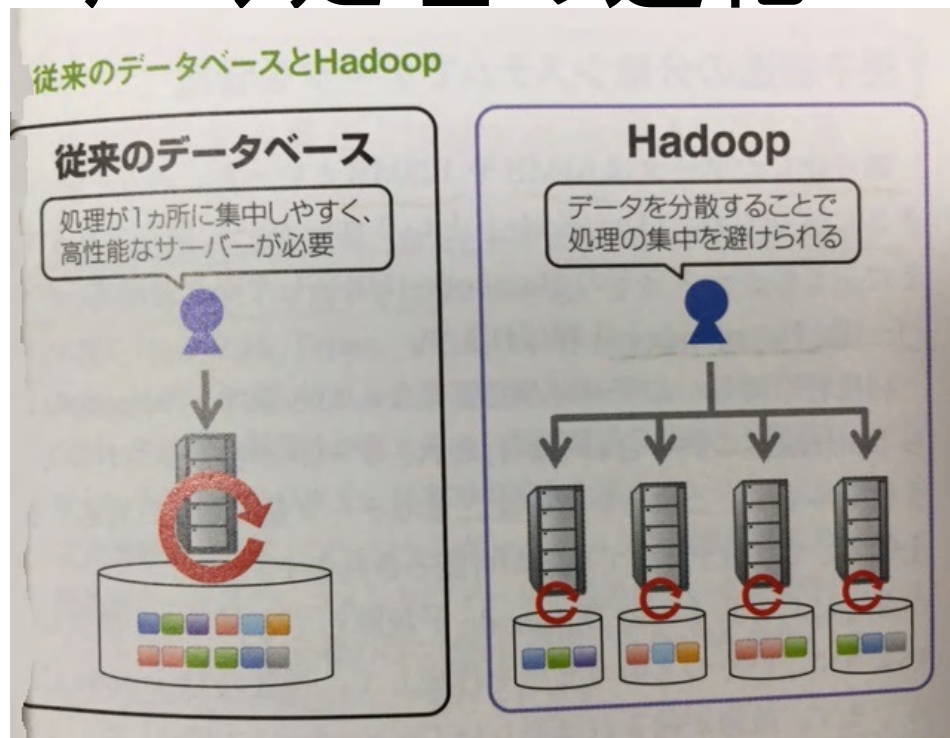
「データサイエンティスト」  
・・・統計・数理  
分析能力

「目利き」・・・データの価値を  
発見  
価値を上げる



## 5.ビッグデータの処理、分析

# データ処理の進化



1. 列指向型・・・データの年齢、購買日時を集計
2. インメモリ型・・・早いが高つく
3. Hadoop・・・最も注目されている技術・・・安価なサーバー群に作業を分担させる、並べ替え、集計する。

# 巨大データの解析技術

- ○hadoop : 大量のデータを複数のマシンに分散して処理できるシステム、（分散ファイルで分散処理する）
- ○データ格納: 22TBのフラッシュメモリーで10倍圧縮する事により200TBのデータをマシン上で動かす事ができる。
- ○ストリームデータ処理: 大量データを一度でバッチ処理すると時間がかかるが、刻々と流れるデータを処理するストリーム処理をする。

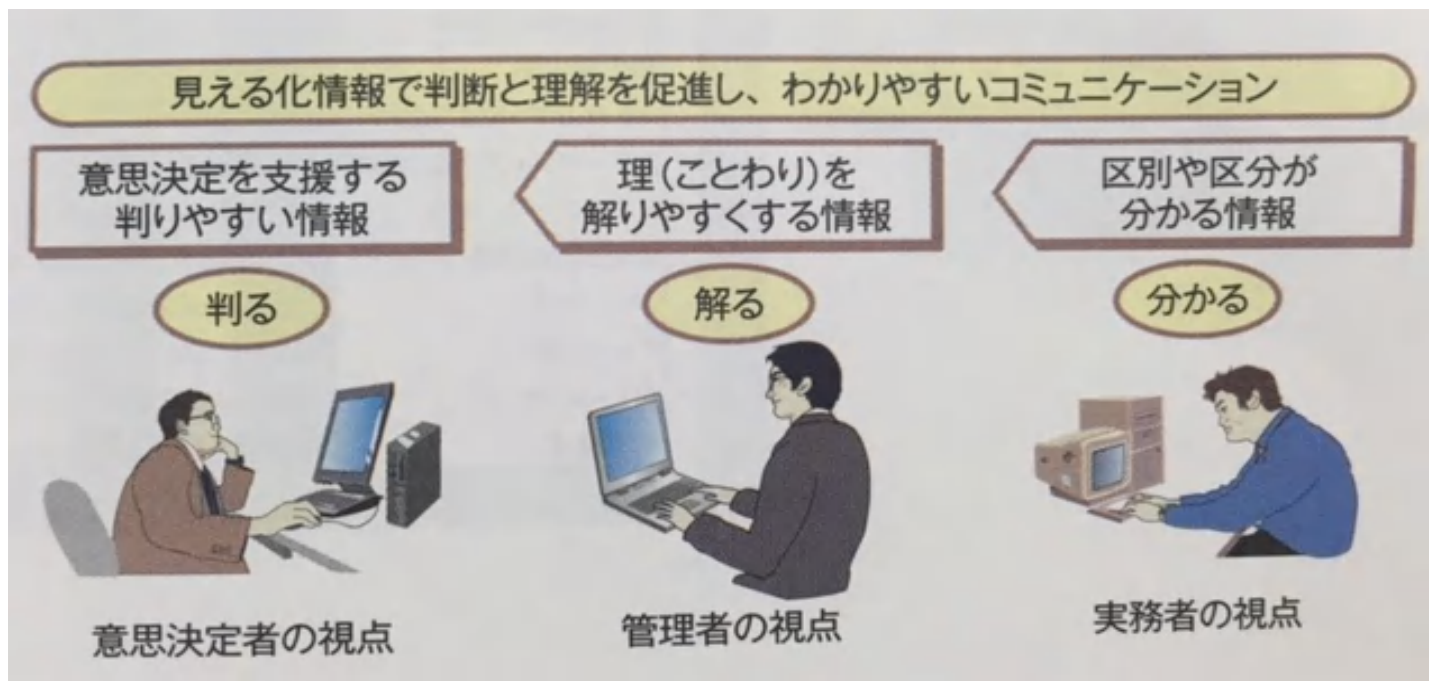
## 5.ビッグデータの処理、分析

# データサイエンティストとは



ビッグデータから価値を引き出す技術を持つ人材が求められている。統計、数理、科学的知識、分析力、大量データ処理の技術を持ち、適切な分析手法を選択出来る人材が必要だが少ない。**データが大量で種類も多くリアルタイムに実行する必要**が高い企業、amazonや楽天(オンラインショップ)、GREEやDeNA(ソーシャルゲーム会社)インターネット広告会社、リクルート社等に在籍している。

# データの見える化



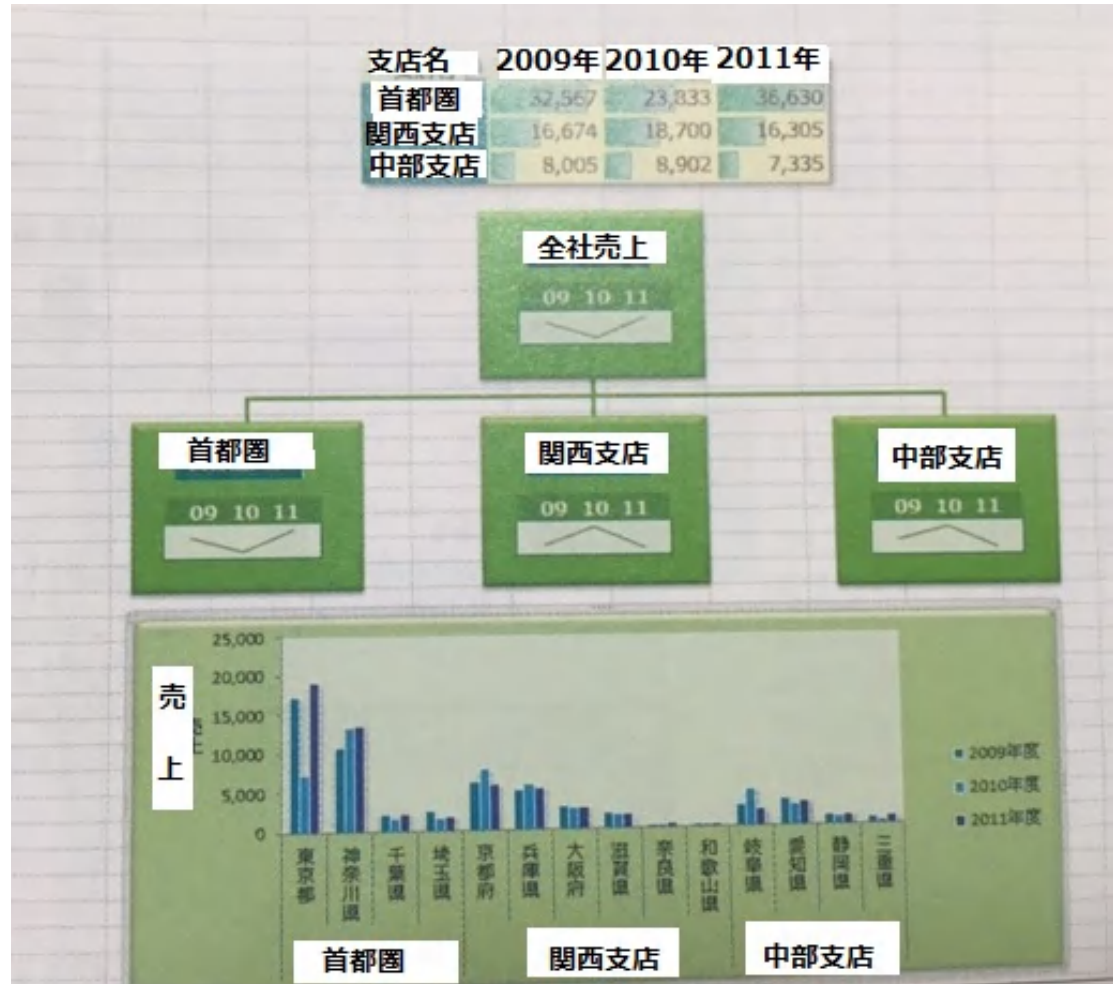
- 1.意思決定者:「判る情報」・・・適確な判断を下す為ひと目で判ること  
物事が判明すること
- 2.管理者層:「解る情報」・・・なぜそうなるかがひと目で解ること  
解きほぐれること
- 3.実務者層:「分かる情報」・・・具体的に区別でき、詳細で見やすい  
意味や区別がつくこと



# データの見える化—経営者層

経営者層：  
判断出来る情報  
効果が判る情報  
損得が判る情報  
可能性が判る情報  
対策立案が出来ること

最新のデータを吟味し、  
質の良い少量の情報を  
きれいに見える化する、  
料理に例えれば旬の  
季節煮出す懐石料理  
のようです



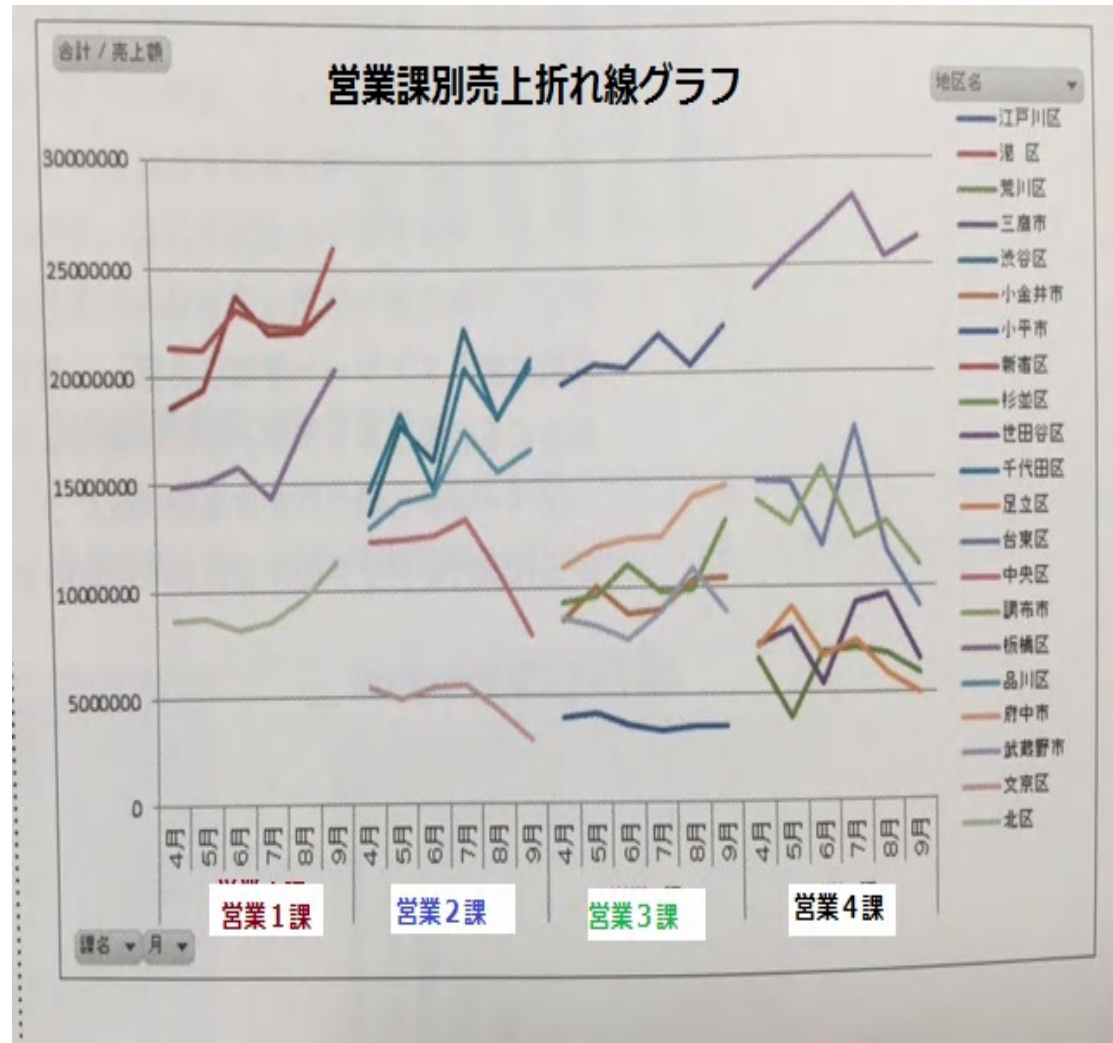


# データの見える化—管理者層

管理者層：

要因が解る情報  
原因が解る情報  
背景が解る情報  
課題が解る情報  
結論が解る情報

必要な要素を全て取り込んだもの、  
料理に例えればコース料理のように各要素を目的に合わせて解りやすく見える化する



# データの見える化—実務者層

実務者層：

- 問題が分かる情報
- 進捗が分かる情報
- 傾向が分かる情報
- 推移が分かる情報
- 比較が分かる情報
- 事実が分かる情報

必要に応じた詳細な情報を見る化する、料理に例えれば注文に応じて出される居酒屋料理のようなものです、

商品区分2 (すべて)		商品別貢献度費率表						
商品名	売上高	粗利高	粗利率	回転率	交差比率	上積成比	利益貢献度	貢献度比率
ミニバーグ グルメ	6,807,454	1,581,454	▲15.4%	155.5	230	7.8%	1,862	3%
味噌シャブシャブの外	7,621,125	1,721,125	▲17.4%	164.5	296	7.7%	2,208	1.5%
純正ラード	6,243,685	1,443,685	▲15.3%	79.0	121	7.3%	882	7.0%
ベーキングパウダ業務用	360	79,000	▼11.3%	59.3	67	5.7%	984	3.0%
ホイルマカロニ 業務用	916	20,530	▼12.6%	83.8	10.6	5.3%	564	4.5%
からあげ粉 小袋タイプ	195	43,958	▲15.6%	99.3	15.6	5.2%	820	6.5%
龍田揚げ粉	500	114,540	▲15.3%	79.0	121	5.0%	604	4.8%
フライトリオ チキンタイプ	980	224,708	▲15.6%	93.0	14.5	5.0%	721	5.7%
ステーキソース 荒引 胡椒味	684	152,046	▲15.3%	41.1	6.3	4.9%	311	2.5%
お好みソース 辛口	4,480	1,044,452	▲15.4%	150.9	23.3	4.9%	1,136	9.0%
たけのこ台湾産 大缶	810	186,754	▲15.3%	111.9	17.1	4.8%	828	6.6%
いりこのだし	3,030	686,582	▲15.5%	84.1	13.0	4.5%	586	4.7%
てんぷら粉	640	147,400	▲15.5%	25.1	3.9	4.4%	173	1.4%
クリームシチュー	9,532	2,197,536	▲15.3%	62.9	9.6	4.3%	1,409	3.3%
徳用醤油	7,740	1,745,000	▲15.3%	130.3	20.0	4.2%	843	6.7%
ツナチキン	19,190	4,395,260	▲15.4%	159.2	24.5	4.0%	980	7.8%
しゃけなんぼん	90,032	21,364,444	▲14.4%	211.0	30.4	3.8%	1,165	9.3%
みりん風調味料	17,444	4,140,996	▲15.7%	180.8	28.3	3.8%	1,084	8.6%
ビッグシュウマイ	33,175	7,227,675	▲15.5%	99.1	15.3	3.7%	563	4.5%
ミートソース	37,320	8,067,280	▲15.4%	62.8	9.7	3.5%	339	2.7%
総計	567,786,798	86,051,980	15.2%	83.1	12.6	100.0%	12,587	100.0%

## 5.ビッグデータの処理、分析

# データ処理の結果から知恵を導き出すには:データマイニング

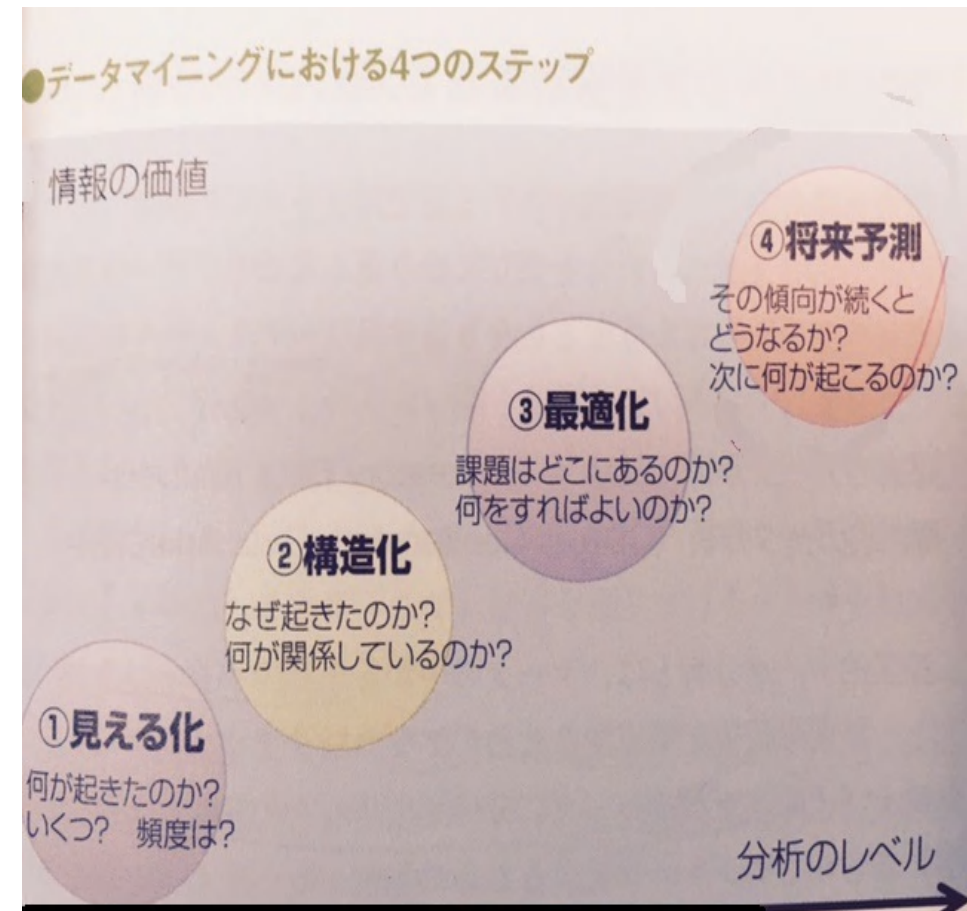
データマイニング:「山の中から宝を採掘する意味」

Step1.「見える化」・・・どんな商品をどんな顧客が何人購入したか等の全体像を把握

Step2.「構造化」・・・なぜ購入したか、天気、同時に買ったもの、WEBアクセスの有無等購入動機に何が影響があるか

Step3.「最適化」・・・構造化で判明した関係をどう組合せたら売上を最大化出来るか

Step4.「将来予測」・・・さらに今後起こる変化を加味して予測する。



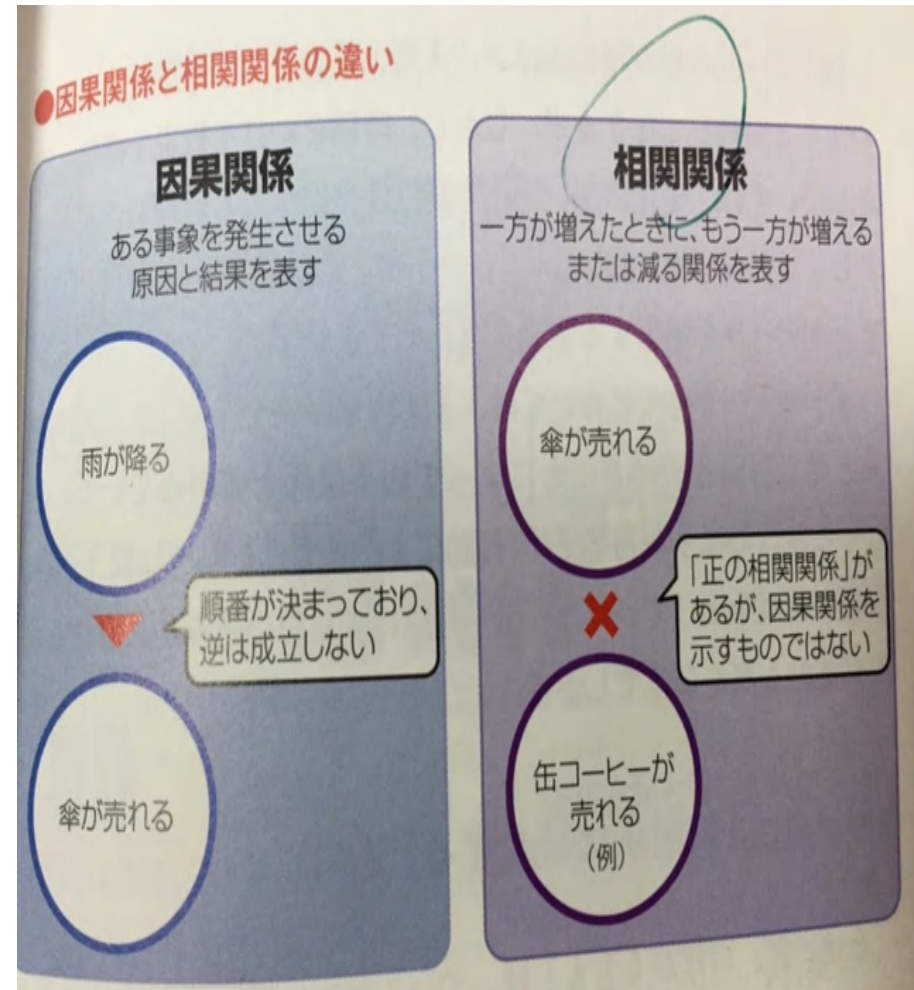


## 5. ビッグデータの処理、分析

# ビッグデータ文化に対応した組織とは

企業は今後ビッグデータ文化を作り上げることも大切。

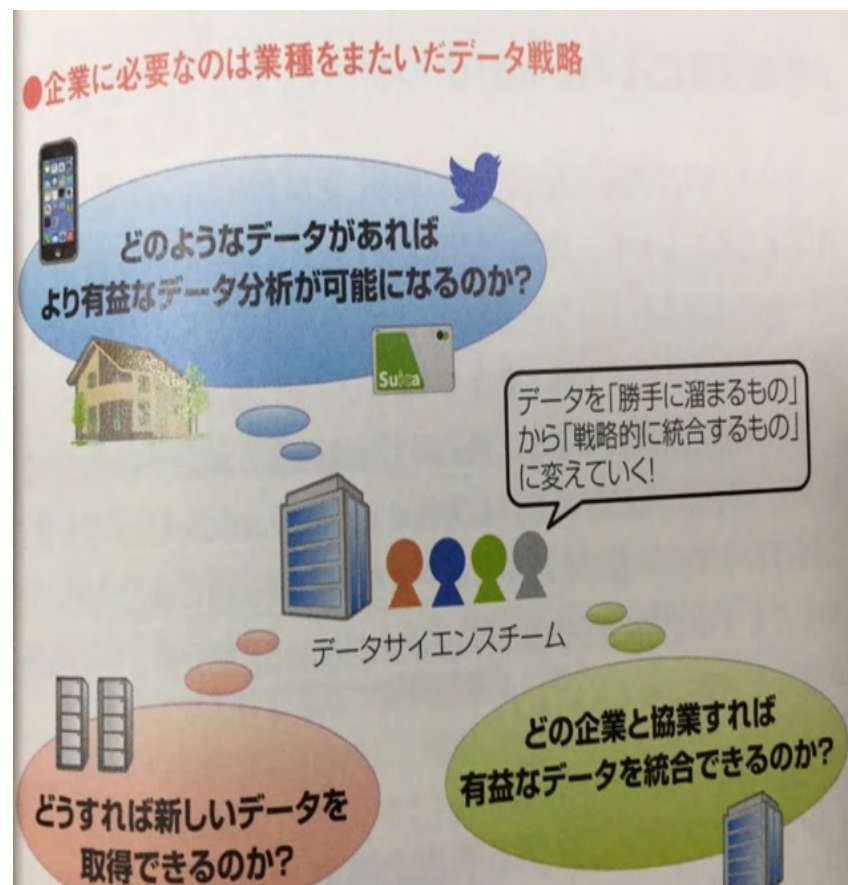
1. 相関関係の活用・・・これまで分析とは**因果関係**を明白にする事だった。しかし因果関係を明らかにするには膨大な時間と費用はかかる。その間にビジネスチャンスを失うこともある。そこでデータの**相関関係**の方法で効果のありそうな結果を抽出して実行に移す事が重要。



## 5.ビッグデータの処理、分析

# ビッグデータ文化に対応した企業間の関係

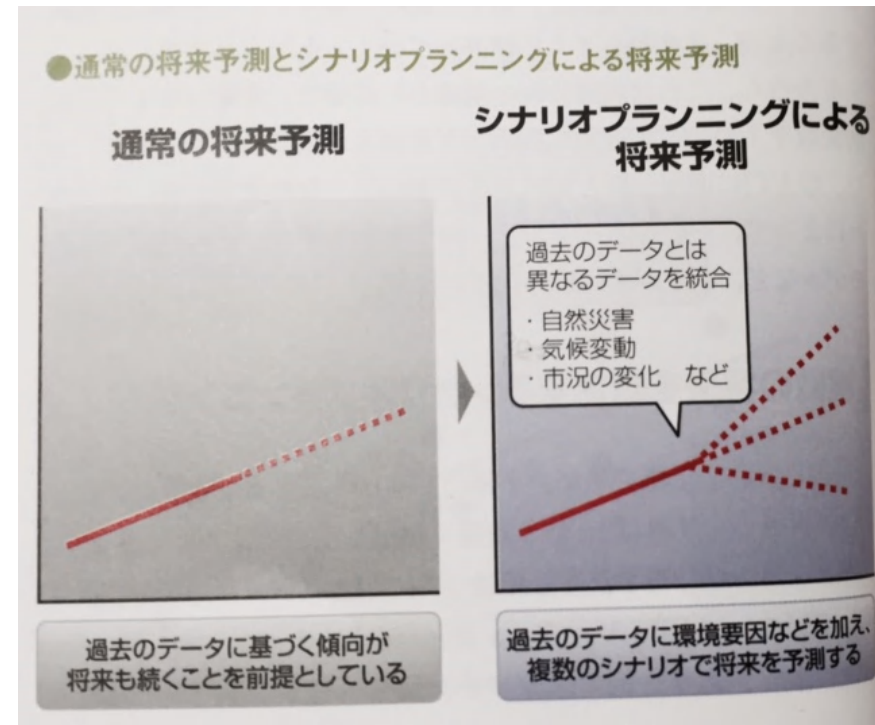
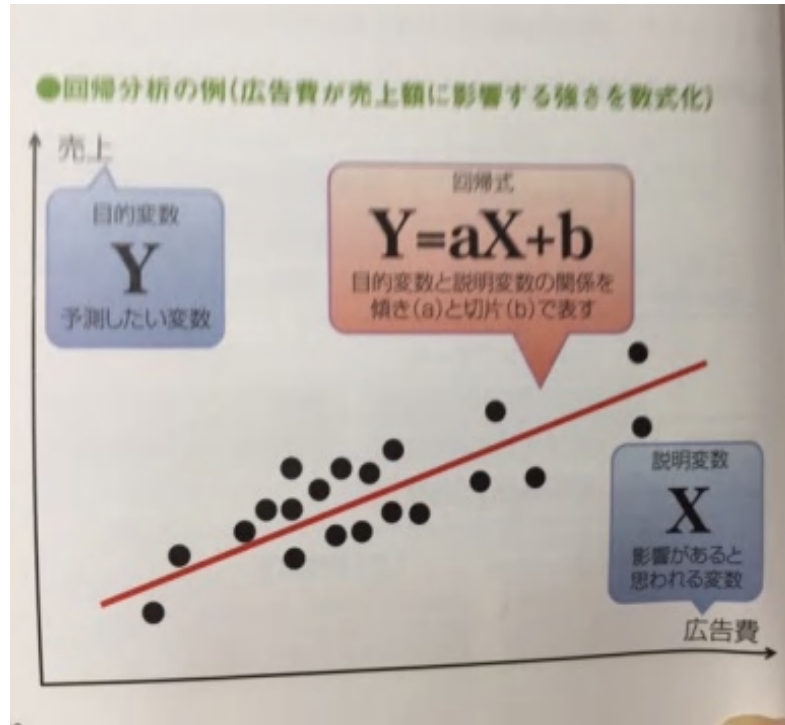
2.異業種間での戦略的なデータ統合・・・競合関係に無い異業種の企業とデータを共用して組み合わせ、利用者の行動の時間的、空間的つながりが解かれれば双方にとって利便性の高い商品・サービスが生まれる可能性が高くなる。欧米では企業間のデータ取引が活発である。データサイエンティストは価値あるサービスの開発の役目もになっている。





## 5.ビッグデータの処理、分析

# 最適化・将来予測



最適化: 構造化で分かった要素の影響度をグラフにする。(例は売上と広告費の関係)

将来予測: 現在まで起きた傾向が続くことを前提にして、将来起こりうる複数のシナリオを設定し、仮想敵に実行して経営の意思決定に生かす。

# ビッグデータは何を生み出すか

- ビッグデータは「これまで見えなかったものが見えてくる」
- センサー技術の進歩も情報収集の手段となっている。
- 冷蔵庫には開閉を感知する、エアコンには温度管理のセンサー、
- 人を感知してスイッチを切る、街中に設置された防犯カメラ もセンサーとおなじ役目を果たしている。
- これからスマート家電、スマートシティの時代にはデータが蓄積される。
- たくさんのデータを収集し、「見える化」することが大切。
- その町で冷蔵庫の開閉が減ったことから、夏祭り等のイベントがあったと推測される。このように推測する力も必要となる。

## 5.ビッグデータの処理、分析

# データの組み合わせ

「**購買履歴**」・・・オンラインショップ、クレジット購買

「**位置情報**」・・・スマートフォン

「**SNS**」・・・ツイッター、FACEBOOK

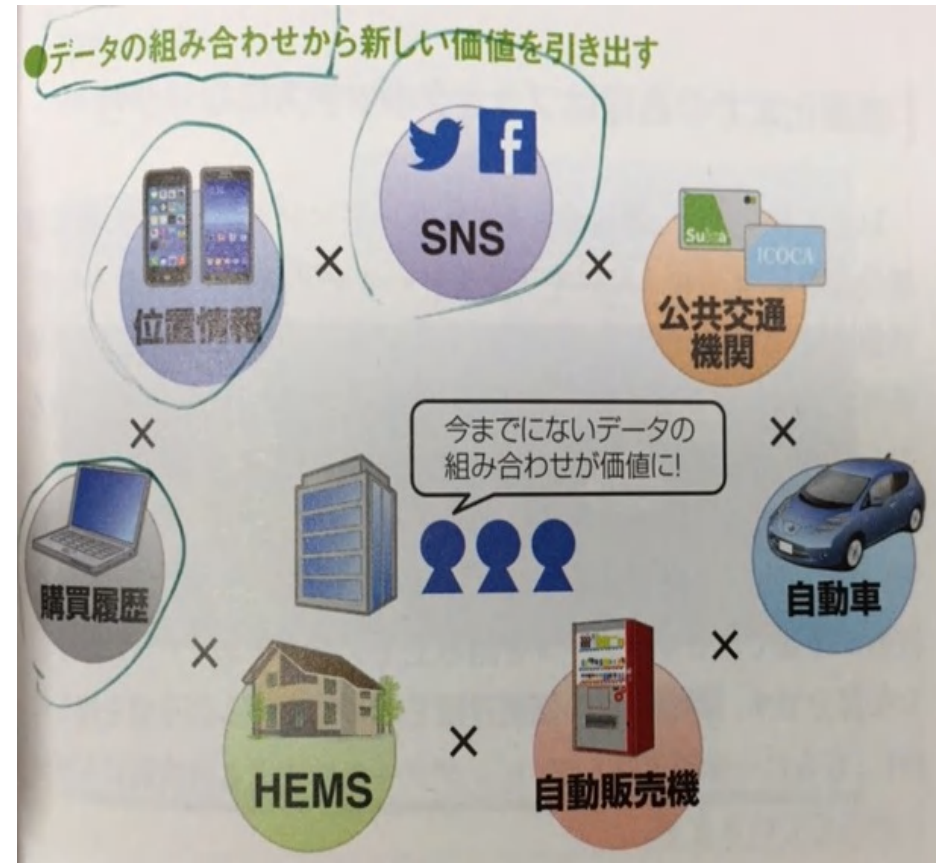
「**交通機関**」・・・ICOKA,SUIKA  
駅の入出情報

「**自動車**」・・・事故情報

「**自動販売機**」・・・在庫データを無線で伝送する

「**HEMS**」・・・

HomeEnergyManagementSystem家電や電気設備をつなぎ電気・ガスの使用料を「見える化」し「自動制御」する。



## 6.企業への活用

# ビッグデータはこう使われる

- アマゾン：蓄積した購買履歴からおすすめ商品を表示する
- ショーウィンド：ショーウィンドに立つとカメラやセンサーが働き性別や年齢を感知、顔認識機能、購入履歴からお得情報を表示する
- ワイパー：自動車の位置情報とワイパーの動作から雨を予測
- クレジットカード：1日1億件の取引データと過去7年の蓄積データから個人モデリングを13分で作り、不正の行動のチェックをする
- つぶやき：シカゴ市警はネガティブなつぶやきを地図上に表示し、デモや犯罪の防止を防ぐ、また新製品発売時につぶやきの反応で売上を予測
- 医療分野：2500名の体重計、歩数計、血圧計の日々データと健康診断データ、等を組み合わせて等糖尿病の発症を予測
- 肌メモリー：スマートフォンのカメラで写した肌画像から肌情報を測定しクラウドに登録して、過去と比較

# NHK震災ビッグデータ-桃



震災後福島県産の桃が汚染されているかどうかで、ツイッターが増える、桃の出荷時に増える。



# NHK震災ビッグデータ-桃2



桃に肯定的な  
ツイッターと否  
定的なツイッタ  
ーが議論され、  
リツイッターさ  
れている。

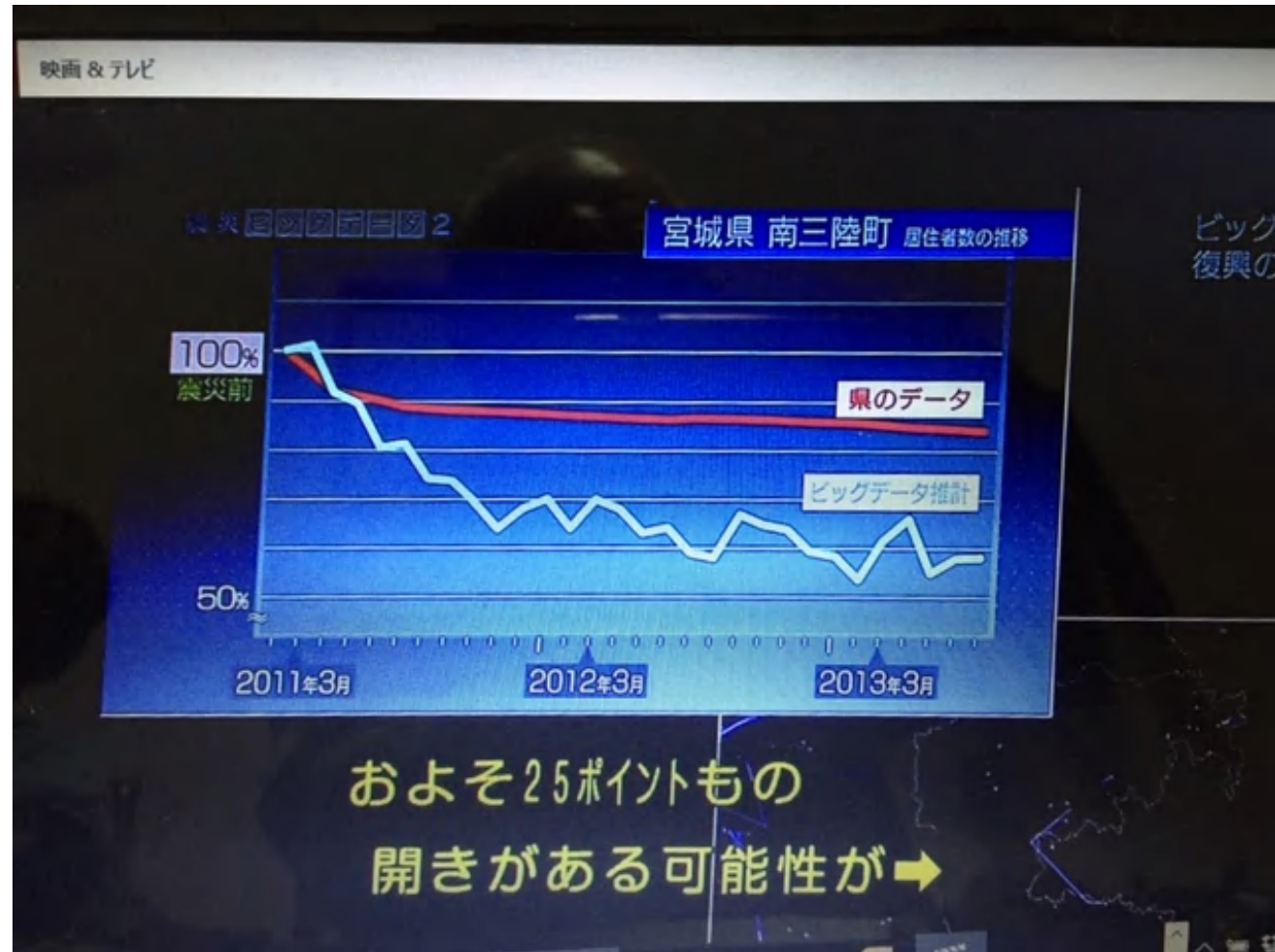
# NHK震災ビッグデータ-位置情報

スマートフォンの情報を持つ会社から位置情報を確保して被災地に留まる人工を可視化した。



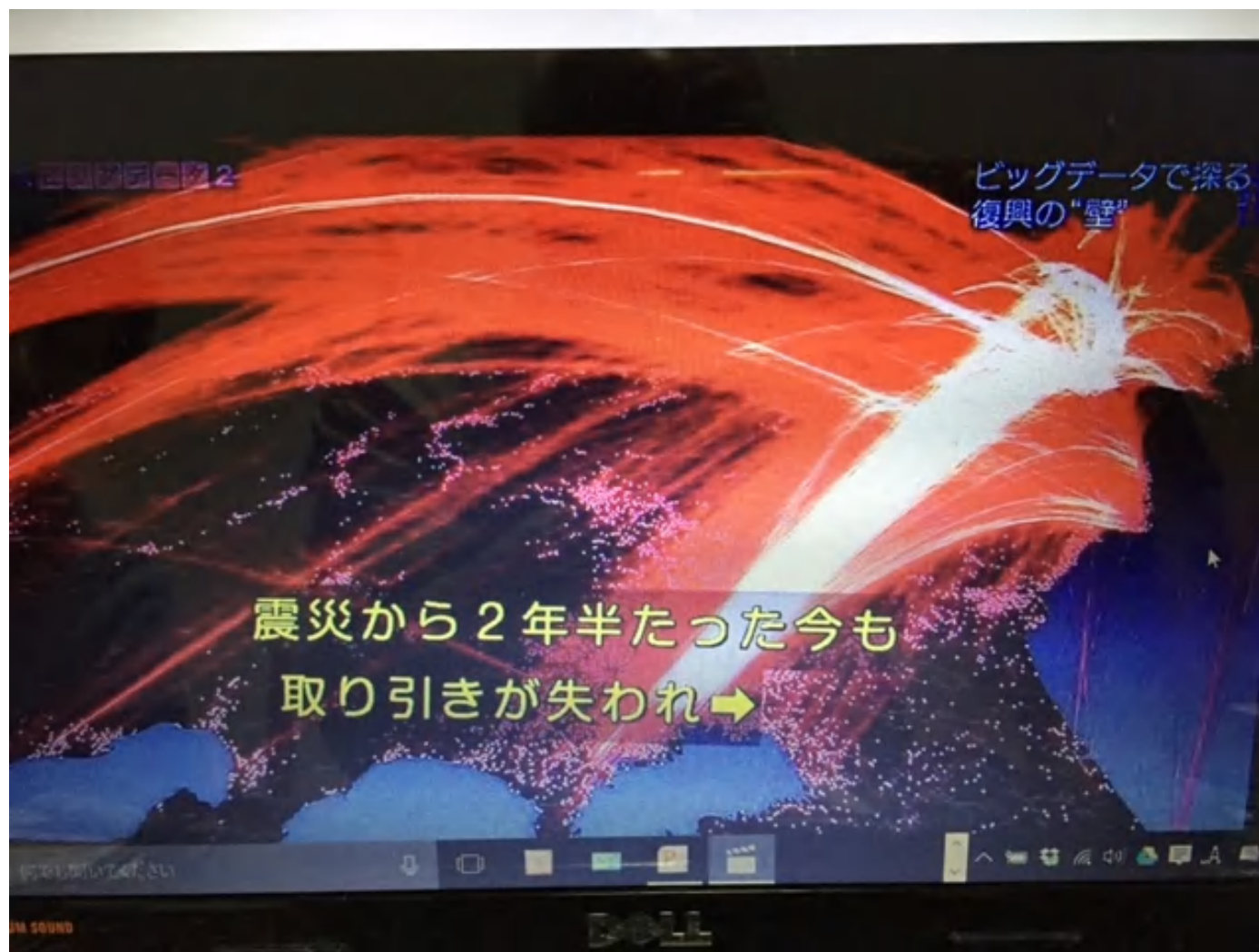
# NHK震災ビッグデータ-位置情報2

夜6時から朝6時まで位置が動かないのはそこに居住しているとすると、自身体では分からない人の動きが判明する。





# NHK震災ビッグデータ-企業ネットワーク



帝国データバンクの75万社の企業情報から震災後失われたネットワークを可視化

## NHK震災ビッグデータ-企業ネットワーク2



震災後残された企業の1次ネットワークからさらに初めて2次ネットワークを可視化した。

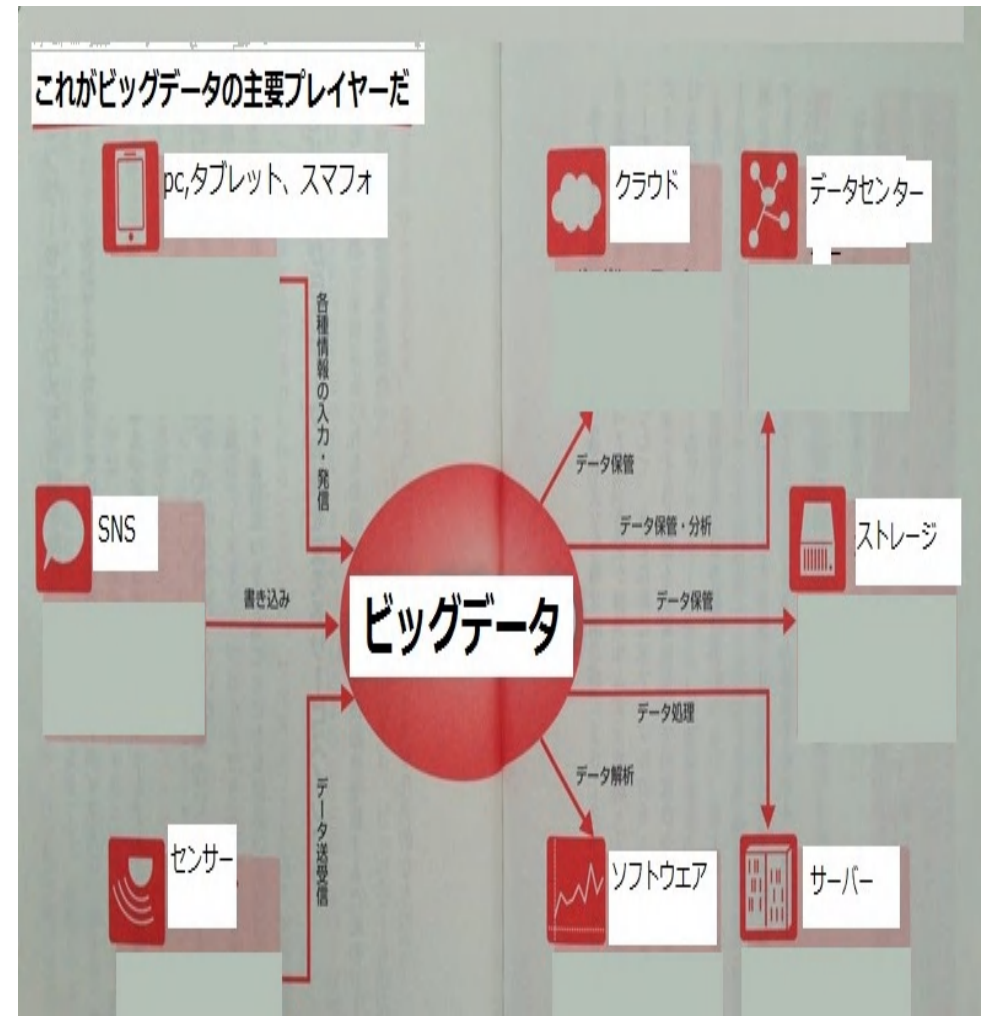


## 7.今後の展望

# ビッグデータの主要な業界

## ビッグデータに関連する業界は

- 1.情報の入力側:  
PC,タブレット、スマホ
- 2.書込:SMS
- 3.データ送受信:  
センサー、カメラ
- 4.データ保管:クラウド
- 5.データ保管、分析:データセンター
- 6.データ保管:ストレージ
- 7.データ処理:サーバー
- 8.データ解析:ソフトウェア



## 7.今後の展望

# ビッグデータの世界を動かす企業

1. PCやスマホ：アップル、富士通、NEC、サムソン、ノキア

2. SNS：Facebook, twitter, mixi, Line

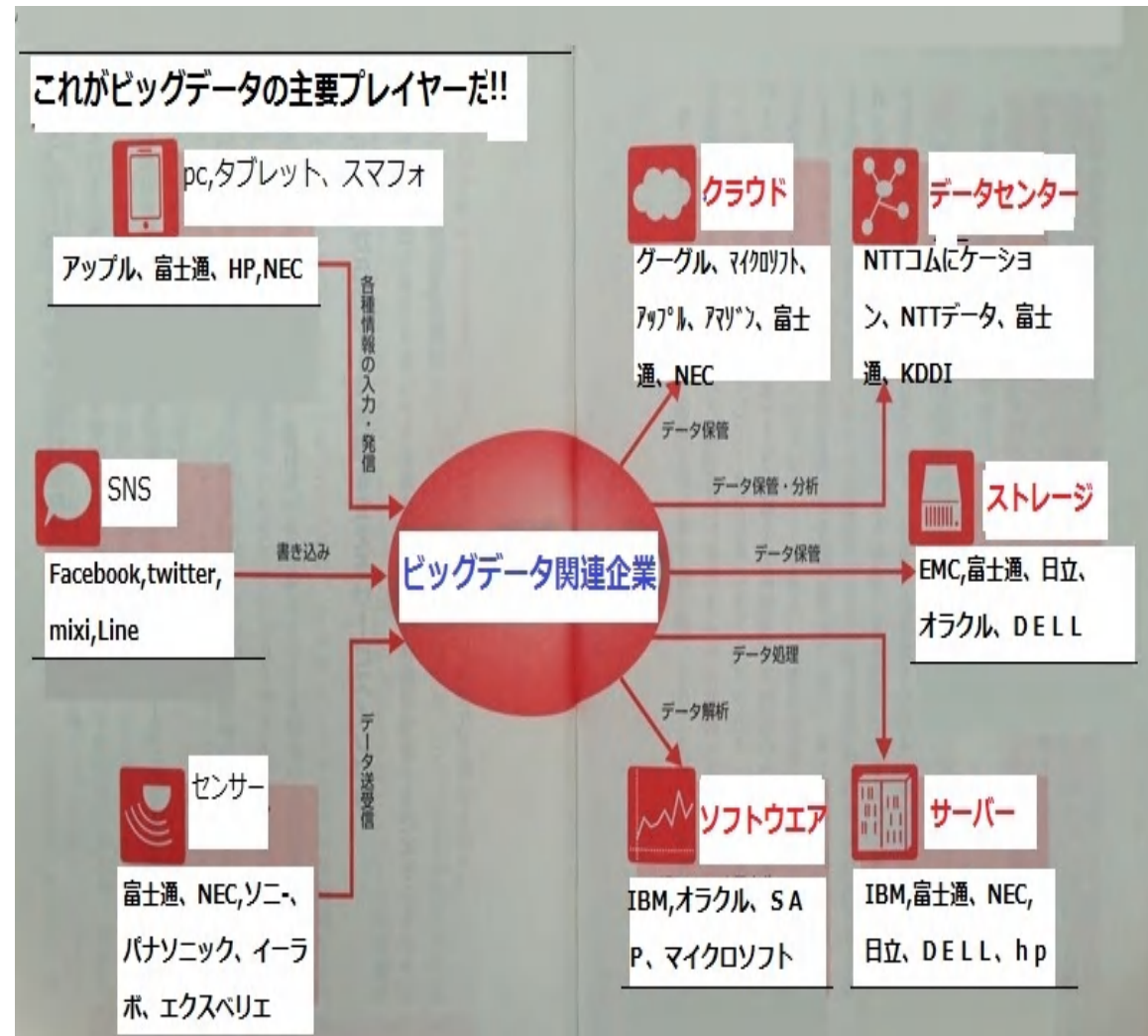
3. センサーやカメラ：NEC、富士通、ソニー、キャノン、ニコン

4. クラウド：Google、アマゾン、マイクロソフト、IBM、NEC、日立、富士通

5. データセンター：NTT、KDDI

6. ストレージ：EMC、富士通、オラクル

7. サーバー：IBM、富士通、NEC



## 7. 今後の展望

# ビッグデータの今後の展開

### 1. 企業の意思決定の際、

データドリブン(効果測定等で得られたデータをもとに次のアクションを起こす事)・・・が採用

2. 企業の内部情報を整理、分析して、ラージデータ(売上明細や入出金明細—これまで一部を除いてあまり活用されてこなかった)を意思決定に役立たせる体制を作る必要がある。

3. 企業に役立つビッグデータのありかを探し出し、変動するビッグデータを早く取り込む組織、体制を作る。

4. ビッグデータは今後「IOT」(internet of things)と結びついて進化、拡大する。

5. さらにその後は人工頭脳(AI-artificial intelligence)が組み込まれ、さらに質の高い判断がされる。

以上で終わりです、ご清聴有難うございました！！