Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Big Data

Er Big Data bare en døgnflue?

Lasse Bache-Mathiesen
CTO BIM Norway

Big Data – What is all the fuss about?

"The effective use of Big Data has the potential to transform economies, delivering a new wave of productivity growth…Using Big Data will become a key basis for competition…"

"We estimate that a retailer embracing Big Data has the potential to increase operating margin by more than 60%"

"$300bn – the potential saving in US healthcare"

"$250bn – the potential saving in European Public Sector"

McKinsey Institute – Big Data: The next frontier for innovation, competition and productivity – May 2011

McKinsey&Company

"Data-Driven Decision-making can explain a 5-6% increase in output and productivity, beyond what can be explained by traditional inputs and IT usage."

MIT – Strength in Numbers – April 2011

Massachusetts Institute of Technology

"Survey participants estimate that, for processes where Big Data analytics has been applied, on average, they have seen a 26% improvement in performance over the past three years, and they expect it will improve by 41% over the next three."

Economist Intelligence Unit | The Economist & Capgemini CONSULTING.TECHNOLOGY.OUTSOURCING

What is Big Data?

NEW DATA SOURCES

25+ TBs of log data every day

12+ TBs of tweet data every day

30 billion RFID tags today (1.3B in 2005)

4.6 billion camera phones world wide

100s of millions of GPS enabled devices

2+ billion people on the Web by end 2011

76 million smart meters in 2009

NEW DATA SOURCES

80% Of world's data is unstructured

BIG DATA OPPORTUNITY

TECHNOLOGY SHIFT

TECHNOLOGY SHIFT

Low cost, massive volumes (Hadoop)

- Came out of Yahoo, Facebook…
- Distributed data on low cost server farms
- Opensource (free software)

Volume

1Pb of data: €5m → €500k

Easy analysis of unstructured data

- From the Social Media players… Facebook…
- Text analysis, voice analysis , sentiment …

Variety

Insight from web logs, text, images, call logs, …

Insight integrated into Business Processes

- SAP & Oracle – integrating BI into apps
- InMemory / columnar storage → Can do real-time

Velocity

Insight delivered at the point of action

The BI-Study 2012/2013 has been conducted by the Business Information Management department at Capgemini Norway

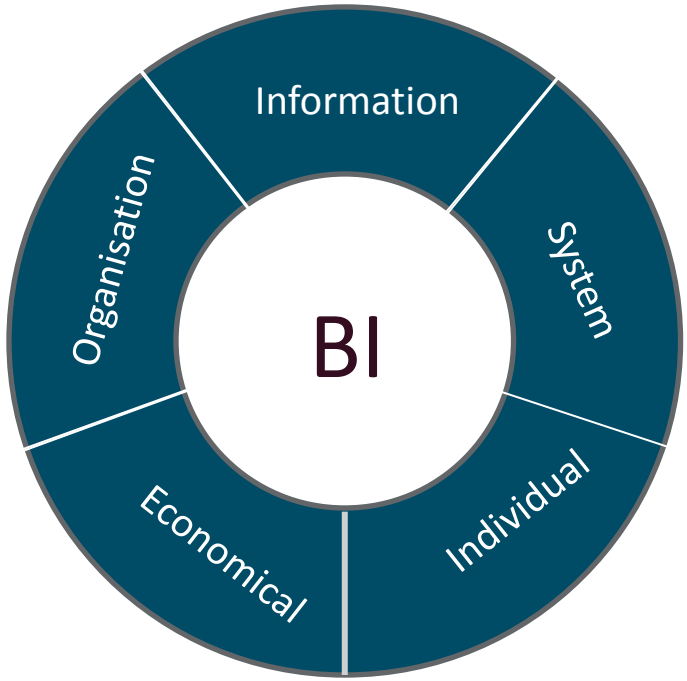
Main Findings

- **Critical success factors in developing a BI solution**
  1. Factor 1: Information Quality
  2. Factor 2: Organizational Impacts
  3. Factor 3: Economic Factors
  4. Factor 4: Individual Impacts
  5. Factor 5: System Quality

The main purpose of the BI-Study is to understand how to establish a successful BI-solution

Success Factors

Prior to initiating this project, a literature study was performed. This resulted in five factors for successful BI-solutions that we wished to study in detail.



Hypotheses

Furthermore, we defined six hypotheses related to the success factors. By trying to answer these, we try to gain further knowledge on how to establish a successful BI-solution.

1. Understanding the business is an important factor in a BI-solution
2. Technology is not as important as experience
3. A successful BI-solution has active users
4. The BI-solution gives the business added value and is a good investment
5. A successful BI-solution should improve the organization
6. Responding companies in the analysis are reluctant to adopt new technologies

55% of respondents are not happy with the pace of change and the adaptability of their BI solution

Findings

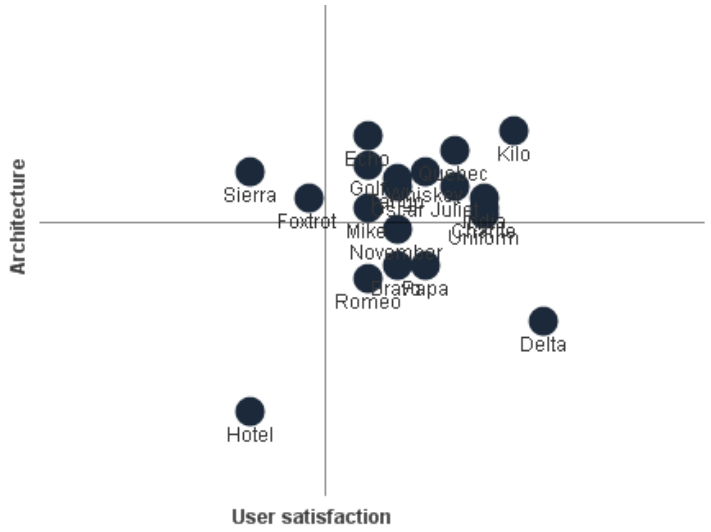
- 75% of the asked companies state that they use agile/scrum as their project management/software development approach when it comes to their BI solution. Only 10% say they use the waterfall approach.
- Further we see that while 35% of the responding businesses are happy with the pace of change and adaptability of their BI solution, 55% are simply NOT happy and believe things are not happening fast enough.
- Only 1 company states to be happy with the adaptability of their BI solution, and that company uses the waterfall approach as their project management/development method. This can be explained by the fact that the waterfall approach takes more time handling the user requirements before continuing with development.
- Pace of change and adaptability is not just a result of project management, but it also indicates that model and architecture has to be adjusted to make the solution responsive to fast change.

Focus on key business requirements is vital

- When using the waterfall approach in software development, there is a broad focus on the user requirements and the team takes time to work with them.
- In order to keep agile software development from evolving into an ongoing farce, it is important for data warehousing professionals to focus on the key business requirements of the project.

Future BI architecture

Data warehouse architecture has no impact on satisfaction



Feature architectural principles

- The established centre for data integration, storage, and exchange in BI environments is traditionally the Data Warehouse. Recently, new architectural principles have been introduced in the landscape of BI and data warehouse. This section will briefly tap into some:
- **Transformation Hub**: A logically central component that concentrates functions for data integration, enrichment, and exchange. It is designed to serve for managerial, analytical and operational applications alike.
- **Logical Data Warehouse**: A mixed data management architecture that simplifies and accelerates information access for business consumers. The logical data warehouse leverages traditional warehouses, big data and data virtualization to manage data.
- **Data Federation**: The ability to aggregate data from disparate sources in a virtual database so it can be used for business intelligence or other analysis.

Findings

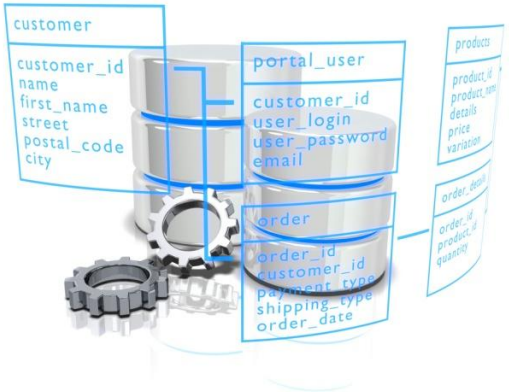
- All the participants where mainly utilizing traditional data warehouse architecture, described by Inmon and Kimball. In total 63,16 % of the respondents had build their warehouse according to Inmon´s principles.
- Despite of that, 65 % of the respondent believed that traditional data warehouse architecture is in need of modernization.

Few have adopted new technologies, yet a joint opinion is that modernization is required

Adoption of new technologies

- The analysis revieals that Norwegian companies are reluctant to acquire new BI technologies:

➢ 10 % has Mobile BI
➢ 5 % uses Cloud-services in the a data warehouse context
➢ 15 % has InMemory
➢ 10 % manage unstructured data



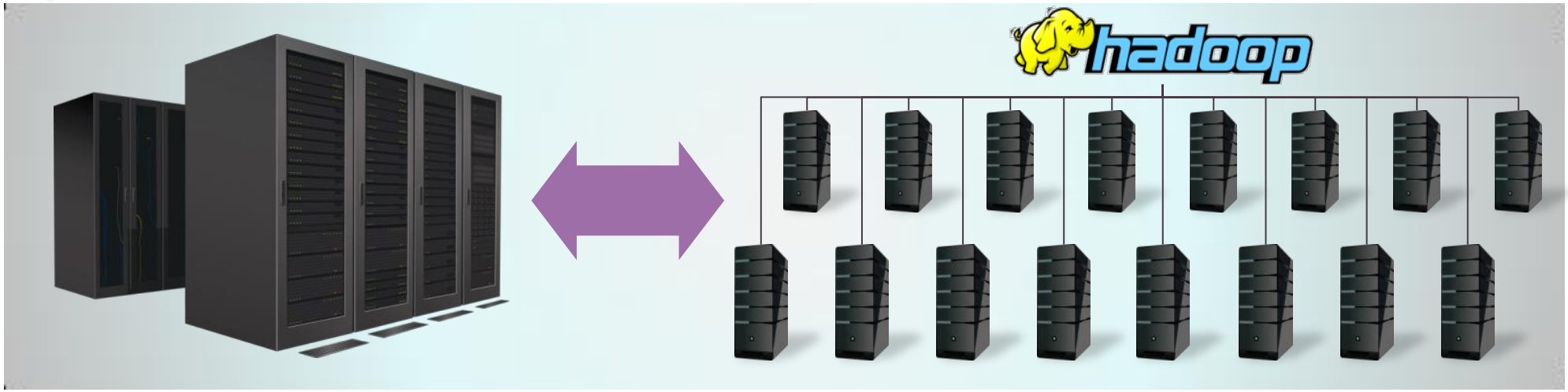
Big Data

- The companies have not adopted any Big Data technologies. Big Data are often taken in conjunction with large amounts of data. The analysis revealed that 60 % of the respondents has a data amount less than 1 TB. This indicates that Big Data technologies used in Norway will be used for other reasons than the unmanageable amount of data.

- The Thre V´s of Big Data:
  - **Volume**: Large amount of data to processes.
  - **Variety**: Data are collected from a variety of sources, allowing the data to be both structured and unstructured.
  - **Velocity**: Data is generated faster than ever.

.

BI Appliances ⇔ Hadoop



- **Expensive** dedicated HW
- Built for **performance**
- Designed for high volumes (e.g. 10s of TB)
- High availability
- Initially developed using Relational Data bases
- **Very mature** solutions (skills, SW, HW, administration)

- Designed for modelled and structured data
- Business As Usual ways to design, build and deliver
- Teradata, Exadata, Netezza, HANA...

- **Commodity** PCs
- Built for **extreme scalability** (Batch oriented)
- Designed for extreme volumes (10s of PB and more)
- Very high availability
- Initially developed for web ranking
- **Not yet fully mature**

- Hadoop = Data is distributed over many machines
- MapReduce = Computing is distributed and executed where data is (grid solution)

In-memory is changing the game

An in-memory appliance

40 x86 cores, 1TB of RAM

For only 100 K EUR !

Performance improvement means:

1 to 10 ratio: 10'' and 20'' become instantaneous

1 to 100 ratio: 2 minutes become 1 second

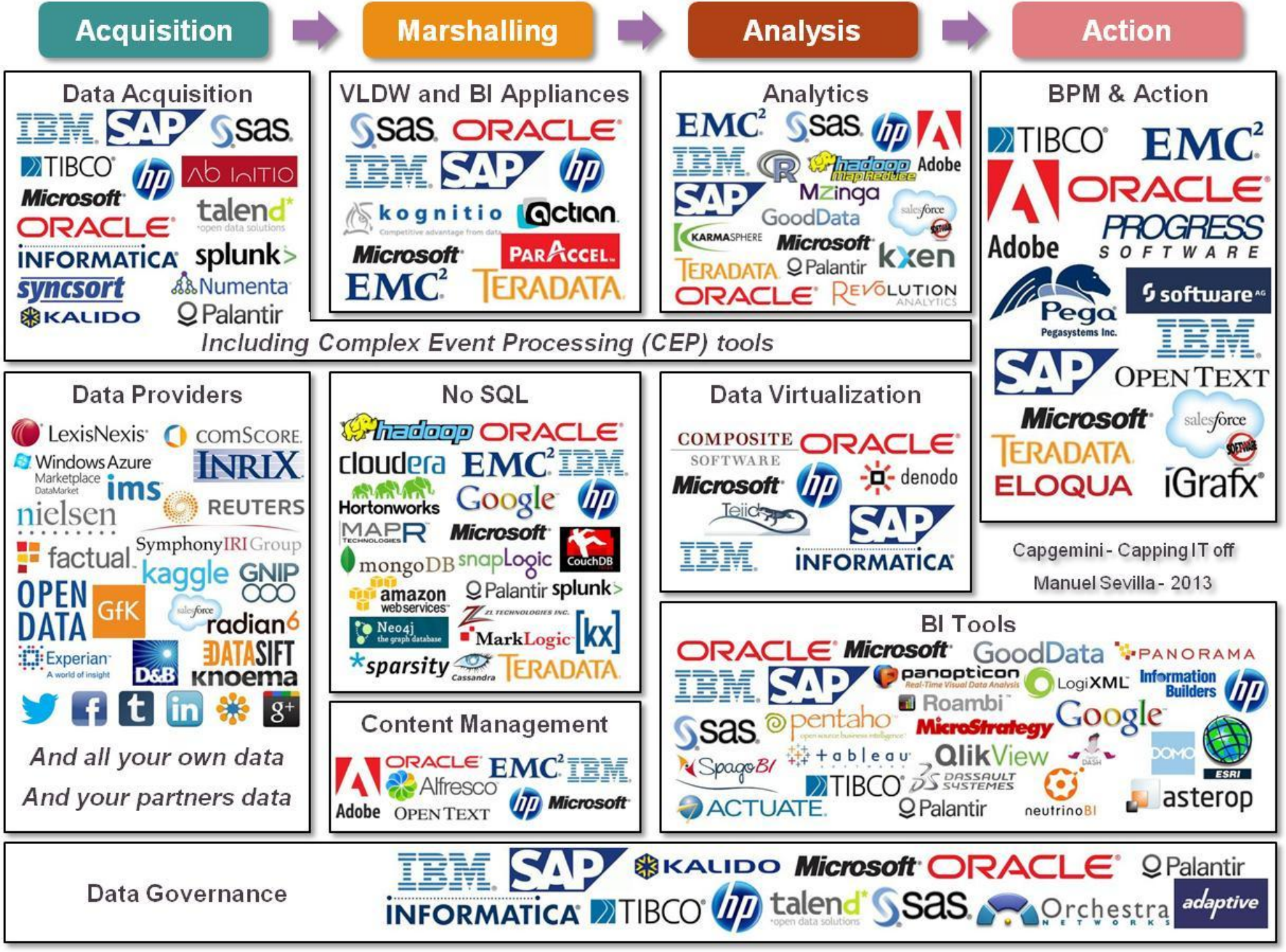
1 to 1000: 2 hours are only 10 seconds

48 hours process should run in 3 minutes !

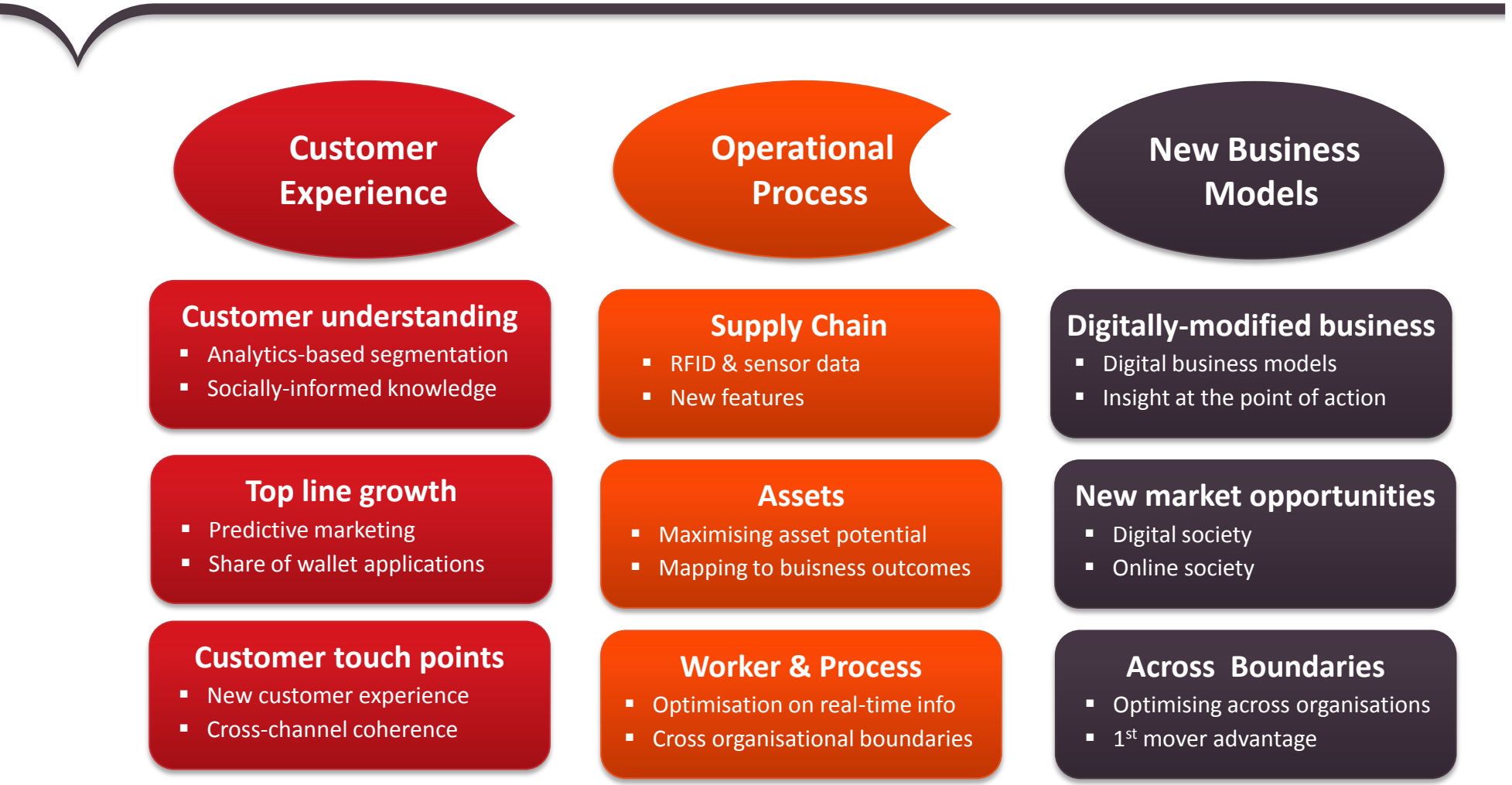
Technology Conclusion

- Big Data technologies allow you to handle data without limit
  - ➔Petabytes and Exabytes of data at a low business cost.
- In-memory means real-time feedback
  - ➔Operations can be guided, automated and optimized in real-time
- Predictive analytics means looking forwards not back
  - ➔Providing the business with daily and strategy guidance not just feedback
- Big Data is designed for the cloud
  - ➔Dynamic scale in processing and storage – no more procurement cycles

The era of back office BI is definitively over. Organizations that do not embrace this change will see themselves outrun by their competitors in the next 3 years


Acquisition
Marshalling
Analysis
Action
Data Acquisition
VLDW and BI Appliances
Analytics
BPM & Action
Including Complex Event Processing (CEP) tools
Data Providers
No SQL
Data Virtualization
Capgemini - Capping IT off
Manuel Sevilla - 2013
BI Tools
Content Management
And all your own data
And your partners data
Data Governance

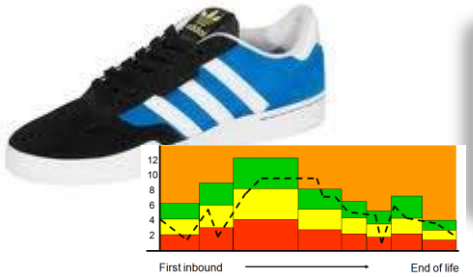
Big Data will impact all businesses...



Big Data & Analytics is able to enhance business models across all industries

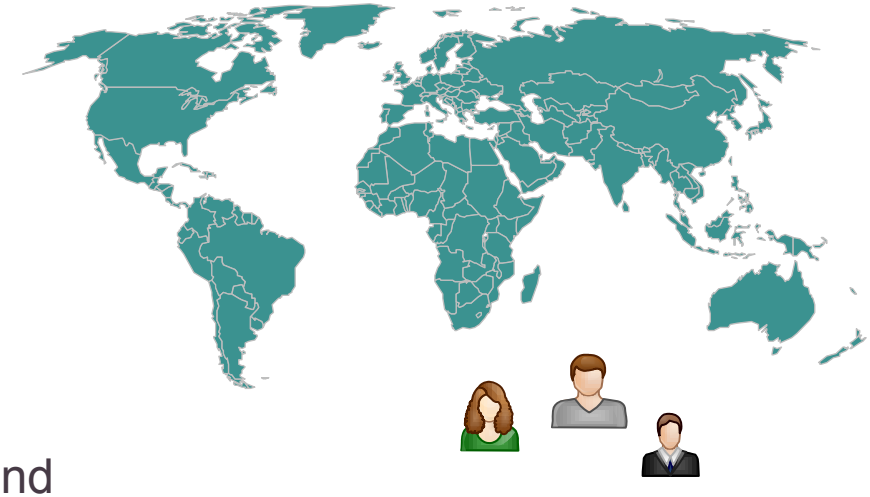
Telco: - Call Centre optimisation
- CDR (Call Data Record) – churn models, revenue optimisation..
- Data traffic analysis

Public Sector: European Space agency: Flood warning and information system – using satellite image data for Disaster Management Centers and Rescue Teams (who access flood information via mobile devices).



Retail: Combining ERP data with multiple POS data to extend the sourcing business model for own retail + wholesale clients. Shortages, outages and returns could be reduced by dynamic buffers.

Healthcare: Workforce Optimisation

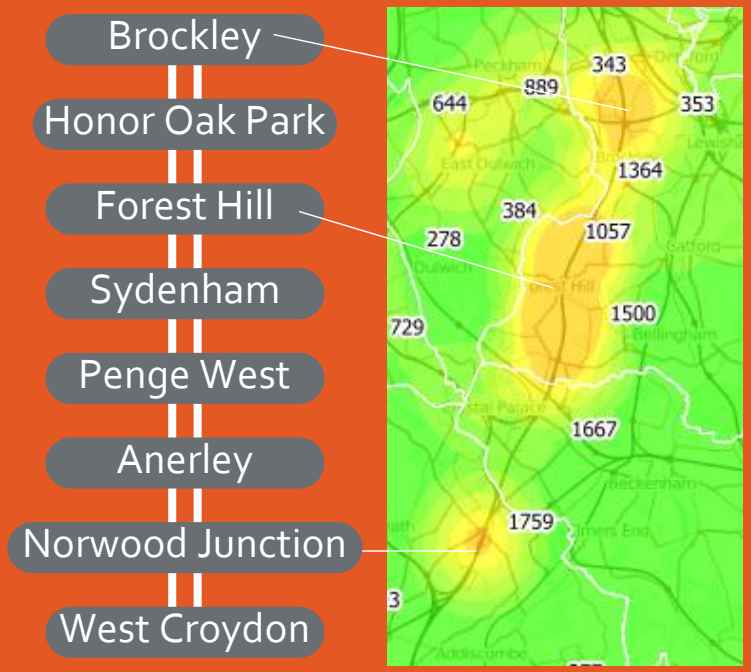


Healthcare organization delivers, maintains and repairs medical equipment all over the world with 6,000 field service engineers across 32 countries.

10x improvement in response times

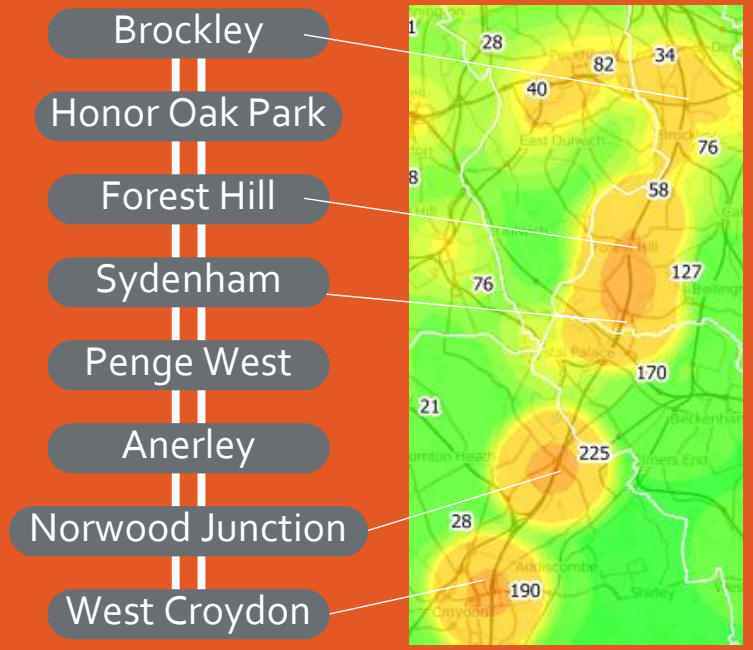
Transport: Customers Affected by London Bridge Closure

Journey Origins



The customers coming into London Bridge in the morning are coming from Brockley, Forest Hill and Norwood Junction.

Journey Destinations



In the AM peak customers are travelling through London Bridge and on to locations in South London such as Croydon.

Real Time Fraud Analytics

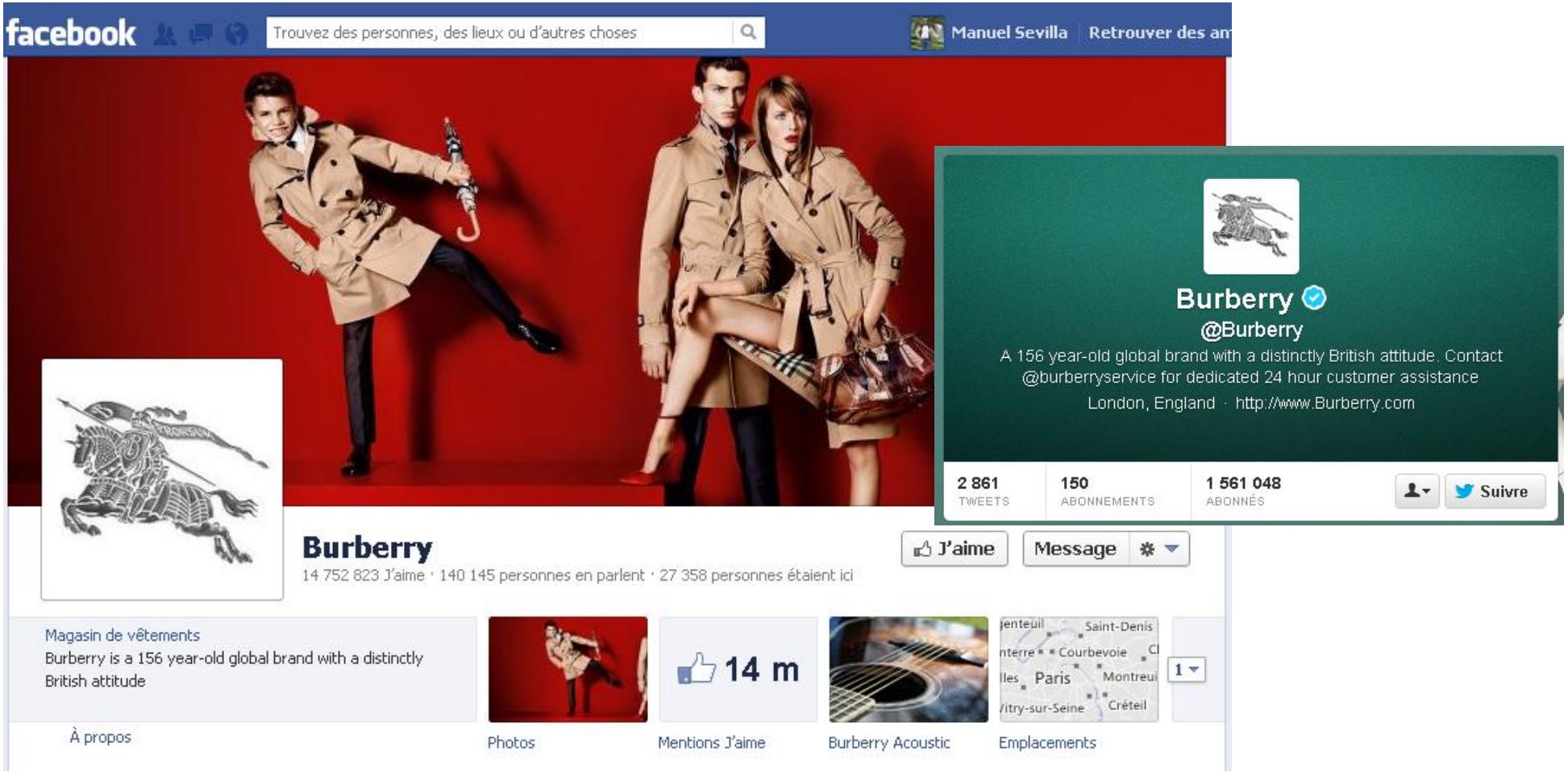
PayPal™

- **Business Challenge**
  - Fraud Detection was done too late (weekly batch)
  - Internet Frauds are often sudden and massive
  - Data sources are very diverse and come from both in or out of the company
  - Long Fraud Analytics process have lead to transaction timeouts ….
    ….and huge loss for the business

- **Solution**
  - Centralize all useful data and provide a real-time access in Big Data solution

- **Business Value**
  - 99,999% Completed Transactions
  - Reduced end-to-end Fraud processing time to 800 ms
  - Fraud processing on Credit Cards down from 45 minutes to under 4 seconds

Retail: A digitalized end-to-end company

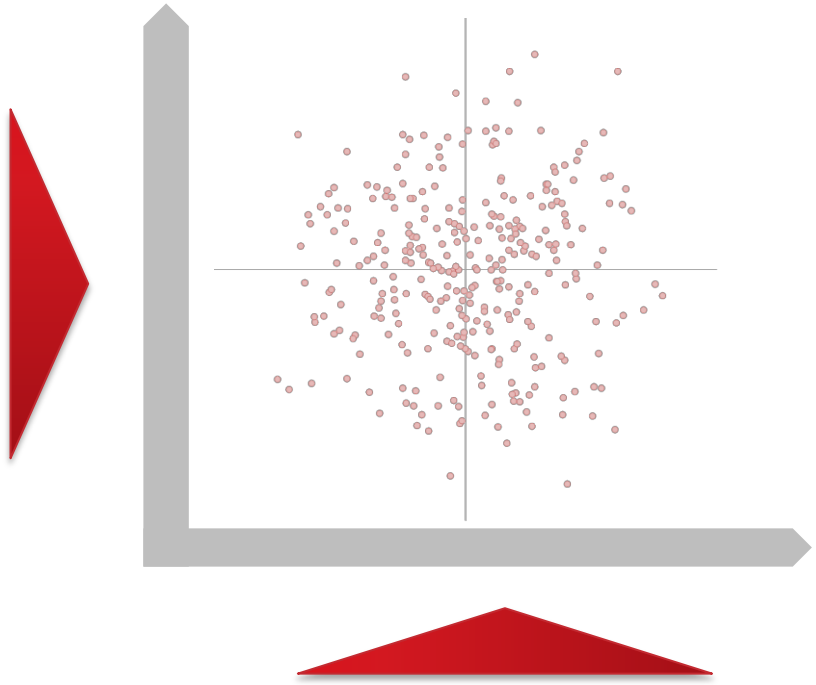


MIT Study: Digital Advantage*

The MIT Center for Digital Business

Digital Intensity is investment in technology-enabled initiatives to change how the company operates – its customer engagements, internal operations, and even business models.

Based on a study of over 400 large companies over 2 years *

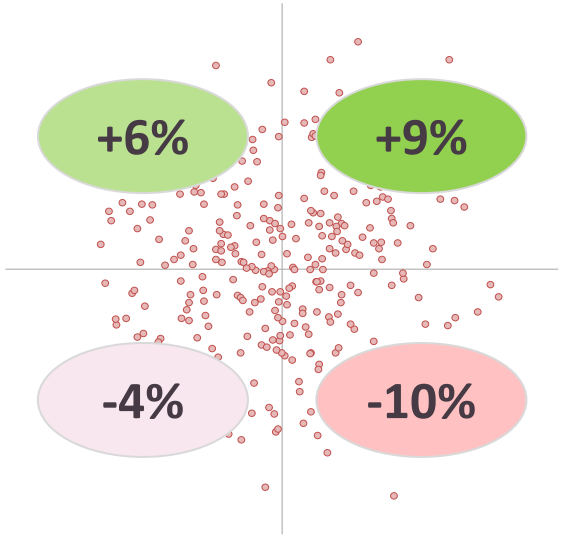


Transformation Management Intensity consists in creating the leadership capabilities necessary to drive digital transformation in the organization.

***The Digital Advantage:** How digital leaders outperform their peers in every industry MIT Sloan & Capgemini - March 2013

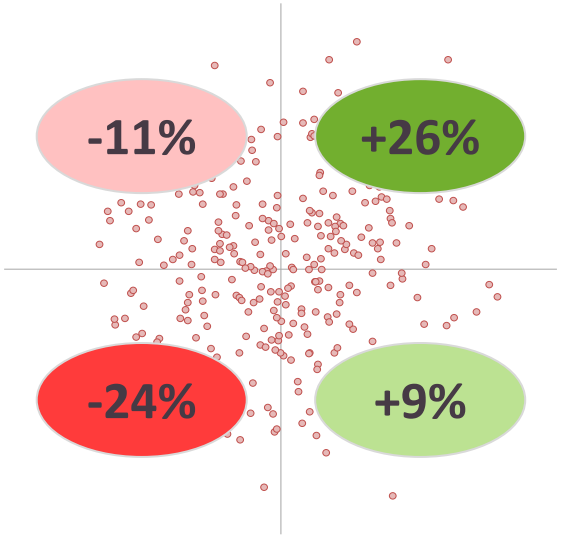
Digitally-mature companies have significantly better financial performance*

Revenue Generation Efficiency



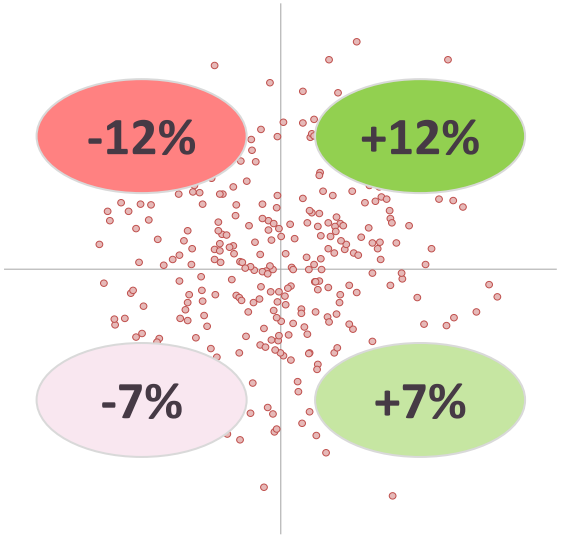
Basket of indicators:
- Revenue / Employee
- Fixed Asset Turnover

Profitability



Basket of indicators:
- EBIT Margin
- Net Profit Margin

Market Valuation



Basket of indicators:
- Tobin's Q Ratio
- Price / book ratio

* Average performance difference for firms in each quadrant versus the average performance of all large firms in the same industry for the 184 publicly-traded companies in our sample

Capgemini

CONSULTING.TECHNOLOGY.OUTSOURCING

People matter, results count.

About Capgemini

With more than 120,000 people in 40 countries, Capgemini is one of the world's foremost providers of consulting, technology and outsourcing services. The Group reported 2011 global revenues of EUR 9.7 billion.

Together with its clients, Capgemini creates and delivers business and technology solutions that fit their needs and drive the results they want. A deeply multicultural organization, Capgemini has developed its own way of working, the Collaborative Business Experience™, and draws on Rightshore®, its worldwide delivery model.

Learn more about us at **www.capgemini.com**.

www.capgemini.com