Briefing:
An Introduction to BigData (.. & Analytics)

November, 2013

CONTENTS

1. Introduction to Bigdata & Analytics 3

1.1 So What is BigData? 3

1.2 Sources & Further Reading: 5

2. Business Examples of BigData 7

2.1 Stylised Example #1: Logistics/Fleet Management 7

2.2 Stylised Example #2: Customer Loyalty & Retention 9

2.3 Further Reading: 11

3. BigData: The Size of the Prize 12

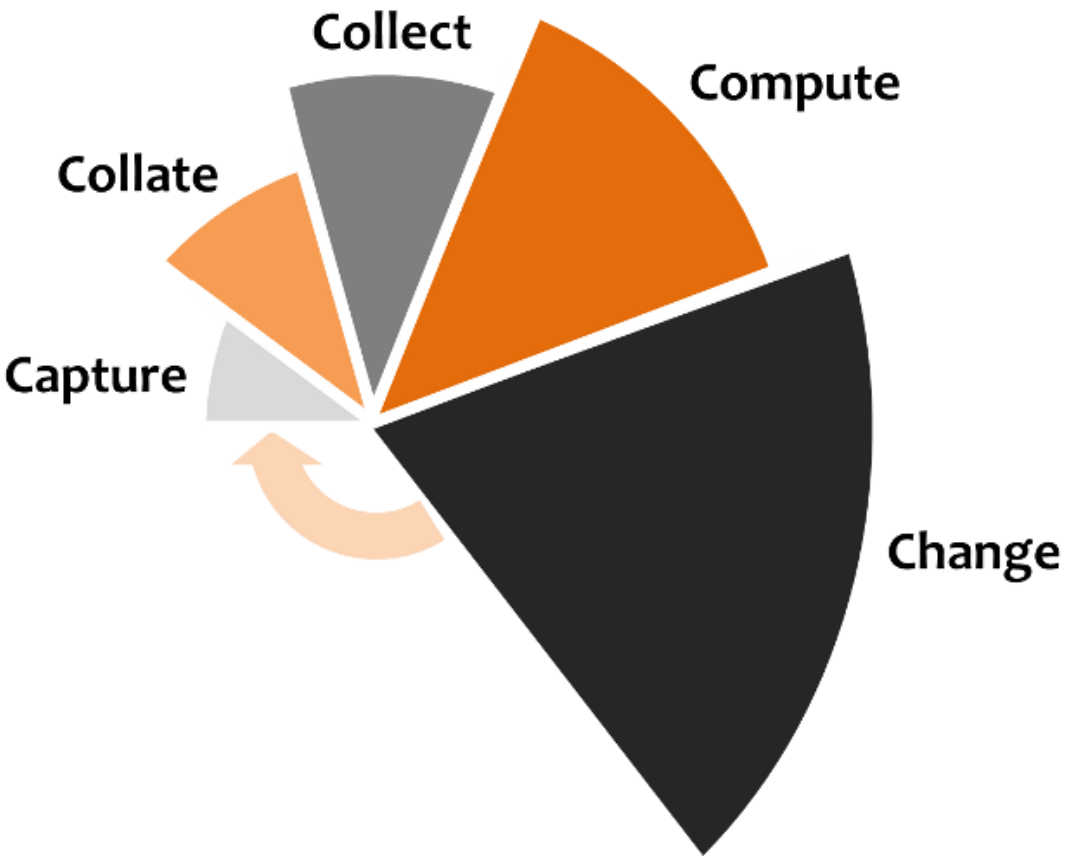
3.1 Sources & Further Reading: 12

1. INTRODUCTION TO BIGDATA & ANALYTICS

Where on Earth to begin describing such a topic as BigData? That's probably one of the biggest early challenges in starting to trawl this subject. As subjects, even Cloud, MMOG and Gamification pale into insignificance when addressing the breadth and depth of complexity involved in BigData. Indeed we touched upon an element of BigData in Gamification (see Long Introduction for more details) alluding to the fact that until the advent (in many cases the depth of this has yet to be achieved) of BigData we did not have the ability; in anything more than a rudimentary sense; to understand, contextualise and predict the behaviours of customers at all. The evolution of computing which was covered very briefly under the Cloud heading of this website brought with it a new age of innovative capabilities and efficiencies with first governments, then companies and latterly individuals the world over able to harness computer processing to drive better services and be more organised than ever before. The arrival recently of SmartPhones has now taken the individual element of this evolution to new heights with the millennial generation expectant of immediacy of knowledge or facts wherever they are at whatever time. Even as recently as 2010, some commentators even mused that we were living through a stagnant boring age. Now though, any thoughts such as these are all about to irrevocably and stupendously change. We are truly witnessing "The Beginning of Infinity".

1.1 So What is BigData?

At the highest level of aggregation BigData can be viewed as comprising of five specific functions each of which requires careful design across people, process and technology. In the simplest terms those functions can be viewed as being:



Each of the five functions is described in more detail below but it is worth pointing out at this juncture that the size of the pie above is indicative of the size of the challenge posed by that function for the implementing organisation.

Capture – the switch from analogue to digital data which began during the 1990s provided the first portents for the today's tsunami of data growth. From 1996 to 2007 digital storage in enterprises went from 0.3 Exabytes to 276 Exabytes. This foundation was then further fuelled by the growth in mobile devices, laptops, sensors & actuators, Machine-to-Machine (M2M), the availability of cheap storage disks, and latterly the promiscuity of data within social media resulting in an estimated total data storage need of approximately 35 Zettabytes by 2020. To further compound the difficulties of managing this level of data, the vast majority of this data (~85%) is unstructured data such as photographs, videos, tweets, posts, blogs, logs, email and webpages. Even conceptualising this growth is difficult, by way of tangible examples the number of live websites has risen from 1.2 Million to 367 Million between 1996 and 2011, and in the same period the number of mobile phones from 145 Million to ~6 Billion. The bottom line is that simply capturing a subset of this data is a monumental task for any company.

Collate – of course having the data is one thing, having it in a useful format is quite another. The second function of the lifecycle is the ordering of this data into some form of computing ingestible structure. This is the domain (in business intelligence vernacular) of Extract, Transform & Load (ETL). In the structured data world the problems of duplication, erroneous inputs, duplications, and absent cells are well known and to a greater or lesser extent the modern day ETL systems are capable of handling these issues. But as we have seen the true prize in BigData is handling the immensity of unstructured data which is now beginning to be addressed through textual analysis techniques, advanced data cleansing/pre-processing, contextual analysis techniques, and advances in natural language processing. To make this more real when we are scanning unstructured inputs a common issue is one of acronyms. To a human the "context" of the document, post, blog or article denotes the probable translation but this is very much more difficult for a computer, especially at the desired processing speed.

Collect – once we have developed and bounded the expansiveness of the capture, and then decided the specifics of collation, we then have to stage it in vast quantities in some form of computer & human usable containers. The actual choice of containment will greatly affect the efficacy of the ability to trawl the data. The physical choices come first to entertain answers to questions such as - How much should be cached? How much on what sort of disks? Considering an offline archive for later retrieval? Are we using de-duplication and compression? Beyond the physical layer choices are the decisions concerning which data store and which collated data to go into which store. The first data store is the continued leverage of the traditional relational database which remains a key asset in the BigData arsenal today. Developed during the 1970s based on the initial work of E F Codd, Donald Chamberlain and Raymond Boyce at IBM, the Relational Database Management System (RDBMS) is a powerful tool for structuring and querying data. Today's databases are highly optimised systems that can search and find data using the Structured Query Language (SQL) within the clean carefully crafted, filled, indexed, non-duplicated table structure of an RDBMS. The second data store is the NoSQL database which emerged in 1998 but has really gained traction in the last five years. It departs from the relational approach, does not use tables, and does not use SQL to manipulate data. Instead the approach is to utilise "key, value" stores, document stores, and graphs. NoSQL is much faster and scales to enormous sizes for the types of work with unstructured data that today's BigData demands.

Compute – if you were being pernickety or strict in terms of creating a linkage between the title 'BigData' and its function you could arguably stop now at the three functions described thus far. This is not the case though, so far we only have data, lots of data but data nonetheless, and data is not in-and-of-itself very useful to human beings. What we actually want at the end of this is

information and better still 'actionable' information. This is where the compute function comes into play. Without getting into a debate on taxonomy from a business perspective compute is really the juncture of processes, programs, mathematics and visualisation techniques employed to deliver meaningful insights, models and recommendations. Together these techniques have come to be called "Business Analytics" but are very closely related and depending on your viewpoint either overlapping or a subset of "Business Intelligence".

Change – at this stage the scene is set for the final function of BigData. We have captured, collated, collected and computed the data into information. Change is about using that information to take decisions that result in action. All of the preceding steps are for nought if the information you obtain cannot be translated into actual transformations within the business to deliver efficiencies, optimisations and ultimately profits. The goal of the change aspect, and the rational in the graphical view of the five functions as increasingly large segments of pie, is to deliver programme(s) for organisational change. This is by far the most difficult of the five functions of BigData to achieve. However, McKinsey estimated that impact of these changes as capable of delivering value of up to $300 Billion of value in US Healthcare, €250 Billion in European Public Sector and $600 Billion consumer surplus globally.

1.2 Sources & Further Reading:

Other links to excellent introductions to BigData & Analytics are:

http://logisticsviewpoints.com/2012/12/05/the-problem-with-big-data/

http://www.oxbridgebiotech.com/review/featured/commercial-potential-big-data-biotech-panel-discussion-obr-oxford/

http://blog.gopivotal.com/news-2/20-examples-of-getting-results-with-big-data

http://content.time.com/time/specials/packages/printout/0,29239,1971133_1971110_1971120,00.html

http://casnocha.com/2010/05/slowing-rate-of-change-and-tech-innovation.html

http://www.prweb.com/releases/2013/9/prweb11116362.htm

http://marketing555.wordpress.com/2012/10/02/the-big-and-small-of-data/

http://practicalanalytics.wordpress.com/2013/10/23/market-sizing-analytics-and-big-data/

http://www.forbes.com/sites/gregsatell/2013/09/13/confused-about-big-data-here-are-5-things-you-need-to-do/

http://www.statisticbrain.com/total-number-of-websites/

http://data.worldbank.org/indicator/IT.CEL.SETS.P2?page=3

http://stats.areppim.com/stats/stats_mobilex2012.htm

http://www.couchbase.com/sites/default/files/uploads/all/whitepapers/NoSQL-Whitepaper.pdf

https://archive.org/stream/cihm_39184

http://en.wikipedia.org/wiki/NoSQL

http://www.datastax.com/wp-content/uploads/2013/06/DS_WP_Implementing_a_NoSQL_Strategy.pdf

http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Needham, Jeff ‘Disruptive Possibilities’

Deutsch, David ‘The Beginning of Infinity’ (2011, Allen Lane)

Mohan, Gary ‘The Big Data Gold Rush’ (2013, www.plainprocess.com)

Barlow, Mike ‘The Culture of Big Data’ (2013, O’Reilly)

Dumbill, Edd ‘Planning for Big Data’ (2012, O’Reilly)

Slocum, Mac (Editor) ‘Big Data Now: 2012 Edition’ (2012, O’Reilly)

2. BUSINESS EXAMPLES OF BIGDATA

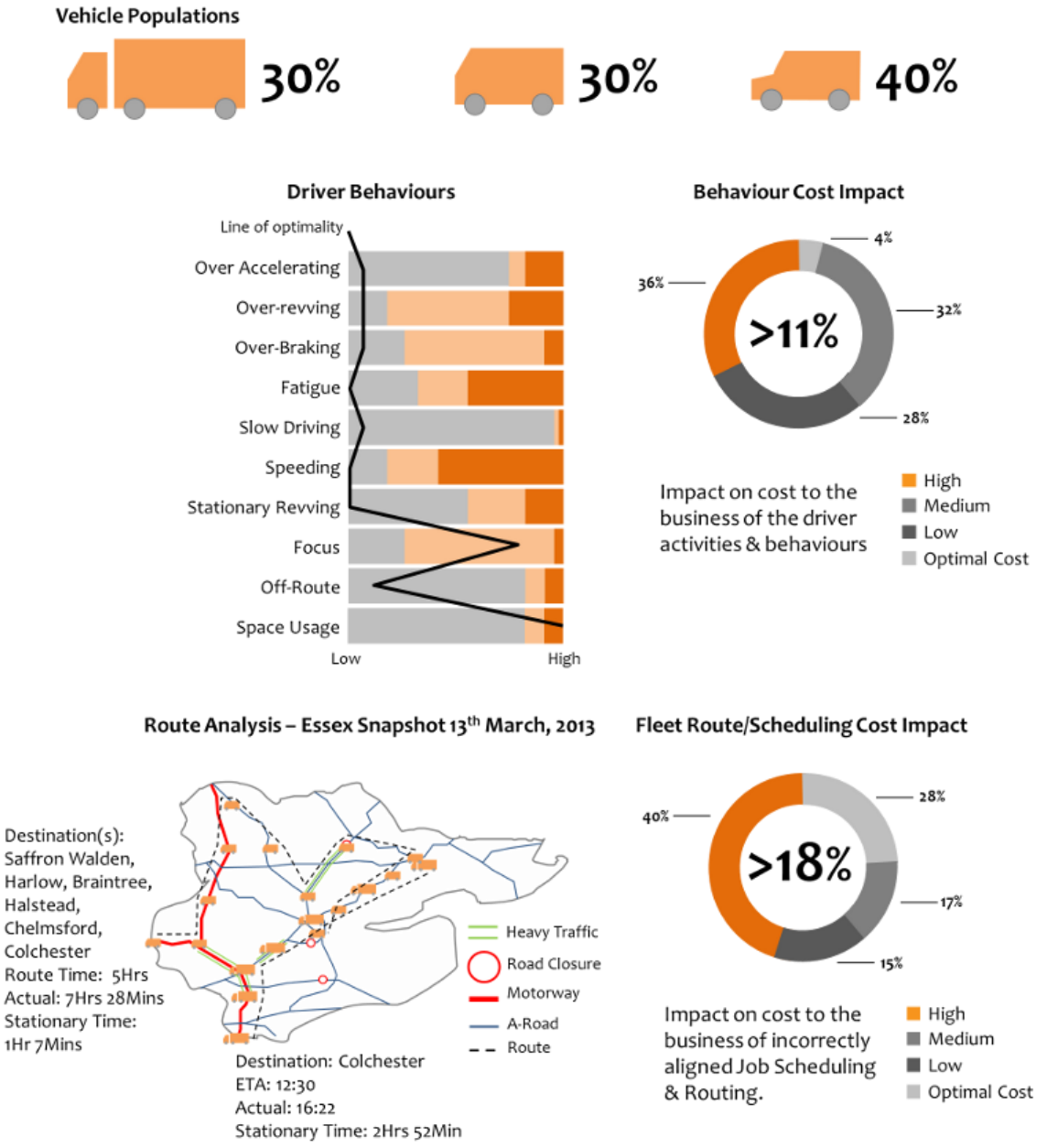
Before delving more deeply into each of the constituent elements that comprise each function of BigData and their importance in the cycle, it is worth expounding what precisely can be done using this BigData then? NOTE: The following examples are not real world case studies – they are provided to help readers understand "how" BigData could be used across the five functions and to help those trying to build a business case for a BigData project.

2.1 Stylised Example #1: Logistics/Fleet Management

John is the CEO of a Logistics company which has seen enormous growth of more than 260% over the past five years. This explosive growth has come at the expense of profitability which has seen a decline from peak values during the full financial year of 2008, which yielded a Gross Margin of 36.4%. At the final board meeting in 2012 John announced that Gross Margin had now declined to 28.6% and that the company needed to take stock of their situation to reluctantly slow their growth trajectory to focus on margin gain. Jane the CIO convinced the board that a BigData initiative could deliver the savings required without the need to slow their growth. The Board agreed to invest in a BigData initiative on the basis of a top & bottom line business case she presented which would deliver at least 250% Return on Investment (ROI) with a payback of only 12 months founded on fuel costs optimisation, fleet optimisation and increased customer satisfaction. An examination of the business case revealed:

- **Fuel cost optimisation** – the first element of the business case outlined an expected hard tangible saving of 7% on fuel costs on the basis that the BigData capture with the correct analytics would enable better driver training, improved driving techniques and optimised intra-day ad-hoc job routing;
- **Fleet optimisation** – the second element provided for another hard tangible saving expectation of 14% from a re-alignment of their fleet encompassing a divestiture of some larger assets coupled with investment in smaller vehicles;
- **Increased customer satisfaction** – the final element of the case was viewed by the CFO and COO as more "woolly". They were unwilling to wholeheartedly accept a direct causal link between improvements in customer satisfaction and top line revenue enhancement. However on the basis of fuel and fleet optimisation savings they were willing to support the plan and any potential revenue gains would be a "bonus".

With investment in place, Jane could set her plan in motion during January and February to fit out every vehicle in the fleet with Global Positioning System (GPS) trackers, Engine Management System (EMS) trackers, upgraded the central job scheduler, routing and manifest production systems and rolled out enhanced robust handheld devices with Electronic Proof of Delivery/Collection (ePOD/ePOC). With the "capture" devices in place Jane focused the BigData team on "collate" and "collect" aspects of the initiative. In particular they built a large scale NoSQL capability which enabled each vehicles positional and engine control unit data to be stored during each second of operation. This equated to almost 25GB of data per vehicle per day. The final IT centric aspect of the project "compute" was completed during March which analysed the all the input data "captured, collated and collected" during March to produce a series on insights into driver behaviours, fleet assignment, job scheduling and much more. An example of the type of output she could present to the Board is presented below in infographic form:



On the basis of the presentation and the findings of the BigData initiative the Board agreed to commence with the final step "Change". In particular the CFO & COO were ardent supporters of the "Change" initiative driving the business to deliver the transformation necessary to derive the results. By the end of November 2013 the business could have already recouped the cost of the BigData project, customer satisfaction had increased by 25%, gross margin had improved and the business had continued to grow.

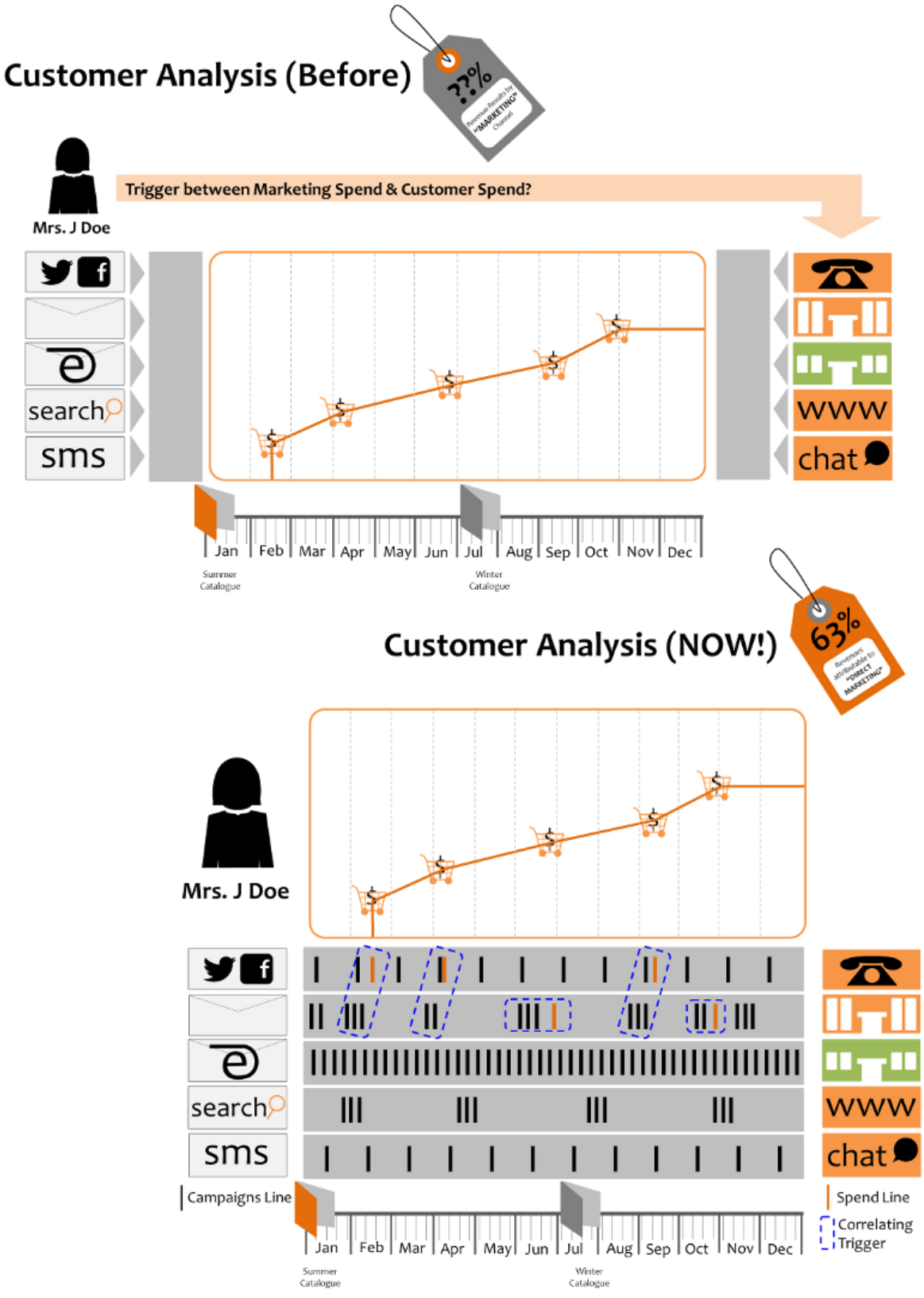
2.2 Stylised Example #2: Customer Loyalty & Retention

Jenny is the CEO of a large retail company which has suffered during the recession period following the global financial crisis. Sales have dropped by almost 2% Year on Year (YoY) for the past three years. Worryingly, the latest trend hints that the downward momentum may be increasing. At a Board meeting in December of 2012 she asked for suggestions from each of the Board members. David the CIO presented a plan based on a BigData & Gamification initiative that he called "Project Cassiopeia". The Board agreed to invest in this BigData initiative on the basis of the top line business case he presented which would deliver at least 175% Return on Investment (ROI) with a payback of only 12 months founded on revenue increases from channel realignment, enhanced loyalty rewards and point of sale upsells. An examination of the business case revealed:

- **Channel realignment** – the first element of the business case outlined a shift of focus between the existing retail channels. Although already a multi-channel retail business there was little understanding of the channel sales split for individual customers. Specifically, customer Mrs. J Doe was known to have spent $625.00 but the business was not aware as standard by which channel Mrs. Doe had made those purchases. More importantly, there was even less knowledge of whether a channel specific campaign had had any cross-channel purchase correlation or any insight into the potential benefit of coupon issuance across the channels. David suggested that with BigData he could derive a unique customer insight for the business to help decisions on cross-channel offers and campaigns;
- **Loyalty rewards** – feedback on the current loyalty reward scheme was not positive with less than 5% of customers reporting high degrees of satisfaction. This was being addresses by marketing but this exercise was leaning towards a commercial enhancement of the offer. David had studied online the results of gamification in loyalty schemes and was uncertain that simply enhancing the commercial aspect of the offer would derive the results required. He planned to revamp the entire loyalty scheme using BigData together with some game mechanics (friends, feedback and fun) and the deployment of a new Mobile Application to deliver greater customer engagement and sales;
- **Point of sale upsells** – the final element of the case was to leverage the BigData understanding of customer purchase history together with predictive analysis to issue "on-the-fly" offers over web and in-store channels, such that at the point of purchase of any item the customer would be recommended a complimentary product based on a unique (to the customer) high propensity for purchase at point of sale.

Taken together these areas would be expected to deliver a 3% increase in revenues over the 12 months following rollout. With investment in place, David set his plan in motion during January and February to build a new customer focused unstructured data store capable of storing inputs from a diverse set of in-house systems but also linking to external social media sites to retrieve customer postings relating to their business. In parallel with this "capture" phase the "collate" and "collect" teams focused on how they could clean and convert this data into usable chunks for collection into data stores that would to enable real-time review. A separate gamification team designed a new upgraded loyalty programme linked to social media that provided status rewards to customers based on their continued spend and linked this to external partner organisations to increase the utility to the customer. This was communicated to customers across multiple channels with immediate uptake. The final IT centric aspect of the project "compute" was completed during April which analysed the all the input data "captured, collated and collected" from 2012 to produce a series on insights into customer sentiment, customer

behaviour across the channels and predicted spending scenarios based on enhanced offers at point of sale. Some of the output she presented to the Board is presented below in infographic form:



On the basis of this and other similar presentations of the findings of the BigData initiative the Board agreed to commence with the final step "Change". The business engaged on a six month multi-streamed transformation programme across marketing, external partners, sales, customer service and e-channels. By the end of November 2013 the business had already begun to experienced visible results in the third quarter with sales revenues no rising slightly. The drive during the fourth quarter is now to focus marketing on those channels that work for each specific customer on an individual basis – for instance in the example of Mrs. J Doe this could revolve around the introduction of additional one-on-one "coupon" offers as there is a strong correlation between her spend and a coupon direct mail campaign. In addition Mrs. Doe clearly spends very little time online and on electronic channels so a considering an inducement strategy via direct mail to elicit her participation in online world should be considered. Also during the fourth quarter the introduction of point of sale coupon or loyalty reward points generation with be introduced linked to the SmartPhone apps to provide further tangible direct linkages between the e-channels, customer interaction and customer spend.

2.3 Further Reading:

http://gigaom.com/2011/07/17/5-real-world-uses-of-big-data/

http://smartdatacollective.com/bernardmarr/150126/revealed-top-5-big-data-use-cases-your-ceo-will-love

http://iveybusinessjournal.com/topics/strategy/four-strategies-to-capture-and-create-value-from-big-data#.UpIKT-IQdmo

http://www.theguardian.com/news/2013/nov/22/rapid-development-in-big-data-analytics-has-led-to-increased-investment

http://www.datasciencecentral.com/profiles/blogs/bi-vs-big-data-vs-data-analytics-by-example

http://www.sas.com/big-data/

http://data-informed.com/business-problems-suited-to-big-data-analytics/

http://www.keywebmetrics.com/2013/07/big-data-visualizations/

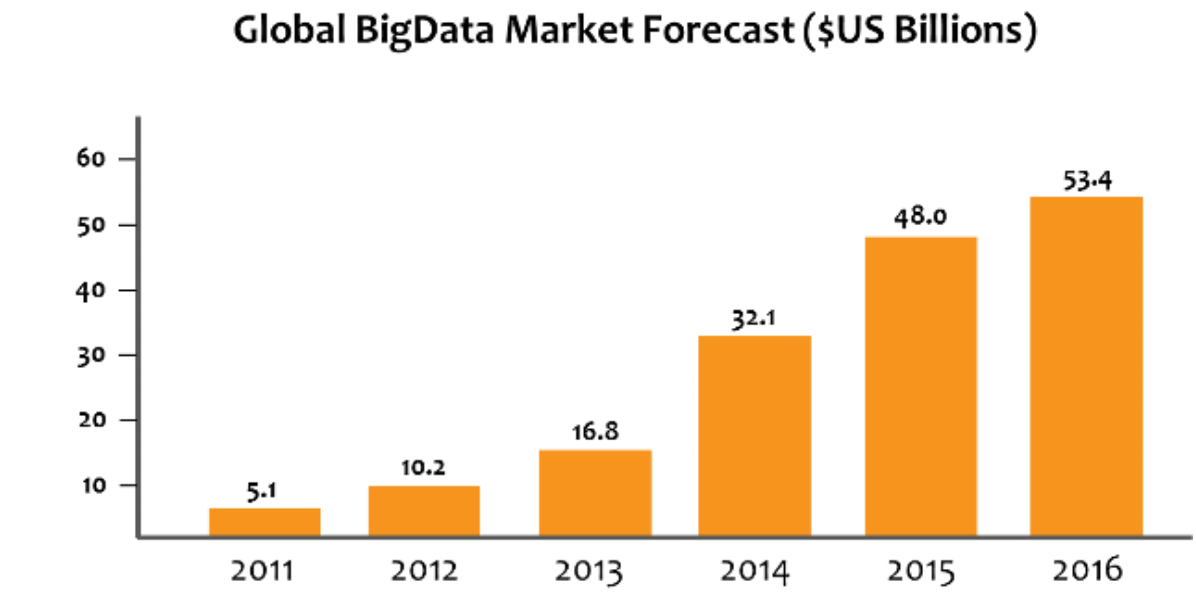
http://www.analyticbridge.com/profiles/blogs/bi-vs-big-data-vs-data-analytics-by-example

http://www.ibmbigdatahub.com/presentation/big-data-retail-examples-action

.. and there are many many more online ..

3. BIGDATA: THE SIZE OF THE PRIZE

The final part of this introduction provides a high level view of the estimated size of the BigData market as a guideline to weigh the significance it plays globally already. The BigData market one of the fastest growing IT markets in the world (arguably the fastest) and is estimated to achieve global revenues of $xxx Billion by 2020 with some projecting a Compound Annual Growth Rate (CAGR) of 58%:



http://wikibon.org/wiki/v/Big_Data_Market_Size_and_Vendor_Revenues

3.1 Sources & Further Reading:

http://www.forbes.com/sites/siliconangle/2012/02/17/big-data-is-big-market-big-business/

http://wikibon.org/wiki/v/Big_Data_Market_Size_and_Vendor_Revenues

http://gigaom.com/2013/01/08/idc-says-big-data-will-be-24b-market-in-2016-i-say-its-bigger/

http://www.prweb.com/releases/2013/9/prweb11116362.htm

http://bits.blogs.nytimes.com/2012/03/07/idc-sizes-up-the-big-data-market/?_r=0

http://www.digitaljournal.com/pr/1591025

http://www.marketsandmarkets.com/Market-Reports/big-data-market-1068.html?gclid=CMvwvYXu_boCFWkCwwod13kAzg