BIG DATA STRATEGY

SELECTED CASE STUDIES

Edited by
Debapratim Purkayastha

Big Data Strategy

Selected Case Studies

Edited by

Debapratim Purkayastha

www.icmrindia.org | www.ibscdc.org

ICMR – IBS City Office, #52, Nagarjuna Hills, Punjagutta,
Hyderabad – 500 082, Telangana, India.

ISBN: 978-81-314-2846-7

Published by
Case Research Centre
ICMR – IBS City Office, #52, Nagarjuna Hills, Punjagutta,
Hyderabad – 500 082, Telangana, India.

Mobile: +91 9640901313.

e-mail: casehelpdesk@ibsindia.org

www.icmrindia.org, www.ibscdc.org

Layout and Cover Page: P.S.R.A.S.V. Prasad

PREFACE

Today 'big data' is one of the most used corporate buzzwords. The term entered the corporate lexicon shortly after Roger Magoulas of O'Reilly Media coined it in 2005 to refer to a large set of data that is nearly impossible to manage and process using traditional business intelligence tools and techniques. Later, an October 2012 *Harvard Business Review* article likened the advent of big data to a management revolution. The following year, 'big data' was declared as the most used corporate buzzword, and by that time, it had also become amply clear that it was not merely a buzzword – it was here to stay.

We are living in the information age, and the last decade has witnessed a tremendous proliferation of big data. The world's leading information technology (IT) research and advisory company, Gartner, Inc., defines big data as "high-volume, high-velocity, and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making". To volume, velocity, and variety, Mark van Rijmenam added another 4Vs in his 2014 book "Think Bigger: Developing a Successful Big Data Strategy for Your Business" – veracity, variability, visualization, and value. According to him, these 7Vs are essential to "better explain the impact and implications of a well-thought through Big Data strategy."

The explosion of digital data in recent years has provided the opportunity to organizations to leverage this information for real time and effective decision making. In the new milieu, the constraints of collecting and storing information have been largely taken care of, making data widely accessible. However, what organizations actually do with this information is what is more important. The challenges regarding how to make sense of the information and leverage it for effective decision-making still remain.

Both the resource-based view of the firm and the knowledge-based view of the firm emphasize the importance of information/knowledge as a key resource that can be leveraged by the firm to gain a competitive advantage. In this era of digitization, it is getting increasingly more necessary for organizations to get their big data strategy right. So, what is big data strategy? According to Datamation, an organization's big data strategy "encompasses its approach to storage, analysis, data frameworks, decisions about data models, and lots – and lots – of careful planning."

Of late, many organizations cutting across industries have jumped on the big data bandwagon. Notwithstanding the various benefits of having a big data strategy, companies have to navigate various new challenges when facing the big data phenomenon. These challenges range from capture, curation, storage, search, and sharing, to analysis and visualization. According to some experts, the complications are not "just the storage, I/O, query, and performance, but also the integration across heterogeneous, interdependent complex data resources for real-time decision making, collaboration, and ultimately value co-creation". These challenges are not IT problems alone; they involve an information problem and a strategic issue. So in addition to technology, a company's big data strategy has important implications for its leadership, talent management, decision making, and organizational culture. Moreover, privacy and data security have also become key concerns in the wake of evolving cyber threats such as malware, ransomware, and increased instances of data breaches. So, the issue of 'vulnerability' also needs to be addressed and there is a good case for adding this to the existing 7Vs of big data strategy.

The case studies included in this book look at issues and challenges in formulating and implementing a big data strategy in some of the top companies in the world. The book starts with cases that show the growing importance of data for organizations cutting across different industries. It looks at ways in which organizations are trying to capture and employ data and the associated challenges. For instance, there are case studies on global furniture retailer IKEA; an online fashion startup, Stitch Fix; and a leader in parks and resorts, World Disney World. The book then presents case studies on IBM, an IT behemoth; Amazon, a leading e-commerce and technology company; Netflix, a leader among digital streaming companies; Procter & Gamble (P&G), a leader in the consumer packaged goods industry; and, Baidu, a Chinese tech giant. Together, these case studies highlight different aspects of big data strategy.

The collection is a good mix of popular contemporary case studies as well as classic case studies that have withstood the test of time. For instance, the Netflix case study has been a bestselling case study right from the time it was published in 2013. Of the eight featured case studies, five are global bestselling case studies on the topic. In addition to raising critical issues, these cases are also a source of best practices in this emerging area.

The first case in this collection is about IKEA. Using social intelligence acquired through social media monitoring and incorporating this intelligence to

strategically change the approach to evolve and attain a business objective empirically, is a challenge. IKEA embraced social media by making its presence felt on popular websites like Facebook, Twitter, Pinterest, and Google+. For IKEA, unique in its corporate structure, embracing social media was a turning point. IKEA formally set up its "Listening Hub" in 2014, acknowledging the importance of social media listening. Prior to setting up the "Listening Hub", IKEA encountered situations that underlined the need to set up a formal Social Listening Hub. This case attempts to understand the importance of social media monitoring and the advantages associated with it, specific to IKEA, illustrated through situations identified and confronted by IKEA in different countries, prior to the setting up of "The Listening Hub". It also helps to understand the various objectives of social media monitoring that IKEA tried to achieve, and the challenges faced by the furniture retailer in implementing its social media strategy.

The next case study discusses how Stitch Fix, an online personal styling service founded by Katrina Lake, employed algorithms to select apparel for its customers according to their style, fit, and price preferences. In the course of recommending apparel for its customers, the company employed algorithms related to, among others, collaborative filtering and mixed-effects modeling. The case also describes how human beings were used to augment algorithms in styling customers. It also points out how the algorithms learnt from the feedback given by the customers for a specific shipment to select styles for subsequent shipments. The case describes how algorithms were employed by the company to perform other key tasks such as allocating a shipment request to a warehouse and computing the amount of and the type of inventory that the company should purchase.

The case goes on to describe how machine learning algorithms were employed by Stitch Fix in new product development – the company's in-house brands eventually contributed over 20% to its turnover. Data mining was employed by Stitch Fix to enter new categories such as plus size for women and premium-brand offerings. The case also describes the challenges faced by the company such as limits to the amount of shopping that its customers could do, intense competition from e-commerce players such as Amazon, and increasing advertising expenses eating into the company's profits.

The next case study is about Walt Disney World in Orlando, Florida. Over the years, Disney had been focusing on improving customers' experience at its parks and various other attractions. It had taken care of minute details to

ensure the comfort and entertainment of customers and had always updated its services to match the changing environment. Moving in alignment with its focus, in 2013, Disney launched a vacation management system, MyMagic+ (MM+), which could streamline its parks' operations and enhance customer experience. The MM+ system had three components — the My Disney Experience website, a mobile app, and the technology driven devices, MagicBands. The case focuses on the benefits and drawbacks of this MM+ system. From the customers' perspective, the MM+ system offered a comfortable all-in-one radio-frequency identification device, MagicBands, which could deliver a more enjoyable and well-planned Walt Disney World experience, whereas for Disney, the system could track the visitors' movements and thus help in generating a vast database. However, concerns were raised about the possibility of the model succeeding in future. Analysts studied these technology driven devices and their working and expressed concern over the possibility of the bands being cloned or the information being retrieved by smartphones. Questions were raised on whether the use of these bands would jeopardize the privacy and security of the personal and financial information of customers. Although Disney tried to put these doubts to rest, the MM+ being a project in its trial phase could not guarantee success in the coming days. Further, analysts were skeptical about customers' willingness to enjoy the Disney services, while putting information valuable to them at risk.

IBM has implemented a big data strategy where the company offers solutions for storing, managing, and analyzing the huge amounts of data generated daily and equips large and small companies to make informed business decisions. The company believes that its big data and analytics products and service will help its clients become more competitive and drive growth. In order to have an extensive portfolio of big data technologies and solutions, spanning services, software, research, and hardware, the company acquired around 30 companies, spending US$16 billion over these big data acquisitions. Cognos, SPSS, Netezza, and Vivismo were a few of the many companies that became integrated into its big data platform. On October 29, 2013, IBM completed the acquisition of the Ireland-based The Now Factory. In 2011, the company also invested US$100 million in the research and development of big data services and solutions.

Leading e-commerce company Amazon.com, Inc. utilizes its big data resources to improve its performance. Being the dominant retailer on the Internet, Amazon has a vast database regarding the tastes, preferences, and

previous purchasing history of its customers. Amazon leverages its big data resources to give more relevant product recommendations and improve its customer care quality. Banking heavily on its big data resources, it upgrades its customer recommendation system. The 360 degree customer profiles developed using big data resources enable Amazon to create highly personalized marketing messages for its customers. Apart from improving its business performance, big data resources are also put to some innovative uses like preventing theft of merchandise from its warehouses. Like IBM, Amazon has also made its big data services available to other smaller e-commerce companies with limited resources. In 2010, Amazon launched an innovative service called Amazon Webstore that allowed smaller companies to build shopping portals based on Amazon's e-commerce platform. This service was highly successful and was adopted by even big companies like Timex and Samsonite who did not want to develop their own e-commerce system. Amazon's suite of cloud based Internet services known as Amazon Web Services (AWS) came out with solutions like Kinesis to help smaller companies implement big data easily. Some analysts opined that as of end 2013, Amazon had all the pieces of the big data puzzle but the firm would have to put these together effectively to emerge as a dominant player in this space and challenge the likes of IBM.

Netflix, the largest player in USA in video streaming services, has a global streaming subscriber base of around 33 million. Netflix has the complete details of the viewing patterns of each of its subscribers, including aspects such as when they hit the pause button and whether they switch off before the credits roll. It has deployed this data to come up with recommendations for each of its subscribers. According to experts, Netflix's data mining competencies received a boost when the company shifted its information technology infrastructure to the cloud. This gave Netflix the flexibility to scale up — and it opted for the Hadoop architecture. Netflix also employed the huge dataset it had about the viewing patterns of its subscribers to get into original programming. It was so confident about the popularity of the original version of the television show "House of Cards" and of the director and the lead actor of the show's remake among its subscribers, that it bought the exclusive licensing rights of the show's remake for US$100 million. The show, as Netflix's executives had predicted, proved to be a success. Netflix planned to come up with more original content shows relying on its assessments of the viewing patterns of each of its subscribers. The case also discusses some of the concerns that experts had about Netflix's big data technologies infringing on the privacy of its subscribers. Experts also raised concerns about the outages faced by Amazon Web Services,

the vendor of cloud computing services to Netflix, which resulted in the latter's site being down three times. Industry observers were apprehensive that players such as Netflix would constrain artistic creativity by employing big data to come up with predictable content merely on the basis of the past viewing patterns of their audiences.

P&G, the world's largest consumer packaged goods company, was on course to achieve 100% digitization of all its activities: collaborations among employees, market research, evolution of new products, production systems, supply structures, distribution, advertising, and customer relationship management. This case study is about the challenges faced by the new CIO of P&G, Linda W. Clement-Holmes, in taking the big data strategy of the leading consumer packaged goods company forward. Under her predecessor, Filippo Passerini, P&G had leveraged big data successfully in all its business decisions – marketing, product development, supply chain, etc. The company collected consumer data and other data from multiple touch points as part of its digitization drive. The case elucidates the benefits that the company had derived from big data, including cost savings and the speedier roll-out of new products. The case also explores the potential downside of an excessive focus on big data and digitization, including diluting the human touch in the company's interactions with its customers, and other potential risks of employing big data. Linda faced a challenging situation as she tried to take the company's big data strategy forward at a time when P&G was facing difficult times.

Lastly, this selection also has two case studies on Chinese tech giants, Baidu and Alibaba. Baidu, which had been betting big on bringing about a global revolution in Artificial Intelligence (AI) through some of its prestigious projects like Apollo, its autonomous driving program and DuerOS, an open platform to support applications in AI. With a share of over 44.5% of the mobile search market and 80.8% of the internet search market in China, Baidu had access to colossal amounts of demographic data. The company, which had over 70 million daily active mobile search users, 700 million internet users, and 6 billion search queries per day, had gathered massive amounts of data backstage, which it strategically channelized into the big data segment. Baidu's big data vision was driven by three core strategic functions which were classified into three functional components – Open Cloud, Data Factory, and Baidu Brain. It built three state-of-the-art research laboratories in Silicon Valley. In 2013, it built Deep Learning Institute with its focus on certain areas – image recognition, machine learning, robotics, human-computer interaction, 3D

vision, and heterogeneous computing. Baidu also established the Silicon Valley Artificial Intelligence Laboratory and the Beijing Big Data Lab. The laboratories were driven by the mission of developing AI technologies which would impact at least 100 million people. The infrastructure was equipped with a powerful combination of deep learning, large datasets, and high performance computing.

The three clusters of Baidu's big data vision were Tiansuan, Tianxiang, and Tiangong. Tiansuan, in principle, integrated large datasets with AI. Tianxiang was a multimedia platform that enhanced user experiences in online interactions. Tiangong, was the Internet-of-Things (IoT). It integrated its cloud applications with internet connected devices, and offered a range of IoT services for users in a cost-effective way. In addition to big investments in Apollo, Baidu had partnered with more than 130 enterprises to roll out the DuerOS applications globally. The technology was already integrated into devices like smartphones, televisions, and home appliances. Baidu's big data initiatives were not confined to business and commercial utilities alone. Many of its applications demonstrated and promised to make an impact on several sensitive social and economic issues like controlling poverty, diseases, stampedes etc.

In the final case study, we have Alibaba, the world's largest retail platform, which started off in 1999 as an e-commerce company involved in overseas and domestic wholesale trade. It went on to expand its business to include consumer e-commerce, online payments, online technology marketing, a web browser, logistics, mobile apps, online entertainment, social networking, etc. Being active in so many spheres meant millions of transactions, buyers, and sellers. Alibaba captured data about the buying / selling habits, transactions, orders, etc. of these buyers and developed data mining capabilities to recover such data and build profiles of the buyers. The data thus collected was shared with the sellers to help them bring out better products for the customers and tweak their products based on demand. This led to the building of a big data driven marketing platform called Alimama, which helped advertisers with digital marketing efforts.

The next step was cloud computing through Aliyun and a cloud computing platform through which customers were provided with customized landing pages and choices that suited them the most. Alibaba used these big data capabilities to develop a credit scoring system to provide financial assistance to merchants who had no credit history and to venture into micro lending and credit scoring services. It went on to develop a fraud detection

management system from its huge data sources to identify suspicious activities and fraudulent users online and immediately take counter measures. Another major use of big data was in the area of detecting and blocking counterfeit and pirated goods. Big data was also used in several other areas like smart traffic management, stock market predictions, etc. As the Chinese economy grew, it became one of the largest in the world and the most advanced mobile economy. Alibaba then set its sights on international markets. Setting the agenda for Alibaba's future, Jack Ma, its founder and Chairman, wanted to build capabilities that would enable the company to venture into global e-commerce in a big way and help six million small businesses sell their products to 1.2 billion customers outside China. In the process, by 2036, Alibaba wanted to become the fifth largest 'economy' in the world. Will big data help Alibaba fulfill its ambitions?

Clearly, organizations have understood the role of big data and its strategic importance. While some companies have made an early foray into the big data space, others are trying to develop their own big data strategies and are also exploring ways in which they can benefit from using big data. According to a Cisco estimate, global IP Traffic will reach 3.3 zettabytes (1 zettabyte = 1 billion terabytes) by 2021, and by this time, the number of devices connected to IP networks will be more than three times the global population. In 2022, global IP Traffic is slated to reach 4.8 zettabytes. The implications for the big data revolution are vast, and organizations cutting across industries can ill-afford to not have a holistic big data strategy in place aligned to their business strategy.

Debapratim Purkayastha

INDEX

ABOUT THE EDITOR

Dr. Debapratim Purkayastha is a Professor and Dean at ICFAI Business School, Hyderabad (The ICFAI Foundation for Higher Education). Purkayastha has more than 19 years' teaching, research and industry experience. He teaches Business Strategy at IBS Hyderabad and heads its world-class Case Research Centre. He is the Consulting Editor of *'Case Folio – The IUP Journal of Management Case Studies'*; Editorial Board member for *'Case Research Journal'* published by North American Case Research Association (NACRA); and an Editorial Advisory Board of *'The CASE Journal'*, the official journal of The CASE Association (published by the Emerald Group, UK).

Purkayastha is an internationally renowned case method expert. He has won numerous case method awards including the highly prestigious 'Outstanding Contribution to the Case Method' award from the Executive Committee of the UK-based The Case Centre. In addition to featuring in the list of all time bestselling authors of case studies written over the past four decades, Purkayastha was declared as The Case Centre's #1 bestselling author of management case studies in 2016, 2017, 2018 and 2019.

He has won multiple case method awards and recognition from the Academy of Management (AOM, North America), Association of MBAs (AMBA, the UK), CEEMAN (Slovenia), EFMD (Belgium), Emerald Group Publishing (the UK), NACRA, oikos International (Switzerland), The Case Centre (the UK), John Molson School of Business (University of Concordia, Canada), China Europe International Business School (CEIBS), AESE Business School (Portugal), etc. In March 2018, Purkayastha was recognized among India's top faculty at the Career360's "Faculty Research Awards", where Mr. Prakash Javadekar, the Union Minister of Human Resource Development, Government of India, gave away the award.

Many of Purkayastha's cases are published in global editions of textbooks of renowned authors and publishers. He has conducted a number of case method workshops in leading B-schools and organizations in Asia, North America and Europe.

Case 1

IKEA's Social Media Listening & Monitoring Initiatives

Debapratim Purkayastha and Jayasri Bhimalapuram

"Social Media has opened up a unique platform for us to interact directly with our customers. Listening to what they want is what we do best, and the Big Sleepover is just one example of how we're using such instant and open feedback to better inform our marketing activity."[1]

– Lois Blenkinsop, PR and Internal Communications Manager, IKEA United Kingdom

In December 2011, IKEA launched "The Big Sleepover" happy to bed campaign in the UK. According to IKEA, its team had uncovered a group on Facebook called "I want to have a sleepover in IKEA". The group was started by IKEA fans and identification of the fan group was the seed for the campaign. IKEA wanted its customers to look at the company not only as a furniture retailer but also as a brand that guided customers at every step of the redesigning and furnishing process while creating an ideal haven and helping them go "happy to bed". IKEA developed a fully integrated plan for the "Big Sleepover" at the Lakeside IKEA store in the UK for its customers. The campaign was broadcast online, on social media (Facebook, Twitter, Flickr, and YouTube), and across print media and was considered a huge success.

Industry observers felt that by monitoring social media, IKEA had benefited through fact based decision making, sales recovery, waste reduction, and opportunity identification to avoid potential public relation crises. IKEA's stated business concept was "It is all about finding simple solutions and saving on every method, process, or approach adopted – but not on ideas."[2] IKEA stated that it had turned to social media to discover who was talking about it, what was being

talked about, and which product was being talked about, and to begin the process of social media listening. According to Hans Gartner (Gartner), Digital Development team, Inter IKEA systems, IKEA understood the value of listening to social media to understand customer experiences, opinions, and demographics. He also observed that the relevant social media data was not always communicated to other departments. So, IKEA teamed up with "The Socializers"[a] to speed up the social listening process by setting up the Listening Hub in 2014 at Inter IKEA Systems Headquarters in the Netherlands. According to Gartner, "*The Vizia/Brandwatch Listening Hub at Inter IKEA Systems allows key IKEA stakeholders to gain fuller insight into global conversations about the brand. The Brandwatch analytics platform provides senior stakeholders with an effective planning tool for multiple silos and regions.*"[3]

While some industry observers lauded IKEA's social monitoring initiatives, others felt that the company had to overcome many challenges to make these initiatives more effective. Gartner himself admitted that many challenges lay ahead for IKEA in relation to social media monitoring. These related to big data impact, data analysis, and IKEA's corporate structure. Data from many websites needed to be extracted and compiled as IKEA had separate pages for each of the countries in which it operated. Further, it had a product range of around 9500 products, which meant filtering the mentions and tagging of products was also a challenge.

About IKEA

Ingvar Kamprad established IKEA in the year 1943 when he was just 17. The name IKEA was formed by a combination of the first letters of his name (Ingvar Kamprad) with the first letters of the farm and village in which he grew up (Elmtaryd and Agunnaryd). IKEA was begun in a shed that was just two square meters in size. Kamprad bought pens, Christmas cards, matches, cigarette lighters, nylon stockings, and other items in bulk at a reasonable price and sold them to the residents of Smaland and made a profit. He used trucks that delivered milk to transport his packages to save money. As his business grew, however, making individual sales calls became cumbersome. So, in 1945, Kamprad started promoting his business through mail order catalogs. In 1947, IKEA introduced furniture in its portfolio. With this leading to a tremendous increase in sales volume, Kamprad discontinued all other

[a] The Socializers is an agency that had been working with organisations to provide them with social data acquired through online research.

products in 1951 to concentrate entirely on furniture sales. His furniture was reasonably priced, yet stylish. In the same year, IKEA launched its product catalog. According to IKEA, the catalog had been its inspiration since its launch. The catalog had been compiled with exciting stories behind its products and new ideas that made its customers love its products and inspired them to buy the products.

The first IKEA shop was opened in Almhult in 1953. By 1955, IKEA started designing its own furniture. It also introduced its famous flat-packaging[b]. In 1958, the first IKEA showroom was opened. In 1963, the company opened its first overseas store – in Norway. In 1965, on the outskirts of Stockholm, IKEA opened its first suburban store. It also introduced the self-service concept and distribution of catalog. In 1969, IKEA ventured into Denmark.

IKEA's vision of "creating a better everyday life for the many people" guided its operations globally. In the long run, IKEA believed, what was good for the people would also be good for it. IKEA also recognized that the impact its business had on society was also its responsibility.

Critics observed that IKEA's organization structure was complicated. The company was centrally structured through Inter IKEA Systems B.V., which worked alongside the IKEA group and IKEA franchises. In 1982, the Stichting INGKA Foundation, based in the Netherlands, became the owner of the newly formed "The IKEA Group" *(Refer to Exhibit I for IKEA's corporate structure).* IKEA, which catered to the retail furnishing needs of customers globally, came to be identified as one of the most recognizable brands in retailer home furnishing brands. It was observed that in each of the countries in which IKEA operated, it tailored its products to suit the requirements of the customers, thus building a rich heritage of customer experience. In 1984, the "IKEA Family", its customer club, was launched. In 1991, IKEA started "Swedwood"[c] to produce wood based furniture and wooden components. Over the years, IKEA came to be known not only for its marketing, but also for other distinctive features. It had a play area at each of its stores, which was supervised by its staff while the parents enjoyed their shopping experience. Each IKEA store included a restaurant that served traditional Swedish food such as potatoes with Swedish meatballs, cream sauce, and lingonberry jam. IKEA made variations to the menu in its restaurants based on their location.

[b] Flat pack furniture is a form of furniture that requires customer assembly.

[c] Swedwood is the industrial group within IKEA. Its primary function is to manufacture and distribute furniture.

IKEA was well known for its furniture that required self-assembly. It was also known for its "flat-packaging"[d]. According to the company, self-assembly and the flat-packaging concept reduced its costs. In all its markets, IKEA stated that its aim was to offer simple furniture designs at affordable prices. The Research & Developmental activities were centralized in Sweden. In 2014, IKEA reported a product range of 9500. According to the company, its product range was renewed every year, with about 2000 new products being launched each year.

IKEA's global sales in the financial year 2014 were reported at €28.7 billion and global revenues[e] at €29.3 billion. In the financial year 2013, global sales were reported at €27.9 billion and revenue at €28.5 billion. *(Refer to Exhibit II for IKEA's global revenues from 2003-2014).* IKEA's food turnover[f] in 2014 was reported at €1.46 billion. As of 2014, IKEA had a presence in 46 countries with 315 stores and 147,000 employees. Its product catalog was printed 217 million times and store visitors were reported at 716 million globally. IKEA stated that it wanted to make the company "twice as good and twice as big" by 2020. Its stated mission was "to build a better everyday life".

IKEA and Social Media

In 1997, IKEA launched its website, www.ikea.com. It reported that in 2011, its website had had 904 million visitors globally. In 2013, visitors to IKEA's website were reported at 1.3 billion and in 2014 at 1.5 billion globally.

IKEA's fans followed it online through different websites. Critics observed that IKEA's internet approach strategy was different across the different locations it operated in. It was present on social networking sites such as Facebook, Twitter, Google+, and Pinterest, among others. On Facebook and Twitter, IKEA launched separate Facebook pages and a Twitter tweet blog for each of the countries in which it operated. Also, the engagement model with its customers was country specific.

On Facebook, IKEA's followers observed that the company's social team posed questions and shared links for photo albums and YouTube

[d] Furniture that was fabricated in flat parts and designed to be quickly and easily assembled was called Flat Furniture. It is also referred to as ready-to-assemble furniture or knock-down furniture.

[e] Global revenue includes revenue from sale of articles and the rental income from IKEA's shipping center operations.

[f] IKEA Food comprised the IKEA restaurant, IKEA Bistro, IKEA Swedish Food market, and the IKEA co-worker restaurant.

clips to increase customer engagement. It was observed that IKEA's Facebook had at least 160,000 fans on each of its country pages with the US page having the highest number at 3.4 million. IKEA involved its Twitter fans in celebrations like "Bring your Own Friends Day" and other campaigns. Its Twitter US page had 220,000 followers and the Canada one had 660,000 followers.

Pinterest, also described as a "Visual Discovery Tool", was a favorite choice of any home furnishing retailer. Critics observed that IKEA, however, was not very active on "Pinterest". Only in the US, the UK, Canada, Germany, and a few other countries did it use Pinterest to visually present its merchandise to the customers. The first ever Google+ page of IKEA was launched in 2011. It was observed that the Google+ page of the US launched in January 2014 had the lowest following compared to IKEA's other country pages.

Embracing the developments in technology, IKEA launched a digital catalog application in 2013. In 2013, the application recorded a download of 10 million. In 2014, IKEA reported that the catalog application was opened 46 million times.

IKEA observed that its customers were tech savvy and were constantly increasing the benchmark for standards of quality and service. It also observed that its customers were well connected on social media on different websites and publicly expressed their opinions. IKEA realized that these mentions and conversations about it across social media needed to be monitored to understand and derive inputs about what was being said about it, who was talking about it, and from which geographic area the customers were talking. For instance, in 2013, IKEA changed its strategy in China after listening to conversations about the brand in the country. IKEA's furniture prices were considered low in Europe and the US, but were higher than the average furniture prices in China. Gathering information from social media, IKEA reduced its prices. It further observed that its product catalog provided opportunities to its competitors to imitate its product and pricing strategy in the country. So, it removed the product prices from the catalog. It also started using Weibo, a leading Chinese micro-blogging website, to target young Chinese customers, as this platform was more popular in the country.

Social Media Monitoring & the Socializers

By mid-2014, IKEA felt the need for an effective customer relationship management program to establish and promote client communication, connect the different business units and departments

to develop and implement its marketing plan, and also to develop a strategic partnership. IKEA stated that through "The Socializers", it aimed to derive real time-insights from social media on who was speaking about it, where it was being spoken about, and which product was being mentioned. The Socializers was founded in 2009 by Marvin Towler, Nathaniel Hansen, and Jeff Snyder. According to The Socializers, people and technology were their focus.[4] They identified key influencers and thought leaders by analyzing social data after conducting extensive online research. According to The Socializers, they worked with brands that lacked a strong social media presence. The Socializers stated that they acquainted their clients with the best tools and provided consulting and training services by using a robust social media strategy. Introduction to new customers and the creation of an audience and a fan following increased a company's presence on social media, thereby increasing its social media footprint and revenue generation, the Socializers stated. The Socializers educated and trained the staff of the client's organization with tools to establish an online identity and grow.

According to Gartner, IKEA teamed up with "The Socializers" and formulated a strategy to help the company emerge as an intelligent operation. IKEA decided to use Brandwatch's Vizia[g], to monitor social media. Vizia, a "self-serve application" or software as a service, archived social media data and conversations happening online to provide brands with information and the means to track specific segments to analyze the brand's presence. Experts observed that Vizia at IKEA was customized by Brandwatch and was a design driven interface, making it user friendly. According to Brandwatch, Vizia allowed IKEA to create data visualizations through drag and drop from multiple screens, and a dashboard to suit the company's requirement for social data interpretation.

IKEA's Listening Hub was a multi-screen social dashboard which displayed many minority reports. The screens illuminated "mentions", across the globe. On analyzing these mentions, Vizia gave insights into who was talking and what was being discussed about IKEA *(Refer to Exhibit III for an image of a typical social command centre (Listening HUB) at IKEA).*

Experts commented that even without notifications, stakeholders looking at the social command center derived an understanding of the customers' thoughts on social media. After the Listening Hub was set up, it was observed that these mentions were sent to the relevant

[g] Brandwatch is a social technology company, was found in 2005.

departments from the Hub, without modifications being made. As Gartner explained, the Listening Hub's function was to detect, in real-time, customer insights from across the globe and derive inputs that needed to be shared and distributed with different departments, markets, and countries.[5] Organization wide, "The Listening Hub" helped encourage the change. It was observed that social media information was being shared and redirected to appropriate departments for appropriate action.

IKEA Monitoring Customers

In November 2014, Forbes[h] named IKEA as the "40th most valuable brand in the world" in its annual list. Marketers considered IKEA as the most recognizable brand on the planet. According to Forbes, IKEA's brand value as of November 2014 was estimated at US$12.5 billion.[6] In 2013, according to Millward Brown Optimor[i] (including data from BrandZ, Kantar Retail, and Bloomberg) IKEA's brand value was estimated at US$12.04 billion, and in the year 2012, at US$ 9.2 billion.[7]

Marketing experts observed that IKEA was "A retail giant with a personal touch" with an increasing social media footprint. It was observed that IKEA's US page alone as of September 16, 2014, had 4.2 million Facebook likes, 291,026 Twitter followers, 81,541 Instagram followers, and 157,124 Pinterest followers.[8] According to Leontyne Green Sykes, chief marketing officer of Ikea North America, the challenge to IKEA's marketing team was listening to customers and providing solutions. In her words, "I challenge myself and my team to be anthropological marketers, listening to customers and providing solutions to their challenges."[9]

According to industry observers, though, the "Listening Hub "was formally set up only in 2014, IKEA was deriving insights from social media about its brands and on what the customers were talking about them much earlier. For instance, in addition to the success of the 'Big Sleepover' campaign, IKEA also tasted success with its Malmo store launch on Facebook. Then there was the crisis situation that IKEA averted in 2013 related to the Great Horsemeat Scandal.

---

[h] Forbes, an American business magazine owned by Forbes Incorporation, is published biweekly. It is known for its lists and rankings including its lists of the richest Americans (the Forbes 400) and rankings of world's top companies (the Forbes Global 2000).

[i] Millward Brown is a global research agency specializing in advertising, marketing communications, media, and brand equity research.

IKEA Malmo Store Launch on Facebook

In 2009, IKEA launched its first Social CRM project in Malmo, the commercial capital of southern Sweden. In that year, IKEA decided to launch its store in Malmo through Facebook. This was considered one of the most innovative launches on Facebook by industry observers. IKEA worked with Forsman & Bodenfors (F&B)[j] for the launch. A Facebook page was created for the IKEA Malmo store manager, Gordon Gustavsson. Over a two-week period, his pictures taken in the stores were uploaded on his Facebook page and his friends were tagged. According to IKEA, anyone who tagged their name to the products displayed in the picture won the product. The campaign went viral. Soon, IKEA observed that there were requests from the fans to upload new pictures.

Marketers found the statistics of this campaign astounding. The promotion crossed geographic barriers, just as IKEA had intended, and reached other Facebook members, Tweeters, and bloggers around the world. By 2010, the video "IKEA – Facebook showroom" had been viewed 283,859 times and its Facebook page was liked 307,664 times. The campaign was awarded "Most Contagious 2009" by the international Contagious Magazine. In year 2010, the www.ikeafans.com website was launched by IKEA fans, and had 1,724 likes for the year. The campaign won a Titanium Lion at the Cannes International Advertising Festival in 2010. In May 16, 2011, Cannes awarded IKEA the 'advertiser of the year' award.

Marketing experts commented that IKEA had demonstrated the art of making money through Facebook's existing features by spreading the word for a store opening in Malmo, Sweden. Facebook changed its terms and conditions after the campaign, to accommodate further launches. IKEA stated that it had achieved its goals on brand imaging, targeting customers beyond a specific geographical location, and store promotion with the store's Facebook launch in Malmo.

Big Sleep Over

In December 2011, IKEA launched "The Big Sleepover" happy to bed campaign in the UK. According to IKEA, its team uncovered a group on Facebook called "I want to have a Sleepover in IKEA", which had been started by IKEA fans. Identification of the fan group was the

[j] A Swedish advertising agency founded in 1986. It had won Big Won Rankings for 2012. It was ranked 10th among the top agencies in the world, and second among Top Digital agencies.

seed for the campaign. According to the company, it had identified a need to change the perception of customers on the IKEA brand as customers were viewing it only as a retail furniture shop. IKEA wanted its customers to see it not only as a furniture retailer but also as a brand that guided customers at every step of the redesigning and furnishing process in creating an ideal haven and then going "happy to bed". IKEA observed that people in the UK were too embarrassed to test out a mattress in-store properly. With these inputs, IKEA developed a fully integrated plan, for the "Big Sleepover" at its stores for its customers. It invited people to sign up for the group "I wanna have a sleepover at IKEA". One hundred entrants (aged 25 years or more with disposable incomes) were randomly picked by IKEA from 100,000 entrants for the campaign.

On the campaign day, according to IKEA spokesperson, Lewis Marshall, ,at the Lakeside IKEA store in the UK, the participants not only got to choose their beds, bedspreads, and linen, but also to get answers to their questions from sleep experts and mattress professionals, for whom IKEA had arranged. The participants were treated to relaxing massages, manicures, warm drinks, and even bedtime stories. The campaign was broadcast online, on social media (Facebook, Twitter, Flickr, and YouTube), and across print media.

Industry experts observed that the IKEA Big Sleepover campaign was a sign that the company was changing its marketing strategy to accommodate the changes in consumer behaviour. IKEA's Big Sleepover won five international awards. IKEA was also seen to be duplicating its Sleepover campaign at other locations across the globe.[10] Experts opined that the Sleepover campaign had shown that IKEA was successful not only in its marketing strategy but also in positioning the brand among the younger generations.

IKEA and the Great Horsemeat Scandal

Another instance that demonstrated that IKEA was monitoring the social media even prior to 2014 was during the Great Horsemeat Scandal. The Great Horsemeat Scandal went public on January 15, 2013, with the Food Safety Authority (FSA) of Ireland announcing that pig and horse DNA had been found in products specified to have been made from beef. Subsequent to the announcement, consumers across the world observed that retailers including Tesco, Lidl, Aldi, Iceland, and Dunnes had recalled 10 million burgers from their store shelves. FSA authorities stated after rigorous monitoring and testing, that horsemeat had been found in beef and beef products in at least 14 European Union (EU) countries.

IKEA launched meatballs in its stores in 1987. It was observed that IKEA's meatballs were tremendously popular, and critics commented that this popularity would have continued if not for the horsemeat scandal. IKEA'S Swedish meatballs, supposed to contain only beef and pork, were found to contain horsemeat in tests conducted by the Czech Republic. It was observed that 81,107 comments were posted regarding the horsemeat scandal incident on social media. Further analysis of these comments (sentiment analysis) was done to ascertain whether the comments were negative, neutral or positive by a global social media monitoring expert. On February 25, 2013, the Czech Republic's Veterinary administration announced IKEA's frozen meatballs had traces of horse DNA. The same day, 56,362 comments were posted online about IKEA of which 5,153 were found to be negative. IKEA immediately recalled a batch of frozen meatballs that had been shipped to the Czech Republic, Slovakia, Hungary, France, the UK, Portugal, Italy, the Netherlands, Ireland, Cyprus, Greece, Spain, and Belgium. Further, the product was removed from its Sweden stores.

The company observed that IKEA had got another 4,000 negative comments from the UK on social websites over the next week though the meatballs had been recalled.[11] In addition to IKEA, a few other brands were involved in the scandal such as Burger King, Nestle, Bird's Eye, and Findus. However, IKEA denied the findings and said, "*Our own checks have shown no traces of horse meat. Now we must of course look into this further*,"[12] in a statement posted on its website.

Facebook, Twitter, and several blogs were monitored as part of social media monitoring for the Horsemeat Scandal. English and non-English language websites like Weibo were also monitored globally. Despite the scandal, critics observed that there were not too many negative comments about IKEA. *(Refer to Exhibit IV for comparison of food sales to global sales).* Critics observed that social monitoring had helped IKEA to handle the crisis situation promptly and suitably.

The Issue with Ikeahackers.net

In 2014, IKEA confronted a situation that amused the Public Relations professionals globally. On the IKEAhackers.net website, IKEA users posted ideas for revamping and remodeling furniture and merchandise that were already available at IKEA stores. The website invited ideas for new products. IKEAhackers.net was founded by Jules Yap (Yap) (actual name Mi Mi Yap) in 2006. According to Jules Yap, she was searching for IKEA hacks and had observed that the internet had many amazing ideas on products at IKEA stores. In an effort to collate and compile these ideas, she floated a webpage, IKEAhackers.net,

which hosted product innovation ideas for IKEA's products. According to Yap, the only source of revenue for the website was the advertisements.

IKEA B.V. Systems promptly sent her a Cease and Desist order, stating that IKEAhackers.net had infringed upon IKEA's intellectual property rights. "*I am afraid, I have a bit of bad news,*"[13] wrote Jules Yap -in her blog on the website after receiving the order. "*I don't have an issue with them protecting their trademark but I think they could have handled it better,*"[14] she said and sought moral support from her fans on IKEAhacker.net. The news went viral and likes were shared on websites like Gizmodo, Slashdot, and Lifehacker. Some fans even called into IKEA to complain, some wrote emails, or chatted with IKEA representatives online.

IKEA's social team followed the incident closely. The company observed that the original blog post was shared more than 5,000 times and retweeted 1,500 times on Twitter within 10 days of its being posted.[15] According to critics, there were more than 10,000 mentions that were not considered on different websites. Following the uproar the blog caused in social media, IKEA stated that it had revisited its decision. *The company clarified to CNBC on email.* "….It has of course never been our ambition to stop their webpage. On the contrary, we very much appreciate the interest in our products and the fact that there are people around the world that love our products as much as we do."[16] *Initially, IKEA insisted that Yap transfer the domain name to IKEA. It stated that Yap had* committed trademark infringement by using its products on a site for commercial purposes. Yap later said that she had worked out a deal with IKEA that allowed her to keep the site only if all advertising on the website was dropped by June 23, 2014.

According to IKEA, its concern was that its brand name was being used for economic gain on the website, though the website hosted its branded products. IKEA stated that it felt customers could get confused. Mack Collier (Collier), renowned social media strategist, mentioned in this blog that IKEA, instead of running into negative publicity with its IKEA fan club for having issued the C & D letter, could have sorted the matter out with Jules Yap more discretely. By issuing a C & D letter, IKEA had created a threat for itself According to Collier, "*If the brand was really worried about advertisements on the site, then IKEA might have considered making a deal with the fan running it to have her remove all ads, and in exchange Ikea could have sponsored the site for the amount she would have earned in ad revenue. That turns a negative PR event into an incredibly positive one for Ikea. It generates new fans for the brand, and everyone wins.*"[17]

Challenges before IKEA

Industry observers felt that by monitoring social media, companies such as IKEA benefited through fact-based decision making, sales recovery, waste reduction, and opportunity identification to avoid a potential public relation crisis. *(Refer to Exhibit V for benefits of social media monitoring).* With 315 operational stores across the world, IKEA's digital marketing strategy remained uniform – extensive use of social media with separate country pages for respective markets and an ever increasing stress on digital catalog applications which could be downloaded on app stores such as Google Play and the Apple store. According to marketing experts, IKEA's customers & fans were characterized as highly educated customers, who constantly dictated high standards of quality and service for products and were well-connected owing to easy access to media to publicly express their opinions. IKEA stated that it looked forward to the social space for a future filled with innovation. However, social media listening and monitoring also posed many challenges to the company. According to Gartner, the challenges ahead of IKEA in social media monitoring were many.

Researchers opined that prior to the advent of social media monitoring tools, organizations had very scanty information about their customers and product users. Surveys or interviews were the only sources of information for organizations. An information deluge was created through content that was written, exchanged in blogs, status updates, forum discussions, video, audio, and images. Organizations could increase productivity, reduce costs, and arrive at better decisions with proper utilization of information, according to a McKinsey[k] report. According to McKinsey & Co, Cluster Analysis, Semantic Analysis, and Sentiment Analysis were possible due to availability of large data. Apart from these advantages, the Big Data impact also posed a competitive threat to businesses that ignored any particular trend on social media. According to IKEA, data analysis was a real challenge, as its products were tailored to local markets. For each country in which it operated, IKEA had a separate Facebook page, Twitter tweet blog, a separate Pinterest page, and a different page on Google +. Not only was the volume of data huge, but it was also scattered. To collate this kind of data from across countries and products was a real challenge for IKEA.

According to industry observers, IKEA was present in many countries across the globe and each of its stores functioned locally.

[k] McKinsey & Company are consultants and are into publishing books, research and articles about business and management.

IKEA realized that many successful and innovative practices that were being undertaken in some parts of the business were not reaching other departments which would also have benefited from such information. IKEA had expanded its presence to many countries globally. IKEA was known for its furniture that required assembling. It was observed to follow Vertical Integration in its organization structure; 90% of its components were outsourced, and the remaining 10% were produced internally (Swedwood). According to IKEA, its Design unit, a cost centered unit, did not take up any design if it was not economical. Since most of the components were outsourced, IKEA enjoyed the benefits of economies of scale with respect to product manufacture, labor, supply and distribution. Any change or innovation in the product needed to be analyzed by all the departments, but idea exchange and development was not simple within IKEA. IKEA's organizational structure was an impediment.

Moreover, by 2014, IKEA had about 9,500 products in its catalog. Some of the products were customized for the local markets. For the Social Monitoring Hub, it was a challenge to analyze what was being discussed about one particular product in a particular market, and to apply analytics to arrive at a logical conclusion as to whether that discussion was spam or not. Each of IKEA's local stores had a website accessed by geographic domicile. Filtering of spam sites would be an added challenge for IKEA's Social Monitoring Hub. According to social media tool analysts, spam was rampant in all of the social monitoring tools. Companies had changed their variables for search in order to avoid spam messages, but had ended up missing a major chunk of the conversations pertinent to their own brands.

Looking Ahead

Industry experts observed that the sequence of events at IKEA from 2012 to 2013 had dented the customer's perception of IKEA. In 2012, IKEA released its Saudi Arabia catalog after removing the women in it via photo retouching. IKEA in its other countries' catalogs had women. The altered images caused a great uproar when revealed in the Swedish media, prompting an apology from IKEA's head office in Sweden. IKEA apologized for removing all women from its Saudi Arabia catalog. "*We should have reacted and realized that excluding women from the Saudi Arabian version of the catalog is in conflict with the IKEA Group values*,"[18] the company said in a press statement.

In November 2012, Ernst & Young revealed that IKEA was associated with suppliers in Germany, who had employed forced labor

before 30 years. The company released an independent report after the E&Y revelation. The statement said East German prisoners, mostly political dissidents, had been involved in the manufacture of goods that were supplied to IKEA 25 to 30 years ago. Jeanette Skjelmose, an IKEA manager, deeply regretted the incident as IKEA did not accept the usage of political prisoners for manufacturing. She further stated that "*at the time we didn't have the well-developed control system that we have today and we clearly did too little to prevent such production methods.*"[19]

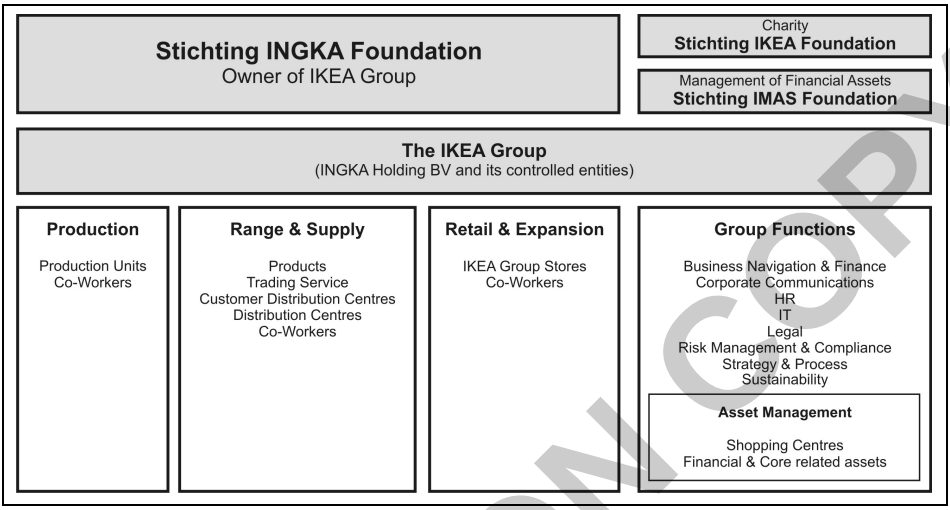
It was observed that the Horsemeat Scandal and subsequent recall of meat balls by IKEA had tarnished the reputation of the company even more in the eyes of customers during February 2013. YouGov's index[1] showed that IKEAs Brand Index had been fluctuating due to the sequence of events from 2012 to 2013. *(Refer to Exhibit VI for the YouGov's Brand Index).*

Industry observers felt that IKEA had already learnt crisis management & response, campaign management, and lead generation through social media listening. They observed that IKEA setting up "The Listening Hub" was a step in the right direction. But how far IKEA could leverage its social media monitoring and social media listening for improving its business prospects depended on whether it was able to measure up to the challenges posed by the new media.

[1] YouGov is an international internet-based market research firm, YouGov's methodology is to obtain responses from an invited group of internet users, and then to weight these responses in line with demographic information.

Exhibit I

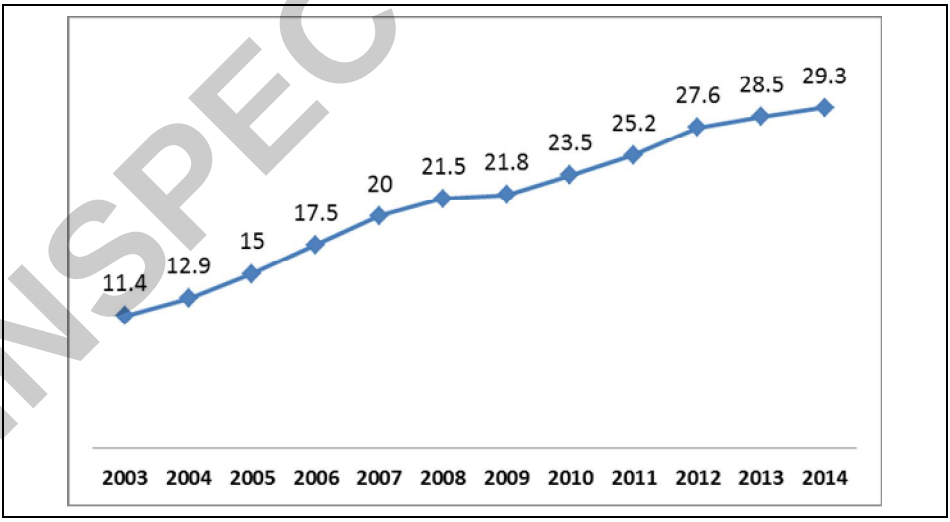
IKEA's Corporate Structure



Source: www.ikea.com/ms/en_CA/pdf/sustainability_report/sustainability_report_2014.pdf

Exhibit II

Global revenues of IKEA (Year 2003-2014, in € billions)



Source: www.ikea.com/ms/en_CA/pdf/yearly_summary/ikea-group-yearly-summaryfy13.pdf

Exhibit III

Image of a Typical IKEA Listening Hub

Source: www.distilled.net/resources/how-to-master-social-listening-lessons-from-ikea/

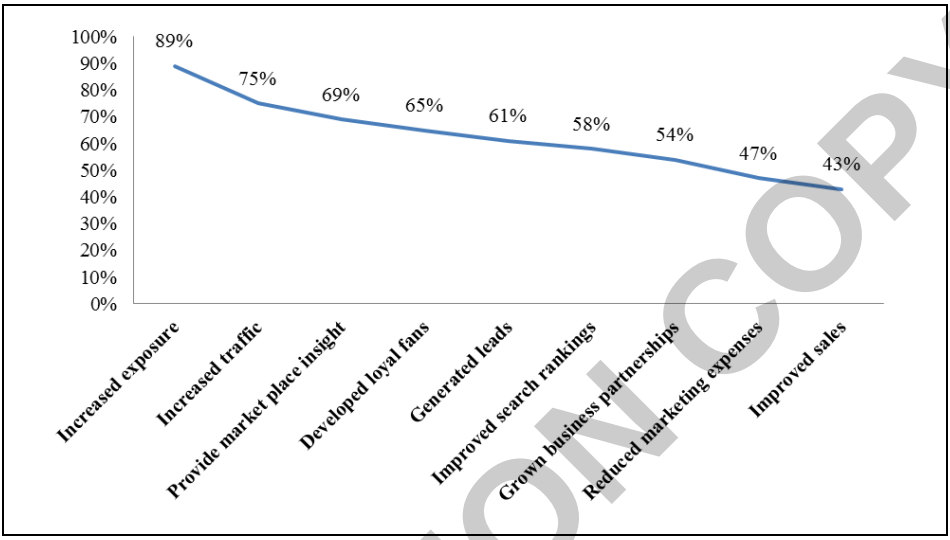
Exhibit IV

Comparison of IKEA's Global Sales and Food Sales

Source: www.ikea.com/ms/en_CA/pdf/yearly_summary/ikea-group-yearly-summaryfy13.pdf

Exhibit V

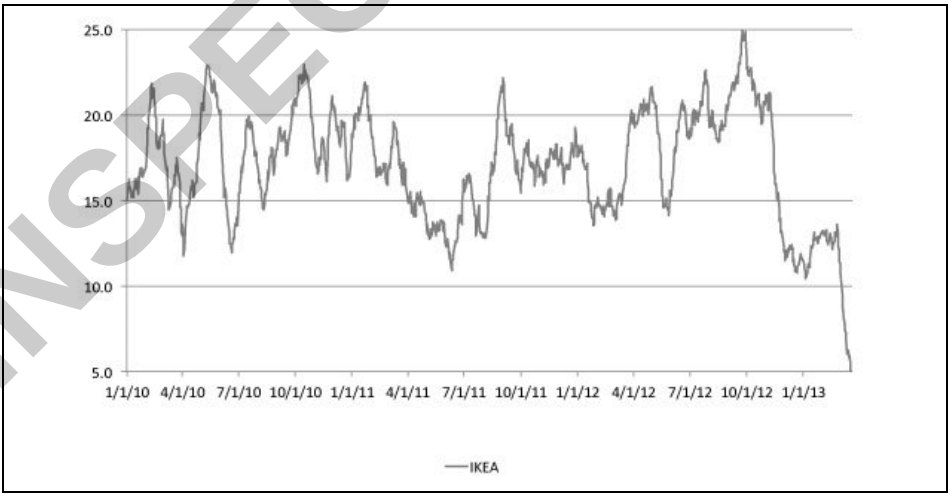
Benefits of Social Media Marketing according to SMB marketers worldwide January 2013



Source: "Social Media is a Brand Builder," www.marketingxlerator.com, May 28, 2013.

Exhibit VI

IKEA's Reputation (according to YouGov's Brand index during September 2012- February 2013)



Source: www.businessinsider.in/IKEAs-Reputation-Has-Taken-A-Serious-Beating/articleshow/21232035.cms

End Notes:

[1] Tabatha Laverty, "In Bed with Brand – How IKEA Turned a Facebook Group Into Customer Engagement Success," http://smbp.uwaterloo.ca, February 5, 2015.

[2] Hannah Tregear, "KEA & The Socializers: Building Social into the Heart of a Global Business," www.brandwatch.com, December 9, 2014.

[3] Ibid.

[4] "About The Socializers," http://thesocializers.com, February 19, 2015.

[5] "IKEA & The Socializers: Building Social into the Heart of a Global Business," www.brandwatch.com, December 9, 2014.

[6] "The World's Most Valuable Brands," www.forbes.com/powerful-brands/list/.

[7] "Global Brands," www.ft.com, May 20, 2013.

[8] "Brand of the Day: How Ikea Brings Design Ideas to Life in Social," www.adweek.com, September 16, 2014.

[9] *Ibid.*

[10] "In Bed with Brand – How IKEA Turned a Facebook Group Into Customer Engagement Success," http://smbp.uwaterloo.ca, February 5, 2015.

[11] "Horsemeat Scandal Hits Tesco And IKEA Across Social Media," www.thegrocer.co.uk, March 7, 2013.

[12] "Horse Meat Found In Ikea's Meatballs," www.npr.org, February 25, 2013.

[13] Jules IKEAHacker, "Big Changes Coming To IKEAHackers," http://news.jchk.net, June 14, 2014.

[14] "Ikea Makes Fan Site Remove Adverts in Rights Row," www.bbc.com, June 16, 2014.

[15] "Suing Your Brand's Fan Site and other Stupid Things to Do on Social Media," http://enterpriseinnovation.net, June 25, 2014.

[16] Mike Willee, "Blogging + IKEA = Intellectual Property Infringement," http://blog.traklight.com, June 27, 2014.

[17] Diana Bradley, "Ikea's Blogger Relations Criticized For 'Crushing' Fan Site Owner," www.prweek.com, June 19, 2014.

[18] "IKEA Regrets Women Erased from Saudi Catalog," http://abcnews.go.com, October 1, 2012.

[19] "Ikea Apologizes for Using East German Prison Labor Camps," www.businessinsider.com, November 16, 2012.

Case 2

Algorithms to Style People: Stitch Fix Applies Data to Fashion

Debapratim Purkayastha and Tangirala Vijay Kumar

> *"One common misconception about fashion is that you can't apply data to it. Everyone thinks that their fashion is so unique – that it's a very fuzzy, subjective thing. In reality, the fundamental things are quantitative."*[1]
>
> ***– Katrina Lake, Stitch Fix's founder and CEO, in March 2018***
>
> *"Data is really Stitch Fix's key differentiator…Stitch Fix knows who's buying what, who's trying and not buying what, what they like and what they don't like."*[2]
>
> ***– Eric Kim, Managing Partner, Goodwater Capital[a], in November 2017***

Stitch Fix, Inc. (Stitch Fix), an online personal styling service, employed algorithms and human stylists to deliver curated apparel to customers. It also employed algorithms to allocate orders to warehouses and to determine the amount and type of inventory to order. The company's algorithms learnt about customer preferences from each shipment and improved its selections over time. The styling service initially offered its service to women and then expanded into the men's segment as well. The company also employed machine learning[b] algorithms to come up with its in-house brands. Over time Stitch Fix started offering its styling service for the plus-size segment for women

[a] Goodwater Capital is a venture capital firm.

[b] Machine learning is a mechanism of data analysis that automates analytical model building. A branch of artificial intelligence it is based on the thought process that systems can learn from data, recognize patterns, and make decisions with minimal human intervention. (Source: "Machine Learning: What It Is and Why It Matters," www.sas.com, https://www.sas.com/en_in/insights/analytics/machine-learning.html)

and included premium wear branded garments in its shipments. Katrina Lake (Lake), Stitch Fix's founder and CEO, saw the valuation of the company increase from US$4.75 million in February 2013 to over US$2 billion as of March 2018.[3,4] However, she was concerned that there was only so much shopping that consumers seemed to be willing to do through Stitch Fix. Also, the company's revenue growth rates were under stress and Stitch Fix had to keep expanding its customer base. Besides, increasing advertising expenditure was eating into the company's profits. Stitch Fix also faced the heat of increased competition from the likes of Amazon[c]. Eric Colson (Colson), Stitch Fix's Chief Algorithms Officer, who was previously the Vice President of Data Science and Engineering at Netflix[d], also faced the challenge of garnering more data about customers and of finding ways to utilize the data for business decisions.

The Origins

After completing her BA in Economics (with Honors) from Stanford University in 2005, Lake took up a consulting job where she assessed data for the restaurant and retail industry.[5] In the course of her work, Lake saw that though several consumers wanted to purchase apparel online, it was tough for them to filter the thousands of choices available and narrow down to the ones that suited their requirements. She recognized that the challenge was, *"How can we marry the ease of shopping online with what people want in clothes, which is really about fit and style?"*[6]

When Lake was not able to find any company that tackled the issue the way she liked, she joined a venture capital firm hoping to come across and join the apparel retailer of the future. A couple of years down the line, she was still no nearer finding such a company and so she decided to start her own venture. In 2009, she joined Harvard Business School with the intention of having a fully worked-out business idea at the end of two years.

Lake began testing the Stitch Fix concept in her second year at Harvard Business School.[7] She requested friends of friends in the Boston region to fill out style surveys. She would then personally go and

[c] Amazon is an e-commerce and cloud computing company. For the year ended December 31, 2017, Amazon had net sales of US$177.9 billion and net income of US$3 billion.

[d] Netflix is a streaming service that enables its customers to view a variety of television shows, movies, and documentaries. It employs big data to suggest television shows and movies to its users and also to create original content.

purchase apparel using her credit card and deliver them in a box to people's homes. Lake's customers gave her a personal check for the clothes that they retained and returned the remaining ones. Lake kept a record of individuals' preferences in an Excel spreadsheet. In the same year (2011), Lake launched Stitch Fix, investing US$750,000 that she had secured from the venture capital company, Baseline Ventures.[8]

Algorithms and the Human Touch

Stitch Fix's initial target customers were women who either didn't enjoy shopping or were too busy to go to a physical store or to narrow down their choices from several online pages of apparel.[9]

The following was how Stitch Fix generally functioned: Customers had to sign up on its website and complete a detailed questionnaire regarding their size, whether they were mothers or currently pregnant and also their due date, how they wanted their clothing to fit, what their style was akin to, what were the parts of the body that they wanted to flaunt or play down, what colors they liked and hated, how frequently they dressed for specific occasions (such as work, events, and dates), and what their preferred price points were.[10] This style profile collected up to 85 data points.[11] Customers were also encouraged to share links to their Pinterest[e] boards comprising pictures of their favorite styles or fashions.[12] This helped the customers communicate aspects such as personal style that were difficult to express in words.[13]

Employing the entire data, an algorithm then mined Stitch Fix's inventory of around 700 brands to select pieces that would suit the customer's profile *(Refer to Exhibit I for a look into Stitch Fix's recommendation system and to Exhibit II for an image of the visualization of Stitch Fix's algorithms.)*.[14] Lake explained how there was a prevailing misconception that data could not be applied to fashion, while actually the opposite was true.

The selection was then shown to a human stylist, who was usually matched to the customer by employing an algorithm (depending on the stylist's expertise and the customer's needs and likings).[15] The stylist reviewed the computer's selections to pick the final assortment of five pieces of garments, shoes, and accessories for the customer.[16] This was based on the holistic context, which took into account trends and

[e] Pinterest is a social network wherein users can visually share and explore new interests by posting (called 'pinning' on Pinterest) images or videos to their own or others' boards (a group of 'pins' generally having a common theme) and browsing the content pinned by other users.

the comments or requests made by the customer (such as "I require more clothes for work", "I need to attend a wedding in August", and so on).[17] The human stylists also included in the shipment a personal note that provided styling recommendations such as how to accessorize or wear the items.[18] *"...there are things humans can do much better, and are likely to remain better at for a long time – things like curation, the ability to see things as a cohesive set, and to improvise. Not to mention being able to relate to other human beings,"*[19] noted Colson.

Stitch Fix then shipped the items to the customers in a box – the company referred to each shipment as a 'Fix' (*Refer to Exhibit III for an image of a Stitch Fix box*). Customers tried out the pieces at home and returned the items that they did not like.[20] "The concept has always been personalization...but trying to figure out which one is best for you is really the challenge,"[21] noted Lake. There were no shipping charges both ways. A prepaid return mailer accompanied every order, which made it easy for customers to return the pieces that they did not want.[22] However, the customers had to pay a US$20 styling fee for every shipment. The US$20 was then applied to any item or items that they purchased.[23] This implied that if a customer wanted to purchase a blouse costing US$50, she had to only pay US$30 in addition to the US$20 fee.[24] Customers got a 25% discount if they purchased all the five pieces.[25] However, the company did not make it mandatory for customers to purchase at least one item; they could return the entire lot if they wanted, and had to merely pay the US$20 fee. Customers could order a box whenever they wanted or by a set schedule.[26]

As customers continued to get shipments, each piece they decided to retain or return offered a new data set.[27] Even returns were considered to be important, with the company's stylists taking in the negative feedback in a customer's remarks to better gauge the customer's style or to adjust a piece's description.[28] Stitch Fix asked what the customer liked or disliked about every piece and employed natural language processing[f] to decode the written answers.[29] Its algorithms processed customer remarks and feedback and could easily and speedily compute how customers felt about specific products and styles. *"Humans are better at interpreting the tone and meaning of textual feedback, but they can't do many. After a few thousand, they might get really bored or have a hard time distilling it down. Machines can hold in their memories far more than we can,"*[30] remarked Colson.

[f] Natural language processing is a branch of artificial intelligence that aids computers to understand, interpret, and manipulate human language. (Source: "Natural Language Processing: What It Is and Why It Matters," https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html)

(Refer to Exhibit IV for an instance of algorithms compensating for a shortcoming of human stylists)

According to Lake, since Stitch Fix used stylists and not merely an algorithm, it obtained superior-quality data and more involved and accurate data points. For instance, a customer might purchase a pair of jeans from the company and then inform its stylist that she had only purchased the jeans as she was in the course of shedding baby weight and wanted the jeans to fit. Right then, however, they did not actually fit – so the stylist could then tell the algorithm to ignore the jeans' data and not assign it to the customer's profile.[31]

Customers volunteered a lot of information to ensure that every delivery was worthwhile.[32] *"In our model...we know exactly why you bought the things you bought and why you didn't buy the things you didn't buy,"*[33] said Lake. The data about what items a customer kept and what she returned was incorporated into her profile and improved the capability of an algorithm and the stylist to predict the pieces she would like, going forward.[34] Explained a lead stylist at Stitch Fix, *"Once we build a relationship with a client, we get smarter every time [we style her] and the algorithm gets smarter every time. When a client decides which items to keep or send back, she can go through her profile and let us know item by item if she liked the fit, the price, the quality. That goes into the algorithm and helps it suggest more for the next time."*[35] A customer who ordered a "Fix" each month said, *"I've seen the things that come in my box start to adapt to more what my personal style is."*[36] However, according to the company, its customers were prone to spending more in the initial six months of availing of the service than they did in the subsequent six months and to spending more in the first year than in the second year.[37]

Predicting Demand and Fashion Trends

Stitch Fix used hundreds of algorithms such as an algorithm that allocated a shipment request to a warehouse (on the basis of available inventory and distance to customer), an algorithm that computed the amount of and the type of inventory the company should purchase, and an algorithm that computed a customer's level of satisfaction with the service *(Refer to Exhibit V for details of how algorithms were used for some of the tasks).*[38,39,40]

The company controlled inventory costs by employing algorithms. A conventional retailer had to chiefly guess the number of pants it would sell at the beginning of the season and place an order. Stitch Fix, however, could forecast based on the number of customers buying

pants and tell vendors to manufacture specific styles, colors, or patterns when placing an order.[41] Lake added, *"In traditional retail, stores know that some things sell and some things don't, but they don't know why. In our model we actually know why. If we have a sweater that doesn't work, for example, we can isolate whether [the problem] was the color, fit, or fabric."*[42]

Stitch Fix's data also helped it to identify fashion trends. In December 2016, Colson said that "cold shoulder" tops – shirts with a portion of the shoulder or sleeve cut out – were particularly hot at that point of time.[43]

Launch of In-house Brands

Based on the customer response data and the data about their style and fit preferences, Stitch Fix's data scientists saw that the algorithms could analyze the data and decide, by a set of traits, what customers were looking for in a skirt or a pair of trousers. However, the issue was that sometimes there was no suitable item in the company's inventory. The algorithms spotted a gap in Stitch Fix's merchandise.[44] In June 2016, Stitch Fix launched an in-house brand of garments under the Hybrid Design initiative. In the next 12 months, the company employed machine learning to create more than 30 items under this initiative. According to Colson, *"We noticed gaps in the market and an opportunity to produce something that doesn't exist, but should."*[45]

The following was how new product development under the Hybrid Design initiative functioned: A group of three algorithms produced a starting point. The first algorithm selected three "parents" – recommended items of clothing that could be either combined or employed as a template for a new item. The second recommended three distinct attributes that had been shown to enhance the parents' style – perhaps a different neckline or sleeves. And the third threw in some amount of randomness, a recommendation that was not typical for the earlier recommended style but could be a fascinating design. Combined, these algorithms searched through a set of 30 trillion possible combinations of, for instance, blouse attributes, to finally come up with just nine recommendations.[46]

Human designers then organized the algorithms' recommendations into an apparel piece that a consumer would like to wear. In other words, software might be capable of creating a mathematically perfect apparel item, but it still was not capable of assessing the cultural context that made the specific item into a fashion success.[47] Stitch Fix's in-house brands accounted for over 20% of the company's revenue in fiscal 2017.[48]

Expansion into New Categories

In September 2016, Stitch Fix launched the styling service for men as well.[49] In February 2017, it launched Stitch Fix Plus that offered plus-size garments to women. At the time of launch, the company had a waitlist of over 75,000 women who wanted to avail of the service.[50] In the plus-size segment for women where there were very few shopping options – whether online or brick and mortar – Stitch Fix Plus debuted with over 90 brands. With Stitch Fix educating its vendor partners who had previously not manufactured plus-size apparel on fit, there were nearly 15 brands that launched plus-size offerings for the first time through the collaboration. *"They have plus-size offerings that some of these designers don't normally make, but are making for their customers so they can make sure they have a huge variety of options,"*[51] explained one of the plus-size bloggers who had advised the company on its plus-size strategy. Stitch Fix also introduced plus-size options for a few of its in-house brands.

In August 2017, Stitch Fix added 100 premium brands to its portfolio of partner brands. The majority of pieces that Stitch Fix sent to its customers were priced below US$100. However, each of the items belonging to the premium brands were priced between US$100 and US$600. One of the reasons for the company adding premium brands was that customers were asking for them. When filling out their profile, customers had to indicate their budget and they were increasingly mentioning that they were ready to pay more. And in the feedback that customers gave post purchase, they often mentioned the brands that they wanted to see included in the box.[52]

According to an industry observer, Stitch Fix's adding premium brands to its portfolio was an indication that the company was making "serious moves in the fashion space" (It had earlier appointed a director of brand development).[53] The observer also remarked that the introduction of premium brands said a lot about how loyal Stitch Fix's customers were; they had sufficient faith in the service to purchase expensive items without selecting the pieces themselves and without taking a look at the item first.[54]

According to Lake, 50% of the items offered by the premium brands would be exclusive to Stitch Fix, and in several instances, the pieces would be made based on the data given by the company. For instance, the luxury denim brand, Paige, unveiled a petite offering for the first time exclusively for Stitch Fix. Stitch Fix provided Paige data comprising Stitch Fix customer feedback that showed that there were no jeans available in the market for short women. This gave the brand the opportunity to enter that category.[55]

Growing from Strength to Strength

Stitch Fix's revenue increased from US$342.8 million in fiscal 2015 to US$730.3 million in fiscal 2016 and further to US$977.1 million in fiscal 2017 *(Refer to Exhibit VI for Stitch Fix's financial performance from 2014 to 2017)*.[56] The company was on course to achieve revenue of US$1.2 billion in fiscal 2018.[57] The number of active customers of the company had increased from 867,000 in August 2015 to over 2.19 million in June 2017 to a further 2.5 million by the end of January 2018 – Stitch Fix defined an active customer as an individual who had reviewed an order in the previous 12 months.[58,59,60] By the end of February 2018, the company had 5,800 employees, including 3,400 stylists, the majority of them working part time from home, interacting with customers.[61,62] As of August 2017, it had a team of over 80 data scientists. Over 50% of the team had PhDs from different disciplines, including astrophysics and computational neuroscience.[63] In November 2017, Stitch Fix went public with a market capitalization of $1.46 billion – the company was valued at over US$2 billion as of March 2018.[64,65]

The percentage of revenue generated from existing customers increased from 83% in 2016 to 86% in 2017.[66,67] Stitch Fix was a hit in middle America among busy working women and among those who had access to a limited range of shopping choices.[68] As per Slice Intelligence[g], Stitch Fix cornered 3% of the entire online purchases made by millennials[h] in 2016.[69] As per The NPD Group[i], Stitch Fix was among the top 10 retailers selling clothes online.[70]

Chinks in Stitch Fix's Armor?

There was, however, skepticism on whether customers who, at the start, were elated to get a customized box of clothing would continue to buy over the months and years. A Forrester[j] analyst, who had classified Stitch Fix as a subscription-box service, said, *"People will try something because they're intrigued--but then realize, 'This isn't something I need on a regular basis.'"*[71] For instance, a specific customer of Stitch Fix since 2015, who loved shopping from the comfort of her home and who liked the pieces that Stitch Fix selected for her, ceased using the service in 2016. Explaining the reason for not continuing with the service, she

[g] Slice Intelligence is a market research firm.

[h] Millennials are those born between the early 1980s and mid-1990s.

[i] The NPD Group is a market research company.

[j] Forrester is a business and technology research and consulting company.

wrote, *"I just don't shop a lot."*[72] And regarding Stitch Fix's selections getting better with the feedback that customers gave after each delivery, a woman who tried out the service remarked, *"...I don't have the patience for that."*[73]

Observers were concerned about Stitch Fix's decreasing revenue growth – sales growth went down from 113% in fiscal 2016 to 34% in fiscal 2017 to an expected 20%-25% for fiscal 2018.[74] Also. pointing out to the phenomenon that Stitch Fix's customers were prone to spending more in the first six months than in the subsequent six months and in the first year than in the subsequent year, an observer said that Stitch Fix was under pressure to ensure that customers continued to be amazed and satisfied order after order and that it had to keep acquiring more customers who were fine with spending money for someone else to pick their clothes.[75] Critics wondered whether Stitch Fix could keep expanding its customer base.[76]

Collection of data about consumers' bodies by clothing companies such as Stitch Fix also raised privacy concerns. "These body measurements look a lot like medical records,"[77] observed a legal expert. Growing competition in the online retail segment was also a concern. According to an observer, Stitch Fix faced a threat from *"incumbent giants that have the money and the hunger to crush upstart rivals."*[78] For instance, in June 2017, e-commerce giant Amazon launched a service called Prime Wardrobe that allowed subscribers to try on apparel at home and return pieces that they did not want at no cost.

Stitch Fix was also challenged by increasing advertising expenses that ate into its profit margins – the advertising expenses of the company increased from US$25.0 million in 2016 to US$70.5 million in 2017.[79,80] The company also stated that it had to increase its marketing expenditure to acquire new customers. *"The easiest customers they've already acquired. Now they (Stitch Fix) have to spend more and more to acquire these new customers,"*[81] pointed out an analyst.

According to Colson, however, there was scope for growth in areas the company hadn't yet entered (for instance, children's apparel or international customers). *"When we launch in new categories, we don't need to build new algorithms. They're there and ready for the data to start streaming through, and then they start training themselves,"*[82] noted Colson. Stitch Fix planned to better the experience for existing customers by introducing new ways to garner information and add more interesting products. For instance, the company unveiled a Facebook Messenger game that displayed styles to customers and allowed them to give a thumbs-up or thumbs-down to what they loved

and did not love in quick succession. "...through gamified experiences, we can collect information while the client is having fun," said Colson.[83]

And Lake, undeterred by the challenges, envisioned a future where "we're (Stitch Fix) able to fill many more parts of our clients' lives with personalization, and being able to meet her shopping needs, and being able to serve her whole family."[84]

Exhibit I

Stitch Fix's Recommendation System

Stitch Fix employed a range of algorithms to generate rank-ordered lists of the inventory and provide options to the human stylists. A filtering step eliminated the styles that the customer had received earlier or that contained attributes which s/he had asked the company to avoid.

For every balance style, the systems assessed the relative probability that the customer would like it. Stitch Fix tagged every piece multiple times with match scores from diverse algorithms and subsequently ranked them. The company used collaborative filtering[k] algorithms to forecast the outcome of sending a style to a customer who had previously not received it (for instance, the customers who had liked what you liked also liked...).

The company employed algorithms to process the tangible data collected from customers' self-descriptions and from apparel attributes. The mixed-effects modeling[l] mechanism was very helpful due to the longitudinal nature of the issue. The model allowed the company to learn (and monitor) the customers' preferences over time for each person and as a whole.

Stitch Fix deduced relevant unstated features of both customers and styles from other data (structured and/or unstructured) and employed them to improve its predictions. For instance, a new customer might indicate that she wore medium-sized blouses, but where would her preference lie on a spectrum of smallish mediums to largish mediums? The company inferred the preference from the customer's fit feedback and past purchases. It then employed the unstated features in, among others, the mixed-effects models.

Stitch Fix had photographic data in the form of inventory style photos and Pinterest boards. The company's systems scanned photos of apparel 'pinned' by customers and searched for similar-looking pieces in the company's inventory. The company employed trained neural networks[m] to generate vector[n] descriptions of 'pinned' images, and then calculated a cosine

[k] Collaborative filtering filters information by employing the recommendations of other individuals. The thought process is that people who were alike in their assessment of certain items earlier will probably agree again in the future. (Source: "Collaborative Filtering," http://recommender-systems.org/collaborative-filtering/)

[l] A mixed-effects model comprises fixed-effects and random-effects terms. Fixed-effects terms are generally the conventional linear regression portion of the model. Random-effects terms are associated with individual experimental units selected at random from a population and are responsible for variations between groups that might affect the response. (Source: "Generalized Linear Mixed-Effects Models," https://www.mathworks.com/help/stats/generalized-linear-mixed-effects-models.html)

[m] A neural network, or an artificial neural network, is an information processing mechanism that simulates the way biological nervous systems, such as the brain, process information. The important component of this mechanism is the novel structure of the information processing system. It consists of several highly interconnected processing elements (neurones) functioning together to solve specific problems. Neural networks, akin to human beings, learn by example. A neural network is designed for a specific application such as pattern recognition or data classification. (Source: Christos Stergiou and Dimitrios Siganos, "Neural Networks," https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html)

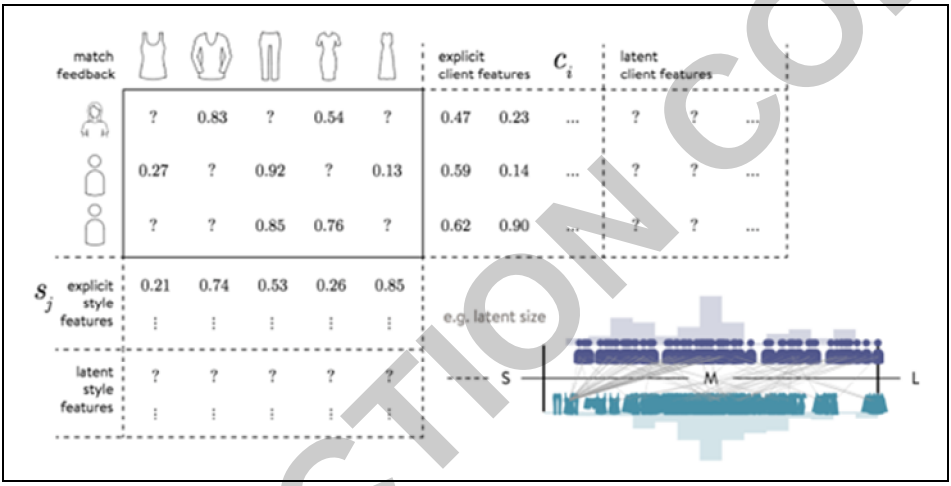
Contd...

similarity[o] between these vectors and pre-calculated vectors for every piece in Stitch Fix's inventory. The company employed natural language processing to score pieces based on the customer's request note and textual inputs from other customers about the same piece.

Source: http://algorithms-tour.stitchfix.com/#recommendation-systems

Exhibit II

A Visualization of Stitch Fix's Algorithms



Source: http://algorithms-tour.stitchfix.com/

[n] A vector is a quantity that has magnitude and direction and that is generally represented by a directed line segment whose length represents the magnitude and whose orientation in space represents the direction. (Source: https://www.merriam-webster.com/dictionary/vector)

[o] The cosine similarity between two vectors is a measure that computes the cosine of the angle between them. This metric measures the orientation and not the magnitude. (Source: Christian S. Perone, "Machine Learning :: Cosine Similarity for Vector Space Models (Part III)," http://blog.christianperone. com, December 9, 2013)

Exhibit III

A Stitch Fix Box

Source: www.stitchfix.com

Exhibit IV

An Instance of Algorithms Compensating for a Human Stylist's Shortcoming

According to Colson, human beings, by default, did not think statistically and were prone to construct narratives around the aspects most glaring in their consciousness. For instance, a human stylist, assigned to a tall customer, recommended an item she thought suited tall individuals and the item was liked by the customer. The human stylist would then reinforce the link between tallness and the product feature from this single instance, frequently ignoring the hundreds of other cases where the recommendation did not work. Algorithms were therefore an essential aid to reduce association fallacies.

Source: "Human-Machine Algorithms: Interview with Eric Colson," https://blog.fastforwardlabs.com, May 25, 2016.

Compiled from various sources.

Exhibit V:

Employment of Algorithms under Specific Scenarios

Problem	What the Algorithms Computed
Warehouse assignment	An algorithm computed a cost function for every warehouse based on a combination of its distance from the customer and the extent of matching of the inventories in various warehouses to the customer requirements. A cost matrix was generated by doing the computation for every customer. The allocation subsequently was a binary optimization problem.
Assigning a human stylist to a shipment	A match score was computed between every available stylist and every customer based on the past shipments, if any, handled by the stylist for the customer and the similarities between the expressed and latent style preferences of the customer and the preferences of the stylist.
Once the pieces for a shipment were chosen, what was the route that pickers had to travel by through the warehouse to fill the box?	Apart from employing the NP-hard Traveling Salesman Problem technique, Stitch Fix identified optimal groupings of shipments that could be picked together.
What quantity of what styles to buy? Which pieces should be sent to which warehouse?	Stitch Fix employed a system dynamics model, feeding it with historical data and employing it for robust optimization because of quantified uncertainties in its predictions.

Source: http://algorithms-tour.stitchfix.com/

Exhibit VI

Stitch Fix's Financial Performance (2014-2017)

Category	2014	2015	2016	2017
Revenue, net	73,227	342,803	730,313	977,139
Gross Profit	25,802	144,749	323,249	434,421
Selling, general and administrative expenses	30,242	108,562	259,021	402,781
Operating income (loss)	(4,440)	36,187	64,228	31,640
Net income (loss)	(6,333)	20,929	33,181	(594)
Cash	34,636	68,449	91,488	110,608
Total stockholders' equity (deficit)	(12,945)	8,539	49,947	61,861

** Thousands of US$*

Source: https://www.sec.gov/Archives/edgar/data/1576942/000119312517313629/d400510ds1.htm#toc400510_2

End Notes:

[1] Bridget Brennan, “The Retailer Redefining Personal Service in Ecommerce,” www.forbes.com, December 4, 2014.

[2] Hilary Milnes, “Now Public, Stitch Fix’s Success Strategy Lies in its Customer Data,” https://db.glossy.co, November 17, 2017.

[3] “Is Stitch Fix’s IPO The Latest E-Commerce Dud Or Is Its Styling Algorithm A Game-Changer?,” www.cbinsights.com, November 17, 2017.

[4] Jessica Pressler, “How Stitch Fix's CEO Katrina Lake Built a $2 Billion Company,” www.elle.com, March 1, 2018.

[5] Tracey Lien, “Stitch Fix Founder Katrina Lake Built One of the Few Successful E-Commerce Subscription Services,” www.latimes.com, June 9, 2017.

[6] Jessica Pressler, “How Stitch Fix’s CEO Katrina Lake Built a $2 Billion Company,” www.elle.com, March 1, 2018.

[7] Dhani Mau, “Stitch Fix Ceo Katrina Lake Wanted to Work at the 'Apparel Retailer of the Future,' So She Founded it,” https://fashionista.com, January 29, 2018.

[8] Tracey Lien, “Stitch Fix Founder Katrina Lake Built One of the Few Successful E-Commerce Subscription Services,” www.latimes.com, June 9, 2017.

[9] Sharon Gaudin, “At Stitch Fix, Data Scientists and A.I. Become Personal Stylists,” www.computerworld.com, May 6, 2016.

[10] Diana Budds, “How Stitch Fix is Using Algorithmic Design to Become the Netflix of Fashion,” www.fastcodesign.com, June 8, 2017.

[11] Drew Harwell, “Companies Race to Gather a Newly Prized Currency: Our Body Measurements,” www.chicagotribune.com, January 17, 2018.

[12] Anmol Rajpurohit, “Interview: Brad Klingenberg, StitchFix on Building Analytics-powered Personal Stylist,” https://www.kdnuggets.com, March 2015.

[13] Hilary Milnes, “Now Public, Stitch Fix’s Success Strategy Lies in Its Customer Data,” https://db.glossy.co, November 17, 2017.

[14] Jessica Pressler, “How Stitch Fix's CEO Katrina Lake Built a $2 Billion Company,” www.elle.com, March 1, 2018.

[15] Hilary Milnes, “Inside Stitch Fix’s Experiment to Design Clothing with an Algorithm,” www.glossy.co, April 17, 2017.

[16] Marisa Kendall, “Computing an Outfit: Stitch Fix Uses Algorithms, Machine Learning to Dress Its Customers,” www.mercurynews.com, December 15, 2016.

[17] Veronika Sonsev, “Can Algorithms Replace Humans At Stitch Fix?,” www.forbes.com, March 15, 2018.

[18] Denise Lee Yohn, “Stitch Fix Combines High Tech And High Touch To Transform Retail,” www.forbes.com, December 2, 2015.

[19] David Kirkpatrick, “For Stitch Fix, the AI Future Includes Jobs,” https://techonomy.com, October 2, 2017.

[20] Nina Zipkin, “Stitch Fix Founder Explains Why the Worst Piece of Advice She Ever Got Was to Raise A Lot of Money,” www.entrepreneur.com, July 20, 2017.

[21] Ryan Mac, “Stitch Fix: The $250 Million Startup Playing Fashionista Moneyball,” www.forbes.com, June 1, 2016.

[22] Kevin Duffy, “People Build Culture at Stitch Fix,” www.mcall.com, March 15, 2018.

[23] Nina Zipkin, "Stitch Fix Founder Explains Why the Worst Piece of Advice She Ever Got Was to Raise A Lot of Money," www.entrepreneur.com, July 20, 2017.

[24] Cathaleen Chen, "Stitch Fix Goes Public at $15. But What the Heck Is It, You Ask?," www.thestreet.com, November 17, 2017.

[25] Michael J. de la Merced and Katie Benner, "As Department Stores Close, Stitch Fix Expands Online," www.nytimes.com, May 10, 2017.

[26] Dennis Green, "Stitch Fix Files for IPO," www.inc.com, October 19, 2017.

[27] Hilary Milnes, "Inside Stitch Fix's Experiment to Design Clothing with an Algorithm," www.glossy.co, April 17, 2017.

[28] Ryan Mac, "Stitch Fix: The $250 Million Startup Playing Fashionista Moneyball," www.forbes.com, June 1, 2016.

[29] Marisa Kendall, "Computing an Outfit: Stitch Fix Uses Algorithms, Machine Learning to Dress Its Customers," www.mercurynews.com, December 15, 2016.

[30] Sharon Gaudin, "At Stitch Fix, Data Scientists and A.I. Become Personal Stylists," www.computerworld.com, May 6, 2016.

[31] Hilary Milnes, "Stitch Fix CEO Katrina Lake Predicts AI's Impact on Fashion," www.glossy.co, January 25, 2018.

[32] Sapna Maheshwari, "Stitch Fix And The New Science Behind What Women Want To Wear," www.buzzfeed.com, September 24, 2014.

[33] Alison Griswold, "Are You There, Margaret? It's Me, Ali.," www.slate.com, April 7, 2014.

[34] Denise Lee Yohn, "Stitch Fix Combines High Tech And High Touch To Transform Retail," www.forbes.com, December 2, 2015.

[35] Sharon Gaudin, "At Stitch Fix, Data Scientists and A.I. Become Personal Stylists," www.computerworld.com, May 6, 2016.

[36] Marisa Kendall, "Computing an Outfit: Stitch Fix Uses Algorithms, Machine Learning to Dress Its Customers," www.mercurynews.com, December 15, 2016.

[37] Shira Ovide, "Stitch Fix Is the Anti-Uber Silicon Valley Startup," www.bloomberg.com, October 20, 2017.

[38] Sharon Gaudin, "At Stitch Fix, Data Scientists and A.I. Become Personal Stylists," www.computerworld.com, May 6, 2016.

[39] "Algorithms Tour," http://algorithms-tour.stitchfix.com.

[40] Sophia Stuart, "How a Camping Trip Gone Awry Turned Into a Personal Shopping Start-Up," https://in.pcmag.com, February 11, 2016.

[41] Ryan Mac, "Stitch Fix: The $250 Million Startup Playing Fashionista Moneyball," www.forbes.com, June 1, 2016.

[42] Hayley Peterson, "This Hot Fashion Startup Eliminates the Hardest Part of Shopping," www.businessinsider.in, March 12, 2015.

[43] Marisa Kendall, "Computing an Outfit: Stitch Fix Uses Algorithms, Machine Learning to Dress Its Customers," www.mercurynews.com, December 15, 2016.

[44] Hilary Milnes, "Inside Stitch Fix's Experiment to Design Clothing with an Algorithm," www.glossy.co, April 17, 2017.

[45] Diana Budds, "How Stitch Fix Is Using Algorithmic Design To Become The Netflix Of Fashion," www.fastcodesign.com, June 8, 2017.

[46] Dave Gershgorn, "Stitch Fix Is Letting Algorithms Help Design New Clothes—and They're Allegedly Flying off the Digital Racks," https://qz.com, July 16, 2017.

[47] *Ibid.*

[48] Shira Ovide, "Stitch Fix Is Retail-Ish, But Wants to Be Tech," www.bloomberg.com, November 16, 2017.

[49] Hilary Milnes, "With a $300 Million Valuation, Stitch Fix Expands to Menswear," www.glossy.co, September 20, 2016.

[50] Bethany Biron, "Stitch Fix Is Latest to Start Selling Clothing in Bigger Sizes," www.glossy.co, February 22, 2017.

[51] Tiffany Yannetta, "Stitch Fix Launches Plus-Size With Over 90 Brands," www.racked.com, February 21, 2017.

[52] Dhani Mau, "Stitch Fix Goes after a Wealthier, More Fashion-Forward Customer," https://fashionista. com, August 22, 2017.

[53] Chavie Lieber, "Stitch Fix Is Going High-End," www.racked.com, August 22, 2017.

[54] *Ibid.*

[55] *Ibid.*

[56] Katie Roof, "Stitch Fix up Just 1% on First Day of Trading, after Reducing Size of IPO," https://beta.techcrunch.com, November 18, 2017.

[57] Katy Steinmetz, "Stitch Fix Has One of Silicon Valley's Few Female CEOs. But The Company Stands Out For More Than That," https://time.com, May 3, 2018.

[58] https://www.sec.gov/Archives/edgar/data/1576942/000119312517313629/d400510ds1.htm#toc400510_2.

[59] Elizabeth Winkler, "Stitch Fix Stays in Fashion," www.wsj.com, March 13, 2018.

[60] Khadeeja Safdar, "Stitch Fix's Customers and Sales Rise, but Profits Are Pinched," www.wsj.com, December 19, 2017.

[61] Jessica Pressler, "How Stitch Fix's CEO Katrina Lake Built a $2 Billion Company," www.elle.com, March 1, 2018.

[62] David Kirkpatrick, "For Stitch Fix, the AI Future Includes Jobs," https://techonomy.com, October 2, 2017.

[63] Pavithra Mohan, "Here's Why Stitch Fix's IPO Is The Tech Exit You Should Actually Care About," www.fastcompany.com, August 2, 2017.

[64] Lucinda Shen, "Stitch Fix's IPO Is More Exciting Than Its Stock Would Have You Believe," https://fortune.com, November 17, 2017.

[65] Jessica Pressler, "How Stitch Fix's CEO Katrina Lake Built a $2 Billion Company," www.elle.com, March 1, 2018.

[66] Ezra Galston, "Unboxing Stitch Fix's S-1," www.ft.com, https://techcrunch.com, October 23, 2017.

[67] Pavithra Mohan, "Report: Stitch Fix Customers, New and Existing, Are Spending Less," www.fastcompany.com, February 28, 2018.

[68] Jason Del Rey, "Stitch Fix Has Filed Confidentially for an IPO," www.recode.net, July 29, 2017.

[69] Hilary Milnes, "Cheat sheet: The Road to Stitch Fix's Blockbuster IPO," www.glossy.co, August 1, 2017.

[70] Lauren Thomas, "Amazon, Stitch Fix Already Rank among the Top Online Apparel Sellers," www.cnbc.com, July 31, 2017.

[71] Ryan Mac, "Stitch Fix: The $250 Million Startup Playing Fashionista Moneyball," www.forbes.com, June 1, 2016.

[72] Michael J. de la Merced and Katie Benner, "As Department Stores Close, Stitch Fix Expands Online," www.nytimes.com, May 10, 2017.

[73] Valentina Zarya, "What to Wear to Work: Fortune Tests Women's Online Styling Services," https://fortune.com, September 21, 2015.

[74] Jeremy Bowman, "Why Stitch Fix Could Be the Next Netflix," www.fool.com, December 27, 2017.

[75] Shira Ovide, "Stitch Fix Is the Anti-Uber Silicon Valley Startup," www.bloomberg.com, October 20, 2017.

[76] Jing Cao, "Stitch Fix Tumbles as Spending for New Clothing Lines Rises," www.bloombergquint.com, December 21, 2017.

[77] Drew Harwell, "Companies Race to Gather a Newly Prized Currency: Our Body Measurements," www.chicagotribune.com, January 17, 2018.

[78] Michael J. de la Merced, "Stitch Fix Prepares an I.P.O. in the Shadow of Amazon," www.nytimes.com, November 16, 2017.

[79] Eugene Kim, "Amazon Makes Life Hard for E-Commerce Start-Ups, but Stitch Fix Has Found a Path," www.cnbc.com, October 21, 2017.

[80] https://www.sec.gov/Archives/edgar/data/1576942/000119312517313629/d400510ds1.htm#toc400510_2

[81] Lauren Hirsch, "Stitch Fix Prices IPO of 8 Million Shares at $15, Below Expectations," www.cnbc.com, November 16, 2017.

[82] Hilary Milnes, "Now Public, Stitch Fix's Success Strategy Lies in Its Customer Data," https://db.glossy.co, November 17, 2017.

[83] *Ibid.*

[84] Dhani Mau, "Stitch Fix Ceo Katrina Lake Wanted to Work at the 'Apparel Retailer of the Future,' So She Founded it," https://fashionista.com, January 29, 2018.

Case 3

Disney's MagicBand System: Leveraging Technology to Enhance Customer Experience

Debapratim Purkayastha and Geeta Singh

> *"Disney could theoretically tailor the visitor's experience more effectively. . . MM+ is not solely for the benefit of Disney. Its intention is to simplify and enhance the visitor's experience. So while technology of this sort certainly benefits the business side, ultimately the consumer is the intended beneficiary."*[1]
>
> ***– James Crompton, Industry Analysts at IBIS World[a], in 2014***
>
> *"Disney's ability to do this, do it well, and do it in a way that doesn't jeopardize the information of guests isn't something I'm entirely confident about. It's also not entirely clear just what kind of data they'll be collecting from guests, what they'll be doing with this data, and who will have access to this data."*[2]
>
> ***– Ingram Connor[b], a Visitor of WDW, in 2013***

The world-renowned Walt Disney World (Disney) in Orlando, Florida, had always been a massive tourist attraction, and it had expanded into different sectors across the globe. Its theme parks, resort hotels, and transportation system had succeeded in attracting millions of visitors annually. For Disney, the focus had always been on improving customers' experience at its parks and its various other attractions. It believed in taking care of the minutest details to ensure the comfort and entertainment of customers and it had always updated its services to match the demands of a changing environment.

---

[a] IBIS World is an Australia-based research company.

[b] Ingram Connor, 25, from Casselberry, visits Disney World about twice a month.

In a bid to streamline the operations of its parks and enhance customer experience, Disney launched a vacation management system, MyMagic+ (MM+) in 2013. With its components, the My Disney Experience website, a mobile app, and MagicBands, the MM+ system was expected to enhance customer experiences through technology driven devices.

For the customers, the MM+ system offered a comfortable all-in-one radio-frequency identification device, the MagicBand, which would help deliver a more enjoyable, well-planned Walt Disney World experience. The company, on its part, planned to track the visitors' movements and get real time data on consumer behavior, which could further be used for various marketing activities. However, concerns were raised whether such a model could succeed in future. Analysts studied the pros and cons of the technology driven services and voiced their concern about the danger of the bands being cloned or of the information being retrievable by smartphones. There were also apprehensions about the privacy and security of the personal and financial information of customers.

Though Disney tried to assuage such fears, the MM+ was a project in its trial phase, and its success in the future could not be guaranteed. Until 2014, Disney was just testing the MagicBand system with some preselected customers. But even as it was preparing to roll out the project on a larger scale, the question before the team was whether customers would really be able to enjoy themselves while putting sensitive data about themselves at risk.

Background Note

As of 2014, Disney was the second largest media conglomerate in the world with its business split up into four major divisions –Studio Entertainment, Parks & Resorts, Consumer Products, and Media Networks. This massive company was world-renowned for its theme parks, welcoming characters, its attractions, and exceptional customer service. *(Refer to Exhibit I for revenues of the Walt Disney Company, and to Exhibit II for the World Disney World facilities at Orlando)*

In 1964, Walt Disney (Walt) purchased 27, 433 acres of swampland in Central Florida.[3] During a press conference, Walt described about his dreams of building a unique entertainment and vacation center. Disney also initiated a project called EPCOT – Experimental Prototype Community of Tomorrow[c]. In 1965, Walt Disney

[c] EPCOT is the second of four theme parks built at Walt Disney World in Florida, near the city of Orlando.

announced plans for building a theme park in Central Florida.[4] In October 1971, the Magic Kingdom Park was opened with two resort hotels.[5] The opening ceremony featured a parade of Disney characters and 76 trombones. From the moment the first guest walked in, the Walt Disney World Resort was considered an iconic destination, attracting millions of families around the world every year.[6] For the next few years, many adventures were added in the Magic Kingdom. In 1973, the Golf Resort Hotel (later renamed as Shades of Green) was opened. In 1976, Disney's first water park, River Country, was opened.

In 1982, the EPCOT center was opened, featuring attractions like Spaceship Earth, The American Adventure, Universe of Energy, and Kitchen Kabaret. With the addition of EPCOT, the total investment at The Disney Resort doubled, and the number of visitors crossed the 100-million mark.[7]

Soon, the first 4-D film of the park, Captain EO, made its debut at EPCOT. In 1988, Mickey's Birthdayland was opened at the Magic Kingdom to coincide with the 60th birthday of the iconic Disney character, Mickey Mouse,.[8] The Disney-MGM Studios and Pleasure Island were opened on May 1, 1989. In 1989, two new resorts, the Walt Disney World Dolphin and the Walt Disney World Swan, were opened at the Walt Disney World Resort near EPCOT.[9] The following year, the name of Mickey's Birthdayland was changed to Mickey's Starland.

To celebrate the 20th anniversary of the Magic Kingdom Park, a new daily procession called the Surprise Celebration Parade was launched in 1991. In 1996, on the occasion of the 25th anniversary of the park, the front of the Cinderella Castle was turned into an 18-story birthday cake. Disney World thus thought up new ways to celebrate different occasions and delighted its visitors.

By 1998, the Disney's Animal Kingdom Theme Park was opened themed entirely around animal conservation.

In order to make its operations more efficient, Disney launched the Fastpass service in 1999. The FastPass system was a type of line-skipping system, which allowed guests to reserve a time to visit an attraction. However, in 2002, the River Country was closed down because it caused death of an 11-year old boy from amoebas in the non-chemically treated water of the water park in 1980 and since then, there was considerable fall in the visitors' attendance.[10] In the same year, new attractions like the Buzz Lightyear Space Ranger Spin ride and The Many Adventures of Winnie the Pooh ride were added at the Magic Kingdom. In 2003, a baby boy elephant, Tufani, was born in Disney's Animal Kingdom Theme Park.[11] Another famous Disney World character, Donald Duck,

started accompanying the guests on an adventure at the Mickey's Philhar Magic. By 2006, some new features like the new Audio-Animatronics of Jack Sparrow and Captain Barbossa, were added in the existing adventures based on fantasy swashbuckler films 'Pirates of the Caribbean'. In January 2008, the name, Disney-MGM Studios was changed to Disney's Hollywood Studios.[12]

Working on an expansion strategy, the Disney management team announced the expansion of Fantasyland at Magic Kingdom Park in 2009. The expansion was expected to almost double the existing area, adding new attractions based on 'Beauty and the Beast,' 'Snow White', and 'The Little Mermaid'. By 2011, the guests were allowed to use their Fastpass tickets to watch Mickey Mouse at the Magic Park. The same year, a new 3-D version of Star Tours, 'Star Tours – The Adventure Continues' was opened at the Disney's Hollywood Studios.[13]

In 2011, in order to manage the huge footfall of visitors across its theme parks, Disney announced the launch of a billion-dollar high-technology project, 'NextGen'. At an investors' conference, Thomas O. Staggs (Staggs), Chairman, Walt Disney Parks and Resorts, announced in 2011 some major changes and said, "Guests will be able to reserve times for their favorite attractions and character interactions...secure seats at our shows and spectaculars...make dining reservations... and pre-book many other favorite guest experiences -- all before even leaving their house."[14]

Scott Smith, a former Disney employee who later on started teaching theme park management at the University of Central Florida, said, "NextGen's goal is to eliminate the wait time, streamline the experience for guests --and more importantly for Disney to make money."[15] As a part of the NextGen project, there was a plan to develop wristbands embedded with radio-frequency identification (RFID) microchips. Moreover, at EPCOT, Disney was already experimenting with RFID technology.

By 2012, the Disney's Toontown Fair was closed down following the expansion of Fantasyland. The following year, the Villas were opened at the Grand Floridian Resort.

The Walt Disney Resort contained 4 theme parks, 2 water parks, 5 golf courses, 26 owned-and-operated resorts, and hundreds of merchandise and dining experiences. The cast members represented more than 80 nationalities and could speak more than 50 languages.[16] As of 2014, the Walt Disney World was the largest single-site employer in the United States, and the resort was still undergoing a massive expansion project.

Mymagic+ System

Companies across the globe had been leveraging technologies to add efficiency to their business operations and thereby increase revenues. Similarly, Disney had continuously used updated technologies to enhance its operations and had delivered revolutionary attractions.

In 2013, Disney World introduced a vacation management system called MyMagic+ (MM+) to create a more magical atmosphere at Disney parks. Using this system, the company tried to recreate and enhance the visitors' vacation experience. Through wireless technology, Disney introduced the MM+ system, ending many traditional practices. The system allowed users to select FastPasses[d] and book their tickets through a website and mobile application 'My Disney Experience'. After more than five years of development, the MM+ was finally launched in January 2013.[17] The system was developed initially using 1,000 testers.

The initiative was aligned with modern consumer behavior and the company believed that "happier customers would spend more money". Staggs opined, "If we can enhance the experience, more people will spend more of their leisure time with us."[18] He further believed that MyMagic+ could prove 'transformational' for Disney.

John Padgett (Padgett), vice president of experience development for Walt Disney Parks and Resorts in Lake Buena Vista, Florida, who had been engaged in enhancing customer experience by creating next-level new business models focused on guest-experience, was working on this radio frequency technology backed new project of MagicBands.

This completely new range of technology with an online presence, My Disney Experience, blended the already existing FastPass system with MagicBands. Although the MM+ system included an updated My Disney Experience website and mobile app, the MagicBands were the highlight of the system. The website and the app allowed users to preselect their rides and VIP seats for fireworks & parades. They could also opt for the character by whom they wanted to be met and greeted.[19] The Disney parks in Orlando offered free Wi-Fi access to its visitors, which along with smartphone applications, enhanced the customized Disney experience.[20] The phone app enabled users to explore reservations for meals, rides, etc. Users also received text messages for ride openings.[21]

[d] FastPass is a service that reduces the waiting time in line for rides and facilitates reservation of seats for attractions like fireworks, parade viewing, shows, and character greetings.

The Great Wolf Resorts[e], another vacation company, had pioneered the use of RFID chips driven wristbands in 2006, but for Disney World, this was a huge project. For Disney's huge global operations, with 121.4 million admissions per year and revenue of almost US$12.9 billion being generated, the RFID technology would be a huge investment.[22] Disney planned to utilize the maximum share of its visitors' time by introducing this system.[23]

Magic Bands

The MagicBands were the centerpiece of Disney's expansive program for changing visitors' experience dramatically at the Florida resort. These wristbands were at the heart of Disney's MM+ technology. Disney believed that the MagicBands could create deeper and more unforgettable memories for its customers.

For approval of these wireless wristbands, Disney filed paperwork in October 2012 with the Federal Communications Commission (FCC). The bands were actually digital IDs, which visitors would wear throughout their Disney trip.

The MagicBands were thick rubberized plastic bracelets, which were expected to upgrade Disney's management while enhancing customer experience. These bands looked like accessories, but were actually developed on radio-frequency identification (RFID) transmitters. They could transmit data on the 2.4GHz spectrum. Microchips installed in these wristbands, would read the radio frequency identification.[24] At the Disney resorts, the bands could be read by short and long-range readers. *(Refer to Exhibit III for Working of Magic Bands)*

The MagicBands were durable and comfortable, with ridges on the inside to facilitate air circulation between the band and the wrist. The bands came in a universal fit that could be adapted for use by all the guests, with an adjustable peel-away layer to make it smaller for children. In addition, the guests could choose one of the seven colors available (pink, red, blue, orange, green, yellow, and gray) and could personalize the band with their encrypted name or nickname, as they wished. The bands were priced at US$12.95 plus tax, and various character-related bling and colorful accessories were sold at various Disney parks to customize these MagicBands.[25] *(Refer to Exhibit IV for the cost of MagicBands and where it could be purchased)*

[e] Great Wolf Resorts is an operator of 11 water parks in North America.

The bands could be purchased at Walt Disney World theme parks and the Downtown Disney area. Initially, they were available to visitors who were either Walt Disney World annual passholders or were staying at the Disney Resort hotels.[26] In case the bookings were made in advance, the bands were mailed to the visitor's home or else they were handed over at check-in points at the Disney resort.[27] The bands were linked to the customer's Disney account and were non-transferable. However, they could be re-used.[28] *(Refer to Exhibit II for details of obtaining MagicBands)*

Using MagicBands was simple; one only needed to touch the band against a sensor, called a touch point at places like resort room doors, theme parks, water park entrances, etc.[29] *(Refer to Exhibit V for an image of the MagicBand and the touchpoint).* While purchasing the MagicBand, the guests were required to enter a PIN code as their unique identity code. The bands were operational all the time and did not have an on/off switch.

In case the band was lost, it could be deactivated at the website 'My Disney Experience'. This would unlink the band from tickets and other entitlements.

Additionally, as of August 2014, the MagicBands were available as stand-alone item at the selected parks, Downtown Disney, Disney's Art of Animation Resort, Disney's Contemporary Resort, Disney's Pop Century Resort, and Disney's Port Orleans Resort.

A Win-win Technology

The technology driven theme parks were expected to modernize the old Disney system. Under the traditional process, visitors had to purchase paper tickets, wait in line for their favorite rides, and buy merchandise either through cash or credit cards.[30] The new system was designed to offer an enhanced experience to visitors.

The bands would serve as tamper-proof, waterproof tickets. Earlier, there had been the real possibility of losing tokens, like the room keys, park admission ticket, photo pass, Fastpass tickets, or even individual credit/debit cards. The bands would eliminate the chances of the tickets getting lost or wet as all the different tickets would be clubbed in the plastic band worn on the wrist. The program enabled the visitors to have paperless access to hotels and various attractions of the Disney Park; also, it provided hassle-free pre-trip reservations with an additional option of a wireless payment system. Being all-in-one devices, the MagicBands would free the visitors from having to carry

multiple paper tickets or cash while enjoying the rides, so they could travel lighter. The bands would allow check-in at FastPass+ locations and the Disney PhotoPass card.[31] Highlighting the multi-purpose features of the bands, Staggs said, "The MagicBand gives you all the entitlements of the park, without having to carry 10 to 15 different pieces of paper. This also allows us to unlock more special things for you."[32] The RFID equipped bands were expected to create a paperless environment for the Disney parks.

Staggs emphasized the fact that while collecting Fastpasses tickets, many family groups became separated. The advance online booking through MM+ would help them stay together during the whole trip while having more time for enjoyment.[33] From travelling, staying together to enjoying the various Disney attractions, the MagicBands would help all the family members and groups enjoy keep together and enjoy themselves.

Guests with accommodation at the Walt Disney World Resort hotels could unlock their rooms using their MagicBands, to gain more comfortable access to hotel services. The bands would act as credit cards for food and merchandise purchases made at Disney Resort hotel rooms.[34] In order to avoid frauds, purchases that could be made without the pin code being entered were limited to US$50 only.[35] Parents who wanted to cap children's spending limit could use the MagicBands to place restrictions that ranged from no purchases to any maximum purchase limit. The MagicBands would also help keep track of the pictures taken by Disney's in-park photographers.[36]

In addition, the visitors would have the option of changing their plans without any penalty being charged through an app on their cellphone or at the kiosks in the Disney parks[37]. It was believed that Disney was building massive incentives for its customers to use the MM+ system. According to Duncan Dickson, associate professor at the Rosen College of Hospitality Management at the University of Central Florida and former director of casting for Walt Disney World, "It's a new way of doing things. When you do new things, it's hard to put your arms around it, [but] standing in line is unproductive for theme parks and guests, so they are trying to create a better experience for them, not just in line but utilizing other areas of the park."[38]

Disney World had expanded through four theme parks, two water parks, retail areas, golf courses and many more attractions, and therefore, it was difficult for guests to enjoy everything in a single trip.[39] Now, with the MM+ system, the visitors could plan their rides with its advance ride reservations feature.

For Disney, this was a giant leap forward as it could drive it to reach the heights of customer service, said some industry observers. The intent of the RFID device was to integrate disparate system elements of Disney together and thereby streamline its operations and manage customers'activities in the finest possible way.

Moreover, the most attractive feature of this wireless system was the establishment of a single platform, which would help develop a vast customer database. The system would enable Disney to monitor its traffic and flow on an individual basis, thus generating a fresh database. The coded information of the wristbands would contain details of the attractions visited, purchases done, and even the transit paths of the parks taken by the visitors. In short, the company could track everything the visitors did in Disney Magic World. Disney would get to know the customers' details and their preferences and therefore could approach them more efficiently. Thus, the RFID bands would track the personal identity, movements, and financial status of the guests. This would further allow Disney to refine its marketing strategies. The rich data was expected to help Disney in evolving better market segmentation, marketing, and brand promotion, enabling it to deliver more sophisticated and personalized marketing.[40] Through wearable RFID bands, the travel and purchasing habits of the customers could be identified and thus, targeting of particular products to identified guests would become easier for the company. In the past, Disney had collected data for framing its various sales campaigns, but the MagicBands system would enrich its database with genuine data on a real time basis.

The data could further be used to establish personalized dialogue with the customers and thereby gain valuable feedback. Also, the visitors would become the live sample of surveys for Disney as they were traced by the radio frequency device readers throughout the park. This could help Disney reshape its offerings, inventory, and supply chain controls. The MagicBand system could allow the company to enhance guest experience while streamlining its operations through highly detailed consumer insights.

Other potential benefits could be realized from such wearable technology. Henry Harteveldt, a travel industry analyst, opined that the RFID enabled system could help to recognize special guests who were celebrating a birthday or marriage anniversary. The bands, loaded with the bearer's information, could be accessed by the various Disney characters. These characters could randomly greet the visitors by name, even wishing them on their anniversary or birthday, leaving them surprised. This would result in personalized services. As Staggs said,

"The more that their visit can seem personalized, the better. If, by virtue of the MagicBand, the princess knows the kid's name is Suzy. . . the experience becomes more personalized."[41]

Jay Rasulo, Chief Financial Officer, Disney Co. , said that the MM+ system would enable the company to sell "services that we can now offer on a personalized basis, because we know who you are, where you are and – if you tell us why are you are coming to visit Walt Disney World for this vacation – whether you're a first-time visitor, a 50th-time visitor, it is your child's fifth birthday, it is a graduation, it's an anniversary.... The more you share with us as a guest, the more we are able to tailor services and, we think, get a lift in selling those services."[42]

The bands could become a source of customized and mass marketing. With massive data, Disney could easily identify visitors' preferences and send them offers specifically designed as per their interests.[43] Based on the customers' browsing history, Disney could trigger impulse buying.

It was estimated that Disney had already made a huge investment of almost US$ 1 billion and would require further funds to train its 60,000 employees. Although, the implementation of the MagicBand system would be a costly venture for the company, it could result in reaping unexpected benefits in long run. In order to accommodate the MM+ system, Disney installed RFID touch points at its 28,000 resort rooms and replaced its entry turnstiles with 283 touch points equipped with RFID readers and biometric finger scanners.[44]

The bracelets would further help Disney harness the true potential of technology. The theme parks experts expected the MM+ would boost Disney's revenue significantly.[45] A travel agent, Dianne Newcomer, opined, "This new program was no small undertaking; these bands are changing the whole dynamics of a Disney World theme park visit."[46]

The MM+ system was believed to be a win-win situation both for the customers and for Disney. The system would enhance the consumer experience by minimizing the obstacles to enjoyment, while providing Disney access to consumer behavior. Further, Jim MacPhee, Senior Vice President of Walt Disney World Parks, said, "Over time, we hope to make sure that everything is incorporated into the MagicBand."[47]

The Other Side

The plastic flexible bands were easy to handle and carry, but being made of non-latex material, they could become irritating and sweaty.[48] Also, the all-in-one device could not be taken off as it was an

entry ticket as well. However, instead of taking a MagicBand, the Disney guests could opt for using a RF-enabled card that functioned in similar way.

Apart from this, the MagicBands bothered many with regard to the degree of privacy it could maintain. The bracelets made Disney enter into the debated terrain of personal data collection and its pros and cons. Analysts remarked that the hesitation of many visitors was justified as no company could actually guarantee information security.[49] The RF equipped bands were linked to visitors' credit card and biometric information, which many found an area of concern. People were worried as Disney World could track one's purchases as well.[50] Although corporate tracking had been an old norm in American cities, the customers were worried about who could access the tracked data.[51] Thus, many visitors were hesitant to use the MagicBands. However, the company gave the assurance that personal data would not be stored on such bands, and promised to take extensive measures to protect the visitor's information.[52]

In January 2013, Edward J. Markey (Markey)[f], Co-Chairman of the Congressional Bi-partisan Privacy Caucus, raised questions about the implications of Disney's new technology. Markey wrote to Robert Iger (Iger), Chairman and CEO at Disney Corporation, asking how the company would track and collect information, and target its guests. He wrote, "Collecting information about how guests use Disney amusement parks could improve the company's ability to target advertisements at its guests, including children. Although kids should have the chance to meet Mickey Mouse, this memorable meeting should not be manipulated through surreptitious use of a child's personal information."[53]

Responding to Markey's letter, Disney issued a public statement, "MyMagic+ is a completely optional program. Disney's privacy policies and practices are fully transparent and guests can choose whether or not to participate in MyMagic+. In addition, guests control whether their personal information is used for promotional purposes and no data collected is ever used to market to children. MyMagic+ is designed to make a visit to Walt Disney World more personalized, seamless, and customized than ever before."[54] Additionally, as a part of Disney's privacy policy, the company could not use personal information to market to under-13 kids.[55]

[f] In May 2011, Edward J. Markey introduced the Do Not Track Kids Act, legislation, which amended the Children's Online Privacy Protection Act of 1998 to extend, enhance, and update the provisions relating to the collection, use and disclosure of children's personal information and establishes new protections for personal information of children and teens

Customers could opt to not use the MagicBand service, and they could also decline to share personal information while enjoying all the services.[56] Guests could also participate in MyMagic+ without using the MagicBands; they could choose a card, which could not be detected by the long-range readers; however, some personalized services were dependent upon long-range readers and were available only to guests using a MagicBand.[57]

Clarifying on the privacy issue, Staggs said, "We designed this with privacy in mind from the get go. Walt Disney World Parks and Resort's vision is simple, it's to be the most trusted provider of shared family travel and leisure experiences throughout the world. The key word there is trusted. We won't betray that trust."[58] In order to protect guest information from unauthorized access, the company said it had implemented many technical, administrative, and physical security measures.

Researchers had raised doubts over another security issue – the possibility of a MagicBand being cloned, providing access to the customers' protected data.[59]

While Disney tried to put such doubts to rest, some of them remained. For instance, it was unclear whether the information would be stored forever with Disney or would eventually be deleted.[60] Analysts agreed that the venture involved a great degree of substantial risks; and the personal information and credit card details of the visitors needed to be secured.[61]

In addition, some visitors were worried about older family members, who were not very comfortable with technology and were not familiar with even smartphones, would use the system. Brittney Karpovich, a Disney lover who had been visiting Disney World three to five times a year, said, "At first I was excited that Disney was making changes; because I go so often, it is nice to see a change. But I was then scared about the new changes. I am a planner … but I fear the MagicBands and MyMagic+ will take away the little spontaneity that I still experience at the parks. More than that, I fear MyMagic+ will be a problem for my parents' generation and older who are not attached to their smartphones. I also have friends who are cast members, and I fear that jobs will be cut at Disney now with the new program."[62]

Moreover, the online component of the MM+ system, My Disney Experience, had some flaws as well; it was reported to be very slow at times; users were signed out unexpectedly; and changes made were sometimes not saved.[63]

Looking Ahead

The MagicBand was still in its infancy. Disney and its hardworking Imagineers[g] had conceived and come out with a product that had huge potential. Besides enhancing visitors' experience, Disney increased its revenue as well. In the first quarter of 2014 when the MagicBands were still in the testing phase, the revenue of the Disney parks and resorts segment jumped 6%.[64] *(Refer to Exhibit III for Revenues of The Walt Disney Company for 2009-2013)* However, the increase could not be solely attributed to MagicBands.

Many Disney fans were impressed with the convenience offered by the MagicBands and believed Disney's promise to bring 'an even more immersive, personalized, and seamless Walt Disney World Resort experience than ever before.'[65] Analysts viewed the MM+ system as a game changer, which could revolutionize the travel industry and how visitors spent their vacations. Many predicted that consumers' expectations could also be molded with such technology driven operations. Douglas Quinby, a senior analyst of a travel industry research authority, PhoCusWright, pointed out, "This initiative will move travel agencies and the industry from a volume-based model to a specialized and personal model. How do you make 1 million customers feel like one in a million?"[66]

Although Disney tried to address various privacy and security concerns, many customers were still hesitant to use the MagicBands.[67] The MyMagic+ system had turned the spotlight on whether the technology driven solutions could help any company extend customer engagement beyond traditional methods.[68] Many wondered how justified any company was to offer personalized services to its customers while putting their privacy at stake.[69]

The second beta phase of the MM+ system started on August 2, 2014. Disney was confident that it would attract a number of visitors through its MM+ system. As Iger said, "This is a very significant undertaking from a technological perspective, and we really want to make sure that we walk before we run because we don't really want to overload our technological backbone..... Right now, we're mostly adding costs associated with Magic+ ahead of what will be, we believe, some interesting revenue generating opportunities. I can also say that it's working, meaning those that are using it – and we've got a number of

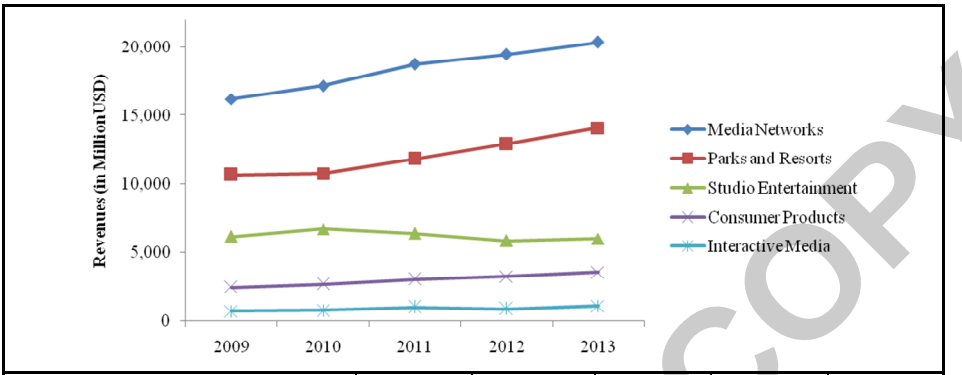
[g] Imagineers at Walt Disney World work mainly from the company's headquarters in Glendale, California, and are engaged in developing new concepts and improving existing attractions.

people that have used it – have reacted very well. This test that we're in right now will probably be used by over 80, 000 guests based on the reservations that we've written with this initiative attached to them."[70]

Adding up all the flaws and bugs, the offer still looked attractive and promising; however, it also put customers' privacy at risk. However, whether the MagicBands would be successful would become evident only after the system was fully rolled out. The question was, could Disney convince its visitors to accept the technology which would be optional, at least initially?

Exhibit I

Revenue of the Walt Disney Company

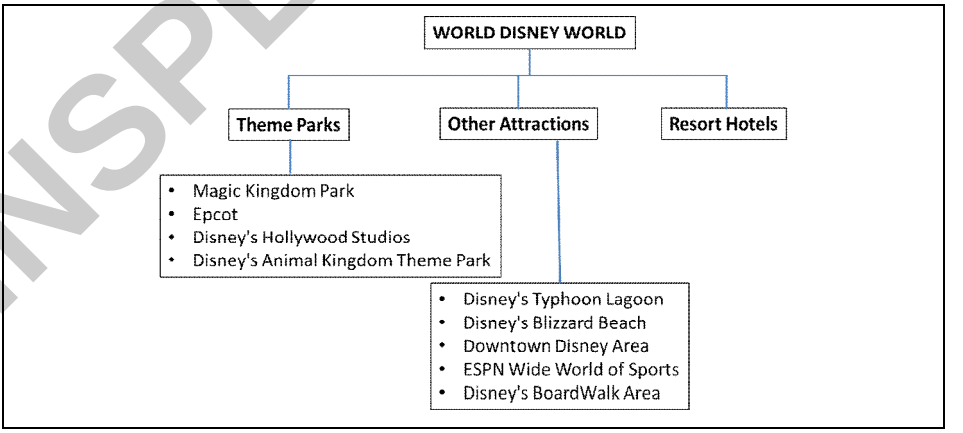


Revenues (in million USD)	2009	2010	2011	2012	2013
Media Networks	16,209	17, 162	18,714	19,436	**20,356**
Parks and Resorts	10,667	10, 761	11,797	12,920	**14,087**
Studio Entertainment	6,136	6, 701	6,351	5,825	**5,979**
Consumer Products	2,425	2, 678	3,049	3,252	**3,555**
Interactive Media	712	761	982	845	**1,064**
Total Consolidated Revenues	36,149	38,063	40,893	42,278	**45,041**

Source: Annual Reports 2010-11, 2011-12 and 2012-13

Exhibit II

World Disney World Facilities



Compiled from various sources.

Exhibit III

The MyMagic+ System

1. Customers were not forced to use the MM+ system; they could choose to participate in MyMagic+ system.
2. First information was collected, as per the company's online privacy policy, when users used the My Disney Experience website and mobile app.
3. After ordering the bands, the visitors would receive them through mail, before even leaving for the Disney resort. The guests could also opt to collect the bands from the resort itself.
4. The MagicBands containeda HF (high frequency) radio frequency (RF) device and transmitter, which could send and receive RF signals through a small antenna.
5. The bands could be read by short and long-range readers to deliver personalized services.
6. The devices were not GPS-based and therefore, did not enable collection of continuous location signals.
7. The devices did not store personal information; they contained only a randomly assigned code, which was linked to an encrypted database.
8. All the transactions were recorded. The information collected could spell out the details of visitors' experience like the waiting time for rides, restaurants, and other attractions.
9. Guests could choose not to receive any marketing information from Disney. Also, the company promised that the collected information would not be used to target advertising to children under age of 13.
10. The company said it had implemented various technical, administrative and physical security measures to protect guest information from unauthorized access, disclosure, use, and modification.

Source: https://disneyworld. disney. go. com/faq/my-disney-experience/privacy-policy/

Exhibit IV

Availability of Disney MagicBands

Magic Kingdom	Emporium on Main Street U. S. A.
	Frontierland Trading Post
	Space Mountain Gift Shop
	Mickey's Gift Station at the TTC
	Big Top Souvenirs
Epcot	Disney Traders in World Showcase
	Camera Center under Spaceship Earth
	World Traveler at the International Gateway
Disney's Hollywood Studios	Legends of Hollywood
	Stage One
	The Darkroom
Disney's Animal Kingdom	Chester and Hester's Dinosaur Treasures
	Disney Outfitters
	Creature Comforts
Downtown Disney	Disney's Pin Traders in the Marketplace
	DisneyQuest gift shop on the West Side

Source:http://www.wdwmagic.com/other/mymagicplus/news/31mar2014-for-day-guests-staying-offsite---everything-you-need-to-know-about-buying-magicbands.Htm

Cost of MagicBands

Guest type	MagicBands' Cost
Disney hotel Guests	Free by Post
Guests without park tickets	$12. 95 from various locations in WDW
Annual pass holders	Free by Post or $11. 65 if purchased in park (10% saving)
Regular park visitors	$12. 95 available at parks, WDW hotels, Partner hotels and stores

Source:http://www.totalorlando.com/blog/disney-parks-magicband-fastpassplus-mymagicplus-photopass-explained/

Exhibit V

Image of the MagicBand and the Touchpoint

Source: Tom Staggs, "Taking the Disney Guest Experience to the Next Level, " http://disneyparks. disney. go. com, January 7, 2013

End Notes:

[1] Sarah Sekula, "Disney Gets Personal with New Mymagic+ System," www.usatoday.com, February 25, 2014.

[2] "What People are Saying about MyMagic+ , " http://articles. orlandosentinel.com, August 17, 2013.

[3] "Walt Disney World Timeline," www.orlando-florida. net, Accessed on November 14, 2014.

[4] Jennifer Fickley-Baker, "Timeline: Celebrating 40 Years at Walt Disney World," http://disneyparks. disney. go.com, September 29, 2011.

[5] "Walt Disney World History," www.wdwmagic.com

[6] "Walt Disney World Resort," http://thewaltdisneycompany.com

[7] "Walt Disney World History," www.wdwmagic.com

[8] Jennifer Fickley-Baker, "Timeline: Celebrating 40 Years at Walt Disney World," http://disneyparks. disney. go.com, September 29, 2011.

[9] "Walt Disney World Timeline," www.orlando-florida. net

[10] Alisha Grauso, "Check Out the Eerie Abandoned Disney Park That Time Forgot," http://moviepilot.com, September 26, 2014 .

[11] Mai Hoang, "Success Born to Disney's Animal Park," http://articles. orlandosentinel.com, July 8, 2004.

[12] www.wondersofdisney. disneyfansites.com/timeline. html

[13] *Ibid.*

[14] Mary Quinn O'Connor, "Disney's 'NextGen' Plan is Expected Cut Wait Times for Rides and More," www.foxnews.com, April 18, 2012.

[15] *Ibid.*

[16] "Walt Disney World Resort," http://thewaltdisneycompany.com

[17] "Privacy Concerns Over Disney MagicBand," www.securitymagazine.com, January 1, 2014.

[18] Brooks Barnes, "At Disney Parks, a Bracelet Meant to Build Loyalty (and Sales)" www.nytimes.com, January 7, 2014.

[19] *Ibid.*

[20] Donald Livengood, "Disney's Magical Big Data Idea," www.enterprisecioforum.com, November 16, 2013.

[21] Mark Wilson, "A $1 Billion Project to Remake The Disney World Experience, Using RFID," www.fastcodesign.com, January 11, 2013.

[22] Brooks Barnes, "At Disney Parks, a Bracelet Meant to Build Loyalty (and Sales)," www.nytimes.com, January 7, 2014.

[23] Jason Garcia, "With MyMagic+, Mickey Will Watch Where You Go, What You Do," http://articles. orlandosentinel.com, October 5, 2013.

[24] "Disney World's New RFID 'MagicBands' Track Guests," www.cbn.com, August 1, 2014.

[25] "Fact Sheet: MagicBand," http://wdwnews.com

[26] "MagicBands& Cards – Frequently Asked Questions," https://disneyworld.disney.go.com

[27] "Fact Sheet: MagicBand," http://wdwnews.com

[28] https://disneyworld. disney. go.com/faq/bands-cards/transferring-to-friend/

[29] https://disneyworld. disney. go.com/faq/bands-cards/how-to-use-magic-band/

[30] Christina Farr, "Disney World's RFID Bracelets Use Data to Bring Fantasy To Life," http://venturebeat.com, January 7, 2013.

[31] "Unlock the Magic with Your MagicBand or Card," www.mickeyvacations.com/Packages/MagicBand. htm

[32] Bonnie Cha, "Tomorrowland Today: Disney MagicBand Unlocks New Guest Experience for Park Goers," http://allthingsd.com, May 29, 2013.

[33] Matthew Panzarino, "Disney Gets into Wearable Tech with the MagicBand," http://thenextweb.com, May 29, 2013.

[34] "Unlock the Magic with Your MagicBand or Card," https://disneyworld. disney.go.com/plan/my-disney-experience/bands-cards/

[35] Matthew Panzarino, "Disney Gets into Wearable Tech with the MagicBand" http://thenextweb.com, May 29, 2013.

[36] Dewayne Bevil, "Disney's MagicBands -- What Visitors Need to Know," http://articles.orlandosentinel.com, March 18, 2014.

[37] "Disney World Website Makes Trip a Breeze," http://archive. thenewsstar.com, October 31, 2013.

[38] Katherine Ferrara Johnson, "Updated: Testing Nearing Final Stages with Disney's MyMagic+," www.travelweekly.com, December 17, 2013.

[39] Jason Garcia, "With MyMagic+, Mickey Will Watch Where You Go, What You Do," http://articles. orlandosentinel.com, October 5, 2013.

[40] Elizabeth A., "Disney's MagicBand:The Theme Park World's Hottest Mobile Tech Innovation," www.instant. ly, January 10, 2013.

[41] Matthew Panzarino, "Disney Gets into Wearable Tech with the MagicBand," http://knotop.com,

[42] "'MagicBands' Open a New World of Marketing for Disney," www.retailcustomerexperience.com, October 7, 2013.

[43] Carolyn Heneghan, "Why Disney's MagicBands are a Cash Cow in the Making," www.fool.com, July 15, 2014.

[44] Claire Swedberg, "MagicBands Bring Convenience, New Services to Walt Disney World," www.rfidjournal.com, June 16, 2014.

[45] Richard Bilbao, "Expert: Disney's MyMagic+, MagicBandsCould Increase Impulse Sales in Big Way," www.bizjournals.com, January 18, 2013.

[46] "Disney World Website Makes Trip A Breeze," http://archive. thenewsstar.com, October 31, 2013.

[47] Dewayne Bevil, "Disney's MagicBands -- What Visitors Need to Know," http://articles. orlandosentinel.com, March 18, 2014.

[48] Drew Taylor, 'Disney World's New MagicBand: Pros and Cons" www.mapquest.com, June 4, 2014.

[49] *Ibid.*

[50] Jon Burstein, "Disney's Latest Experiment with Technology: MagicBands," www.miamiherald.com, November 23, 2013.

[51] Robert Niles, "Are Disney MagicBand Privacy Concerns Legit?" www.themepark insider.com, December 04, 2013.

[52] "Fact Sheet: MagicBand," http://wdwnews.com/releases/2014/04/30/fact-sheet-magicband/

[53] "Markey to Disney: Kid-Tracking Bracelet Raises Privacy Questions," www.markey. senate. gov, January 24, 2013.

[54] Robert Brigante, "Congress Questions Disney Regarding MagicBand / MyMagic+ Privacy Concerns, Marketing Toward Children," http://www.insidethemagic. net, January 24, 2013.

[55] Katherine Mangu-Ward, "Mickey Mouse Is Watching You," www.slate.com, January 31, 2013.

[56] Christina Farr, "Disney World's RFID Bracelets Use Data to Bring Fantasy to Life," http://venturebeat.com, January 7, 2013.

[57] "My Disney Experience – Frequently Asked Questions," https://disneyworld. disney. go.com/faq/my-disney-experience/privacy-policy/

[58] Sarah Sekula, "Disney Gets Personal with New MyMagic+ System," www.usatoday. com, February 25, 2014.

[59] Steve Griswold, "Can Disney's Magic Bands be Hacked?" http://mousechat. net, December 15, 2013.

[60] Ricky Brigante, "Congress Questions Disney Regarding MagicBand / MyMagic+ Privacy Concerns, Marketing Toward Children," www.insidethemagic. net, January 24, 2014.

[61] "Privacy Concerns Over Disney MagicBand," www.securitymagazine.com, January 1, 2014.

[62] "What People are Saying about MyMagic+," http://articles. orlandosentinel.com, August 17, 2013

[63] Mary Aviles, "The Federal Government Should Talk to the Mouse," http:// baumanresearch.com, November 25, 2013.

[64] Carolyn Heneghan, "Why Disney's MagicBands are a Cash Cow in the Making," www.fool.com, July 15, 2014.

[65] Adam Weinstein, "Disney World Creepily Tracks Visitors NSA-Style with MagicBands," http://gawker.com, January 2, 2014.

[66] Katherine Ferrara Johnson, "Testing Nearing Final Stages with Disney's MyMagic+" www.travelweekly.com, December 16, 2013.

[67] Robert Niles, "Are Disney MagicBandPrivacy Concerns Legit?" www.themepar kinsider.com, December 4, 2013.

[68] Mike Clark, "Disney MyMagic+ Sets A New Bar for Mobile Wallet Design," www.nfcworld.com, March 12, 2013.

[69] Kirsty Styles, "Disney World Goes Paperless with MagicBands," http://mobilemarket ingmagazine.com, January 9, 2013.

[70] Gina Jordan, "Disney CEO Says MagicBands are a Success," http://secureidnews.com, August 13, 2014.

Case 4

Big Data Strategy of IBM

Debapratim Purkayastha and Syed Abdul Samad

"Our goal is to offer the broadest and biggest depth and breadth of any vendor in the marketplace. As we fill up this portfolio, we're really filling in the holes that we had in (our) integration."[1]

– Nancy Kopp-Hensley, Director of Big Data strategy at IBM, 2013.

On October 29, 2013, International Business Machines Corporation (commonly referred to as IBM) completed the acquisition of The Now Factory. The Now Factory, headquartered in Dublin, Ireland, was a provider of analytics software that helped communications service providers (CSPs) in delivering enhanced customer experiences and offered new revenue opportunities. With the new acquisition IBM expected to enhance its Big Data platform by augmenting its existing products *(Refer to Exhibit I for a note on Big Data).* For instance, the acquisition complemented the IBM MobileFirst analytics portfolio. Similarly, the IBM InfoSphere Streams augmented The Now Factory's data mediation and analytics and helped IBM expand into more real-time analytics.

IBM, a US-based computer hardware and software major, had been on an acquisition spree in the Big Data space since 2006. The company had implemented a Big Data strategy, where it offered solutions to store, manage, and analyze huge amounts of data that was generated daily and equipped large and small companies to make informed business decisions. It had acquired around 30 companies to expand its Big Data portfolio of products and services. The company believed that Big Data had the potential to analyze the huge amounts of data available with organizations and create value for them. The company believed that its Big Data and analytics products and service would help its clients become more competitive and drive growth. The Big Data platform offered valuable insights to companies and

transformed their businesses. IBM established a broad portfolio of Big Data technologies and solutions, spanning services, software, research, and hardware. In 2011, the company invested US$100 million in the research and development of Big Data services and solutions. It also spent US$16 billion over these Big Data acquisitions. In the past, the company had integrated companies like Cognos, SPSS, Netezza, and Vivismo into its Big Data platform.

About IBM

IBM was a technology and consulting corporation headquartered in Armonk, New York, US. It manufactured computer hardware and software and offered infrastructure, hosting, and consulting services for products ranging from mainframe computers to nanotechnology *(Refer to Exhibit II for IBM's products and services).* Founded in 1911 as the Computing Tabulating Recording Company (CTR), IBM was the result of a merger of three companies namely the Tabulating Machine Company (founded in 1889), the International Time Recording Company (founded in 1889), and the Computing Scale Company (founded in 1885). However, the name IBM was adopted only in 1924. After that, the company went on to become the 19th largest firm in terms of revenue in the fortune rankings of 2012. For the year ended December 31, 2012, the company had revenues of US$ 104.5 billion along with an operating income of US$ 21.90 billion and net income of US$ 17.60 billion[2] *(Refer to Exhibit III for IBM's Financial Data).* Further, it ranked 1st and 5th on *Fortune's Company for Leaders and Most Admired Company* lists respectively. It was also recognized as the no. 1 *Green Company worldwide by Newsweek* for 2011-2012.

As of 2013, the company had 12 research laboratories worldwide – Almaden, Austin, Australia, Brazil, China, Dublin, Israel, India, Tokyo, Watson (New York), Zurich, and Nairobi – with Watson as its research headquarters. The company held the record for the most number of patents generated by a company for 20 consecutive years. Over a period of time, it had come up with inventions like the Automated Teller Machine (ATM), Floppy Disk, Hard Disk Drive, Electronic Keypunch, Relational Database, Dynamic Random Access Memory (DRAM), and Watson Artificial Intelligence. The company had also developed a unique enterprise class Big Data platform that blended traditional technologies (suited for structured, repeatable tasks) with complementary new technologies (ideal for ad hoc data exploration, discovery, and unstructured analysis) and allowed users to perform analytics, stream computing, data warehousing, and information integration and governance.

Big Data and Analytics: Origins

According to the *Oxford English Dictionary,* the term 'Big Data' was first used in 1941 to quantify the growth rate in the volume of data – alternatively known as the information explosion. In his article, *The Scholar and the Future of the Research Library (1944),* Fremont Rider, a Wesleyan University librarian, estimated that in every 16 years, the university libraries in US doubled in size. He used the term Big Data to refer to this increase. In 1967, in an article called Automatic data compression published by BA Marron and PAD de Maine, the term was used for the first time to indicate the information explosion on external storage devices through computers. Later, the term became more and more related to information in the digital format and the Information Technology sector.

Similarly, analytics too was not a new concept. Its evolution began with Analytics 1.0, which started in mid-1950. It was primarily descriptive analytics and reporting activity based on structured internal data. The analysis was used mainly for internal decision support. Various versions of Spreadsheets were examples of Analytics 1.0. By early 2000s, Analytics 1.0 evolved into Analytics 2.0, which could churn the unstructured information generated from big offline companies and Internet. Many new database systems, analytical software, and consulting services were developed over the years. In addition, a new role too emerged called Data Scientist, who could analyze the available data. By 2010, social media and a network of connecting devices had increased rapidly, leading to the availability of a huge amount of free real-time data. Then emerged Analytics 3.0, which did not look into databases for stored historical data, but worked based on data produced in real-time. Embedded in products and services, Analytics 3.0 thought and communicated back to the enterprise. For instance, sensors in fruit/cement containers collected data, determined whether the fruit was spoiling or cement was beginning to dry, and passed on the information to the company, so that the company could take the necessary action or plan preventive measures.

Data and analytics were interrelated. As the amount of data generated increased, the need for better analytics grew. Data in itself was not very useful as such, depending in turn on analytics for a meaningful conclusion or decision to be derived from it.

IBM and Big Data

In the 1950s, John Hancock Mutual Life Insurance Co. collected 600 Megabytes of corporate data. This was the largest amount of corporate data collected till then. The company was one of the pioneers

of digitization. It collected and stored information of two million policy holders on a Univac computing system. During the 1960s, American Airlines developed a flight reservation system using IBM computing systems and stored around 807 Megabytes of data. Federal Express, with its scanning and tracking, collected 80 Gigabytes of data during the 1970s. In the 1980s, with its focus on analyzing ATM transactions, CitiCorp., gathered 450 Gigabytes of data. Data from the Wal-Mart warehouse reached 180 Terabytes during the 1990s. With the explosion of the Internet, Internet software company Google generated 25 Petabytes of data in the 2000s and Facebook's data hoard in the 2010s was around 100 Petabytes. Further, the analysis of Facebook data generated 500 Terabytes of additional information per day.[3]

In 1956, IBM introduced 305 and 650 RAMAC (Random Access Memory Accounting) data processing machines. They had the first-ever disk storage. The 305 had fifty 24-inch disks with a total capacity of 5 Megabytes and the system weighed one ton; it was leased for US$3,200 per month (US$27,482 as of 2013).[4] IBM described it as "a stack of disks that stores millions of facts and figures less than a second from management's reach. Because transactions are processed as they occur, the fresh facts held in a random access memory show business as it is right now, not as it was hours or weeks ago."[5] That was the definition of Big Data in 1956. In 2013, the corporate data volumes ran into hundreds of Petabytes and Exabytes and customers could get a 1 Terabyte memory device for personal computers in the form of a 3.5-inch disk drive for just US$85 at any online store. However, with the introduction of its Data Processing Machine, IBM had created a market based on huge amounts of digital data storage and their fast access.

Big Data Strategy

Smarter Planet was a corporate initiative of IBM, which sought to highlight how government and business leaders were capturing the potential of smarter systems to achieve economic and sustainable growth and societal progress. In November 2008, in his speech at the Council on Foreign Relations, IBM's Chairman, CEO and President Sam Palmisano, outlined an agenda for building a 'Smarter Planet'. He emphasized how the world's various systems – like traffic, water management, communication technology, smart grids, healthcare solutions, and rail transportation – were struggling to function effectively. And he suggested that leaders, with the help of instrumentation, interconnectedness, and intelligence, should try to improve these systems and industries. He highlighted dozens of initiatives in which leaders had created smarter systems to solve the

planet's most pressing problems. IBM had created its Smarter Cities portal, which tracked progress on these issues in several key cities around the world. The company had been working with companies, cities, and communities around the world to build a Smarter Planet.

Smarter Planet was the cornerstone of IBM's business strategy. Over the years, it had come up with various initiatives like Smarter Cities, Smarter Commerce, Smarter Traffic, etc, which generated a lot of data from their complex processes. Judith S. Hurwitz, President and CEO of Hurwitz & Associates, LLC[a], opined that the Smarter Planet initiative was the design point for IBM's Big Data strategy though Big Data was not a new thing for the company. Discussing the Smarter Planet, IBM in its website said, "We've seen enormous advances, as leaders are using an explosion of data to transform their enterprises and institutions through analytics, mobile technology, social business, and the cloud. We've also seen how this new era is starting to create winners. They're changing how their decisions are made. They're redesigning how their teams work, reassessing how to serve their customers, and changing the very nature of business. It's the ability to harness data that gives these leaders their competitive advantage in the era of 'smart'."[6] IBM recognized that 'data' played a crucial role for these leaders in bringing about these changes. The company, which already had an Information On Demand strategy, evolved it into an Information Agenda to create a set of common products and services to support data usage, which became the Big Data Strategy. IBM's Big Data strategy was not just a product strategy; rather, it was a strategy based on the concepts of data in motion, data at rest, and data variety used such that it could change the way the world worked.

The company believed that Big Data had the potential to change the way organizations made decisions using analytics and demonstrate leadership in their respective sector while increasing shareholder value. With the surge in data volume and variety, organizations needed to analyze complex data and therefore favored more integrated and socially aware systems over transaction-oriented systems like enterprise resource planning (ERP), customer relationship management (CRM), and dealer management systems. Analytics and Big Data had become the big bets for IBM, as it recognized the huge market potential of being a leader in the field. IBM expected strong growth in the Big Data and Analytics areas. It expected the business to reach US$20 billion in revenue by 2015.[7] Technology law expert Andrew Brydon of Pinsent

[a] Hurwitz & Associates, LLC is a research and consulting firm focused on emerging technology including Big Data, cloud computing, service management, software development, computing management, and security.

Masons[b], said, "Businesses need to be aware and devise a 'big data' strategy. If they don't, they are going to fall behind market competitors."[8]

IBM's Big Data Platform

IBM committed itself to Big Data and Analytics through sustained investments and strategic acquisitions. In 2011, it invested US$100 million in the research and development of services and solutions that facilitated Big Data analytics. In addition, it had been bringing together as many Big Data technologies as possible under its roof. The Big Data strategy of the company was to combine a wide array of the Big Data analytic solutions and conquer the Big Data market. The company's goal was to offer the broadest portfolio of products and solutions with the depth and breadth that no other company could match. While the company was collecting (through acquisitions) and developing various products, it was actually trying to fill all the gaps it could find in its integration of Big Data solutions. The company's acquisition spree was to follow the strategy and capture the market. Since 2008, the company had spent over US$16 billion across 30 analytics-based acquisitions.[9] Its most recent acquisitions were The Now Factory and Daeja Image Systems *(Refer to Exhibit IV for some of IBM's Important Big Data Acquisitions).*

In addition, hundreds of mathematicians and data scientists were developing leading-edge analytics at IBM's commercial research organization. With so many acquisitions, product development, and integrations, the company had developed the largest patent portfolio in the world. IBM had incorporated many of the innovations relating to unstructured data management, text analytics, image feature extraction, and large-scale data processing into its Big Data platform. It had more than 27,000 business partners, 500 patents, 9,000 dedicated business analytics and optimization consultants, 400 researchers, and nine global analytics solution centers in Berlin, Beijing, London, New York, Ohio, Tokyo, Washington, and Zurich.[10]

IBM had developed a unique enterprise class Big Data platform that addressed Big Data business challenges, combining traditional technologies (for structured, repeatable tasks) with complementary new technologies to explore and analyze unstructured data. The core capabilities of its Big Data platform included Hadoop-based analytics,

[b] Pinsent Masons LLP is a full-service international law firm. The firm ranks amongst the top 20 law firms in the United Kingdom.

stream computing, data warehousing, and information integration and governance *(Refer to Exhibit V for the Capabilities of IBM's Big Data Platform).* Incorporating these core capabilities and supporting services, IBM created many products based on the platform *(Refer to Exhibit VI for IBM's Big Data Portfolio of Products).* It understood the business dynamics and challenges of each industry and helped businesses across industries to analyze large volumes of data and analyze data in motion (network traffic) to capture opportunities that they had earlier lost due to the lack of such a platform *(Refer to Exhibit VII for the List of Problems Addressed by IBM's Big Data Solutions).* The company had the technology and expertise to address specific business issues and deliver a rapid return on investment. A Forrester Research report issued in January 2013 termed IBM as an 'unshakable leader' of Big Data analytics solutions. IBM's Big Data strategy included the expansion of R&D, and acquisitions and business initiatives across its software, hardware, and services portfolio.

Adding to the Big Data platform capabilities were the open source software frameworks like Hadoop that played a major role in driving innovation at IBM. It enabled the processing of data-intensive computational tasks on a large scale. It also had other supporting tools like Hadoop's Distributed File System (HDFS), which enabled better access to data and supported data movement from Hadoop to other systems. Hadoop actually acted as a cost effective data warehouse. The Hadoop technology belonged to Apache Software Foundation and was in the early stages of its evolution. Hence, IBM tried to bridge the gaps in it and make it suitable for enterprise adoption. Bridging the gaps could be done in two ways. The first was to take the Hadoop code and modify it according to enterprise requirement thereby creating proprietary software insulated from further improvement, which would make it difficult to operate with complementary technologies. The second approach was to retain the code but add more layers and optional components that augmented and enriched the software. IBM followed the second approach with its Big Data Analytics products like InfoSphere BigInsights. Such an approach enabled it to adopt and add innovations or changes to the product. While IBM addressed Hadoop's limitations when adding changes, it also contributed bits of code to Hadoop like the Apache Derby, Apache Geronimo, Apache Jakarta, DRDA, XERCES, etc. The company constantly kept modifying and enhancing its analytics software and Big Data systems. For instance, in 2011, it updated its SPSS (SPSS Statistics 20.0) and added its SPSS Predictive Analysis Software to the cloud environment.

IBM's Big Data platform was not an adopt-it-all product. It had the convenience of implementing a single product, a set of products, or the entire range into the project or enterprise. People could start with

unlocking Big Data, reducing costs with Hadoop, analyzing raw data, analyzing streaming data, or simplifying their warehouse or a combination of a few or implement all at once.

The company, in October 2012, introduced a new acceleration technology for its portfolio of Big Data software. The IBM Digital Analytics Accelerator was based on Netezza (analytics software) and Unica (marketing software) technology and was capable of deriving insights from web traffic, customer emails, and social media to run more efficient marketing campaigns. The Accelarator also enhanced IBM's analytics capabilities offered through its SmartCloud platform. It further added new features to its InfoSphere Streams software, which helped communications providers measure network performance, customer behavior on social media, and geospatial information. Another enhanced product from IBM was the new mobile analytics software, Cognos Mobile, intended for Apple iPad users.

The company had made a clear distinction between its traditional business intelligence solutions and Big Data offerings. Michelle Warren, analyst and principal of Toronto-based MW Research & Consulting, said, "While data management and security is an issue for IT departments, accessing the information quickly is the biggest pain point. IBM's approach offers solutions to address all three concerns, allowing for businesses to run more cohesively."[11]

Addressing Emerging Business Requirements

In 2013, IBM was awarded the contract to support Thames Water Utilities Limited's[c] (Thames Water) Big Data project. The UK government planned to install smart meters in every home by 2020. Using these meters, the company would be able to collect a lot of data about the consumption patterns of its customers. As a part of its next five-year plan, Thames Water planned to invest in Big Data analytics to improve its operations, customer communication, services, and customer satisfaction using this data. It chose IBM as an alliance partner for the project to support technology and innovation.

[c] Thames Water Utilities Ltd, known as Thames Water, is a private utility company responsible for the public water supply and waste water treatment in large parts of Greater London, the Thames Valley, Surrey, Gloucestershire, Wiltshire, Kent, and some other areas of in the UK. It was founded in 1989 and head quartered in Reading, Berkshire, UK.

SHOP.CA[d] wanted to deliver a meaningful, customer-centric experience that delivered the right message at the right time and in an easy way to create and analyze marketing campaigns and promotions. It needed a scalable and flexible e-commerce solution. It chose IBM WebSphere Commerce to manage the online marketplace, product catalog, and rewards program. It also chose IBM Digital Marketing Optimization, including IBM Digital Analytics and IBM LIVEmail, to evaluate customer behavior to design tailored marketing programs.

Another company that adopted IBM's Big Data technologies to solve its problems was Eircom Group Ltd.[e] The company was facing fierce competition as many of its customers were changing their service provider and had recorded an annual loss of €1.5 billion due to the customer churn. The company then took the help of IBM's products like SPSS Collaboration and Deployment Services, SPSS Statistics Standard, and SPSS Modeler to analyze customer sentiment and used predictive modeling to determine factors that led to unfavorable customer experiences and to identify customers who were likely to change their service provider. As a result, the company was able to increase customer retention by 6%.

Another area that IBM's Big Data platform could cater to was the biopharmaceutical industry. Companies in the sector were getting increasing volumes of fast-moving and diverse data. They were also facing increasing pressure to deliver greater value and efficacy in medicines. At the same time, the price cuts were squeezing their operating revenues. IBM's portfolio of products offered business intelligence (BI) tools from the Cognos and advanced analytics and decision-management tools from the SPSS, which were designed to manage and analyze large sets of diverse types of data to support decision-making and deliver greater value to the companies in the pharmaceutical sector. These tools were also used in analyzing the data generated from clinical trials.

Similarly, there were many other companies that used IBM's Big Data platform and tools to overcome their business challenges. For instance, Vestas Wind Systems[f] used IBM big data analytics software

[d] SHOP.CA is Canada's largest e-commerce marketplace. It offers Canadians one-stop shopping for national and international brands with free shipping, free returns, and no cross-border fees. SHOP.CA provides millions of products across 26 categories.

[e] Founded in 1984, Eircom Group Ltd. provides communication services such as fixed-line, mobile, and broadband services.

[f] Vestas Wind Systems A/S is a Danish manufacturer, seller, installer, and servicer of wind turbines. With more than 47000 wind turbines installed over 30 years, it is the largest such company in the world.

and powerful IBM systems to improve wind turbine placement for optimal energy output. MultiView,Inc., a leading business-to-business (B2B) digital media publisher, with the support of IBM's InfoSphere Data Explorer, improved the relevance of its search results, reduced the time to create and update buyers guides (85% improvement), and projected more than a 10% sales increase with improved discovery capabilities. The Dublin City Council deployed intelligent traffic control systems using IBM's InfoSphere Streams. ASTRON, the National Institute for Radio Astronomy in the Netherlands, built a streaming analytics platform that ran on energy-efficient exascale supercomputing technology using IBM's InfoSphere Streams and SPSS modeler. These were only a few of the many businesses that solved their business challenges using IBM's Big Data platform.

Big Data Challenges

Over the years, consumer attention had shifted from radio, print, and television to the digital media as it facilitated real-time engagement of consumers. Brands competed for consumer attention through such media and relied on them for data and analytics for customer acquisition and retention and to offer tailor-made products and services to them. However, some analysts believed that relying on such data was a mistake. They said the real challenge with such data was that it was mostly unstructured and the difficulty lay in structuring it and filtering the genuine data. To sell a product to a customer, a brand had to first understand who the customer really was and to achieve that it had to connect with the customer on a personal level. There was a lot of data available in the media but it did not necessarily mean it was genuine or that it guaranteed meaningful insights. For instance, people maintained social media profiles where they engaged in liking and sharing of information. Though it might seem like a lot of useful data to the brands, there was a flipside to it in the form of private reality. People ensured that their social media profiles presented them in the best possible way but they did not necessarily reveal their inner emotions. Even though they volunteered to share information about themselves, the information was not necessarily accurate nor did it reflect their personal preferences. Brands found this lack of accuracy of the collected data a challenge. Even if valid data was collected, brands faced the challenge of developing a meaningful analytics strategy, as big brands like IBM had to measure outcomes across many channels, and most were not adept at tracking online metrics. Another ongoing challenge for IBM was keeping track of all the company's information warehouses.

The term 'Big Data' itself brought the focus on the volume of data which tempted companies to lose control on other aspects of this data and led to short-sighted decisions. According to Gartner, Inc.[g], the volume of worldwide information was growing at a rate of 59% annually.[12] However, the challenge lay in the three aspects of such data – volume, velocity, and variety. Many companies attempted to manage the Big Data challenge by excluding many other dimensions and focusing only on volume. Speaking about the issue of making sense out of Big Data, Yvonne Genovese, vice president and analyst at Gartner, said, "The ability to manage extreme data will be a core competency of enterprises that are increasingly using new forms of information – such as text, social, and context – to look for patterns that support business decisions in what we call Pattern-Based Strategy. Pattern-Based Strategy, as an engine of change, utilizes all the dimensions in its pattern-seeking process. It then provides the basis of the modeling for new business solutions, which allows the business to adapt. The seek-model-and-adapt cycle can then be completed in various mediums, such as social computing analysis or context-aware computing engines."[13]

While Big Data seemed to be exciting stuff for companies, developing a winning Big Data strategy posed the real challenge, as it involved investing in new technologies and talent. Though Big Data was in the nascent stage, small, medium, and big organizations were eagerly embracing it to understand the massive amounts of data associated with processes and costs related to supplier parts, manufacturing, logistics, quality control, customer service, and more. They were using it to establish predictive performance models to address quality issues and ensure customer satisfaction. However, the main challenge lay in finding people who were comfortable with business concepts and operations as well as adept with the analytical tools required to process the data. These people who could make quantifiable connections between the cause and effects were known as Data Scientists. However, McKinsey estimated that by 2018, the industry would face a shortage of more than 1.5 million data scientists.[14] Many businesses across the world had begun implementing Big Data projects. However, only a third of them were confident that they could carry it off with their existing staff. Analysts opined that until universities created the data science disciplines and produced data scientists to manage Big Data, the industry would face a dearth of talented individuals.

[g] Gartner, Inc. is the world's leading information technology research and advisory company. It was founded in 1979 and is headquartered in Stamford, Connecticut, US.

Addressing the Challenges

IBM had brought in new systems, software, and services to complement its Big Data platform. With these products it helped its customers to access and analyze data and use it to make informed decisions for the betterment of their businesses. The Big Data solutions were also meant to protect data and identify and restrict suspicious activity and block access to company data.

In 2012, IBM introduced Data Click analytical software to support unstructured data. It provided users with self-service data provisioning to support Big Data repositories like Hadoop, JSON, NoSQL, and JDBC. Speaking about Big Data and Data Click, Nancy Kopp-Hensley (Kopp-Hensley), a director in product marketing and strategy for Big Data Systems at IBM, said, "It has moved from what people refer to as 'the science experiment' into our enterprises and we have to hold that data to the same standards, manage it, and govern it appropriately for the enterprise. You can't drop those standards because it is unstructured data. This allows us to provide self service business intelligence capabilities but maintain data lineage."[15]

With the increased use of Big Data technology, there was also a rise in data security issues. IBM's InfoSphere BigInsights had built-in features that could be configured during the installation process for the security of data. It allowed four authentication options – No Authentication, Flat File authentication, LDAP authentication, and PAM authentication. Further, it had the option to configure HTTPS to provide security when the user connected to the BigInsights web console.

It introduced an Information Governance Dashboard, which provided a set of sample reports that one could use or customize for many common governance reporting tasks. The dashboard provided insights into policies, information governance rules, categories, and custom attributes, implemented data resources, etc., to allow the user to know the level of trust in a data source. With Big Data technologies, it had become easier to monitor and prevent unauthorized access of data. IBM also introduced a Big Data appliance known as PureData system for Hadoop. This helped customers in the faster deployment of Hadoop in their enterprises and reduced complexities. It also included many built-in archiving tools and a higher level of security and offered performance, usability, integration, and analytics functions. IBM was also launching a program called Big Data Stampede. "It is designed to bring all of the skills and expertise that you need to get these projects started but with a focus on skills transfer,"[16] said Kopp-Hensley. To

provide self-sufficiency with Big Data projects, IBM also included education and training resources for its clients. The company already had a Big Data University online and was also working with around 1000 universities to include this training in their curriculum.

Looking Ahead

In the past few years, Big Data had been the most hyped technology trend, and by 2013 it had started to gain acceptance as it held promising opportunities for businesses. It was showing its impact on the healthcare, industrial, retail, and financial sectors to name a few. It enabled companies to run live simulations of trading strategies, geological and astronomical data, and stock brokers could analyze public sentiment about a company from social media. Emerging technologies such as Hadoop, NoSQL, and Storm made such analytics possible. According to a Gartner survey in 2013, 64% of organizations had invested or planned to invest in the technology, but only 8% of them had actually begun deployment.[17] Many businesses were in the process of gathering information as to which business problems Big Data could solve for them.

IBM's advanced business analytics software services had become essentials for many businesses and the government. IBM provided a mix of Big Data analytic products to tap, analyze, measure, and predict patterns from the available data. It gave a competitive edge to many big and small businesses through its products and services. The company had been in the computing business for over a century and had significantly shifted from hardware to software products and services. Big Data has become the next big avenue of growth for IBM. The company's cognitive computer systems (computers modeled on the human brain) were in the testing stage. But the company believed that its future would be based on four data factors – social, mobile, analytics, and cloud, or 'SMAC' – combined with cognitive systems that would have a major impact on businesses, government, and society. And the IBM management was gearing up for a Big Data-driven future. In March 2013, Ginni Rometty, CEO, told business leaders at the New York-based Council on Foreign Relations that, "Big Data and predictive analytics would play a major role in how organizations make key decisions."[18] The development and working of the Watson computer system – that answered questions posed in natural language – was evidence of the company's vision toward building Big Data and cognitive technologies.

In March 2013, IBM along with Memorial Sloan-Kettering Cancer Center[h] planned to develop a cognitive system that combined Watson's capabilities with MSKCC's data to help provide oncologists a quick way to access diagnostic and treatment options and help doctors choose the best care for the patient. IBM was also involved in a research project named Deep Thunder, which used Big Data analytics to create precise weather forecasts and predict severe storms three days in advance. Kerrie Holley, a research fellow at IBM, said, "The current era of programmable computing had a good run, but it's coming to an end. Before that there was the tabulating era, where we used tabulating machines – pre-transistors, pre-computers. We see each of these eras lasting about 40 to 50 years. This new era of computing will require more innovation and invention. We're seeing more devices connected to the Internet. There's a lot of machine-to-machine interaction that's made possible because we're beginning to exploit the Web as a programmable, open platform. This technology will in no way replace doctors. It'll be an aid to doctors, and it will (enable) them to cut down on error rates."[19]

[h] Memorial Sloan-Kettering Cancer Center had extensive medical data from its history of treating cancer patients.

Exhibit I

A Brief Note on Big Data

What is Big Data?

Every day, the world creates 2.5 quintillion bytes (quintillion=10^{18} bytes or 1 billion gigabytes) of data. This data comes from text, pictures, videos, sales transactions, GPS signals, sensors gathering climate information, social media sites, log files, click streams, etc. This data is both in the structured and unstructured forms. Big Data is a popular term used to describe this exponential growth and availability of structured and unstructured data. This data had been ever-growing, coming at much faster rates and from a wide-variety of sources. With such a huge amount of data, it becomes complex for companies to handle and process it using traditional databases and data processing applications. So in terms of technology companies, the term Big Data refers to the technology (which includes tools and processes) that an organization requires to handle large amounts of data and their storage facilities. Big Data requires exceptional technologies to efficiently process large quantities of data like A/B testing, crowdsourcing, data fusion and integration, genetic algorithms, machine learning, natural language processing, signal processing, simulation, time series analysis, and visualization.

Examples of Big Data

- The NASA Center for Climate Simulation (NCCS) stores 32 petabytes of climate observations and simulations on the Discover supercomputing cluster.
- Decoding the human genome originally took 10 years to process. Now, it can be achieved in less than a week: the DNA sequencers have divided the sequencing cost by 10,000 in the last ten years, which is 100 times faster than the reduction in cost predicted by Moore's Law.
- Big Data analysis played a large role in Barack Obama's successful 2012 re-election campaign.
- Walmart handles more than 1 million customer transactions every hour, which is imported into databases estimated to contain more than 2.5 petabytes (petabyte= 10^{15} bytes; 2.5 PB=2560 terabytes) of data – the equivalent of 167 times the information contained in all the books in the US Library of Congress.
- Facebook handles 50 billion photos from its user base. It also has more than 1.15 billion active users generating social interaction data.
- More than 5 billion people call, text, tweet, and browse websites on mobile phones daily.

Big Data: Importance and Capability

Acquiring data was not the real issue for businesses and other organizations; what they could do using it was more of a concern. Technological advancements had made it possible for organizations to store most of the big data and analyze it with the availability of cheap and abundant storage facilities, faster processors, affordable open source, and distributed big data

Contd…

Contd…

platforms (such as Hadoop and SAS), parallel processing, clustering, virtualization, large grid environments, high connectivity, and high throughputs and cloud computing and other flexible resource allocation arrangements. Using Big Data and Big Data analytics, organizations could analyze the available data and reduce their costs and time, develop new products, and make smarter business decisions. They could also determine the root causes of failures, defects, and issues; optimize routes for package deliveries; send recommendations and advertisements to targeted segment of customers; etc.

Big Data Usage

Data had grown at an exponential rate in the world and had swept into every industry and business function. It had become an important factor of production, alongside labor and capital and helped companies in facing market competition. Big Data usage could create value – by making information transparent and usable at a much higher frequency; by collecting more accurate and detailed performance information on every aspect of business to boost performance; enabling narrower segmentation of customers and offering precisely tailored products or services for them. Enabling sophisticated analytics can substantially improve decision-making and help in the development of the next generation of products and services. Many small companies too want to improve their marketing by pinpointing the best sales prospects through social media.

Big Data Market

In 2010, the industry was valued at US$100 billion and was growing at 10% annually. Organizations like Software AG, Oracle Corporation, IBM, Microsoft, SAP, EMC, HP, and Dell have spent more than US$15 billion on software to specialize in data management and analytics. Big Data was used across sectors but a few sectors gained more from its use. The computer and electronic products and information sectors, as well as finance and insurance, and government are poised to gain substantially from the use of Big Data. Big Data has increased the demand for information management specialists. McKinsey has estimated that by 2018 there will be a shortage of 140,000-190,000 people with analytical skills to analyze Big Data. However, issues related to privacy, security, intellectual property, and even liability will need to be addressed.

Big Data Companies

The following 10 vendors were making a name in the Big Data space and helping businesses get more value from their unstructured and voluminous data sets.

- **Hadapt** specializes in integrating SQL with Apache Hadoop
- **Precog** comes with purpose-built business applications
- **Platfora** brings business intelligence capabilities to big data pools in Hadoop
- **YarcData** brings a small appliance into the enterprise to maximize the interesting capabilities of graphical search across messy and disparate data sets.

Contd…

Contd…

- **Datameer** has branched out with a namesake data integration and quality solution that bypasses a traditional data warehousing/data mart approach.
- **SiSense** has its Prism product used by Target and Merck
- **Kapow Software** is making some big value-adds in terms of integration and automation capabilities
- **ZettaSet** has a support system for Intel's Hadoop launch
- **Space-Time Insight** has also produced a real-world "situational intelligence" suite
- **ClearStory Data** offers a scalable application for data discovery and analysis across sources

Compiled from various sources

Exhibit II

IBM's Products and Services

Products	Services
Systems and servers by processor	**IT services**
• AMD processor-based servers	• Business continuity and resiliency services
• Intel processor-based servers	• End user services
• POWER systems	• Integrated communications services
• z-processor based	• IT strategy and architecture services
Systems and servers by operating system	• Maintenance and technical support services
• AIX	• Middleware services
• IBM i (i5/OS)	• Security and privacy services
• Linux	• Server services
• UNIX	• Site and facilities services
• Windows	• Storage and data services
• z operating systems	

Contd…

Contd…

Systems

- BladeCenter
- Blue Gene (US)
- Cluster systems
- Power Systems
- System x
- System z

Storage

- Disk systems
- Network attached storage
- Storage area networks
- Storage software
- Tape systems
- Media
- Storage A-Z

Software

- Software product finder
- Systems software

Software by brand

- CICS (US)
- Cognos
- DB2
- IMS
- Informix

Other services

- Application innovation services (US)
- Application management (US)
- Outsourcing

Business consulting

- Business analytics and optimization
- Customer relationship management
- Strategy and transformation
- Supply chain

Self Assessment Tools

- Business continuity and resilience (US)
- Data mobility (US)
- Energy Efficiency Assessment (US)
- Information infrastructure (US)
- Maintenance and Technical Support (US)
- Risk management (US)
- Storage optimization (US)

Contd…

Contd…

- Infosphere
- Lotus
- Rational
- SPSS
- Tivoli
- Websphere

Products for

- Small & Medium Business
- Business Partners

Printing Products

- Printing systems
- Printing paper (US)

Other products

- Networking
- Point-of-sale systems
- Semi-conductors (US)
- Workstations (US)

Resources

- IBM Certified pre-owned equipment
- Recycling
- Financing options
- Systems Advisor Tool (US)
- Warranty information

Resources

- IBM Institute for Business Value (US)
- IBM Press publications (US)
- Redbooks (US)
- Success stories (US)
- Training

Source: www.ibm.com

Exhibit III

IBM's Financial Data (2008-2012)

For the year ended December 31	2012	2011	2010	2009	2008
					US$ in millions except per share amounts
Revenue	104,507	106,916	99,870	95,758	103,630
Net Income	16,604	15,855	14,833	13,425	12,334
Operating (non-GAAP) earnings	17,627	16,318	15,023	13,452	12,293
Earnings per share of common stock					
Assuming dilution	14.37	13.06	11.52	10.01	8.89
Basic	14.53	13.25	11.69	10.12	9.02
Diluted operating (non-GAAP)	15.25	13.44	11.67	10.03	8.86
Cash dividends paid on common stock	3,773	3,473	3,177	2,860	2,585
Per share of common stock	3.30	2.90	2.50	2.15	1.90
Investment in property, plant and equipment	4,082	4,108	4,185	3,447	4,171
Return on IBM stockholders' equity	81.6%	71.2%	66.8%	80.4%	48.7%
At December 31					
Total assets	119,213	116,433	113,452	109,022	109,524
Net investment in property, plant and equipment	13,996	13,883	14,096	14,165	14,305
Working capital	5,807	8,805	7,554	12,933	6,568
Total debt	33,269	31,320	28,624	26,099	33,926
Total equity	18,984	20,236	23,172	22,755	13,584

Source: "2012 IBM Annjual Report", http://www.ibm.com/annualreport/2012/bin/assets/2012_ibm_annual.pdf, 2012

Exhibit IV

IBM's Important Big Data Acquisitions

Cognos	Jan. 31, 2008	IBM completed its acquisition of Cognos, a Canadian company whose products range from business intelligence to financial performance and strategy management to analytics applications.
SPSS	Oct. 2, 2009	IBM closed its acquisition of SPSS. The purchase of the Chicago company strengthened IBM's business analytics and optimization strategy by providing predictive analytic capabilities that can help customers predict future events and proactively act upon that insight to drive better business outcomes.
Coremetrics	Aug. 2, 2010	IBM completed the deal for San Mateo, Calif.-based Coremetrics, a leading Web analytics firm. The acquisition brought IBM business analytics capabilities that enable organizations to gain real-time insight into consumer interactions.
Netezza	Sept. 9, 2010	IBM closed its acquisition of Netezza, whose technology helps clients gain faster insights into their business information, resulting in increased performance at a lower cost of ownership. Netezza's data warehouse appliances bring analytics directly into the hands of business users within every department of an organization, including sales, marketing, product development and human resources.
Algorithmics	Oct. 21, 2011	IBM closed its acquisition of Algorithmics. The company's technology enables IBM to offer improved business insights at banks and investment and insurance organizations to assess risk and address regulatory challenges.
Emptoris	Feb. 1, 2012	IBM completed its acquisition of Emptoris, which expanded Big Blue's cloud-based analytics offerings by providing supply chain intelligence, leading to better inventory management and cost efficiencies.
DemandTec	Feb. 15, 2012	IBM completed its acquisition of DemandTec, whose cloud-based software helps companies drive profitability through measurable pricing, promotion, and assortment planning.
		Contd…

Contd…		
Varicent	May 23, 2012	IBM closed the deal for Varicent Software, which provides analytics software for compensation and sales performance management.
Vivisimo	May 29, 2012	IBM completed the acquisition of Vivisimo, which offers federated discovery and navigation software that helps organizations access and analyze Big Data.
Star Analytics	Feb. 1, 2013	IBM announced it is buying the software portfolio of Star Analytics, a privately held business analytics company in Redwood City, Calif. The combination of IBM and Star Analytics software will advance IBM's business analytics initiatives by allowing organizations to gain faster access and real-time insight into specialized data sources.
StoredIQ	Feb. 7, 2013	IBM completed its acquisition of StoredIQ. The firm's software will enable IBM to help organizations derive value from big data and respond more efficiently to litigation and regulations, dispose of information that has outlived its purpose, and lower data storage costs.
Daeja Image Systems	Sept. 30, 2013	Daeja's imaging tech allows IBM's enterprise content management customers to view and annotate unstructured data on mobile devices. The acquisition will enhance IBM's ability to provide customers with instant access to critical information, including archived data, from multiple computing devices such as tablets, smartphones, and PCs.
The Now Factory	Oct. 29, 2013	This acquisition complements the IBM MobileFirst Analytics portfolio, which is designed to enhance the way organizations analyze mobile device usage to provide exceptional customer experiences. The Now Factory's clients can also access IBM PureData System for Analytics and Hadoop-based IBM InfoSphere BigInsights for exploration and deep analysis of subscriber and network data.

Adapted from Darryl K. Taft, "IBM and Big Data: Acquisitions Drive Big Blue's Strategy", www.eweek.com, February 11, 2013, and other sources

Exhibit V

Capabilities of IBM's Big Data Platform

Core Capabilities	Supporting Platform Services
• **Hadoop-based analytics:** Processes and analyzes any data type across commodity server clusters. • **Stream Computing:** Drives continuous analysis of massive volumes of streaming data with sub-millisecond response times. • **Data Warehousing:** Delivers deep operational insight with advanced in-database analytics. • **Information Integration and Governance:** Allows you to understand, cleanse, transform, govern, and deliver trusted information to your critical business initiatives.	• **Visualization & Discovery:** Helps end users explore large, complex data sets. • **Application Development:** Streamlines the process of developing big data applications. • **Systems Management:** Monitors and manages big data systems for secure and optimized performance. • **Accelerators:** Speeds time to value with analytical and industry-specific modules.

Adapted from "IBM Big Data Platform", http://www-01.ibm.com/software/in/data/bigdata/enterprise.html

Exhibit VI

IBM's Big Data Portfolio of Products

- **InfoSphere Streams:** Enables continuous analysis of massive volumes of streaming data with sub-millisecond response times.
- **InfoSphere BigInsights:** An enterprise-ready, Apache Hadoop-based solution for managing and analyzing massive volumes of structured and unstructured data.
- **InfoSphere Data Explorer:** Discovery and navigation software that provides real-time access and fusion of Big Data with rich and varied data from enterprise applications for greater insight and ROI.
- **IBM PureData powered by Netezza technology:** Simplifies and optimizes performance of data services for analytic applications, enabling very complex algorithms to run in minutes not days.
- **DB2 with BLU Acceleration:** Advanced, innovative capabilities to accelerate analytic workloads for databases and data warehouses.
- **IBM Smart Analytics System:** Provides a comprehensive portfolio of data management, hardware, software, & services capabilities that modularly delivers a wide assortment of business changing analytics.
- **InfoSphere Master Data Management:** Creates trusted views of your master data for improving your applications and business processes.
- **InfoSphere Information Server:** Helps to understand, cleanse, transform, and deliver trusted information to your critical business initiatives, integrating Big Data into the rest of a company's IT systems.
- **IBM Security Intelligence with Big Data:** Provides exceptional threat and risk detection, combining deep security expertise with analytical insights on a massive scale.
- **Business Analytics for Big Data:** Extracts significant value from Big Data as part of the company's overall analytics initiative.
- **Big Data on Power Systems:** Delivering next generation applications for Big Data and analytics
- **Big Data solutions on System x:** Delivering insights with faster time to value.
- **Big Data on System z:** Leveraging the power and performance of today's mainframe to drive advanced analytics for a company's business.

Adapted from "IBM's big data portfolio of products for the big data platform", http://www-01.ibm.com/software/data/bigdata/platform/product.html

Exhibit VII

Problems Addressed by IBM's Big Data Solutions

Financial Services • Risk and fraud management • Customer analytics **Transportation** • Logistics optimization • Traffic congestion **Healthcare/Life Sciences** • Medical record text analytics • Genomic analytics **Telecommunications** • Call detail record processing • Customer profile monetization	**Energy and Utilities** • Smart meter analytics • Asset management **Digital Media** • Real-time ad targeting • Website analysis **Retail** • Omni-channel marketing • Click-stream analysis **Law Enforcement** • Real-time multimodal surveillance • Cyber security detection

Adapted from "Big data in action", http://www-01.ibm.com/software/in/data/bigdata/industry.html

End Notes:

[1] Jeff Bertolucci, " IBM's Big Data Strategy: Combine And Conquer", www.information week.com, February 13, 2013

[2] "2012 IBM Annjual Report", www.ibm.com, 2012

[3] "The Evolution of Big Data in U.S. Corporations", http://whatsthebigdata.com, March 10, 2013

[4] "Big Data in 1956: IBM Launches the Disk Drive Industry", http://whatsthe bigdata.com, September 14, 2013

[5] *Ibid.*

[6] "What is a Smarter Planet", www.ibm.com/smarterplanet/in/en/overview/ideas/ index.html?re=spf

[7] "IBM Completes Acquisition of The Now Factory", http://online.wsj.com, October 29, 2013

[8] "Every company needs a 'big data' strategy to avoid falling behind rivals, expert says", www.out-law.com, September 24, 2013

[9] Krishnan Parasuraman, "Part III: IBM's Strategy for Big Data and Analytics", www.ibmbigdatahub.com, November 7, 2012

[10] "IBM Completes Acquisition of The Now Factory", http://online.wsj.com, October 29, 2013

[11] Ryan Patrick, "IBM bets big on big data strategy", www.itworldcanada.com, October 25th, 2011

[12] "Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data" www.gartner.com, June 27, 2011

[13] *Ibid.*

[14] Howard D. Elias, "The Big Data Challenge: How to Develop a Winning Strategy", www.cio.com, June 14, 2012

[15] "With Big Data Offerings, IBM Helps Customers Address Emerging Challenges", www.dbta.com, September 16, 2013

[16] *Ibid.*

[17] Jody Gilbert, "Executive's Guide to Big Data Strategies and Best Practices", www.zdnet.com, October 1, 2013

[18] Jeff Bertolucci, "IBM's Vision For Cognitive Computing Era", www.information week.com, May 29, 2013

[19] *Ibid.*

Case 5

Amazon's Big Data Strategy

Debapratim Purkayastha and Adapa Srinivasa Rao

"At Amazon's recommendation team, we asked ourselves, how do we take this data and make a little bit more money, how do we apply it in this channel differently and it was really neat. Now it is being called 'big data' in the marketing and media world, but at the time we were doing this stuff, it was just kind of putting one foot in front of the other."[1]

– David Sellinger, Former Software Manager (Customer Behavior Research), Amazon.com, Inc. (Amazon), in October 2013.

Leading e-commerce company Amazon.com, Inc. (Amazon) and its subsidiary Zappos were ranked among the top ten retailers in the National Retail Federation Foundation/American Express Customers' Choice Awards[a] for two years (2010 and 2011) in a row. Industry observers felt that the coveted recognition was the result of Amazon's use of its big data[b] resources to provide superior service quality. Right from the time it had emerged as a dominant provider of Internet services in the early 2000s, Amazon had started to focus on big data to improve its performance. Along with many other major Internet

[a] The National Retail Federation Foundation/American Express Customers' Choice Awards were designed to know the consumer attitudes toward retailers' customer service. The survey collects data by polling consumers and is conducted by the marketing intelligence firm BIGinsight.

[b] Big data refers to the growth and availability of large volumes of data, both structured and unstructured. Such an exponential volume of data could not be analyzed by the traditional software used to handle databases. The latest trends in technology allowed decision making to be done largely based on data and analysis instead of past experience and intuition. According to a definition given by industry analyst, Doug Laney, big data spans three key dimensions, viz. Volume (amount of data generated), Velocity (speed at which data is streamed), and Variety (formats in which data comes in).

companies, it realized the importance of big data in the early 2000s, and had since then, focused on properly utilizing the huge databases of people who were shopping on its e-commerce portals.

Amazon leveraged its big data sources to give its customers good product recommendations and thereby improve the relationship with them. It utilized its big data resources to meticulously upgrade its famed customer recommendation system. Data on past purchases made by customers was used to give them highly customized product suggestions. Analysis of past customer data also helped Amazon in giving suggestions to new customers who were buying from its portal for the first time. Big data helped Amazon in developing 360 degree customer profiles and to create hyper-personalized marketing messages regarding the products based on the needs and preferences of individual customers. On the customer side, Amazon also utilized its big data resources to improve the quality of its customer care. Easy access to the profiles of customers and their past purchasing/browsing habits made it easy for the company's customer service executives to provide quick solutions to the complaints of customers. The acquisition of Zappos[c] by Amazon in the year 2009 further facilitated the use of big data in improving customer service quality. Big data resources were also put to some innovative uses like checking fraud at the organizational level. Product catalogue data was analyzed thoroughly to identify which of the items were more likely to be stolen. The results of this analysis were fed back to the warehouses of the company to limit the theft of items.

Other than improving its own performance, Amazon also helped other smaller e-commerce companies by allowing them to use its big data resources and improve their performance. An innovative service called Amazon Webstore, launched in 2010, allowed smaller companies to build their portals around Amazon's e-commerce platform. Users of Amazon Webstore could advertise their products on Amazon's portals by paying a small part of the sales proceeds as a commission to Amazon. For a fixed monthly fee for utilizing the service, partnering businesses could use Amazon's big data resources. Amazon Webstore was quite successful and was adopted by both small as well as big retailers such as Timex and Samsonite who did not want to have their own e-commerce system. Amazon's suite of cloud based Internet services known as Amazon Web Services (AWS) had also come out with solutions for small companies so that they could implement big data easily. A new service known as Kinesis announced in November 2013, could process

[c] Zappos, headquartered in Las Vegas, Nevada, USA was a leading online retailer of shoes and clothing products. This online shopping portal was founded by Nick Swinmurn in the year 1999.

high volumes of data flowing into AWS on a real time basis. According to some industry observers, this was Amazon's bid to close the loop on its integrated cloud stack and deliver an end-to-end solution for collecting and processing data. They felt that just by taking a relook at the various aspects of its big data capabilities and effectively leveraging on these, the company could emerge as a threat to the entire analytics eco-system.[2]

Background Note

Amazon was founded in the year 1994 by Jeffrey Preston Bezos (Bezos). It started its operations at a time when the reach of the Internet was increasing and the Internet was being considered as a potential business medium. Understanding the trend, Bezos came up with the idea of selling books through the Internet. He felt that books were the best products to sell online as millions of titles were in print and an e-commerce site could house and sell many more books than the conventional brick-and mortar bookstores. Bezos calculated that the common brick-and-mortar stores could not house more than 200,000 books at a time[3] and aimed to build a large online bookstore which would be bigger than any physical bookstore in the world. Amazon was initially funded with the money that Bezos borrowed from friends and relatives. Bezos and his wife, along with some employees, built the website and tested it for over a year before launching it *(Refer to Exhibit I for Timeline of Amazon).*

Amazon was finally opened to customers in the year 1995. Like many other technology giants, it was initially run from a garage – the one in Bezos' Washington home. At the time when Amazon started its operations, the book retailing market was highly fragmented and there was no major player except Barnes & Noble, Inc.[d] Barnes & Noble had one-tenth of the total market share but no online presence. Amazon thus got the first mover advantage and faced very little competition in its initial days of operation. Right from when it began its business operation, Bezos focused on customers and believed that customer loyalty was the key to penetrating the market and increasing sales. Amazon started to ship goods to all the 50 states in the US and 45 other countries within a month of its launch – and all this while still working from Bezos' garage. Amazon's popularity grew through word-of-mouth as customers recommended it to others. Within four months of its launch, Amazon was selling more than 100 books a day. The company's

[d] Barnes & Noble, Inc., headquartered in Manhattan, New York City, USA, is the largest book retailers in the United States.

impressive performance attracted investors and Amazon got its first big investment of US$ 100,000 from Madrona Venture Group, Inc.[e] in 1995. The company reported net sales of US$ 511,000 during the first six months of its operations and Bezos' confidence that he could make a success of the company increased.

At the beginning of 1996, Amazon moved to new headquarters – a small warehouse in Seattle. The company employed 11 people and offered 2.5 million book titles. Following the Japanese model, Amazon had very limited inventory and thereby kept its costs under control. It started an innovative affiliate marketing program called Amazon Associates Program in July 1996. The program allowed third party websites to sell books through links to Amazon posted on their sites for a commission of 15 percent on the total sales made. The program was a huge success and helped in expanding Amazon's reach without the company having to spend much on advertising. Experts opined that the program not only generated traffic to Amazon but enhanced the brand's presence online as these third party sites carried Amazon's logo on their pages. The Amazon Associates Program was later extended to all the products sold on Amazon's portal. Amazon went public in the year 1997 and offered 3 million of its shares for sale. The shares opened at US$ 18 a share and the IPO raised US$ 54 million for the company. In the year 1998, Amazon started selling DVDs with the opening of its video store which was followed by the launch of Amazon.com auctions in March 1999.[4]

Amazon's success attracted many new competitors like Book Stacks and Book Zone to the market and this led to higher competition for the company. To counter the competition effectively, Amazon introduced new features like online product reviews where customers could write their own book review as well as read reviews written by others. By the year 2000, Amazon had made a big change in its business model and started selling other products. In 2000, it also expanded its presence and launched sites in France and Japan. In the year 2001, Amazon allowed other retailers to sell their products through its site and took a part of the sales proceeds as its commission. In the last quarter of the year 2001, Amazon reported its first profit.[5]

Amazon's operations were further expanded in the year 2003 as it opened new websites in Asia-Pacific and European countries. In the year 2006, it launched a key subsidiary called Amazon Web Services.[6]

[e] Madrona Venture Group, Inc., headquartered in Seattle, Washington, USA, is a venture capital which primarily focuses on investing in early-stage technology companies.

Amazon Web Services provided an array of cloud based remote computing services to its customers. The advent of the digital era was changing the content consumption patterns of people. Many people started reading books and magazines on their desktops and laptops instead of buying physical copies. Responding to this change, Amazon introduced an e-book reader called Kindle in the year 2007. Kindle was a big hit in the market and heralded a new era of digital reading. Kindle was later released as an app for other devices working on operating systems like Android and iOS[f]. By the year 2011, the market capitalization of Amazon had reached the US$ 100 billion mark, making it one of the leading technology companies in the world. For the fiscal year 2012, Amazon had revenue of US$ 61.09 billion and a net loss of US$ 39 billion *(Refer to Exhibit II for the financials of Amazon).*

Big Data at Amazon

Over the years, Amazon had evolved from being a pure e-commerce player into a giant Internet services firm which offered a large range of services for individuals and corporations. It started to focus heavily on big data and embarked on its transition from a pure online retailer into a giant big data company.[7]Amazon along with other major Internet giants like Yahoo! Inc.[g] (Yahoo) and Twitter, Inc.[h] (Twitter) realized in the early 2000s that they had huge amounts of data about their users which they could put to valuable use.[8] While the other companies did not concentrate on the importance of big data, Amazon was quick to cash in on the invaluable database of people who shopped on its e-commerce portals around the world. The product recommendation team at Amazon thought of innovative ways in which it could use the data accumulated by the company.[9] The result was the big data revolution which transformed the way Amazon did business.

As an e-commerce giant, Amazon's success had always depended on making the right products available to its customers. Making the right products available in turn depended on understanding the precise products that customers wanted. Understanding the needs and tastes of customers involved doing proper market research as well as

[f] Android and iOS are the two leading mobile OS promoted by Google and Apple respectively. They are used in mobile devices like smartphones and tablets.

[g] Yahoo, Inc., headquartered in Sunnyvale, California, USA, is a leading multinational Internet company.

[h] Twitter, Inc., headquartered in San Francisco, California, USA, is a leading social networking and microblogging service. It allows its users to send and receive text messages which are limited to 140 characters.

analyzing its own customer base. Since its inception, Amazon had been renowned for its product recommender system which provided product suggestions to customers depending upon their past purchasing behavior. Data collected from its customers was the primary driving force behind Amazon's recommender system. Being the leading e-commerce player, Amazon had a large bank of data regarding the likes and the past purchasing behavior of its customer base. It had used this data bank to build its recommender system. Its earlier recommender system had been based on showing more items similar to the ones which were being looked for by its customers. This item-by-item similarity method was built on the basis of collaborative filtering[i] and was hugely successful in deepening the relationship with its customers. Its recommender engine had since been improved and perfected to give better results.

Amazon later started utilizing the historical purchase data of consumers as well and the click-stream data of all its customers to show webpages with uniquely customized information.[10] Using such data helped Amazon in many ways other than showing the related and alternative products that the consumers had been looking for. Mining the vast amount of data helped in understanding the inner feelings and likings of customers which they could not express themselves. Commenting on the importance of data in understanding the behavior of customers, Michael Driscoll of Dataspora[j] said, "You can ask people what influences their desire to renew their cell phone contract, and what people say and what they do are often very different. Data is the key to differentiating between what people say in terms of sentiment and what they do in terms of actions."[11]

Utilizing Big Data

Amazon leveraged on big data to improve its relationship with its customers and provide superior customer service. The online retailer built a vast database of its customers and their buying preferences over a long period of time. It was one of the first e-commerce companies to start using the cross-selling/up-selling method. This customer recommendation system was later augmented by utilizing its big data resources. Using big data, Amazon started analyzing the past product

[i] The Collaborative filtering system is a technique used in many recommender systems of e-commerce portals. Collaborative filtering involves filtering information or usage patterns through techniques involving multiple data bases, viewpoints, and agents.

[j] Dataspora, headquartered in Cambridge, Massachusetts, US is a leading big data and analytics consultancy.

purchases made from its online store by its customers and the other items that were purchased along with them. Data collected from its customers was used to give silent but highly customized suggestions to make them buy more. This analysis of data helped Amazon give product suggestions to existing as well as new customers who might not have otherwise bought a complementary product. This was the reason why Amazon's sites displayed 'Other customers who bought this item also purchased that item' kind of cross-selling recommendations.[12] Product related recommendations were also customized based on many factors such as the customer's location and demography. "It can even cross-correlate buying behavior between home and garden sales,"[13] said Jeff Kelly, lead big data analyst at Wikibon[k].

The bewildering range of products that were showcased on e-commerce portals made them seem unwieldy and incomprehensible to many customers. According to analysts, mining the treasure trove of information and providing relevant product recommendations could make e-commerce sites feel smaller and more intimate to the consumers.[14] Big data also helped Amazon in the development of its personalized marketing strategy – a tactic in which it excelled. Many e-commerce firms resorted to generic mass emailing of the products and offers that were available with them. This strategy led to the wastage of marketing efforts of many firms and the labeling of e-commerce mails as spam. Amazon created 360 degree customer profiles which tracked and stored everything related to customers like their browsing history, social data, tastes and preferences, past purchase history, etc. These 360 degree customer profiles facilitated the identification of discerning groups of customers who could be well targeted.[15] Amazon could create hyper-personalized marketing messages regarding the products based on the individual customer's needs and interests.[16]

Amazon also relied on big data to improve the quality of its after sales service to its customers. Most American customers were known to have a largely negative experience in their service interactions.[17] Amazon tried to solve this problem by leveraging on the large data it had regarding its customers. Having the right data helped it have a favorable exchange with its customers and to solve their problems quickly. Amazon's customer service executives had speedy access to data regarding the past purchases and browsing history of its customers. This enabled the company to provide quicker solutions to the problems and complaints of its customers. Complainants did not

[k] Wikibon, headquartered in Marlborough, Massachusetts, USA, is a community of practitioners and consultants on technology and business systems that use open source sharing of free advisory knowledge.

have to spell out their details like last names, contact numbers, and addresses repeatedly before their queries/problems were solved. This unique approach of utilizing big data to improve service quality gave spectacular results. After having a positive experience with Amazon's support team in one such encounter, Sean Madden, a top business blogger, said, "After nearly a decade of ordering stuff from Amazon, I never loved the company as much as I did at that moment."[18]

One of the factors which facilitated the use of big data for customer service was Amazon's acquisition of Zappos, the largest online retailer of shoes, in the year 2009. Amazon acquired Zappos for US$ 1.2 billion to expand its reach in product categories in which it was not strong.[19] Zappos was famed for using its customer database to provide a personal touch to its customers and turn them into its fans and cheerleaders.[20] Amazon adopted the customer service strategies of Zappos after it took over the company. The application of big data for improving customer service made Amazon and its subsidiary Zappos to rank among the top ten 10 retailers in National Retail Federation Foundation/American Express Customers' Choice Awards for the years 2010 and 2011 *(Refer to Exhibit III for top 10 retailers in National Retail Federation Foundation/American Express Customers' Choice Awards for 2010 and 2011).*[21]

Rather than using big data to just provide better product suggestions and improve the quality of service, Amazon used it to check fraud in the organization. An interesting area where Amazon benefited through using big data was in preventing warehouse theft. At any given point of time, Amazon had 1.5 billion items in its catalogues across its 200 fulfillment centers across the world. Theft of these items was a big threat to Amazon. The problem with identifying which of these items were more sought after by thieves was that both expensive and low-priced items were stolen. Inexpensive items too were often stolen due to reasons like their scarcity. To solve this problem, Amazon used big data and updated its product catalogue data nearly 50 million times a week.[22] Product catalogue data was collected, stored, and analyzed to identify which of the items were more likely to be stolen and the information was fed back to the warehouses *(Refer to Exhibit IV for the five components of big data process).* This helped Amazon in preventing the theft of items in its catalogues. Werner Vogels (Vogels), Chief Technology Officer and Vice President of Amazon.com, felt that data and storage should be unconstrained. "In the old world of data analysis you knew exactly which questions you wanted to ask, which drove a very predictable collection and storage model. In the new world of data analysis your questions are going to evolve and change over time and as such you need to be able to collect, store, and analyze data without being constrained by resources."[23]

Aiding other Companies with Big Data

Other than utilizing big data for improving its own performance, Amazon also helped other e-commerce portals to leverage its big data resources. This it achieved through an innovative service for smaller ecommerce businesses called Amazon Webstore. Amazon Webstore, launched in 2010, allowed retailers to build their portals around Amazon's e-commerce platform.[24] Amazon Webstore was an independent store outside of Amazon's official e-commerce store built on the third party domain name and brand. Users of Amazon Webstore could place Amazon Product Ads[l] on Amazon's portals. People who clicked on these product ads would be redirected to the site of the partnering site where the sale would be completed. As part of this partnership, Amazon allowed its partnering sites to use its big data while maintaining their independent identity as a small online e-commerce store.[25] Amazon charged its merchant partners a fixed monthly fee as well as a fixed commission for using its resources and big data resources.[26]

An interesting case where Amazon Webstore improved the performance of a small retailer was Anaconda Sports. Anaconda Sports, a successful sports retailer from New York, USA, found itself stuck with an inefficient and expensive e-commerce system with issues like inability to store all the customer information, lack of unique experience based on customer preferences, and poor customer service quality.[27] Modifying its e-commerce portal through Amazon Webstore made it possible for it to develop an efficient store which solved all the problems it had been facing as well as increase its sales substantially.

Commenting on the benefits of Amazon's Webstore for small and medium businesses, Scott Pulsipher, director of Amazon Webstore, said, "By leveraging Amazon's technology and infrastructure, Amazon Webstore levels the playing field for small- and medium-sized businesses, helping them quickly and easily build their businesses and improve the customer experience."[28] Amazon Webstore was implemented even by big brands like Timex, MTV, Boeing, and Samsonite which allowed them to improve their engagement with their customers *(Refer to Exhibit V for the screenshot of Samsonite's website built using Amazon Webstore).* Commenting on how Amazon Webstore helped to increase sales and cut costs at Timex[m], its e-commerce director Cal Crouch said, "When we launched our new Amazon

[l] Amazon Product Ads was an advertising program that allowed sellers to promote their products on the official e-commerce portals of Amazon in different countries.

[m] Timex, headquartered in Hoofdorp, Netherlands, is a maker of timepieces and luxury goods.

Webstore, we saw an immediate lift of 40 percent in revenue and average order size. And on the support side, we have gained the flexibility to make most changes to content as well as brand ourselves – saving us thousands (of dollars) in development costs."[29]

Amazon's Amazon Web Services (AWS) helped a lot of companies to develop better applications, deploy new products and services, and cut their costs *(Refer to Exhibit VI for AWS architecture).* Amazon offered its solution using familiar tools such as Oracle Database and Microsoft SQL Server, while also pioneering and promoting new platforms such as DynamoDB[n], Hadoop[o], and Redshift[p].[30] "One of the core concepts of Big Data is being able to evolve analytics over time. For that, a company cannot be constrained by any resource. As such, Cloud Computing and Big Data are closely linked because for a company to be able to collect, store, organize, analyze, and share data, they need access to infinite resources,"[31] said Vogels.

Small companies faced a lot of difficulties in adopting and deploying big data due to the limited resources at their disposal. Amazon came out with solutions for such companies so that they could implement big data easily. In November 2013, Amazon Web Service announced a new service for real time processing of big data. The service known as Kinesis, processed the high volumes of data flowing into Amazon's web-based storehouses on a real time basis. The tool had the capability to accept any number of data sources and could process terabytes of data per hour. It was intended to allow developers to create applications that worked on a real-time basis for tasks like website traffic analysis, business transactions related to marketing and finance, social media data, and logs *(Refer to Exhibit VII for Kinesis' architecture).* Commenting on the flexibility Kinesis would bring to businesses, an analyst at Neovise[q] Paul Burns said, "Sometimes people spend hours or days just collecting the data, then coming back and processing it, so it's out of date.... So Amazon said we'll take care of all that for you, just write your own program and connect to us."[32] The ability to create big data apps through Kinesis was expected to remove one of the biggest bottlenecks for smaller companies in adopting big data for their businesses. But one limitation of Kinesis was that all the data processing would be done at the data centers of Amazon itself instead of at the clients' location.

---

[n] DynamoDB is a managed NoSQL database service which makes it simple and cheap to store and retrieve large amounts of data.

[o] Hadoop is an open-source software framework for storing and processing large data-sets.

[p] Redshift is a fast and powerful data warehouse service which is a part of AWS.

[q] Neovise, headquartered in Fort Collins, Colorado, USA, is an IT industry analyst firm.

Looking Ahead

Analysts came up with suggestions on more ways in which Amazon could benefit from big data. Having its roots in selling books, Amazon had built a review system for the books sold through its website. Amazon's review system was mainly based on text reviews written by customers and the number of stars (from one to five) given to a book or author. This review system allowed Amazon to build a community and a loyal customer base. Over the years, there were allegations that many authors had found a way to manipulate Amazon's review system and to get paid reviews for their books. Such paid reviews tended to be biased and in turn, they impacted the reliability of the review system.[33] To solve this problem, some industry experts suggested that Amazon create a big data solution which would allow readers to give a vast range of additional feedback and comments which could be used to check the veracity of the reviews. Analysts opined that apart from improving the reliability of its review system, a big data based review system would also make it possible for Amazon to show more relevant reviews to the customers just as it suggested relevant products.

Another suggestion regarding the use of big data to further Amazon's prospects was in giving better competition to other big Internet companies like Google Inc.[r] and Facebook, Inc.[s] The business models of many of these Internet companies were based on online advertising. And according to an estimate by Google, 30 to 40 percent of its revenue from search advertising came from e-commerce sources.[34] The growth of Amazon as the Internet's one-stop shop and its increasing product base made it the primary destination for product searches, clearly bypassing Google. This unique position left Amazon in possession of more shopping data of people than any other Internet company. Some analysts were of the view that using big data, Amazon could beat Google and Facebook in the long term. Bezos acquired The Washington Post[t] for US$ 250 million in August 2013.[35] This acquisition sparked speculation among industry observers that Bezos would use big data to revitalize the news business and find new revenue sources for the ageing business. The use of big data analytics could give better insights into the readers of The Washington Post. Amazon could

[r] Google Inc., headquartered in Mountain View, California, USA, is a leading Internet-related products and services firm.

[s] Facebook, Inc., headquartered in Menlo Park, California, USA, is a leading social networking service.

[t] The Washington Post, headquartered in Washington, D.C., USA is a leading American newspaper.

thereby integrate the likings of readers in developing new products in the news business. According to analysts, Amazon's expertise in big data could transform the online news business in the same way as Bezos had transformed the 500-year-old book publishing business.[36] According to Wikibon's big data analyst Jeff Kelly, as of end 2013, Amazon had all the pieces of the big data puzzle but the firm would have put these together effectively to emerge as a dominant player in this space.[37]

Exhibit I

Timeline of Amazon.com

Year	Month	Event
1994	July	Amazon incorporated in Delaware
1995	July	Amazon.com launched. Sells its first book, "Fluid Concepts & Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought"
1996	July	Amazon.com Associates Program launched
1997	May	Announces IPO and begins trading on NASDAQ
1998	June	Launches music store
1999	March	Launches Amazon.com Auctions, the company's Web auctions service
	December	Jeff Bezos named Time Magazine "Person of the Year"
2000	August	Launches Amazon.fr (France)
	November	Launches Amazon.co.jp (Japan)
2001	April	Amazon partners with Borders Group to run the company's online bookselling business
2002	July	Launches Amazon Web Services
	November	Opens Apparel & Accessories Store
2003	June	Launches Amazon Services, Inc. subsidiary
2004	April	Opens Jewelery Store
2005	February	Introduces Amazon Prime
2006	September	Launches digital video download service, Amazon UnboxTM
2007	November	Launches Amazon Kindle
2008	November	Announces Frustration-Free Packaging initiative
2009	May	Introduces Kindle DX
2010	April	Amazon moves to new HQ in South Lake Union, Seattle
2011	July	Market capitalization of Amazon tops US$ 100 billion
2012	February	Amazon launches Sports Collectibles Store
2013	August	Amazon buys Washington Post

Source: Compiled from various sources.

Exhibit II

Financials of Amazon from 2008-2012 (Figures in US$ Millions)

	2012	2011	2010	2009	2008
Total net sales	61,093	48,077	34,204	24,509	19,166
Total operating expenses	60,417	47,215	32,798	23,380	18,324
Income from operations	676	862	1,406	1,129	842
Interest income	40	61	51	37	83
Interest expense	(92)	(65)	(39)	(34)	(71)
Other income (expense), net	(80)	76	79	29	47
Total non-operating income (expense)	(132)	72	91	32	59
Income before income taxes	544	934	1,497	1,161	901
Provision for income taxes	(428)	(291)	(352)	(253)	(247)
Equity-method investment activity, net of tax	(155)	(12)	7	(6)	(9)
Net income (loss)	(39)	631	1,152	902	645

Source: http://phx.corporate-ir.net/phoenix.zhtml?c=97664&p=irol-reportsannual.

Exhibit III

Top 10 Retailers in National Retail Federation Foundation/ American Express Customers' Choice Awards for 2010 and 2011

Sl. No.	2011	2010
1	Amazon.com	Zappos
2	L.L. Bean	Amazon.com
3	Zappos	L.L. Bean
4	Overstock.com	Overstock.com
5	QVC	Lands' End
6	Kohl's Department Stores	JCPenney
7	Land's End	Kohl's
8	JCPenney	QVC
9	Newegg	Nordstorm
10	Nordstorm	Newegg

Source: "Customers' Choice Awards," http://www.nrffoundation.com.

Exhibit IV

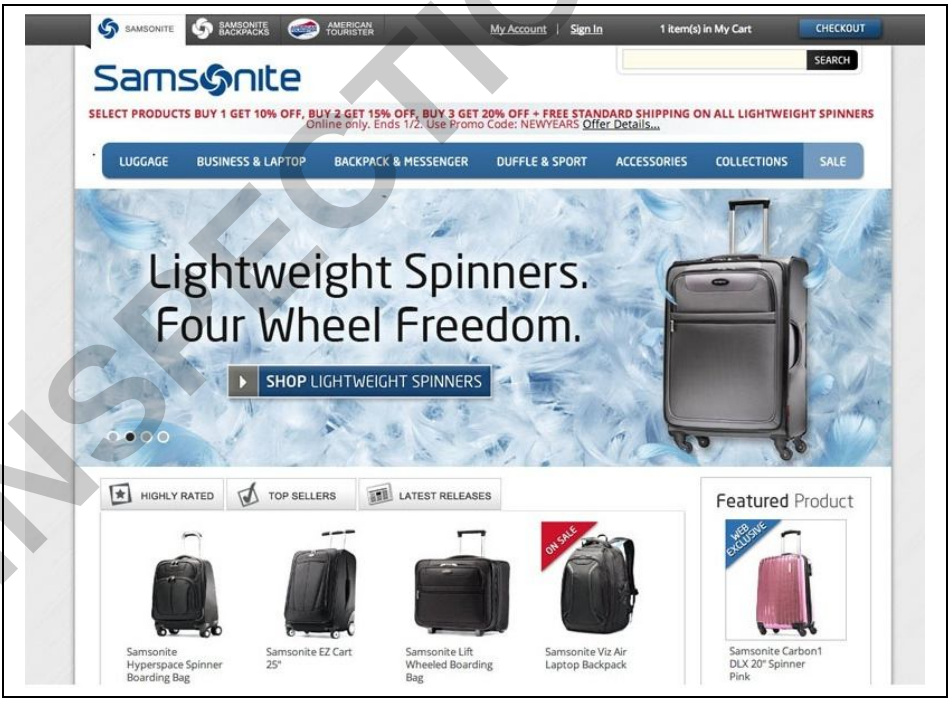
Five Components of the Big Data Process

Collect	Collecting and getting the data to the place where the process can be started.
Store	Storing the collected data before it is put to proper use.
Organize	Controlling the quality of data by knowing which data to include in the stream. Organizing also involves validating data in order to make sure that correct data is used.
Analytics	Analysis of well-organized data to create usable information.
Share	Information that is created through analytics is shared with those who need it.

Source: Eric Savitz, "CeBIT: Amazon CTO Werner Vogels Talks Big Data," www.forbes.com, August 3, 2012.

Exhibit V

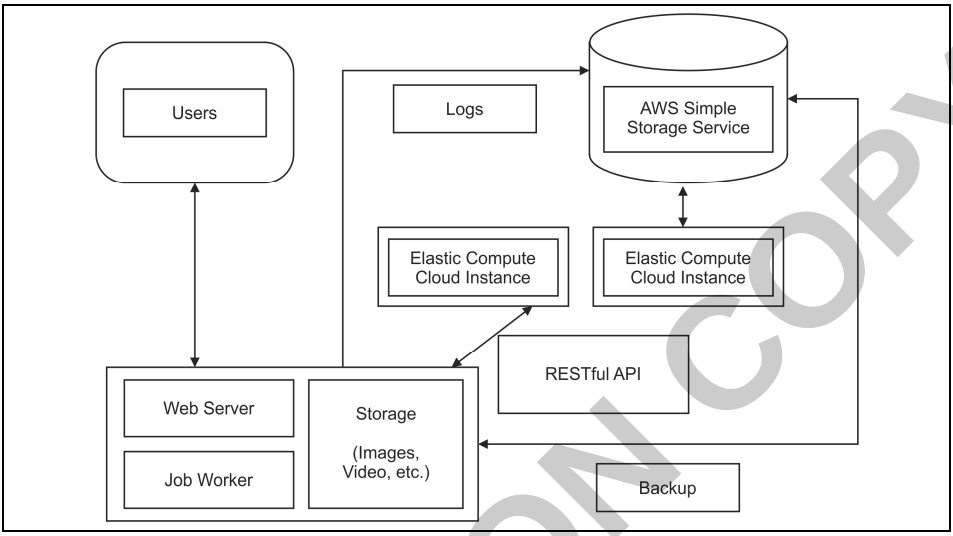
Screenshot of Samsonite's Website Built Using Amazon Webstore



Source: http://webstore.amazon.com/client-showcase/b/6254207011.

Exhibit VI

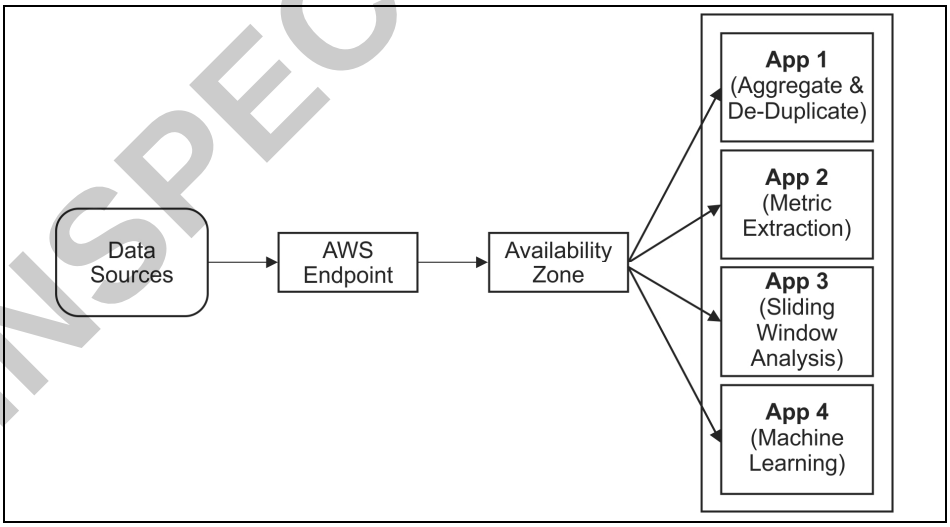
AWS Infrastructure



Source: http://d36cz9buwru1tt.cloudfront.net/pixnet_diagram_2.jpg.

Exhibit VII

Kinesis' Architecture



Source: http://arstechnica.com/information-technology/2013/11/amazon-wades-into-big-data-streams-with-kinesis/.

End Notes:

1 Teresa Novellino, "At War with Amazon? Rich Relevance Offers Big Data Weaponry," http://upstart.bizjournals.com, October 9, 2013.

2 Maria Deutscher, "Amazon Closes the Loop on Big Data," http://siliconangle.com, November 22, 2013

3 "History of AMAZON.COM," http://btmaushart.iweb.bsu.edu.

4 "Timeline History Amazon.com," http://amazongenius.com.

5 "History of AMAZON.COM," http://btmaushart.iweb.bsu.edu.

6 "History and Timeline," http://phx.corporate-ir.net.

7 "How Amazon is Leveraging Big Data," http://www.bigdata-startups.com.

8 Bill Vorhies, "A Brief History of Big Data Technologies – From SQL to NoSQLtpHadoop and Beyond," http://data-magnum.com, October 31, 2013.

9 Teresa Novellino, "At War with Amazon? RichRelevance Offers Big Data Weaponry," http://upstart.bizjournals.com,October 9, 2013.

10 "How Amazon is Leveraging Big Data," http://www.bigdata-startups.com.

11 "Q&A: What Can Non-IT Companies Learn from Amazon and Facebook about How to Leverage Big Data?" www.hightable.com.

12 Jodi Beuder, "How Can Big Data Improve the Customer Experience this Holiday Season?" http://www.icmi.com, November 5, 2013.

13 Constance Gustke, "Retail Goes Shopping Through Big Data," www.cnbc.com, April 15, 2013.

14 AdriaSaracino, "Interesting Ways Businesses Use Big Data to Improve Personalization," http://www.clickz.com, April 23, 2013.

15 Lisa Desjardins, "How Amazon Uses Marketing Personalization," http://www.nectarom. com, October 7, 2013.

16 Lisa Desjardins, "How Amazon Uses Marketing Personalization," http://www.nectarom. com, October 7, 2013.

17 Sean Madden, "How Companies like Amazon Use Big Data to Make You Love Them," http://www.fastcodesign.com, May 2, 2012.

18 Sean Madden, "How Companies like Amazon Use Big Data to Make You Love Them," http://www.fastcodesign.com, May 2, 2012.

19 "History and Timeline," http://phx.corporate-ir.net.

20 Justin Amendola, "What Zappos.com can Teach you About Turning Customers into Mega Fans," http://www.business2community.com, May 19, 2012.

21 Frank Reed, "Amazon Retains Top Spot in Customer Service Poll, Zappos Third," http://www.marketingpilgrim.com, January 19, 2012.

22 Ryan Lawler, "How Amazon Uses Big Data to Prevent Warehouse Theft," www.gigaom. com, October 18, 2011.

23 Roberto V. Zicari, "On Big Data: Interview with Dr. Werner Vogels, CTO and VP of Amazon.com," www.odbms.org, November 2, 2011.

24 Chris Crum, "Amazon Launches New Webstore E-Commerce Product," http://www. webpronews.com, May 24, 2010.

[25] "Power of Amazon," http://webstore.amazon.com/power-of-amazon-your-brand/b/6254196011.

[26] "Amazon Webstore- Pricing," http://webstore.amazon.com/amazon-webstore-pricing/b/6368778011.

[27] AdriaSaracino, "Interesting Ways Businesses Use Big Data to Improve Personalization," http://www.clickz.com, April 23, 2013.

[28] Chris Crum, "Amazon Launches New Webstore E-Commerce Product," http://www.webpronews.com, May 24, 2010.

[29] Chris Crum, "Amazon Launches New Webstore E-Commerce Product," http://www.webpronews.com, May 24, 2010.

[30] Doug Henschen, "Amazon's Vogels: Big Data Belongs In The Cloud," www.informationweek.com, April 19, 2013

[31] Roberto V. Zicari, "On Big Data: Interview with Dr. Werner Vogels, CTO and VP of Amazon.com," www.odbms.org, November 2, 2011

[32] Andy Patrizio, "Why Amazon's Kinesis Tool is a Big Deal for Working with Big Data,"www.citeworld.com,November 22, 2013.

[33] Amir Kurtovic, "A Better Review: Why Amazon Should Embrace Big Data to Fix Its Ratings System," http://www.amirkurtovic.com.

[34] David Hughes, "Big Data is the Only Way to Compete with Google," http://allthingsd.com, July 18, 2013.

[35] "Amazon Boss Jeff Bezos Buys Washington Post for $250m," http://www.bbc.co.uk, August 6, 2013.

[36] Karyl Scott, "How Bezos Could Apply Big Data and Other Amazon Tactics to the News Business," http://www.citeworld.com, August 14, 2013.

[37] Maria Deutscher, "Amazon Closes the Loop on Big Data," http://siliconangle.com, November 22, 2013

Case 6

Netflix: Leveraging Big Data to Predict Entertainment Hits

Debapratim Purkayastha and Tangirala Vijay Kumar

"Netflix has so thoroughly analyzed viewer patterns that they now know what viewers enjoy and don't enjoy — before viewers even see the show."[1]

– Joe Kukura, in February 2013.

"We've entered an era when product decisions can be made not by analysis of demographics or user testing but by extremely fine-tuned measurements of current user activities designed to predict what new activities they will be eager to engage in."[2]

– Damian Rollison, Vice President of Product and Technology, Universal Business Listing[a], in February 2013.

"My worry about BD (big data) was that it will steal spontaneity from our lives by figuring out what we like before we've even had a chance to decide."[3]

– Mike Cassidy, in February 2013.

Netflix, a company which operated in the DVD-by-rental segment and in the video streaming segment, had, since its inception, garnered significant insights into the viewing patterns of its customers. It knew what shows they preferred viewing, the time at which they watched the shows, the devices on which they streamed the videos, and other such viewing nuances. According to observers, the company's big data capabilities had received a shot in the arm from the cloud computing

[a] Universal Business Listing is a company focused on promoting online visibility for local businesses.

infrastructure it sourced from Amazon Web Services. Netflix had initially employed this data to recommend movies and TV shows to its subscribers. However, with increased competition for broadcast rights of movies and TV shows which had their initial run and the reluctance on the part of some potential competitors to equip Netflix with movie and TV program ammunition, Netflix plunged into original programming. The company clinched the exclusive rights to the first and second seasons of the show "House of Cards" for a staggering US$100 million. According to industry observers, the company bid this amount after confirming that many of its subscribers were fans of the director, David Fincher, and the lead actor of the show, Kevin Spacey, and had been followers of the original "House of Cards" show[b]. The show proved to be a roaring success, but the most incredible aspect, according to observers, was that Netflix executives were sure that it would succeed even before the first scene was shot.

However, some industry observers were concerned that Netflix would find itself on the wrong side of the law with privacy concerns being raised by some observers over the employment of big data capabilities to gauge the personal viewing habits of its subscribers. Others were apprehensive about the potential damage to Netflix's reputation due to its site being down on account of outages in Amazon Web Services. Analysts were also concerned about Netflix stretching itself financially to create a credible image for itself through original programming. Some industry observers were apprehensive that as companies such as Netflix, which had greater awareness about the audience's viewing habits, gained more prominence in the generation of original programming, they might begin influencing the creative decisions of even directors and writers.[4]

Background Note

Netflix was started in 1997. In 1999, the company introduced a DVD subscription service, providing unlimited rentals on a monthly subscription basis. In 2000, it introduced the personalized movie recommendation mechanism that employed subscribers' ratings to forecast the preferences of all Netflix subscribers.[5] Toward the close of 2007, Netflix was shipping around 1.6 million DVDs on a daily basis.[6] In 2007, the company also launched its video streaming services in the

[b] House of Cards was based on a 1989 novel by British writer Michael Dobbs. It revolved around one politician's ruthless methods to clinch power and handicap his rivals. The original "House of Cards" was a mini-series which was broadcast by the BBC in 1990.

US, and, in 2010, began the global expansion of this business. As of March 2013, Netflix, Inc. functioned in three categories: Domestic Streaming, International Streaming, and Domestic DVD.

The Domestic DVD business provided DVDs-by-mail subscription services in the US. Under the streaming services, Netflix provided an Internet television network service that allowed subscribers to stream TV programs and films directly on TVs, computers, and mobile gadgets in the US and globally.[7] As of December 31, 2012, Netflix had 27.15 million subscribers for its streaming service in USA. Subscribers in other parts of the world totaled 6.12 million. The subscribers to Netflix's mail order business totaled 8.22 million.[8] As of November 2012, Netflix accounted for 33% of the peak period video streaming traffic in North America.[9] As of March 2013, the company's subscribers in nearly 40 nations were viewing close to 1 billion hours of TV shows and movies on a monthly basis.[10] As of March 2013, Netflix was charging US$ 7.99 per month from each of its video streaming subscribers in USA. For the year ended December 31, 2012, Netflix had generated revenues of US$3,609.3 million and net income of US$17.1 million *(Refer to Exhibit I for Netflix's financial performance from 2005 to 2012).*

Big Data at Netflix

Netflix heralded the "big data" approach to Internet media, wherein each likely metric of experience was recorded, assessed, and inventoried forever *(Refer to Exhibit II for a brief note on big data).*[11]

Tracking Subscribers' Viewing Patterns

Netflix tracked each search that a viewer made, each good or bad rating attributed by a viewer to what s/he had just seen, besides the ratings data from third-party providers such as Nielsen. These were in addition to the location data, device data, remarks on the social media, etc. Netflix was also aware of what a subscriber most probably viewed on each of his/her devices like mobile or laptop or tablet and what viewers belonging to a specific ZIP code preferred viewing on their laptops on a Saturday night. It even kept a tab on the number of subscribers who switched off before the credits began to roll.[12] The company also captured how a subscriber watched programs, including the number of times s/he paused the show, whether s/he was a binge viewer, whether s/he often exited the show to view others, and, whether s/he left shows incomplete.[13] Explaining how Netflix's ability to gauge viewer preferences had evolved over time, Ted Sarandos (Sarandos), Netflix's Chief Content Officer, stated, "Here is what the data from our

DVD business tells us: we know what we shipped to you and we know when you returned it. I have no idea if you watched it. I have no idea if you watched it 20 times. With streaming, we have insight into every second of the viewing experience. I know what you have tried and what you have turned off. I know at what point you turned it off."[14]

In gauging the preferences of its subscribers, Netflix also took into account their ratings, the queuing up of movies/ shows by subscribers, and other viewer-driven actions. It combined these findings with the type of content viewed by them, as for instance, action or comedy.[15] While the significance of isolated incidents such as taking a bathroom break, sleep hours, etc., could be minimized, Netflix was focused on the trends that began to become visible over time. According to an information technology expert, this level of garnering data bestowed upon Netflix a level of insight into its subscribers' needs and preferences that none of the television or studio networks could ever tap into (but which they all wanted, as was evident from the innumerable surveys and viewer polls that Hollywood[c] had administered over the years).[16]

Netflix's data about its subscribers had increased at least 10 times from 2007 till 2013, as it focused more on the Internet streaming business than on its DVD rental business.[17] As noted by an information technology expert, Netflix's strategy ultimately boiled down to "capture data about 'What' you offer your users, and 'How' they interact with it, then you can infer 'Why' things work or don't work."[18]

Inventory of Data and Algorithm

In the last quarter of 2011, Netflix's streaming service subscribers had seen over 2 billion hours of streaming video. As of June 2012, Netflix had over 25 million subscribers to its streaming service. Given this subscriber base and usage, Netflix captured, among others, around 30 million user actions on a daily basis such as rewinds, fast forwards, and pauses, 4 million ratings on a daily basis, and, searches numbering 3 million on a daily basis.[19]

As of July 2011, Netflix was storing 50 distinct files for each viewable item, including 3 copies of each film, viewer ratings for ten years, comprehensive viewer account information, and metadata comprising intricate licensing rights for everything, audio files, log files,

[c] The term Hollywood is used to refer to the American film industry and the lifestyles of the individuals connected to it.

subtitles, etc. Netflix had globally placed cloud data centers for minimum latency[d], verification data for subscribers, data about the test cell[e] that every viewer was in, and the gadget that each account activated, bookmarks[f], personalization, and category preferences.[20]

Netflix compiled all of its subscribers' viewing patterns and took all of its customers' viewing habits and show ratings and applied a sophisticated algorithm to them.[21] A Netflix blog post dated April 6, 2012, stated, "Streaming has not only changed the way our members interact with the service, but also the type of data available to use in our algorithms."[22] The algorithm employed by Netflix derived data from multiple sources and factored in different behaviors. The star rating given by viewers was an important source. If a subscriber rated a show without streaming it as s/he had viewed it in a theater or on DVD or s/he rated it casually, the algorithm allotted a lower weight to the rating. In case a subscriber saw only a part of the show and rated it, the weight given to the rating got proportionally reduced.[23]

Personalized Recommendations

Netflix gave viewing suggestions by employing algorithms that came up with apt content before the appropriate viewer with the help of data such as their prior viewing patterns and also taking into account the viewing behaviors and recommendations of other subscribers.[24] As of April 2012, three-fourth of the shows viewed by subscribers resulted from some kind of recommendation.[25] The company was reputed for its recommendation and targeting technologies that helped it dish out personalized content to the audience that had the highest probability of viewing it. If a subscriber of Netflix's streaming service wanted to watch a movie, the company employed its mammoth product-attribute datasets and the ratings of viewers numbering millions to narrow down on the appropriate movie. By assessing data regarding the movies that subscribers ordered for, it could suggest the films that they could watch subsequently. Subscribers liked it as they got recommendations about

[d] Latency implies any of various types of delays usually involved in processing of network data. (Source: Bradley Mitchell, "Network Bandwidth and Latency," http://compnetworking.about.com/od/speedtests/a/ network_latency.htm)

[e] A test cell is a group of real and virtual machines with particular hardware and software configurations that comply with the requirements of a particular test environment. Teams or individual testers can reserve the cell to conduct tests. (Source: Pietro Marella, "Extend Rational Quality Manager to Manage Lab Assets in the Cloud," www.ibm.com/developerworks/rational/library/quality-manager-manage-lab-assets-cloud/index.html)

[f] Bookmark is the address of a web page that is saved on the computing device so that the user can access it again easily.

movies they would potentially be interested in. The benefit for Netflix was that subscribers ordered for more movies.[26] Explained Sarandos, "We've made really incredible strides to predict what people watch right after they finish something else. Basically, it's a statistical push based on what other people have watched and really enjoyed immediately after viewing the same program you just finished."[27]

The homepages of Netflix for each of its subscribers were personalized to the maximum extent possible. The homepage comprised groups of videos arrayed in horizontal rows. Every row had a label that indicated the intended meaningful link among the videos in that cluster. A major portion of Netflix's personalization boiled down to the manner the company chose the rows, the way it arrived at the videos to be a part of the rows, and upon the sequencing of the videos. A common basis for personalization was the compilation of rows based on genres like comedy or drama. Every row comprised three tiers of personalization: the selection of the genre itself, the sub-category of videos included in that genre, and the sequencing of those videos. The basis for arriving at the genre preferences of a subscriber were the recent videos s/he had viewed, his/ her ratings, and his/ her actions such as stopping the show midway, or direct feedback garnered through viewing preferences survey. Similarity was also a key basis for personalization in Netflix delivering its service. Similarity was perceived in different ways, for instance, in searching or including videos in the queue.[28]

The basic ranking model employed by Netflix for each individual balanced his/ her predicted rating and the show/ movie's popularity *(Refer to Exhibit III for the basic ranking model employed by the company).* Netflix further fine-tuned its findings by employing A/B testing[g] *(Refer to Exhibit IV for visual representation of the testing which was carried out).*[29] According to experts, a key part of the strategy that Netflix employed to provide a rich experience to its viewers in terms of user interactions and recommendations was the software architecture that it had put in place *(Refer to Exhibit V for Netflix's software architecture).*

[g] A/B testing is a simple technique of testing alterations to a webpage against the present design and finding out which of the alterations generate positive results. It is a technique of verifying that any new outlay or alteration to an element on a webpage is enhancing the conversion rate prior to making the alteration to the site code. An A/B test comprises testing two versions of a web page — an A version (the control) and a B version (the variation) — with real-time traffic and assessing the impact each version has on the conversion rate. (Source: "What is A/B Testing?," https://www.optimizely.com/ab-testing)

Todd Yellin, Vice President of Product Innovation at Netflix, detailed the intricate process by which Netflix arrived at the suggestions for its Watch Instantly menu. According to him, "One of the challenges of big data is we'll have all these terabytes flowing at us, and we have to figure out which data to pick."[30] This, according to him, implied fragmenting the data further into, for instance, what a specific viewer watched the previous night and what s/he watched the previous year. He pointed out that what a subscriber viewed in the recent past was of more value. Then Netflix combined it with other information such as ratings, if any, ascribed by the viewer at any point in time and also the likely demographics of the family constituting the account. In the summer of 2012, Netflix also introduced a "post play" feature wherein after one of the episodes of a show ended and credits rolled, a pop-up window suggested to the subscribers that they view the subsequent episode. The timing of the pop-up was carefully adjusted by an algorithm which worked out when a viewer would most probably switch off. If a viewer did not engage in any actions, the subsequent episode began automatically.[31]

Shift from Own Data Centers to Amazon Web Services

According to observers, Netflix's big data capabilities received a boost with the gradual shift, beginning in 2009, of the company's computing infrastructure from its own data centers to the cloud[h], by engaging the services of Amazon Web Services (AWS), the cloud computing subsidiary of Amazon[i]. Netflix's search and recommendation engine and its streaming servers started utilizing AWS's cloud infrastructure.[32,33] The principal reason for Netflix's shift was that its pace of setting up data centers could not catch up with the increase in demand for its streaming services.[34] The convenience of scaling up information technology infrastructure on cloud, payment for only the computing resources used, and allowing internal IT staff to concentrate on operations of greater significance than just ensuring that servers were working fine were other reasons for Netflix's tapping the services of AWS.

[h] Cloud computing is an Internet technology that enables one to accomplish one's work online without having to set up hardware or software and without storing data at one's end. A user can access any of his/ her applications through any device having an Internet browser and online connectivity.

[i] Amazon.com, Inc. is an American multinational e-commerce company. It is an important player in the business of providing cloud computing services as well.

Netflix employed Amazon Elastic MapReduce[j] to assess streaming sessions and derive business metrics regarding performance, viewing patterns, etc., which allowed it to continue to enhance the quality of streaming.[35] Hadoop's[k] processing capabilities enabled Netflix to conduct huge data analyses like creating visual representations of traffic patterns for each kind of device traversing several markets. The increased processing competencies bestowed by Hadoop also enabled Netflix's engineers to find out where traffic on the network was slowing down, enabling it to order for extra network capacity *(Refer to Exhibit VI for Netflix's cloud-based Hadoop architecture).* The technology—which could process bigger data sets — also aided Netflix to gauge customer choices better so that it could come out with better recommendations.[36] As of November 2012, AWS was managing 95% of the computation and storage requirements of Netflix.[37]

The Need for Original Content at Netflix

Prior to March 2011, Netflix had had no plans of becoming the first avenue where a TV show or a movie was beamed. Rather, it concentrated on shows and movies having their second, third, or even fourth runs. In other words, Netflix dealt with content subsequent to its release either in theaters or on television. And occasionally, Netflix could lay its hands on content after it played in theaters and subsequently on television for quite a while.[38] In July 2011, a prominent securities company stated that Netflix required more streaming content to appeal to customers, and that only around one-fifth of Netflix movies could be streamed online. At that time, a financial analyst had observed, "The bigger Netflix gets, the more content deals they can do. But they'll never be online what they are in DVDs, because of rights and distribution windows…"[39]

The prime programming which was critical to Netflix's streaming service was becoming difficult to access.[40] Content owners raised the fees they charged from Netflix and in some cases cancelled their licensing

[j] Amazon Elastic MapReduce is a web service employing which a user can immediately provision as much or as little capacity as he/ she requires to perform data-intensive operations for applications like web indexing, data mining, log file analysis, data warehousing, machine learning, financial analysis, scientific simulation, and bioinformatics research. It is not necessary to worry about issues such as time-consuming setup and handling. (Source: http://aws.amazon.com/elasticmapreduce/)

[k] Hadoop or Apache Hadoop is an open-source software framework that helped the processing of large data sets in a distributed computing environment. Amazon Elastic MapReduce made use of the Hadoop framework.

arrangements with Netflix.[41] HBO, a creator of well-known television shows, declined to license any of its shows to Netflix. Comcast, a big cable company with a strong portfolio of sports shows, also showed no inclination to join hands with Netflix. Showtime did not license its shows before the passage of a couple of seasons. Industry observers expected such tactics to leave Netflix with the option of only streaming old content. Recalled Sarandos, "I looked at that and realized we were faced with a supply source that wasn't reliable."[42] In the meantime, Netflix was also facing increased rivalry. Hulu[l], HBO, Amazon.com, Comcast, Apple[m], Google[n], and many start-ups were creating competing businesses. Analysts expected these companies to clash with Netflix for content and also subscribers; they also expected them to vie with Netflix in auctions for exclusive rights to television shows and movies.[43,44] Hence, Netflix's CEO, Reed Hastings, decided that the company would finance original shows on its own. Also, by creating original content, Netflix expected to enjoy the type of credibility generally associated with an AMC[o] or with HBO. Explained Sarandos, "If they like it they watch more. If they watch more, they will value the service more."[45]

Enter the 'House of Cards'

Identifying the Show

Netflix made viewers label films and TV shows with hundreds of metadata[p] descriptors[q] to comment, among others, about the actors, the screenplay, the overall appeal, and the category. Previously, these labels

l Hulu is a website and over-the-top (OTT) subscription service providing ad-supported on-demand streaming video of TV shows, films, webisodes and other new media, trailers, clips, and behind-the-scenes footage from networks and studios. (Source: Daniel Ferry, "A Bear Gives a Bull Case for Netflix," www.fool.com, February 24, 2013)

m Apple Inc., is an American multinational company that designs, develops, and sells consumer electronics, computer software, and personal computers. For the year ended September 29, 2012, Apple had generated net sales of US$156,508 million and net income of US$41,733 million.

n Google Inc. is an American multinational company specializing in Internet-related services and products. These comprise search, cloud computing, software, and online advertising technologies. For the year ended December 31, 2012, Google had generated total revenues of US$50,175 million and net income of US$10,737 million.

o AMC is an American cable and satellite television channel that is owned by AMC Networks. The channel basically broadcasts theatrically released movies, coupled with a limited number of original shows.

p Metadata represents a set of data that describes and provides information about other data.

q A descriptor is an item of stored data that points out how other data is stored.

were employed to suggest other shows available on the service, fundamentally constructing outlines of individual viewers according to their choices. However, Netflix began gradually ordering original content as it was aware of what subscribers desired prior to their being in the know.[46]

Netflix, akin to other networks, desired a series of hits. The challenge for it was never to have access to viewers' homes — its streaming service could be accessed on any gadget that could be connected to television. The challenge was to create content that viewers were keen to watch and would not have access to elsewhere. Rather than create lots of content and wishing that a part of it would be lapped up by viewers, it utilized its huge data repository of 29 million subscribers' viewer choices and patterns.[47,48] When the show 'House of Cards', a remake of a BBC mini-series, was auctioned in 2011 with the combination of David Fincher and Kevin Spacey, Netflix executives mined their mammoth data repository. The company's assessment of viewer data helped it realize that a remake of the famous BBC show on political intrigue could deliver the goods. Taking into account the content and the team involved, Netflix was confident about the potential viewership being substantial enough for the show to be a success.[49,50] Unlike other networks which conventionally approved a show subject to their being satisfied with a trial episode, Netflix approved two complete seasons (26 episodes) of House of Cards without viewing even one scene.[51] It gained exclusive rights for the two seasons reportedly for a price tag of over US$100 million.[52]

Netflix also employed data to target different viewers with different trailers. The total number of distinct trailers numbered ten. The trailer beamed to a viewer was based on his/ her viewing history. If a subscriber viewed action movies more, s/he was beamed a racy, energetic version. If a subscriber viewed more of romances or dramas, s/he was streamed a trailer which emphasized characters.[53]

Netflix started making the program 'House of Cards' available on February 1, 2013, to the 33 million subscribers of its streaming service globally.[54] 'House of Cards' was the first show wherein the TV program rolled out the complete season comprising thirteen episodes at one time for subscribers to view as per their convenience.[55]

Results

Netflix did not come out with data regarding either the number of its subscribers who viewed House of Cards or viewer ratings of the show as it did not follow the business model of selling advertising slots.[56,57]

However, based on circumstantial evidence, industry observers proclaimed 'House of Cards' to be a hit. Over 10,000 comments were posted on social media within one day of 'House of Cards' debuting. According to social media analysts at the media research firm, Fizziology, in the first 24 hours of its debut, *House of Cards* figured thousands of times per hour in social media chats. Fizziology also observed that 62 percent of these comments were positive.[58] House of Cards garnered outstanding reviews with a total of 9 on IMDb.com[r].[59]

Unmetric, a social media benchmarking company, stated that the trailer of House of Cards was the most viewed video on Netflix's YouTube Channel accounting for over 1.2 million views – one million of these views were garnered from January 16 to February 8. According to experts, this program was largely responsible for Netflix's enhanced footprint on the social media. They also opined that by attracting more viewers to Netflix's YouTube Channel, House of Cards was considerably expanding the company's marketing avenues. Experts expected the show to aid Netflix in increasing the number of its subscribers as it created additional content. According to Trendrr, a company focused on deciphering social media data, 63% of the comments within the first three days of debut of House of Cards i.e. from February 1 to February 3, was positive.[60] On February 4, i.e. on the fourth day of the release of House of Cards, wired.com conducted a poll of its readers to view how many among them had viewed House of Cards. Of the 1,900 readers who participated in the poll, over 26% confirmed that they had watched all the 13 episodes.[61]

On February 12, 2013, Sarandos declared that House of Cards was the most-viewed show, both in terms of number of distinct viewers and overall hours streamed, in each of the countries where Netflix offered its streaming services.[62] He also expected the number of viewers of the show to increase with the passage of time. According to him, Netflix would expand its portfolio of original shows as it would give it bargaining power when negotiating with studios on getting exclusive access to original shows. According to him, this did not, however, mean an increase in Netflix's subscription fees given the company's subscriber volumes.[63] According to an investment analyst, House of Cards bestowed upon Netflix the credibility that it was capable of producing original content on its own. The company would have a receptive audience for any of the original shows that it came up with.[64]

[r] IMDb or Internet Movie Database (IMDb) is a popular online database of information regarding movies, television shows, and video games.

The Reasons behind the Show's Success

According to observers, the success of House of Cards was not a freak incident. To collect real-time data about the programs its subscribers were watching, their demographics and viewing patterns, Netflix used the NoSQL database Apache Cassandra[s].[65] Hence, Netflix executives were aware of precisely what their customers numbering millions were viewing; they knew exactly how famous the creations of David Fincher were, and the number of their subscribers who were Kevin Spacey's fans, and the number of their subscribers who had watched the original British version of the famous and critically acclaimed House of Cards on the streaming service.

Netflix's data signaled that Spacey or Fincher followers were also fans of House of Cards, broadcast by the BBC in 1990. Scrutinizing the company's data trove, Netflix managers were able to forecast that a House of Cards version starring Spacey and directed by Fincher was the one that Netflix viewers would be keen on viewing.[66,67] Explained Steve Swasey (Swasey), Netflix's VP of Corporate Communications as of 2011, "We have a high degree of confidence in [House of Cards] based on the director, the producer, and the stars. We can look at consumer data and see what the appeal is for the director, for the stars, and for similar dramas."[68]

Industry observers too appreciated Netflix for employing data about viewers' preferences to target them. The company had employed its viewers' data and recommendations mechanism to channelize viewers to the show, instead of devoting resources to promote House of Cards via online and TV ads on other websites. Swasey added, "Through our algorithms we can determine who might be interested in Kevin Spacey or political drama and say to them, 'You might want to watch this.'"[69] An analyst noted that a big banner advertisement that popped up every time he switched on Netflix and which drew his attention to the next episode of House of Cards was, to a certain extent, responsible for his binge-viewing *House of Cards*.[70]

How Netflix Benefited from Big Data

According to industry observers, instead of creating shows based on viewers' response to trial episodes akin to the route taken by conventional stations, the video-streaming business enabled Netflix to assess the viewing patterns and choices of its subscribers. It was aware

[s] Apache Cassandra is an open source distributed management system.

of what the subscribers viewed, the time at which they viewed the shows, and even when they paused to take a recess. By incorporating the data into an algorithm, it could forecast their tastes and what would keep them hooked.[71]

Observers opined that Netflix had sufficient information about the viewing choices and patterns of its subscribers. So it chose movies and shows suitably. Explaining the advantage of Netflix basing its decisions on data, Sarandos commented, "You get a very addressable audience. Better than that, I know exactly who they are."[72] Netflix executives claimed that the data they had about the number of subscribers who viewed a specific show and the number of times they viewed it helped them in arriving at the amount they would pay to license it. Explained Sarandos, "Traditionally, content is worth what the buyer says it is worth, but our data draws from viewer behavior to bring a bit more science into the calculation."[73]

The company was also employing the entire big data to generate content that subscribers desired to view deploying its analytics to devise the original videos that it was financing.[74] Explained one of the experts, "It also allows them to bid on and acquire scripts and content with insights that others may not have…"[75] Jonathan Friedland, Netflix's communications head, explained, "We know what people watch on Netflix and we're able with a high degree of confidence to understand how big a likely audience is for a given show based on people's viewing habits. We want to continue to have something for everybody. But as time goes on, we get better at selecting what that something for everybody is that gets high engagement."[76]

According to industry observers, the conventional studios were at a disadvantage as they clearly had no feedback mechanisms with the movie watchers except for data regarding tickets sold and DVDs ordered for. They were concerned with studios frittering away billions of dollars financing shows which did not appeal to viewers. In 2011, two-thirds of the 43 new shows financed by the large broadcast networks were withdrawn after one season, fundamentally due to poor ratings.[77] A reputed economist's research established that merely 22% of Hollywood movies proved to be profitable.[78] An industry observer opined that the broadcast networks had been following the wrong strategy by spending financial resources running into millions of dollars which could have gone into creating better content. Netflix, on the other hand, did not have to expend huge resources marketing its shows. It was aware of who would like its shows, and its content delivery infrastructure could direct the targeted viewers to the concerned show. An industry insider explained, "They're also not spending $40 million a show on a marketing campaign. They have a guy in a room who writes an

algorithm."[79,80] Analysts pointed out that, in future, if Netflix were to be sure that there would be audience for a show before giving it the go-ahead, it could drastically decrease the proportion of content that would prove to be a dud. Explaining the advantage that Netflix potentially had in creating original programming, given its treasure trove of big data, an expert observed, "They'll know when to launch a drama or a comedy or a show aimed at teens (or tweens, or women, or…) based on data, not the hunches of some network programming head…"[81]

Looking Ahead

Netflix was to debut four more original shows in 2013 including the shows "Hemlock Grove" and "Arrested Development".[82] The latter would be a remake of 'Arrested Development', a popular television show.[83] With Netflix tasting success with House of Cards, Amazon and Microsoft were quick off-the-block in announcing their entry into the creation of original content.[84] Amazon was also funding, on a trial basis, six comedy shows.[85] Microsoft which had previously started tying up programming deals for its Xbox Live platform, revealed that it had purchased exclusive rights to an Indie film.[86] It was reportedly making a big foray into original content creation.[87] But some industry observers also expressed concern about Netflix stretching itself financially, given the huge resources it had to spend on original programming *(Refer to Exhibit VII for concerns expressed in this regard).*

Some experts were concerned about Netflix's dependence on AWS for its mission-critical operations given that during 2011 and 2012, AWS faced outage once and twice respectively. On all these occasions, Netflix's streaming service was down. Experts considered the last Netflix outage which occurred on Christmas Eve to be the most damaging for the company as it happened at a time when many Americans spent relaxed hours watching television along with their families. Many had then opted for Netflix's competitors to access video streaming services.[88]

Some observers were also concerned that Netflix, which was intimately aware of the viewing patterns of its subscribers, would get into trouble if it did not carefully handle the data and hence the privacy of its customers. To substantiate their concern, they pointed out to Netflix's previous brush with the law. In 2006, Netflix launched the Netflix Prize, a competition inviting developers to create algorithms that recommended movies Netflix subscribers might want to watch. However, Netflix cancelled the competition in 2010 after some of its customers took the company to court for infringing on their privacy. The plaintiffs claimed that the anonymous user IDs, film titles, and ratings employed for the competition, when tallied against publicly accessible

data, revealed their identities.[89] According to experts, the greater access to personal data that big data frequently entailed would focus attention on the tension between privacy and convenience.[90]

Some industry observers also foresaw a potential conflict of interest in Netflix's deployment of big data to create original programming as the company's streaming traffic was majorly driven by its recommendations of shows and movies to its viewers. They hinted at a potential controversy emerging if Netflix's portfolio of original programming got bigger and the company's algorithms diverted traffic toward its shows.[91] Also, the dependence of companies such as Netflix on big data to create original programming generated concerns over creativity. Questioned an industry observer, "Will screenplays some day be written to meet the whims of data-driven media streaming companies?"[92] Lamented another industry observer, "If Netflix perfects the job of giving us exactly what we want, when and how will we be exposed to things that are new and different, the movies and TV shows we would never imagine we might like unless given the chance? Can the auteur survive in an age when computer algorithms are the ultimate focus group?"[93]

Analysts expected the use of big data to grow immensely in the coming years. According to Matt Pfeil, co-founder and VP of Customer Solutions at big data software company DataStax which worked with Netflix to implement Apache Cassandra, "If you talk about this age as the data age, we're still in the teenage years, and as it matures there's going to be orders or magnitudes of different types of technologies that just encompass big data. The more data you have and the more you can do with it, the smarter this business decision."[94] However, others were concerned about the limitations of using big data. According to an industry observer, "Data can only tell you what people have liked before, not what they don't know they are going to like in the future."[95]

Exhibit I

Netflix's Financial Performance (2005-2012)

Category	2005	2006	2007	2008	2009	2010	2011	2012
Revenues	682.2	996.7	1,205.3	1,364.7	1,670.3	2,162.6	3,204.6	3,609.3
Operating Income	2.9	65.2	91.8	121.5	191.9	283.6	376.1	49.9
Net Income	42	48.9	66.7	83	115.9	160.8	226.1	17.1
Earnings per Share (Diluted)	0.64	0.71	0.97	1.32	1.98	2.96	4.16	0.29
Net Cash Provided by Operating Activities	157.5	248.2	277.4	284	325.1	276.4	317.7	22.8
Total Liabilities	138.4	194.6	249.2	268.3	480.6	691.9	2,426.4	3,223.2
Total Shareholders' Equity	226.3	414.2	429.8	347.2	199.1	290.2	642.8	744.7

** Million of US$, except earnings per share (diluted)*

Source: http://ir.netflix.com

Exhibit II

What is Big Data?

The term big data referred to the phenomenon of advancing trends in technology that enabled a fresh approach to assessing the world and arriving at decisions. It represented an approach wherein decision-making was largely dependent upon data and analysis instead of experience and intuition.[96] Some experts also employed the term 'big data' to describe datasets whose size was beyond the capability of normal database software devices to capture, store, handle, and assess.[97] The majority of the big data emanated from unstructured content such as words, images, and Internet video and the surge of sensor data. The improvements in the field of big data were largely due to the advances in the field of computing and artificial intelligence such as natural-language processing, pattern identification, and machine learning.[98]

Big data was expected to generate value in five fundamental ways. First, it was expected to make information transparent and to employ the same more often. Second, as organizations generated and saved greater transactional information in digital mode, they were expected to garner more precise and finer performance data on all aspects from inventories to sick leave, and hence gauge variability and bolster performance. Third, big data enabled greater fragmentation of customers and hence delivery of more acutely customized products/ services. Fourth, sophisticated analytics could considerably enhance decision making, mitigate risks, and discover very important insights that would otherwise have been unknown. Fifth, big data was expected to allow companies to evolve new products and services, improve existing ones, and invent completely new business models.[99]

According to experts, big data could be employed to increase efficiencies in diverse industries. For instance, a next-generation retailer could monitor the behavior of individual customers from their Internet click patterns, prepare a database of their choices, and forecast their probable behavior in real time. It would then be in a position to identify when customers were closing in on a purchase decision and prod the sealing of the deal by bundling preferred products, combined with loyalty program savings. This real-time targeting, which would also employ data from the retailer's multi-layered membership rewards scheme, would enhance purchases of merchandise with greater margins by its most valuable patrons.[100] An insurance company ran though the figures about the finer details of customer behavior which enabled it to price risk better.[101] Some manufacturers employed algorithms to assess sensor data related to production lines, developing self-regulating processes that eliminated waste, prevented expensive (and occasionally dangerous) human interventions, and finally increased output.[102]

However, before going full steam ahead with the implementation of big data technologies, an organization had to have a well-laid out plan in place. This plan had to detail the process by which the data lying hidden in different business units or divisions was mined and integrated, the areas or business functions in which the analyzed data would be deployed, and the tools that would enable the front-line employees to assimilate the data into their daily operations so that the envisioned benefits were realized.[103] Every big data plan had to tackle three key challenges. First, the final benefits envisioned had to be prioritized. Second, the organizational

Contd…

Contd…

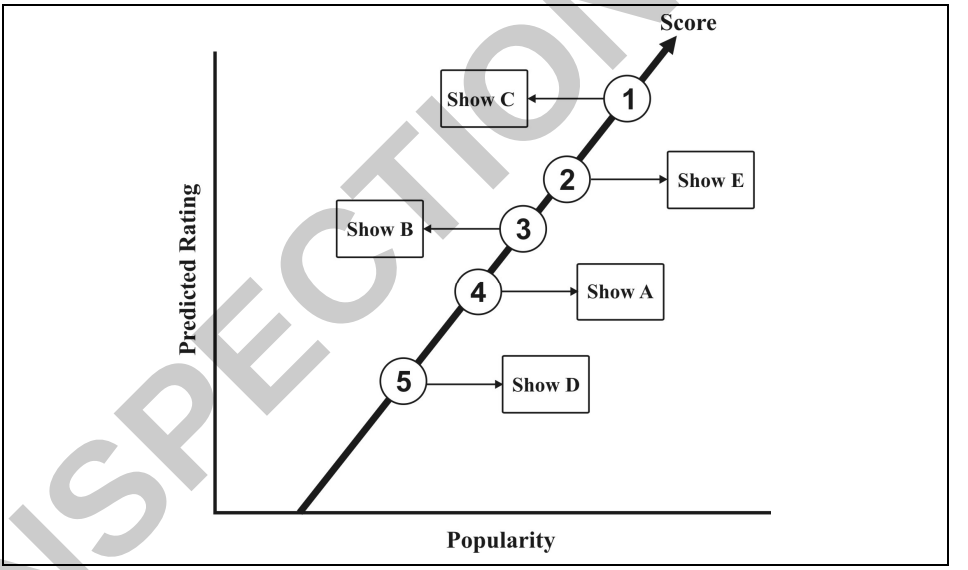
and the employee sensitivities had to be taken into account. For this, simpler analytical tools had to be employed in the initial stages where the benefits derived were modest but which would be required to boost the confidence of the employees or to take the organization to the next level wherein more complex data would be mined and analyzed. Third, the skill-sets of employees who were expected to deploy the data had to be developed.[104]

According to observers, big data could be deployed to push the envelope in the field of health care. To illustrate this argument, they pointed out to the example of Google which had scrutinized the time and location of online searches on flu and was able to spot probable flu outbreaks two weeks before some national governments did.[105]

Compiled from various sources

Exhibit III

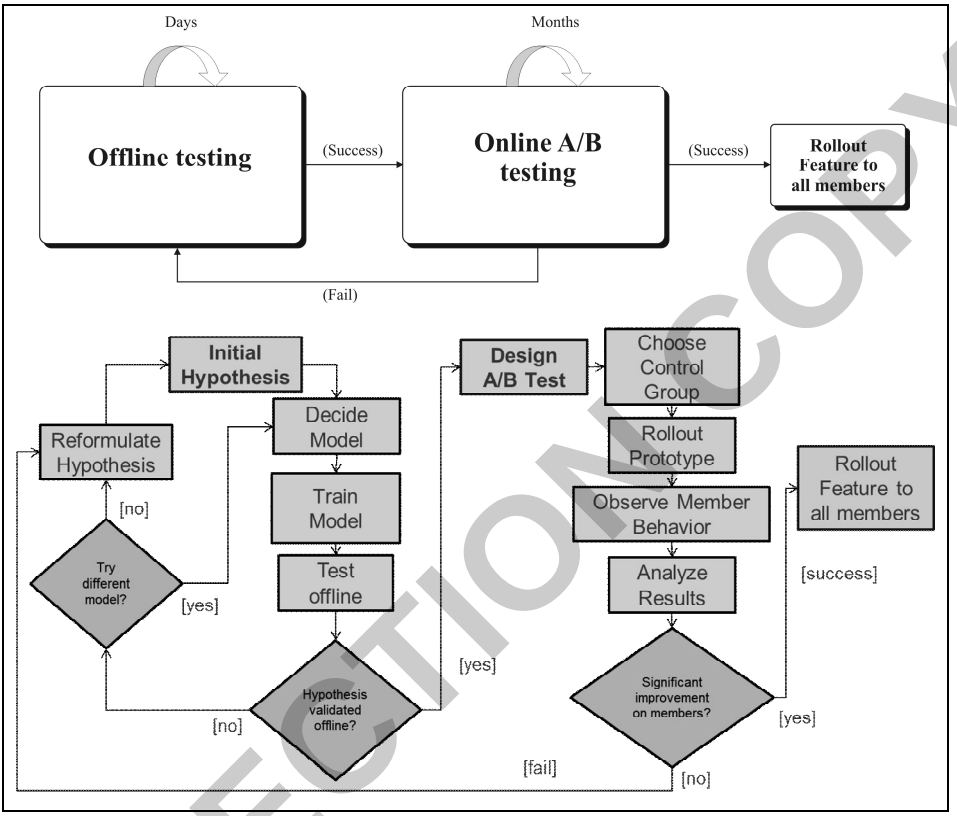
Netflix's Basic Ranking Model



Source: Xavier Amatriain and Justin Basilico, "Netflix Recommendations: Beyond the 5 Stars (Part 2)," http://techblog.netflix.com, June 20, 2012.

Exhibit IV

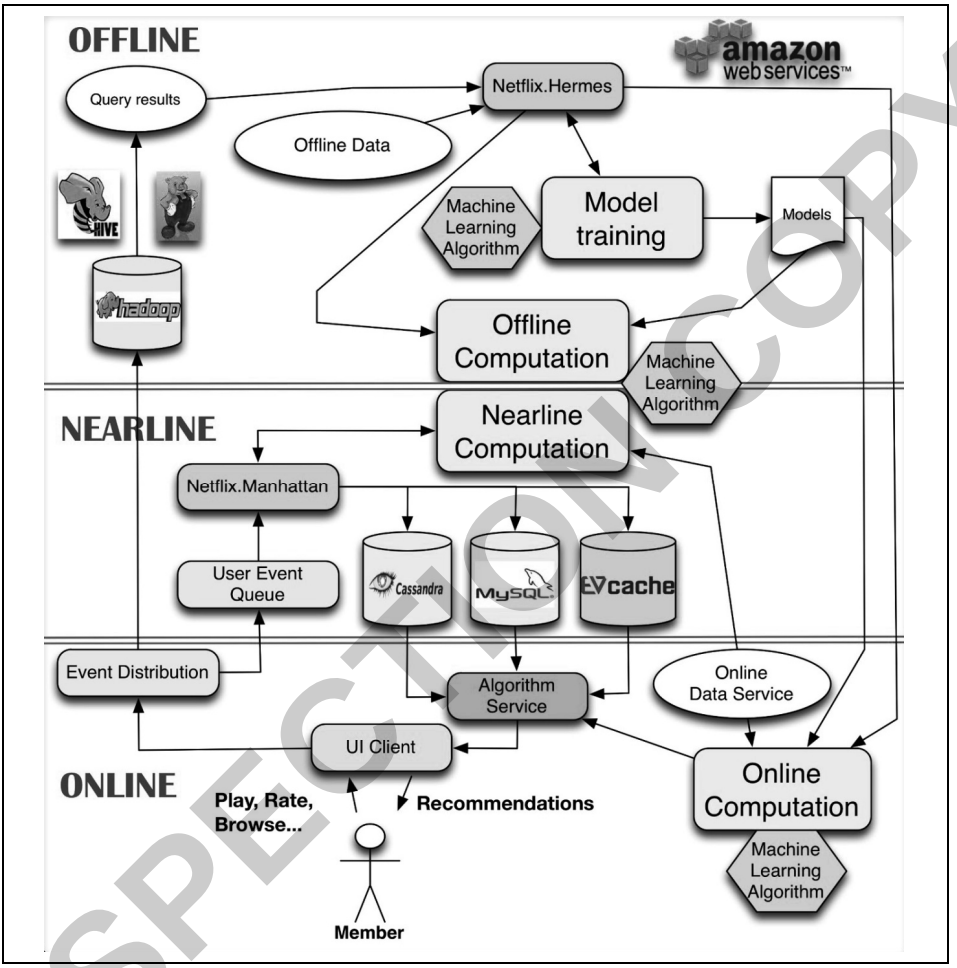
Visual Representation of the Testing Conducted by Netfllix



Source: Xavier Amatriain and Justin Basilico, "Netflix Recommendations: Beyond the 5 Stars (Part 2)," http://techblog.netflix.com, June 20, 2012.

Exhibit V

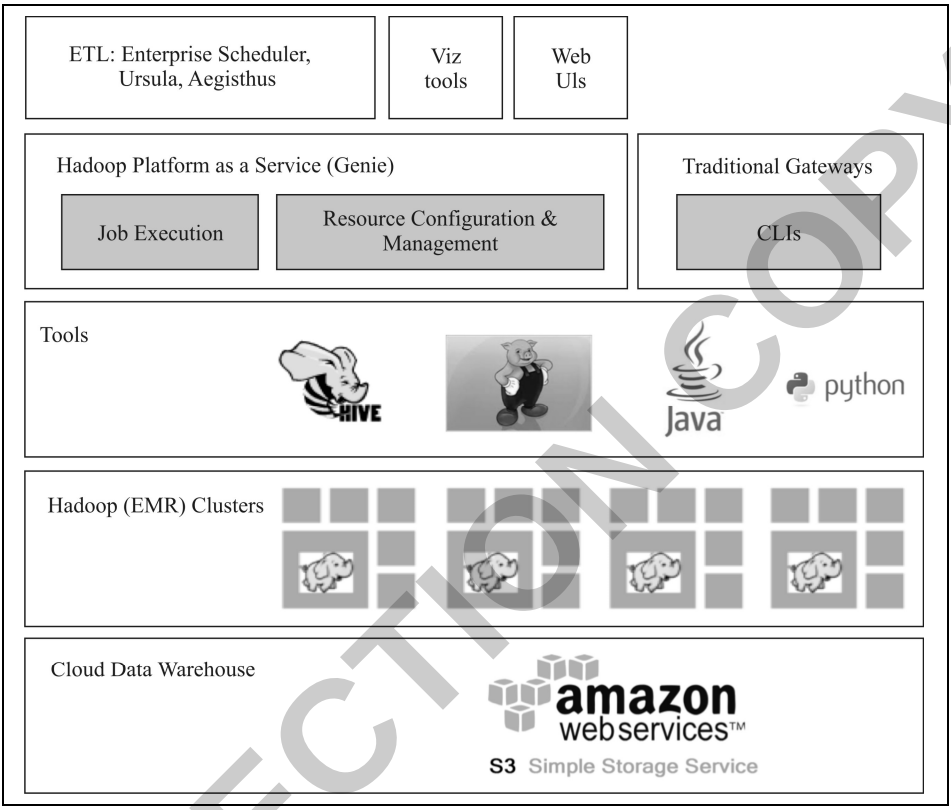
Netflix's Software Architecture



Source: Xavier Amatriain and Justin Basilico, "System Architectures for Personalization and Recommendation," http://techblog.netflix.com, March 27, 2013.

Exhibit VI

Netflix's Cloud-based Hadoop Architecture



Source: Sriram Krishnan and Eva Tse, "Hadoop Platform as a Service in the Cloud," http://techblog.netflix.com, January 10, 2013.

Exhibit VII

Will Netflix Survive the Original Programming Game?

Netflix was reported to have shelled out US$40 million for 13 episodes of 'Hemlock Grove' which was expected to make its debut post House of Cards.[106] For each episode of 'House of Cards' too, Netflix was reported to have coughed up US$3.8 million, around twice the cost generally incurred on television shows.[107] Industry observers were skeptical about Netflix's ability to sustain the success that its first foray into original content had produced. A media analyst pondered about the consequences if a show such as Arrested Development did not prove to be as much a success as Netflix expected. Netflix had a lot at stake as it was ordering a complete season or more at a single time. It would not be able to abort a show if it turned out to be a debacle, as other networks did.[108] Also, according to observers, the entry of new rivals such as Amazon Prime would make it tougher for Netflix to retain the same share of consumers' discretionary income.[109]

According to analysts, Netflix was shelling out more than its competitors to prevent them from accessing shows such as 'House of Cards' but this would mean that it would have fewer shows in its portfolio as its finances were limited. To maintain its leadership position, analysts opined, Netflix was compromising its near-term profitability to invest more in original content. On January 29, 2013, the company announced its plans to borrow $500 million to refinance around US$225 million in debt and put more financial resources behind original programming. The company had made commitments of over US$5 billion for the streaming content services. A financial analyst expressed concern that Netflix's content expenses were increasing more speedily than its profits.[110] Standard & Poor's also expressed concern over the rise in debt leverage for Netflix. It forecast insufficient cash flows for Netflix for 2013 and probably for the first six months of 2014 due to enhanced expenditure on original programming. S&P pointed out that original programming entailed greater upfront payments and the returns could not be predicted with certainty.[111] An investment firm forecast that Netflix would shell out US$5 billion content costs between 2013 and 2017, with US$4.5 billion of the expenditure taking place by 2015 itself.[112] Netflix itself estimated that it would be incurring an expenditure of US$2.1 billion on licensing content in 2013.[113] However, analysts were skeptical about whether Netflix's original programming play at a huge cost would actually generate sufficient returns. The company, for instance, required 520,834 more subscribers to avail of Netflix's streaming services for two years at a cost of US$7.99 to break even on its US$100 million investment on House of Cards.[114]

Analysts were also skeptical about Netflix's ability to compete with competitors such as Amazon and Microsoft in the original programming space, given their financial wherewithal (Refer to Table I for a comparison of some of the financial figures of Netflix with those of Amazon and Microsoft for the year ended December 31, 2012).

Contd…

Contd…

Table I

A Comparison of Some of the Financial Figures of Netflix with those of Amazon and Microsoft

Category	Netflix	Amazon	Microsoft
Revenues	3,609.3	61,093	73,723
Net Income (Loss)	17.1	(39)	16,978
Net Cash Provided by Operating Activities	22.8	4,180	31,626
Total Shareholders' Equity	744.7	8,192	66,363

* *million of US$, except earnings per share (diluted)*

Source: http://ir.netflix.com; http://phx.corporate-ir.net; www.micorsoft.com

Industry observers expected Amazon, particularly, to give Netflix's executives sleepless nights. Amazon was partly employing the considerable cash flows generated by its ecommerce business and its capital to provide competition to Netflix's streaming service through Amazon Prime Video. During the third quarter of 2012, Amazon clinched a licensing deal with EPIX enabling its users to stream movies such as The Avengers, Iron Man 2, and the Hunger Games, which also implied that Netflix would no longer have the exclusivity that it had enjoyed with the cable channel.[115] According to a cable executive who has spoken to Amazon, its Prime service had also started heavily employing data-driven approaches to decide its programming.[116] And hinting that Netflix might lose its first-mover advantage to players with deeper pockets, an industry observer remarked, "Maybe Amazon will go beyond its tentative investments and throw a hundred million at a different A-list series…"[117] Also, Hulu, one of Netflix's chief competitors, was jointly owned by Fox, ABC, and CBS.[118]

Compiled from various sources

End Notes

1 Joe Kukura, "How Big Data Stacked the Deck to Guarantee 'House of Cards' would be a Hit," www.allvoices.com, February 25, 2013.

2 Damian Rollison, "Big Data and Local Search: The Netflix Precedent," http://streetfightmag.com, February 7, 2013.

3 Mike Cassidy, "Netflix "House of Cards" Offers NYT Columnist a Peek at the Real "Must See" TV," www.siliconbeat.com, February 25, 2013.

4 Jessica Leber, ""House of Cards" and Our Future of Algorithmic Programming," www.technologyreview.com, February 26, 2013.

5 "A Brief History of the Company that Revolutionized Watching of Movies and TV Shows," https://signup.netflix.com/MediaCenter/Timeline.

6 Joanne L. Kaufman, "Post Office Drawback Cited in Dark Forecast for Netflix," www.nytimes.com, December 6, 2007.

7 http://investing.businessweek.com/research/stocks/snapshot/snapshot.asp?ticker=NFLX

8 Janko Roettgers, "Netflix Ends Year on a High Note, Calls House of Cards "Defining Moment for Internet TV"," http://paidcontent.org, January 23, 2013.

9 "Sandvine Global Report: Internet Data Usage up 120 Percent in North America," www.sandvine.com, November 7, 2012.

10 http://investing.businessweek.com/research/stocks/snapshot/snapshot_article.asp?ticker= NFLX

11 Christopher Williams, "House of Cards: Netflix's Experiment in Binging Pays off," www.telegraph.co.uk, February 8, 2013.

12 Andrew Leonard, "How Netflix is Turning Viewers into Puppets," www.salon.com, February 1, 2013.

13 Michael Nixon, "House of Cards: Dealing with Big Data," www.business2community.com, February 6, 2013.

14 "Ted Sarandos – Audience Tastes," www.carseywolf.ucsb.edu, June 2012.

15 Adam Huttler, "Is House of Cards the Future of Cultural Programming?," www.huffingtonpost.com, February 4, 2013.

16 Justin Sulhoff, "Netflix Uses Big Data to Produce Original Content," https://megaplanit.com, February 13, 2013.

17 Jessica Leber, ""House of Cards" and Our Future of Algorithmic Programming," www.technologyreview.com, February 26, 2013.

18 Michael Cutler, "What Netflix Knows about you and why it's a Lesson to others...," http://cotdp.com, June 19, 2012.

19 Derrick Harris, "Netflix Analyzes a Lot of Data about your Viewing Habits," http://gigaom.com, June 14, 2012.

20 Marshall Kirkpatrick, "Netflix's Big Data Plans to Take over the World," http://readwrite.com, July 26, 2011.

21 Bruce Upbin, "How Intuit Uses Big Data for the Little Guy," www.forbes.com, April 26, 2012.

22 Stephanie Overby, "In Awarding Prize for Analytics, Netflix Failed to Predict it wouldn't be Used," https://data-informed.com, April 30, 2012.

[23] "Ted Sarandos – Audience Tastes," www.carseywolf.ucsb.edu, June 2012.

[24] Andrew Wallenstein, "Friending Facebook a Big Deal for Netflix," http://variety.com, March 13, 2013.

[25] Xavier Amatriain and Justin Basilico, "Netflix Recommendations: Beyond the 5 Stars (Part 1)," http://techblog.netflix.com, April 6, 2012.

[26] Dave Feinleib, "How the Netflix Big Data Approach is Transforming Education," www.forbes.com, July 20, 2012.

[27] "Ted Sarandos – Audience Tastes," www.carseywolf.ucsb.edu, June 2012.

[28] Xavier Amatriain and Justin Basilico, "Netflix Recommendations: Beyond the 5 Stars (Part 1)," http://techblog.netflix.com, April 6, 2012.

[29] *Ibid.*

[30] Rachel King, "Panel: Netflix, StubHub, IBM Execs Discuss Value of Big Data," www.zdnet.com, August 10, 2012.

[31] Kat Ascharya, "Media Mind: Binge Viewing and the Shifting TV Landscape," www.mobiledia.com, October 10, 2012.

[32] Ryan Lawler, "Netflix Moves into the Cloud with Amazon Web Services," http://gigaom.com, May 7, 2010.

[33] John Ciancutti, "Four Reasons we Choose Amazon's Cloud as our Computing Platform," http://techblog.netflix.com, December 14, 2010.

[34] Barb Darrow, "Netflix: We don't Need No Stinkin' Data Centers!," http://gigaom.com, March 28, 2012.

[35] "Netflix Selects Amazon Web Services to Power Mission-Critical Technology Infrastructure," http://phx.corporate-ir.net, May 7, 2010.

[36] Joel Schectman, "Netflix Uses Big Data to Improve Streaming Video," http://blogs.wsj.com, October 26, 2012.

[37] Dan Gallagher, "Amazon Outage Shines Light on Netflix's Big Bet," http://blogs.marketwatch.com, December 26, 2012.

[38] MG Siegler, "Netflix Original Content is Much More than a Strategy Shift — it Could Shift an Industry," http://techcrunch.com, March 18, 2011.

[39] Reinhardt Krause, "Netflix's Bigger Issue than Price Hikes: Data Caps," http://news.investors.com, July 15, 2011.

[40] Alex Ben Block, "Netflix's Ted Sarandos Explains Original Content Strategy," www.hollywoodreporter.com, April 7, 2012.

[41] Steven Rosenbaum, "Netflix's Risky Strategy for 'House of Cards'," www.forbes.com, February 5, 2013.

[42] Alex Ben Block, "Netflix's Ted Sarandos Explains Original Content Strategy," www.hollywoodreporter.com, April 7, 2012.

[43] Christopher Williams, "Is Netflix about to be a Has-Been? Suddenly, Problems Seem Overwhelming," www.forbes.com, May 2, 2012.

[44] Dean Takahashi, "Netflix Confirms Deal to Launch Kevin Spacey Series via Video Streaming," http://venturebeat.com, March 21, 2011.

[45] Alex Ben Block, "Netflix's Ted Sarandos Explains Original Content Strategy," www.hollywoodreporter.com, April 7, 2012.

[46] David Carr, "Giving Viewers What they Want," http://nytimes.com, February 24, 2013.

[47] Roberto Baldwin, "Netflix Gambles on Big Data to Become the HBO of Streaming," www.wired.com, November 29, 2012.

[48] Roberto Baldwin, "With *House of Cards*, Netflix Bets on Creative Freedom," www.wired.com, February 1, 2013.

[49] Kyle Vanhemert, "The Secret Sauce Behind Netflix's Hit, "House of Cards": Big Data," www.fastcodesign.com, February 19, 2013.

[50] Thomas H. Davenport, "The Future of Entertainment is Analytical," http://mobile.blogs.wsj.com, March 6, 2013.

[51] Roberto Baldwin, "Netflix Gambles on Big Data to Become the HBO of Streaming," www.wired.com, November 29, 2012.

[52] Nellie Andreeva, "Netflix to Enter Original Programming with Mega Deal for David Fincher-Kevin Spacey Series 'House of Cards'," www.deadline.com, March 15, 2011.

[53] Michael Berliner, "Internet TV Commissioning Round-Up," www.guardian.co.uk, February 21, 2013.

[54] Cliff Edwards, "Netflix's Big Gamble with 'House of Cards'," www.sfgate.com, February 4, 2013.

[55] Greg Satell, "What Netflix's 'House of Cards' Means for the Future of TV," www.forbes.com, March 4, 2013.

[56] Dieter Bohn, "House of Cards the 'Most-Watched' Show on Netflix, will 'Arrested Development' Follow?," www.theverge.com, February 12, 2013.

[57] Brian Stelter, "Release of 13 Episodes Redefines Spoiler Alert," www.nytimes.com, February 5, 2013.

[58] "'House of Cards' Getting Thousands of Mentions per Hour on Social Media," www.allvoices.com, February 2, 2013.

[59] "'House Of Cards': A Sign that Netflix can Become a Content Creator," http://seekingalpha.com, March 13, 2013.

[60] Natan Edelsburg, "Is 'House of Cards' a Success? Social TV has the Answer," http://lostremote.com, February 18, 2013.

[61] Laura Hudson, "The Poll Results are in: You're Totally Binge-Watching House of Cards," www.wired.com, February 5, 2013.

[62] Luke Westaway, "House of Cards is Most-Watched Show on Netflix, Everywhere," http://crave.cnet.co.uk, February 13, 2013.

[63] Dieter Bohn, "House of Cards the 'Most-Watched' Show on Netflix, will 'Arrested Development' Follow?," www.theverge.com, February 12, 2013.

[64] "'House Of Cards': A Sign that Netflix can Become a Content Creator," http://seekingalpha.com, March 13, 2013.

[65] Sophie Curtis, "Netflix Foretells 'House of Cards' Success with Cassandra Big Data Engine," Techworld, March 28, 2013.

[66] "The 'Big Data' Revolution: How Number Crunchers Can Predict our Lives," www.npr.org, March 7, 2013.

[67] Andrew Leonard, "How Netflix is Turning Viewers into Puppets," www.salon.com, February 1, 2013.

[68] Ryan Lawler, "How Netflix Will Use Big Data to Push House of Cards," http://gigaom.com, March 18, 2011.

[69] *Ibid.*

[70] Daniel Ferry, "A Bear Gives a Bull Case for Netflix," www.fool.com, February 24, 2013.

[71] Margaret Rock, "The Year Ahead: Finding Wisdom in Big Data," www.mobiledia.com, December 17, 2012.

[72] Jason Lange, "Netflix, Big Data, and What Hollywood is Missing out on.," www.jasonlange.me, September 24, 2012.

[73] "Ted Sarandos – Audience Tastes," www.carseywolf.ucsb.edu, June 2012.

[74] Peter Kafka, "Big Data, Soft Sell: Netflix Pitches a Hands-Off Approach to Hollywood," http://allthingsd.com, March 1, 2013.

[75] Jesse Hirsh, "Netflix, Big Data, and House of Cards," http://jessehirsh.ca, February 5, 2013.

[76] Roberto Baldwin, "Netflix Gambles on Big Data to Become the HBO of Streaming," www.wired.com, November 29, 2012.

[77] Daniel Ferry, "A Bear Gives a Bull Case for Netflix," www.fool.com, February 24, 2013.

[78] Thomas H. Davenport, "The Future of Entertainment is Analytical," http://mobile.blogs.wsj.com, March 6, 2013.

[79] Andrew Wallenstein, "Netflix Series Spending Revealed," http://variety.com, March 8, 2013.

[80] Daniel Ferry, "A Bear Gives a Bull Case for Netflix," www.fool.com, February 24, 2013.

[81] Rick Liebling, "The Media House of Cards and Netflix's Big Disruption," www.fastcocreate.com, February 28, 2013.

[82] Cliff Edwards, "Netflix's Big Gamble with 'House of Cards'," www.sfgate.com, February 4, 2013.

[83] Peter Pham, "The Future of Netflix is Data Driven," http://beta.fool.com, March 9, 2013.

[84] Greg Satell, "5 Trends that Will Drive the Future of Technology," www.forbes.com, March 12, 2013.

[85] Brian Stelter, "Release of 13 Episodes Redefines Spoiler Alert," www.nytimes.com, February 5, 2013.

[86] Greg Satell, "What Netflix's 'House of Cards' Means for the Future of TV," www.forbes.com, March 4, 2013.

[87] Andrew Wallenstein, "Netflix Series Spending Revealed," http://variety.com, March 8, 2013.

[88] Joel Schectman, "Netflix Amazon Outage Shows 'any Company can Fail'," http://blogs.wsj.com, December 27, 2012.

[89] Clint Boulton, "CIOs must Keep Big Data Private," http://mobile.blogs.wsj.com, October 23, 2012.

[90] Brad Brown, Michael Chui and James Manyika, "Are you Ready for the Era of 'Big Data'?," www.mckinseyquarterly.com, October 2011.

[91] David Carr, "Giving Viewers What they Want," http://nytimes.com, February 24, 2013.

[92] Jessica Leber, ""House of Cards" and Our Future of Algorithmic Programming," www.technologyreview.com, February 26, 2013.

[93] Andrew Leonard, "How Netflix is Turning Viewers into Puppets," www.salon.com, February 1, 2013.

[94] Sophie Curtis, "Netflix Foretells 'House of Cards' Success with Cassandra Big Data Engine," Techworld, March 28, 2013.

[95] David Carr, "Giving Viewers What they Want," http://nytimes.com, February 24, 2013.

[96] Steve Lohr, "The Age of Big Data," www.nytimes.com, February 11, 2012.

[97] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh and Angela Hung Byers, "Big data: The Next Frontier for Innovation, Competition, and Productivity," www.mckinsey.com, May 2011.

[98] Steve Lohr, "The Age of Big Data," www.nytimes.com, February 11, 2012.

[99] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh and Angela Hung Byers, "Big data: The Next Frontier for Innovation, Competition, and Productivity," www.mckinsey.com, May 2011.

[100] Brad Brown, Michael Chui and James Manyika, "Are you Ready for the Era of 'Big Data'?," www.mckinseyquarterly.com, October 2011.

[101] Jacques Bughin, John Livingston and Sam Marwaha, "Seizing the potential of 'big data'," www.mckinseyquarterly.com, October 2011.

[102] Brad Brown, Michael Chui and James Manyika, "Are you Ready for the Era of 'Big Data'?," www.mckinseyquarterly.com, October 2011.

[103] Stefan Biesdorf, David Court and Paul Willmott, "Big Data: What's your Plan?," www.mckinseyquarterly.com, March 2013.

[104] *Ibid.*

[105] Nayan Chanda, "Can Big Data Save Lives?," www.businessworld.in, June 22, 2012.

[106] Kevin Fitzpatrick, "How Much Money is Netflix Putting into 'Hemlock Grove?'," http://screencrush.com, June 12, 2012.

[107] Roberto Baldwin, "With *House of Cards*, Netflix Bets on Creative Freedom," www.wired.com, February 1, 2013.

[108] Roberto Baldwin, "Netflix Gambles on Big Data to Become the HBO of Streaming," www.wired.com, November 29, 2012.

[109] "Analysis: Why Netflix must Rethink Binge Viewing," http://variety.com, January 29, 2013.

[110] Cliff Edwards, "Netflix's Big Gamble with 'House of Cards'," www.sfgate.com, February 4, 2013.

[111] Debbie Cai, "S&P Lowers Netflix Outlook to Negative on Higher Debt Leverage," http://online.wsj.com, January 29, 2013.

[112] Tom Cheredar, "What the Disney Streaming Deal Means for Netflix's Future," http://venturebeat.com, December 5, 2012.

[113] Tom Cheredar, "Netflix's Reed Hastings Says Amazon Prime Loses $1B Annually," http://venturebeat.com, November 17, 2012.

[114] Rebecca Greenfield, "The Economics of Netflix's $100 Million New Show," www.theatlanticwire.com, February 1, 2013.

[115] Agustino Fontevecchia, "Amazon Eyeing Everyone's Lunch: Bezos Going after Apple, Netflix, Big Data, and Cloud Computing," www.forbes.com, October 25, 2012.

[116] David Carr, "Giving Viewers What they Want," http://nytimes.com, February 24, 2013.

[117] Tim Wu, ""House of Cards" and the Decline of Cable," www.newyorker.com, February 4, 2013.

[118] Daniel Ferry, "A Bear Gives a Bull Case for Netflix," www.fool.com, February 24, 2013.

Case 7

Big Data Strategy of Procter & Gamble: Turning Big Data into Big Value

Debapratim Purkayastha and Koti Vinodbabu

The Procter & Gamble Company (P&G), a leading consumer packaged goods company, was regarded as a pioneer in extensively adopting big data and digitization to understand consumer behavior and facilitate quick decision making. All its activities – internal communications, consumer research, product development, production systems, supply chains, marketing and promotions, and customer relationship management – sought to leverage data. According to analysts, this digitization drive had resulted in P&G becoming more nimble and efficient. The individuals mainly credited for this feat were the company's former Chairman and CEO, Bob McDonald (McDonald), and Chief Information Officer (CIO), Filippo Passerini (Passerini).

With Passerini stepping down in June 2015, Linda W. Clement-Holmes (Linda) who had previously acted as P&G's Global Information & Decision Solutions Officer, took charge as the new CIO. According to P&G's top management, digitization and analytics had helped the company save billions of dollars and was a key to its quick launch of products. Experts too thought that P&G *"is doing big data right"* and *"is building tremendous organizational capabilities around big data. While P&G appears to have fully embraced big data, they perhaps uniquely benefited from a forward-thinking CEO and nearly unlimited corporate resources to build out the organizational capability."*[1]

But some experts were skeptical about P&G's obsession with digitization, however well-intentioned it might be. They felt that the company might consume more time in analyzing the big data and this would hamper the speed of its decision making. Big data processing might even damage the company's operations and its reputation in the

event of a privacy breach of consumer data. The issues before Linda were how to ensure P&G struck a balance between efficiently leveraging big data and ensuring the security and reliability of its systems and how she could work with the top management to reform P&G's organizational structure according to big data requirements.

Big Data Origins

Big data referred to extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions. It denoted the huge amount of data recurrently collected through devices and technologies such as credit cards and customer loyalty cards, social media from the internet, Wi-Fi sensors, and electronic devices. Much of this data was unstructured, that is, not shaped in a specific, predefined data model. According to the Oxford English Dictionary, the term 'Big Data' was first used in 1941 to quantify the growth rate in the volume of data – alternatively known as the information explosion. Later, the term became more and more related to information in the digital format and the Information Technology sector.

By 2013, big data had started to gain acceptance as it held promising opportunities for businesses. It was showing its impact on the healthcare, industrial, retail, and financial sectors, to name a few. It enabled companies to run live simulations of trading strategies, geological and astronomical data, and helping stock brokers analyze public sentiment about a company from social media. Emerging technologies such as Hadoop [a], NoSQL[b], and Apache Storm[c] made such analytics possible. According to a Gartner survey in 2013, 64% of organizations had invested or planned to invest in the technology, but only 8% of them had actually begun deployment. Many businesses were in the process of gathering information as to which business problems big data could solve for them. In 2015, annual revenue from the global big data market was nearly US$22 billion, with predictions suggesting revenue doubling in size by 2020.[2]

[a] Hadoop is an open-source framework that allows the storing and processing of Big Data in a distributed environment across clusters of computers using simple programming models.

[b] A NoSQL database provides a mechanism for storage and retrieval of data which is modeled by means other than the tabular relations used in relational databases.

[c] Apache Storm is a distributed stream processing computation framework written predominantly in the Clojure programming language. It uses custom created 'spouts' and 'bolts' to define information sources and manipulations to allow batch, distributed processing of streaming data.

About P&G

P&G, headquartered in Cincinnati, Ohio, marketed more than 300 products in over 180 countries. Its entire operations were classified into six categories – beauty, grooming, health care, snacks & pet care, fabric care & home care, and baby care & family care. For the year ended June 30, 2016, P&G had generated net sales of US$65.3 billion and net profit of US$9.93 billion *(See Exhibit I)*.

P&G was considered one of the most innovative companies globally in terms of applying technology to developing products. The company's innovative culture was attributed to its Connect+Develop (C+D) initiative launched in 2000. The aim of the initiative was to grow the company by tapping innovation from outside, and, using such innovation to introduce new products.

The goal was that 50% of new innovations should be sourced externally.[3] P&G deployed information technology to partner with companies, entrepreneurs, and, research institutions.[4] It spotted brilliant concepts globally and used its in-house capabilities to make tangible products. IT operations were outsourced. According to analysts, the biggest and the most enduring benefit that accrued to the company from the C+D initiative was the change in the thought processes of its employees. P&G's employees became open to experimenting with new ideas and did not shy away from collaborating with external partners to enhance the company's business.

In 1999, P&G formed its shared services division, Global Business Services (GBS), to provide cost-effective business assistance globally – the company, at the same time, shifted from a regional structure to a product-oriented approach with the creation of four Global Business Units – Fabric and Home Care; Paper; Beauty Care; Food and Beverage and Health Care.[5] GBS had to equip P&G with superior IT, finance, facilities, and human resources. Technology formed an essential part of its operations.

McDonald and Passerini (who was also the Group President of GBS) had singled out 88 specific business processes that steered the company's activities. For each of the processes the interval between the time the data flowed in and the time it was acted upon was calculated. Utilizing behavioral forecasting models technologies, it was established that the speed of taking decisions could be enhanced and the time lag in implementation condensed. The goal was to operate the business in real-time and enable information access to everyone simultaneously.[6]

Big Data at P&G

At P&G, predictive analytics[d] formed a vital component of all significant decisions which had a bearing on sales and margins *(See Exhibit II)*. P&G started employing analytics in 1992 when it had an excess number of production units in USA and was required to cut down on the surplus production infrastructure and devise mechanisms to right-size its global supply chain. At that time, P&G's analytics team worked on aspects like the North American Free Trade Agreement's effect on the business dealings, trucking deregulation, and excess manufacturing infrastructure, and suggested different condensation possibilities. A number of manufacturing units were closed down, and, by early 2008, P&G had achieved expense reduction in excess of US$1 billion.

By 2004, operations research (OR) was embedded into the decision making architecture of P&G. Analytics was deployed to take decisions on handling inventory, drafting the supply chain, deciding the quantum of stocks to be supplied to retail stores, determining the most favorable hours in a day to display television ads, and the timing of new product roll-outs.[7]

According to analysts, P&G was an example of a company for whom big data was capability augmenting. The company had given preference to analytics and had had an analytics group since 1992. Besides, it had been making efforts to understand consumer behavior on a priority basis. As a result, P&G was implementing big data practices ahead of most other companies. The big data initiative was led by Passerini and supported by McDonald. P&G had also modified its hiring practices to fetch in more data scientists. Since 2011, P&G and Google Inc. had been exchanging experts annually. While Google sought to learn about advertising, P&G wanted to learn from Google's digital intelligence. At P&G, all managers were improving their digital skills, and every manager's digital and analytical performance was evaluated in the performance management process.

P&G needed to clearly and easily understand its rapidly mounting and huge amount of structured and unstructured data across research and development, the supply chain, and customer relationship operations both from traditional data sources and new sources of online

[d] Predictive analytics comprises a host of statistical methodologies from modeling, computer algorithms, data mining, and interactive decision theories that analyze current and historical facts to make predictions about future events. Models grasp the connections between diverse variables to enable evaluation of risk or prospects pertaining to specific situations and aid choice making.

data. In addition, business teams at P&G required access to more data about sales, market capacities, demographics, weather, etc. with more depth to help the experts understand the data. P&G faced challenges in providing a clarifying vision of *"why is this happening"* in every division to its experts. It also wanted to quickly obtain and assimilate multiple structured and unstructured data sources at big volumes for quick decision making.

P&G also needed to be able to incorporate a variety of new data types to get to know current trends in the markets as the traditional approaches of data analytics were not working. And growing data volumes were leading to rising storage costs. P&G thus needed the intervention of experts to pick the right data at the right time. It was for these reasons that P&G ventured into big data analytics.

P&G checked different data analytics companies and associated with Oracle Corporation (Oracle) in November 2014. P&G and Oracle decided to test with a Big Data Cloud platform for P&G by April 2015. Oracle functioned directly with P&G to help the launch of the service. It agreed to set up a provisional Big Data Cluster by April 1, 2015, with the understanding that P&G would move to the Public Cloud platform. Oracle turned over the Big Data Public Cloud Server to P&G on April 1, 2015.

However, the IT team from GBS had been working with Hadoop (Cloudera on Oracle appliances) since 2014. The first year's efforts were based on exploratory analytics and proof of concepts while the second year's activities were based on big data-based applications. In this process, the team recognized a need for a Business Intelligence (BI) tool more appropriate to the Hadoop ecosystem. In 2015, the team boarded on a project to allow 300 users worldwide to start developing insights from 200 Tera bytes of raw data stored in Hadoop. The particular application was built using Arcadia Data's[e] Hadoop-converged BI technology that permitted the team to construct an application meeting their goals.

P&G was able to find solutions for many business problems with new analytics competencies resulting from the implementation of the Oracle Big Data Appliance, combined with the Cloudera Hadoop ecosystem. The company assimilated vast amounts of structured and unstructured data across customer interactions, customer facing operations, supply chain, and research and development, from both traditional data sources and new sources of online data. The solution

[e] Arcadia Data unifies data discovery, visual analytics and business intelligence in a single, integrated platform that runs natively on Hadoop clusters.

helped the company to serve its customers efficiently wherever they were located. The Oracle Big Data Appliance, which was a pre-engineered solution, supported the company's capability to conduct analysis on multiple data sets including those from industry sources, retail data, and internal enterprise information within no time and to react more quickly. P&G was able to load and integrate data faster and perform trustworthy analysis at scales that were formerly not possible.

P&G had refined its hiring practices to recruit data scientists, digitally skilled experts, software developers, social media experts, and managers who were well versed with quantitative decision practices. In 2015, Guy Peri was appointed as P&G's Chief Data Officer. His job was to make sure that P&G leveraged data in the best and most secure way. His responsibility was to work with the company's data science team which applied advanced analytics against P&G's largest business opportunities across marketing, sales, eBusiness, product innovation, and the supply chain.[8]

Marketing

P&G made increasing use of Web 2.0 technologies to promote its products and augment their brand recall. An important medium was the factual accounts narrated by consumers about their product experiences. McDonald, who wanted P&G's brands to have a distinct rapport with each customer, said, *"Advertising is very different today. You don't talk to somebody. You engage them in a discussion and you give them the freedom to participate in that discussion, and actually to advertise for you."*[9]

Though P&G sold its products through about 1500 websites, it sold its products more exclusively through retailers and that made it a tough task for the company to gather customer data. P&G had earlier depended largely on television for promoting its products. But as consumers started spending more time online through smartphones, tablets, and PCs, it realized TV had lost its effectiveness. As part of its efforts to collect customer data, P&G's global nappy brand Pampers created Pampers.com, a website that offered advice, free samples, and money saving coupons personalized to each stage of a woman's pregnancy or the baby's age, which users selected with a graphic slider at the top of the page. In return, consumers had to submit personal details as they logged in to the site. Dennis Devine (Devine), associate director, consumer solutions, P&G, said, *"We had huge amounts of data and it was all over the place. There were no standards for it; there was no way to view it. We took it on ourselves to set up a vision of how we could bring all this information together."*[10]

The GBS's consumer relationship management wing started a conceptual architecture, which positioned the consumer at the center, and enumerated each possible 'touch point' around them. A touch point might be online, in a physical store, or on television. Each was classified and put up as a predominant view of consumer interaction. In addition, P&G began to gather data for all named consumers into a single source from consumer websites, with the information of all the choices of customers and about how they wanted to interact with P&G brands or any programs they took part in. For this, P&G selected Teradata to construct and host a data warehouse, campaign management and analytics system in the cloud, on a pay per usage basis. Devine said, *"We can look at demographics, which consumers are engaged with a marketing program, what are they engaged with, and what is driving them. We have information on transactions, can see who is most loyal, and see how we can make other consumers more loyal. All these questions we can answer with this tool."*[11]

P&G used Secure File Transfer Protocol[f] to extract transfer and load data to its data warehouse. Master data management was supported by software from Trillium Software. The cloud-based system was built on a Teradata data warehouse, integrating Teradata Customer Interaction Manager, the Retail Logical Data Model, Teradata Master Data Management, and analytics tools including Teradata Warehouse Miner.

P&G made efforts to obtain quicker feedback on the performance of its marketing campaigns and switched to constant tracking and scenario planning. Though it had previously been employing marketing-mix modeling[g] to determine the pay-offs on marketing expenditure, dissections poured in late. P&G moved to specific marketing-mix assessment software and obtained analyses within one and a half months post-initiative.[12]

Decision Making

McDonald said, "By getting the right data to the right decision makers at the right time, we can become increasingly efficient and productive."[13] Data management at P&G was founded on three

[f] Secure File Transfer Protocol is a network protocol that provides file access, file transfer, and file management over any reliable data stream.

[g] Market Mix Modeling is the determination of additional sales resulting from marketing spends like couponing, advertising, promotional price reductions, etc. This is achieved by considering business performance for a specific time-frame, employing advanced mathematical and statistical tools, and ascribing sales variance to the various expenditures.

principles: Openness of data (making the same information available to all concerned employees), well-timed data (giving the data as soon as possible), and transmission of data (making the data available through various media). The company extensively used online tools like high-speed networking, data visualization, and quick assessment of diverse data strands. Alike data was remitted simultaneously on mobile gadgets, desktops, and P&G's Business Sphere set-up. This enabled the managers to analyze information irrespective of their location. As a result, the time taken for decision-taking was reduced from weeks or months to minutes. P&G was reaching the stage where it could monitor each employee's competence, where product-wise weekly sales targets could be set, and the manufacturing gauged real-time globally.[14] It had developed the competence to analyze the profit estimates and the unit sale estimates real-time and make amendments immediately.[15]

When Passerini took over as CIO of P&G in 2004, he renamed the IT department "Information and Decision Solutions (IDS)". The renaming was based on Passerini's conviction that data and analytics needed to play a more vital role in decision making at P&G. Since then, the IDS unit had planned several initiatives that had changed the way decisions were taken at P&G.

Business Sphere

In 2008, P&G's GBS started looking for BI software with the objective of providing a flexible BI tool to its decision makers and facilitating access to real time data so that they could perform predictive analytics and answer the 'what if' questions that arose in decision making. Another objective was to find a system that allowed P&G employees all over the world to attend the same meetings and view the same data. *"Senior executives had been limited to running the business by looking in the rearview mirror, and we wanted a forward-looking focus with better speed and quality of the decision-making,"*[16] said Alan Falkingham, director of BI and analytics at P&G. The BI initiative resulted in the creation of innovative meeting rooms called Business Spheres in partnership with Cisco, SAP, Nielsen, and TIBCO Spotfire.

GBS wanted to equip the top management teams with all the information required to reach resolutions during meetings. It developed a conference room called Business Sphere which comprised a round table circumscribed by two television displays, each 32 by 8 foot capable of displaying charts, graphs, heat maps and other visuals *(See Exhibit III)*. The members could view a world map of markets expanding or contracting depending upon their performance vis-à-vis the targets. The view could be narrowed down to specific countries and product

categories. By clicking a button, micro-level data from any of the countries that P&G operated in, such as the prices of the company's skin cream in Brazil and its sales data vis-à-vis other brands, could be displayed. More clicks could lead to the details of its detergent sales in Lebanon, if required.

P&G executives across the globe met in the Business Spheres each week, to appraise the latest results and forecasts available through the Decision Cockpit dashboards. The targets were like adjusting pricing, changing the product mix, changing marketing approaches, or increasing marketing expenditures to regain market share where there were losses or to improve margins where the settings were strong. These might indicate that an additional 770,000 cartons had to be sold to match the previous quarter's sales. The advancement of the strategy could be monitored by all. Every corner of the Business Sphere room had a mini screen for senior management in remote regions to participate in via video. These managers could view the data images on their laptops or iPads.

P&G experts used the Spotfire BI tool to study data and compiled the reports to present them to the managers at meetings in the Spheres. Spotfire could analyze the agenda of the employees in the Sphere and create 'what if' scenarios to exemplify the consequences of different approaches. The program analyzed and connected as much as 200 terabytes of data to answer a set of questions, allowing for in-depth analysis and customization.

Decision Cockpit

The Decision Cockpit involved placing the key information on computer desktops of the decision makers to help over 50,000 P&G employees make better decisions quickly. For instance, a salesman might get updates about important clients, sales opportunities that could be pursued, and product performance trends. The objective was to enable employees to concentrate on finding solutions instead of deliberating on what the issues were. The data presented on the heat map[h] *(See Exhibit IV)* explained what was happening in different markets and allowed team members, supported by data analysts, to know why variations in market penetration were taking place so that they could work together to resolve how to increase market penetration.

[h] A heat map shows all the markets in which P&G products compete and their relative share (red indicating low market share and green indicating high market share), and also puts in clear perspective the importance of growing the share of any one of those markets.

The technology powering the Decision Cockpit, known as 'Consumer Pulse', employed analytics on customer feedback garnered from blogs, tweets, and social networking websites.[17] Remarks were segregated brand-wise and were displayed on the concerned brand manager's computer screen for an immediate response to the customer.[18] Consumer Pulse used Bayesian Analysis/ Bayesian Inference[i]. At P&G, BayesiaLab facilitated superior reasoning by enabling faster establishment of connections between diverse variables in consumer experiments and by instantly breaking down consumer data.[19]

Senior-level managers were equipped with wireless applications to study performance briefs, permit expenses, and be informed if things went off the track, irrespective of their locations. In July 2011, P&G equipped 18,000 of its employees, scattered worldwide, with cloud computing applications to enable information exchange and partnerships. P&G employees could access this data exchange service on any wireless device.

Product Development

By 2011, P&G was extensively using computer-based simulation to design and evaluate virtual replicas prior to their manufacture. Only when the virtual replicas satisfied the quality parameters did P&G progress to the real-world prototype phase.[20]

P&G leveraged big data effectively in its new product development process by combining consumer data from multiple brand outlets and using it for innovations. It used modeling and simulation tools to minimize prototyping expenses. For instance, it had used them to decide how the molecules in a dishwashing liquid would react over time to improve the product. Apart from this, as of 2013, the C+D program was receiving more than 4,000 ideas a year from around the world for new products and product enhancements.

From creating the blueprint for a product or a container through process manufacturing and the manufacturing unit's output, modeling and simulation played a crucial part in P&G's activities.[21] The objective was to employ simulation to enable the visualization to the maximum extent possible in all activities. For instance, in July 2011, P&G launched a high-end version of its detergent Ariel, which contained

[i] Bayesian inference is a technique of statistical inference wherein data is employed to revise the imprecision of parameters and estimates in a probability design. As data accrues, the confidence level in a hypothesis either soars or abates.

'micro boosters' that the company claimed penetrated deep and removed dirt. These boosters were evolved by employing 3D technology, which helped P&G's scientists view new dimensions, outlines, and tiers of the cloth and devise methods to penetrate deep to give a good wash.[22] The company used big data tools and testing in product development. For instance, Tide dry cleaners, a branded dry cleaning franchise, was developed by leveraging consumer visions about the shortfalls of the existing dry cleaning products, and P&G's own insights into consumer household cleaning habits.

P&G's next goal was to reproduce the whole store and simulate a consumer's experience with its products *(See Exhibit V)*. These set-ups were equipped with giant-sized video displays, powered by computer modeling, simulation, and 3D technologies on all walls, replicating a retail store. Consumer focus group members looked at and evaluated virtual product depictions. Advanced software enabled documentation of consumer responses to product arrangement, colors, and molds. Eye-monitoring software was used to gauge the product and packing features to which the buyers paid heed. Virtual solutions applications were applied in nearly four-fifths of all the company's projects. P&G disseminated the findings of its virtual reality studies to its retailers to assist them in arriving at more sensible choices regarding the brands and volumes that they had to stock. P&G also employed these set-ups to construct virtual 3-D shop surroundings. This helped the retailers visualize how product arrangements appeared in their shops and improve the current stacking pattern and in-store promotions. Its eye-tracking software also helped evaluate the store shelf palpability of its products.[23] Bernard Eloy, Director, Virtual Solutions, P&G, said, *"One of the key benefits of virtualization is that we are gathering key data and insights, which we can pull together to create new products and new brand experiences."*[24]

Supply Chain

P&G started employing demand sensing[j] software (DSS) worldwide to increase short-run estimate precision and reduce safety stock. DSS obtained the existent estimation data from the demand plan, synthesized it with the real-time information of every day indents and deliveries, and employed advanced trend identification formulas to make daily estimates for each product line. On its own, it made changes

[j] Demand sensing varies from demand planning as it is about collecting information as quickly as possible from the time of occurrence of the event or in real-time and using it immediately to amend the supply chain.

to these estimates each day considering real-time information that was constantly fed into the system. By 2007, in the P&G businesses where DSS was deployed, the reduction in estimation error and safety stock exceeded 30% and 10% respectively, over 2006.[25]

P&G installed 'control towers' to handle its distribution structures globally. A control tower, in the context of a supply chain, was a site that allowed a view of inward and outward movement of goods, akin to an airport control tower that directed the maneuver of an airplane. This technology offered a comprehensive view of the supply chain and removed the flaws in freight arrangement to enable compressed conveyance time and enhanced protection. It also helped in reducing inventory and decreasing the number of inventory cycles.[26] As of August 2010, the Control Tower know-how had reduced the number of empty load trips by more than 15% when compared to 2009.[27]

P&G also used a system called Distributor Connect, which let it manage inventory in real time. Moreover, the firm kept connected to its retailers through a worldwide coordinated data warehouse that allowed it to manage commercial transactions in a completely automatic fashion. These systems aggregated data, which were compiled by data scientists, P&G's marketers, and engineers to develop new products and improve existing ones. The firm spent US$2 billion annually on R&D, and operated to structure the innovation by forming multiple groups accountable for generating new products and development. To spread the innovation to all departments, P&G assigned business and innovation strategies for senior executive leadership to review.

P&G analyzed consumer-behavior data to forecast orders worldwide to cut its distribution costs. It was implementing Terra Technology owned by supply chain analytics company E2open across its markets. P&G had deployed E2open's demand sensing analytics in 75% of its global business and had also implemented multi-enterprise demand sensing (MDS), inventory optimization, and transportation forecasting. MDS leveraged downstream data like feedback from point-of-sale (POS), channel inventory, warehouse withdrawals, distributor data, and retailer forecasts to better understand customer demand and quickly respond to sudden shifts in demand.[28]

P&G announced a partnership with Ecodesk in 2015, to launch a leading edge, cloud-based supplier assessment scorecard system to drive increased engagement and analysis of supplier sustainability performance. P&G had been using a supplier scorecard but in a spreadsheet format since 2010. The scorecard was built based on the association of P&G's granular, tried and tested approach for measuring

supplier sustainability performance with Ecodesk's know-how in cloud-based sustainable supply chain technology platforms, supplier engagement, and data collection analysis.

Results

In the early 2000s, P&G confronted the problem of rising R&D costs. It implemented the C+D program which leveraged InnoCentive, a web-based platform that invited experts to solve technical challenges that P&G was facing. By 2011, half of the new products had elements that had originated from outside the company, up from 15 percent in 2000. R&D efficiency at P&G was up at 60 percent, and R&D as a share of revenue had tumbled from 4.8 to 3.4 percent.

P&G was among the winners of the 2016 Oracle Cloud Platform Innovation Award in using Oracle's platform as a service (PaaS), and infrastructure as a service (IaaS) in new and advanced ways to give its business a makeover, drive innovation, and gain a competitive advantage. P&G was able to use Oracle's solutions to develop marketing insights from growing volumes of point of sale data. Jim Fortner, Vice President of IT Development and Operations at P&G, said, *"We can conduct analysis and respond to the data much quicker than ever before."*[29]

By leveraging big data, P&G was able to save inventory costs of over US$1billion.[30] Commenting on the results of using Oracle Big Data Appliance, Terry McFadden, associate director, GBS, P&G said, *"We wanted an end-to-end solution to provide visibility into structured and unstructured data, giving us the bigger picture and details about our products. Using the R language and the broader Hadoop ecosystem on Oracle Big Data Appliance gives us a solution that scales linearly. It is an extremely reliable platform that provides an excellent foundation for an analytics factory. I was convinced that Hadoop was the answer to our big data challenge. We did not want to knit our own cluster. Oracle Big Data Appliance gave us the shortest path to value. We could drop the implementation in and start chopping on problems."*[31]

Analytics was employed to identify the best supplier for every product ingredient globally. Each supplier was assessed to determine whether it could honor the commitment. As of 2007, P&G USA's retailers had been expending US$140 million annually on in-store exhibits. Employing analytics, P&G started advising its retailers online on these exhibits.[32] It earned US$67 million as fees per annum by cutting retailers' expenses, and, reduced the lead time for exhibits from five to two months. Due to the continuous movement of its products

across countries, P&G had to face substantial currency fluctuations. Analytics software was used to forecast the best exchange rates, and accordingly shuffle production and ingredient quantity procurement between nations.

Analytics was employed to determine inventory requirements and the places where they were required. New warehouses were added to the supply chains to optimize inventory. Analytics was also used to study recruitments, the number of employees resigning, retirements, job rotations, and career advancements, to develop a 'flow model' which informed managers about the inflow and outflow of employees during the subsequent period and helped them recruit accordingly.[33]

However, P&G's endeavor to quantify everything also came in for criticism from some observers. A marketing expert doubted the ability of Bayesian analysis to really decode human dynamics and, hence, of reliably measuring customer satisfaction. He felt that consumers wanted to be viewed as individuals and not as numbers and that P&G might not have got its priorities right by applying number-crunching to social networking platforms. Instead, it required brand managers who listened to consumers to really comprehend their wants. Regarding P&G's efforts to have a one-to-one customer connection through digitization, he doubted whether customers actually desired this or would prefer to simply use social media to convey product aspects that they disliked.[34]

Challenges

P&G sales peaked at US$83.7 billion in 2012 before stagnating and then declining as the company began offloading brands. As of mid-2016, for 14 of the previous 17 quarters, P&G had been losing market share in half or more of its products. The company's stock had stagnated, inching up a cumulative 6.5% since 2014.[35] In this situation, Linda, who had assumed the role of CIO in June 2015, had the big responsibility of continuing and leveraging the big data initiatives started by Passerini. Linda had varied experience in P&G and had served as Chief Diversity Officer, Senior Vice President of GBS, and as the Global Information & Decision Solutions Officer.[36] Analysts felt that succeeding a successful and long-tenured CIO was a double edged sword as Linda would inherit well run operations and a strong team but might find it challenging to chart her own path.[37]

Analysts felt that P&G's leadership had a vital role to play in adopting a culture of data-driven decision-making within the organization to reverse the situation in which it found itself. The head of

analytics at a leading logistics company described his efforts at driving a data-driven culture, *"Change management is one of the biggest challenges of Big Data implementation. Analytics needs to be integrated with processes. We had to educate and train our field force over and over again in order to make analytics a part of their daily routine."*[38]

Big data initiatives were needed to be implemented in different departments, which would make it difficult to define implementation goals and deadlines. Highlighting this challenge, Eric Spiegel, CEO of Siemens USA, said, *"Leveraging big data often means working across functions like IT, engineering, finance and procurement, and the ownership of data is fragmented across the organization. To address these organizational challenges means finding new ways of collaborating across functions and businesses."*[39]

Even as P&G was on course in its deployment of big data, concerns were expressed about the company's vulnerability to technological crashes, given that the dependability of big data was still in question. Passerini had earlier said, *"…we are at a point where if we had a major outage in our systems, it would be material instantaneously. If we have a security issue on one of our websites, that could have an impact on the company's reputation instantaneously. We live in a world where there is less margin for error than there was just a few years ago…because just one problem, but the 'wrong' one, could be one too many."*[40] However, analysts opined that big data could only advocate trends, authenticate claims, and reduce the amount of human error in decision making procedures; it was the human decisions and strategies that would turn the wheel. David Scott Taylor who became the CEO and President of P&G in 2015 admitted that big data had the potential to change P&G's fortunes with respect to its lagging revenues. However, he acknowledged, *"Until we put up numbers, I don't expect anyone to believe anything."*[41]

Exhibit I

Key Financials of P&G (2012-2016)

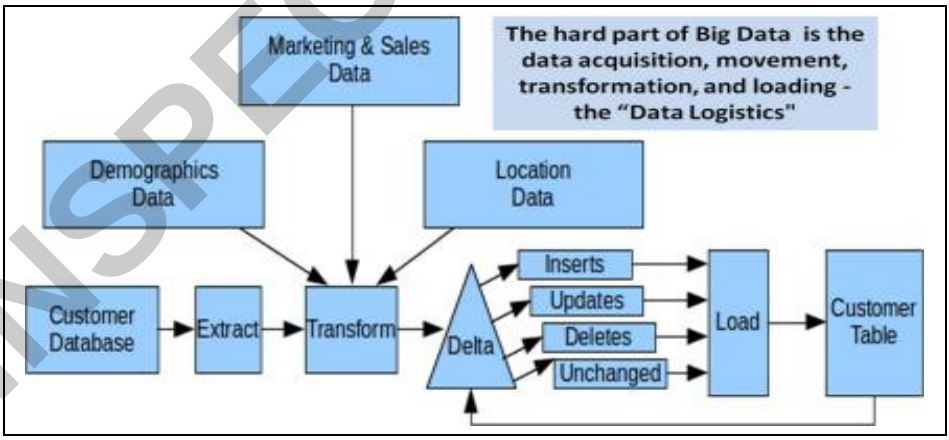
Fiscal year is July-June. All values US$	2012	2013	2014	2015	2016
Sales/Revenue	83.68B	82.58B	80.51B	70.75B	65.3B
Gross Income	41.64B	41.54B	39.9B	34.31B	33.11B
Net Income	9.17B	11.21B	11.2B	8.19B	9.93B
Preferred Dividends	256M	244M	253M	259M	255M
Net Income Available to Common	9.08B	11.07B	11.39B	6.72B	9.83B
EPS (Basic)	3.39	4.04	4.19	2.48	3.64
Basic Shares Outstanding	2.75B	2.74B	2.72B	2.71B	2.7B
EPS (Diluted)	3.18	3.86	4.01	2.33	3.55
Diluted Shares Outstanding	2.94B	2.93B	2.9B	2.88B	2.84B
EBITDA	20.29B	18.38B	18.48B	17.87B	17.56B

(M=Million), (B=Billion)

Source: www.marketwatch.com

Exhibit II

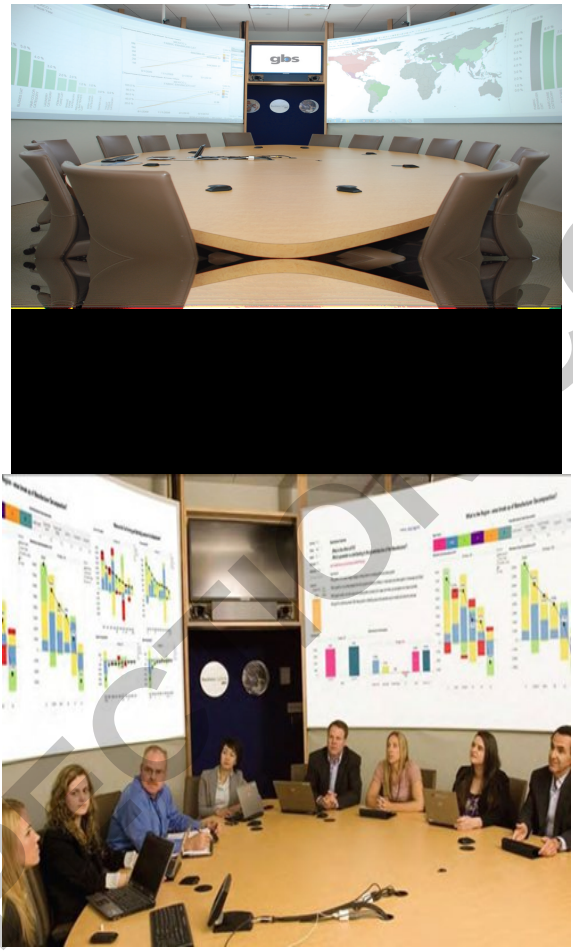
P&G Predictive Analytics



Source: practicalanalytics.co

Exhibit III

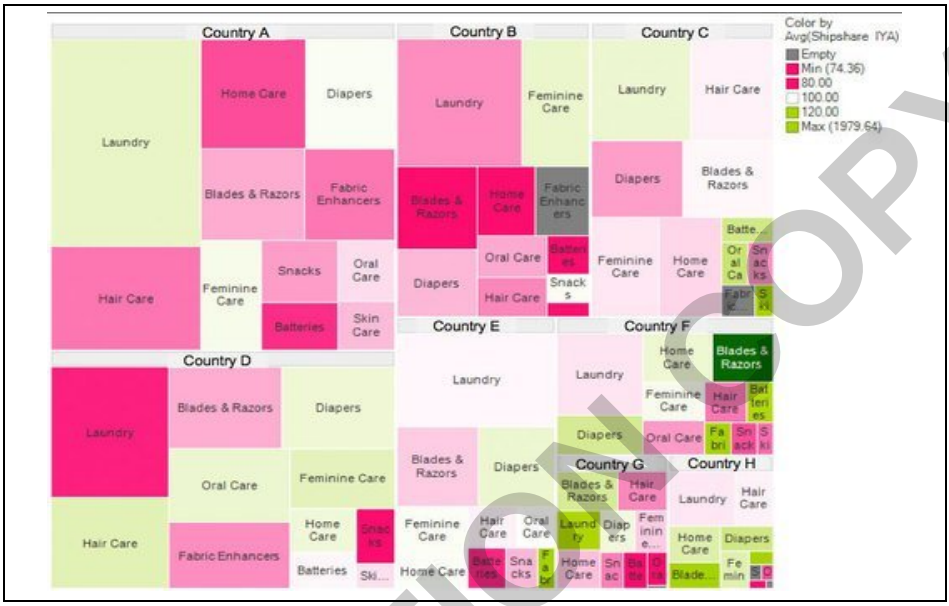
P&G's Business Sphere



Source: http://spotfire.tibco.com and www.fusioncharts.com

Exhibit IV

P&G Heat Map



Source: www.practicalanalytics.co

Exhibit V

P&G's Simulated Store

Source: www.accenture.com

End Notes:

1 John Egan, "Procter & Gamble: A Major Player In Data-Driven Marketing," www.create.pulsepoint.com, October 10, 2014.

2 Swapnil Bhagwat, "Why Big Data Is a Game Changer for Marketers?" www.customerthink.com, September 19, 2016.

3 James Joia, "Connect + Develop℠ with Procter & Gamble," www.innovatio management.se, May 28, 2009.

4 Sandy Staggs, "Connect + Develop with Procter & Gamble," www.ideacon nection.com, November 20, 2008.

5 www.pg.com/en_US/downloads/investors/annual_reports/1999/annual_ report_ 1999.pdf

6 Jessica Shambora, "Is P&G the Next Digital Giant?," http://tech.fortune. cnn.com, July 19, 2011.

7 "Procter & Gamble Wins Top Decision-making Award at Operations Research Meeting," http://scienceblog.com, October 2004.

8 http://www.indybigdata.com/guy-peri/

9 Jessica Shambora, "Is P&G the Next Digital Giant?," http://tech.fortune. cnn.com, July 19, 2011.

10 "Procter & Gamble Uses Teradata Cloud Analytics for Global Marketing," www.emilapld.wordpress.com, November 13, 2013.

11 Lindsay Clark, "Procter & Gamble Uses Teradata Cloud Analytics For Global Marketing," www.computerweekly.com, 2016.

12 Jack Neff, "P&G Shifts Marketing-Mix Biz to Nielsen, DemandTec for Faster ROI Reads," http://adage.com, August 12, 2011.

13 www.pg.com/en_US/downloads/investors/annual_reports/2010/PG_2010_ AnnualReport.pdf

14 "The Idea of an Information Democracy...," http://thefuturevalueof business.com, November 21, 2011.

15 Jessica Shambora, "Is P&G the Next Digital Giant?," http://tech. fortune.cnn.com, July 19, 2011.

16 Sue Hildreth, "Data+ Awards: Procter & Gamble Puts Worldwide BI Data In Executives' Hands," www.zoominfo.com, August 26, 2013.

17 Gib Bassett, "Indispensable Consumer Relationships at Heart of P&G Strategy," http://smartdatacollective.com, December 6, 2011.

18 "P&G Seeks "One-on-One" Relationship," www.warc.com, December 2, 2011.

19 "Strategic Partnership with Procter & Gamble," www.bayesia.com, June 15, 2009.

20 Bill Lydon, "Modeling & Simulation Speed Process Development," www. automation.com/resources-tools/articles-white-papers/articles-by-bill-lydon/modeling-simulation-speed-process-development.

21 John Clisham, "P&G: Leveraging Software Assets to Increase the Bottom Line," www.easasoftware.com/articles/CXO_editorial_on_EASA_at_P&G.pdf.

22 Dan Sewell, "P&G Readies New Detergent Global Launch," http://articles. boston.com, June 14, 2011.

[23] "Tobii Announces Collaboration Agreement with P&G," www.cisionwire.com, June 2, 2010.

[24] "Walmart, P&G, Kraft Share Sales & Marketing Tactics," http://consumer goods.edgl.com, June 8, 2011.

[25] "P&G Shaves Safety Stock with Terra Technology," www.supplychain brain.com, July 1, 2007.

[26] James A. Cooke, "The Next Big Things: "Control Towers" and Demand Shaping," www.supplychainquarterly.com, June 20, 2011.

[27] "Procter & Gamble Plans Digitization Drive," www.warc.com, August 31, 2010.

[28] "Terra Technology's Multi-Enterprise Demand Sensing Implemented Globally by Procter & Gamble," www.marketwatch.com, June 27, 2013.

[29] Murthy Mathiprakasam, "Big Data at Work in Consumer Goods and Retail with Oracle Data Warehousing," www.blogs.oracle.com, March 14, 2014.

[30] "Strategies to use Big Data in FMCG: How FMCG businesses are utilizing big data to gain a commercial advantage," www.prnewswire.com, November 12, 2015.

[31] "With Big Data, Procter & Gamble Gains Insights for Quicker, Smarter Business Decisions," www.oracle.com, November 12, 2015.

[32] Tuomas Sandholm, David Levine, Michael Concordia, Paul Martyn, Rick Hughes, Jim Jacobs and Dennis Begg, "Changing the Game in Strategic Sourcing at Procter & Gamble: Expressive Competition Enabled by Optimization," www.cs.cmu.edu, January-February, 2006.

[33] Andrew Hines, "How Operations Research Drives Success at P&G," www.cbsnews.com, February 13, 2008.

[34] "It's Now Possible to Have a One-on-One Relationship with Every Consumer.," http://newmediaandmarketing.com, December 1, 2011.

[35] Jennifer Reingold, "Can Procter & Gamble Find Its Aim Again?" www.fortune.com, June 9, 2016.

[36] Peter High, "Linda Clement-Holmes Ascends To CIO Role At P&G With Board-Level Membership Already In Hand," www.forbes.com, October 19, 2015.

[37] Martha Heller, "CIO Succession is a Double-Edge Sword," www.cio.com, November 11, 2014.

[38] https://www.capgemini-consulting.com/resource-file-access/resource/pdf/big_data_pov_03-02-15.pdf

[39] "Slogging Through Data? You're not Alone," www.domo.com, February 12, 2015.

[40] Michael Totty, "The View from the CIO's Office," http://online.wsj.com, April 25, 2011.

[41] "Can Procter & Gamble Find Its Aim Again?" www.inbusiness.ae, June 9, 2016.

Case 8

Baidu's Big Data Strategy – Impacting Chinese and Global AI Landscape

K B S Kumar and Indu Perepu

"Today, we look for information, In the future, information will look for us." [1]

– Ya-Qin Zhang, President, Baidu

In 2017, Baidu, China's biggest search engine increased its R&D spending to $464m, 28 per cent more than it had spent a year earlier. Baidu, which covered about 95% of China's population, had been betting big on bringing about a global revolution in Artificial Intelligence (AI) through some of its prestigious projects like Apollo, its autonomous driving program and DuerOS, an open platform to support applications in Artificial Intelligence (AI).

Pinning its hopes on Apollo – in 2017, Baidu floated a $1.52 billion autonomous fund exclusively to speed up the autonomous driving projects compete with its archrivals in the US. This was preceded by a US$2.9 billion spend on research and development in AI in 2016. Baidu roped in about 70 partners, which included big names like Microsoft, Ford, and Daimler, to accelerate the mission.

With over 44.5% of the mobile search market and 80.8%2 of the internet search market in China, Baidu had access to colossal amounts of demographic data. The company, which had over 70 million daily active mobile search users, 700 million internet users, and 6 billion search queries per day, had gathered massive amounts of data backstage, which it strategically channelized into the big data segment. In 2014, influenced by the rapidly transforming $34 billion global big data market3, Baidu shifted its strategic gears from 'All Mobile' to 'All AI' and changed its focus from internet and mobile to cloud

technologies. However, being a slow starter as compared to its peers like Alibaba and Tencent, Baidu lagged behind in terms of market share.

Nevertheless, its cutting edge big data architecture, state-of-the-art R&D facilities, ultra-modern cloud technology infrastructure, its hundred plus and still counting data analytics tools, its three-pronged cloud technology strategy (Tiansun, Tianxin, Tiangong), and its futuristic projects showed that Baidu had a promising future in the Artificial Intelligence segment globally.

About Baidu

Baidu was founded by Robin Li Yanhong (Li), and his friend Eric Xu (Xu) in 2000. Li who studied and worked in the US, developed a search mechanism called 'Link Analysis' during his stint with a company called Infoseek, which developed a search engine in the early 1990s. Link Analysis ranked the popularity of a website based on the number of other websites linked to it. Li got patent for Link Analysis. Later Infoseek was acquired by Walt Disney Company, after which it focused on content. Intending to pursue his interest, Li started his own search engine with Xu, who was also working in the US.

At that time Li analyzed the Chinese market, and realized that a big opportunity awaited him in China, as the number of Chinese using internet was growing rapidly. At the same time local portals could not cater to the needs of the people as they failed to customize the search according to local preferences. Li and Xu then decided to shift to China and started Baidu in a hotel room overlooking the Beijing University. They decided to call it Baidu, which meant hundreds of times, and had roots in Chinese culture[a]. Baidu set up its headquarters in Beijing and raised US$ 1.2 million from venture capitalists[b] in early 2000.

Baidu was launched in 2000, and at that time there were several internet portal in China like Sina, Sohu and Yahoo China. These portals offered services like mail, news, messaging etc. Baidu started offering search services for the other portals, and charged them whenever the service was used. Later on it developed its own search engine. In its own search engine, Baidu offered advertising space. Advertisers could place their ads by bidding for space on its search engine. Every time a user clicked on the ad, the advertiser was charged.

---

[a] It was based on a poem written by Xin Qiji in the 12th century.

[b] These included Integrity Partners and Peninsula Capital. Later in September 2000, two other venture capital firms, Draper Fisher Jurvetson and IDG Technology Venture, invested US$10 million in Baidu. US-based Google invested US$ 5 billion in Baidu.

Baidu grew rapidly and total revenues increased from RMB 10.5 million in 2002 to RMB 110.9 million in 2004. Quickly Baidu consolidated its presence in China's search engine market, and at the same time ventured into several other online services.

In 2005, Baidu was listed on Nasdaq. On the very first day of trading its stock price went up by 354% to reach US$ 122.4. This increased the value of the company to US$ 4 billion. The growth continued at by the end of 2007, it had 210 million users. It then ventured into C2C platform through Baidu Youa. It grew both organically and inorganically. It acquired travel related website Qunar, and launched its own mobile operating system Baidu Yi in 2011.

The success of Baidu Yi along with the changing search habits of Chinese, where they were shifting from desktop search to mobile search, Baidu invested in mobile search, and also launched a mobile phone. It also made a few acquisitions in order to consolidate its position in mobile technologies.4

In 2015, Baidu entered into a deal with US-based technology company Microsoft under which Baidu became the default homepage and search for Microsoft browser in China.

As of 2017, Baidu was the largest search engine in China with a market share of 76.05%. It was also the second largest search engine in the world. It was ranked among the largest internet companies in the world, and was considered to be a leader in AI. Its webcrawler known as spider was capable of trillions of web pages.

Its AI based network, Baidu Brain scored 97% in voice recognition and 99.7% accuracy in facial recognition. It also operated a cloud arm, Baidu Cloud, through which it offered AI, Big Data and Cloud Computing services to customers.

For the year ending December 2016, its total revenues were RMB 70.549 billion. *(Refer to Exhibit I for the Financials of Baidu)*

The Big Data Ecosystem

Baidu's top priority was to build a robust infrastructure compatible with highly advanced features in big data application. In 2000, after the launch of Baidu Search, it released various internet applications at regular intervals. The applications helped in increasing Baidu's traffic on its various product platforms. Many of its tools became the most sought after and most popular internet apps in China. They included Baidu News, Baidu Maps, Baidu Post Bars, Baidu

Knows, Baidu Space, and Baidu Baike. *(Refer to Exhibit II for Various products and services by Baidu).* Each internet tool brought more users to Baidu's database.

Realizing that traffic was growing incessantly and that the database size was increasing exponentially, and inspired by its industry peers, Baidu started building its big data infrastructure. In 2016 inspired by Google's big table, Baidu built its proprietary tool Tera5, an open source-based, internet scale database. Tera was a powerful data storage system that could easily scale to petabytes across numerous commodity servers, with several high-end features like Linear and modular scalability, automatic and configurable sharding, and MVCC (Multi-version Concurrency Control [c]) to name a few. This was the beginning of Baidu's foray into the big data space. *(Refer to Exhibit III for more about Big Data)*

By the middle of 2012, Baidu had grown to become China's largest internet company, and the 4th most trafficked website in the world.[6] With each feature boasting millions of users, Baidu's user address book had more than 300 million users, enough to kick-start the big data initiatives.

Baidu's big data vision was driven by three core strategic functions which were classified into three functional components – .Open Cloud, Data Factory, and Baidu Brain.

In March 2012, Baidu declared its Cloud development strategy[d] by launching Open Cloud on a Developer Center Website[7]. Open Cloud was designed to facilitate ultra-large-scale distributed computing and storage capacities. It was specifically designed to address the issues of large data storage and processing. Open Cloud was characterized by features like a high CPU utilization rate, high flexibility, and cost-effectiveness, low energy consumption and high storage density, cutting-edge graphics processing unit, etc. As a part of its strategy, Baidu's Open Cloud initially provided each user 15 GB of free space, which was increased to 2 terabytes in a year from the launch. Baidu's Open Cloud offered functions like cloud storage service, client software, file management, resources sharing, and Third Party Integration. Baidu's Cloud platform was equipped with cutting edge tools like

[c] Multiversion concurrency control, is a concurrency control method commonly used by database management systems to provide concurrent access to the database and in programming languages to implement transactional memory.

[d] Cloud computing enables ubiquitous access to shared pools of configurable system resources and higher-level services that can be rapidly processed over internet with minimal management effort

Personal Communication Services, Site App applications[e], cloud MTC[f], Location-based services, etc.

Data Factory was the second component in Baidu's Big Data Engine. Data Factory essentially organized and processed massive amounts of data that was at the TB level or even greater. Data Factory supported hundreds of terabytes of heterogeneous data query, addressing a variety of business scenarios.

Lastly, Baidu Brain was to facilitate the application of algorithms to the data. While the data structure and processing were managed by the first two components, Baidu Brain opened Baidu's floodgates to the field of AI. Baidu Brain brought a host of futuristic applications like Machine Learning capabilities and Deep Learning, which were applied in its research areas like speech, image, and text recognition. Baidu Brain enhanced the scope of intelligence and application of Baidu's big data initiatives. It played a significant role in carrying out sharp analysis, deep learning, swift processing, and multiple-utilization of large datasets. With about 200 neural network parameters, Baidu Brain was the world's largest AI-based neural network[8].

With its ultra-modern big data engine, in 2012, Baidu kicked off its journey to harness its strengths in big data. Strong infrastructure offered a healthy platform to exploit the power of big data in numerous ways. In addition to its indigenous infrastructure, it also capitalized on the features of open source platforms like MongoDB, which allowed Baidu to attain strength and scalability in its big data initiatives.

Research Infrastructure

Baidu made no compromises in building the fundamental infrastructure to realize its big data vision. It built three state-of-the-art research laboratories in Silicon Valley. In 2013, it built Deep Learning Institute with its focus on certain areas – image recognition, machine learning, robotics, human-computer interaction, 3D vision, and heterogeneous computing[9].

In 2014, Baidu established the Silicon Valley Artificial Intelligence Laboratory and the Beijing Big Data Lab. The laboratories were driven by the mission of developing AI technologies which would impact at

[e] SiteApps is a platform to install apps and to optimize a site easily

[f] A leading provider of both cloud-based and on-the-ground I.T. services for small business

least 100 million people. The infrastructure was equipped with a powerful combination of deep learning, large datasets, and high performance computing.

By 2016, Baidu had developed more than 100 different products, which included messaging, photo sharing and facial recognition, metadata storage for Cloud, Baidu Maps, and Social forums. Each of the 300 million users was allocated 2 TB of free storage.

Strategic Levers

With high-end research laboratories and robust big data infrastructure, Baidu, after its stupendous success as a search engine, was all set to make its entry into the fast emerging field of big data. Li, strategically classified Baidu's big data vision into three clusters – Tiansuan, Tianxiang, and Tiangong.

Tiansuan, in principle, integrated large datasets with AI. It offered services that encompassed collection, storage, analysis, and application of data. It gave its clients easy access to machine learning and deep learning technologies to build their own services. Tiansuan came with numerous built-in solutions targeting segments such as healthcare, marketing, finance, etc.

Tianxiang was a multimedia platform that enhanced user experiences in online interactions like live streaming, music, movies etc. It was essentially a high-end application of AI in online data and content, equipped with advanced technologies like Media Cloud Transcoder, Video-On-Demand, Optical Character Recognition, etc. Its various offerings included video, image, and document processing services, speech analysis, live streaming, and face recognition.

The exclusive focus of the third strategic lever, Tiangong, was the Internet-of-Things (IoT). It integrated its cloud applications with internet connected devices, and offered a range of IoT services for users in a cost-effective way. Tiangong essentially targeted five industries for its services – Manufacturing, Logistics, Vehicles, Energy, and O2O retail.

By 2014, with state-of-the-art research laboratories, robust big data infrastructure, and a clear vision and well-crafted strategic levers, Baidu was all set to offer competitive products and services broadly encompassing three avenues viz. Data Analytics, AI, and the Internet-of-Things.

Data Analytics

By 2016, Baidu was registering 6 billion computer searches and 4 billion mobile searches per day. It harnessed the colossal amounts of data it was gathering daily through its search platforms and rolled out its Data Analytics solutions as per its first strategic lever – Tiansun.

Baidu classified its data into four broad categories viz. Social Data, Industry Data, Business Data, and Baidu Data *(Refer to Exhibit IV for Baidu's Big Data Process Flow).* The data from various categories was processed through its Big Data engine components, viz. the Open Cloud, Data Factory, and Baidu Brain, and the output was generated in the desired manner. The final products of data analytics could broadly be classified into Analysis and Decision, Data and Statistics, and Platform and Tool. The potential users of the final output were classified as Macro Industries, Medium Industries, and Micro Industries *(Refer to Exhibit V for Baidu's Big Data Output – An Illustration).*

Some of the analytics-based products that Baidu released were Baidu Economic Index Trends, Climate Index, Macro-economic Index, predicting tourist flow, employment trends, and support solutions for policy-making and decision making. Baidu's services supported various governments and business bodies like the Ministry of Transport, the State Food and Drug Administration, the Ministry of Education, and many more *(Refer to Exhibit VI for Baidu's Big Services to Government of China).*

Applying Big Data

Payment Gateways

China's cross-border ecommerce segment was expected to grow at 15 per cent annually[10]. Around 2010, as a result of the rise in domestic as well as cross-border e-commerce segments, the online payment trend in China was on a steep rise. The Chinese payment gateways witnessed fierce competition. The Goliaths of the payment gateways, Alibaba (Alipay) and Tencent (WeChatpay), dominated the payment segment. It was almost a duopoly with a 54% share of mobile transactions being claimed by Alipay and 40% by WeChatpay11. Baidu Wallet – Baidu's proprietary online payment platform, was launched in 2014. By 2016, being a new and late entrant into the e-commerce segment, Baidu Wallet was only the 10th-biggest payment platform in China, with over 100 million activated accounts to its credit. By the end of 2016, Baidu claimed about 0.4 per cent of the Chinese mobile payments market.

In July 2017, in a strategic move, Baidu partnered with PayPal to grab a piece of China's payments market. The alliance gave Baidu access to more than 17 million PayPal merchants, allowing Baidu Wallet users to exercise online purchases outside China by linking their Baidu Wallets to PayPal globally. The alliance was expected to give a boost to Baidu Wallet in the wake of tough competition from Alipay and WechatPay. It also promised Baidu a cross-border consumer experience, beyond China. According to Dan Schulman, President and CEO, PayPal "With Baidu, PayPal will be able to offer millions of international merchants increased access to millions of consumers in China who can search, shop and discover the world through Baidu and Baidu Wallet."[12]

Advanced Data Technics

International Data Corporation[g] (IDC) predicted that by 2025, the data on earth would grow to 163 zettabytes – 10 times the 16.1 ZB data generated in 2016. It also indicated that by then, about 20 percent of the data would be critical and about 10 percent would be hypercritical to the daily lives of human beings. This suggested the criticality of big data and the sensitivity of data analytics in times to come. IDC suggested that effective storage, processing, analysis, and integration of data with AI and IoT would determine the very existence of enterprises of the future.

In order to make its data storage, processing, and integration systems more robust, Baidu signed a contract with Seagate Technology in September 2017 to work toward the fields of big data analysis and advanced storage system development and implementation. The contract promised technical cooperation from Seagate for Baidu's business needs, facilitating the development of customized products. The contract assured that new products and technologies would be introduced into Baidu at an accelerated pace. The terms of the pact promised significant benefits with regard to a host of business aspects like optimized supply chain models, advanced development of cloud computing, and big data domestically and globally.

Managing Brands

Baidu offered its big data tools and services to enterprises and businesses seeking to manage and protect their brands. It helped high-end brands combat counterfeits and fraudsters who misused their

[g] IDC (International Data Corporation) is a provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets.

brands online. Baidu's data analysis techniques made it easy for them to protect their copyright on the internet. Baidu's services like Baidu Index, Baidu Sinan, and Baidu Zeitgeist facilitated brand awareness, optimizing search engines, and rendering social presence to enterprises that leveraged those services.

Baidu's data analytics offered advanced methods of understanding consumer behaviors, which helped various enterprises in accessing more customers and building a stronger brand. Baidu identified typical consumer behaviors in four patterns represented by the acronym SIVA (Solution, Information, Value, and Access) based on their search commands. Solution searches indicated consumers who attempted to solve a specific problem that addressed their immediate needs. Information searches suggested such consumers who looked for a specific brand value consumers were the ones who were sensitive to cost and quality of several different brands access customers prioritized their convenience and ease while making the purchase.

Such innovative approaches to understanding customers and integrating the resulting frameworks into Baidu's big data processors helped in managing, protecting, and building the brands of various business enterprises who were Baidu's clients.

Tracking Tourism Sector

With its seamless big data interventions, Baidu also left its footprint on the Chinese tourist sector. It used its big data techniques to map the movements of the Chinese population touring other parts of the country during holidays. Baidu Maps had about 200 million registered users whose movements were tracked by its location-based technology. About 3.5 million users requested location positions every day[13].

Artificial Intelligence

While Baidu was successful in most of its data analytics solutions, it was still not at the stage where it could rub shoulders with the behemoths – Alibaba and Tencent. It was evident from the statistics around the popular 3V of big data (Volume, Velocity, and Variety) that Tencent and Alibaba had transformed their companies to suit mobile, leaving Baidu behind[14]. Baidu, despite its huge data sets, modern research techniques, and data tools, had been struggling to catch up

with the two tech-giants. Owing to the walled garden approach[h] of the content and mobile leaders, Google and Apple, Li admitted that Baidu had slowed down in the race to enable people to create content[15].

However, Li's bet to make it happen in big data landscape hinged on Baidu's potential leadership in AI, as a part of its second strategic lever, Tianxin. Ya-Qin Zhang (Zhang) President, Baidu, who also headed Baidu's AI initiatives, said "AI is the single most transformative force of our time; it is changing everything we do."[16] Zhang believed that though search via keywords through smartphones and computers was prevalent at that point of time, in the future, voice and image recognition features would gain more prominence, and had the potential to be a substitute for keyword-based search. Thus, Zhang saw immense potential for AI and huge scope for Baidu to lead the global AI revolution. According to him, AI technologies and IoT devices would determine the success of any technology company in future. To him, it was all about making information access easier for the consumer, and nothing could match AI and IoT in making that happen. Baidu mobilized the lion's share of its resources to build an ecosystem that supported AI initiatives.

Baidu invested liberally in AI. In 2017, it invested 15% of its revenue, which was about $1.5 billion, in research and development, all on AI. This was most likely the highest such investment among all Chinese companies[17].

Strategic alliances and collaborations were key to Baidu rolling out its AI initiatives. In 2016, the company partnered with the Chinese government to build the country's first national AI research laboratory in Beijing. In early 2017, Baidu acquired Kitt.ai- a US-based AI startup. Strategic alliances were the key approach that Baidu took to drive AI initiatives at an accelerated pace. In July 2017, Baidu announced a tie-up with NVIDIA to accelerate its AI initiatives viz. self-driving vehicles and AI home assistants. The plan included integrating NVIDIA's various platforms like Volta™[i] GPUs, DRIVE™ PX[j], NVIDIA SHIELD™ TV[k],

[h] On the Internet, a walled garden is an environment that controls the user's access to Web content and services. In effect, the walled garden directs the user's navigation within particular areas, to allow access to a selection of material, or prevent access to other material.

[i] Graphics cards that are set to house the next generation of GPUs in computer games.

[j] A series of computers aimed at providing autonomous car and driver assistance functionality powered by deep learning.

[k] An advanced living room streamer built for IoT applications.

NVIDIA HGX™[l], TensorRT™[m] etc. with Baidu's AI components like Baidu Cloud, PaddlePaddle, and DuerOS to enhance the company's various AI initiatives.

Baidu had also been offering its cloud services to enterprises that wanted to reap the benefits of AI but did not have the necessary infrastructure. Most of its AI systems were made open source, allowing business enterprises to make use of its systems at a very reasonable cost. In November 2017, Baidu entered into a contract with Xiaomi to address the requirements of IoT and AI with emphasis on voice recognition, deep learning, and computer vision, robotics, Augmented Reality (AR), Virtual Reality (VR), and self-driving cars[18]. In December 2017, Qualcomm Technologies signed a pact with Baidu to work on voice-enabled solutions. The collaboration was expected to cater to the users' requirements for more comprehensive and better quality information and services needs by integrating Baidu's conversation AI system – DuerOS – with the OEMs of smartphones by Qualcomm. Two of Baidu's most promising applications were Apollo, its operating system for driverless cars, and Machine Learning applications based on its conversation-based AI platform, DuerOS.

Apollo – Autonomous Driving Systems

Autonomous cars comprised a segment that was witnessing fierce competition among global tech giants like Google, Tesla, BMW, and many more. Top thinkers were of the view that driverless cars would revolutionize the automotive sector in future. In a report in 2017, McKinsey estimated that autonomous driving could contribute $1.5 trillion in net new revenues to the automobile market by 2030[19]. It also projected that by 2030, they could account for up to 15% of total new car sales.

In July 2017, Baidu teamed up with Microsoft to invest in an autonomous driving project called Apollo over the next three years. In September 2017, Baidu launched a $1.52 billion autonomous driving fund called 'Apollo Fund' to support and accelerate 100 Autonomous driving projects, over the next three years. Li expected to see mass production of driverless cars by 2021. Years. In order to accelerate its pace, Baidu had even opened up the project, which was initially internal to Baidu, to external stakeholders. By September 2017, Baidu had on

[l] A reference architecture that standardizes the design of data centers accelerating AI in the cloud.

[m] A high-performance deep learning inference library for production environments.

board 70 partners representing various fields in the auto industry, which included companies like Ford, Daimler, Nvidia, Intel, Microsoft, and Velodyne. The partners contributed technological resources to the Apollo platform to support the critical features and functions of vehicles.[20]

Baidu's vision for Apollo had received implicit consent from the regulators and the government too, as the vision was in accordance with the government's goal to surpass the US in the application of modern and emerging technologies. Baidu's vision for most of its technology initiatives including the driverless cars was in accordance with the 'Made in China 2025' policy initiative that required domestic firms to claim dominance over strategically critical industries, which included AI and Robotics. The project was popular among the Chinese people too. A poll conducted in August 2017 by eMarketer suggested 82% of Chinese consumers wanted to ride an autonomous vehicle21, which indicated possible success for Baidu in its driverless car mission.

Machine Learning

In the last several years of its research on AI, Baidu had made significant headway in Machine Learning. It had been integrating into its core services machine and deep learning technology, self-teaching, and neural net technology to reap the benefits in sectors like finance and healthcare. In 2014, Baidu launched a speech recognition technology called 'Deep Speech'[22], which recognized and interpreted voice input in noisy environments without distorting accuracy.

Zero-shot learning – the ability to solve a learning task without having received any previous examples of solving such a task[23] – was a feather in Baidu's cap. Zero-shot learning was a major breakthrough for Baidu in machine learning through language command and processing.

Baidu's machine learning applications boasted of some of the best features in its much anticipated Apollo driverless cars project too. Baidu's car showed a rotating roof-mounted, sensor which gathered and disseminated information about the car's immediate environment. Subsequently, the information was analyzed by machine learning algorithms to guide the car through the best and safest route.

Augmented Reality was another area where Baidu applied machine learning which was focused on marketing and advertising. The initiative brought Baidu global clients like KFC and L'Oreal.

Baidu had been applying Machine Learning and deep learning technology in healthcare too as a part of its Baidu Doctor project toward building a chat program that could diagnose illnesses with the help of the patient's voice input. It had in its pipeline a medical robot that could assist and conduct medical procedures.

Raven was another proprietary hardware of Baidu. Raven was an AI-powered hardware that comprised a smart speaker, a display controller, and a personal robot characterized by six human-like joints. Raven was a personal digital assistant with AI embedded into its functions. With this, Baidu joined the league of Google and Amazon who developed their proprietary hardware as platforms to support their voice-based AI assistants.

Xiaoyu Zaikia (Little Fish) was another home-robot by Baidu, which could turn its "head" to listen to the speaker, gather inputs, and facilitate the control of smart home equipment and assist in instructions like online shopping with the help of its natural language processing interface.

Internet of Things

In a survey-based report, IDC predicted that the IoT market would hit $1.4 trillion by 2021[24]. In September 2017, it was estimated at $800 billion. The third strategic lever of Baidu's Big Data Strategy – Tiangong was about expanding its footprint in the IoT segment.

Baidu's DuerOS platform essentially aimed at helping developers in creating voice-enabled devices that could hear, understand, and follow the users' commands. Baidu had partnered with more than 130 enterprises to roll out the DuerOS applications globally. The technology was already integrated into devices like smartphones, televisions, and home appliances. According to Kun Jing, DuerOS lead at Baidu, "We want to enable everyone on this planet to build DuerOS into devices with a very low barrier."[25]

Baidu's collaboration with Qualcomm included developing an AI-based voice assistant for IoT and mobile devices. Its alliance with Xiaomi included integrating its IoT infrastructure with the latter's mobile devices. The partnership was expected to apply conversational AI to Xiaomi's mobile and other devices and gain a better understanding of how the technology operated in the real world[26]. In November 2017, Baidu joined hands with Philips to integrate Philips devices with its DuerOS conversational AI platform[27]. The initiative would launch connected lighting in China, making lighting devices work seamlessly in

unison with all smart devices from smart speakers to wearable devices, controlled through voice commands The initiatives promised to give a thrust to Baidu's applications in the Chinese smart home market which, by 2020, was estimated to be RMB 350 billion[28].

The National Mission

The Chinese government, as a part of its 'Made in China Policy', declared an ambitious strategy to propel the country toward global technology leadership by 2030. In November 2017, the government officially recruited Baidu along with Alibaba and Tencent, the trio popularly referred to as BAT[29] apart from iFlyTek[n]. China's Ministry of Science and Technology recognized Baidu as one of the first members of the AI national team.

The four companies were entrusted with building "open innovation platforms" in four different fields by leveraging their respective strengths in AI application. (Refer to Table I for the AI Application areas each company was involved in)

Table I

AI Application areas

#	Company	The AI Application Area
1	Baidu	Autonomous driving
2	Alibaba	Project 'city brains': improving urban life, including smart transport
3	Tencent	Medical diagnosis
4	iFlyTek	Voice intelligence

Compiled from various sources

The government's decision to include AI in the national development strategy and to select Baidu as one of the critical players towards the national mission looked immensely encouraging and promising for the company. According to Raymond Wang, a partner at Roland Berger. "*It is the first time some of the country's biggest companies have been named in such a strategy. Government's blessing could give Baidu a leg-up when it came to cooperating with carmakers on self-driving vehicles*."[30]

[n] iFlytek is a Chinese information technology company established in 1999. It creates voice recognition software and voice-based internet / mobile products

Social and Economic Impact

Baidu's big data initiatives were not confined to business and commercial utilities alone. Many of its applications demonstrated and promised to make an impact on several sensitive social and economic issues also.

Baidu's big data initiatives promised to help alleviate poverty. In a country with more than 50 million people living in poverty[31] Baidu had a critical role to play in making a positive impact on the economic conditions of the people. It, had been partnering with government in a nationwide drive to alleviate poverty.

Baidu formed a big data network by setting up various experimental units throughout the country. Its analysis around poverty was based on eight fundamental constructs that explained the living conditions of people – access to running water, toilets, kitchens, mobile internet, service facilities, road infrastructure, electric lighting, and financial services[32].

The initiative promised to help the government in implementing policies to alleviate poverty more efficiently. Baidu's intervention in Zhongyang county was a successful initiative as it had a profound effect on the lives of the people. With Baidu's guidance, the people of Zhongyang switched from basic agriculture practices to mushroom farming, which proved high yielding, and resulted in a marked rise in income levels.

Baidu also identified various factors that caused economic imbalance in the farmers' community, which included out-of-date information, lack of market access, inconvenient transportation, and overstocked products leading to poor sales and dwindling incomes. After analyzing the situation, Baidu launched the Baidu Support Plan, which helped farmers who had an overstock of goods to sell their excess produce online. Baidu also offered the farmers training in social media marketing and the basic resources to build offline sales channels for these farmers. The initiatives contributed to the alleviation of poverty.

Baidu targeted poverty through education too. It launched channels to improve the quality of education and the required facilities in poor counties. It offered the teachers in the counties an opportunity to partner with it, as a part of which it donated customized Baidu Education membership cards in counties with poor educational facilities. With the help of the cards the teachers could enhance their teaching skills by gathering superior quality information about education from Baidu Library. Baidu's strategy to improve educational

resources was highly inclusive. It also aimed to improve the conditions of the educational resources in the locations inhabited by minorities. Baidu Wenku[o], as a part of its public welfare plan called Elieen Public Welfare Plan, invited bilingual teachers belonging to Uyghur, Kazakh, Tibetan, Korean, and Mongolian origins to translate educational resources. The project that had several pre-defined phases, aimed at uploading tens of thousands of documents related to minority educational resources. As more users liked the online pages of the translated documents, more money kept flowing into the welfare plan @ one yuan per like. As of October 2017, the welfare plan had mustered more than $1 million.

Baidu's Dengue Baidu Search Index (DBSI) played a crucial role in responding to the epidemic of dengue more efficiently[33]. DBSI in combination with meteorological and demographic factors, led to the development of a predictive model for dengue fever. DBSI showed a positive and reliable linear relationship with dengue fever occurrence, leading to a better prediction capability.

Many times, at overcrowded locations stampedes occurred. Baidu's machine learning program demonstrated successful predictions about crowds gathering at certain locations, prompting precautionary measures to be taken to prevent deaths and injuries in stampedes.[34] Baidu Maps, which had more than 300 million active users by 2017, claimed a reliable correlation between number of map queries and the number of users in an area. When the number of queries from a location exceeded a predetermined threshold, it indicated a possible large crowd gathering and a potential threat of a stampede. This discovery led to a mechanism for triggering warnings much ahead of time when large crowds were expected to gather in a certain area. This machine learning based initiative promised to help in better event-managements and facilitate smooth flow of human traffic at times of large gatherings.

Baidu Index[p] promised to predict Chinese stock returns based on a host of constructs analyzing the investor behavioral patterns. The Search Frequency of Baidu Index (SFBI) was found to predict next day's price changes. The trading strategies based on the SFBI index were found to deliver the desired performance, predict stock returns, and offer guidance in asset pricing and risk management[35].

[o] An open platform constructed and operated by Baidu Company for Internet users to share online documents.

[p] The internet information, on Chinese stock returns.

Baidu's AI applications influenced the food sector too. In alliance with KFC[q], it built an physiognomy[r] application. The facial-recognition technology backed by AI analyzed customers' faces and evaluated constructs like age, gender, and facial expression to determine their choice of order. The restaurant stored the picture of each customer while receiving an order, so that they could predict the customer's order during their next visit to the store. The initiative prompted KFC to store an image database of customers and this helped it in gaining a much deeper understanding of consumer behavior, and in offering a much more customer-centric service.

In 2014, Baidu, with the help of its AI applications, successfully predicted that Germany would win the Football World Cup 2014[36], based on five critical factors viz. team strength, home court advantage, recent performance, overall World Cup performance, and bookmaker odds.

The host of Cloud technology applications from Baidu in the field of data analytics, AI, and IoT promised to bring about a positive, visible, and significant change in the social, economic, and business spheres of China and beyond.

The Road Ahead

In 2011, Baidu was China's most valuable internet company. It had a market of US$46 billion in 2011 which increased to US$64.6 billion by 2017 – a 40 per cent rise. However, Baidu had to face fierce competition from its domestic peers Alibaba and Tencent. In 2014, Alibaba went public on the New York Stock Exchange and grew at a rapid pace to a market cap of US$300 billion by 2017.

Competition was cut-throat, both on the global and the domestic turf. Baidu was constrained on all sides by a tough fight for market share. Having responded late to the big data market, Baidu lagged behind in capitalizing on the cloud technology based opportunities. According to Philip Beck, an early investor in Baidu, "Somewhere along the line, Baidu forgot that they were an advertising platform like all internet giants."[37] The e-commerce Goliaths Alibaba and Tencent were nimble in responding to the emerging data analytics sector and grabbed a significant chunk of the market. Other players like JD.com and iFlyTek were also constantly threatening to overtake Baidu.

[q] KFC previously known as Kentucky Fried Chicken is an American fast food restaurant chain. KFC was the largest restaurant Chain in China as of 2017

[r] Predicting behavior based on physical traits is called physiognomy

Nevertheless, over a period of time, Baidu emerged as one of the top global cloud-based technology companies, which had been foraying at a brisk pace into applications of high-end technology viz. big data analytics, AI, and IoT. Its bold and ambitious initiatives like Apollo and cutting edge technology platforms like DuerOS offered Baidu a huge opportunity to dominate the AI segment on both the domestic and global fronts.

With its cutting edge cloud technologies and infrastructure, Baidu had been moving from strength to strength toward building ultra-modern predictive frameworks, futuristic devices, and paradigm shifting solutions to various individual, business, economic, and systemic challenges. With a timely shift in its strategy from 'all mobile' to 'all AI', Baidu demonstrated noticeable results in its efforts to catch up with the technology trends.

However, its colossal investments in research and development were putting immense pressure on Baidu to make its projects profitable sooner rather than later. It remained to be seen whether Baidu would succeed in its ambitious AI projects to emerge as a global leader in AI-based products and services.

Exhibit I

Baidu Financial Data

(Year ended March 31)	2012	2013	2014	2015	2016
					(in RMB 000s)
Revenue					
Online Marketing Services	22,245,643	31,802,219	48,495,215	64,037,006	64,525,115
Others	60,383	141,705	557,103	2,344,723	6,024,249
Total Revenue	**22,306,026**	**31,943,924**	**49,052,318**	**66,381,729**	**70,549,364**
Operating Costs and Expenses					
Cost of revenue	(6,448,545)	(11,471,839)	(18,885,450)	(27,458,030)	(35,278,945)
Selling, General and Admn	(2,501,336)	(5,173,533)	(10,382,142)	(17,076,383)	(15,070,586)
Research and Development	(2,304,825)	(4,106,832)	(6,980,962)	(10,175,762)	(10,150,753)
Total Operating Costs and Expenses	**(11,254,706)**	**(20,752,204)**	**(36,248,554)**	**(54,710,175)**	**(60,500,284)**
Operating Profit	**11,051,320**	**11,191,720**	**12,803,764**	**11,671,554**	**10,049,080**
Interest Income	866,465	1,308,542	1,992,818	2,362,632	2,341,631
Interest Expense	(107,857)	(447,084)	(628,571)	(1,041,394)	(1,157,562)
Income (loss) from equity investments	(294,229)	22,578	(19,943)	3,867	(1,025,727)
Other income net,incl. exchange gains	449,738	140,951	336,338	24,909,964	4,301,785
Income before income taxes	**11,965,437**	**12,216,707**	**14,484,406**	**37,906,623**	**14,509,207**
Income taxes	(1,574,159)	(1,828,930)	(2,231,172)	(2,474,377)	(2,913,594)
Net Income	10,391,278	10,387,777	12,253,234	35,432,246	11,595,613

Source: Baidu – SEC Filings

Exhibit II

Baidu – Products and Services

Search Products	Web Search, Image Search, Video Search, News, Web Directory, Hao123.com, Dictionary, Top Searches & Search Index, Open Platform
Social Products	Post Bar, Space, Album
UGC-based Knowledge Products	Knows, Encyclopedia, Wenku, Experience
Location-based products & services	Maps, Groups Buy Directory, Travel
Music Products	Baidu Music, Baidu FM, TT Player
PC Client Software	Browser, Input Method Editor, Toolbar and Baidu Companion, Baidu Hi, Media Player, Reader
Mobile Related Products & Services	Mobile Search, Cloud Smart Terminal platform, Mobile browser, Palm, Mobile phone input method editor, Contacts, Netdisk, Photo Wonder, Wallpaper, Desktop,One-Click-Root, Voice Assistant

Contd…

Contd...	
Products & Services for Developers	Developer Center, Personal Cloud Storage, Baidu App Engine, TS Browsing engine, Mobile Test Center, LBS Open Platform, Baidu Webmaster Platform, Statistics, Share
Other Products & Services	Qunar, IQiyi, Baijob, BaiduPay, Games, Search & Store, Application Store, Ads Manager, Data Research Center, Sky, Senior citizen search, Search for Visually impaired, Patent Search, Translation, Missing Person search site

Source: www.ir.baidu.com

Exhibit III

About Big Data

According to the McKinsey Global Institute, "Big data refers to data sets whose size is beyond the ability of typical data software tools to capture, store, manage and analyze." Such a data is collected through devices and technologies such as credit cards and customer loyalty cards, social media from the internet, Wi-Fi sensors, and electronic devices.

Advancement in technology over the years had a huge impact in the way data was collected and disseminated. Earlier data was more structured, as it was captured in pre-determined formats and could be fixed into particular fields in databases. Thus it was easy to analyze and compare. With the sources of data becoming varied, the data became unstructured, and started to include other forms like images, audio, and video files, data from sensors, geospatial data, radio frequency identification, social media data, etc. Such data did not possess a particular format or length, was difficult to fit into specific fixed formats, so it was difficult to store and archive. To analyze and process such data traditional tools like relational databases and desktop software for statistics and visualization were inadequate, and needed parallel software running on thousands of servers.

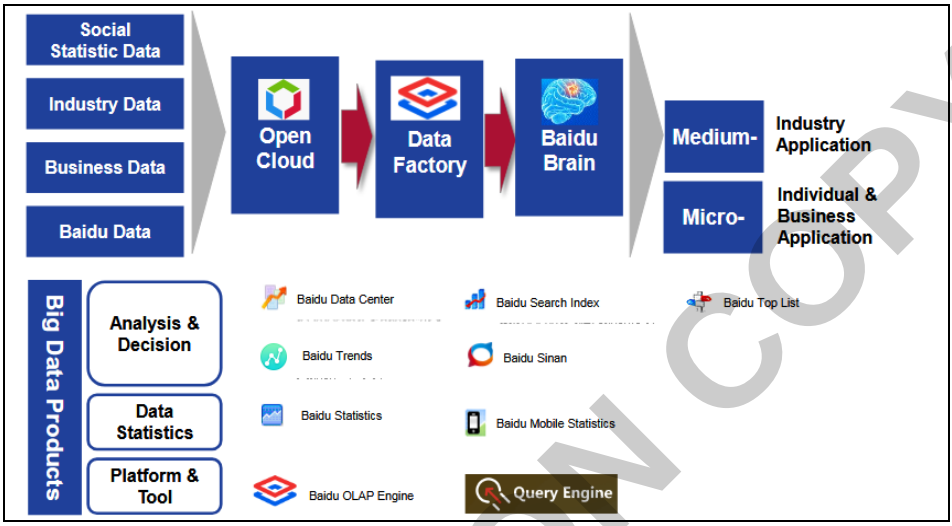
Business intelligence was widely used to analyze such data. It involved gathering, storing and providing access to data. this was done through decision support systems, online analytical processing, forecasting and statistical analysis. Business intelligence identified the data patterns of the past data. Data analytics looked at how to use the data in the future. Big data analytics involved using algorithms to drive decision making. Those used for predicting behavior are called predictive analytics. Big data analytics helps in finding new patterns and co-relations.

Experts believed that any organization imbibing big data in its activities becomes innovative, and will be able to deliver more effectively, which would result in more sales and more profits.

Compiled from various sources

Exhibit IV

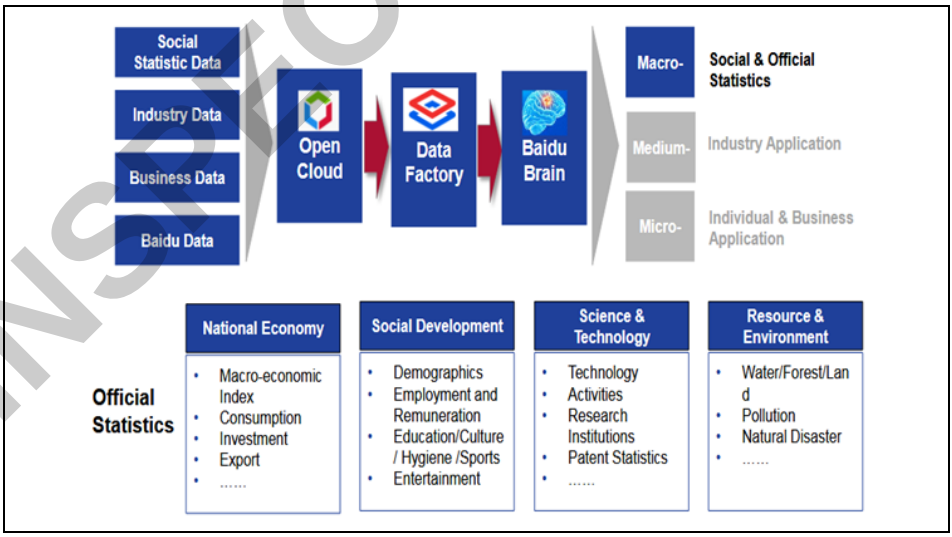
Baidu's Big Data Process Flow



Source: www.baidu.com

Exhibit V

Baidu's Big Data Output- An Illustration



Source: www.baidu.com

Exhibit VI

Baidu's Big Services to Government of China

Source: www.baidu.com

End Notes:

1 Nat Levy, "Baidu President calls AI the 'Single most Transformative Force of our Time'," www.geekwire.com, October 09, 2017

2 Craig Smith, "80 Amazing Baidu Statistics and Facts," https://expandedramblings.com, December 23, 2017

3 https://www.statista.com/statistics/254266/global-big-data-market-forecast/

4 These included 91 wireless (appstore), iQuiyi (online video platform), PPStream (online video streaming)

5 https://github.com/baidu/tera

6 "Multi-Tenant Cluster Storing 200+ Billion Documents across 160 Shards," www.mongodb.com, July 08, 2016

7 www.developer.baidu.com/en/

8 Eileen McNulty, "10 Machine Learning Experts you need to Know," http://dataconomy.com, September 16, 2014.

9 http://research.baidu.com/institute-of-deep-learning/

10 Yingzhi Yang, Yuan Yang, "PayPal joins Baidu in fight for China's Payments Market," www.ft.com, July 27, 2017

11 "Is WeChat Pay taking over Alipay?" www.forbes.com, June 20, 2017. (originally appeared on Quora)

12 Yingzhi Yang, Yuan Yang, "PayPal joins Baidu in fight for China's Payments Market," www.ft.com, July 27, 2017

13 "Baidu's Big Data Tracks Chinese Travel," http://bigdata-madesimple.com, January 28, 2014

14 Jessi Hempel, "Inside Baidu's Bid to Lead the AI Revolution," www.wired.com, December 06, 2017

15 Andrew Nusca, "How Baidu Plans to Win the Artificial Intelligence Market," http://fortune.com, October 18, 2017

16 Nat Levy, "Baidu President calls AI the 'Single most Transformative Force of our Time'," www.geekwire.com, October 09, 2017

17 Andrew Nusca, "How Baidu Plans to Win the Artificial Intelligence Market," http://fortune.com, October 18, 2017

18 Jonathan Shieber, Jon Russell, "Baidu and Xiaomi are Working Together on Internet of Things and Artificial Intelligence," https://techcrunch.com, November 27, 2017

19 Rachel Gunter, "Behind the Microsoft-Baidu Cloud Partnership," https://beta.marketrealist.com, October 02, 2017

20 Rachel Gunter, "Behind the Microsoft-Baidu Cloud Partnership," https://beta.marketrealist.com, October 02, 2017

21 Yue Wang, "Why China's Baidu Could Beat Google In The Race For Self Driving Cars," www.forbes.com, November 22, 2017

22 Josh Hortwitz, "Baidu's Research Lab Announces 'Deep Speech' Recognition System," www.techinasia.com, December 19, 2014

23 Aaron Kurmins, "Baidu AI achieves 'Zero Shot Learning' Ability using Natural Language," www.extremetech.com, April 06, 2017

[24] Barb Darrow, “How the Internet of Things Can Be Leveraged in Health Care, Retail, and Manufacturing,” http://fortune.com, September 27, 2017.

[25] Khari Johnson, “Baidu’s Ambitious Plan for ‘Everyone on this Planet’ to use its Conversational AI DuerOS,” www.venturebeat.com, August 19, 2017

[26] Barb Darrow, “How the Internet of Things Can Be Leveraged in Health Care, Retail, and Manufacturing,” http://fortune.com, September 27, 2017

[27] “Philips Lighting Partners with Baidu to bring Connected Lighting to Smart Homes in China,” www.newsroom.lighting.philips.com. November 16, 2017

[28] “Philips Lighting Partners with Baidu to bring Connected Lighting to Smart Homes in China,” www.newsroom.lighting.philips.com. November 16, 2017

[29] Meng Jing, Sarah Dai, “China Recruits Baidu, Alibaba and Tencent to AI ‘National Team’,” www.scmp.com, November 21, 2017

[30] Meng Jing, Sarah Dai, “China recruits Baidu, Alibaba and Tencent to AI ‘National Team’,” www.scmp.com, November 21, 2017

[31] “Baidu’s Great Contribution-Using Big Data help Alleviate Poverty, www.chinadaily.com.cn, October 26, 2017

[32] “Baidu’s Great Contribution-Using Big Data help Alleviate Poverty, www.chinadaily.com.cn, October 26, 2017

[33] “Dengue Baidu Search Index data can Improve the Prediction of Local Dengue Epidemic: A Case Study in Guangzhou, China,” http://journals.plos.org, March 06, 2017

[34] Zeb Soo, “Chinese Search Engine Giant Baidu Develops AI Algorithm to Predict Crowds, Pre-Empt Stampedes, www.scmp.com, March 24, 2016

[35] “Baidu Index and Predictability of Chinese Stock Returns,” https://link.springer.com, March 24, 2017

[36] Steven Millward, “Baidu Claims its Data is More Accurate than Microsoft, Paul the Octopus, at World Cup Predictions,” www.techinasia.com, July 09, 2014

[37] Yingzhi Yang, Yuan Yang, “Baidu bets its Future on AI Revolution,” www.ft.com, August 30, 2017

Case 9

Alibaba's Big Data – A Global Game Changer

K B S Kumar and Indu Perepu

"Alibaba is not a retail business, we are a data business."[1]

– Jack Ma, Founder and Executive Chairman of Alibaba Group.

The world's biggest online commerce company, China-based Alibaba Group Holding Limited (Alibaba), had a market share of 75% in Chinese online retail as of 2017 and was touted as the fastest growing tech-giant in the world. Alibaba, started in 1999 by an English teacher Jack Ma as an e-commerce company, went on to become an internet behemoth. It operated e-commerce platforms, each acting as a marketplace connecting buyers and sellers[a]. *(Refer to Exhibit I for different businesses of Alibaba).* Alibaba's retail marketplace platforms had 454 million active buyers, 1.5 billion listings and a Gross Merchandise Volume[b] (GMV) of US$ 547 billion as of March 2017.[2] According to estimates by the company, if Alibaba were a country, its economy would be 22nd largest in the world[3], (considering the GMV). *(Refer to Exhibit II for Annual Active buyers).*

Such a wide presence and usage generated massive data about the users. Alibaba collected data from the platforms in its system. The data was mined and analyzed to better understand the markets, consumers, and their preferences. Such data was shared with the vendors who were a part of the Alibaba system to help them improve

[a] Alibaba was active in the areas of shopping, mobile wallet, digital entertainment, travel booking, wholesale and retail marketplace, logistics, and financial services, to name just a few

[b] GMV is a measure of the total value of third party sellers' *transactions on the websites*.

their offerings and advertise to the right target audience. With the huge amount of data, Alibaba moved toward making itself a data-commerce company. The data helped Alibaba progress into cloud computing. Its success in cloud computing made the company provide its cloud computing and data management services to other organizations, institutions, and governments across the world.

Alibaba put the data it had captured to several uses. The data was used to improve its services and the predictive data generated was used to place before the consumers their choice of products. It was used to give loans at a very low processing cost, to eliminate counterfeit goods from the e-commerce chain, and even to detect fraudsters. The data was also put to use in its foray into physical retail as part of its 'new retail' strategy. Alibaba planned to use technology to spruce up the US$ 4.9 trillion retail industry in China by removing the barriers between online and physical retail through the use of big data.

According to Daniel Zhang (Zhang), Group CEO, Alibaba, *"Data is the blood of the new economy. We use data to refuel our business, and refuel the participants in the ecosystems to help our partners to do business easier anywhere in the fast changing digital world."*[4]

As Alibaba moved ahead and set its sights on going global and having a customer base of around 2 billion consumers by 2036, experts said big data would play a major role in its quest to become a global e-commerce superpower.

Inception and Growth

Alibaba was started by Jack Ma, an English teacher from Hangzhou, and seventeen of his friends in 1999. The founders always believed that "*the Internet would level the playing field by enabling small enterprises to leverage innovation and technology to grow and compete more effectively in the domestic and global economies.*"[5]

Alibaba started off with an English language website Alibaba.com, a global wholesale marketplace. It also started a website for domestic wholesale trade. During the year it raised US$ 5 million from a consortium of investors. In 2000, Alibaba raised funding of US$ 25 million from Softbank, Goldman Sachs, Fidelity, and other institutions. By 2001, it had one million registered users. In 2003, it founded Taobao a consumer e-commerce website on the lines of eBay.

In 2004, Alipay was started as a third party online payment platform. During the year, Alibaba raised US$ 82 million from different

investors. It was the largest private equity investment in China. (Alipay was rebranded as Ant Financial Services Group[c] in 2014).

In 2005, Alibaba formed a strategic partnership with Yahoo China and took over its operations. Yahoo invested US$ 1 billion in Alibaba for a stake of 40%. In 2007, Alibaba was listed on the Hong Kong stock exchange. In 2007, Alimama was launched as an online marketing technology platform.

In 2009, the Big Taobao strategy was launched to position Taobao as a one-stop e-commerce service provider to promote the wider use of e-commerce. In the next year, an independent web domain Tmall.com was launched. *(Refer to Exhibit III for Alibaba's Milestones)*

In 2014, Alibaba listed on the New York Stock Exchange. The US$ 25 billion issue was the largest US IPO of all time. After the IPO, Alibaba went on to become the fourth largest tech company in the world. It was ahead of US giants like IBM, Oracle, and Facebook.

By May 2015, Alipay had become the most frequently used form of payment, surpassing even cash in China. In first tier cities, 80% of the consumers used Alipay. As of 2016, websites related to Alibaba accounted for 75% of all the online shopping in China and around 10% of the total retail in the country. The vast ecosystem in which Alibaba operated included billions of orders, millions of sellers, and thousands of logistics companies, and worked seamlessly to provide products and services to customers.

The growth of Alibaba coincided with the Rise of China. In the 18 years of Alibaba's existence, China's per capita GDP grew by a compounded annual rate of 14%[d], which meant that per capita income grew from US$ 870 in 1999 to US$ 8100 in 2017. China also developed technologically, and became the most advanced mobile economy. Meanwhile, the country emerged as a 'manufacturer for the world'. With the growth of the internet in China, Chinese exports grew rapidly with several manufacturers and suppliers entering the market as also buyers. There was a need for a platform that brought them all together.

That was when Alibaba set its eyes on becoming a global internet company emerging from China. Right from the beginning, Alibaba looked at being a 24X7 global meeting platform for suppliers and buyers from across the world, and not just China. It followed a model where

[c] Ant Financial consisted of Alipay, Yu'wBao Money market fund, and a credit rating system Sesame.

[d] As against US which grew at 3% during the same period.

buyers used the site free and the sellers paid to get in touch with the buyers. Thus, Alibaba was able to attract 4.4 million registered users from 200 countries by 2009.

Alibaba, on its way to growth, took on several established, international players in the Chinese market. For example, Taobao, which was started in 2003, became highly popular and marched ahead of the global e-commerce giant e-Bay in China. Till then e-Bay had been the leader in the C2C market.

For the year 2016, Alibaba exceeded Gross Merchandise Volume of RMB 3.7 trillion[e] in China to become the largest retail economy in the world. *(Refer to Table I for Alibaba's GMV from 2015 to 2017). (Refer to Exhibit IV for Alibaba's financial details).*

Table I:

Alilbaba's GMV

(RMB Billion)

As of	Taobao Marketplace	TMall	Total
June 30 2014	342	159	501
September 30, 2014	380	176	556
December 31, 2014	494	293	787
March 31, 2015	381	219	600
June 30, 2015	427	246	673
September 30, 2015	438	275	713
December 31, 2015	563	401	964
March 31, 2016	449	293	742

Compiled from www.alibaba.com

The Making of a Big Data Behemoth

Capturing and Disseminating Data

In 2005, the transactions on Alibaba's websites were less than 10,000 per day. By 2007, they increased to around 500,000 per day. As its popularity grew, the number of transactions grew, reaching even 1 billion transactions per day during popular shopping days.

[e] 1 US$ ≈ RMB or Yuan 6.55

From the beginning, Alibaba captured all the data pertaining to the buyers, sellers, and other data about frequency of order, average order size, etc. This was captured from its different sources. *(Refer to Exhibit V for different internal sources from which Alibaba captured the data)*

The massive data generated needed systems that analyzed millions of transactions. Alibaba started capturing data using data platforms like RAC 2009 when data was about 100 TB[f] and Hadoop[g] in 2012 with data capacity of about 100[h] PB. With the increase in data, it started using Open Data Processing Service (ODPS) with data capacity of about 10 EB. According to the company, as of 2014, the server at Alibaba Data Platform Division had more than 100 PB of processed data. Every day, the websites continued to give huge amounts of data on monetary transactions, finance, social networking services, online services, map tools, etc.

Apart from its internal data from different sources, Alibaba also started using data from external bureaus, thus building an integrated data platform. At the same time, it focused on cyber security for all the transactions and used high security standards for the data lifecycle – data collection, use, transfer, sharing, and disposal. It ensured that individual data of the customers was not shared.

Alibaba also developed data mining capabilities to recover the constantly accumulating data. This helped it build detailed profiles of the customers and make their buying behavior forecasts.

The three key features that made Alibaba's data effective were: derivation of data from the shopping actions of the customers on its e-commerce platforms, which made it authentic; the data having around hundred attributes and being highly structured, which made it of high quality; having comprehensive data of several petabytes. *(Refer to Exhibit VI for more about big data)*

All the data that Alibaba collected was processed and the relevant sets of data were sent to the merchants, helping them come out with better products to target customers better, and to sell more goods through Alibaba. The same data was routed to manufacturers to help them change and tweak products based on demand. According to Scott Likens, head of PwC's analytics and emerging technologies business in

[f] 1PB (Petabyte) = 1024 TB (Terabyte); 1EB (Exabyte) = 1024 PB

[g] Hadoop is open-source software framework for distribution, storage and processing of data.

[h] 100 PB of data is equal to data of 100 million HD films

Hong Kong, "Now ecommerce giants can see exactly what to manufacture. You have manufacturing on demand because you have data on consumers."[6]

The data was also used to build a big data driven marketing platform called Alimama. Alimama connected all the data assets of Alibaba like demographic attributes, consumption data, physical location, browsing behavior, payment method, and social data of almost 730 million internet users in China. Combining e-commerce and brand marketing with big data, Alimama helped advertisers in digital marketing. *(Refer to Exhibit VII for more about Alimama)*

In 2009, Aliyun, the cloud services arm of Alibaba (which was later renamed AliCloud) was started. It was incorporated to handle the huge volumes of traffic and also to manage the data of the e-commerce platforms of Alibaba and Alipay. Aliyun had R&D centers and operations centers in Beijing, Hangzhou, and Silicon Valley. A big data cloud platform from Alibaba, MaxCompute, was released in 2013 and was used initially to support Alibaba's internal operations, including processing hi-definition product images and videos. When customers used mobile phones to access TMall or Taobao they were taken to their personalized landing pages and given recommendations for products. This was made possible due to the real time calculation of users' past shopping behavior that was done by MaxCompute. *(Refer to Exhibit VIII for more about Alibaba's cloud computing)*

Using the Expertise

Putting all the data it had collected to further use, Alibaba built personalized search and recommendation engines to predict the likes and intentions of users. By using billions of data samples and attributes effectively, Alibaba came out with machine learning models. It developed server architecture to handle large models and billions of parameters. These were distributed into large sets of servers for parallel computing. Implementation of the machine learning algorithms produced manifold benefits. This created the best match for a particular user in a particular context, and produced targeted advertisements, resulting in increased sales.

For the benefit of users, Alibaba also used Virtual Reality, where computer generated 3-D images along with Augmented Reality that gave a realistic view of the products. It used machine learning technologies including high-dimensional statistics, online learning, transfer learning, and deep learning to improve its e-commerce to meet customer demand.

Alibaba was able to tailor its content and products according to the behavioral characteristics of the users. The company studied user behavior data at different stages of their shopping and derived algorithms that helped it determine and narrow down the intention of the shoppers. An app then pushed the products that users would prefer and provided content to help them make purchasing decisions at a particular point of time. This process helped not only the merchants but also Alibaba to attract more digital advertisements, with the result it was able to garner the largest share in digital ad spend in China. According to Zhang, "*Sophisticated real-time personalized product recommendations, subscription-based content and short-form videos all helped to create a stimulating and fun journey for consumers across our retail marketplaces. We are not only meeting the demand of the customers, but also creating the demand.*"[7] *(Refer to Exhibit IX for Alibaba's businesses and flow of data)*

Singles' Day

Singles' Day was created by students in China in the mid-1990s as anti-Valentine's Day. Celebrated on November 11, it was a huge business day in China, with people who were single buying gifts for themselves and for other singles. Alibaba started celebrating the event as a major online shopping festival in 2009. This generated massive revenues, making the world sit up and take notice of the buying power of the Chinese.

Its popularity and the sheer size of the business that took place on that day could be gauged by the fact that the Singles' Day promotion that Alibaba carried out generated revenue of US$ 14 billion on a single day in 2015. This was equal to a year's revenue of Facebook. In 2015, Alibaba recorded over 175,000 transactions per second, with data of 180 PB. In 2016, GMV was US$ 17.8 billion. Such volumes meant a huge number of transactions and payments. In 2016, the peak transaction activity was 175,000 transactions and 120,000 payments per second. For the frenzy that the Singles' Day shopping created, Alibaba used big data algorithms, predictive data, sophisticated software, and data applications and informed the merchants about the kind of goods they should be ready with and the location of the warehouses where those goods should be stored. Alibaba also tried uni-marketing[i] for Singles' Day in 2017.

[i] Uni-marketing was a data-driven marketing technique to help brands target the right consumers by matching buyers and sellers precisely.

To meet the demand from millions of customers and the enormous online traffic generated, Alibaba used cloud computing technologies like elastic computing[j], virtualization[k], and real-time data processing. It also used a super computational engine called Apsara, to ensure that the systems were ready to process data on that day and to handle a sudden spike in traffic. It used Artificial Intelligence systems to handle customer service related work during Singles' Day and automated almost 95% of customer service related tasks.

For Singles' Day in 2017, sales reached US$ 25.3 billion as against US$ 17.8 billionl the previous year. Observers said that providing a personalized experience to each shopper and making intelligent recommendations for all the consumers had helped hugely in this and that the company had big data to thank.

On the day, 812 million orders were processed. Alibaba tried its big data in logistics during Singles' Day to deliver the goods from the warehouses located nearby. Using this technique, the first order was delivered in 12.18 minutes of the purchase.

Alibaba planned to extend the Singles' Day sale to Hong Kong and Taiwan by 2018, and then to other South East Asian countries over the next few years.

Financial Services

Alibaba launched Alipay[m] in 2004 as a third party mobile and online payment platform. At that time, only 0.2% of the Chinese had check books or credit cards. In such a scenario, Alipay acted as a facilitator on Taobao. When a buyer purchased an item on Taobao, Alipay got the money. It was held there till the shipment was completed by the vendor and the buyer had confirmed that the product received was what had been promised and was in good condition. After the confirmation, Alipay transferred the money to the vendor.

---

j Elastic computing is a concept in cloud computing in which computing resources can be scaled up and down easily by the cloud service provider.

k Big data virtualization is a process that focuses on creating virtual structures for big data system.

l To understand the scale of sales it can be compared to the sales of Black Friday and Cyber Monday in the US in 2016, which stood at US$ 6.79 billion.

m Alipay was linked to 108 partner banks in China. This was also linked to payment services like Western Union and VISA. Alipay could be used to pay utility bills, medical bills, air time, make credit card payments. etc.

In 2007, Alibaba launched AliLoan backed by China Construction Bank (CCB)[n], which had huge funds but was not able to find the right kind of borrowers. Like any other bank, it was also cautious about lending to small businesses as they did not have any credit history. Alibaba provided its data to CCB to help it make informed decisions. The relationship was terminated in 2011, after which Alibaba started lending through its AliFinance website. In the next one year. AliFinance extended loans of RMB 28 billion to 130,000 small businesses. By 2015, it had issued credit of RMB 100 billion.

Even by 2106, 25% of the population in China had a credit history. The remaining 75% of the people who did not own a credit card or have any credit record were an important segment for the banks which were looking for new avenues of growth. The pre-lending assessment was highly important for this kind of a loan as it was completely unsecured, and there were no assets that the lender could seize in case of loan default.

As a part of the pre-lending assessment, to evaluate the borrower, Open Data Processing Service drew data from the shopping websites of Alibaba and also Alipay. It used more than 100 computing models to process 80 billion data entries to determine the ability of the borrower to repay loans. All the decisions about lending were made by the system using the data, without any human intervention. The data also allowed Alibaba to find out when more loans were required (typically July and August) and lend accordingly. Using data and minimal human intervention reduced the cost of each loan to only RMB 0.3, which was far less than the cost of traditional bank loans, which usually cost RMB 300 to process.

After its success with lending to outside borrowers, Alibaba set up micro loan companies, Zhenjiang and Chongqing, in 2010 and 2011 respectively to provide loans to small vendors and micro enterprises which were part of Taobao and Alibaba.

Most of the banks which gave loans whether to a huge company or to an individual in the form of a microloan, needed to authenticate the client through review, risk profiling, etc. This deterred banks from giving small loans as the cost of processing these was huge. To provide loans, Alibaba used its own credit scoring model that was based on the online activity of the vendors. It started using data to understand client behavior to offer tailored financial services. The data came from several

[n] China Construction Bank is one of the 'Big Four' banks in China. As of 2015, it was the second largest bank in the world by market capitalization. It was also the sixth largest company in the world.

sources and used several sets of data like revenue growth, transactions, rating by clients in the industry, rating by users, records on Taobao, repeat buyers, etc. to rate them before giving loans. *(Refer to Table II for the sources of data)*

Table II:

Sources of Data

- Sales Data from the Sellers
- Authentication and Registration Information from the Platform
- Psychological Test Results
- Credit Rating
- Historical Transaction Records
- Marital Status
- Import and Export from Customers
- Customer Interaction Behavior
- Utility Bill Payments.

Compiled from various sources

Those who were involved with Alibaba or Taobao for three months were given loans of about RMB 20,000-30,000. Some of the features of microcredit from Alibaba were daily interest calculation, online payment, and no collateral requirements. The loans were completely unsecured and were given without any contact between borrower and lender.

Micro-lending witnessed a sudden spurt after 2013, after which the new loans increased by RMB 100 billion. As of February 2014, Alibaba had offered loans to more than 700,000 small businesses.

Many traditional banks also started cooperating with Alibaba to extend credit and other services to low-end individuals and SMEs. Association with Alibaba provided the banks access to huge amounts of data and also gave them several new business opportunities. Alibaba's use of IT helped the banks reduce their client data acquisition and operation costs. The banks could access the data and inventory details of the vendors on Taobao and TMall, along with details of their transactions, sales, inventory, etc. This helped them understand their financial situation. According to Chen Weiru, Associate Professor of Strategy, China Europe International Business School, Shanghai, "*Alibaba's big data processing capability is at the very forefront worldwide: they can process 120,000 orders within a second. So if*

Alibaba can maintain its edge in big data, we can expect them to resolve two difficulties in the financial sector: credit and cost."[8]

Credit Rating

In 2015, drawing on big data, Alibaba launched a credit scoring service Sesame Credit Management Co Ltd. This was the first agency in China to use online and offline data to generate individual credit scores. The consumers could use this score to access credit services like home and vehicle loans, and other loans. According to Sesame Credit's chief data scientist Yu Wujie, "*Sesame Credit is focused on those who may have little credit history at traditional credit agencies. They may have never obtained bank loans or applied for credit cards. However, they might be active internet users who shop online a lot, e-pay their utility bills on time, have a stable residential status and have been using their mobile phone numbers for a long time. We will take these and other factors into consideration when assessing consumers' creditworthiness.*"[9]

Sesame Credit obtained data from millions of registered and small businesses that transacted on Alibaba Group's websites. The data about payment history was obtained from Alipay. This data was further augmented by including data from external sources like public agencies, financial institutions, and other merchants.

Alibaba was of the view that this credit score had a host of uses, right from checking potential tenants and potential employees to making online bookings.

Fraud Risk Monitoring and Management System

Using big data processing and risk models, Alibaba built a fraud risk monitoring and management system. The system had the capability to capture fraud signals analyzing data of user behavior. It analyzed the data in real time through machine learning to predict fraudulent users. This led to the real-time payment fraud prevention monitoring system called Counter Terrorist Unit (CTU). The system had the capability to track and analyze users' behavior, identify suspicious activities, and use intelligent arbitration to apply different counter measures.

The first version of CTU was launched in 2005, and the main focus was on investigating unusually large transactions, frequent refunds, etc. Later on with more data coming in, CTU was extended to include money laundering, stolen / lost cards, marketing frauds, etc. This became a system that ran continuously to provide protection. Any activity like logging in, changing profiles, withdrawing money,

transferring money to other accounts, etc., were called events. Any event that was deemed suspicious triggered CTU and a risk decision was returned within 100 milliseconds. If the event was of low risk, it was processed; if it was high risk, it was stopped from being further processed.

Alibaba came out with real time fraud detection techniques using big data and built a risk prevention framework. *(Refer to Exhibit X for Multi-layer risk prevention framework)*

CTU went on to become one of the most advanced online payment fraud management systems. Within Alibaba and its subsidiaries, these models were used in every procedure like opening accounts, verifying identities, placing orders, withdrawing money and other transactions.

After successfully using it, Alibaba extended its process of building a clean payment environment to its external customers and came out with a product called AntBuckler for banks and merchants. Fully developed by Alipay, AntBuckler helped banks and merchants identify fraudulent activities and instances of cyber-crime.

This product used the RAIN score engine (Risk of Activity, Identity and Network) to quantify risk levels. This generated a risk score which enabled users to know the risk level of the borrowers. The RAIN score ranged from 0 -100 with a higher score meaning higher risk. It was also color coded with green meaning safe and red meaning risky.

In the RAIN model, several variables were selected to interpret the status and behavior of an object. The variables were given weight within the dimensions of activity, identity, and network. Depending on the risk, the variables and the weight given to variables changed. It could identify and prevent fraudulent and malicious behaviors. This was extremely useful for online payment systems and services to discover frauds, and enable both buyers and sellers to carry out safe transactions.

Operation Cloud Sword

All over the world, a huge problem was counterfeit and pirated goods, which the Organization for Economic Cooperation and Development estimated at US$ 461 billion in 2015. This accounted for 2.7% of global trade and it had doubled in a span of seven years. This was one problem which governments could not address despite their best efforts. China, being the leading manufacturing center in the world, needed to address the problem too.

In 2016, the government pulled up Alibaba for its lax attitude toward the sale of fake goods. Investors reacted negatively to the news and Alibaba's stock plunged by 9% on a single day. Jack Ma realized the gravity of the situation and acted upon it immediately. The company used big data technologies to monitor, track, and detect counterfeit goods and their manufacturers.

Alibaba's Taobao consisted of millions of vendors of which a small number were the ones selling counterfeit goods. Alibaba felt that the presence of these vendors, though miniscule, could prove to be harmful for its reputation and business, and also for the customers, who had reposed trust in Alibaba.

Alibaba had been actively involved in eliminating such vendors for a long time but, attracted by huge margins, these vendors concealed their true identity and continued selling such products by changing the storefront on e-commerce websites like Taobao or TMall.

Alibaba had a program that continuously scanned its e-commerce websites, taking data from big data sources. It used 600 data analytics models to analyze in real-time the merchants and product listings. The data points used included product specification, customer reviews, and user reviews. All the new products that entered the listings had to go through a scanning system that used an Artificial Intelligence program. Other data analyzed included relational data pertaining to user behavior, merchandise, return address of the company, etc. to find suspected fake goods. Optical character recognition was another technique which detected anomalies like a huge difference between the original and selling price, which showed a potentially fake product. Using big data, Alibaba deployed an app to distinguish the data patterns to identify the counterfeiters. It found that 90% of the fake goods originated from 10 regions and concentrated on those.

In 2015, Alibaba announced that it was assembling a team of 300 people to work exclusively on eliminating fake goods. It launched an IP protection technique called CloudSword, using which it was able to accurately find and delete the listing of counterfeit goods. Using the vast amounts of data it had, Alibaba was able to effectively track the whole supply chain of counterfeiters and crack down on them. It made a few more checks and ensured the information it had gathered was authentic before submitting it to the authorities. This was initially conducted as a pilot program in 11 cities, and the local Economic Investigation Team was informed. In 2016, an enhanced version of CloudSword was launched. It was used by the China Patent and Trademark Office of Zhejiang province, where the pilot program was carried out. Through this, many more counterfeit manufacturing and

sales locations were uncovered. The success of this led to the formation of the CloudSword Alliance, which several provinces and cities like Shanghai, Anhui, Jiangxi, and Jiangsu joined.

Alibaba identified the fakes that were listed by collating data obtained from customer complaints, observations from the brand owners, optical character recognition, and photos and logos displayed online and received from consumers. Again through Alipay account information, bank details and other financial information were obtained. This was again matched with the registration information of the vendors to pinpoint their exact location. After finding the location of the counterfeit vendor and confirming that they were selling fake goods, the appropriate authorities were informed. Alibaba also provided the authorities with evidence, which was used to prosecute the counterfeit goods vendors. According to Matthew Bassiur, head of global intellectual property enforcement at Alibaba Group, "*Big data alone isn't going to bring an end to all counterfeiting of brands and products on anyone's platform. But the good news is that its use greatly evens the odds of catching counterfeits and counterfeiters, maybe even predicting where they're most likely to appear.*"[10]

Within a year from August 2015, Alibaba had removed more than 380 million product listings and closed down 180,000 third-party seller stores. It also sued a few vendors who were found to be selling such products for violation of contract and goodwill.

In 2017, the company formed a counterfeiting alliance called Alibaba Big Data Anti-Counterfeiting Alliance. Companies like Mars, Amway, Swarovski, Louis Vuitton, Huawei, and Samsung were part of the alliance. The alliance obtained support from trade associations, intellectual property experts, and regulators, law enforcement agencies, and government bodies. The alliance agreed to pool its resources and collaborate to eliminate fake goods. All the members also agreed to share their expertise in the areas of IP authentication and anti-counterfeiting. All the members of the alliance were given access to the anti-piracy platform of Alibaba. The technology that Alibaba used scanned more than 10 million product listings every day and removed several spurious listings.

The New Retail

The most populous country in the world, China had more than 500 million middle class consumers and the total retail market was valued at US$ 5 trillion in 2016. By 2019, online shopping in China was expected to cross US$ 2 trillion, making the market 3.6 times larger

than the US market. By 2020, China was estimated to account for 60% of the global e-commerce.

As of 2016, online sales accounted for 15% of all the retail sales. Shopping habits had changed quite a bit over the last decade, with more and more people opting for e-commerce. However, physical stores still accounted for a major share in retail, with shoppers continuing to show a preference for touching and feeling the goods they wanted to buy and for interacting with shop assistants.

With physical retail continuing to boom, Alibaba made its move toward bricks and mortar stores. Analysts termed it as online-to-offline (O2O) revolution. It was felt that this could hold huge challenges for the existing physical retailers with a limited online presence.

Marking its move into physical retail, Alibaba acquired a department store chain, Intime Retail, for US$ 2.6 billion in January 2017 and also entered into a strategic alliance with the state-owned supermarket Bailian Group in February 2017. Bailian had a huge bank of underused retail space, including 4,700 outlets across 25 provinces, which Alibaba planned to put to good use. Bailian and Alibaba planned to share stores, merchandising capability, logistics, and technology. Both the companies wanted to leverage their online and offline capabilities to integrate merchandise, logistics, payment tools, etc. The other areas of cooperation included designing cross-channel stores, R&D in retail technology, big data, customer relationship management, supply chain management, etc.

Alibaba planned to bring its big data capabilities into physical retail also and to bring out a new uni-channel combining online and offline stores, offering traditional brick and mortar retailers tools to transform the way they connected to consumers. Its O2O strategy was expected to seamlessly bring together both types of retail. The foray into retail was expected to help Alibaba accumulate more data on consumer behavior. At these stores, Alibaba planned to use beacons and wi-fi sniffers to capture data about traffic flows in stores, areas of longer engagement, people leaving without buying, average time spent, etc., and to collate it with the data it had available.

Both the companies planned to provide real time service solutions for customers at stores by designing cross-channel store operation and ordering systems. They planned to integrate the membership systems to provide big data driven sales. To improve merchandise selection and reduce costs, the logistics of Bailian were to be combined with that of Cainaino, the logistics platform. The payment tools from Bailian Safepass and Bailian OK were to be integrated with Alipay.

The big box retailers were watching Alibaba's moves with a mix of interest and trepidation. The CEO of one of the retailers said, "*It is changing the whole landscape.... It's a reality which you have to deal with.*"[11] Some of them felt that in order to ward of the threat from online retailers, the physical retailers needed to cooperate with each other. A few retailers opined that aligning with one of the e-commerce retailers was necessary for survival and success.

Analysts said that the concept of 'New Retail' would transform retail in China and take it ahead of the highly developed retail markets in the US and Europe in terms of retail innovation. Alibaba would have a major role to play in this through its data centered shopping approach. This approach was expected to help Alibaba identify the kind of stores for every demographic, and also improve in-store experience, making it more personalized and convenient.

Alibaba also entered into a data-sharing scheme with several retailers, where it provided detailed information on the buying behavior of customers within a specific radius of a mall or a store. The details included gender, preferences, frequency of buying, kind of products bought, etc. The hypermarkets or stores were asked to provide the database of their own customers in return. Thus, Alibaba was able to build its database on retail customers.

Alibaba came up with a grocery store called HEMA, which was an online-offline hybrid store. It was started in 10 locations in Shanghai and Beijing. The merchandise included 3,000 curated high-end dining products from 100 countries. After seeing the products, customers could order them using a mobile app and have them delivered at home within 30 minutes (within a five-kilometer radius). These stores had digital tags that were updated real time. The mobile app captured detailed shopping behavior including movement around the store, time spent near each rack, and purchases. Analysts said that this would provide data that was more useful than traditional metrics like value per order. Thus, by using online technology for the collection and application of big data and by bringing it into the offline world, Alibaba would personalize the shopping experience, which would give it a competitive edge over other retailers.

Taking Big Data to the Next Level

The Artificial Intelligence (AI) program, that was an offshoot of big data, was used widely in China to tackle problems like traffic congestion, easing traffic, suggesting alternate routes, medical imaging, etc. AI and MaxCompute were deployed to support local governments in

several smart city projects in mainland China. For example, as a part of the 'Hangzhou City Brain' project, videos were taken from more than 50,000 surveillance cameras in the city, and processed to take decisions pertaining to traffic signals. "*What MaxCompute does in City Brain projects is to combine all the public data sources in real time, perform data analytics, and generate actionable insights. This helps the government to take preemptive action in situations such as traffic control*,"[12] said Wanli Min, artificial intelligence scientist, Alibaba Cloud.

The automated traffic system capabilities made intelligent adjustments to the traffic lights according to traffic movement. For example, if a huge vehicle was crossing the light, it remained green till the vehicle had crossed. The system also gave users real time traffic recommendations and suggested routes with less traffic. After the commencement of the Hangzhou City Brain project, traffic speeds increased by 11% in a span of just one month.

The efficacy of Alibaba's AI program could be gauged from the fact that in April 2016, Aliyun predicted the winner of a popular reality TV show 'I am a Singer' using factors like popularity of the songs performed, social media buzz, feedback, studio audience views, and the performance of the singers. The winner was decided by a panel of 500 judges, and the outcome could swing any way as all the singers who reached the finals were equally talented. The predictions from Aliyun were shown live on TV. Analysts who pitted Aliyun against Google's Alpha Go (which was used to play ancient Chinese board game Go and beat a human player), said that it was far more difficult to use AI in programs like 'I am a Singer' as several emotional parameters were included in determining the results.

In 2014, Alibaba entered into an agreement with Youku Tudou Inc., China's leading internet television company. The two companies cooperated to drive adoption of big data in marketing in China. Alimama was to use the data and technology to support the 'View and Buy' video marketing technique developed by Youku. Using this technique, consumers could click on merchandise that appeared in the video and purchase it, while continuing to watch the video. It also created a Merchants' Video Channel, where merchants could market their products and services themselves by creating videos to promote them.

Earlier, Alimama launched video marketing services in collaboration with major video websites to improve the reach and relevance of online video marketing to help the merchants who listed their products on Taobao and TMall.

In 2015, global FMCG giant Unilever entered into a strategic partnership with Alibaba. Alibaba through its data driven ecosystem helped Unilever reach customers across China precisely and effectively. This helped Unilever penetrate into rural China. The collaboration also included e-commerce and big data. Both the companies planned to innovate in the area of big data analytics, and supply chain management.

In 2017, American drinks brand Pepsi partnered with Alibaba to increase its reach in Asia. Pepsi planned to improve consumer experience by leveraging Alibaba's big data to introduce innovative marketing initiatives, customize products, and improve distribution. According to Mike Spanos, PepsiCo GCR president & CEO, "*Through this collaboration, we will fully take advantage of Alibaba's platform and data to carry out more innovative experiments, perfect PepsiCo's products and services, and enhance online consumer experiences. Our joint effort with Alibaba will help us lead the ever-changing consumer trend and better serve Chinese consumers.*"[13]

In 2015, Alibaba launched a stock market index based on its proprietary data. This would use big data in investment decisions. Ant Financial, along with a company called Gildata, unveiled the China Securities Index (CSI) Taojin Big Data 100 Index. The index weighed industries based on the number and growth of transactions taking place through Alipay, price levels, and the supply demand situation. These were then used to calculate stock positions. An investment company based in Shenzhen, Bosera Asset Management, launched a passive fund based on this index.

In November 2017, Alibaba announced that Alibaba Cloud would partner with FIFA[o] Club World Cup as a presenting partner until 2022. According to FIFA's Director Marketing Sales and Strategy, Iain Downie, "*We are aware of the substantial value that Alibaba Cloud provides to business in terms of technical support and global reaching services. We can only see great potential in this fresh collaboration related to the FIFA Club World Cup.*"[14] Alibaba would design and also present the award for the most valuable player. Though its infrastructure, Alibaba Cloud would help manage massive traffic during the peak hours of the game, and also provide big data analytics requirements.

[o] International Federation of Association Football or FIFA is responsible for organizing major international tournaments in football, including the World Cup

Looking Ahead

According to Jack Ma, pure e-commerce would vanish in the future and five trends would dominate – 'new retail', 'new production', 'new finance', 'new technology', and 'new resources'. These new trends would bring with them several opportunities and challenges, especially for online entrepreneurs, to use the internet to meet increasing demand. According to Jack Ma, 'New retail' meant "the integration of online, offline, logistics, and data across a single value chain." Alibaba tried out 'New retail' during the Singles' Day 2017, where big data was used to integrate online and offline activities. Its strategic alliance with Bailian also showed that it was moving in the right direction as far as 'New Retail' was concerned.

In March 2017, Alibaba announced that it planned to restructure its e-commerce business to meet the demands of new retail. This could lead to the merger of TMall and Juhuasun. Experts said that this move clearly indicated that Alibaba planned to remove the barriers between online and offline retail and instead segment retail into digitally enabled and traditional markets. "*New retail reimagines the relationship between consumers, merchandise and retail space by leveraging mobile internet and big data. It will upend the traditional manufacturing and supply chain, the connection between merchant and consumer, as well as the overall identify, reach, analyse and serve their customers, and their digital transformation will be empowered by Alibaba's ecosystem*,"[15] explained Zhang.

He also said that the data would soon be used to reshape manufacturing and it would be an era of 'making in internet', which would replace 'selling at internet', and 'marketing through internet'. The company had plans to come up with a new retail strategy where online, offline, logistics, and data would be integrated across a single value chain. It was up to the companies to take advantage of billions of digital footprints that consumers left and continue to leave in the digital world and come out with the right products, meet demand, and also create new demand.

With the expertise it had gained in all the areas related to big data, Alibaba was looking to venture into global commerce, hit a GMV of US$ 1 trillion by 2020, and reach 2 billion consumers by 2036. According to Zhang, "*Our ultimate strategy is to build the future infrastructure for global commerce. That's an undertaking that, while highly ambitious, remains true to Alibaba's core mission of making it easier for consumers and companies alike to do business anywhere.*"[16]

Ma said globalization was one of the key growth engines for Alibaba. He wanted to grow Alibaba into a platform that would enable six million small businesses to sell across borders, deliver products anywhere in the world within 72 hours, and eventually serve 1.2 billion people outside China.

Experts said that big data had helped Alibaba become the world's largest retail platform and the largest retailer ahead of US-based Walmart, and it would have a major role to play in Alibaba's goal of becoming the equivalent of the world's fifth largest country by GDP (after the United States, China, the European Union, and Japan) by 2020. Is big data the trump card that would help Alibaba realize its ambitions in global commerce?

Exhibit I:

Alibaba – Business

Business	Activity
Alibaba.com	• Imports and exports
TMall	• For reputed brands to sell in China
1688.com	• b2b trade in China
Aliexpress.com	• Retail website
AutoNavi	• Map Supplier
Taobao Marketplace	• C2C online shopping platform
Alipay	• Third party online payment
Aliyun	• Cloud computing service platform
AliExpress	• Online retail service from Chinese businesses for international buyers
Yahoo! China	• Internet services like news, email, search
Aliwangwang	• Instant messaging
AliHealth	• Pharmaceutical e-commerce
UCWeb	• Mobile internet software technology

Adapted from www.alibaba.com

Exhibit II:

Alibaba – Annual Active Buyers

As of (12 months ended)	Number (in Million)
June 30, 2014	279
September 30, 2014	307
December 31, 2014	334
March 31, 2015	350
June 30, 2015	367
September 30, 2015	386
December 31, 2015	407

As of (12 months ended)	Number (in Million)
March 31, 2016	423
June 30, 2016	434
September 30, 2016	439
December 31, 2016	443
March 31, 2017	454

Compiled from Annual Reports, Alibaba

Exhibit III:

Alibaba – Milestones

Year	Event
1999	China marketplace (later called 1688.com) launched
2004	Aliwangwang, an instant messaging tool for buyers and sellers in Taobao mall, was launched
2006	Strategic Investment in Koubei.com
2008	Taobao Mall, a B2C platform, was founded
	Koubei merged with China Yahoo
	Alimama integrated with Taobao
	Alibaba R&D Group established
2009	Alibaba Cloud computing established
2010	Juhuasuan, a sales and marketing platform for flash sales Established
	AliExpress – enabling exporters in China to transact with consumers across the world – was launched
	Mobile Taobao App launched
2011	Alipay obtained online payment license
2014	Tmall Global – enabling international brands to offer products in China – was launched
	UCWeb acquired and integrated
	Ant Financial Group was established

Compiled from various sources

Exhibit IV:

Alibaba Financial Data

(in RMB Millions)

Year ended March 31	2013	2014	2015	2016	2017
Revenue	**34,517**	**52,504**	**76,204**	**101,143**	**158,273**
Cost of revenue	(9,719)	(13,369)	(23,834)	(34,355)	(59,483)
Product development expenses	(3,753)	(5,093)	(10,658)	(13,788)	(17,060)
Sales and marketing expenses	(3,613)	(4,545)	(8,513)	(11,307)	(16,314)
General and administrative expenses(1)	(2,889)	(4,218)	(7,800)	(9,205)	(12,239)
Amortization of intangible assets	(130)	(315)	(2,089)	(2,931)	(5,122)
Impairment of goodwill and intangible assets	(175)	(44)	(175)	(455)	-
Yahoo TIPLA amendment payment(2)	(3,487)	-	-	-	-
Income from operations	**10,751**	**24,920**	**23,135**	**29,102**	**48,055**
Interest and investment income, net	39	1,648	9,455	52,254	8,559
Interest expense	(1,572)	(2,195)	(2,750)	(1,946)	(2,671)
Other income, net	894	2,429	2,486	2,058	6,086
Income before income tax and share of results of equity investees	10,112	26,802	32,326	81,468	60,029
Income tax expenses	(1,457)	(3,196)	(6,416)	(8,449)	(13,776)
Share of results of equity investees	(6)	(203)	(1,590)	(1,730)	(5,027)
Net income	**8,649**	**23,403**	**24,320**	**71,289**	**41,226**

Source: Annual Reports, Alibaba

Exhibit V:

Alibaba Internal Sources of Data

Source	Type of Data	Description
CAINIAO Global	Logistics Data	Official global parcel tracking platform of Alibaba
Amap.com	Geographical and Vehicle data	Web mapping, navigation and location-based services. The company, Autonavi Software, was acquired by Alibaba in 2014.
Alijk.com	Health & Life data	The online pharmacy business of Tmall, was transferred to Alijk.com. It later launched a mobile app for buying medicines.

Source	Type of Data	Description
CNZZ	PC Website monitor Data	Online statistical analysis services and web analytics provider. It was acquired by Alibaba in 2011.
UMENG	Mobile app monitor and browser data	Analytics provider acquired by Alibaba in 2013
UC Browser		Mobile browser owned by Alibaba
Alisports.com	Entertainment Data	Sports unit of Alibaba
Alibaba Pictures		China Vision Media (acquired in 2014), involved in movie and TV programs, magazine publishing, advertising
Alibaba Literature		Content production, online publishing
Youku Tudou		Video streaming portal
Xiami		Music streaming platform
Damai		Event Ticketing
Weibo.com	Social Network Data	Microblogging
Etao		Online shopping search
Adchina.com	Marketing Data	Internet advertising platform
Alimama.com		Marketing platform for big data
Taobao.com	e-commerce Data	Online shopping website
Alitrip		Travel booking website
Tmall.com		B2C online retail
Koubei		Local commerce platform
Juhuasuan		Group shopping site
1688.com		Online wholesale marketplace

Compiled from various sources

Exhibit VI:

About Big Data

Big data refers to large amount of structured and unstructured data that can be mined for information. Big data is extremely large and complex data that can be analyzed to reveal patterns, trends and associations etc. and can be especially be used for analyzing human behavior and interactions. Such a data is collected through devices and technologies such as credit cards and customer loyalty cards, social media from the internet, Wi-Fi sensors, and electronic devices. According to the Oxford English Dictionary, the term 'big data' was first used in 1941 to quantify the growth rate in the volume of data – alternatively known as the information explosion. Later, the term became more and more related to information in the digital format and the Information Technology sector.

Advancement in technology over the years had a huge impact in the way data was collected and disseminated. Earlier data was more structured, as it was captured in pre-determined formats and could be fixed into particular fields in databases. Thus it was easy to analyze and compare. With the sources of data becoming varied, the data became unstructured, and started to include other forms like images, audio, and video files, data from sensors, geospatial data, radio frequency identification, social media data, etc. Such data did not possess a particular format or length, was difficult to fit into specific fixed formats, so it was difficult to store and archive. Apart from these, there was semi-structured data, which was structured data but had a few attributes of unstructured data. According to IBM as of 2013 2.5 quintillion bytes of data was present in the world, of which 90% of the data was created in a span of two years. According to an analyst, "Personally I see big data as a meshing of information from every possible touchpoint…and the activities to evaluate, interpret and glean value from this data by turning it into information…to enhance any organization's decision making capabilities."

In 2001, Gartner analyst Doug Laney came up with three Vs of Big data high-volume, -velocity and –variety. Volume refers to increasing amount of data being generated by companies about the customers internally and also from external sources like the social media, sensors, monitoring devices, mobile networks, government systems, etc. Velocity refers to the speed at which data is collected. With growing sophistication in the data collection methodologies, more data sources came under the purview of collection, and data could be collected and processed at high speed. Variety refers to different types of data. Data could be generated from conversations, opinions posted, in the form of pictures, videos, audio, machine data, system data etc.

Over the next few years other 'V's were added. These included Veracity (trustworthy data), Value (importance of data to the organization), Visualization (the way data is presented), Vocabulary (metadata – data about data), Venue (geolocation), and Variability.

According to the McKinsey Global Institute, "Big data refers to data sets whose size is beyond the ability of typical data software tools to capture, store, manage and analyze." To analyze and process such vast amounts of data traditional tools like relational databases and desktop software for statistics and visualization were inadequate, and needed parallel software running on thousands of servers.

Contd…

Contd…

Business intelligence was widely used to analyze such data. It involved gathering, storing and providing access to data. This was done through decision support systems, online analytical processing, forecasting and statistical analysis. Business intelligence identified the data patterns of the past data. Data analytics looked at how to use the data in the future. Big data analytics involved using algorithms to drive decision making. Those used for predicting behavior are called predictive analytics. Big data analytics helps in finding new patterns and co-relations.

Vast sources of data and different ways of deriving meaning from them meant that companies have access and insights into their customers and suppliers. This provides companies opportunities to customize the products or services to meet the needs of the customers, and better integrate the suppliers for developing, making and delivering the products in demand at the right time.

Experts believed that any organization imbibing big data in its activities becomes innovative, and will be able to deliver more effectively, which would result in more sales and more profits.

Compiled from various sources

Exhibit VII:

About Alimama

Alimama was started by Alibaba in 2007 as an online exchange company, which allowed advertisers to reach the customers effectively. It provided services such as customer segmentation, product positioning, running marketing campaign, and more. It also helped international brands launch their products on TMall by helping them reach the right kind of customers. To do this, from the big data sources, Alimama used identity, browsing history, search, transactions, social data, payments, shopping cart activity, video consumption, cookies and map data from smartphones. According to the CMO of Alimama, "Like the mother in a family who links together all of the family members, we integrate all of the data within the Alibaba group."

Previously, while collecting data, every device – computer, mobile phone, tablet – was considered to be a unique user. If a person used all three devices, it was considered as three different usersp. So the advertisements targeted the same individual several times, wasting the advertisers' budget.

Alimama addressed this challenge by bringing together a single user's data and creating cross-screen tracking delivery. The technology it developed helped it determine whether the impression was from the same user on a different device, and whether to project the ad or not. Instead of depending on purely demographic attribute data, Alimama based its advertisements on time-defined moments, like buying travel related things when people were planning a holiday, and predicted their future buying habits.

Contd…

[p] This phenomenon was termed as 'swimming lane effect' and limited the effectiveness of the data collected.

Contd…

In 2015, Alimama launched an open Data Management Platform or DMP, through which Alibaba planned to work with brands and merchants to understand the customer data better. Through this it wanted to market to a broad base of consumers and create a better shopping experience for buyers by providing personalized shopping recommendations.

Compiled from various sources

Exhibit VIII:

Cloud Computing

In 2010, Alibaba started offering cloud computing services to businesses and the government. By June 2014, Aliyun was serving over 1400,000 clients from a wide range of industries, including mobile application development, internet gaming, digital entertainment, financial services, consumer electronics, and governmental bodies. These companies paid fees to gain access to the data centers, and were charged based on the computing power they consumed. AliCloud offered 20 products, solutions, and services covering the data development chain, like data services, data processing, analysis, and capabilities for machine learning.

In 2014, Aliyun partnered with Donghua Software, an ICT service provider, to establish a smart city. Both parties jointly provided consultation, design, construction, operation, and application development services for smart cities with big data based on Aliyun's cloud computing platform. The solutions obtained from the project were to be sold through Aliyun.

Alibaba released MaxCompute for external users too, where enterprise customers could store data including images and audio and video files. A single MaxCompute cluster could be scaled to 10,000 servers. It could store up to a peta byte of data to support different computing models, and enable query on massive datasets. It processed the data at the lowest computing cost in the world at US$ 1.44 per TB of data.

By 2016, AliCloud had 500,000 paying customers using it for storage, security, database management, and cloud computing services. The demand for cloud computing in China was expected to reach US$ 20 billion by 2020.

By 2017, it had data centers across the world including the USA, Australia, the Middle East, Europe, Singapore, China, and Japan. Alibaba's revenue from cloud computing stood at RMB 3019 million in 2016 and increased to RMB 6663 million by 2017.

Compiled from various sources

Exhibit IX:

Alibaba – Businesses and Data Flow

Annual Report, Alibaba

Exhibit X:

Multi-Layer Risk Prevention Framework

Alibaba's multi-layer risk prevention framework was extensively used in Alipay. In this system there were five layers that helped in preventing fraudulent transactions.

1. Account Check
2. Device Check
3. Activity Check
4. Risk Strategy
5. Manual Review

Whenever a buying was initiated, the first step was the account check. During this step, checks were conducted to see if the buyer or seller account had any suspicious activity, or if any such activities had been reported earlier. A check was done to confirm that the buyer's account was not being used by another party. Then a device check was performed to check the IP address and device. These included the number of transactions from the same device / IP Address; etc.

In the third layer – also known as behavior check – the historical records of the buyers and sellers were checked, and also their buying pattern, and the links among different accounts and devices. At the same time, a check was also done to see whether any of the accounts were linked to bad accounts. During the fourth check – risk strategy – the final decision about the authenticity of the account was taken. In this stage, the results from all the previous stages were aggregated. If anything was found to be fraudulent, it was sent to auto-decision. If there were a few which needed more checking or were found to be suspicious, they were subjected to manual checking. During this stage, the buyers or sellers were contacted directly for verification and checking.

Compiled from various sources

End Notes:

[1] "Jack Ma: Alibaba Is Not A Retail Company, But A Data Company," TMT Post, June 27, 2016.

[2] http://www.alizila.com/

[3] Alex Linder, Jack Ma says Alibaba will be world's 5th largest economy by 2036, http://shanghaiist.com, June 15, 2017

[4] Five Reasons why Alibaba is a Data (Not e-commerce) Company, www.alizila.com, October 17, 2016

[5] http://www.alibabagroup.com/en/about/history

[6] Alibaba Taps User Data to Drive Growth Spurt, www.ft.com, June 21, 2017

[7] Earnings Call Transcript, Alibaba Group Holding, Q2 2018 Results

[8] Can anyone Dethrone the Mighty Alibaba? The Link, www.ceibs.edu, 2016.

[9] China's Alibaba draws on 'Big Data' to Launch Credit-Scoring System, www.out-law.com, February 03, 2015

[10] Matthew Bassiur, How Alibaba deployed Big Data to combat China IPR pirates and counterfeiters, www.recode.net, December 20, 2016

[11] Paul Liu, Xuemei Bennink Bai, Jason Jia, Eva Wang, The Accelerating Disruption of China's Economy, www.fortune.com, June 26, 2017.

[12] Carol Ko, Big Data Processing: from Smart Retailing to Smart Traffic Mgmt, Computerworld Hong Kong, October 25, 2017

[13] Jack Ridsdale, Pepsi Signs Strategic Partnership with Alibaba, www.licensing.biz, May 12, 2017

[14] Alibaba Cloud Becomes Partner of the FIFA Club World Cup Until 2022, Al Bawaba, November 23, 2017

[15] Nayela Deeba, Alibaba and Bailian Leverage Big Data to Support New Retail Model, www.cio-asia.com, February 21, 2017

[16] Big Data Game-Changer: Alibaba's Double 11 Event Raises the Bar, www.technologyreview.com, November 14, 2016.

The last decade has witnessed a tremendous proliferation of **big data,** providing organizations with the opportunity to leverage this information for real time and effective decision making. In the new milieu, the constraints of collecting and storing information have been largely taken care of, making data widely accessible. However, what organizations actually do with this information is more important, and there still remain challenges regarding how to make sense of the information and leverage it for effective decision-making.

In addition to technology, a company's **big data strategy** has important implications for its leadership, talent management, decision making, and organizational culture. Moreover, privacy and data security have also become key concerns in the wake of evolving cyber threats such as malware, ransomware, and increased instances of data breaches.

In this era of digitization, it is getting increasingly more necessary for organizations to get their big data strategy right. So, how are leading companies giving shape to their big data strategy?

The case studies included in this collection look at issues and challenges in formulating and implementing a big data strategy in some of the world's top companies. The first three case studies show ways in which organizations cutting across different industries are trying to capture and employ data and the associated challenges. For instance, there are case studies on global furniture retailer **IKEA;** an online fashion startup, **Stitch Fix;** and a leader in parks and resorts, **World Disney World.**

The next six case studies are on **IBM,** an IT behemoth; **Amazon.com,** a leading e-commerce and technology company; **Netflix,** a leader in digital streaming; **Procter & Gamble,** a leader in the consumer packaged goods industry; and, **Baidu** and **Alibaba,** two Chinese tech giants. Together, these case studies highlight different aspects of big data strategy.

The collection includes five **prize winning** case studies. It is a good mix of **contemporary** as well as **classic** case studies on big data strategy that have withstood the test of time. For instance, the Netflix case study has been a **global bestseller** right from the time it was published in 2013.

In addition to raising critical issues, these cases are also a source of **best practices** in this emerging area. With the amount of digital data set to explode in coming years, the implications for the big data revolution are vast, and organizations cutting across industries can ill-afford to not have a holistic big data strategy in place aligned to their business strategy.

CASE RESEARCH CENTER
IBS
HYDERABAD
www.icmrindia.org | www.ibscdc.org

ISBN 978-81-314-2846-7
9 788131 428467