



Naga Lakshmi Damaraju | Navneet Kaur Khangura

Modak Analytics: Shaping the Future in Digital India?

It was August 2016. Aarti Joshi, the founder and Chief Executive Officer of Modak Analytics, was gazing at the skyline from her office, enjoying the cool August evening breeze. The skyline was a reflection of the dramatic transformation of Hyderabad, once a laid-back and relaxed city in the southern part of India to house one of the country's prime digital capitals, Cyberabad. She was pondering the future of her firm, Modak Analytics. Established in the last quarter of 2010 and barely five years old, Modak's emergence and its rapid catapult into the limelight was no less dramatic than the emergence of Cyberabad. The firm already had a big achievement to its credit. In 2014, it had executed one of the "biggest big data" projects in the world. The project involved voter data analysis for the Indian general elections, the world's largest election exercise with nearly 814 million voters on the electoral rolls. The volume of data involved was more than four times that of the United States presidential election. The challenge of that project kept the team charged and focused.

Two years later, although other interesting projects were coming the team's way and Modak's top management was finding new uses for their accumulated data and experience, they were palpably ready for the next big challenge. For Joshi, the question was not just one of making the next "big leap", but also of making sure that Modak Analytics was pivoting in the right direction. Clearly, there were many options on the table. Modak could capitalize on its big data expertise to become a data integration company and move into Western markets. Or, it could use its extensive database to become a leading data aggregation company in India providing services to banks, insurance companies, fast-moving consumer goods (FMCG) companies, and other segments. It was a matter of finding a sustainable growth path for the firm that would hold the interest of the team for some years to come. Hyderabad's rising skyline seemed to be challenging Joshi to scale up Modak to equal, or greater, heights.

EMERGENCE OF MODAK ANALYTICS

Joshi was a professional in the field of data analytics. In the seven years she had spent in the United States, she gained great exposure to and a deep appreciation of the power of analytics for decision making in all kinds of firms—large, medium and small. Upon her return to India, she was excited to see something she did not quite expect. There were many small- and medium-sized Indian retailers that had digitized point of sales (POS) data—data that could be used for making informed business decisions. However, these firms were not using their collected data to that end. She quickly realized that since data analytics was traditionally dominated by big players like Microsoft, Infosys and IBM, these small- and medium-sized firms (SMEs) were not able to afford these services and, consequently, were unable to tap the potential of the data that was readily available to them. Joshi sensed an

Professor Naga Lakshmi Damaraju and Navneet Kaur Khangura prepared this case solely as a basis for class discussion. This case is not intended to serve as an endorsement, a source of primary data, or an illustration of effective or ineffective management. The authors acknowledge inputs from Yashaswini Mulagund and Archana Malyala, academic associates at the Indian School of Business, during the development of this case study. This case was developed under the aegis of the Centre for Learning and Management Practice, ISB.

Copyright © 2017 Indian School of Business. The publication may not be digitised, photocopied, or otherwise reproduced, posted or transmitted, without the permission of the Indian School of Business.

ISB077

opportunity to establish a company that could provide affordable data analytics services to these Indian SMEs (see Exhibit 1). Modak Analytics was born to serve this opportunity, at least initially.

Joshi soon realized that SMEs were not as attractive a segment as she had thought. They lacked a strategic approach to business, had wafer-thin profit margins (partly due to high real estate costs), were unaware of the benefits of data analytics and, consequently, were reluctant to invest in data analytics. Joshi needed a new profitable target market for Modak Analytics.

Exploring the BFSI space and the launch of new products

The banking and financial services industry (BFSI) was the next pasture Joshi explored. Interestingly, most banks and companies in the BFSI sector had vast amounts of data but they lacked the requisite data analytics infrastructure. Major banks such as State Bank of India (SBI), HDFC Bank, ICICI Bank, Axis Bank, etc., were exceptions to the rule and had some analytics infrastructure. BFSI companies had the money to invest and great scope for using analytics for their improvement (see Exhibit 2).

Modak Analytics was able to break into this target market with an opportunity to work for one of the largest cooperative banks in South India. The bank had decade-old software—a core banking system (CBS)—which connected its 40 branches and supported common banking transactions. There was no significant competitive advantage in using it since all medium-scale banks in India had implemented core banking systems and were providing almost the same level of customer service. Therefore, the cooperative bank was keen on using Modak's expertise in data analytics in its bid to gain a competitive advantage.

Modak Analytics implemented Enterprise Data Warehouse (EDW), a unified database analysis and reporting solution using open source software, at the bank. While implementing EDW, Modak Analytics developed RapidETL, a data integration product that could be used to extract, transform and load data into the EDW. RapidETL could also be used to automate code generation for any other ETL software such as Informatica, DataStage, and others (see Exhibit 3). With automated ETL Solution accelerators that used RapidETL, a separate team to select and correlate the data was not needed. RapidETL did it automatically and converted all relevant data to appropriate formats without the need for consistent human involvement. Hence, it improved efficiency and reduced time and costs dramatically.

Yet another insufficiency of the CBS provided Modak with the opportunity to develop its Master Data Management (MDM) product. Although the bank's branches were linked through the CBS, there were no standardized processes across the various branches. Therefore, it was often the case that a customer had multiple customer identification numbers and only minimal and incomplete data was being entered to maintain customer records. The CBS was not equipped to check for spelling errors in names or addresses. Consequently, due to spelling mistakes or variations, an existing customer was regarded as a new one and given a different ID when buying a new product. This type of duplication made it impossible to ensure good quality data. Observing this flaw in the CBS, Modak Analytics created MDM, which essentially provided a common reference point for all data through a "master file". By providing a single version of data, it facilitated data sharing among personnel and departments and ensured consistency in information (see Exhibit 4).

The MDM product helped banks differentiate their offerings in the marketplace. An effective MDM strategy enabled an organization to gain a 360-degree view of customers to better understand their needs and consequently increase cross-selling and trading-up opportunities across multiple platforms—branch, ATM, Internet and mobile. It helped in understanding the profile of a complete household and not just one individual, along with the share of the wallet in the household. Since marketing was done at the household level, understanding the needs and profile of the complete household was useful for targeted marketing and providing differentiated offerings. MDM also enabled banks to regulate compliance and reduce risk. Operational efficiency improved, thereby reducing costs. In effect, MDM helped banks with their growth and retention goals by giving quick access to customer and household profiles. Overall, the strategic agility of these organizations improved with a more realistic picture of the data.

Further product development opportunities came along in an assignment at a growing regional cooperative bank based out of Hyderabad. Modak developed the Campaign Management System (CMS) during this assignment. With the CMS, customers could be geocoded on a map, giving the bank the capability to perform geospatial analysis. The bank could select customers from the map (by point and click) and send emails to only those selected customers. They could also "slice and dice" the customer segment according to age, products already purchased, family profile, lifetime value of the customer, etc. The CMS was designed for marketing/ sales professionals (business users) who thought differently from IT analysts. It avoided the need for expensive IT resources and cumbersome IT-driven processes to garner insights into the business (see Exhibit 4).

The project also involved profiling all the customers of the bank. Such profiling helped the bank answer such questions as: *Who are our customers? Where do they live? What products and services do they buy in specific branches? What are the common attributes of our customers? and What demographic and socio-economic groups do they belong to?* Modak Analytics delineated all the characteristics of the existing customers of the bank and presented the information to its top management team. They were impressed and wanted Modak to give them a way to find people with similar profiles as their current customers. Such an exercise would enable the bank to engage in targeted marketing, which would have higher conversion rates than a regular marketing campaign.

Although it appeared simple, this was not an easy task in India. Unlike the US market where companies like Acxiom, Experian, TransUnion, etc., provided demographic and socio-economic data about any individual residing in a neighborhood, such core data was not available in the Indian market. While marketers in the US had ready access to relevant data about each household, such as the number of children, annual income, value of the house, cars owned, pets owned, etc., such data was non-existent in the Indian context. This paved Modak's way towards Big data.

Experiments with Big Data

At this juncture, Milind Chitgupakar, a veteran business consultant with 15 years at IBM, joined Modak Analytics as its Chief Analytics Officer (CAO). Chitgupakar had served IBM as a global leader for customer and marketing analytics and had rich and varied work experience in the banking and insurance verticals. He had extensive domain knowledge in building large data repositories and held six patents in his name in the field of data analytics. While at IBM, Chitgupakar was part of setting up a 300-member data analytics practice in Latin America. After years of being with a big brand company, he was looking to work on something new and entrepreneurial. This led him to take up Modak's offer.

Meanwhile, Modak Analytics was scouting for open data sources that were available in India to understand if it could gather some core data about consumers for the bank. In the process, it stumbled upon the electoral database. Since the bank was based out of Hyderabad, the team at Modak expected that gathering the city's electoral data would be sufficient for the task on hand and would be a simple exercise.

However, the task was not simple to execute. The electoral rolls were only available in PDF format and the job of copying and pasting the information from PDF files to text files was arduous and time-consuming. Modak Analytics needed technology to decipher the PDFs and extract relevant information from them. With automation and innovation being the core of Modak Analytics, it developed a proprietary technology that helped automate the process of deciphering PDFs in 12 different regional languages, and then converting and storing the information in a database. Modak first processed the electoral rolls of the city of Hyderabad, and then of the entire state of Andhra Pradesh. In the process of analyzing the electoral data, the company gained some interesting insights into voter attributes and voter preferences for political parties.

At one of the conferences later on, Chitgupakar described the technology Modak had developed for tapping open data sources. He also presented the insights gathered from the whole exercise. The following day, he received a call from the IT wing of a leading pan-India political party, asking him to present this work. The CAO and his team presented an in-depth analysis to members of the party. Impressed by the presentation, the party requested Modak Analytics to take up the task of preparing a

ISB077

voter profile analysis for the entire country and providing election insights at the booth level. It was to be a "pro-bono" project. The understanding was that Modak Analytics could go to the media and talk about the project in the industry after the elections were over. However, there was neither a contract, nor a memorandum of understanding (MoU) signed between the political party and Modak Analytics in this regard. It was a big moment for the team at Modak Analytics, as they were taking on the biggest Big Data challenge in the recorded political history of the world. If they did this right, they would be front page news.

The Big Challenge

General elections in India were the largest in the world with nearly 814 mn voters on the electoral list, while the US had 193.6 mn voters, Indonesia 171 mn, Brazil 135.8 mn and the UK 45.5 mn voters. Unlike the US, where data was pretty much homogenous, data in India was largely heterogeneous because of the varied cultural, demographic, economic and social differences in the country. Owing to these hugely differential factors, each parliamentary constituency in India was unique and demanded a customized advanced analytics model that would collate data, conduct profiling, and classify each polling booth along with every voter. The Modak team undertook the mammoth exercise of collating, transforming and fusing details about each voter relating to demographics, economic status, caste and religion obtained from various data sources and then combining them with voter roll data (see Exhibit 5A and Exhibit 5B). This was done to help the political party in microtargeting and decision making.

The project had the following characteristics:

- The voter data was from close to 930,000 polling booths across 543 parliamentary constituencies, including 4,120 assembly constituencies.
- The information relating to 812 mn voters was available in 900,000 PDF documents, i.e., approximately 25 mn PDF pages. Data from these PDFs had to be deciphered and collated into a single user-friendly document. This exercise was further complicated by frequent updates that were made by various election commissions.
- Voter rolls were published in PDF format in 12 languages — English, Hindi, Telugu, Tamil, Kannada, Malayalam, Oriya, Bengali, Oriya, Punjabi, Marathi and Gujarati.
- Diversity of other names and information added to the challenge.

In addition, census data, economic surveys, and National Sample Survey Organisation (NSSO) data that were available at the village or district levels comprised the other sources of data. These data sources had to be mapped to polling booths. There were also some external data and the political party's proprietary data that had to be combined with individual voter data. The team realized that transliteration of voter rolls was necessary for mapping and combining these data. "This task is very complex and no other IT company has ventured into this. Or, even if they have ventured, they might not have succeeded," said Joshi.

With the help of its proprietary technologies, Modak Analytics processed this unstructured data, did fast transliteration, and also engaged in "fuzzy matching" based on Indian names, addresses, villages, etc. Data dictionaries were also developed for each state (based on the language), while automation tools such as RapidETL were used for transforming, fusing and merging massive datasets (see Exhibit 5C). Math quants and data crunchers enabled the conversion of mounds of electoral data into easily understandable information that could be used to create booth-level strategies for the political party's electoral campaign. As they progressed, Modak realized that these technologies, created in-house, had the potential to be patented.

Through this endeavor, Modak Analytics built the first-of-its-kind single repository of electoral data. Technology was a significant impediment to processing and managing the huge volume (18 TB) and variety of data. Modak Analytics had to build its own 64 core Hadoop servers. It created a Big Data infrastructure that included Hadoop and PostgreSQL to store and process this massive and frequently changing data (see Exhibit 5D). The master file that Modak created contained more than eight terabytes of data that was updated from time to time. Testing and validating the data was another important

challenge. To address this issue, heuristic (machine learning) algorithms were developed for classifying people based on name and geography. The heuristic algorithms made it simpler to identify religion, caste and ethnicity based on name and geography.

While conducting this data analysis, Modak Analytics offered several insights to the political party—insights that had important implications for the party's performance in the elections.

A few examples are as follows:

- Data on young voter population was missing. Based on this insight, the political party launched a pan-India campaign to reach young voters.
- The probable voting percentage at each polling booth was calculated using a predictive model with about 25 attributes.
- Classification of polling booths based on their affinity to vote for the political party.
- A comprehensive booth profile for each booth that helped the party come up with ground tactics and messaging strategies for each booth.

After the task was completed, Modak was ready to talk to the media about its work. However, the political party asked Modak to keep its work confidential and not go public. This was a big disappointment to the team at Modak since showcasing their work and gaining the limelight was important for them, particularly as Modak was a small start-up company. However, the brighter side of not being in any contractual agreement with the political party was that Modak could hold the data it had collected and analyzed as its own intellectual property. No one except Modak had any right over the 814 mn voter data that were processed and transliterated. Eventually, however, Modak did speak about the Big Data project to the media and got some publicity, while maintaining secrecy with regard to the political party.

Modak Analytics landed a couple of projects in the field of analytics in 2014, immediately after the election assignment. One was to build a compliance infrastructure for a major public sector bank in South India with total business approaching \$40 billion. Although Modak was a small player, it won against significantly larger players due to its domain expertise, technology and low-cost structure. As of August 2016, the project was on target for on-time delivery, and the bank viewed Modak as a niche analytics player with great expertise. The other data analytics engagement was the implementation of a Big Data platform for a credit risk company.

THE ROAD AHEAD

While data analytics projects were certainly coming Modak's way, the top management team was occupied with answering the big questions of positioning and sustainable growth for Modak. There were two paths visible. First, Modak could continue to be in the big data space and work on analytics projects from within and outside India. It could also look for a market for its RapidETL technology outside India. Second, Modak could be a leading data aggregation company in India, processing data from open sources and providing consumer information to a variety of companies such as FMCG companies, consumer packaged goods (CPG) companies, and others.

The Big Data and Advanced Data Analytics Space

The Big Data Market is expected to grow at 21.8 % compounded annual growth rate to \$ 41.5 billion in 2018.¹ In India, it was estimated to be about \$ 1 billion in 2015, according to NASSCOM estimates.² The benefits from a BDA (Big Data and Analytics) solution were mainly two: improved efficiencies

¹ IDC. *Big data and analytics is increasingly a game of inches*. Retrieved from <https://www.idc.com/prodserv/4Pillars/bigdata>.

² Agarwal, M. (2013, July 29). *How these startups are betting on growing needs for big data analysis*. *Iamwire.com*. Retrieved from <http://www.iamwire.com/2013/07/how-these-5-startups-are-betting-on-growing-needs-for-big-data-analysis/17700>.

ISB077

resulting in cost savings, and increase in revenue due to more sophisticated business insights. A survey conducted by Frost and Sullivan found that about 85% of the respondents recovered their costs within three years.³ Companies in the EMEA region (Europe, Middle East and Africa) and the Americas considered analytics to be more important than did firms in the Asia-Pacific region (see Figure 1). The major players in this industry were MuSigma, AbsolutData, IBM and Infosys.

MuSigma

MuSigma was an Indian company that provided analytics services to companies and boasted of having more than 140 Fortune 500 clients. MuSigma was ISO 27001 certified. Its strategy was to institutionalize decision support through solutions in data engineering, data sciences and decision sciences.

Absolutdata

Absolutdata was a global analytics products and services leader for more than 15 years specializing in big data analytics, marketing analytics and customer analytics. A privately-owned US firm based in California, its clients included companies in the retail, technology, pharmaceutical, CPG, travel and telecom sectors.

IBM

International Business Machines (IBM) was an American company headquartered in New York with a strong presence in the big data and information management space. Data management and warehousing, Hadoop, stream computing, content management and information integration, and governance were among its key capabilities in this space. As per its 2014 annual report, its total revenue was \$92.8 billion.⁴ IBM's analytics business grew by 7% to \$17 billion, i.e., about 18% of the total revenue.⁵

Infosys

Infosys was among the global leaders in consulting, technology and outsourcing. The company employed over 193,000 people in over 50 countries and reported revenues of \$8,711 million in 2015.⁶ The range of services offered by Infosys included Big Data strategy, architecture strategy and advanced analytics strategy.

Others

CapGemini, Wipro Analytics, Accenture, TCS Analytics, HCL Analytics, Cognizant Analytics and LatentView were also important players in this domain in India.

In any Big Data or analytics platform project, data integration—sourcing, transforming and fusing datasets—constituted about 80% of the work. Data integration in technical parlance was referred to 'ETL', which stood for extraction, transformation and loading. This work required extremely skilled resources and proprietary tools. Success in this space depended on creating a unique proposition where automation of key processes provided differentiated services to clients.

While working for the electoral campaign, the 10-member team at Modak was able to build one of the biggest and most complex databases in the world by analyzing Indian voter data. This was possible since Modak built and deployed innovative tools and methodology to significantly automate the data

³ Frost and Sullivan. (2015, December 30). 2015 big data and analytics market survey: Initial observations.

⁴ IBM. (2014). Annual report. Retrieved from http://www.ibm.com/investor/att/pdf/IBM_Annual_Report_2014.pdf.

⁵ IBM. (2014). Annual Report.

⁶ Infosys website. Financial snapshot: Fiscal 2015-2016, Q1 2017. Retrieved from <https://www.infosys.com/about/Pages/last-fiscal.aspx>.

integration processes. This automation resulted in considerable cost reduction (about 10 times) and time saving compared to the cost and time that would be required if traditional methodology for data integration had been used. Modak's RapidETL product had the potential to become a disruptive technology in this space. Modak wanted to market RapidETL to mature data analytics markets, such as those in the United States, Europe and Australia. It was also dreaming of becoming the biggest data integration company in the world. Further, Modak could also provide "end-to-end" services, including building predictive models, to help their clients. While India presented an interesting market for Big Data and Analytics, the US and Europe still remained the biggest target markets.

The Data Aggregation Space

Data aggregation companies identified multiple useful data sources and helped figure out consumer trends and patterns by aggregating information from these sources. The major players in this industry worldwide were Nielsen Holdings Plc and Experian Plc.

Nielsen Holdings Plc⁷

Nielsen was a global, independent measurement company based out of the United Kingdom. Its business was organized into two reporting segments—the "buy" segment and the "watch" segment. The "buy" business focused on total consumer measurement. It aimed to measure all consumer purchases at every touchpoint, (e.g. e-commerce, mobile devices and subscription models). While the United States was its primary market, Nielsen covered over 106 countries representing more than 90% of the world's population in its "buy" business. This gave it a significant competitive advantage because it was able to provide multinational corporations with a global view of their market share.

In the "watch" business, by 2015, Nielsen measured media in 47 countries. This represented roughly 80% of global advertising spending. Its digital audience measurement capability was also rapidly growing. By the end of 2015, Nielsen's Digital Ad Ratings metric covered 17 countries. This represented about 85% of global digital advertising spending. In the year ended December 31, 2015, the company reported revenues of \$6,172 million.⁸

Experian PLC⁹

Experian was a leading global information services company, providing data and analytical tools to clients around the world. It was headquartered in Dublin, Ireland and operated in North America, Latin America, the UK and Ireland, Europe, the Middle East, Africa and the Asia Pacific region. It had four business lines: credit services, decision analytics, marketing services and consumer services. It helped businesses manage credit risk, prevent fraud, target marketing offers and automate decision making. It collected information on people, businesses, motor vehicles and insurance as well as 'lifestyle' information from on and offline surveys. The US and the UK contributed towards a major share of its revenues. It employed about 16,000 people in 39 countries. Experian plc was listed on the London Stock Exchange and was a constituent of the FTSE 100 index. Total revenues for the year ended March 31, 2016, were US\$4.5 billion.¹⁰ It was rated by Forbes in 2014 as one of the "world's most innovative companies".

Acxiom

Acxiom was a 40-year-old, US-based enterprise data, analytics and software as a service (SaaS) company. It processed data from open sources and made information and insights available to companies. Acxiom provided marketing and information management services in multichannel

⁷ Nielsen. (2015). Annual report. Retrieved from <http://s1.q4cdn.com/199638165/files/online/annual/report/2015ar/assets/pdfs/Nielsen2015AnnualReport.pdf>.

⁸ Nielsen. (2015). Annual report. Retrieved from http://s1.q4cdn.com/199638165/files/doc_downloads/annual_meeting/2015/Nielsen10K2015.pdf

⁹ Experian website. Retrieved from <https://www.experian.com/corporate/experian-profile.html>.

¹⁰ Experian. (2016). Annual report. Retrieved from <https://www.experianplc.com/media/2733/experian-ar2016.pdf>.

ISB077

marketing, addressable advertising and database management.¹¹ It had a huge presence in the US, Europe, parts of Asia and South America.¹²

Modak's second option was to play in this space and become a leading data aggregating company in India that would provide demographic and socio-economic customer segmentation information to banks and insurance, FMCG, CPG and pharmaceutical companies. On the lines of companies like Acxiom, Modak Analytics could focus on processing more open data sources that were available and then provide a consumer index to marketing companies.

During its project with the political party, Modak used geocoding and created an enhanced demographic database that contained 814 mn potential prospects at the sub-neighborhood level. The company was already working on enhancing this database by adding two to three layers of additional open data to get the consumer index of a sub-neighborhood. These additional layers would enhance granularity and incorporate a lifestyle index of the consumers. The idea was to create a 250×250-meter data grid of India and identify the characteristics of people belonging to each grid. Companies could use this database for analyzing customer profiles and developing micro markets and associated marketing and sales strategies. This demographic and lifestyle database could be made available to companies through a subscription model.

Modak Analytics had already been approached in this regard by a leading market research firm that provided consumer indices based on market research and surveys to companies. The research firm took sample interviews of people in various cities and extrapolated the sample results to a pan-India level. It had proposed to work with Modak on a subscription model based on the electoral data that Modak already had. Modak would be required to give the research firm a sample based on certain criteria such as age, gender, education, etc., so that the interview sample would be accurate. As each interview cost the firm INR 500 (about \$10), the research firm was very keen on getting the right sample for the surveys. Apart from this, Modak was also working on developing proofs of concepts to validate the use of external data to profile customers for a leading bank and an Internet service provider firm.

In this approach, Modak would face competition from companies such as Nielsen and Experian, which were also trying to obtain socio-economic segmentation data in India. In addition, Acxiom and other similar firms were also likely to enter India in the next couple of years and could pose a huge challenge to Modak in the Indian market.

AT THE MOMENT

Joshi was struggling with the multiple alternatives that were further embedded in the two options; for example, if Modak pursued the first option, should it remain a domestic Indian operator or leap into international markets? Would a joint venture be an appropriate mode to pursue these growth options or should Modak "go it alone"? Which sources of funds would be feasible and appropriate for pursuing these options? Would venture capital funding be appropriate and what were the associated risks? Should Modak focus on banking or should it move into other verticals such as retail, healthcare, FMCG and pharmaceuticals?

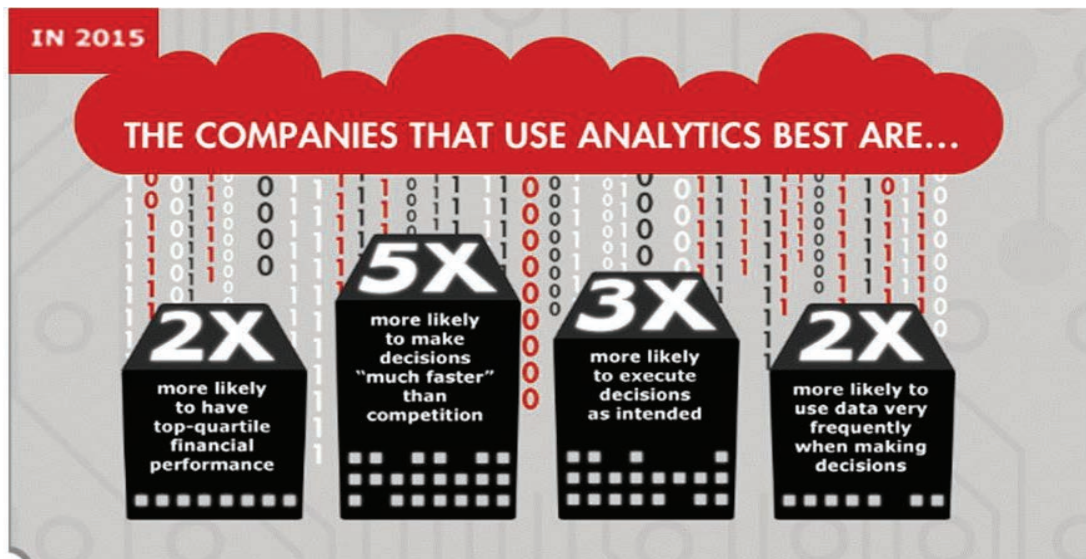
The future was looking a lot like a complicated open data source that needed a clean-up and streamlining to make sense. Joshi's analytical mind was energized by the challenge.

¹¹ Acxiom website. Retrieved from <http://www.acxiom.com/about-acxiom/>.

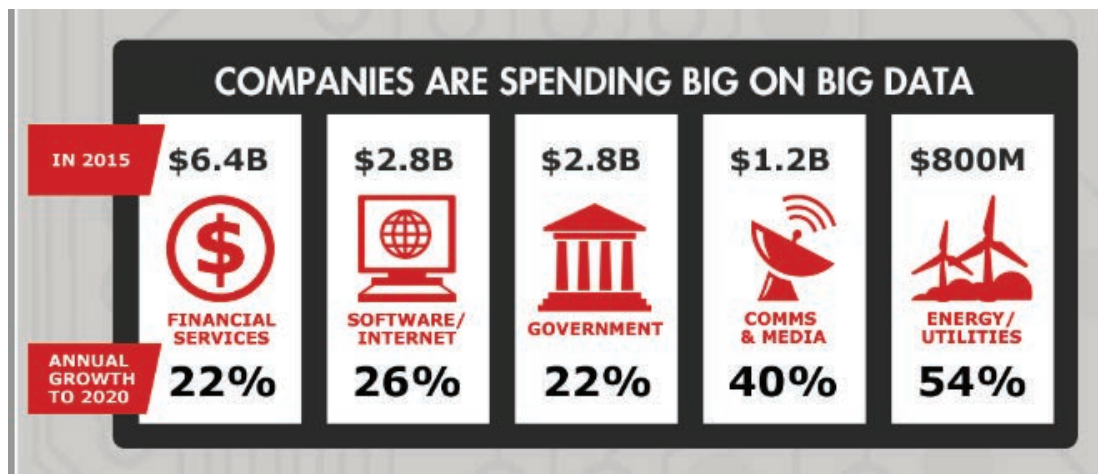
¹² Acxiom website. Retrieved from <http://www.acxiom.com/about-acxiom/locations/>.

EXHIBIT 1

BENEFITS OF DATA ANALYTICS FOR COMPANIES



SPENDING ON BIG DATA

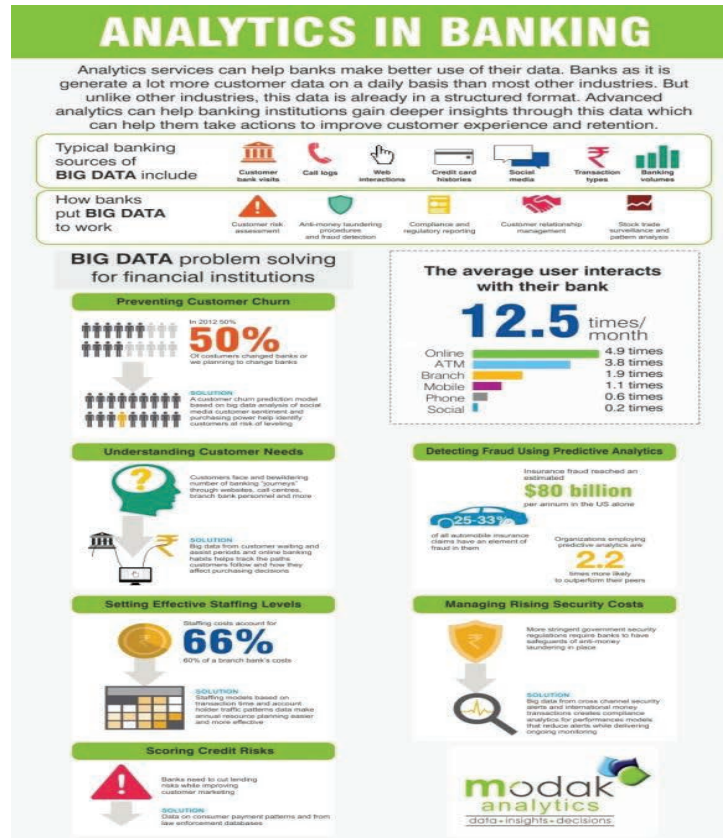


Source: Louis Columbus (2014, June 24). Roundup of Analytics, Big Data, Business Intelligence and Market Estimates. *Forbes*. Retrieved from <https://www.forbes.com/sites/louiscolumbus/2014/06/24/roundup-of-analytics-big-data-business-intelligence-forecasts-and-market-estimates-2014/#770f9e02388e>.

ISB077

EXHIBIT 2

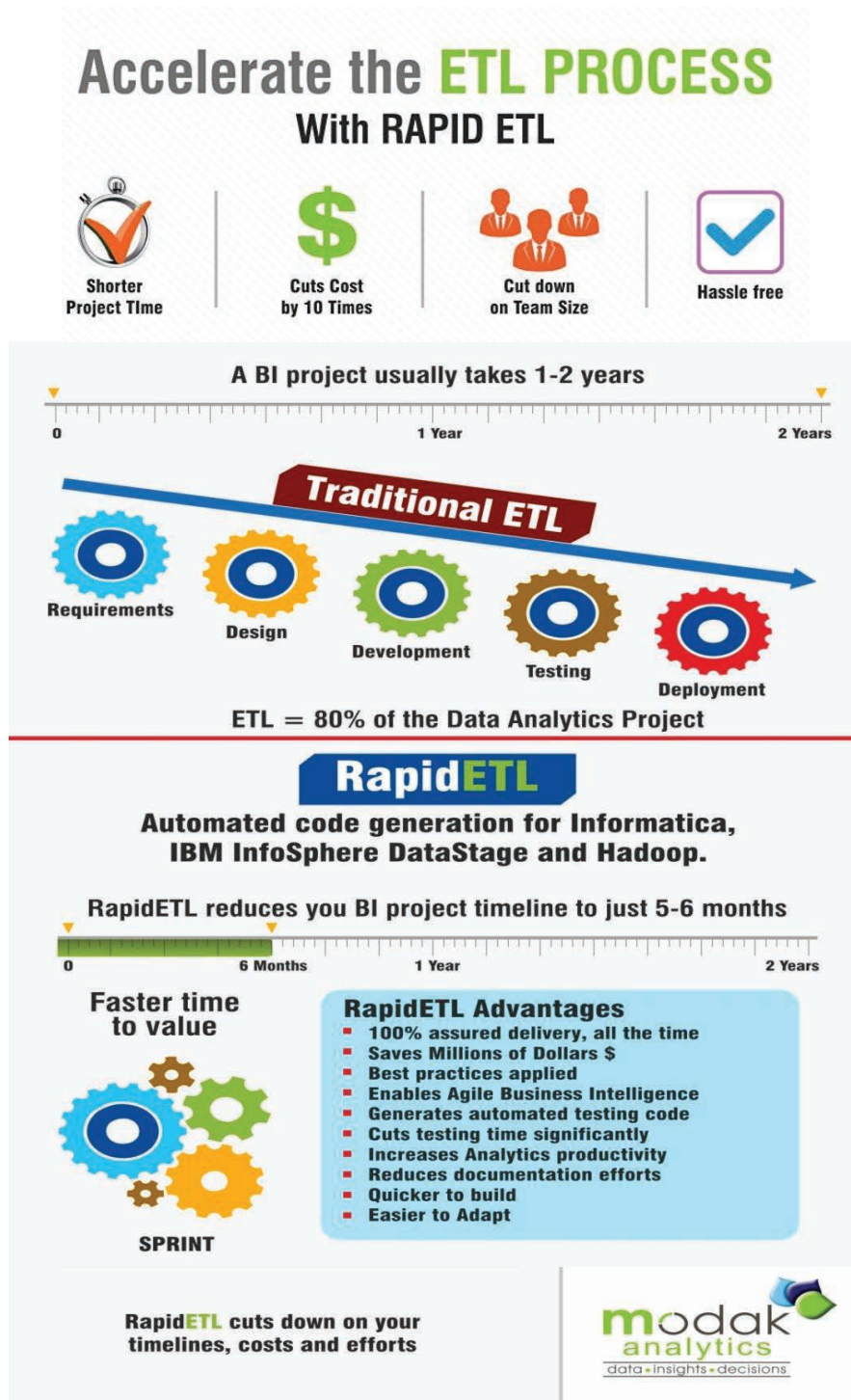
USE OF ANALYTICS IN THE BANKING SECTOR



Source: Company sources, Modak Analytics.

EXHIBIT 3

ETL AND RAPIDETL

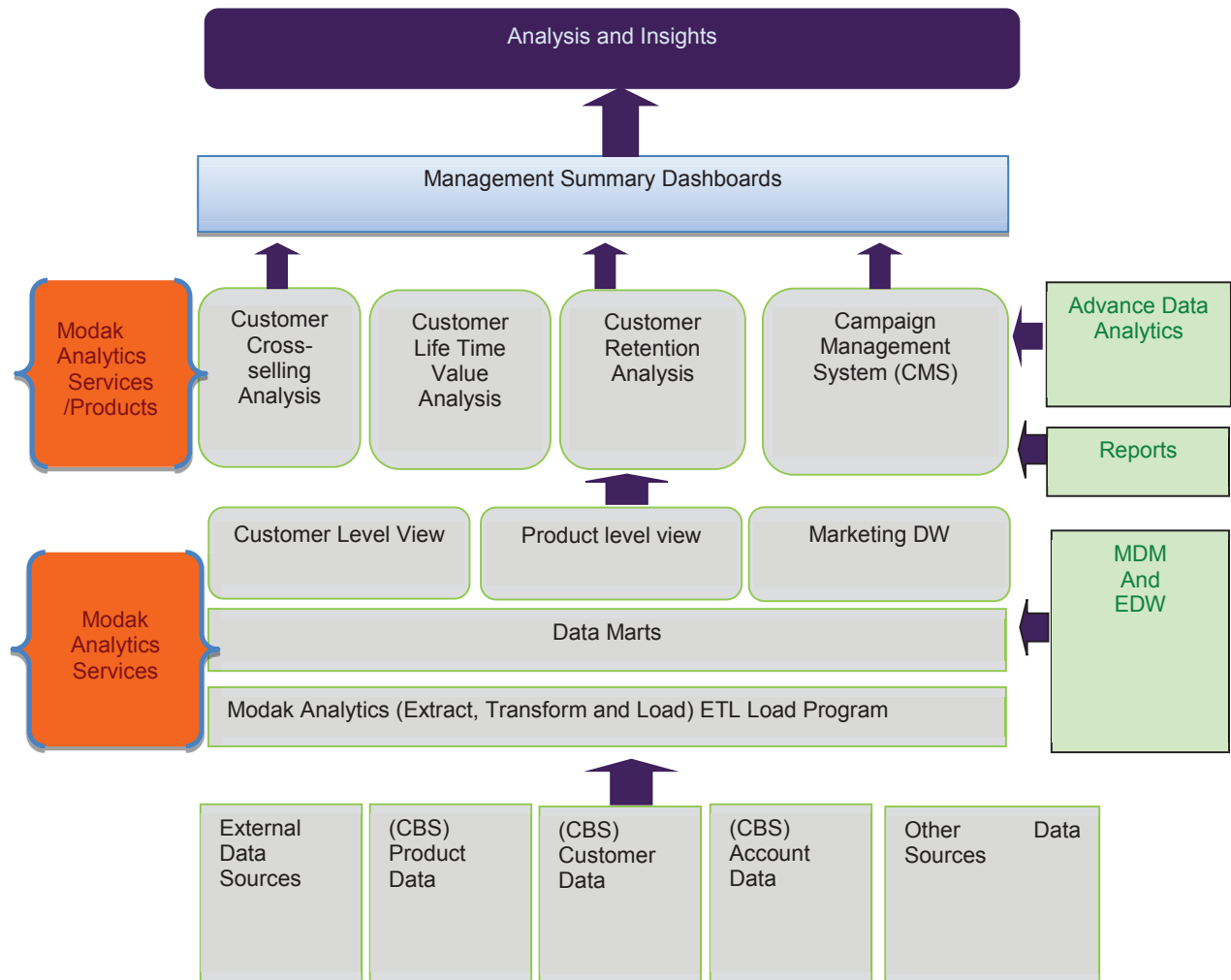


Source: Company sources, Modak Analytics.

ISB077

EXHIBIT 4

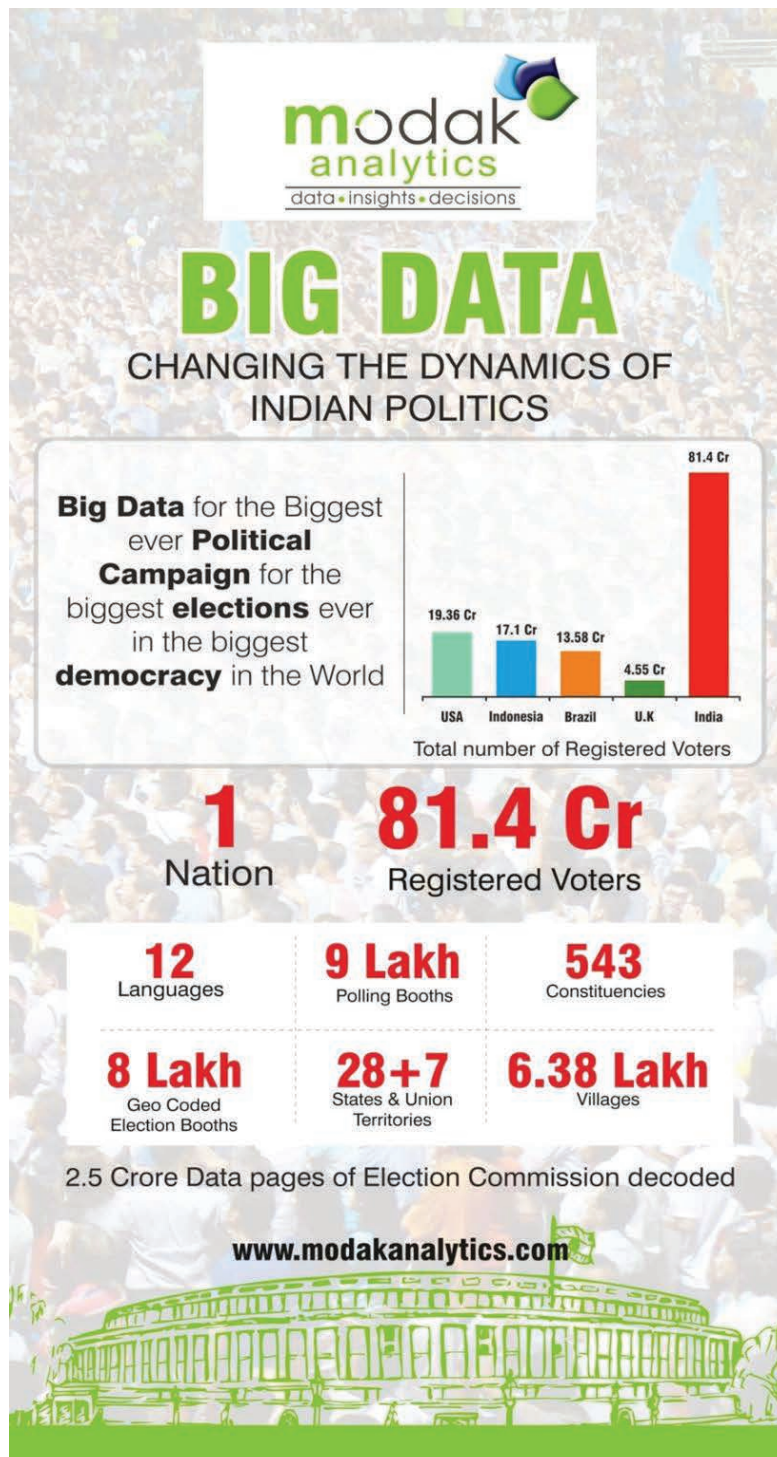
MODAK'S ENTERPRISE DATA WAREHOUSE



Source: Company sources, Modak Analytics.

EXHIBIT 5A

INDIA'S ELECTORAL BIG DATA



Source: Company sources, Modak Analytics.

ISB077

EXHIBIT 5B

DECODING INDIAN ELECTORAL DATA FROM 12 LANGUAGES

**DECODING
INDIAN ELECTION DIVERSITY WITH
BIG DATA**

2.5 Crore Data pages
of Election Commission decoded from
12 Languages

Overlaid on the background are several sample Indian voter ID cards in different languages. Visible text on the cards includes:

- Card 1: CQV0169607, नाम: बलराम (नववर्मी), पिता: अनिलराम (नववर्मी), आयु: 55, लिंग: पुरुष, जिला: 021/000255
- Card 2: AYR0320473, नाम: अरुण कुमार, पिता: विष्णु कुमार, आयु: 53, लिंग: पुरुष
- Card 3: IFC1407972, नाम: प्रकाश सिंह, पिता: निरंजन सिंह, आयु: 29, लिंग: पुरुष
- Card 4: LXV1268127, नाम: लक्ष्मी, पिता: राम, आयु: 44, लिंग: महिला
- Card 5: HHK1234566, नाम: हरीश, पिता: राम, आयु: 44, लिंग: पुरुष
- Card 6: KMY0267211, नाम: कमल, पिता: राम, आयु: 25, लिंग: महिला
- Card 7: KMB1364876, नाम: अतार सिंह, पिता: खेम चंद, आयु: 69, लिंग: पुरुष

www.modakanalytics.com

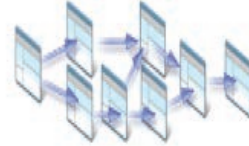
Source: Company sources, Modak Analytics.

EXHIBIT 5C

ILLUSTRATION OF A COMPLEXITY IN NAME MATCHING THAT AMOUNTED TO
“FUZZY MATCHING” IN THE INDIAN CONTEXT

Same names (Indian challenge)

srinivas
sreenivas
shrinivas
shreenivas
sriniwas
श्रीनिवास
ശ്രീനിവാസ്

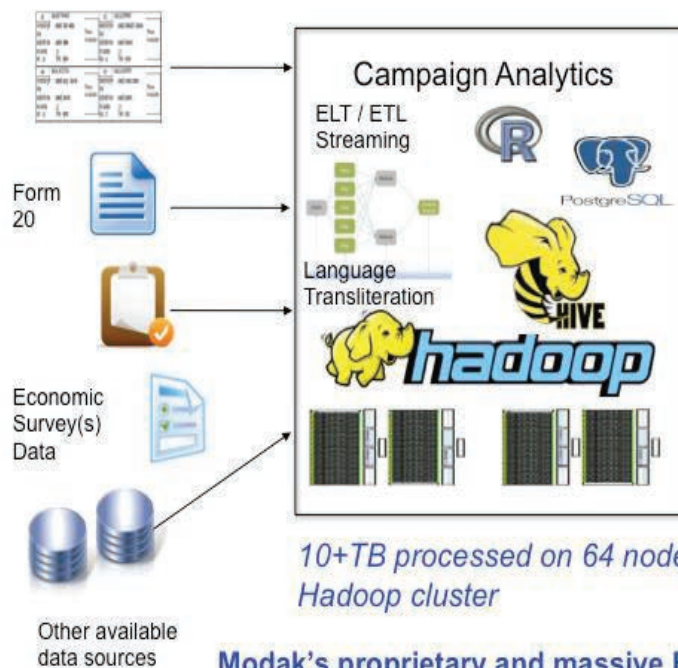


Fusing and Linking Data from various sources is massive challenge in India

Village Shekhpura, Nawada, Bihar (Indian challenge)



(Source: Company sources, Modak Analytics)



- Close to 81.1 crore voter rolls are processed, cleaned, standardized, and then loaded into the database
- Depending on the External Data sources, crores of voters are matched using Name, and Address, etc
- Advanced Heuristic based Name Analysis and matching for finding out relevant parameters
- House-holding done based on Addresses
- Name matching also done across languages
- R are used for Advanced Analytics on structured data and also for Text Analytics

Source: Company sources, Modak Analytics.

ISB077

EXHIBIT 5D

HADOOP INFRASTRUCTURE USED FOR THE BIG DATA PROJECT



Source: Company sources, Modak Analytics.

EXHIBIT 6

MODAK ANALYTICS SOFTWARE PRIVATE LIMITED COMPANY LOGO



Source: Company sources, Modak Analytics.