GIVING DATA A VOICE: THE RISE OF TALKINGDATA

Sia Siew Kien, Jun Jie Yang, Wee-Kiat Lim

On any given month in late 2017, more than 0.7 billion active mobile devices sent their data through different apps to TalkingData.[1] As one of the largest third-party Big Data service firms in China, its revenue had increased by 140% from 2015 to 2016, and jumped another 90% in 2017. More than 500 employees filled its Beijing headquarters in 2017, marking a 400% increase in headcount since 2015. Many of its new hires were data scientists and data engineers. Its partners and clients spanned traditional industries and newer internet-based sectors, including global giants such as Google, Qualcomm, L'Oréal, and Yum!, as well as established Chinese firms, such as Tencent and Ping An Insurance.

Diverse forays into financial services, real estate, and retail brought considerable promise and peril to TalkingData, especially with the increasing complexity in managing a fast-growing company and surging demand for deep expertise across different applications and segments. The startup had also ventured abroad in search of partnership and collaboration opportunities.

How should TalkingData's CEO, Leo Cui, continue to grow the company, sharpen its innovative edge, and take it beyond China's shores? What issues should he anticipate and address over the next few years?

"IN GOD WE TRUST; ALL OTHERS MUST BRING DATA"

The inspiration for the startup name, TalkingData, came from W. Edwards Deming, a famous quality management "guru". As recounted by Leo, "I read one of his books. It said, 'In God we trust; all others must bring data'. Thus, we named [our company] TalkingData, i.e., Talking with Data."[2]

---

[1] TalkingData. (2017, May). *Introduction of TalkingData* [corporate document].

[2] Chen, X. Y. (2014, February 12). TalkingData CEO 崔晓波深度专访：真正懂大数据的公司不说大数据. *Chinese Software Developer Network*. Retrieved from http://www.csdn.net/article/2014-02-12/2818242

The Asian Business Case Centre, Nanyang Business School, Nanyang Technological University, Nanyang Avenue, Singapore 639798. Phone: +65-6790-4864, E-mail: asiacasecentre@ntu.edu.sg

TalkingData's value proposition was clear from its early days. Leo and his co-founders saw abundant opportunities in filling a data analytics void in China. It was a paradox. On one hand, digitalisation was becoming more pervasive, resulting in a proliferation of data.[3] On the other hand, despite the availability of data and sophisticated data analytics tools, most businesses in China rarely used much data in their decision-making. Instead, they continued to make decisions based on their management instincts.[4] TalkingData wanted to help businesses cross the chasm to leverage data and data analytics to make better-informed decisions.

SCALING UP AND DIVING DEEP

As the mobile internet business began to take off in 2012, many new industries were emerging; online gaming was one of them. An online game provided a closed and well-defined virtual environment that could not only collect data on user behaviour, but also standardise and quantify user activities. TalkingData could generate indicators such as application usage, retention rate according to specified periods, levels of play, and purchases made. The company also analysed users' behaviour during special in-game events that ran for limited periods of time such as a day, a week, or a month; this allowed TalkingData and game developers to glean more insights on player behaviour to enable benchmarking.

The gaming industry turned out to be the perfect laboratory for TalkingData to test its ideas, products, and business analytics models. One of the innovations arising from its experience with the sector was a dashboard that could organise a set of key indicators and methodologies to analyse the game and in provide facilitate TalkingData's consulting services.

TalkingData was able to collect data when developers used its software development kit (SDK) to create their applications, which, in this context, were games. The code embedded in the games would collect data from the devices in which the games were activated and played. TalkingData would receive information, or "data breadcrumbs" in its parlance, gathered at the device level.

According to Dai Min, TalkingData's marketing director for overseas businesses, the gaming industry had a business model that was suitable for data analytics: *"Gaming is actually a business that is heavily data-driven. It has a clear business model on how to go about making profits [by responding to data]."*[5]

TalkingData quickly cornered China's gaming analytics market, with a market share of over 70% in less than two years.[6] Gradually, Leo realised its approach toward the gaming sector could also be applied in other industries. The company used the methodologies it had refined in its gaming analytics operations to help clients make sense of their mobile app metrics and identify relevant key performance indicators (KPIs).

Due to their technological advantage, TalkingData soon attracted a few large companies as their clients, such as China Merchants Bank (CMB). In 2013, the bank approached TalkingData after experiencing technical problems in producing a data analytics SDK. This was TalkingData's first time working with a traditional company.

During its experience with gaming analytics in its early days, TalkingData developed several tools and refined its data collection and management methodologies as well as its data analytics capabilities. Its consulting services were fast becoming its core business. This came as no surprise. With a wide repertoire of tools, products, and platforms for data analytics under its belt, TalkingData had been advocating consulting as a means to thread them together and provide a customised, end-to-end package to its clients.

---

[3] Zhang, H. (2017, February 8). 拒绝 BAT, 它要为超过 40 亿移动终端提供服务. *界面*. Retrieved from
    https://www.jiemian.com/article/1100353.html
[4] ibid.
[5] Dai, M. (2017, December 6). Personal interview.
[6] Dai, M. (2017, September 7). Personal interview.

Consulting was especially valuable to large enterprises as it allowed TalkingData to go beyond ad-hoc and piecemeal tools or data services and offer its clients integrated solutions that aligned with their respective strategies and goals. It would help enterprises figure out how to deploy infrastructure and processes to collect and manage data, as well as analyse user preferences and establish operational KPIs. TalkingData would also collaborate with partners to deploy infrastructure (e.g., cameras and face recognition software) and technologies (e.g., face recognition artificial intelligence) for specific requirements.

To impart knowledge on data science and to build a community of innovation and collaboration, TalkingData also established TalkingData University (TDU). This community would consequently grow the Big Data analytics market in China, notwithstanding TalkingData's share of the pie.

TDID: ONE ID TO FIND THEM, ONE ID TO RULE THEM ALL

TalkingData collected a significant volume of data through mobile devices. When consolidated and purposefully integrated, it created a holistic picture of consumers across a wide range of domains. For example, the gyroscope and accelerometer in mobile devices allowed TalkingData to guess with some level of confidence the gender of a device's main user. This was possible as men and women typically treated their devices differently: men tended to keep their phones in the front pockets of their pants; women, on the other hand, usually carried theirs in purses or handbags.

TalkingData consolidated these data from multiple sources using TDID, a unique data identification number (ID) system it had developed in-house. The TDID was also a secured and depersonalized identifier used to consolidate data across disparate devices (smartphone, tablet, and laptop), functions (e.g., payment for a belt at an apparel store or for a cup of coffee in a café), and applications (e.g., web browser). (See **Exhibit 1** for an overview on TDID.) This approach produced rich records of individuals which could be mined for insights on anything: from consumption and gaming preferences to physical stores visited. TDID was not just a back-end data analytics tool; it was robust enough to be employed as an identity verification service for multiple partners and suppliers such as telecommunications providers (e.g., China Mobile), map suppliers (e.g., SuperMap) and e-commerce retailers (e.g., Jindong).

NEUTRALITY AND INDEPENDENCE AS ITS STRENGTH

TalkingData merged data from its collaborators to provide its clients with more comprehensive insights. It built its own data transaction platform, the Smart Data Market, by encouraging its collaborators and data suppliers to join its platform. This enabled TalkingData to organise and analyse data legally and efficiently despite their diverse origins.

It also offered advanced data analytics services, such as predictive engines as well as lookalike and context-aware modelling. With the aim of gathering reliable, standardised, and validated data from various sources at one location (i.e., TalkingData), the company hoped that more developers and enterprises would become partners and data suppliers, and welcomed collaborators from multiple domains, such as finance, automobile, real estate, and travel. This way, clients could understand their businesses not only from data generated in-house (with the help of TalkingData), but also from data provided by other data suppliers. For instance, Fangjia.com, a real estate startup, contributed data on property sales and rental prices of specific locales.[7] TalkingData was proud to proclaim that it was the only neutral third-party data management platform (DMP) in China.

> "Data requires one to remain neutral. Once a giant making a strategic investment invests in you or acquires you, your business loses its meaning. We would rather (leverage on neutrality to)

---

[7] TalkingData (2017, April 25). *Product overview* [internal document].

build higher barriers of entry. As long as we have higher barriers of entry, it is hard for others to catch up with us."[8]

Leo Cui, CEO of TalkingData

For this reason, TalkingData had resisted acquisition offers from several large companies, including Baidu, Alibaba, and Tencent, choosing to remain an independent company to deepen its "neutral" edge and expand on its own.

"IT'S LIKE YOU'RE ALWAYS STUCK IN SECOND GEAR…" TACKLING THE CHALLENGES OF INDEPENDENCE[9]

Striking out on its own was no easy feat. Cooperation among data providers, data platforms, and service providers was fast becoming an industry norm, and TalkingData realised that there was a pressing need to transform itself to make it to the top. Instead of being just another IT and data technology (DT) solutions provider, it had to position itself as the architect of the data analytics ecosystem in which it had thrived.

TalkingData realised the need for collaboration whilst servicing a real estate company. TalkingData did not have the requisite expertise to fully address the client's concerns, particularly in engaging and managing various stakeholder and user groups, such as developers, constructors, and government sectors. As such, it sought out professional teams to work together on the project. Since then, TalkingData realised that strategic partnerships could help it better fulfil its clients' business demands in a comprehensive manner. Over time, together with partners, it built extensive data analytics capabilities in the real estate sector, e.g., in profiling property buyers, analysing human traffic, and understanding migration trends in a city.

> *"[Our partners'] strengths lie in [the fact] that they understand businesses, they know what clients want and they possess algorithm modelling capabilities. Yet they lack data, which we have. We then realised that an ecosystem would be the solution to such a business situation."*[10]

Jiang Qi, Vice President and Head of TalkingData's Data Commercialisation Unit (DCU)

To act on this realisation, TalkingData reorganised itself into three business units: the (1) data commercialisation unit (DCU, 数据商业化单元), (2) data transaction unit (DTU, 数据交易单元), and (3) data partner unit (DPU, 数据应用合作单元).

DCU provided full business solutions for clients in the same way that the consulting arm of a traditional software company would. In other words, it provided data services such as building data sets, and was also involved in implementing the recommended solutions on clients' premises and helping them transform their data capabilities, such as building their in-house data management platform to manage their own data assets, performing customer profiling, and managing marketing campaigns.

DTU offered standardised data products and other data-related services such as preparing data for advertising purposes or operations like customer profiling. Its activities included tagging data attributes as well as encrypting and sending data sets to clients to better inform their marketing campaigns.

DPU's *raison d'être* was direct revenue or profit generation. It focused on building more and better connections with TalkingData's technology partners and data partners, such as telecommunications operators. This involved improving partners' understanding of data operations and quality control, and

---

[8] Chen, X. Y. (2014, February 12). TalkingData CEO 崔晓波深度专访：真正懂大数据的公司不说大数据. *Chinese Software Developer Network*. Retrieved from http://www.csdn.net/article/2014-02-12/2818242

[9] TalkingData. (2017, July 21). *2017 年年中总结和下半年计划 – 战略和海外 v3* [internal document].

[10] Jiang, Q. (2017, December 6). Personal interview.

advising them on how to structure their service licensing agreements, especially for those new to such collaborative arrangements. In fact, DPU was the *de facto* representative of TalkingData's partners, and even helped the latter in identifying revenue generating and sharing opportunities with TalkingData's other partners and clients.

Focusing on "openness, connection, security, and intelligence", TalkingData continued to build a suite of products and services that catered to businesses using data and data analytics. While many of its customers were plain end-users (e.g., banks using data for new financial app development), some that approached TalkingData were active players in the data value chain, performing diverse roles of data production and collection, processing, integration, analysis, and presentation. Many became its partners. Most of these firms were tech giants or prominent industry players, such as Tencent, CMB, Didi Chuxing, and China Telecom.

TalkingData also put in place privacy policies explaining how they collected, used, stored, and transferred information and highlighting its compliance with laws and regulations in China and overseas (e.g. GDPR[11]).[12] In addition, it created a new data governance team (under their Research and Development department) that would attend to privacy issues, among other critical issues concerning data and quality control. Together with legal affairs, the data governance team would supervise and ensure the implementation of processes pertaining to data security management, such as data protection, desensitisation, and defence. Data desensitisation, for instance, involved the application of standardised labels to ensure that a user's identity was kept confidential.

LOOKING FORWARD: NEW GROWTH OPPORTUNITIES

Exhibit 2 shows the shift in TalkingData's revenue mix over the years, reflecting its change in strategy and reference point for benchmarking. As the CEO's Assistant, Hui Yang, said, "Previously, our benchmark was Palantir[13]…We can't find a firm to benchmark [against] in China in terms of the business model. Now, our potential competitors include consulting firms and digital service providers".[14]

TalkingData did not merely perform Big Data analytics or offer AI services. It also had unique first-party data, and, unlike telecommunications operators, weather bureaus, and insurance companies (which also had first-party data), it built data marketing and partnership capabilities to create value for its clients. It was not a pure enterprise solutions company like SAP and IBM, even though it had data scientists, engineers, and other experts to implement digital solutions. TalkingData was capable of engaging various verticals as it could rely upon partners for specialised applications, such as risk management and advertising tracking. Its unique resource and key differentiating factor was the data analytics platform that Leo and his team had assiduously built, refined, and even rebuilt over the years. The company keenly monitored potential competitors that could emerge as "data juggernauts" like itself.

In addition, the company wanted to be more strategic in identifying and structuring its relationships with clients and partners. For example, TalkingData explored revenue sharing with clients and partners (e.g., licensing) in creating and deploying new products, tools, and data services that might potentially be redeployed in the future as part of its standard suite of offerings. Going beyond mobile data, it also planned to explore new data sources including Internet-of-Things (IoT) sensors and data streams from over-the-top

---

[11] GDPR refers to the General Data Protection Regulation (GDPR) (EU) 2016/679, which is a European Union (EU) regulation on data protection and privacy for all individuals within the EU and the European Economic Area (EEA).

[12] TalkingData. (2018, June 28). *TalkingData 发布 2018 年最新战略布局，探索发展新路径* [WeChat post].

[13] Palantir Technologies was a U.S. company founded by several tech entrepreneurs including Peter Thiel, a co-founder of PayPal, in 2003. Named after the all-seeing crystal ball in J. R. R. Tolkien's legendarium, the company specialised in Big Data analytics.

[14] Hui, Y. (2017, September 4). Personal interview.

(OTT) applications, such as media content delivered via the Internet instead of traditional subscription providers (e.g., pay TV or satellite providers).[15]

TalkingData was shifting its focus to two growth areas, namely marketing and financial businesses. The former covered all marketing activities and operations across various industries, from retail to food and beverages, particularly in segments where the company believed it could leverage on its data market platforms, tools, and partners to deliver value that no others could. The latter area included banking and securities firms. TalkingData aimed to lead the market in helping financial institutions solve their pain points in digital transformation, in areas such as in mobile finance, risk control, and creditworthiness in retail banking.

Moreover, TalkingData was keenly aware that it would have to contend with global competitors. Since completing its Series B fundraising in late 2014, TalkingData had set its sights on global markets, with an eye on leveraging other companies' data technology and data science in China and overseas. To prepare for more intense competition on the horizon and a more global outlook, TalkingData had been working on its investment strategy. Two investment funds were established. One focused on tech startups and companies based overseas, in markets such as the U.S. The other fund focused on Chinese companies that successfully applied Big Data technology to specific domains, such as real estate, finance, and fast-moving consumer goods.[16]

Leo acknowledged that TalkingData's technological edge still lagged behind similar U.S. firms, but he believed that China should gain a significant lead over other countries in the next three to five years, given that it had the largest market for data application.[17]

Meanwhile, TalkingData would step up its efforts to explore partnerships with international tech firms such as Kaggle, Kochava, and Google[18] as well as with traditional global brands, including Tiffany & Co, Prudential, and Mindshare.[19]

From its humble beginnings to boasting a staff strength of 500, and from applying its technology solely in the online gaming domain to an extensive reach spanning finance, real estate, new retail sectors, and markets beyond China, TalkingData had grown significantly over the past six years. As the company poised itself to expand further, how could Leo continue to grow the company, sharpen its innovative edge, and take it further beyond China's shores? What issues should he anticipate and address over the next two to three years? How much potential does TalkingData have, and what will its next growth trajectory be?

---

[15] A more specific example would be the launch of the service HBO Now in 2015. HBO, a traditional cable TV channel, marketed itself directly to consumers instead of distributing its content only via pay TV providers (e.g., Comcast in the U.S.). This move was seen as a response to over-the-top (OTT) entrants such as Netflix and Hulu. There are different revenue models for such OTT services. Some are subscription based, while others may offer tiered access (e.g., free versus ad-supported subscriptions) and on-demand content.

[16] Zhang, H. (2017, February 8). 拒绝 BAT, 它要为超过 40 亿移动终端提供服务. *界面*. Retrieved from https://www.jiemian.com/article/1100353.html

[17] ibid.

[18] Kaggle was a predictive modeling and data analytics platform founded in 2010 and based in San Francisco. It was acquired by Google in 2017 but remained a distinct brand. Kochava was a mobile app analytics provider founded in 2011 and based in the U.S.

[19] TalkingData. (2017, July 21). *2017 年年中总结和下半年计划 – 战略和海外 v3* [internal document].

ASIACASE.com
THE ASIAN BUSINESS CASE CENTRE

EXHIBIT 1: TDID
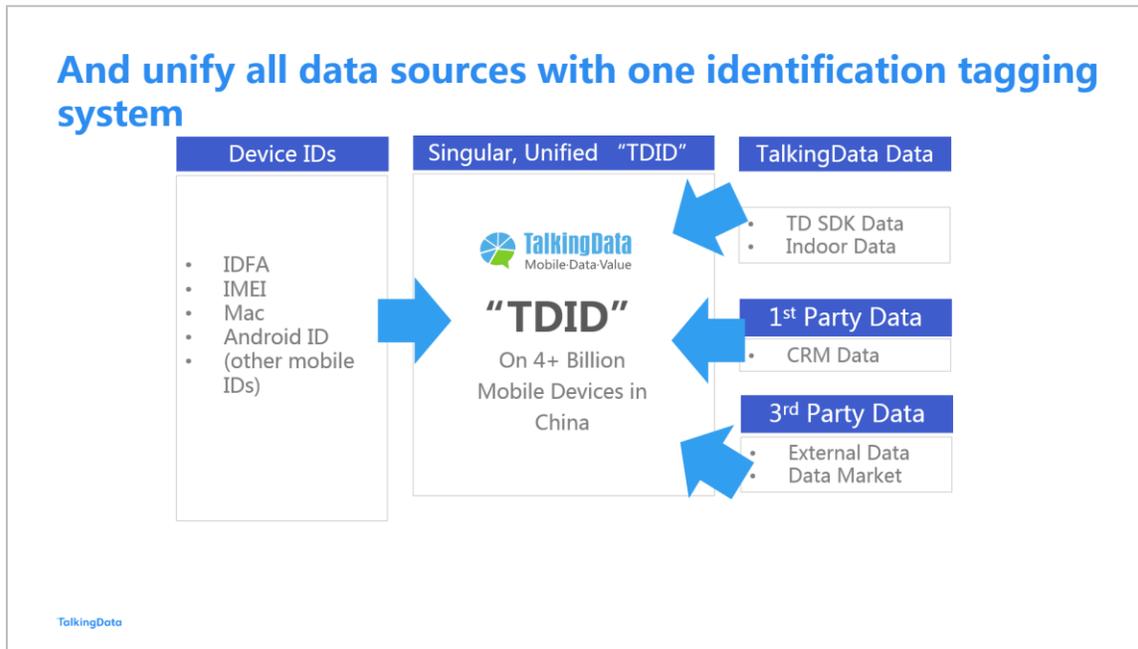
Note: IDFA = identifier for advisers, an IOS system identifier. IMEI = International Mobile Station Equipment Identity, a unique identity for used for phones, Media Access Control address (MAC address) is a unique identifier assigned to network interfaces for communications on the physical network segment. Android Device ID is the specific alpha-numeric Identification code associated with a mobile device in Android phones.
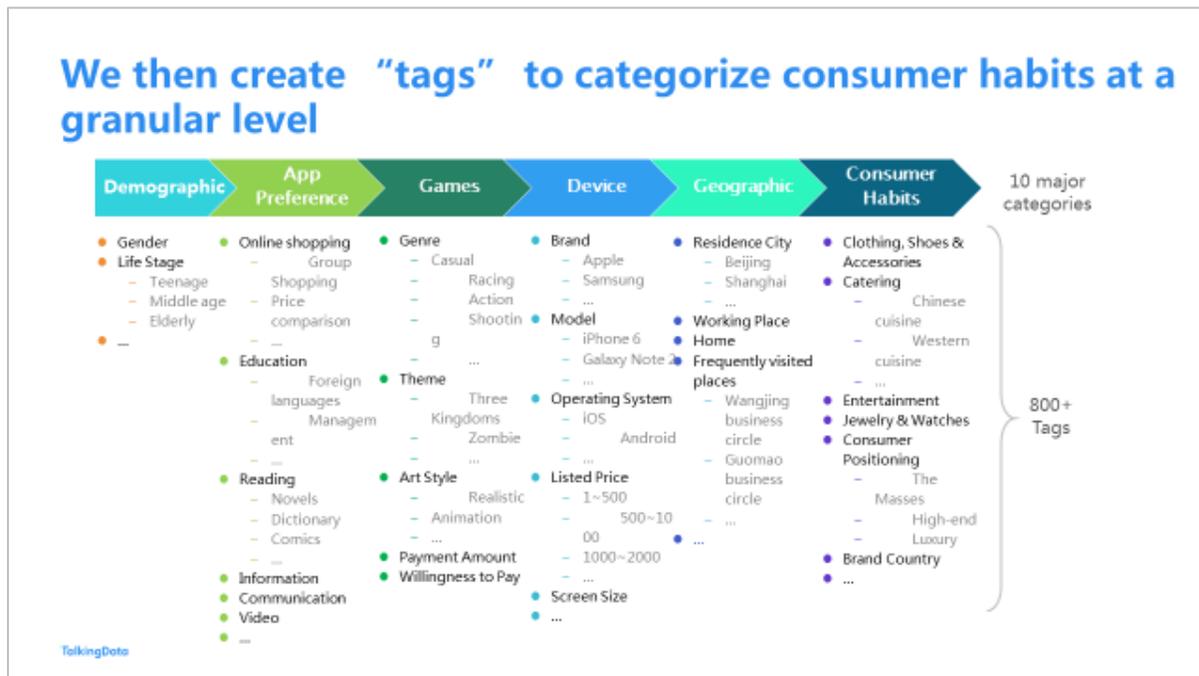
Source: TalkingData

EXHIBIT 2: REVENUE PROFILE OF TALKINGDATA

Source: TalkingData

AsiaCase.com
THE ASIAN BUSINESS CASE CENTRE
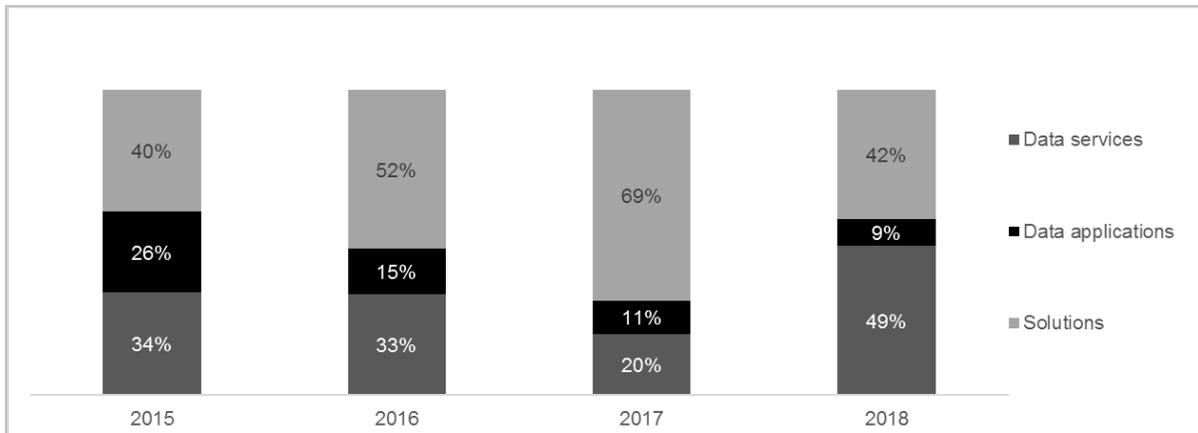
APPENDIX 1: INTRODUCTION TO BIG DATA ANALYTICS

While Big Data was an attractive proposition for many companies, it presented tremendous challenges, especially to those that were new to Big Data and data analytics. Few could handle the overwhelming volumes of data that streamed in at a breakneck pace from their business operations. Even fewer could make sense of the complex and multidimensional[20] data originating from multiple sources.

Take the retail sector, for example. What types of data might an apparel company have collected as part of its routine business operations? It might have captured the purchases that customers have made online and in its outlets in shopping malls. Online purchases could have been made through its own online store as well as third-party online portals such as Amazon. It also might have recorded customer feedback via online reviews of their products, email correspondences, chat bot conversations, and even audio content of customer helpdesk telephone calls. It might have a database of newsletter subscribers keen on updates on its sales and special events. It might also keep track of operations information such as inventory levels in-store or in storage for a diverse range of apparel, from blouses, skirts, to trousers of various designs and seasons.

What if the company decided to put in place new sensors and advanced cameras that could capture how shoppers – especially its most valuable customers – behaved on their shop floors? Where did they linger the longest? Which corners were ignored? Similarly, what could the company learn from consumers' online browsing behaviour? What were the typical items "saved for later" in online shopping carts? Which customers ultimately made purchases online and how did they make their buying decisions? What about those who browsed the online collection before, after, or even during a visit to a physical store? Could data (such as footfall) help the company negotiate its retail rentals with the mall management and help it plan its resources (e.g., store attendants), especially during peak periods? How could the company gather and make sense of the various types of data without needing to overhaul or even replace its existing IT systems? Furthermore, could data be analysed in real-time, for example, to identify criminal behaviour (e.g., fraud, shoplifting) and troubleshoot poor shelving during sales at specific stores?

It goes without saying that being able to tame the volume, variety, and velocity of Big Data would help companies glean precious insights for evidence-based decisions. Given the myriad benefits of Big Data, it is not surprising that several data analytics firms have popped up to provide expertise on using this new technological resource. TalkingData is one example of a fast-growing data analytics company riding the Big Data wave.

TalkingData's Origins in Gaming Analytics

TalkingData earned its stripes helping game developers achieve better performance with insights distilled from data analytics. Its first forays in online gaming analytics tools such as mobile applications analytics and mobile advertisement tracking formed the foundation of its data analytics capabilities. These tools helped app developers, marketers, and advertisers understand their users better and fine-tune current products and services, and were also useful in developing and introducing new ones.

Gaming app analytics shed light on user behaviour that aided developers in improving engagement and achieving their goals. For example, by analysing user dropout rates across different levels of the game, developers could decide whether to adjust the difficulty level to encourage its users to continue playing instead of leaving the game. They could also draw meaningful insights from information on how much their users spent, how their users made payment, and what their users spent on most readily (see **Figure 1**).

---

[20] Multidimensional data refers to data that typically has more than two attributes (e.g., time, geolocation, intensity of signal, and duration of call).

Figure 1: Screenshot of Revenue Statistics by Time Period
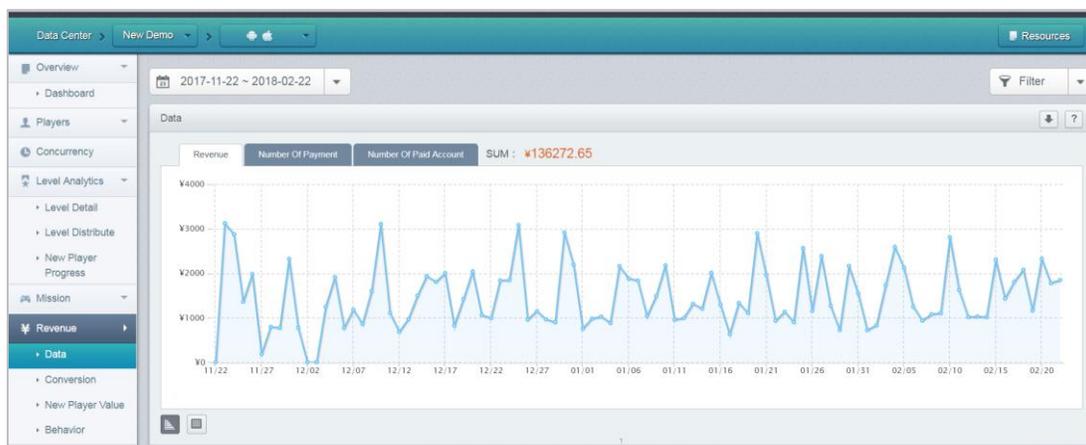
Source: TalkingData

Game developers had at their disposal research methods such as A/B testing to gauge the effectiveness of messaging and content design in achieving their objectives, for example, in prompting players to buy new props or to continue playing the game. App analytics enabled developers to go one step further: they could refine and optimise the flow of their in-app processes to make them more efficient and user-friendly, especially to new registrants during sign-up. With ad tracking analytics, marketers and advertisers would be able to tell at one glance how well their ads performed (e.g., based on the number of click-throughs) across multiple platforms. They could drill deeper into real-time data to obtain specific user information, such as the type of operating system. In addition to tracking by platform and channel, they could also track the performance of each ad campaign, and glean insights from the more successful ones to inform their future marketing strategies.

In essence, TalkingData relied on its own unique methodology that was developed in-house to understand specific target groups. Through building profiles using dimensions such as demographics, wealth, hobbies and interests, and brand preferences. Combining such information with other data sources, the company could learn about a lot about a city, including its human traffic patterns, state of urban development, and general social trends such as population migration. These led to valuable insights such as the most effective promotional channels to reach and influence the city's population. By performing competitive analysis, TalkingData would be able to build a richer picture of its clients' target audience and competitors, and thereby create the most value for its customers and partners.