



LYNDA M. APPLGATE  
KARIM R. LAKHANI  
NICOLE BUCALA

## Podium Data: Harnessing the Power of Big Data Analytics

*There is always value to be found in data but the question your organization has to address is this – do we allow or even encourage one-off, independent analysis of information in silos . . . or do we formalize that effort, and try to sustain the value-generating skills we develop?*

– Andrew White, Gartner, July 2014<sup>1</sup>

In early March 2015, two veteran entrepreneurs sat in their one-room headquarters in a turn-of-the-century building in downtown Lowell, Massachusetts, and debated the next steps for the growth of their new venture, Podium Data. The company website summarized Podium Data's mission and value proposition:<sup>2</sup>

PODIUM DATA HAS ONE GOAL: To dramatically lower the time, cost, and effort of managing and using business data. Creators of industry-leading Podium and DARE products, Podium Data has embedded decades of know-how onto a disruptive big data platform that accelerates time-to-answer 10100x.

Posted on their office walls were logos of Podium's first few customers, stretching across industries including financial services, media, and pharmaceuticals. Since launching the new business one year earlier, Podium Data cofounders CEO Paul Barth, COO Steve Richards, and Technologists Bob Vecchione, Michael Howard, and Atif Majid had developed what they believed to be an innovative software platform to collect, manage, and transform data from a wide variety of sources that solved their customers' most pressing business problems in a fraction of the time and at a fraction of the cost when compared to existing solutions. (See **Exhibit 1** for the bios of the Podium Data founding team.)

The Podium Data team had just celebrated the first birthday of their new venture, a software company capitalizing on the application of disruptive "big data" technology in large businesses. The product vision was born out of years of strategic consulting to Fortune 500 companies at NewVantage Partners, a data- and analytics-driven consultancy that helped companies create and capture value from the vast stores of information generated within large enterprises and their ecosystem partners (e.g., customers, suppliers, strategic partners).

---

Professors Lynda M. Applegate and Karim R. Lakhani and Nicole Bucala (MBA 2015) wrote the original version of this case "Podium Data: Harnessing the Power of Big Data Analytics," HBS No. 815-128 which is being replaced by this version prepared by the same authors. It was reviewed and approved before publication by a company designate. Funding for the development of this case was provided by Harvard Business School and not by the company. Certain details have been disguised. HBS cases are developed solely as the basis for class discussion. Cases are not intended to serve as endorsements, sources of primary data, or illustrations of effective or ineffective management.

Copyright © 2015, 2017 President and Fellows of Harvard College. To order copies or request permission to reproduce materials, call 1-800-545-7685, write Harvard Business School Publishing, Boston, MA 02163, or go to [www.hbsp.harvard.edu](http://www.hbsp.harvard.edu). This publication may not be digitized, photocopied, or otherwise reproduced, posted, or transmitted, without the permission of Harvard Business School.

During its first year of operation, Podium Data's cofounders self-financed their venture through salary cuts and consulting revenue from early customers. One year later, in March 2015, Podium Data closed on its first round of external financing. At just over \$3 million, the round was expected to fund growth over the next 12–18 months. (See **Exhibit 2** for the company's financing history and its plan for future financing.)

CommonAngels Ventures, an early-stage technology venture capital firm, led the seed round. James Geshwiler, Managing Director, explained why CommonAngels Ventures was eager to finance Podium Data:

We see many companies that call themselves “big data” companies. Podium Data stood apart because of the length of time and depth of IT and data management experience of the founding team and because they solved very specific problems for customers seeking to leverage big data analytics and were delivering significant value to paying customers.

While many technology companies expected to raise several rounds of financing, Barth and Richards were hopeful that their Series A round of \$20–\$25 million, which they planned to raise in mid-2016, would enable them to establish a leading position in the big data/analytics space. As they considered their next stage of growth, the founding team contemplated several important questions. Should they dive deep into one vertical, or serve customers across multiple industries? This decision also raised the question of which industry or industries they should target, and the order in which they should target them.

The team was also interested in ensuring that their early customers would opt for a subscription fee-based license model rather than a perpetual license model. Traditionally, enterprise customers paid for database software and applications with a licensing fee, but the shift from running applications in a company-owned or outsourced data center to delivering applications and hosting data on secure, on-demand cloud services platforms was challenging this traditional model. Would their current customers accept the shift in the pricing model from perpetual license to subscription fees for a hosted service, which Podium Data would run either within the customer's data center, or in a Podium Services cloud computing platform? Additionally, would revenue from subscription licenses keep pace with the funding needed to rapidly grow the company?

Finally, Barth and Richards knew that the decision to pursue equity financing from venture capital investors meant that they would also need to consider that these investors would most likely push Podium Data to consider an “exit” in three to five years after the last round of financing was raised. These exits usually involved an acquisition, an IPO, or a significant investment from a private equity firm that would buy out the initial investors and enable the company to rapidly consolidate an industry sector before considering an exit for the late-stage investors. Depending on the type of exit, different decisions would need to be made about how to create maximum value for the Podium Data's founding team and its employees and customers, in addition to its investors. Clearly, a challenging few years lay ahead.

## Becoming an Entrepreneurial Leader

A fourth-generation entrepreneur from the Midwest, Barth graduated from Hampshire College in Amherst, Massachusetts, and then moved two hours south to New Haven, Connecticut, where he received a Master of Science in Artificial Intelligence from Yale. After graduating from Yale, Barth joined Schlumberger's Connecticut research labs, where he developed, patented, and published

innovations in artificial intelligence, computer graphics, and parallel processing. Then, with the support of a full fellowship from Schlumberger, he moved two hours northeast and, in 1992, received a PhD from MIT's Electrical Engineering and Computer Science Department, or in MIT lingo, "Course 6."<sup>3</sup> Barth's dissertation on dataflow computing established new methods to coordinate thousands of parallel processing tasks, enabling high-performance "supercomputing."<sup>4</sup> He also attended courses at MIT's Sloan School of Management, where he developed a keen interest in business and was inspired to pursue a career in entrepreneurship.

After graduating with his PhD, Barth joined Thinking Machines, a pioneer in supercomputing, located in Cambridge, Massachusetts. There, he developed one of the earliest commercial marketing databases for American Express, which hired him to lead technology at the American Express subsidiary Epsilon, a database marketing firm. Shortly after, in 1995, Barth cofounded his first technology venture, Tessera Enterprise Systems, which leveraged emerging, commercial, parallel processing databases to enable businesses to analyze the vast stores of customer information in their production systems.<sup>5</sup> As Chief Technology Officer (CTO) of Tessera, Barth helped shape some of the first commercial data warehousing technologies. Tessera was successfully acquired by iXL four years later, and Barth stayed on as CTO until 2001.

Ready for another taste of early-stage entrepreneurship, Barth then cofounded, with Randy Bean, his second start-up, NewVantage Partners. Initially, NewVantage advised IT leaders and business executives of Fortune 500 businesses on how to harness the power of information in their organizations and the role of IT in delivering the business value of IT. By 2014, NewVantage described its mission as "providing data and analytics-driven strategic consulting services."<sup>6</sup>

Central to the NewVantage model from the earliest days was the mission to become a thought leader in the field. Initially, thought leadership took the form of a regular series of NewVantage dinners and breakfasts, during which a small group of C-level business and IT executives, along with the founders and thought leader guests, discussed key issues in harnessing the value of business information as they also shared insights on how to address those issues. In addition, the NewVantage Partners' founders also formed an advisory board of distinguished thought leaders who also participated in the discussions and on consulting projects. By 2014, the thought leadership advisory board and discussions had been supplemented with a yearly Big Data Survey, conducted for the first time in 2012, and a series of white papers and published articles in prestigious journals, such as *MIT Sloan Management Review*, the *Harvard Business Review*, the *Wall Street Journal*, *CIO Magazine*, and *Forbes*.<sup>7</sup>

Over the course of his work at NewVantage, Barth identified clients' pain points at every part of the "big data and predictive analytics" cycle: data creation, collection, preparation, analysis, and use. In addition, he recognized that, by 2014, big data and predictive analytics technologies had shifted from experimentation to widespread adoption, as enterprises sought to capture the intrinsic, "untapped" value in the vast stores of data that a company generated over the course of doing business.<sup>8</sup> Indeed, the *NewVantage 2014 Big Data Survey Report* found that 67% of the 125 C-level executives<sup>a</sup> who participated in the survey reported that they had big data systems in operation within their firm. An additional 17% said that they had big data initiatives underway, and another 12% were considering big data initiatives. Of the 108 respondents who either had operational big data systems and/or had new initiatives underway, 35% reported that they had invested \$10 million or more in these Big Data initiatives in 2014, and 6% reported that they had invested over \$100 million. More importantly, by 2017, these investments in big data initiatives were estimated to grow, with 74% expecting to invest

---

<sup>a</sup> These 125 executives held C-level positions within 59 large enterprises, 42% from industry leaders such as Aetna, American Express, General Electric, Fidelity Investments, Johnson & Johnson, and Kaiser Permanente.

more than \$10 million and 28% expecting to invest more than \$50 million. In addition, 88% of respondents believed that a strong partnership between business and IT was critical for success.<sup>9</sup> Bean commented on the findings: “Big Data was a new topic just a few years ago, with many companies grappling with its role in their organization. Now Big Data is largely viewed as a mainstream activity that is an integral component of how [enterprises] do business.”<sup>10</sup>

## Podium Data: From Vision to Reality

Most of the companies that sought NewVantage’s services found it increasingly difficult to identify the best approach to harnessing the value of data created within the firm, and the software to use to access, store, and analyze the data.<sup>11</sup> Companies often were not in the habit of collecting data in a form that made it easy to analyze; rather, most companies’ data was scattered across dozens of different systems and was often uncompiled, or “raw.” The task of preparing data for analysis often involved the tedious work of cleaning data, merging datasets, and putting them into a consistent format, which often took months or years of preparation and a significant investment in hardware and software, even before the data was ready to be analyzed to create value.

Barth had struggled with these issues in his work with clients. But, the spark for founding Podium Data came from a key insight that Barth had while working with a large NewVantage banking client. The bank’s compliance director told Barth’s team he could not quickly generate the stress test reports that were required in the aftermath of the debt crisis that had precipitated the global economic recession in 2008. The bank’s compliance team had only 90 days to generate the requested reports, a key component of which was the evaluation of multiple characteristics of every loan in the bank’s portfolio, which were stored in multiple systems in multiple locations across the country.

To solve the bank’s problem, Barth and his NewVantage team proposed a novel strategic plan that included a “data lake” that compiled vast quantities of data from multiple legacy IT systems onto an open-source big data software system called ApacheHadoop® (from here on referred to as Hadoop). The team helped the bank evaluate vendors offering Hadoop platforms, and the bank chose an appliance that included one vendor’s release of Hadoop and a small number of traditional database servers. In addition, Barth’s team went beyond their traditional strategy role to develop a custom application that provided critical data management functionality that was missing in Hadoop. The application greatly expanded the amount of data available to the business, allowing the bank’s compliance team to access the more than 4,000 files, with over 100,000 attributes of loan and customer information refreshed every day. More important, the new application on Hadoop was delivered at a significantly lower cost and in significantly less time than the bank’s IT group had estimated using traditional technologies.

Barth explained that the value delivered to the bank by the NewVantage team was measured in the tremendous savings in cost and time over the traditional technologies and approaches available in the market in 2013: “The client had forecast that the cost of hardware and software for the data storage and access solution would have been over \$10 million using traditional data warehousing technology. In contrast, the Hadoop hardware platform with the database servers cost less than \$1 million. In addition, the time needed to deliver the compliance application to business users was cut from almost two years to less than six months.”

### *Launching Podium Data*

Barth’s consulting work, coupled with surveys and reports that were being published by NewVantage, suggested that the market had never been hungrier for good data tools and services.

During 2013, Barth approached his NewVantage partner Bean to discuss launching a software company to commercialize a new product based on the insights learned during his work with the bank. “We recognized that the technology insights we had during the consulting engagement with the bank could be used to create a new software product that could be a potential game changer in the industry. We knew that the investment, brand, and business model for a software company would vary significantly from a strategic consultancy, so I proposed launching a new entity,” Barth explained. Bean added, “While I hated to lose Paul as a NewVantage principal, I knew he was itching to return to his roots and start another technology company.”

By late January 2014, Barth teamed up with Richards and the other members of the founding team to launch Podium Data. (See again **Exhibit 1** for the bios of the five founding team members.) The five cofounders spent the next few months writing the code to create a robust, proprietary version of the software based on Barth’s vision and then filed for a patent.

The initial financing of Podium Data came from the partnership arrangements worked out between Barth and Bean. Under the terms of this partnership, NewVantage would subcontract work relating to several consulting clients to Podium Data to complete projects already underway with Barth and his team. Consulting revenues from these clients, as well as projects sold by Bob Vecchione, funded Podium Data’s software development during 2014. Barth, Richards, and other members of the Podium Data founding team also took substantial pay cuts during the first year. This self-financing totaled \$1 million and funded office rent as well as the purchase of two used computer racks that served as big data clusters. These used computers were purchased for \$5,000 each; new ones would have cost at least \$50,000 each.

The Podium Data founding team, supported by offshore software developers who were well known to Atif Majid, spent the first six months of 2014 working on the initial release of Podium Data’s first two products.<sup>12</sup>

- **Podium** was an end-to-end data management and preparation platform built from the ground up to leverage low-cost, high-performance, big data technologies. Podium’s core features provided important data management functionality that enabled fast access to data from a wide range of data sources, including the many enterprise IT systems in place in most organizations, and the new sources of unstructured data not captured in existing IT systems. In addition to capturing the data, the Podium data management platform also identified and stored “data about the data” — often called “metadata” — that included the source and owner of the data, when it was created, and its business definitions, and kept a history for how it was used. Business users accessed the Podium data management tools through an integrated, easy-to-use browser interface. This allowed even the most unsophisticated business users to take full advantage of the cost effectiveness of big data technologies without their inherent complexities, significantly shrinking the time to access data and begin creating business value. More importantly, Podium did not require costly and time-consuming changes to existing IT systems, and it also enabled business users to analyze data using the many different data analytic packages (e.g., SAS, Oracle Business Intelligence, and IBM Watson Analytics) that were already in use in most organizations.
- **DARE** was the brainchild of Michael Howard, who discovered a way to transparently encrypt, encode, and compress sensitive data without any modification to existing applications or to the core Hadoop open source code. DARE ran in parallel with Hadoop’s core technologies, such as Map-Reduce, and allowed them to run faster due to Podium Data technology for accelerating

data input and output. DARE enabled users to securely migrate sensitive data to big data platforms located on their premises or in the cloud.

Building on many years of large-scale data management experience and the experience of consulting on big data and analytic solutions, Podium Data's products enabled business users to load and refresh data from many sources; view, search, and browse business and technical data; view data history and origination; create custom datasets; and schedule, load, and transform different analyses or queries. As such, Podium Data provided the platform and tools upon which new customized applications could be built quickly and at a much lower cost. (**Exhibit 3** provides an overview of the architecture.)

With the product platform in place, between July and December 2014, Podium Data ran a beta test of the platform for its first four customers, which stretched horizontally across several industries. The beta customers, which were paid engagements, worked with the Podium Data founding team to build a pilot that would solve a pressing problem for business users who required a big data solution. A short description of each customer and the big data solution that they worked on during the pilot is presented below:

- The head of commercial insurance at an insurance company who wanted to combine years of policy data from multiple, different customer information and claims systems to identify key factors driving policy renewal.
- The head of enterprise architecture at a major bank who wanted to deliver "Data as a Service" to business users throughout the bank.
- The head of marketing at a media company who wanted to combine advertising and customer data with social media and its customers' product data to deliver targeted advertising and marketing.

In addition, by 2015, three additional customers had begun pilot tests with Podium Data. Finally, two of the customers transitioned from the pilot to the production operations phases. (See **Exhibits 4a–4d** for a description of Podium Data's customers in early 2015, sample customer quotes, a Customer Engagement Model, and a comparison of big data and data warehouse approaches.) These three phases—Pilot, Transition, and Operations—formed the foundation for the Podium Data Customer Engagement Model, the details of which were still being worked out in March 2015.

### *Podium Data's Customer Engagement Model*

In early 2015, Podium Data engaged new customers through an eight-week, \$75,000 pilot project. During this pilot phase, the Podium Data team worked with the client to identify a pressing business problem that required that business data be accessed from multiple sources. The team then used its proprietary platform and tools to work with users to identify and access the necessary "data streams," to add the security and compression tools needed to efficiently store them in a Hadoop "data lake," and to simplify analysis and reporting.<sup>b</sup> Gartner analysts explained that, while traditional approaches to storing data in relational databases or data warehouses were built for a specific purpose:<sup>13</sup>

Data lakes focus on storing disparate data and ignore how or why data is used, governed, defined and secured. The data lake concept hopes to solve two problems, one

<sup>b</sup> The term "data lake" refers to a general-purpose data repository that was fed by multiple different types of data from multiple different sources. See Gartner, "Gartner Says Beware of the Data Lake Fallacy," press release, July 28, 2014, <http://www.gartner.com/newsroom/id/2809117>, accessed April 9, 2015.

old and one new. The old problem it tries to solve is the proliferation of information silos. Rather than having dozens of independently managed collections of data, you can combine these sources in a data lake. The consolidation theoretically results in increased information use and sharing, while cutting costs through server and license reduction. The new problem data lakes conceptually tackle pertains to Big Data initiatives. Big Data projects require a large amount of varied information. The information is so varied that it's not clear what it is when it is received, and constraining it in something as structured as a data warehouse or relational database management system (RDBMS) constrains future analysis.

But Gartner analyst Andrew White cautioned that the current approach that many firms took to creating data lakes could be problematic in the long term.<sup>14</sup>

Addressing both of these issues with a data lake certainly benefits IT in the short term in that IT no longer has to spend time understanding how information is used – data is simply dumped into the data lake. However, getting value out of the data remains the responsibility of the business end user. Of course, technology could be applied or added to the lake to do this, but without at least some semblance of information governance, the lake will end up being a collection of disconnected data pools or information silos all in one place.

Gartner analysts went on to caution that Hadoop and other technologies being used to create data lakes before the launch of Podium Data also carried substantial information quality, security, and system performance risks.<sup>15</sup>

[To date, many businesses design] a data lake to accept any data, without oversight or governance. Without descriptive metadata and a mechanism to maintain it, the data lake risks turning into a data swamp. And without metadata, every subsequent use of data means analysts start from scratch. [Note: metadata provides an ongoing definition of individual data items that describe the source of data, when it was created, how it is used over time, and what insights were gained from different uses. These definitions are stored with the data and can be updated each time a data item is accessed and used.]

Another risk is security and access control. Data can be placed into the data lake with no oversight of the contents. Many data lakes are being used for data whose privacy and regulatory requirements are likely to represent risk exposure. The [technical] security capabilities of central data lake technologies are still embryonic. [Note: Security is also a governance issue that links to a company's overall risk management systems—including more recent high-profile problems that are causing many companies to appoint Chief Security Officers that report directly to the CEO and board.]

Finally, performance aspects should not be overlooked. Current tools and data interfaces simply cannot perform at the same level against a general-purpose [data lake] as they can against optimized and purpose-built infrastructures. For these reasons, Gartner recommends that organizations focus on semantic consistency and performance in upstream applications and data streams [prior to creating a data lake].

The Podium Data team believed that they had differentiated their technology and their Client Engagement Model from the current technologies and approaches. (See **Exhibits 5a–5f** for a summary of the business intelligence and data analytics ecosystem and Podium Data's role.) First, its Podium data management system provided a visual set of tools that allowed business users to identify and

manage data as it was accessed from upstream data streams, and this metadata enabled data scientists and IT professionals to ensure semantic consistency and governance of data as it populated centralized data lakes. The Podium data management tools also ensured that business users could directly interact with the data to solve pressing business problems and create business value. The Podium Data team believed that their approach provided the foundation for creating the governance systems that would need to be developed to enable business users to take a more active role in the governance of how data was used to create business value while also providing the strong security features to address the security risks identified by Gartner. (The Podium Data security features are discussed in more detail in the **Appendix**.)

Podium Data customers reported that the Podium Data approach enabled business users to assume responsibility for defining how data would be used to drive business value. Customers were impressed with how easy the Podium Data tools were to use and how fast business users without technical training could access the data and create analytical models that could be used to make decisions and drive business value. (See again **Exhibit 4**.) For many customers, Podium Data provided a bridge to the enterprise's traditional IT systems and infrastructure, including software applications, database management systems, and the licensing and cloud hosting of Hadoop. These data and applications were often managed by different businesses and outsourced partners located in data centers around the world.

Once a customer completed its pilot and entered the operation phase of running and maintaining the initial system, while also expanding into new applications, Podium Data encouraged the customer to sign up to access its system from a cloud services platform that Podium Data would host with a cloud computing partner. (**Exhibit A-1** in the **Appendix** provides an overview of cloud computing service providers at the time of the case and Gartner's evaluation of their offerings.) From a technical perspective, operational big data applications built for clients during a pilot could either be hosted by the client or be offered by Podium Data as a cloud-based service. At the time of the case, Podium Data's founding team hoped that its pilots would transition to the hosted, subscription fee services model.

## Decision Time

In early 2015, the Podium Data founding team was pleased with all that they had accomplished in the year since Podium Data launched. But they knew that many challenges lay ahead, and they would need to make some key decisions.

### *Podium Data Revenue Model*

As mentioned earlier, in March 2015, Podium Data had a tiered revenue model that required customers to pay \$75,000 for an eight-week pilot. Barth and Richards believed that asking customers to pay up front ensured that they "had skin in the game."

A lot of VCs we talked with wanted us to give our pilot away for free to achieve "zero friction sales," but we believe that's the wrong approach for a new software firm that is selling to large enterprise customers. When you are dealing with big companies, it's important to have them write a check up front. This ensures that, psychologically, they believe there is value in the product you will deliver. If they don't pay you for it, to them the software you are delivering is still in the lab.

Once the pilot was completed and the customer began the transition to the operations phase, Podium Data planned to convert these customers from a consulting fee to a subscription fee revenue



model. In January 2015, Podium Data had seven paying customers—two of which were transitioning to the operations phase. By 2016, it hoped to have 10 paying customers with 4 pilots and 6 in the operations phase, which would yield a conversion rate of 75%.

Barth and Richards planned to eventually decrease the pilot period from eight weeks to two weeks. Ultimately, they believed that they would be able to do away with the consulting fee and bundle the development of the pilot within a software subscription fee right from the start by allowing the customer to cancel the contract within 90 days.

While the subscription fee-based business model was becoming more widely accepted in the industry, it contrasted with most traditional packaged software revenue models in important ways. Under U.S. copyright law, software was intellectual property that was protected unless it was in the public domain.<sup>16</sup> Traditionally, software vendors, such as Microsoft and Oracle, sold copyright software using a perpetual software license, which granted the use of one or more copies of software under an end-user license agreement (EULA), while ownership of the copies of the software remained with the software vendor. The typical EULA included terms that defined the uses of the software, such as the number of installations allowed or the terms of distribution. Enterprise software EULAs also specified volume discounts based on the number of users within the enterprise and maintenance agreements, which specified the length of maintenance services provided and the fees paid.<sup>c</sup>

At the time of the case, many large software vendors also had consulting teams that could work with enterprise clients to install and maintain their applications. Some, like IBM Global Services, had dedicated business units that could work with clients—even if they were not buying IBM technology products—to develop customized applications that could be used to solve business problems and create business value. In addition, independent software consulting companies, like Accenture, CapGemini, CSC, Tata Consultancy Services, Infosys, and Wipro Technologies were available to help companies develop custom applications. Many of these vendor-based or independent consulting firms also provided outsourcing services—allowing enterprise clients to run and maintain their IT systems, including data warehouses, in data centers owned by the vendor. Many outsourced service providers were also willing to run and maintain IT systems within data centers owned by the enterprise customers. Consulting fees were often charged to develop custom applications and, if appropriate, ongoing outsourcing/maintenance fees could be charged when a new application became operational.

More recently, the shift to Software as a Service (SaaS) and cloud computing had emerged as “new” approaches that enabled enterprises to receive “on-demand delivery of IT resources and applications via the Internet with subscription based pay-as-you-go pricing.” Interestingly, many experienced IT professionals who lived through the early days of computing in the 1960s and 1970s believed that these time-sharing approaches were actually returning IT service delivery to its roots. Historically, the time-sharing model developed because computing technology of the day was complex and costly, and access to the talent needed to design, develop, and manage early computing systems and solutions was scarce. Many believed that the shift to SaaS and cloud computing also provided an opportunity for service providers to hire the rare, specialized expertise needed to deliver secure, reliable on-demand services while dramatically reducing the cost of building and running the highly specialized and complex technology platforms. This was especially true, analysts said, of emerging big data technologies and applications designed to drive business value.

---

<sup>c</sup> For a sample software licensing agreement, see <http://www.lawdepot.com/contracts/software-license-agreement/?loc=USCA>.

*Horizontal vs. Vertical Industry Penetration: Should Podium Data Stay Broad or Go Deep?*

Podium Data's early beta clients represented a broad range of industries. (Refer to **Exhibit 4** for a summary of these beta customers and the value delivered or expected to be delivered to them by Podium Data's solutions.)

As they implemented solutions for their beta customers in different industries, Podium Data's founding team recognized the advantages to building additional solutions within the same vertical industry group. "To date we have depended on our beta customers and word of mouth to launch seven different pilots—two of which are now transitioning to the operational phase, with two others right behind them," Barth and Richards explained. "Now that we have raised \$3 million in seed financing, we need to decide whether to hire a salesperson with experience in selling to one vertical, and then develop a small consulting team that can work with customers in that vertical to ensure that we can deliver new solutions to new customers in two weeks or less."

Alternatively, Barth and Richards wondered whether they should continue to sell by word of mouth and referrals across diverse industry groups. They then could develop a more general data analytics consulting team that could work with specialists in the client organization to build solutions. Would the decision to remain broad enable Podium Data to establish a leading position as a big data/analytics solution provider? Would the decision to go deep within a specific vertical enable it to more quickly become a leading niche analytics player within a vertical industry? In the end, the founding team recognized that the decision about whether to stay broad or go deep could influence future decisions about the amount of money that would need to be raised to finance growth, the type of exit that investors would seek, and the valuation that would be received.

No matter which approach they took, the Podium Data founding team knew that they needed to continue scaling the organization by building out a sales team and a research and product development team. They also knew that, at least in the next year, they would need to keep their two-phase Customer Engagement Model, which would mean channeling consulting revenue toward the business during the pilot phase, while quickly transitioning each pilot to what they hoped would be a subscription fee-based SaaS/cloud computing model. As a result, in the short term, they knew they needed to be discriminating about the new clients they brought on and the pilot projects they accepted. In an attempt to stay true to the mission of providing a superior technology product at a fraction of the price, Barth and Richards explained their focus as: "A consultant says yes to everything. A product guy says this is my product roadmap."

Even as they turned their attention to what needed to be accomplished in the short term, the founding team contemplated their next round of financing. Decisions made today would influence the amount of money that would need to be raised later and would influence future decisions on whether the company would: (1) remain independent and file for an IPO; (2) hire an investment bank to arrange for an acquisition; or (3) seek private equity financing to buy out the venture capital (VC) investors and enable Podium Data to remain private while consolidating the industry.

As they pondered these longer-term issues, the founding team found themselves returning to the business models they were considering and how today's decisions about product-market positioning and pricing, and the proprietary assets that they built in terms of intellectual property, talent, and capabilities, would influence future Podium Data valuations. (**Exhibit 6** provides a summary of key companies and recent valuations of key players in Podium Data's ecosystem.) Barth and Richards explained that today's capital markets were strongly influenced by the pricing model chosen. "In the enterprise software space today," they said, "there are three distinct business models that are emerging."

The first business model is that of a professional services firm. Using this model, customers pay consultants to provide installation, custom software solutions, and maintenance of software that the client owns. Companies adopting this model are typically worth 1X revenue at the time of a sale since the proprietary assets are usually based on the expertise of the consultants who can walk out the door as soon as the non-compete ends. A sale—probably to an existing software company—would probably be the preferred exit because professional services are hard to scale.

A second business model for software firms in our industry resembles a traditional software vendor. Using this model, customers pay a software-licensing fee—often for a perpetual license—and then pay an annual maintenance fee typically equal to about 20% of the license purchase price. Companies adopting this model are typically worth 5-10X revenue at the time of a sale or IPO based on the value of the intellectual property involved in the software. A drag on the valuation of traditional software firms is that the licensing fee revenue model enables the business to deliver plenty of cash when the company is growing, so there is pressure to continue to sell the same number of accounts each month in order to maintain revenue. Over time, the relationship with customers can erode future value.

The third business model for a software firm, which we are moving toward, resembles emerging SaaS and cloud computing service provider models. Using these models, customers pay a subscription fee to access software and computing services on an “on-demand” basis. Today, companies adopting this model are yielding the highest valuation at the time of a sale, IPO, or private equity financing—often 10-20x revenue. By 2015, subscription models had become very popular because analysts believed that they matched the value of the software with its usage rate, while enabling the service provider to centralize their infrastructure and expertise. This strategy thus reduced costs for both the service provider and for customers. If customers value the service, the service provider can better match new investments to stable returns, making business outcomes easier to predict for both the service provider and customers. Such models may bundle free updates and maintenance into the subscription services, or they may adopt a hybrid approach and use different revenue models for different services.

But even as they considered how the capital markets currently valued these different business models, the Podium Data founding team knew that the revenue growth of the SaaS model was significantly slower than selling perpetual licenses. Staying true to the subscription model committed Podium Data to raising money for several years, with the attendant risks of VC funding cycles. In addition, customers had also expressed concerns about whether Barth and Richards would continue to lead Podium Data over the long term, given that the decision to accept VC financing meant that the VC investors would require that Podium Data’s founders agree to either sell the company or conduct an IPO within the next five years or so. Barth recognized these issues and he and the Podium Data team had worked to address them. “Subscription pricing is attractive to many customers because they can pay for what they use, and are not committed to a big up-front cost if the product is not broadly adopted,” he explained.

However, we knew that larger companies needed to ensure that, if they terminated the subscription or if there was a change in control in the company, they could still access the data managed by Podium. We addressed this concern in the design of our product by publishing all the objects managed by Podium in the Hadoop-standard HCATALOG repository and storing all data in a standard Hadoop format with human-readable naming conventions. If Podium is turned off for any reason, they'll lose its functionality and our easy-to-use interface, but they will have full access to the data they have loaded and prepared with Podium. This was a critical decision that allowed an early-stage company to sell enterprise software to large customer.

As the Podium Data team turned back to the work at hand, they recognized that the relationship and Customer Engagement Model they were developing would be disruptive to many in the established IT industry and knew that they would need to help customers and potential customers understand the benefits and explain them to others in their organization.

**Exhibit 1 Podium Data Founder Bios****Paul Barth, PhD, CO-FOUNDER & CEO**

Paul is a recognized thought-leader and practitioner in leveraging information as a strategic asset and in best practices in data management. Prior to founding Podium Data, Dr. Barth was a co-founder with NewVantage Partners a consultancy that provides expertise and guidance to Fortune 1000 businesses. Paul began his career with Schlumberger. After completing his PhD, he joined Thinking Machines, a pioneer in high-performance computing, serving as lead architect responsible for developing the Quantum system for American Express, one of the earliest commercial marketing databases. Paul later joined the American Express database marketing subsidiary, Epsilon, as technology head. In 1995, Paul co-founded and was CTO of Tessera, a pioneer in data warehousing. Following Tessera's 2000 acquisition by the Internet integration firm, iXL, Paul continued as CTO until 2001. Paul holds a PhD in computer science from the MIT, and an MS from Yale.

**Steve Richards, CO-FOUNDER & COO**

Steve is an experienced operating and financial executive having worked the last 25 years in both small and large software companies. Prior to starting Podium Data, Mr. Richards served as the CFO of Peoplefluent where he was instrumental in driving growth and profitability of this \$115M leading SaaS provider in the HCM software sector. Steve began his software career at Sybase as Senior Finance Manager. During his tenure company revenues grew from \$150 million to nearly \$1.0 billion. Steve then went on to serve with four consecutive Greylock Partners funded companies, most recently as CFO with Awareness, Inc., an innovator in social media marketing software. In addition, he served as CFO at ClearForest Ltd, a leader in text analytics software, which was sold to Reuters in 2007. Prior to ClearForest, Mr. Richards served as COO & CFO of netNumina Solutions and CFO of Tessera Enterprise Systems, both culminating in successful exits. Steve holds a BA in economics and history as well as an MBA in finance from the University of Massachusetts Amherst.

**Bob Vecchione, CO-FOUNDER & TECHNOLOGIST**

Bob has been working on the design and delivery of complex large scale data solutions for over 20 years. Since working at parallel computing pioneer, Thinking Machines, Bob has been leveraging leading edge technologies to solve challenging business problems for the world's leading Financial, Insurance, Pharmaceutical and Retail firms. After studying Electrical Engineering at the University of Massachusetts Lowell, Bob began his career at Prime Computer working in the Integrated Circuit Technology group. Subsequent to Prime, Bob joined Thinking Machines in Cambridge where he held several roles from Systems Integration, Development to Field Services. After Thinking Machines, Bob partnered with Stephen Brobst and built a leading Data Warehouse consulting firm called Strategic Technologies & Systems which was later purchased by Teradata. Bob then joined an early start up named Knowledge Stream Partners that focused on delivering business intelligence to the fingertips of decision makers. Knowledge Stream Partners was acquired by Exchange Applications in 2000. In the 10 years prior to starting Podium Data, Bob, as an independent, has built strong partnerships with among others Teradata, Ab Initio, Oracle and delivered industry leading solutions to a many large enterprises including Travelers, American Express, JP Morgan Chase, Staples, and Bank of America. Bob is recognized as an industry leader in the design, architecture and implementation of large scale data systems.

**Michael Howard, CO-FOUNDER & TECHNOLOGIST**

Michael is an expert in database technology with a broad computer systems background. Prior to starting Podium Data, Mr. Howard was focused on data warehousing applying Big Data technologies

to solve traditional business data processing problems for several major US financial institutions. Early in his career Michael co-authored the Microsoft Applesoft Compiler for the Apple computer.

Michael was then the first non-founder employee of Datest where he architected a specialized time-series database for CD-ROM for the financial services industry. Datest was acquired by Lotus Development Corporation. Continuing his work with CD-ROM databases, Michael joined Ziff Communications as VP Technology for the Computer Library Division. He then founded InterActive WorkPlace, a sales force information system that was acquired by Siebel Systems. Michael later worked with Scalent Systems as Senior Member, Technical Staff. Scalent was later acquired by Dell to become the cornerstone of their data center automation suite. Michael worked as a research intern at Xerox PARC while obtaining BS and MS degrees in Computer Science from MIT.

### Atif Majid, CO-FOUNDER & TECHNOLOGIST

Atif has a wealth of software application design and development experience with expertise in multiple software languages, tools and methodologies. Prior to starting Podium Data, Mr. Majid served as Director of Engineering at Epsilon where he was responsible for the development of terabyte scale data warehouse. Early in his career Atif was the Technical Lead for the Architecture Team at Xchange, Inc where he designed the company's next generation product which included a real-time, rules based recommendation and offer delivery system. He then went on to Astrum Software where he was instrumental in the development of a storage resource management application. Astrum was subsequently acquired by EMC. Atif then served as Vice President of Development at BEZ Systems, Inc. building a team to deliver a predictive performance management application. BEZ was later acquired by Compuware. Atif holds a BS in Electrical Engineering from the University of Engineering and Technology, Lahore and an MS in computer science from the University of North Carolina, Charlotte.

Source: Company documents.

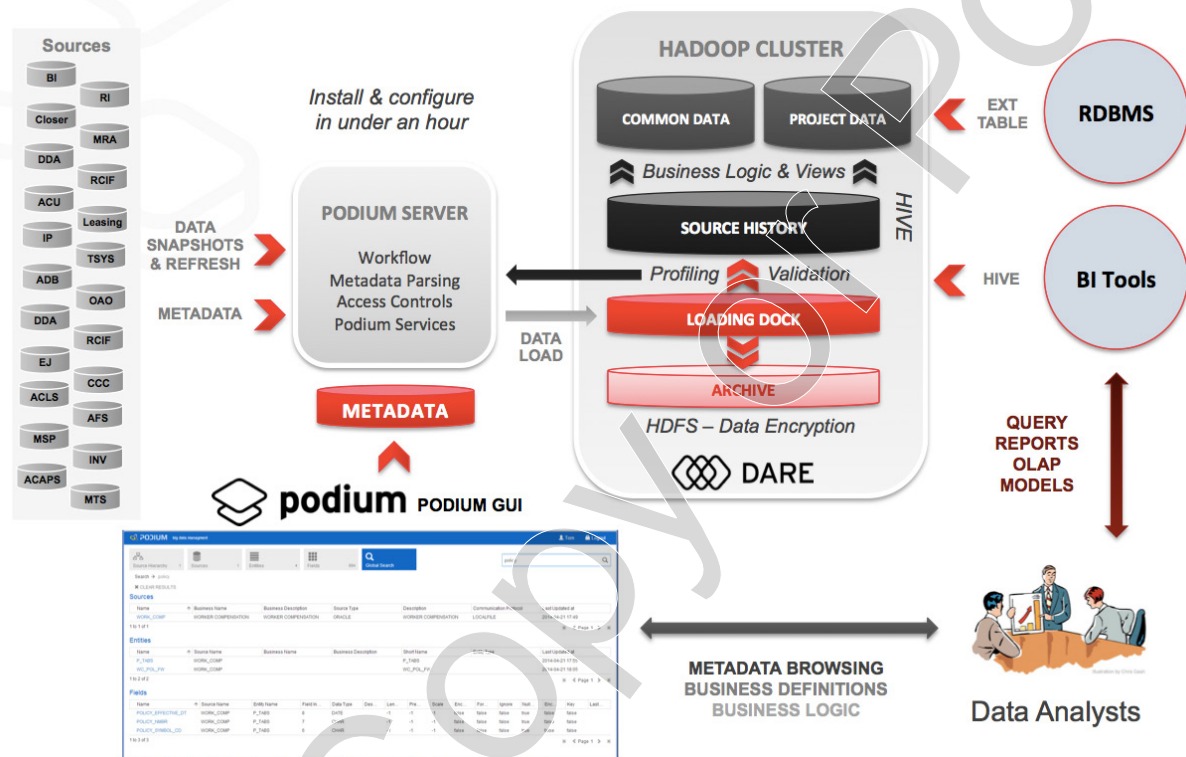
### Exhibit 2 Podium Data Financing Rounds

| Bootstrap Financing by Founders (March 2014 to March 2015)                |           |                      |                  |              |                |                    |
|---|-----------|----------------------|------------------|--------------|----------------|--------------------|
| Pre-Money Valuation   | Financing | Post-Money Valuation | Equity Ownership |              |                |                    |
|   |           |                      | Founding Team    | Options Pool | Seed Investors | Series A Investors |
|   | \$1M      |                      | 100%             |              |                |                    |
| Seed Round Financing (CommonAngels Ventures and 3 high net worth Friends) |           |                      |                  |              |                |                    |
| \$15M   | \$3M      | \$18M                | 79%              | 5%           | 16%            |                    |
| Series A Financing (Estimated June 2016)                                  |           |                      |                  |              |                |                    |
| \$75M   | \$25M     | \$100M               | 56%              | 10%          | 11%            | 23%                |

Source: Company documents.

Note: Data in this exhibit are disguised and are for teaching purposes only.

Exhibit 3 Podium Data Architecture



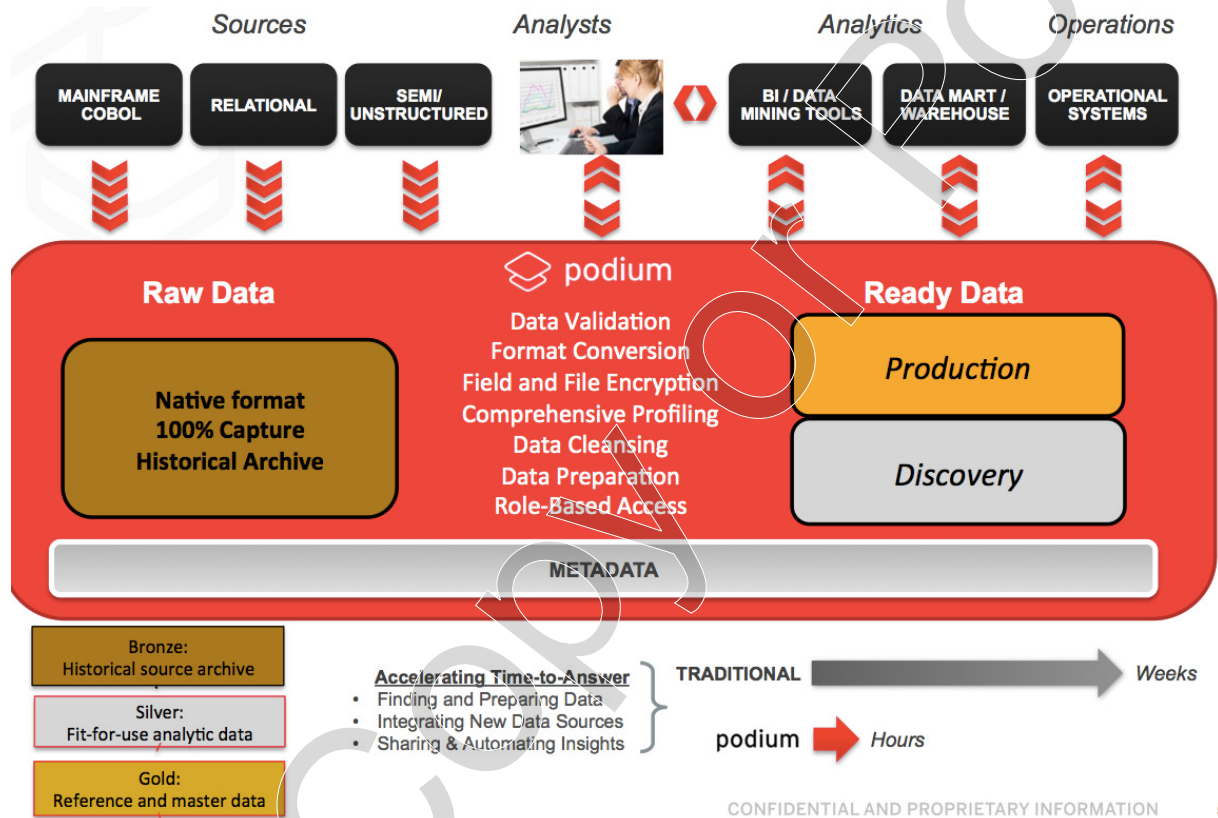
Source: Company documents.

As described on the Apache Hadoop® website:<sup>17</sup>

The Apache Hadoop® project develops open-source software for reliable, scalable, distributed computing. The Apache Hadoop® software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures.



Exhibit 3 (continued)



Source: Company documents.



**Exhibit 4a** Podium Data Customers in Early 2015

| Customer                           | Industry           | Source                   | Date Entered Pilot Phase | Date Production License Purchased | Key Benefits  |
|------------------------------------|--------------------|--------------------------|--------------------------|-----------------------------------|---|
| Head of Commercial Insurance       | Insurance          | NVP                      | March 2014               | NA                                | Data mining to predict root causes for attrition.   |
| Head of Marketing Analytics        | Pharma             | NVP                      | June 2014                | November 2014                     | Accelerate analytics with self-service, hosted data platform. Ability to manage terabytes of clinical data. |
| VP Finance Leading Data Governance | Media              | NVP                      | September 2014           | December 2014                     | Analysis of subscriber behavior across paper and digital delivery. Ability to manage 5TB daily data.        |
| CIO                                | Insurance          | NVP                      | June 2014                | May 2015 (estimate)               | Ability to rapidly integrate policy data to optimize service calls.   |
| CTO                                | Banking            | Large Technology Partner | January 2015             | May 2015 (estimate)               | Platform to provide Data-as-a-Service for new applications.   |
| VP IT                              | Financial Services | Consulting Partner       | January 2015             | NA                                | Rapidly migrate legacy mainframe data to Hadoop for analytics.  |
| VP IT                              | Banking            | Podium                   | February 2015            | September 2015 (estimate)         | Accelerate data sourcing, cleansing, and provisioning for business applications.                            |

Source: Company documents.

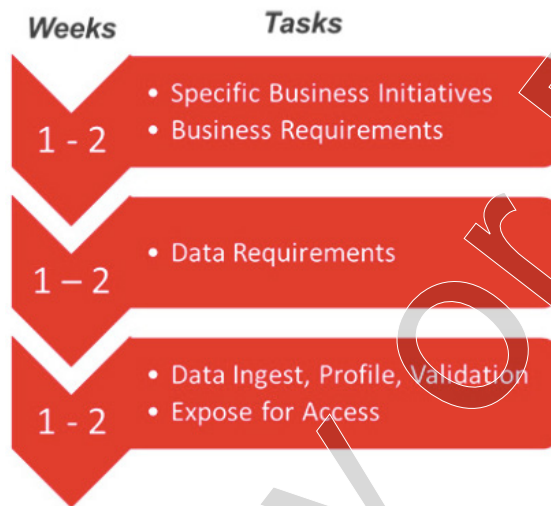
**Exhibit 4b** Podium Data Customer Quotes

Effectively managing data affects the bottom line – 99% of financial services executives surveyed believe they're losing revenue as a result of not effectively managing and leveraging business information.

Source: Wall Street Journal global survey.



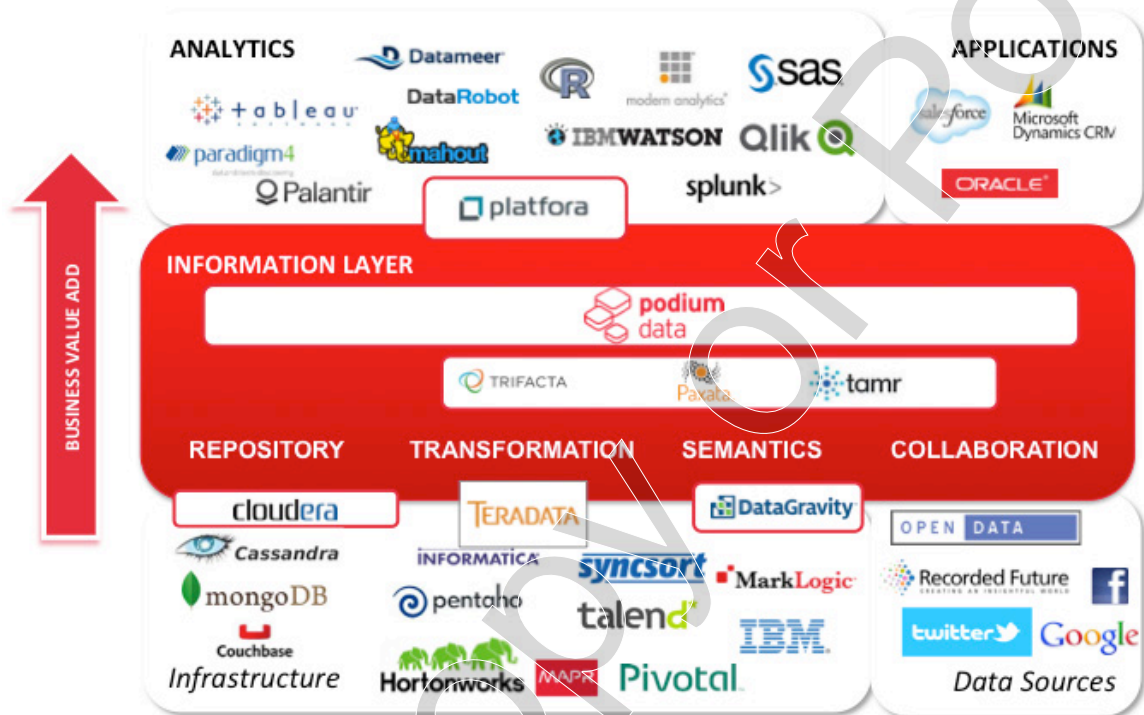
Source: Podium Data website, <http://www.podiumdata.com/engagement-2/>, accessed June 16, 2015.

**Exhibit 4c** Podium Data Customer Engagement Model**Exhibit 4d** Comparing Big Data and Relational Data Warehouse Approaches

|                 | BIG DATA         | RELATIONAL          |
|-----------------|------------------|---------------------|
| PLATFORM COST   | <\$1MM           | >\$5MM              |
| DATA LOADING    | 2 DAYS           | 1 MONTH             |
| ANALYTICS       | 25 LINES OF CODE | 25,000 LINES OF SQL |
| PROCESSING TIME | 40 HOURS         | 2 MONTHS            |

Source: Podium Data website, <http://www.podiumdata.com/engagement-2/>, June 16, 2015.

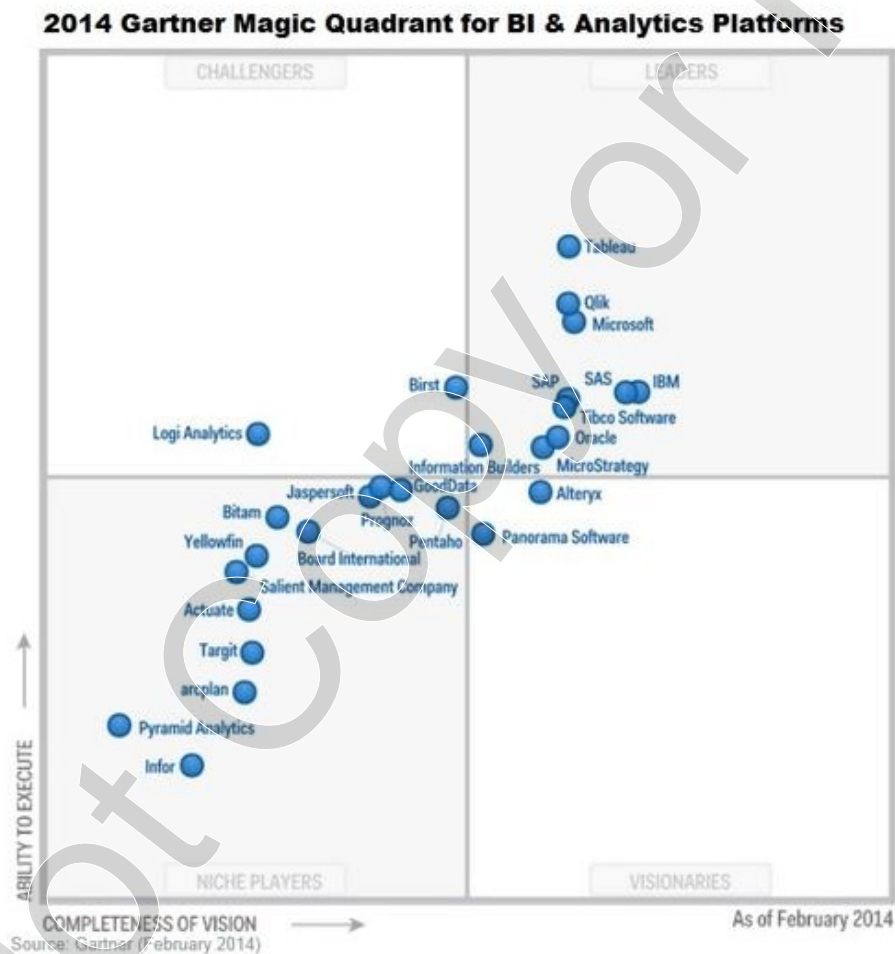
Exhibit 5a Podium Data Ecosystem



Source: Company documents.

**Exhibit 5b** 2014 Gartner Magic Quadrant for BI and Analytics Platforms

**Note:** As shown in **Exhibit 5a**, Podium Data's software creates a new layer in the BI and Analytics Platform. Players listed on the Gartner Magic Quadrant and in **Exhibit 5c** are among the players in the big data ecosystem that Podium Data is entering.



Source: Gartner, "Magic Quadrant for Business Intelligence and Analytics," February 23, 2015, <http://www.gartner.com/technology/reprints.do?id=1-2ADAAYM&ct=150223&st=sb>, accessed April 13, 2015.

**Exhibit 5c** Worldwide Business Analytics Software Revenue and Share %

|                      | Revenue (\$M) |          |          | Share (%) |        |        | Growth (%) |           |
|----------------------|---------------|----------|----------|-----------|--------|--------|------------|-----------|
|                      | 2011          | 2012     | 2013     | 2011      | 2012   | 2013   | 2011-2012  | 2012-2013 |
| Oracle               | 6136.40       | 6477.00  | 6748.70  | 19.10     | 18.60  | 17.90  | 5.50       | 4.20      |
| SAP                  | 4777.50       | 5208.80  | 5627.60  | 14.90     | 14.90  | 14.90  | 9.00       | 8.00      |
| IBM                  | 4449.40       | 4597.60  | 4652.90  | 13.80     | 13.20  | 12.30  | 3.30       | 1.20      |
| Microsoft            | 2347.10       | 2734.90  | 3121.70  | 7.30      | 7.80   | 8.30   | 16.50      | 14.10     |
| SAS                  | 2264.70       | 2422.20  | 2575.50  | 7.00      | 6.90   | 6.80   | 7.00       | 6.30      |
| Teradata             | 928.90        | 1050.70  | 1054.10  | 2.90      | 3.00   | 2.80   | 13.10      | 0.30      |
| Salesforce.com       | 227.40        | 330.90   | 566.00   | 0.70      | 0.90   | 1.50   | 45.50      | 71.10     |
| Informatica          | 440.60        | 455.40   | 525.20   | 1.40      | 1.30   | 1.40   | 3.40       | 15.30     |
| Adobe                | 320.50        | 385.40   | 499.00   | 1.00      | 1.10   | 1.30   | 20.30      | 29.50     |
| Infor                | 443.80        | 461.10   | 487.60   | 1.40      | 1.30   | 1.30   | 3.90       | 5.80      |
| MicroStrategy        | 401.30        | 412.50   | 437.60   | 1.20      | 1.20   | 1.20   | 2.80       | 6.10      |
| QlikTech             | 293.50        | 359.40   | 431.30   | 0.90      | 1.00   | 1.10   | 22.50      | 20.00     |
| FICO                 | 311.60        | 332.90   | 325.40   | 1.00      | 1.00   | 0.90   | 6.80       | -2.30     |
| ESRI                 | 224.70        | 229.80   | 276.20   | 0.70      | 0.70   | 0.70   | 2.30       | 20.20     |
| Tableau Software     | 61.40         | 122.60   | 225.20   | 0.20      | 0.40   | 0.60   | 99.60      | 83.60     |
| Information Builders | 200.30        | 208.60   | 210.10   | 0.60      | 0.60   | 0.60   | 4.10       | 0.70      |
| TIBCO                | 125.60        | 183.40   | 180.10   | 0.40      | 0.50   | 0.50   | 46.10      | -1.80     |
| JDA Software         | 143.70        | 157.60   | 177.30   | 0.40      | 0.50   | 0.50   | 9.70       | 12.50     |
| Autodesk             | 119.80        | 136.90   | 139.50   | 0.40      | 0.40   | 0.40   | 14.20      | 2.00      |
| Epicor Software      | 124.40        | 121.00   | 137.70   | 0.40      | 0.30   | 0.40   | -2.70      | 13.80     |
| Other                | 7800.10       | 8503.80  | 9346.40  | 24.30     | 24.40  | 24.80  | 9.00       | 9.90      |
| Total                | 32142.90      | 34892.50 | 37745.30 | 100.00    | 100.00 | 100.00 | 8.60       | 8.20      |

Source: IDC 2014.

Note: This table does not take into account any mergers and acquisitions that closed after December 31, 2013.

**Exhibit 5d** Worldwide Revenues for Big Data Technology and Services 2011-2018 (\$B)

|   | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2013-2018 CAGR (%) |
|---|------|------|------|------|------|------|------|------|--------------------|
| <b>Infrastructure</b>                                   |      |      |      |      |      |      |      |      |                    |
| Servers   | 0.7  | 0.8  | 1.0  | 1.3  | 1.6  | 1.9  | 2.6  | 3.4  | 26.7               |
| Storage   | 0.7  | 1.5  | 2.5  | 3.5  | 4.6  | 6.1  | 7.7  | 8.9  | 28.8               |
| Networking  | 0.2  | 0.3  | 0.5  | 0.6  | 0.7  | 0.9  | 1.1  | 1.3  | 23.6               |
| HPDA infrastructure (part of HPC)                       | 0.7  | 0.8  | 0.9  | 1.1  | 1.4  | 1.8  | 2.3  | 2.7  | 23.5               |
| Other datacenter infrastructure (hardware and software) | 0.2  | 0.3  | 0.3  | 0.4  | 0.4  | 0.5  | 0.6  | 0.7  | 18.3               |
| Cloud infrastructure                                    | 0.3  | 0.5  | 0.8  | 1.2  | 1.7  | 2.4  | 3.6  | 4.7  | 43.6               |
| Subtotal  | 2.7  | 4.2  | 6.0  | 8.0  | 10.5 | 13.6 | 17.7 | 21.7 | 29.2               |
| <b>Software</b>   |      |      |      |      |      |      |      |      |                    |
| Information management software                         | 1.1  | 1.4  | 1.7  | 2.0  | 2.4  | 2.9  | 3.4  | 4.1  | 19.0               |
| Discovery and analytics software                        | 0.8  | 0.9  | 1.2  | 1.5  | 1.9  | 2.5  | 3.1  | 3.8  | 26.3               |
| Applications software                                   | 0.2  | 0.3  | 0.3  | 0.4  | 0.5  | 0.7  | 0.9  | 1.1  | 27.8               |
| Subtotal  | 2.1  | 2.6  | 3.2  | 4.0  | 4.9  | 6.0  | 7.4  | 9.0  | 22.9               |
| <b>Services</b>   | 2.7  | 3.1  | 3.7  | 4.6  | 5.7  | 7.1  | 8.8  | 10.8 | 23.8               |
| <b>Total</b>  | 7.4  | 9.9  | 13.0 | 16.5 | 21.1 | 26.7 | 34.0 | 41.5 | 26.2               |

Note: See Table 2 for top 3 assumptions and Table 3 for key forecast assumptions.  
Source: IDC, 2014

Source: IDC 2014.

**Exhibit 5e** IDC Big Data and Analytics Opportunity Matrix

| Workload                        | Enterprise Performance Management<br><i>Plan and Analyze</i>                     | Operational Intelligence<br><i>Sense and Respond</i> | Exploration and Discovery<br><i>Model and Predict</i> |
|---------------------------------|--|--|---|
| <b>Business Process</b>         |  |  |   |
| Money<br>(Accounts, cash)       | Financial reporting, consolidation, budgeting, profitability, risk management    |  |   |
| People<br>(Customer, employee)  | Psychologically predictive, optimizing offers and content, experience management |  |   |
| Things<br>(Products, materials) | Predictive maintenance, optimizing assets and operations, demand sensing         |  |   |

Source: IDC, Worldwide Business Analytics Software 2014-2018 Forecast, and 2013 Vendor Shares, #249926.

**Exhibit 5f** IDC Big Data and Analytics Technology Support

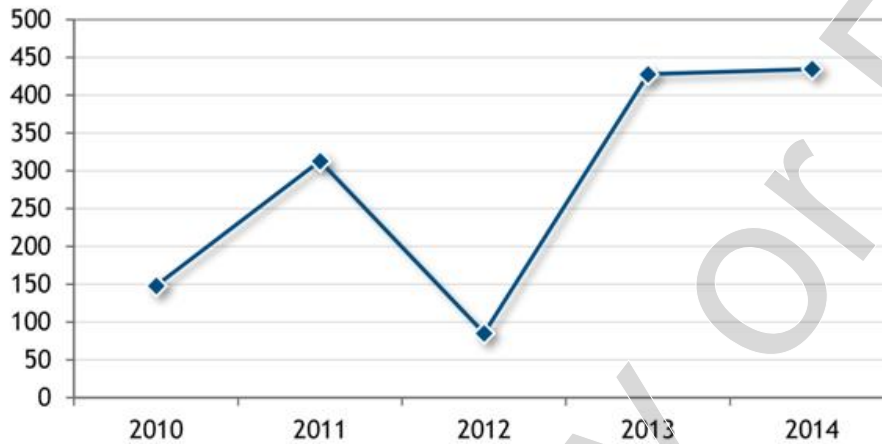
| Workload<br>Business Process    | Enterprise Performance Management<br><i>Plan and Analyze</i>            | Operational Intelligence<br><i>Sense and Respond</i>  | Exploration and Discovery<br><i>Model and Predict</i>           |
|---------------------------------|---|---|---|
| Money<br>(Accounts, cash)       | Relational data warehouse with in-memory technology<br>OLAP, dashboards | Event or stream processing<br>rules engines, workflow | Hadoop, NoSQL, graph, schema-less databases, advanced analytics |
| People<br>(Customer, employee)  |   |   |   |
| Things<br>(Products, materials) |   |   |   |

Source: IDC, Worldwide Business Analytics Software 2014–2018 Forecast, and 2013 Vendor Shares, #249926.



**Exhibit 6** Advanced and Predictive Analytics Venture Capital Financing

*Number of VC Financing Deals: 2014 reflects deals January – September 2014*



Source: IDC, Venture Capital Funding for Advanced and Predictive Analytics Software, #APA52W.

*Average Size of VC Financing Rounds (\$M): 2014 reflects deals January – September 2014*



Source: IDC, Venture Capital Funding for Advanced and Predictive Analytics Software, #APA52W.

## Appendix Podium Data Security Features

Podium Data offered software with a heightened security posture to ensure data could be used safely and to its fullest extent. It built in governance structures that included a system of checks, access controls, and encryption, as well as an effort to remove personally identifiable information (PII) from data before pouring it into the data lake. The company also placed its server behind a firewall, which blocked incoming connections and reduced the likelihood that a hacker could penetrate the network and ensured encryption at the field and file levels.

Encryption technologies used authentication and keys to ensure data could only be read by an intended recipient. The intended recipient would verify he or she was indeed the intended recipient through an authentication protocol, which often required the recipient to enter in a unique user name and at least one password. Then, the intended recipient would use his or her key to unlock the encrypted data; importantly, anyone without the right key would be unable to unlock the data unless a “back door” in the encryption technology existed.

Encryption could occur at three levels: (1) the hardware level, (2) the file level, and (3) the field level. While the first two are self-explanatory, field-level encryption refers to the process of encrypting a column of data within a file. While encryption at levels one and two would help prevent the insider threat from gaining unauthorized access to data, level-three encryption prevents hackers from remotely accessing the data.

A decision to encrypt the data where ever it lay – on a hosted cloud or on internal infrastructure – was a nontrivial, strategic data management decision. In Podium Data’s case, its software ensured that data would land in already encrypted fields. Furthermore, since big data is useful for its aggregated form, much of the statistical analysis on the data can be done while the data is still encrypted. Podium also filled the gap with its own encryption product, DARE, such that the data lay encrypted on the cloud and every field data it landed in after a query was designated a “DARE-ified” field.

In a November 3, 2014, press release, IDC forecasted that public IT cloud services spending would reach \$56.6 billion in 2014 and grow to more than \$127 billion in 2018. This represented a five-year compound annual growth rate (CAGR) of 22.8%, which was about six times the rate of growth for the overall IT market. IDC went on to estimate that, in 2018, public IT cloud services would account for more than half of worldwide software, server, and storage spending growth.

“Over the next four to five years, IDC expects the community of developers to triple and to create a ten-fold increase in the number of new cloud-based solutions,” said Frank Gens, Senior Vice President and Chief Analyst at IDC. “Many of these solutions will become more strategic than traditional IT has ever been. At the same time, there will be unprecedented competition and consolidation among the leading cloud providers. This combination of explosive innovation and intense competition will make the next several years a pivotal period for current and aspiring IT market leaders.”<sup>18</sup> (See **Exhibit A-1** for key cloud services providers.)

In an attempt to lead the shift in enterprise software from data centers to the cloud, in July 2014, Tim Cook, CEO of Apple, and Ginny Rometty, CEO of IBM, shocked the world when they announced a new Apple-IBM strategic partnership to combine Apple’s technology products – especially its mobile broadband iPhone and iPad – and its vast application developer network with IBM’s big data and analytic capabilities. Gregg Keizer, a *ComputerWorld* reporter, discussed the reaction of analysts:<sup>19</sup> “Today’s deal between Apple and IBM was ‘brilliant,’ one analyst said, while others called it a huge win for the former, which now gets into the enterprise *through the front door* rather than having to sneak in through the back.” “This is a very big deal for Apple, which just outsourced a high quality sales,

support and service organization in the likes of IBM,” said Frank Giliet, an analyst with Forrester Research. “What’s brilliant about this is that it lets Apple preserve its focus but partners on a large scale with someone who can totally represent them in the enterprise.”

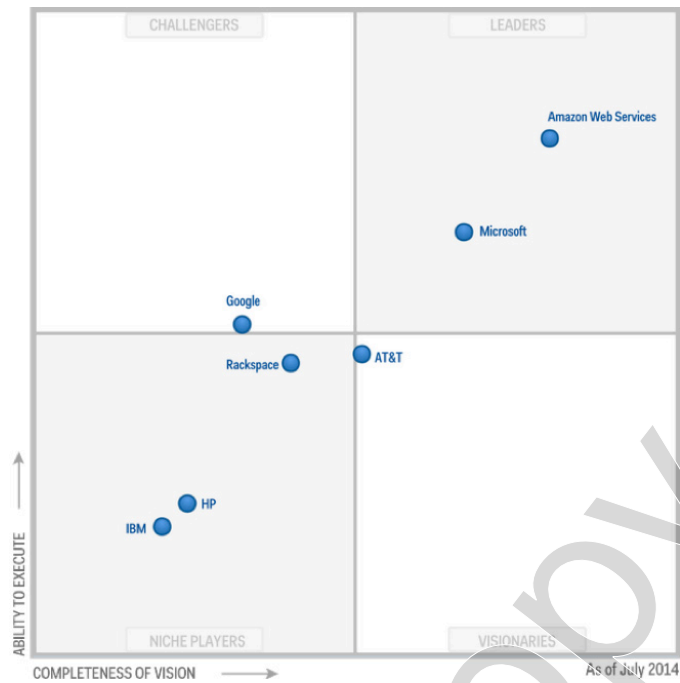
What’s in it for IBM? Keizer continued:<sup>20</sup>

IBM will sell the Apple devices, craft more than 100 industry-specific enterprise solutions that will include native apps, optimize its “cloud services” strategy for iOS [Apple’s mobile operating system], package device supply, activation, and management services, offer financing and leasing plans, and provide on-site support to customers. Apple will handle telephone support with new AppleCare options designed for enterprises.

In an interview after the announcement, Cook said: “It’s huge. It’s landmark. This is all about transforming the enterprise.” Rometty agreed and added, “This is all about unlocking mobility in the enterprise.”<sup>21</sup>

In December 2014, the new partnership announced its first release of 10 enterprise mobile business apps that catered to a wide range of industries, from financial services to law enforcement. Industry analysts and the press were enthusiastic. “Among the 10 apps that Apple and IBM released on Dec. 10 were an airline in-flight customer service and engagement app, another aimed at telecommunications firms for field repair calls and a third for law enforcement incident response. Others targeted business sectors ranging from banking and insurance to retail and government social services.”<sup>22</sup>

In April 2015, IBM announced that it was expanding the Apple partnership even further to launch its Health Cloud and Watson Cognitive Computing capabilities.<sup>23</sup> Health Cloud and Watson Cognitive Computing capabilities would support health data entered by customers in iOS apps using Apple’s ResearchKit and HealthKit frameworks. In addition, IBM would also develop a set of MobileFirst for iOS apps, which would enable enterprises to work with their employees to better manage their health and fitness needs. “Our deep understanding and history in the healthcare industry will help ensure that doctors and researchers can maximize the insights available through Apple’s HealthKit and ResearchKit data,” said John E. Kelly III, Senior Vice President, IBM research and solutions portfolio. “IBM’s secure data storage and analytics solutions will enable doctors and researchers to draw on real-time insights from consumer health and behavioral data at a scale never before possible.” Apple’s Jeff Williams, Apple Senior VP of Operations, agreed: “With Apple’s groundbreaking ResearchKit, researchers can easily create apps that take advantage of the power of mobile devices to give them rich data from a diverse global population. Now IBM’s secure cloud and analytics capabilities provide additional tools to help accelerate discoveries across a wide variety of health issues.”<sup>24</sup>

**Exhibit A-1** Cloud Service Providers in July 2014

Source: G. Ruth and A. Chandrasekaran, "Magic Quadrant for Public Cloud Storage Services," Gartner, 2014, <http://www.gartner.com/technology/reprints.do?id=1-1WWKTQ3&ct=140709&st=sb>, accessed April 14, 2015.

## Endnotes

<sup>1</sup> "Gartner Says 'Beware of the Data Lake Fallacy,'" Gartner Newsroom, July 28, 2014 <http://www.gartner.com/newsroom/id/2809117>, accessed April 9, 2015.

<sup>2</sup> Podium Data, homepage, [www.podiumdata.com](http://www.podiumdata.com), accessed April 9, 2015.

<sup>3</sup> MIT Alumni Database; and CSAIL documents.

<sup>4</sup> MIT CDOIQ, "CDOs Take the Lead in Big Data and Information Sharing," 2014, <http://www.mitcdoiq.org/2014-speakers/>

<sup>5</sup> Podium Data, "Management Team," 2015, <http://www.podiumdata.com/management-team/>; and NewVantage Partners, "People," 2015, <http://newvantage.com/people/paul-barth/>.

<sup>6</sup> NewVantage Partners, "About," n.d., <http://newvantage.com/about/>.

<sup>7</sup> NewVantage, Big Data Thought Leadership, <http://newvantage.com/thought-leadership/>.

<sup>8</sup> NewVantage, *Big Data 2014 Big Data Survey Report*.

<sup>9</sup> NewVantage, *Big Data 2014 Big Data Survey Report*.

<sup>10</sup> NewVantage Partners, "Big Data 2014 Survey Report," press release, <http://newvantage.com/wp-content/uploads/2014/11/NVP-Big-Data-Survey-2014-Press-Release-110314.pdf>, accessed April 10, 2015.

<sup>11</sup> NewVantage Partners, "Big Data 2014 Survey Report," press release.

<sup>12</sup> These definitions from <http://www.podiumdata.com/product/>, accessed April 9, 2015.

<sup>13</sup> "Gartner Says 'Beware of the Data Lake Fallacy,'" Gartner Newsroom.

<sup>14</sup> "Gartner Says 'Beware of the Data Lake Fallacy,'" Gartner Newsroom.

<sup>15</sup> "Gartner Says 'Beware of the Data Lake Fallacy,'" Gartner Newsroom.

<sup>16</sup> See Title 17 of the US Copyright Law (Section 117), <http://www.copyright.gov/title17/92chap1.html#117>, accessed April 9, 2015.

<sup>17</sup> This definition from <https://hadoop.apache.org>, accessed April 9, 2015.

<sup>18</sup> IDC, press release, November 3, 2014, <http://www.idc.com/getdoc.jsp?containerId=prUS25219014>, accessed April 14, 2015.

<sup>19</sup> G. Keizer, "Apple fashions 'brilliant' partnership with IBM for instant enterprise cred," CIO Australia, July 16, 2014, [https://www.cio.com.au/article/550129/reaction\\_apple\\_fashions\\_brilliant\\_partnership\\_ibm\\_instant\\_enterprise\\_cred/](https://www.cio.com.au/article/550129/reaction_apple_fashions_brilliant_partnership_ibm_instant_enterprise_cred/), accessed July 16, 2014.

<sup>20</sup> G. Keizer, "Apple fashions 'brilliant' partnership with IBM for instant enterprise cred."

<sup>21</sup> G. Keizer, "Apple fashions 'brilliant' partnership with IBM for instant enterprise cred."

<sup>22</sup> G. Keizer, "Apple-IBM Partnership off to an impressive start," *ComputerWorld*, December 15, 2014, <http://www.computerworld.com/article/2859480/apple-ibm-partnership-off-to-impressive-start.html>, accessed January 16, 2015.

<sup>23</sup> IBM, "IBM and Apple expand partnership to help transform medical research," press release, April 13, 2015, <http://www-03.ibm.com/press/us/en/pressrelease/46583.wss>, accessed April 14, 2015.

<sup>24</sup> IBM, "IBM and Apple expand partnership to help transform medical research."