WHITE PAPER

Big Pharma, Big Data - Big Deal? Yes, Really!

- Sunil S. Desai and Bill Peer

Abstract

Companies in the Pharmaceutical industry are being increasingly inundated with data that most are simply not capable of leveraging. As a consequence, most companies have vast amounts of un-leveraged and under-leveraged data. Big data analytics provides a way to harness this un-leveraged and under-leveraged data and gain timely insights for making better business decisions. Big data analytics is a big deal - from its potential benefits to the complexities inherent in setting up and sustaining the capability. This paper is intended as a primer for Pharma execs and discusses critical aspects that should be addressed when setting up the capability, a framework for identifying and shortlisting use cases in a sustainable manner, and some sample Pharma use cases.

With so much Big Data, where is the Big Knowledge? Where is the big actionable insight? Pharmaceutical companies of all sizes are increasingly being swamped with data, yet often seem to lack timely actionable insights drawn from the data. Why is this? Our efforts in the field have shown that the primary reason is not technology, nor is it a lack of desire. It is primarily an inability to effectively manage their evolving needs when adopting or deploying big data capabilities, not the least of which is big data analytics.

We believe that some of the major challenges facing the pharmaceutical industry currently -- declining R&D productivity, globalization of the supply chain, country-specific regulatory compliance requirements, rising commercial demands -- can be addressed very effectively by harnessing the power of big data analytics. New sources of information, new models of analysis, and radical reduction in the cost of computing are all creating new potentials for the pharmaceutical industry. However, the journey to roll out big data analytics programs is fraught with organizational, technological, and operational challenges and complexities.

It is the point of this paper to highlight some critical considerations that we have come to understand over the past few years of helping and operating in our clients' big data programs. We will begin by establishing some working definitions, then quickly dive into the five areas that are essential to address. With the necessities shared, we offer use cases that could serve as starting points for pharmaceutical companies and then we conclude the paper.

What is Big Data?

Big data refers to a collection of data that cannot be processed in a timely manner using traditional database applications. Big data is typically characterized by the three Vs: Volume, Variety, and Velocity.

1. **Volume**: Datasets whose sizes run into the petabytes, exabytes, and beyond
2. **Variety**: Data whose form could be structured or unstructured
3. **Velocity**: Data that comes into the processing environment very fast

It is not uncommon for people to add a fourth or even fifth V, veracity and value. Veracity speaks to the truthfulness of the data and represents the accuracy of the data. Value refers to the overall dollar value of big data itself and how valuable a given collection of data is along with its analytic processing.

Big Data Analytics – Simple Concept, Complex Reality

Setting up and institutionalizing a big data analytics capability is non-trivial and necessitates a holistic approach. Organizational dynamics, technology evolution, and the speed of business all play a part in the challenge. Each must be considered and addressed as part of the Big Data Analytics program strategy.

Pharmaceutical industry recognizes the potential benefits of big data analytics. However, our experience with most pharmaceutical companies indicates a gap in translating this recognition into a functional capability that is coordinated and institutionalized effectively across different business units. The challenge stems from insight-centric analytics being traditionally done in silos, be it business units or business functions, research units, or otherwise, where enclaves of knowledge workers and/or their data is partitioned from the whole. This is in direct contrast to the promise and power of big data analytics wherein enormous opportunity rests in large swaths of cross functional data being blended and operated upon. As pharmaceutical firms begin to recognize this and try to make adjustments, they face challenges beyond technical.

Our experience helping clients execute these programs over the past few years indicates there are five critical areas to be addressed when a pharmaceutical company first embarks on a big data analytics journey. These essential elements of a Big Pharma Big Data Analytics strategy are organizational structure, talent, technology, data management, and the operating focus of the big data program itself.

1. **Organizational Structure**: For the program to be successful, deep consideration must be given to the culture of the company, organizational structure, and operating norms. For example, we've found that funding models (e.g., central, business unit, or shared), value expected (e.g. time to market, revenue stream, or margin reduction), and finance orientation (e.g. profit center or cost center) all play a part in picking the right structure for bringing your new big data analytics capabilities to life.

   Structure: In practice, we've see four basic structure models be successful at large firms:

   - **Decentralized**: Business Units (BU) have distinct data sets and cross-organizational scale is not an issue; each BU can make its own big data decisions with limited coordination.
   - **Decentralized with a Lead BU**: Each BU makes its own decisions, but one BU plays the lead role in establishing standards, etc.
   - **Center of Excellence (CoE)**: An independent CoE oversees the big data analytics program; BUs pursue initiatives under the CoE's guidance and coordination.
   - **Centralized**: Corporate center takes direct responsibility for identifying, prioritizing, and implementing initiatives.

2. **Talent**: Big data analytics programs are resource intensive and a diverse set of complex skills are required to support and sustain such programs. For the purposes of this discussion, we will list only the different talent types that would provide these required skills: Program Manager, Infrastructure Manager, Big Data Architect, Data Steward, Domain Analyst, Financial Analyst, Data Scientist, Data Engineer, Data Analyst, and Data Visualization Analyst.

   As the above list of talent indicates, sustaining a big data analytics program requires a serious investment in skilled resources. Additionally, any decision regarding structure of the big data analytics program has implications on the number of each talent type required to sustain the program.

3. **Technology**: There are three broad big data functions that must be present, concurrent, and continuous in any pharmaceutical company: ingesting, processing, and presenting. Spanning across these functional groups, you will find security capabilities, such as data encryption and masking, you will find data and work flow management, and you will find data logging and monitoring.

   - Data Ingesting is the collection of capabilities focused on consuming data for a myriad of sources, volumes, and velocities and preprocessing it (e.g. quality measures, normalization) for usage by the enterprise. There are many open source and vendor offerings in this space to include Flume, Sqoop, Oracle Data Integrator, and SAP BODS. In the pharmaceutical industry, extra attention should be paid to the capabilities around source variety, as there are numerous untapped and emerging data sources where data blending will yield entirely new insights.
   - Data Processing is the collection of capabilities focused on analytic processing. Typically delivered as a platform, with a slew of pattern finding and processing options, the ability to process data as fast as it is ingested gives clear advantage, especially where an actionable insight being processed diminishes in value over time. There are many open source and vendor offerings in this space to include SAP Hana, IBM Netezza, AsterData, and Hadoop based platforms. There are also tools like R and SAS that can give specialists extra fire power in deep analytic pattern finding. Once these specialists

find the patterns, they can be codified and put into the general analytic data processing platform.

- Data Presenting is the collection of capabilities focused on bringing awareness of an actionable insight to an interested party. Most commonly experienced as reports and dashboards, the critical element is being able to help bridge the gap between the data, the actionable insight, and the human. There are many offerings in this space that include Tableau, Qlikview, Microsot BI Platform, Spotfire, and Birst to name a few.

As you might imagine, there are offerings in the marketplace that span one or more, or all, of these functions. Keeping these divisions in mind will allow a firm to better select the right product or product mix for their particular Big Data intentions.

4. **Data Management**: In the healthcare and pharmaceutical settings where there are stringent, and growing in complexity, requirements regarding data privacy and regulatory compliance, there is more to data management than just getting data on to the big data platform. Robust processes would be required to:

   - Create an company-wide data governance plan
   - Create data standards and business rules
   - Establish data access security requirements
   - Create and maintain consistent reference data and master data definitions
   - Administer data in compliance with business and regulatory obligations
   - Resolve data integrity issues across stakeholders

5. **Incubation**: The big data analytics strategy must accommodate the company's own evolving maturity as these powerful new capabilities come to life. There are two distinct phases of this maturation, the Find phase and Refine phase, and each requires a different focus.

   - **Find phase**: In the early stages of maturity (i.e., exploration and piloting phases of the program), the strategic focus should be on finding the right use cases and finding the right users. The key tactic is to empower operational flexibility and aggressively push awareness to the organization (i.e., showcase capability and create a buzz among end users). This phase is where failure is most likely to happen, and it typically happens because the right set of initial use cases are not identified correctly.
   - **Refine phase**: The onset of this phase comes naturally as some of the early use cases that have impact are materialized. In this phase, it is important to shift towards operational efficiency (i.e., scale up and roll-out phases of the program) because you will see a pull effect by users as their demand on your program's capabilities grow (i.e., capability is mature and widely adopted, users reach out with requests for implementation of additional value-added use cases).

Use Cases - Which Ones to Implement?

Big data analytics is a means to an end. The end is to extract value for customers, business partners, and shareholders from the large volumes of potentially un-leveraged and/or under-leveraged data held by pharmaceutical companies. To this end, big data analytics capabilities are typically rolled out by piloting and, subsequently, rolling out specific use cases. As previously noted, the success of a Big Data program depends critically upon the Find phase and finding the right use cases.

So, how do you identify use cases for consideration? Based on our experience, we recommend taking an open approach to identifying use cases. That is, invite stakeholders' companywide to submit questions that big data could help address that could materially impact their area. This sends a strong message that the analytics needs of all stakeholders are being addressed and, thereby, promotes an inclusive approach to building company-wide big data analytics capabilities.

Once the initial set of use cases has been identified and documented at a high level, it would undergo filtering and clustering to narrow the list down to a more manageable set of use cases. Each of these use cases would be further developed to understand basic facets such as: process impacted, process owner, KPIs impacted, data needs, and business benefits. Additionally, each use case would also be assessed and rated along two dimensions - i) business value, and ii) feasibility of implementation. These ratings would be used as inputs for a "use case prioritization framework" that would rank the use cases relative to each other to shortlist a handful of use cases for the next level of in-depth detailing. This additional level of due diligence would drive the final selection of use cases for implementation.

Depending upon the available bandwidth of resources as well as continuing analytics needs, the process described above would be repeated throughout the Find Phase to continue enriching the portfolio of use cases. Newly identified use cases would be added to others that did not make the implementation cut in the previous "review/selection" cycle and the "identification/filtering/assessment/prioritization/ in-depth detailing/selection" cycle would be repeated.

The relevance, appropriateness, and scoring of any use case is company dependent. The final determination of ideal Find phase use cases depends upon the business model of the company, its potential to address a business problem (e.g., revenue generation, cost reduction, operational improvement, customer satisfaction, patient outcome, etc.), its fit in support of the company's strategic goals, and finally the political and funding realities that must be accommodated. That said, in pharmaceutical industry value ecosystem, we've found the following areas particularly ripe for big data analytics opportunities.

For the R&D-intensive pharmaceutical brand drug manufacturers, drug discovery and development lead times and associated high costs have always

been a challenge. Big data analytics, with its ability to rapidly identify associations between large volumes of varied data, can be leveraged in R&D. It can simultaneously process clinical trial, molecular, and publicly available data to either discover or rule out associations between molecules and targeted diseases. In addition, predictive modeling can also be used to predict outcomes related to molecule efficacy, side effects, etc. Such insights would enable pharmaceutical companies to narrow their drug discovery efforts on molecules that exhibit promise in treating the targeted disease and discontinue efforts on the molecules that do not show promise. Thus, usage of big data analytics can accelerate the drug discovery and development process and, thereby, cut down costs and reduce the time-to-market for new drugs.

For the pharmaceutical wholesale distributors, it is crucial for them to understand their medium and small sized retail chain customers and align them with most suitable sales programs. This is especially true in the sales of generic drugs where these distributors make their highest margins. Big data analytics can leverage large volumes of available data to understand customers better and develop precise customer segmentation. Subsequently, big data analytics can leverage this customer segmentation data along with other internal and external data such as call center data, customer survey data, sales visit data, sales program data, syndicated data, and social media data to provide insights into individual customer behavior. These insights can then be used to evaluate the efficacy of existing sales programs and to recommend the most suitable program(s) for specific customers. These insights would also provide an opportunity to cross-sell/up-sell other sales programs that are more profitable.

For the increasingly global pharmaceutical supply chain, a majority of the suppliers are located in Asia. These could be suppliers of ingredients to branded drug manufacturers, suppliers of generic drugs to pharmaceutical distributors, etc. Hence, events such as operational issues, natural disasters, geopolitical events, etc., at these offshore suppliers can lead to supply disruptions and seriously impact the availability of pharmaceuticals in the U.S. down the road. Supply disruptions lead to lower fill rates on customer orders and lost sales. Big data analytics can be used to develop predictive models that would leverage vast amounts available data, both internal (e.g., purchase orders, supplier delivery performance, inventory, sales history, product substitution rules, etc.) and external (e.g., regional media, weather projections, regulatory announcements, etc.), to predict potential supply disruptions well in advance. This would trigger proactive intervention and execution of contingency plans by ordering additional product from alternate suppliers and/or ordering of substitute product.

Conclusion

Pharmaceutical companies are being increasingly inundated with data that most firms are simply not capable of leveraging. As a consequence, most companies have vast amounts of un-leveraged and under-leveraged data. Big data analytics provides a way to harness this un-leveraged and under-leveraged data and gain timely insights for making better business decisions. Big data analytics is a big deal for the pharmaceutical industry - from its potential benefits as well as the challenges and complexities inherent in setting up the capability. Increasingly, companies in the pharmaceutical industry are deploying big data analytics programs. This discussion outlines certain critical considerations that are necessary to ensure successful roll out and derive sustained value from big data analytics programs.

Sunil Desai is Senior Principal, Management Consulting Services, and Bill Peer is Principal Technology Architect, Center for Advanced Technologies, Infosys Limited.

For more information, contact askus@infosys.com