Lightbeam business case: pharma

Executive summary

With the increasing complexity of manufacturing processes in the pharmaceutical industry, pharma companies are finding that supply chains and process engineering are becoming more important to their bottom lines. Both of those departments of a pharma company depend on data analysis. Therefore, the advent of big data technologies coming from a proliferation of IoT sensors and modern machine learning techniques represents a massive strategic opportunity for today's pharma companies.

There are a lot of vendors who claim to help pharma companies capitalize on this opportunity, and many of these vendors have already been able to unlock significant results for their clients. However, there are still a few critical weaknesses in the approaches of existing players: rollout periods over a year long, underutilized software, and unintuitive user experiences.

Lightbeam is a cognitive computing product providing self-serve big data analytics. Lightbeam is unique for both our intuitive conversational interface — we allow users to ask the computer questions about their data in English — and the sophistication of our data tools — our product brings the cutting edge in data integration and machine learning research to everyone. Our technology provides an answer to the difficulties in data integration and user experience associated with existing big data approaches, but is also largely technically complementary to existing software.

In short, we believe that the time has come for big data analytics to be offered in plain English — and to everyone, not just a few experts in your firm.

To save you a bit of time...

Many vendors have been talking about the importance of self-serve analytics in the pharmaceutical industry. If you want, you can skip over the "trends" and "implications" sections if you'd just like to see how our approach is unique.

Trends in process engineering

Big data is important

In the 21st century large molecule drugs have experienced a massive surge in the pipelines of major pharmaceutical companies; the global biological medicines market is projected to reach $287 billion by 2020. Industry leaders typically invest up to $1 billion in total pre-approval R&D costs. With this influx of new molecules requiring more advanced, expensive production lines, supply chain and production line optimization have become critical to the bottom lines of every pharma company.

These kinds of process optimization hinge on the availability of predictive analytics technologies. Debugging a large molecule bioreactor simply isn't possible with traditional Excel-based process analytics — in fact, X study fact. It's clear that the need for big data analytics is here to stay.

Current state of data analytics

However, more advanced predictive analytics is not necessarily making it to the end user in a pharma company. Here, we summarize the results of a survey of dozens of process engineers in leading pharmaceutical companies:

• A large majority of engineers were still using Excel for their data analysis

• Many engineers had access to Alteryx, Tableau, and other business intelligence tools, but those who did commented that they did not find them very intuitive and

that they were forced to consult specialists when they wanted to get things done with those tools

- Some engineers had access to apps built in the R-based Shiny platform. These engineers commented that Shiny was great when an app was available, but that they frequently still had to wait for colleagues to build apps to serve their needs.

- Very few engineers had access to cognitive computing or machine learning. Those who did relied on colleagues who were specially trained in those technologies, and were forced to email them anytime they had a question.

A clear pattern emerges from this survey: no matter what technologies you use, you still need a small group of highly trained specialists to do big data and machine learning. This means that everyone has to wait for this group of specialists to process its work and that turnaround times on routine questions are anywhere from a day to months long. We call this the *specialist bottleneck.*

Implications

Lack of self-serve analytics is costly

We have outlined the specialist bottleneck — the problem of data analytics being bottlenecked on a small group of data specialists — whose solution is the core of our vision, but you might be asking, why is this such a big deal?

It's not just wasted personnel time, it has a massive impact on your bottom line. Every day that a production line is not up and running represents millions of dollars of lost potential revenue, and both process development for new manufacturing efforts and incident management for production line problems are directly reliant on data analytics. This means that the specialist bottleneck is a core contributor to the time it takes to fix process problems.

What would a solution be?

Criteria

So, having outlined the problem, let's take a look at the properties that a potential self-serve data solution must satisfy to be strategically relevant to your business, as well as how other vendors stack up.

Intuitive — Any good self-serve analytics solution should be usable by all process engineers and supply chain analysts within your organization. Otherwise, the specialist bottleneck makes the solution just as slow as having Python or R-based analysts doing manual analytics — negating the promise of "self-serve." Unfortunately, existing solutions struggle with the tradeoff between intuitive interfaces, adaptability and power.

Adaptable — The solution should be able to deal automatically with custom calculations and data sources that the customer provides. Otherwise, change management quickly becomes a problem, leading to low adoption.

Flexible and powerful — The solution should not restrict the user to a few predefined queries. Instead, the end user should ideally be able to do exactly what a coder is able to do. Otherwise, professional data scientists and coders remain a large bottleneck in your organization's analytics pipeline.

Scalable — The solution must be able to support today's large data volumes. This generally requires it to have some sort of query and calculation optimizer, which can get very technically complex. There's a tradeoff here too — between scalability and adaptability.

Existing solutions

Unfortunately, we find that no solution in the market meets all of these criteria. Broadly, there are several kinds of solutions in the process analytics space:

Business intelligence —   These vendors are selling their general-purpose BI software into the process analytics space. While these vendors can easily handle pivot tables and charting, the absence of proper mathematical modeling and machine learning makes them poor end-to-end solutions for process analytics.

Big data contractors —   Cloudera, Pentaho, Hortonworks, and other "big data" contractors are selling impressively broad suites of big data tools. While they provide scale and adaptability, and are intuitive to some degree, they lack the flexibility required to do heavy custom analytics (like machine learning or numerical simulations). In many ways, they're like BI solutions but with big data — they're glorified dashboard generators.

Data studios —   Much like us, these vendors (Alteryx, Informatica) are selling self-serve interfaces to data analytics. However, experience has shown that the complex interfaces they provide are, in practice, too difficult for the average analyst to easily learn. Power users develop around these vendors, leading to the same specialist bottleneck they claim to prevent in the first place.

"Cognitive" —   IBM and Wolfram are selling products that claim to be intelligent enough to support querying data in English. These should be addressed separately.

IBM Watson —   Watson's impressive Jeopardy and oncology demos have defined the category of cognitive computing. However, IBM Watson isn't a product — it's a platform. To build a full solution resembling the Watson demo, you need to hire an expensive army of IBM contractors. Watson's extremely long time-to-value, combined with its vendor lock-in, make it a much less attractive offering than the marketing videos suggest.

WolframAlpha Pro —   Wolfram's offering is intuitive and flashy — as long as you're not doing data integration or basic process engineering models, neither of which are built into the product. Users have to do ETL manually before sending data to WolframAlpha, and have to write their own numerical models. You're also locked into the walled garden of Mathematica, because that's the only way you can extend

WolframAlpha — not to mention that you have to buy Wolfram's proprietary hardware, deepening the vendor lock-in.

App creation tools —   Tools like Shiny and Mode Analytics make it easy for your developers to make apps for the rest of your organization without knowing Javascript or other frontend technologies, so you can make your own software to fit your needs. However, you're still bottlenecked on R specialists to write the apps. Furthermore, each app can only do a very narrow range of things, since trying to implement a complex enough frontend to handle the full spectrum of process analytics questions is a disaster in the making with Shiny — it's not a very maintainable system at scale. This forces your coders to write apps over and over again when users' needs change.

Introducing Lightbeam

Your engineers and analysts can type English questions into Lightbeam and get instant answers. It's not just impressive technology — it's strategically relevant to your business, because Lightbeam satisfies the four important characteristics of a great self-serve analytics solution:

Intuitive — Nothing can be more intuitive than a tool that allows users to type a query in plain English.

Adaptable — Our automatic data integration technology means fewer contractors and shorter time-to-value. We can plug into existing databases, vendor systems, and code, and none of this requires much custom work. Change management is a breeze.

Flexible and powerful — Unlike other analytics systems, Lightbeam is Turing-complete — a term which means that Lightbeam can answer such a large set of questions that your analysts have access to the infinite richness of coding without the effort or coding skills required of Python or R specialists.

Scalable — Our Photon query technology is based on the unparalleled performance of Haskell, which also boasts the most advanced parallelization and cloud libraries out of any programming platform. Lightbeam can scale to petabytes of data.