TIBCO

Delivering Smart Insights and Decisions with Six Essential Smart Capabilities of Hyperconverged Analytics

How and why to democratize AI and machine learning for business and operations

Leaders in every organization are now keenly aware that data and analytics, if used wisely, can help quell uncertainty, reduce risk, and provide a resilient, sustainable competitive advantage. In these times of unprecedented change, many organizations have a strong sense of urgency to understand not only what is happening in their business at this exact moment—but more importantly—what is going to happen in the next second, minute, day, week, month, or even further into the future.

Increasing Data Volumes

Data is expected to grow to 175 zettabytes by 2025, and that magnitude is incomprehensible to most. If you were to store it all on DVDs, you would have 23 stacks up to the moon. Or, it would take you 1.8 billion years to download it all using an average internet connection speed.[2]

However, despite increased spending on AI and related technologies, many organizations are failing in their endeavor to become data-driven.[1] They struggle with using artificial intelligence, data science, and machine learning technologies to extract actionable insights from the ever increasing deluge of data that's available to them. At TIBCO, we have learned that to extract maximum value from business and operational data, you need a new approach: hyperconverged analytics. It's the fastest and easiest way to generate the profound business insights that will transform your business.

Hyperconverged analytics is the next evolution of business intelligence technology. It brings visual analytics, data science, and streaming analytics together in a seamless, easy-to-use, and tailored experience that delivers immersive, smart, and real-time insights.

Hyperconvergence enables you to become AI-driven by infusing data science and machine learning into your business applications. It allows data scientists to be more productive, with technologies like automated machine learning (auto ML), and it empowers them with the flexibility to use any scripting language, cloud service provider, and machine learning service or framework for maximum value and effectiveness.

Why Smart, Hyperconverged Analytics?

You've seen the stats on explosive data growth. Your business is trying to capitalize on ever-increasing amounts of data. Whether that data is internal or external — transactional, text, images, audio, video, social, GPS, sensor, or other — it can be a struggle to provide cohesive and actionable insights to stakeholders.

By using data science and machine learning technology, you can find the proverbial needle in the haystack and separate the insight or signal from the noise.

In addition to the immensity of data, it's distributed, also presenting barriers. Data proliferation and analytic silos create much agony[3]:

- Analytic silos: lack of standards for coordinating and collaborating across descriptive, diagnostic, predictive, and prescriptive analysis methods.
- Decision-making silos: distributed decision-making without broader cross-functional organizational context

A strategic plan and comprehensive approach to hyperconverged analytics can help address these barriers and provide an opportunity to improve business decisions and ultimately business results.

---

1 Bean, Randy and Thomas H. Davenport. Companies Are Failing in Their Efforts to Become Data-Driven, Harvard Business Review, February 5, 2019. https://hbr.org/2019/02/companies-are-failing-in-their-efforts-to-become-data-driven

2 Data Age 2025: The Digitization of the World From Edge to Core. IDC White Paper, sponsored by Seagate, November 2018.

3 IDC. Readying Enterprises for the New Generation of Analytics. September 2020. https://www.tibco.com/resources/whitepaper/readying-enterprises-new-generation-analytics

What Are Business Decisions?

At the core, business decisions are fairly simple to understand.[4] Your organization could be making literally thousands (if not millions) of decisions on any given day.

Figure 2 shows types of decisions (strategic, tactical, operations, execution) that can be made in batch or real time, along with the time increments associated with them. Strategic decisions could be made quarterly while operational and execution decisions may be made in seconds, minutes, hours, or days.

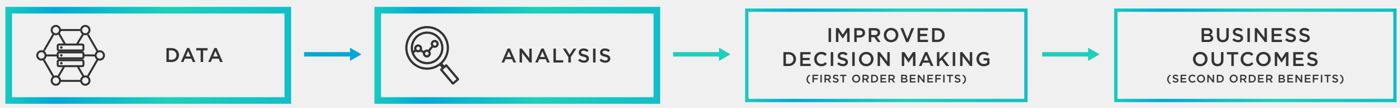

Figure 1: Data and analytics should be used to improve decision-making and business outcomes.

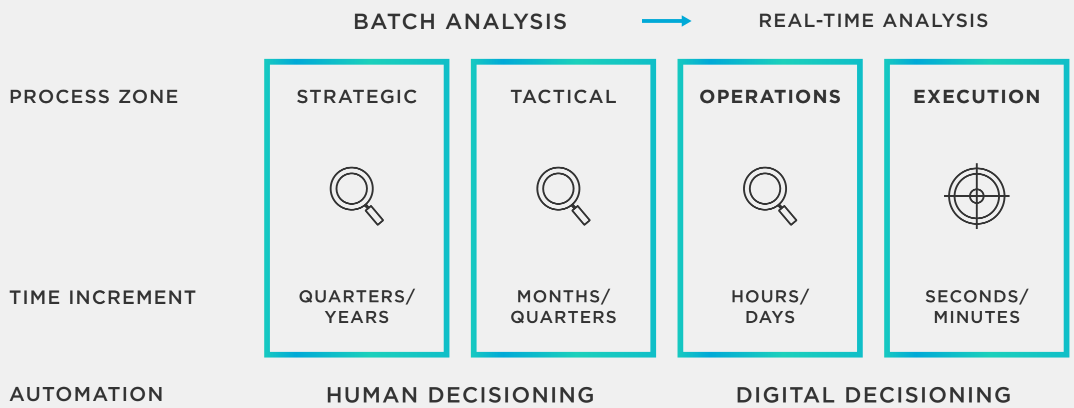

Figure 2: Decision time scales for human and digital decisions.

Organizations that can make repeatable, AI-infused digital decisions in real time outperform those who cannot. Most large organizations have a plethora of dashboards and reports; However, many of these information hubs cannot surface actionable, trustworthy insights that decision-makers can use to change organizational behavior and performance.

4 Ibid. IDC, 2020.

Insight without Action Has Limited Value

If it isn't happening yet, very soon your organization will need to move beyond zombie dashboards and provide actionable insights to the right person, within the right context, at the right time.

Figure 3 represents a typical decision process within an organization with the value associated with the decision on the y-axis and the amount of time required to make the decision on the x-axis.

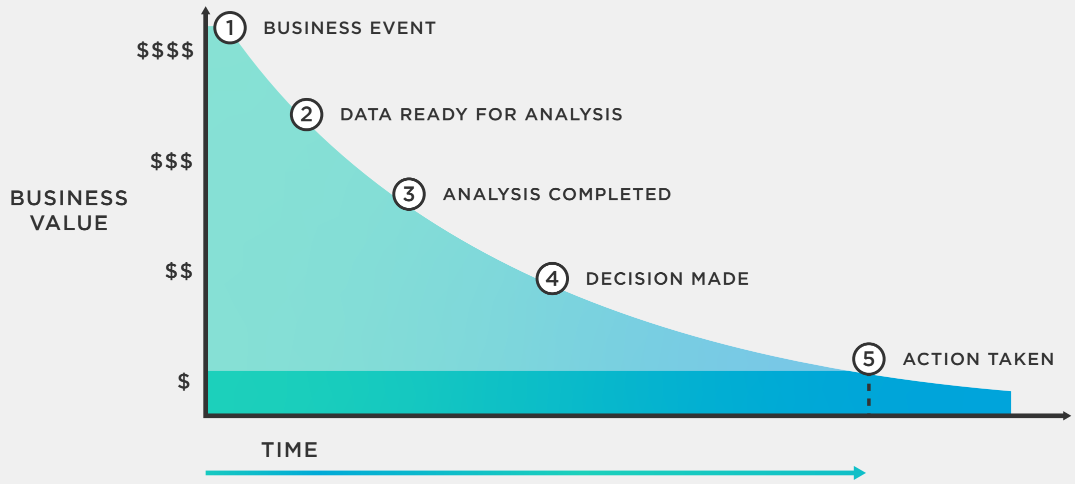

Figure 3: Typical digital decision process within an organization.

The trick is to act on business events before they perish and the information value contained within the data decays. As illustrated in Figure 4, if you can shorten the time between business event creation and action taken, you increase the business value of the decision.

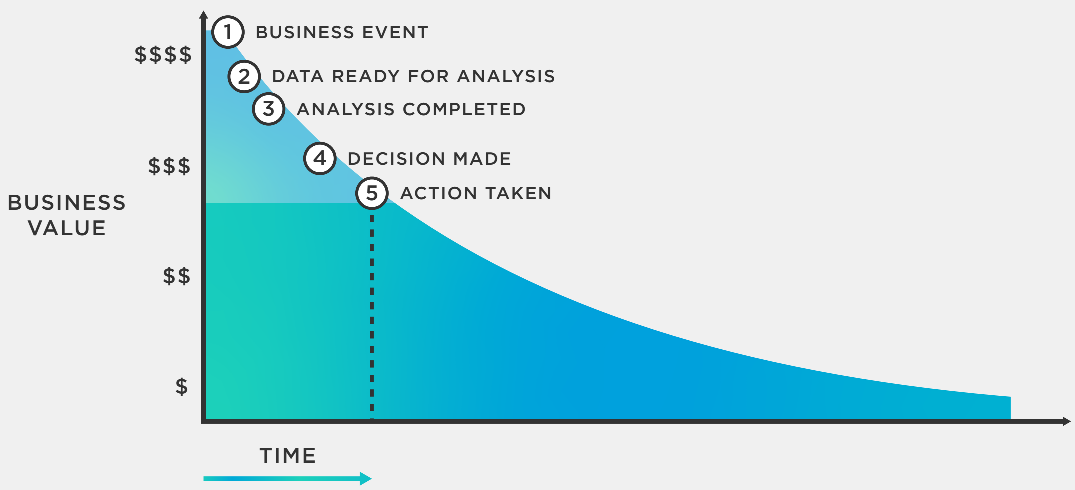

Figure 4: Increase business value by shortening the time from business event to action taken.

To reduce time and increase business value, you need to reimagine your approach to analytics.

Six Essential Smart Capabilities of Hyperconverged Analytics

1. Democratizing AI with ML-infused apps.

As previously mentioned, the goal for any organization is to deliver actionable insights to the right individual, at the right time, within the right context. If done smartly, your users may not even be aware that artificial intelligence is powering their business application. AI-infused applications go beyond the rearview mirror; they provide you with foresight, optimization, and what-if scenario analysis to ensure that the best possible decisions are made.

To accomplish this, TIBCO Spotfire analytics has a unique capability that allows data scientists to create data functions that deliver advanced analytics to business users. As an example (Figure 5), using recent health data from the World Health Organization (WHO), and socioeconomic data from the World Bank, we created a dashboard that allows government health officials to compare life expectancies, and factors that contribute to it, in countries around the world.

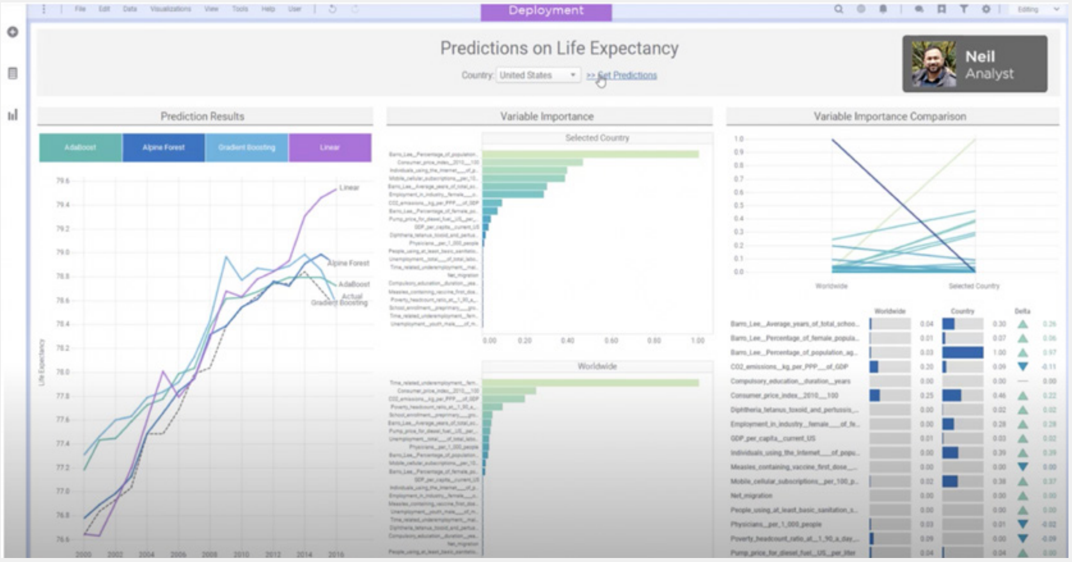

Figure 5: Predicting life expectancy by country, dynamically generated prediction outputs are displayed from a TIBCO Data Science Workflow in Spotfire software.

When a user wants to compare two countries, they simply select the countries within the dashboard, and the data science predictive workflows run behind the scene to display the predictions and factors driving the results. The data science workflows include both native functionality and an embedded Python notebook using an Adaboost model from scikit-learn.

2. Eliminating the mundane with automated AI infused insights for citizen data scientists.

Various surveys and analyst reports note a shortage of qualified data scientists. To help accelerate the data science process, many businesses rely on business users, domain experts, business analysts, and citizen data scientists to find insights from data before engaging data scientists. After the insights are generated, they may be shared with data scientists for additional investigation.

To help business users quickly surface insights from among the quantities of data, TIBCO Spotfire software uses AI to automatically correlate and provide best-practice visualizations so fast insight is possible.

In the following example, a business user is examining data and selects a population of interest. As soon as the data is selected, a list of correlations, relationships, and best practice visualizations are presented for further investigation. The user can quickly add the items of interest to their dashboard and continue the investigation.

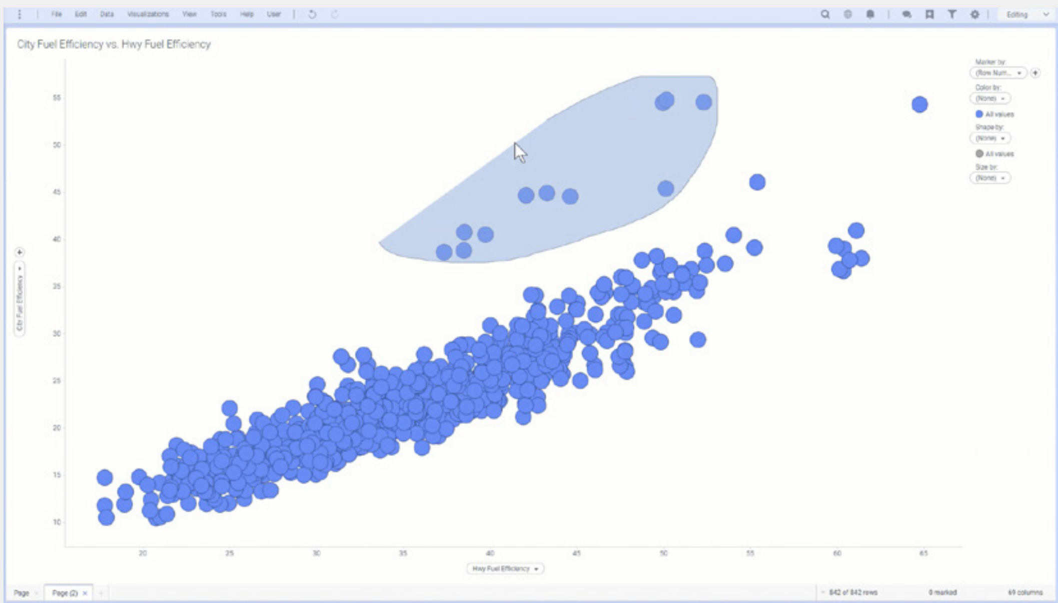

Figure 6: User select data of interest from Spotfire dashboard.

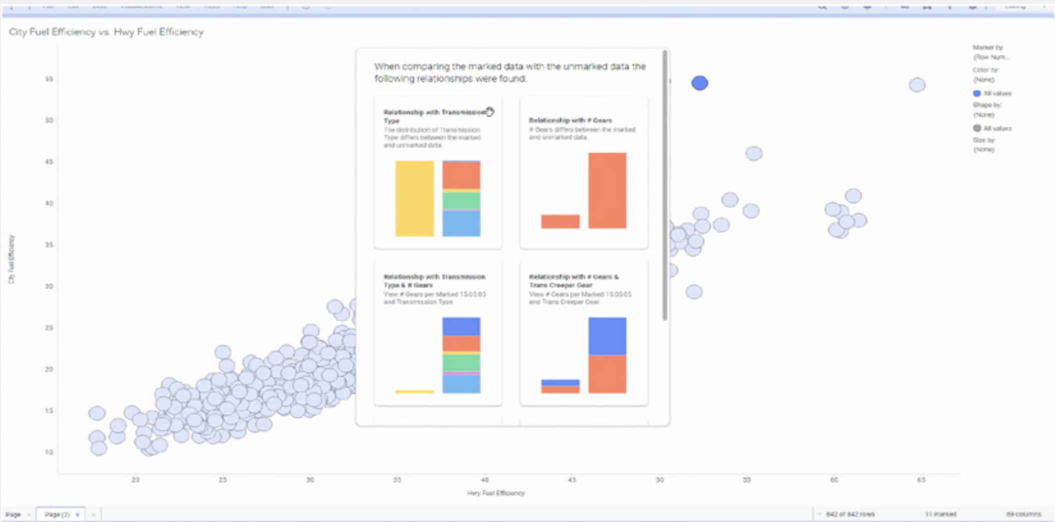

Figure 7: AI-infused insights are possible by automatically suggesting correlations, relationships, and best practice relationships from the selected data.

"By 2022, 40% of machine learning model development and scoring will be done in products that do not have machine learning as their primary goal."[5]

In addition to lassoing points on a graph, the Spotfire natural language query (NLQ) feature provides a search box (similar to Google) for asking questions of the data, with relevant results displayed. With embedded data science and ML algorithms, Spotfire software eliminates many manual steps and automatically presents items of interest.

3. Accelerating data scientist productivity

Low Code or No Code

Modern businesses today, generally have a hodgepodge of separate tools for data management, business intelligence, data science, and application development used by different types of users. Some users prefer to code in Python or R, others prefer a low-code, drag-and-drop interface.

With an end-to-end data science and machine learning platform like TIBCO Data Science software, your data scientists have complete flexibility. The TIBCO Data Science platform allows them to create drag-and-drop ML pipelines on a canvas. This feature accelerates productivity by eliminating the need for code that performs mundane tasks (ingest data, fill in missing values, remove outliers, etc.). Then, when the time is right, the data scientist can collaborate with a machine learning engineer and use an embedded Python notebook directly within the workflow. The neat thing is that this Notebook automatically introspects the upstream and downstream nodes and makes the necessary data connections, which eliminates low-level coding. Users can then use the Notebook as part of the workflow to solve very complex problems.

---

5 Idoine, Carlie. Worlds Collide as Augmented Analytics Draws Analytics, BI and Data Science Together. Gartner, March 10, 2020.

Transparent and Editable AutoML

To accelerate development of data science and ML pipelines, data scientists often use automated machine learning (auto ML) as a quick start guide and best practice. TIBCO Data Science auto ML nodes automate many of the mundane steps of the data science process, including data preparation and transformation, feature generation and selection, and model creation and selection. What sets the TIBCO Data Science platform apart from many others is that the workflows are transparent and editable. In other words, in addition to the "big easy button" that automates the process, data scientists can still edit each of the automatically generated workflows to adjust and add their own expertise for a better result.

Figure 8 shows the different workflows generated when the TIBCO Data Science AutoML Orchestrator is used. Each workflow is fully transparent and editable.

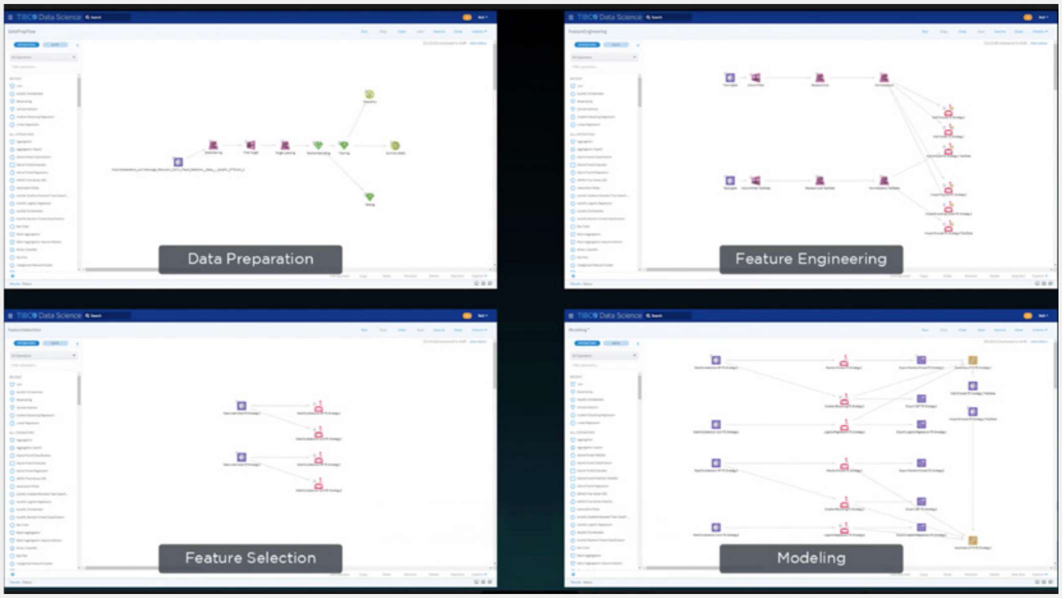

Figure 8: Transparent editable workflows generated from the AutomL Orchestrator in TIBCO Data Science software.

4. Working with Cloud Services: Google, Amazon, and Microsoft

From data storage and scalable computation environments, to algorithms and ML frameworks, many data scientists look to platforms like Amazon, Google, and Microsoft to accomplish a variety of AI use cases. For example, many use the embedded Python Notebooks for computer vision, speech, language, translation, text to speech, and emotion analysis. These include but are not limited to:

- Amazon for Anomaly detection, SageMaker for deep learning, Rekognition for image, video analytics, and object tracking
- Google Cloud for TensorFlow for deep learning, Natural Language API for sentiment analysis, entity analysis, entity sentiment analysis, content classification, and syntax analysis
- Microsoft Cognitive Services for anomaly detection, vision, speech, language, and emotion analytics

With TIBCO Data Science software, you can seamlessly use resources from any and all cloud providers, providing your organization with ultimate flexibility, and alleviating the need for your data scientists to learn complex low-level coding associated with any of them.

As an example, Figure 9 illustrates image recognition using AWS services and results rendered in Spotfire analytics.

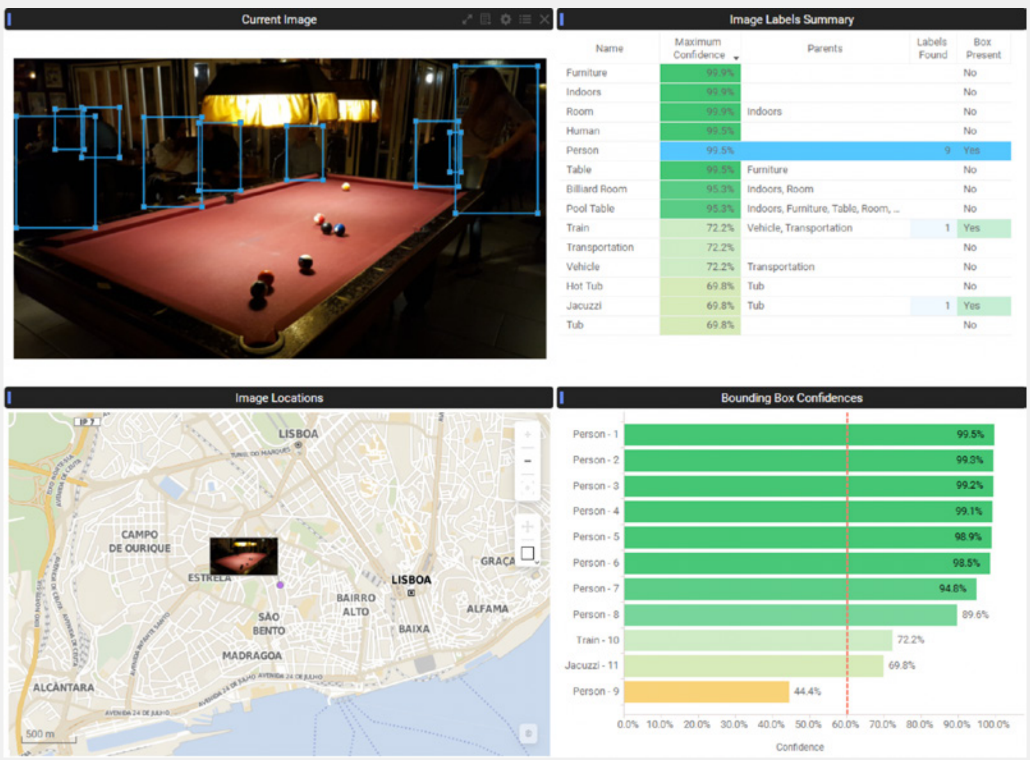

Figure 9: Image recognition with Spotfire Analytics and AWS via Spotfire data functions.

5. Scaling data science across the enterprise with model ops.

For your organization to create AI-infused business applications, you need a way to create models with a data science and ML platform, and deploy, distribute, and monitor the models to target business systems. Machine learning models are very different from traditional software applications.

Model operations (aka model ops, ML ops, AI ops) is a cross-functional, collaborative, continuous business process that operationalizes data science by managing statistical, data science, and machine learning models as reusable, highly available software artifacts via a repeatable deployment process.

Why do you need model operations?

When organizations deploy traditional software within the enterprise, the software will forever continue to work as designed. Yes, there may be bugs and updates necessary, but overall, you simply deploy the software (WorkDay, Salesforce, etc.), and it continues to function.

For data science and machine learning, it's a different story altogether. The process of building a data science or machine learning model generally involves taking historical data and training the model to make a prediction on data it has never seen before. Remember, the formula y = mx + b from middle school? Well, that's a very simple linear model where y is the prediction, m is the slope of the line, and b is the intercept. The challenge you have is, if the underlying data changes (which it always does), your deployed model may not be providing an optimal prediction. In other words, the deployed models within a business will start to decay (make suboptimal predictions) as soon as they are deployed. Therefore, you need a systematic way to monitor the performance of models and update (aka retrain) and redeploy the models as needed based on model monitoring metrics.

What is model operations?

Model operations (aka model ops, ML ops, AI ops) is a cross-functional, collaborative, continuous business process that operationalizes data science by managing statistical, data science, and machine learning models as reusable, highly available software artifacts via a repeatable deployment process. It encompasses unique management aspects that span data prep and transformation steps, model deployment, model inference (aka predictions), scalability, maintenance, auditing, and governance, as well as ongoing model monitoring in production to ensure they continue to deliver business value as conditions change.

What does it take to put a model into production? Figure 10 shows the people involved.

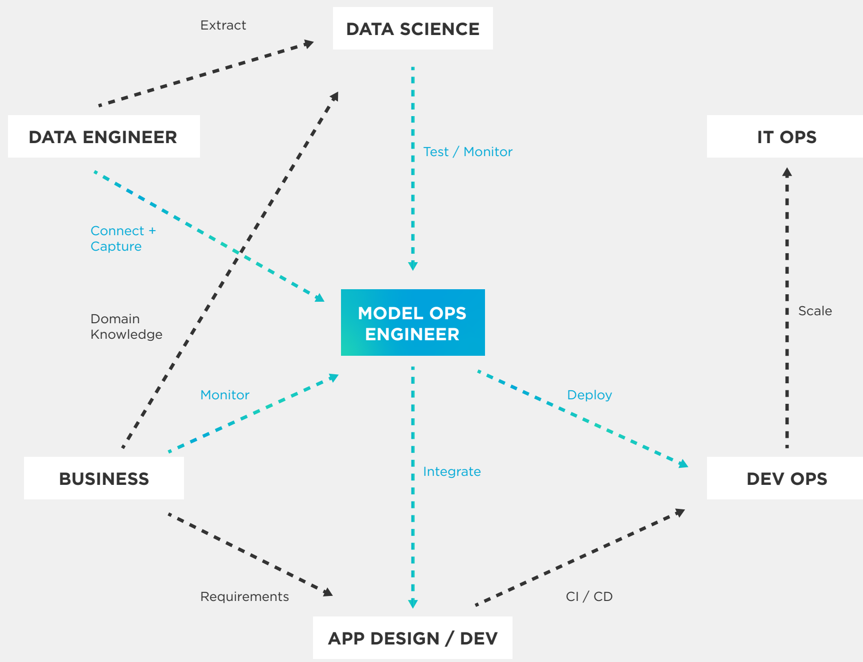

Figure 10: Team members involved in operationalizing data science models.

Operationalizing AI model ops takes a village. Let's say you are an insurance business that wants to deploy machine learning to dynamically price your policies. Here are the roles involved in that project and what they do:

- **Business owner:** Tries to grow the business. Their main objective is to sell insurance policies and grow premiums, determine go to market methods and premium offers. Ultimately, they want to keep good customers and acquire more good customers, those who always pay ontime with the lowest possible risk.
- **Data Scientist:** Fits statistical and data science models for retention and acquisition. Provides micro segments for discounts depending on tenure, claims, payments, etc.
- **Model Ops Engineer:** Sets up infrastructure for saving the models that provide the quotes (and discounts), monitors the acceptance of models, monitors model health and quality, and retrains or remodels as conditions change. The model ops engineer generally reports to the head of data science and head of pricing (business) for approvals.
- **Data Engineer:** Manages the database that stores quote data. Also works with the model ops engineer in creating data preparation pipelines to match event data with data at rest for model inputs and outputs.
- **IT Dev-Ops:** Manages the continuous integration/ continuous delivery (CI/CD) pipeline. Generally owns development and management of the quote manager, model manager, system components, and integration with other systems (for example, fraud detection).

- **IT Ops:** Monitors and manages daily operations of the infrastructure consisting of hardware performance, maintenance, support of pricing systems, and other systems such as claims and fraud.
- **App Design:** Works with dev-ops and the business on the user experience, customer experience, and in-offer management.

6. Hyper-personalizing with advanced spatial analytics.

As mentioned earlier, the trick is to act on business events before they perish and the information value contained within the data decays. Many business events and corresponding data have a geolocation associated with them. Visualizing the data on a map brings it to life, but that really only scratches the surface. The ability to overlay map layers (think street view and/or terrain view in Google maps) is important for many use cases. The ultimate capability is using advanced analytics, data science, and statistical analysis between any two points on the map within and across layers. If generating a Varonoi polygon that summaries multiple points on your map layer would save time and clarify the data, you want to be able to provide that feature.

Beyond geospatial coordinates, advanced spatial analytics is also used to understand and hyper-personalize actions for factory floors, semiconductor wafer map analysis, store layouts, human bodies, and petrochemical geophysics.

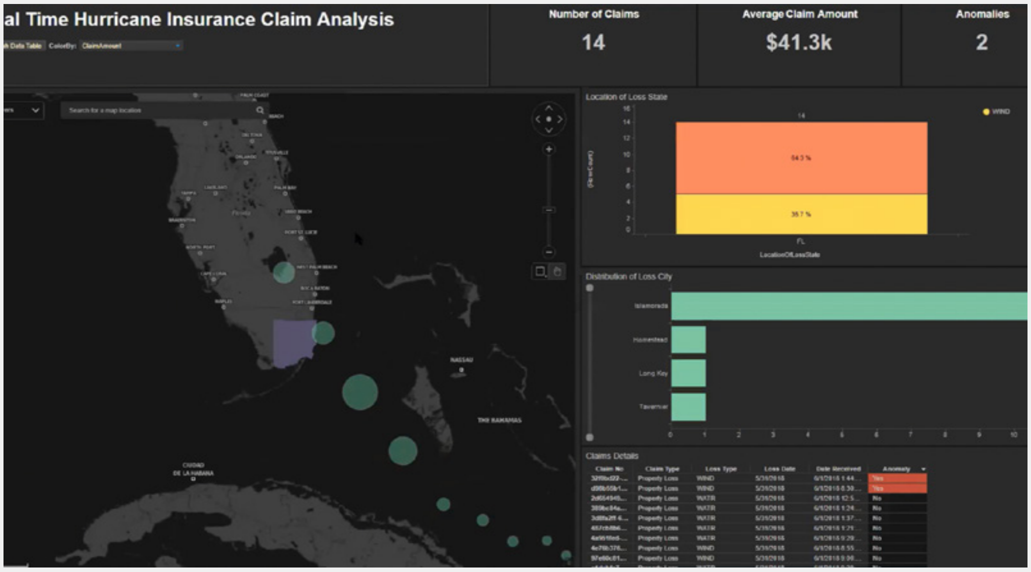

Figure 11: Real-time insurance claim predictions based on a hurricane path.

In Figure 11's example application, analysts are investigating claim transactions across the path of a hurricane in Florida. The insurance company can route claims more efficiently into categories such as: pay, reject, or investigate for fraud.

A logistic regression model that predicts the probability of claim anomaly/fraud, has been configured in Spotfire analytics and published to the Spotfire Server (Spotfire Data Streams software) for scoring claim transactions. Claim scoring, assessment, and routing is performed in real time on an event stream or as a batch process. All Spotfire visualizations are interactive and may be readily configured to respond to real-time data updates.

A significant count of claims in Broward County is represented by the size of the circle in the map chart. The multilayered Spotfire map chart includes the ability to do calculations and modeling within and between layers, and insurance analysts can identify the anomalous claims from the logistic regression model as pink circles on the map chart. Spotfire makes drill-down between visualizations easy, within and between analysis of real-time event streams alongside historical data, updating marking schemes automatically. Analysts may also annotate and assign claims for further investigation. Machine learning models embedded in the application can be called from a case management application to track the business process and the outcome of claims processing in a closed loop.

Smart Hyperconverged Analytics in Action

Norfolk Southern Railway Creates AI-Infused Apps and a Network Operations Control Center

Norfolk Southern is one of the largest railway operators in the United States. With about 20,000 rail miles that serves every major container port, it operates the most extensive intermodal network in the Eastern US.

At $100,000 each, 65,000 railcars—a very expensive $6.5 billion asset—are used to transport automobiles and trucks. The railcars have several IoT sensors so the company can know precisely where they are at any given time, forecast arrivals, and quickly notify customers if their shipments might be delayed. The company built an AI-infused portal for customers to track their shipments.

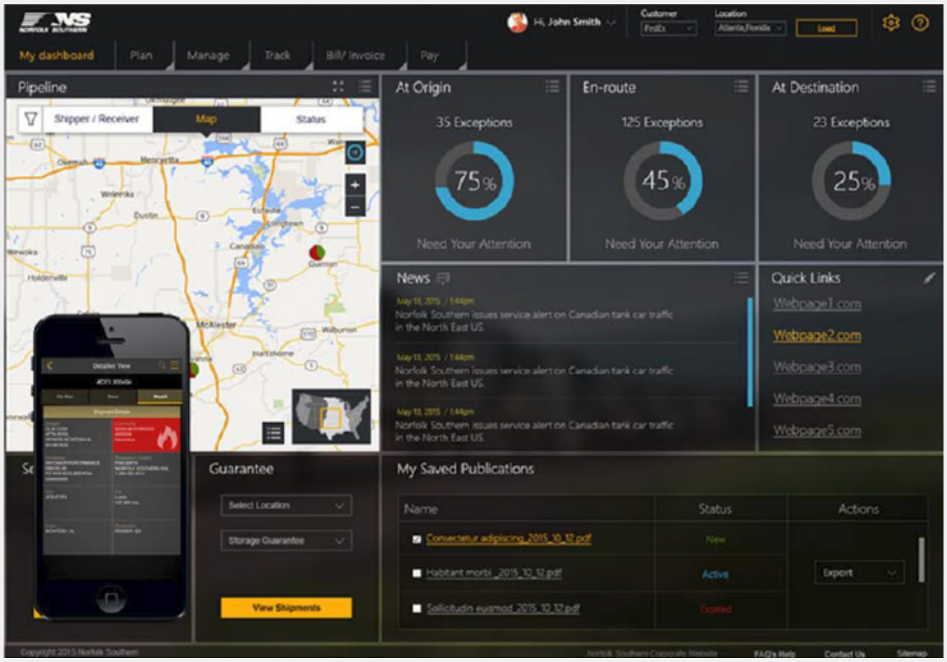

Figure 12: Norfolk Southerm AccessNS customer portal.

Fueled by data science, delivering real-time insights not only increases customer satisfaction, it improves business performance. The company has increased average train speed and operational efficiency, and improved delivery time.

AA Ireland Uses Model Ops to Create a Dynamic Pricing Application for Insurance Quotes

AA Ireland uses model ops for dynamic pricing. Following the four steps of build → manage → deploy → monitor, the company created a closed-loop continuous learning application.

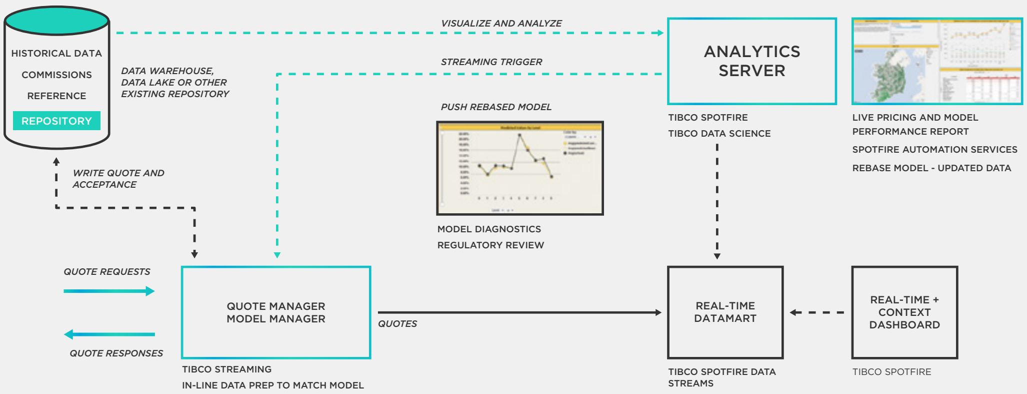

Figure 13: AA Ireland's model ops workflow.

For many years before the company operationalized data science, it used a market-dominant analytics software solution. Change was far riskier than simply hiring people who knew how to use Spotfire software, but since embarking on the TIBCO model ops process, the company can hire data scientists with no coding or math background. They are able to release models within three months.

The insurance company has always had predictive models, but they were subject to three-month update cycles. With model ops and TIBCO software, data goes out to models almost as soon as it comes in, updating them in real time. Rather than generating simple predictions, the company can answer questions like "If we increase or decrease discounts, what is the effect on volume and profitability?" The team can now understand total opportunities and risks in the market and make informed decisions.

PerkinElmer Creates Deep Learning Applications to Analyze Image Data from Cells within the Life Sciences Industry

In its Signals application,[6] PerkinElmer, a life sciences company, makes extensive use of AWS and TIBCO Data Science and TIBCO Spotfire software.

It developed a high content screening application that uses automated microscopes to capture thousands to millions of images of cells as well as deep learning to automatically analyze these images and extract different cell features from them. Scientists subject the cells to various compounds or genes and monitor to understand how they respond to various treatment approaches.

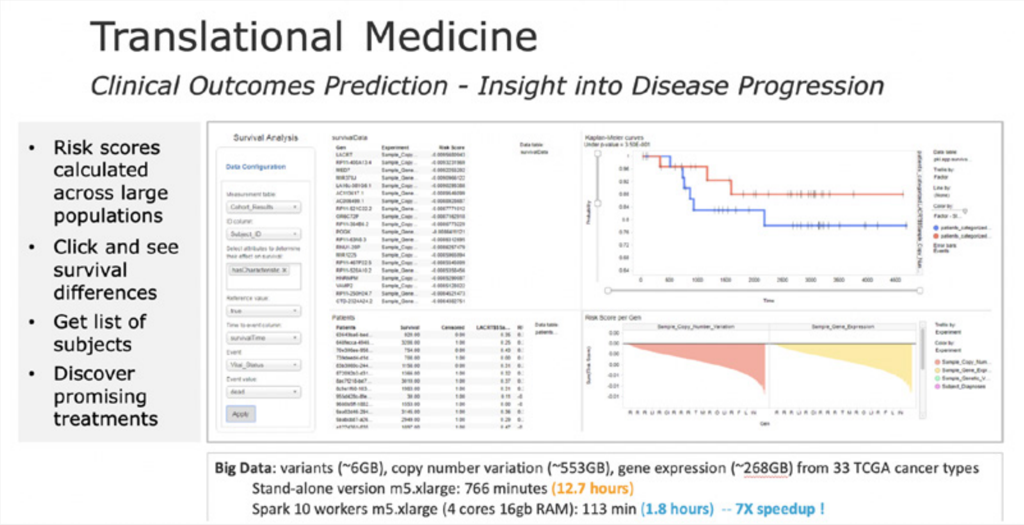

Figure 14: Translational medicine application using deep learning and image analytics.

6 DSC Webinar Series: A Collaborative Approach to Machine Learning, May 6, 2020. https://www.datasciencecentral.com/video/dsc-webinar-series-a-collaborative-approach-to-machine-learning

Smart hyperconverged analytics accelerates the arrival of better business outcomes: increased revenue and reduced costs.

Why it Matters

Whether you are beginning your analytics journey or already an expert, hyperconverged analytics technology will help you accelerate your business. Start with the end in mind. Get the technologies and talent in place, and measure the business value and outcomes of your analytics initiatives. Success breeds success, so start small and pick a tangible use case (in tight alignment with your company's strategic goals) that can be accomplished within six to nine months. After this, you can continue to evolve and build your organization.

If you heed the advice in this paper, you will set your organization on a path to analytics nirvana that will help increase revenue or margins by one to two percent and reduce costs by 5 to 30 percent.[7] If you proceed haphazardly with poor quality data and the technologies and processes of old, you will be outflanked.

With hyperconverged analytics, you can embed analytical findings and intelligence into all business processes for:

- Richer, deeper, immersive discovery—without needing more resources
- Decision support where and when your users need it
- Game-changing new opportunities for cost savings and market opportunities

To learn more about hyperconverged analytics capabilities, check out this comprehensive resource hub.

How to Get Started

TIBCO Data Science and TIBCO Spotfire Analytics are fully integrated. The combination forms the backbone of smart, within smart, immersive, real-time hyperconverged analytics. Learn more about TIBCO Data Science software.

7 Hurtgen, Holger and Niko Mohr. Achieving Business Impact with Data. McKinsey, April 2018. https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/achieving-business-impact-with-data

Global Headquarters
3307 Hillview Avenue
Palo Alto, CA 94304
+1 650-846-1000 TEL
+1 800-420-8450
+1 650-846-1005 FAX
www.tibco.com