RAMP UP YOUR ANALYTICS CAPABILITIES

A practical guide to go from PoC to long-term Advanced Analytics success

POST /DataRetrieve HTTP/1.
Host: 192.168.1.1
Content-Type: ap
Content-Transfer-
$zfile = "uploa
$zmessage = "
<?xml version="1.
$fp = fopen($zfile, "a+
$m:SecureHeader>****</m
?m:SecurityArray>********</
</encrypted-wrapper>
<verifiedToken>
report value 88268;
</verifiedToken>
</xml>

ARE YOU READY TO SCALE UP YOUR ANALYTICS SUCCESS?

Enthusiasm for advanced analytics techniques like predictive analytics and machine learning is mounting with over half of business and technology professionals saying they are investigating artificial intelligence, for example[1]. As the volume and types of data available continue to grow, the only limits on the potential of that data are our own abilities to collect, store and analyze it.

As industries have transformed to become data-driven, many examples of how advanced analytics can drive business success have started to emerge. Some examples include:

- Phone companies use information about customer behavior to identify customers at risk for cancelling their subscription. This allows them to engage those customers, reducing churn and improving revenue.

- Ecommerce companies use data about their consumers to provide tailored online experiences and product recommendations which increase sales

- Marketing agencies are able to identify customers with the highest likelihood to buy and tailor offers to just those customers, increasing the efficiency of marketing investments[2].

Fortunately, it's relatively straightforward to start experimenting with advanced analytics use cases using data you already have. Most organizations will do so through a structured proof of concept (PoC) that addresses a defined business need, and many of these PoCs will be successful in demonstrating the value of advanced analytics. Once over this initial hurdle, however, many organizations struggle to implement at scale, and for the long-term.

Ramping an advanced analytics initiative from PoC to a core capability of the organization presents a number of challenges. For example, while a PoC typically uses a finite amount of historic data – easier for testing and training analytics models with – using larger volumes of streaming, real-life data can put greater pressure on your IT team and technology resources. It also needs buy-in and engagement from key stakeholders to adapt to new systems that enable them to act on the results of advanced analytics solutions. This requires legwork upfront to integrate your analytics tools with key systems and applications, and to drive acceptance of ongoing changes to established workflows across the organization. Also, training the key stakeholders on how the data is being used and how to inteprate the data helps drive adoption of and builds trust in these analytic systems.

Intel has worked closely with many organizations that are attempting the cross to chasm[3] of early adopters to bringing advanced analytics into mainstream use. According to Deloitte, 21 percent of projects are canceled prior to being delivered, or are never used[4]. Not fully planning for the organizational preparedness required to respond to advanced analytics insights is one of the most common reasons we see for PoCs stalling as they move into production. This paper addresses best practices and frameworks related to three exercises that can increase the likelihood of success:

1. Conducting an analytics capabilities assessment

2. Building an insights-focused, cross-organizational team

3. Defining an analytics process

STEP ONE:
CONDUCT AN ANALYTICS CAPABILITIES ASSESSMENT

It can be tempting to jump into a new advanced analytics project and start work on developing a model as quickly as possible. Most organizations have several worthy projects in mind already (more on use case selection later), but even so, it pays to take a strategic view. An effective approach is to adopt a capabilities-driven strategy[5] that maps your analytics roadmap to your overall business strategy.

Examples of organizational capabilities are innovation, customer focus, leadership, and people focus[6]. First, determine your **current strategy, capabilities and future goals** and how they support the organization's business model. This is a critical step in ensuring the right executive sponsorship, strategic focus, and organizational buy-in. Consider what differentiates your organization from others, and how those differentiating factors tie into the services that are offered to customers today. What new services will these capabilities enable your organization to offer next year, or in five, even ten years? Often these capabilities will be communicated by the executive management team. Make sure that the advanced analytics program is additive to these key organizational objectives. You might even draft a team mission statement that explicitly states how the two relate on specific elements like customer service or risk mitigation. Explore opportunities to bring data-centric thinking into your organizational strategy (if it's not there already), or start even simpler by considering how different data sources can be united to help streamline decision making in real time. Many forward-looking organizations today have 'being data-driven' as one of their core competencies, making it easier to align analytics and business strategies.

Assess your analytics capabilities and needs with an annual SWOT analysis. Build an action plan for competency development based on your findings.

Once you have an understanding of how the program will align to the broader organizational strategy (why the program exists), the next level of detail is evaluating the current resources available. At this stage you'll need to assess a number of areas, particularly your human and IT resources. Start by conducting a **SWOT analysis**, again thinking beyond the initial PoC to the long-term operation of your proposed analytics use case. Ideally, you should conduct such an analysis annually, to ensure that evolving needs and capabilities are taken into account as new analytics (and other) initiatives are started.

Then consider what **capabilities and tools** you already have in house to achieve the goals you've identified. This analysis can generally be broken down into human capital and technical resources.

Human Capital

It's important to understand what skills your existing team has or could easily develop, and where it would make more

sense to call upon external vendors or consultants, or to build and strengthen a team within the organization.

For example, some organizations already have strong data science capabilities in-house, so they may enlist external support primarily to help with implementation. Others choose to bring in data science support as well. An important consideration is how much of a change the organization needs to make with regards to adopting analytics. If a significant mindset shift is needed, it may be easier to have this message and the ensuing change management process facilitated by a partner.

If you decide to bring in external data science capabilities, think about whether you'll need your vendor to develop a bespoke solution, or whether you could use (or at least make a start with) an out-of-the-box option. The level of expertise of your vendor(s) will also be critical – do you need them to know your business and industry in-depth, or would you prefer them to bring a breadth of technical experience to the table from other industries, which your

own personnel can translate to work with your environment? Either way, your project will involve close collaboration and multiple meetings between business, technical and managerial stakeholders, so having a vendor that can speak the language of all these groups will accelerate progress.

IT infrastructure

In addition to human resources and skills, the technology available to you is also an important consideration when developing your competencies. When working with our customers on their technology plans, we recommend they make the most of the open nature of the x86 architecture by considering solutions from multiple vendors for each aspect of their data management and analytics environments. Avoiding being 'locked in' to a single vendor helps maintain flexibility as you trial and roll out your analytics programs. For example, if after the initial PoC, you find that an element of your chosen solution doesn't meet your needs, you can try a different one relatively simply and cheaply.
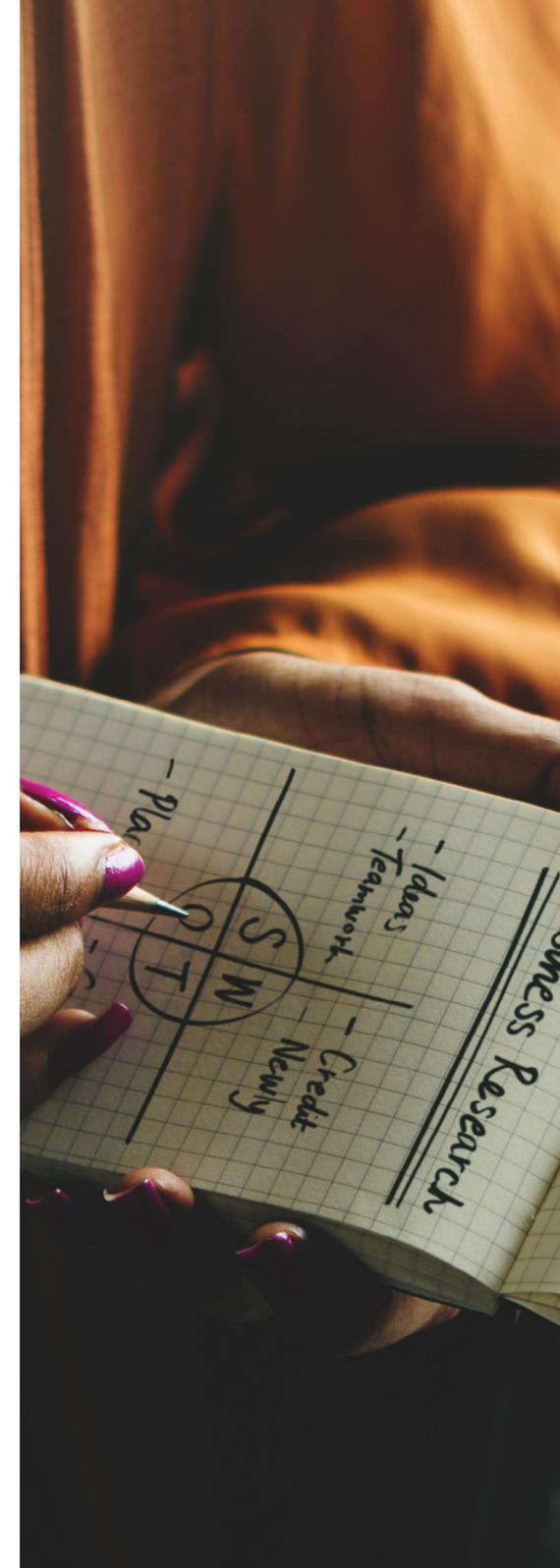
The same goes for software. Many open source frameworks and components – such as Apache Spark*[7] - are available online, offering the ability to adapt the code and develop tailored applications at minimal cost. Alternatively, organizations can adopt a best-of-breed approach, using Software-as-a-Service applications to mix and match the tools they need, and then working with a specialist integrator to stitch them all together. A number of cloud service providers (CSPs) now offer tools and capabilities designed for analytics (such as AWS Sagemaker*), which can help boost your capabilities with relatively little disruption.

Involving some element of open source can be useful for healthcare organizations that have strong ties to universities and other research bodies, which tend to make heavy use of open source toolsets. It's also possible to wrap open source components in an enterprise package (such as those from Red Hat or Cloudera) to achieve the flexibility of open source while also benefitting from the support of a third-party vendor.

Security must also remain a top priority. Whenever you consider using a third-party solution, make sure you carefully explore how any data it handles is secured. Many organizations will have robust security processes in place already that can be used with the analytics program.

Finally, bear in mind what you'll want to do next as you choose your analytics solutions ecosystem. Providers should offer a broad range of capabilities that will enable you to smoothly move on to more mature or complex analytics over time. Look for those that go beyond providing point solutions for a single use case, and rather offer support for whatever challenges or needs you face as you move forward – from basic tools to highly customized solutions.

Once you have assessed and identified your capabilities and technical requirements, you can then develop an **action plan** to address shortcomings. If additional human resources are required will you hire, train existing staff, or work with a consultancy or systems integrator (SI)? For technology, will you build out a new environment on-premise, use cloud resources, or some combination of the two? Involve key business and technical team members in these discussions so that plans align with the strategies for those groups.

BUILD AN INSIGHTS-FOCUSED, CROSS-ORGANIZATIONAL TEAM

The data management and analytics capabilities of your project are of course an essential aspect, but they're not the only one. Any initiative relies on the people that design, implement and use it to make it a success, so it's critical to also think about your organizational structure and how you will build an analytics-enabling team. Unfortunately, it's all too common for an analytics project to fail due to a lack of integration or communication between, say, the IT team that built the solution, the data science team that created the algorithm, and the business unit that needs to use the output. Even if that output is hugely valuable, if the end users have not been involved in its development, they may be resistant to using it.

Executive Sponsorship: For a project to achieve long-term success, it will need an executive sponsor – someone in senior management with passion for and commitment to digital transformation and an understanding of the role that advanced analytics plays in this journey.

This executive sponsor will provide the support, funding and top-down leadership you'll need not only for the PoC but also to enforce any workflow and cultural changes and ensure you can scale up your initiatives over time. In many cases executive sponsorship will need to come from more than one source. In John Kotter's classic text on change management he talks about building a "Guiding Coalition"[8].

Analytics Leadership: Your executive sponsor(s), while being closely involved throughout, will need help in evangelizing analytics across the organization. Identify other individuals in senior and mid-level management that can also help promote the vision and drive others to do their part in helping to achieve it. These should be individuals who see the value that analytics-driven insights will bring to their own daily roles, and so have a vested interest in getting the initiative up and running. They – along with their counterparts in the IT team – will also play an important role in helping work out how and where these insights should be integrated with existing systems and data reporting tools[9]. A good resource on how to maximize the impact of analytics leadership support is HBR's *10 Must-Reads On Change Management*[10] which provides several articles and frameworks for establishing and measuring the success of leadership during organizational change.

Funding: The budget to support analytics projects may come from a variety of sources within the organization. We have seen funding come from sources including lines of business, IT budget, vendors, and industry sponsorships. As the impact of the advanced analytics organization becomes more widely recognized, more funding should become available at an organizational level. Ensure that your funding providers are clear on what the costs of a project will be throughout its lifetime – for example, as a PoC is ramped up to be used across the organization, what additional investments will be needed and when?

Include business, technical, data science and managerial stakeholders from the outset. The more collaborative the process, the better your outcomes will be.

Organizational Design: There are a variety of ways in which you can structure your analytics capabilities, depending on the organization and the type of project. For example, you may choose to set up an independent analytics team within the organization that provides support to business teams for specific projects as needed. Alternatively, analytics responsibility could be split across each department, sitting underneath overall department management (for more on this, see Accenture's paper: *Building an Analytics-Driven Organization*[10]).

However you choose to split it, it's important to make sure roles and responsibilities are clearly defined. An example breakdown of an analytics team uses three job codes to define team responsibilities, (see figure 1):

- Development (technical/IT team members who ensure all systems integrate smoothly) – 50 percent of the team
- Analytics team (made up of data scientists) – 30 percent of the team
- Product (those responsible for ensuring smooth collaboration with internal subject matter experts, third-party stakeholders and other business groups with the analytics and development teams) – 20 percent of the team

Analytics Team Distribution

| 50% Development/IT | 30% Analytics | 20% Product |

Figure 1: Analytics team job codes and distribution

One of the main skills gaps for organizations looking to implement analytics is data science. Attracting and retaining top data scientists can be a challenge, so it's worth also developing a plan for how you will do this. Based on your SWOT analysis, you should have an idea of the specific technical skills you need and can focus recruitment efforts on meeting those requirements. Then also consider how you will retain and nurture that talent over time, including mapping out career paths for data scientists, many of whom do not wish to follow a typical journey that ends in management rather than doing their job function.

Role	Responsibilities
Business / domain experts	• Responsible for defining the context around the problem and what might be reliable data to use. They provide the vision and the use case, communicate any data constraints or assumptions, explain data/variable/feature relationships, and validate the results against business need. • Examples of domain experts (depending on problem context): Hospital management, Health experts (doctors, nurses, support staff, etc.), lab staff, clinical managers.
Data scientists / statisticians / data modelers (collectively known as data scientists)	• Work closely with the domain experts to understand the business problem, how the information will be used, any data constraints or assumptions, data/variable/feature relationships, and how to validate the results. • Help translate the business problem into analytic (mathematical/statistical) terms and explain the data back to the business problem. Should be able to articulate back to the team how any assumptions around the data impact the problem. • Define the data requirements. • Understand how to process data (cleaning, integrating, storing, querying, and analyzing), use a wide range of dimension reduction and predictive modeling techniques, and visualize the data for confirming or challenging data relationships and behavior assumptions.
Data managers / data architects / data engineers	• Responsible for the data management infrastructure, which enables processing and storing of the data for querying in a timely manner. The majority of this work (or automation) should be done as the data is ingested into the system, not after it is stored. • Create flexible data schemas for fast querying of the data. • Support the data requirements of the data scientists.
IT specialists / computer scientists / solution architects	• Bridge any technology gaps to meet the data scientist's computing, data processing, and data storage needs. Responsible for recommending the right hardware and software infrastructure.

STEP THREE:
DEFINE AN ANALYTICS PROCESS

Analytics programs, like any core business function, are a combination of resources and process. In the previous section we discussed several of the key resources. In this section we will address three processes:

1. Aligning with business challenges through a process like the Cross Industry Standard Process for Data Mining (CRISP-DM)[12] model

2. Working up the stages of analytic solution maturity

3. Designing the data process mapping pipeline.

It is important to understand the components, how they come together and which steps are necessary to build a successful solution. Failing fast and learning what to improve is important. Figure 2 shows how these methodology components fit together.

Before beginning the discussion of each process, it's important to point out that having a clear definition of who owns and contributes to each process is critical. This is an unstated assumption throughout the following sections. There is no right answer to this, as team skills, organizational politics and other factors can all play into the final design. Some of the key roles and responsibilities to consider are called out earlier in this paper. What's important is that each process has the right inputs, the right stakeholders, and the right executive sponsorship to ensure that the results are aligned with the business objectives and actionable.

Aligning with business problems

As discussed earlier in this paper, aligning analytics projects with business objectives is paramount for success. The CRISP-DM[13] is one of the more commonly[13] recognized methodologies from the data mining industry that drives business focus and comprehends the iterative nature of analytic solutions. Starting with the business understanding to identify the problems to be solved and then investigating how to solve them through analytic techniques is a journey. This iterative process enables you to:
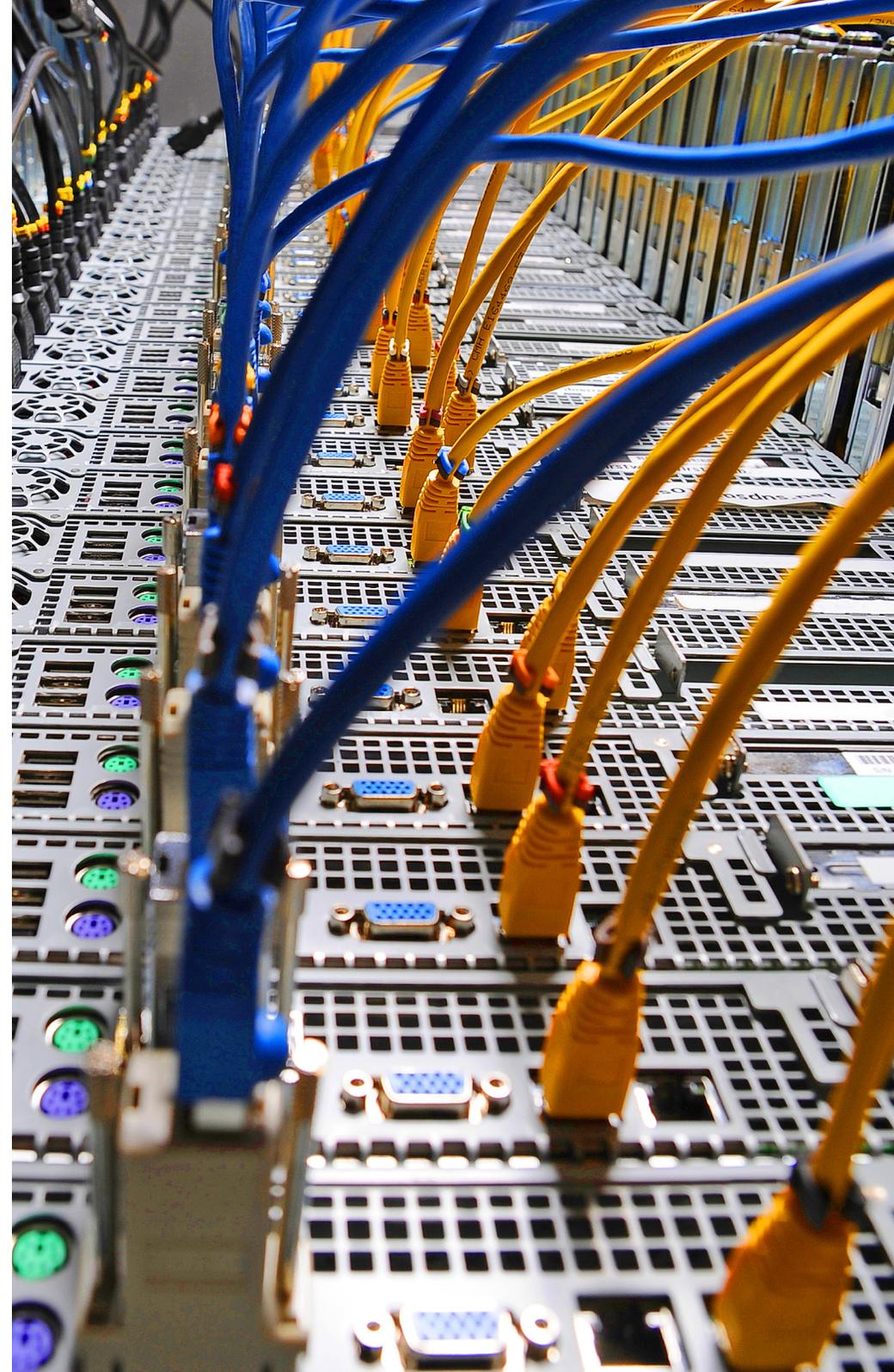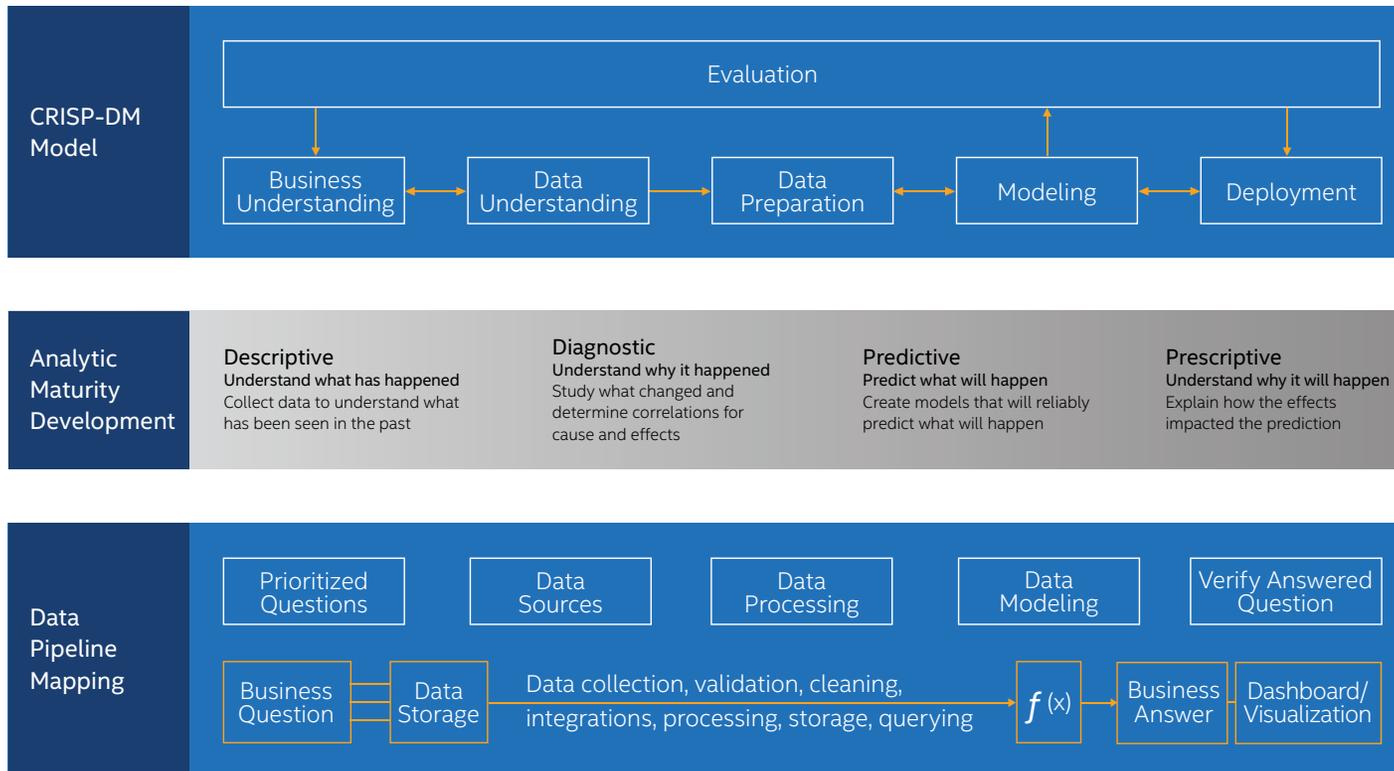
Figure 2: Alignment of the various methodologies described (CRISP-DM, analytics maturity, and data process mapping pipeline).

- Align the solution to business goals
- Evaluate how the right data is creating the desired impact to these business goals
- Adopt a fail-fast approach to learn what is not working and re-evaluate any disconnects to these business goals
- Deploy the agreed upon model for business implementation.

While the CRISP-DM provides a high-level methodology to ensure the business need and data are evaluated and stay aligned, it does not provide details around how this work should develop. The following analytics maturity model shows how the work with the raw data is converted into impactful insights, while the data process mapping pipeline discusses the process infrastructure requirements to support this data conversion. Teams should be able to flow among these three component methodologies to ensure the right evaluation, work, and requirements are being met for a successful solution.

Increasing analytics maturity

Like anything that is built, one needs to start with raw materials and start molding them to create a reliable product. For example, to build a car, metal parts start with raw metal that is melted and cast to build each part of the car. The car is then assembled after careful tolerance design to ensure parts fit and work together to create a drivable car. Analytics solutions are planned in a similar way, where material is raw data and the product is reliable insights to answer a business problem.

Below we'll look at how an organization might progress through the first four stages of analytics maturity as they develop a new prescriptive analytics solution. As mentioned before, an organization may not have the skills or technology to progress through all four stages straight away, but having a roadmap for how this process works can complement the capabilities assessment discussed earlier in this paper.

Descriptive Stage – *Understanding what has happened*

Data scientists work with domain experts to identify what data is needed to be collected to answer the business problem, and the data collection begins. The data scientists look at historical data to describe discernible outcomes or observations of various past data behaviors .

Diagnostic Stage – *Understanding why it has happened*

Once data scientists understand what behaviors have happened in the past, the next step is to understand why these behaviors occurred. This is where feature creation or extraction work begins. These features will be used to help explain why the behaviors occurred and start forming the foundation of the predictive model in the next stage.

Predictive Stage – *Predicting what will happen*

After features are identified and extracted, a data scientist will start to build models using these features to predict desired outcome behaviors. In building these models, many different algorithms can be used to test the accuracy, data assumptions, and best fit of the data. While this work is very much more an art than a science, the choice of the algorithms used is based primarily on data type being used and data assumptions being made. The data scientist will pick the best model by a series of trade-offs of the model complexity, accuracy, and modeling cost.

Prescriptive Stage – *Understanding why it will happen*

This is about understanding why a prediction happened or showcasing the impact of input features on the model's outcomes. This information is important to help migrate unwanted actions and exploit a desired outcome. It gives the business control over what decisions need to happen when, by monitoring (and ultimately controlling) the underlining unwanted causes.

The predictive model's feature parameters provide insight into its importance and impact on the predictions (i.e. its sensitivity). It is these features that data scientists need to explore, understand, and exploit for the prescriptive stage to be successful.

These stages align with the CRISP-DM methodology, showing how insights can be extracted from raw data at the different maturity stages. The descriptive and diagnostic stages follow the Business and Data Understanding and the Data Preparation steps. The predictive and prescriptive stages follow the Modeling, Evaluation, and Deployment steps.

Data process mapping pipeline

As the team works with the data to create and convert it into insights, a map of how the data must be processed needs to take shape. This map is designed as the team works though the CRISP-DM and the analytics maturity of the project. It establishes the data and infrastructure requirements for the overall analytics solution. Like a car manufacturing line, data needs to go through a processing line from its raw state to reliable insights, with many quality checks, processing, and integration steps along with fault tolerance and time-to-insights considerations.

Data collection and storage

Before data work can start, the historical and current data need to be collected from the various sources, the solution architect needs work with the data scientists and domain experts to determine what raw data need to be collected, where to store it and evaluate any gaps with the current system. Often organizations will construct a new unified environment like a data lake to collect information from multiple siloed systems.

This can be a long process to identify the right, reliable data sources. Many times, it is discovered that the right data has never been collected or is hard to measure. Once the data sources are identified, how the data is integrated and stored can be a challenge. For example, trade-offs of what data resolution to store (raw vs summarized data), merging different data types and metadata, or the ability to query data in a timely way can all be difficult. Incorporating quality checks of the data to identify or correct problem data (such as out-of-range data or incomplete/missing fields) must be done in these steps, which are defined by the data scientists through their analytics maturity project work. The sooner these checks are applied, the less

work needs to be done later in the process. A benefit from this up-front investment is that the integration of new data sources in the future will be much more efficient.

This process often involves a variety of skill sets including data scientists, infrastructure engineers, solution architects, and business representatives. The team needs to think about their data's supporting connectivity infrastructure, how the data is moved, stored and queried, and the computing power for data quality checks and summarizations.

> As organizations become more sophisticated in their use of data, their analytics maturity will evolve.

Data processing

Data processing involves taking the raw data and preparing it to train a model or evaluate it through a model (called inference or scoring). Many algorithms (the math engines to train a model) require the data to be structured and formatted in order to make sense of what is being processed. This data structuring is painstaking, especially when it entails integrating raw and metadata of various types.

Further data quality checks need to be done. Depending on the data and its sensitivity, summary statistics or dimension reduction techniques, data standardizations, text stemming and other techniques should be done and automated before potential summarization and modeling of the data.

Data modelling

Data modelling has two components: training and inference (also called scoring). This is where analytics infrastructures can be taxed based on three factors: the amount of data, the model complexity, and the number of models.

Training models is the process of identifying and building models through the use of algorithms. Data scientists send training data through algorithms that output a model with tuned parameters. They then assess these models with test data to determine their accuracy, assumptions, and other evaluations and determine if these models meet the business requirements.

Once trained models pass this evaluation, they are then incorporated into business applications to deliver designed business insights. It is important to understand the amount of data needed and the sensitivity to time to train. Often, data scientists need to try several types of algorithms and then hone the best model from this work. Having an infrastructure that allows the data scientist to scale with this need is important.

Data visualization

Business insights from models can be hard to interpret in their raw forms. From lack of context to large amount of information to decipher, well designed data visualizations help the user make sense of what the data is telling them, creating the necessary context around the outputs. Each data visualization should answer a question that the data consumer is anticipated to ask, allowing them to dig deeper into why a business result happened.

A key consideration in the visualization stage is how to incorporate the results of models back into existing workflow. In some scenarios the best solution is to integrate the results of the model into an existing application. Other times it might make sense to develop an entirely new application, for example if the application needs to be highly optimized for a certain device. Whichever path is chosen, it is important to have a clear feedback loop built into the application so that users can provide input on when the model produces accurate or inaccurate results. It can also be helpful to collect more qualitative data. An example of this would be a scenario where a user chooses not to act on a recommended course of action suggested by the model, opting to go with their intuition. Capturing why they made this decision can be useful to evaluate both model accuracy and the overall user experience to drive higher utilization of the model in the future.

In summary

A team of domain experts, data scientists, data engineers, and solution architects need to work together to ensure analytics solutions are created to support successful business insights. Using several methodology components such as CRISP-DM, the analytics maturity model and data processing pipeline mapping, they can ensure the right business focus, data workloads, and data processing are developed for successful implementation of an analytics solution.

Achieve your goals with Intel® technology:

Intel® Xeon® Scalable processors: This proven, world-class and highly integrated platform delivers highly scalable performance for a wide variety of analytics, AI & other datacenter workloads—including machine and deep learning—and offers low total cost of ownership for batch deep learning inference and classic machine learning workloads.

Intel® Select Solutions: Intel® Select Solutions for Advanced Analytics in the enterprise are verified hardware and software stacks optimized for database performance and flexibility to conduct business operations much faster at the SLAs required for enterprise business.
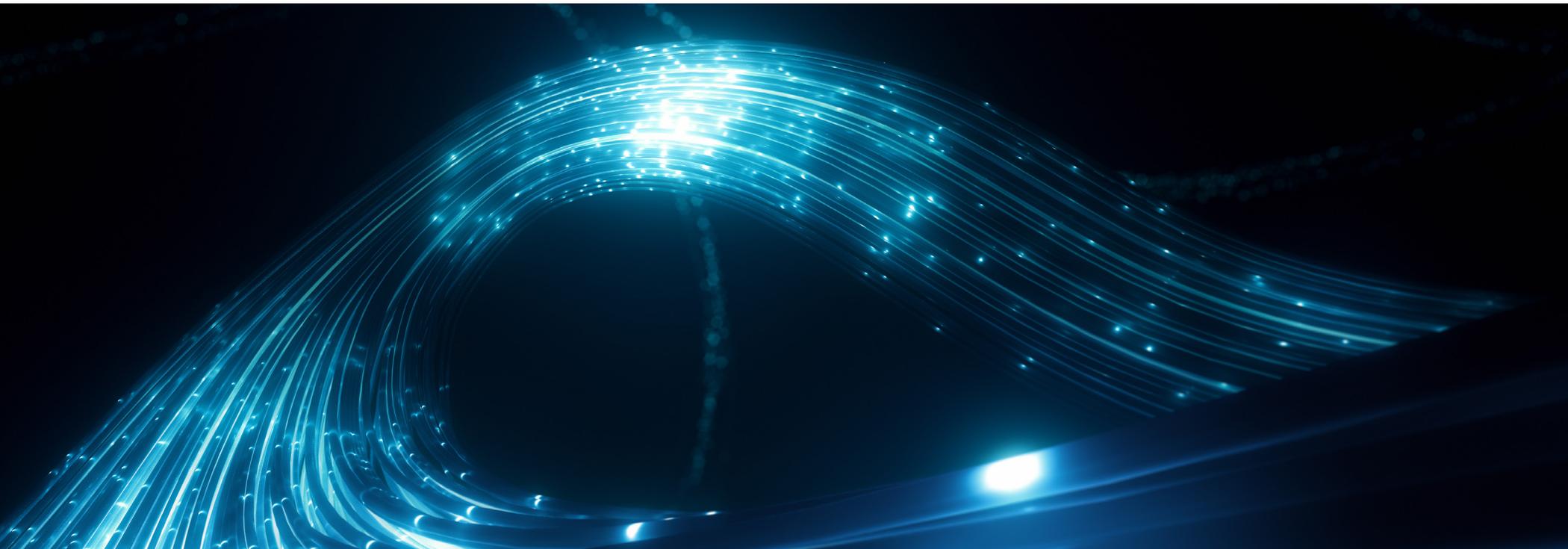
Foundations for the future

The potential for advanced analytics to improve operational, financial and business performance is well recognized, leading to a surge in new advanced analytics projects. With comprehensive planning, organizations can increase their success rate and scale as they move from proof of concepts to production solutions. It is recommended that as part of this planning, organizations conduct an analytics capabilities assessment, work strategically to build an insights-focused, cross-organizational team, and clearly define an analytics process. Doing so will contribute to investments made in advanced analytics providing the highest business return possible.

Where to get more information

- White Paper: Tame the Data Deluge
- White Paper: Five Steps to Delivering the Data-Driven Business
- Solution Brief: Maximize Data Value with a Real Time Streaming Analytics Solution
- White Paper: Magnify Business Outcome: Building a Big Data Analytics Warehouse
- eGuide: Defining a Data Strategy
- Learn more about how Intel® technologies can help you achieve your insights-driven objectives at www.intel.com/analytics

[1] *Forrester Research – Artificial Intelligence: Fact, Fiction. How Enterprises Can Crush It; What's Possible for Enterprises in 2017* https://go.forrester.com/blogs/16-11-02-artificial_intelligence_fact_fiction_how_enterprises_can_crush_it/

[2] These examples from: *Predictive Analytics: The Power to Predict who will Click, Buy, Lie, or Die*, Eric Seigel, Wiley

[3] *Crossing the Chasm*, Geoffrey A. Moore, HarperCollins

[4] Predictive Project Analytics, Will Your Project Be Successful?, Deloitte, https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/risk/ca-en-ers-predictive-project-analytics.pdf

[5] Outlined in: *The Essential Advantage, How to Win with a Capabilities-Driven Strategy*, HBR, Paul Leinwand, Cesare R. Mainardi

[6] http://business.inquirer.net/202970/strategic-hrm-building-organizational-capabilities

[7] https://spark.apache.org/

[8] *Leading Change*, John P. Kotter, Harvard Business School Press

[9] For more on the importance of business/IT alignment: *Why IT Fumbles Analytics*, Harvard Business Review, https://hbr.org/2013/01/why-it-fumbles-analytics

[10] *HBR's 10 Must-Reads On Change Management*, John Kotter, Robert D. Austin, Harvard Business Review

[11] *Building an Analytics-Driven Organization*, Accenture. https://www.accenture.com/dk-en/~/media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Industries_2/Accenture-Building-Analytics-Driven-Organization.pdf

[12] https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

[13] https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html