FROM DATA TO ACTION

The Intel guide to analytics

CONTENTS

EXECUTIVE SUMMARY

Why analytics should top every CIO's agenda

Ron Kasabian, Vice President Data Center Group, Director Big Data Solutions

Years from now, when we look back at the 2010s, it's likely we'll label it the analytics decade.

That's great news. Analytics is a transformational opportunity for all industries. It gives IT departments—like yours and ours—a unique opportunity to deliver even more business value.

IT contributes skills, tools, and insights that help look at every business activity through a new data-driven lens. Making the right investments and reaching the right decisions can increase profitability, lead to better quality products, and help identify areas for expansion.

Of course this won't all happen in a day. Those skills, tools, and insights take time and effort to build.

This special report is designed to help you accelerate that process. First, as a big enterprise, we've done our own trial-and-error work in analytics. Second, our consulting and product teams work with across a wide range of industries—finance, retail, manufacturing, agriculture, you name it. We've gotten an inside view as organizations have tried a lot of different technologies, application stacks, and approaches.

In this report you'll find peer perspective from CIOs, CTOs, and others at some smart organizations we've worked with, including AOL, Booz-Allen Hamilton, Chevron, and EMC.

We founded our Advanced Analytics practice in 2011, working from the success of a pilot project that helped increase the efficiency of a sophisticated computing grid we use to test new chip designs. The project helped us learn how to best work with business units, and how to prioritize a rapidly growing list of project requests.

Another big project for Intel involved using Big Data to do a better job identifying leads for our sales team. Ultimately, that endeavor has refined our ability to find useful signals in noisy data, and taught us more about breaking big ambitions into manageable chunks.

All together, these and other analytics projects yielded $351 million in additional revenue for Intel in 2014. This year will be even bigger. We're learning and improving, working with our partners in Human Resources to expand our analytics capabilities through smart training and hiring.

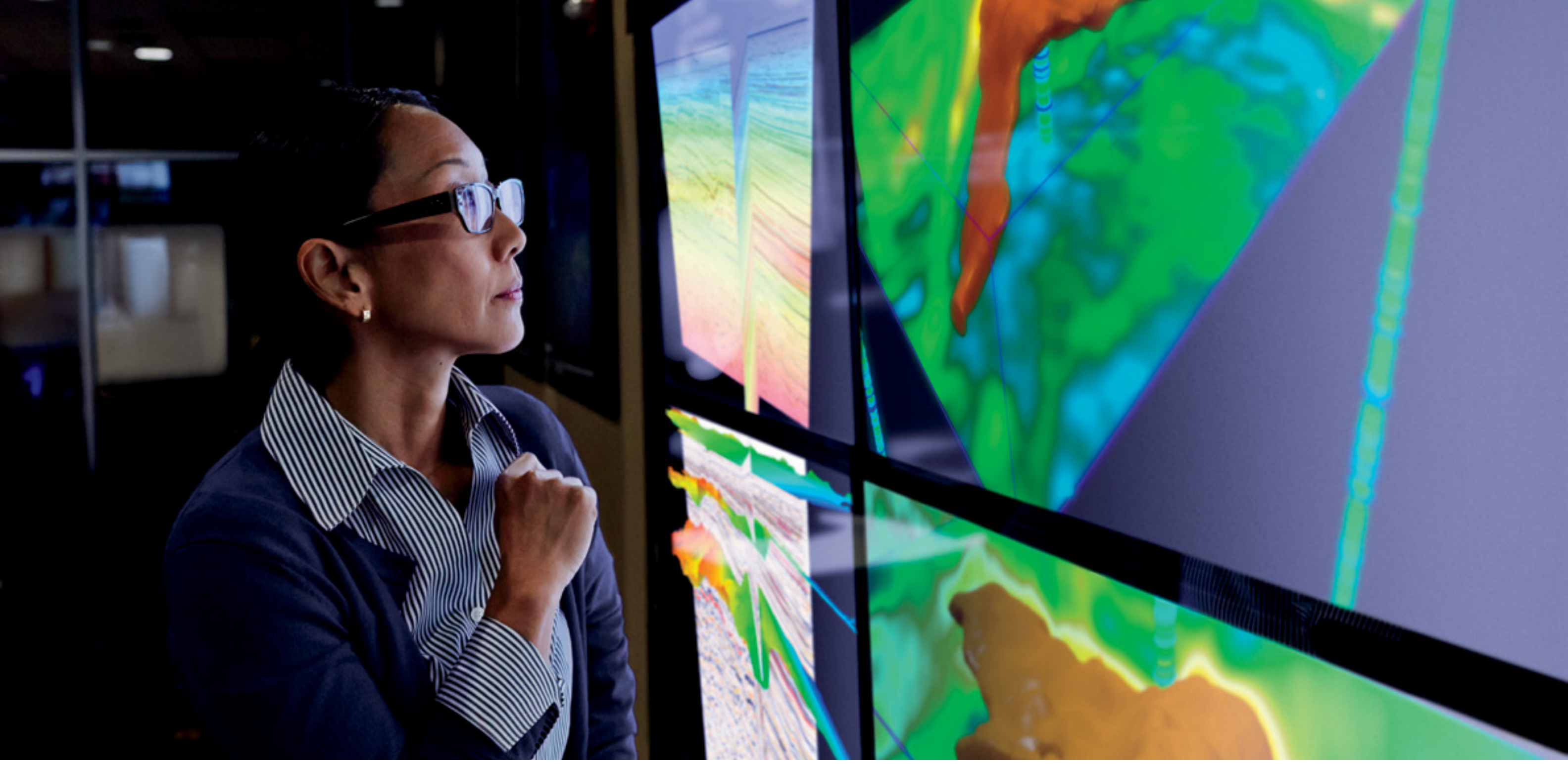
We've captured key lessons in this report, which deals with architecture and how to build analytics capabilities, but also how to build an organization that's equipped to act on the insights gleaned from analytics.

You'll also find food for thought on security, and guidance from Intel's Chief Data Scientist Bob Rogers on determining how to tell when a problem is beyond your organization's immediate capabilities.

Lastly, we've shared a pair of case studies that show how analytics is transforming industries, which we hope will inspire you to think critically about how to change your own business for the better.

The analytics revolution is here, but still in its infancy. The sensors and other systems that will collect data are new and not widely deployed. And the technologies that we will use to store and sift through data, like Hadoop, are still emerging and new to many IT pros. There's going to be lots to learn over the next several years. But without question, this is an opportunity for IT to make organizations more innovative, more efficient, and more successful. Intel is excited to be a part of it and we hope we'll have the opportunity to work with you.

Arm yourself with the insights in this report, and let's keep moving forward.



BUILDING A DATA-DRIVEN BUSINESS

How some companies are putting analytics at the heart of every decision

The amount of data you collect doesn't matter, if your organization doesn't have the skill—and the will—to use it. Here are some best practices we've learned, in working with hundreds of businesses, for building a culture that can make analytics projects truly pay off.

TAKEAWAYS

1. Goals for individuals and groups are critical for building a data-driven organization
2. Hiring for an analytical mind-set is as valuable as hiring specific skills
3. Ongoing training can go a long way, without requiring a huge budget

Like a lot of companies, energy giant Chevron sometimes struggles to find well-rounded data scientists. So last year, the company held a problem-solving contest to identify potential analysts hidden among its 64,000-plus employees.

Chevron promoted the contest to about 350 employees who had expressed an interest in analytics. It held an introductory teleconference explaining the contest, which asked entrants to predict things such as the amount of oil a reservoir would yield or the cost of oil at a certain time.

Among the winners of that contest, now an annual company event, was a woman who had a background in statistics but was working in a supplier management position. She's since become "an analytics rock star," says Margery Connor, head of Chevron's Center of Excellence for Advanced Analytics. "We might never have identified her without this contest."

Over the years, Chevron, based in San Ramon, Calif., has turned itself into a data-driven business, prioritizing analytics and operations research throughout the organization.

In April, the Institute for Operations Research and the Management Sciences, an association of analytics professionals, recognized Chevron with its annual INFORMS Prize. The organization specifically praised the way Chevron's data-focused approach helps it reduce material costs, recover resources, and operate more safely and reliably.

Businesses large and small are increasingly launching analytics programs of their own, often with similar goals. Over the past year, there's been a 125 percent rise in the number of organizations that have deployed or implemented analytics projects, according to IDG Enterprise. Many organizations surveyed said they expect analytics projects to surface strategy-defining insights into customer behavior, sales patterns, or product quality.

Delivering on that promise means finding, hiring, and developing talented data scientists. But that's just the beginning. A culture driven by data has to extend beyond a specialized group employees trained in analytics. It involves the decision-making style of the whole organization, in every function and line of business.

"Being data-driven means drawing conclusions based on the evidence, not on the opinion of the highest-paid person in the room," says Kirk Borne, principal data scientist at technology consultancy Booz Allen Hamilton. To build a culture that routinely makes data-centered commitments, he says, a company's employees must make decisions based on all available and pertinent data rather than on preconceived notions.

Detecting data-driven candidates

The analytics discipline has been around for decades, but in the context of big data it's a relatively new field. There aren't a lot of programs turning out job-ready candidates, and midcareer professionals don't always come with the necessary technical training.

"Sourcing candidates is really much more challenging," says Alexis Fink, talent intelligence and analytics leader at Intel. "You can't just source by a university."

The good news is that you don't have to limit yourself to finding fully-trained candidates. And there are likely scores of individuals scattered across your organization who, with good training and the right tools, can help transform your company into one driven by data.

Fink says strong analytics candidates need more than math and programming skills to succeed, however. Good ones approach data with an open mind, Fink adds.

> "Being data-driven means drawing conclusions based on evidence."
>
> — Kirk Borne

"Most people use data the way a drunk uses a light post," Fink says, "for support rather than illumination." Rather than selectively seek data to uphold an existing belief, a data-driven professional will start with a question and use scientific methods to find data that reveal the right answer.

To help determine whether someone embraces this approach, Fink advocates "a portfolio review model to assess people's data analysis problem-solving skills." By walking through a project in which the candidate changed

Chevron wins INFORMS Prize of excellence in analytics and operations research.

course, for example, she can evaluate his approach in using technology to solve challenges.

"You can gauge their judgment" and fit for a specific position, Fink says. "For example, did they use statistical analysis to understand whether a pattern was forming that justified the change?" Further reviewing candidates' work samples helps her "determine the quality of the code they are writing and the appropriateness of the statistical analysis that they run."

Borne suggests looking for people who espouse a penchant for lifelong learning and have demonstrated an ability to accept change. These qualities tend to align well with data-fueled projects.

At Booz-Allen Hamilton, he says, "We have internal training for employees who want to become data scientists or learn data analytics skills. The course runs an hour a week for nearly a year; to qualify, employees must pass a math and programming test."

Building a data-driven culture

A data-driven culture is one that rewards data collectors across the organization. It's led by executives who want to know what the data suggest, who develop a decision-making structure that includes data analysis, and who base plans on that analysis.

It might sound like a big effort to get there, but to a large extent, it's just a formalization of common behaviors.

For example, Chevron is now rolling out a training program that uses real data and actual problems facing line-of-business employees.

"Most people see a lot of data in their daily jobs," she says. "That's the data we want them to work with" in the training program.

Chevron pairs training participants with analysts who can use algorithms to come up with answers. This practical approach to training sharpens employees' ability to frame business challenges and identify how the data can help solve them.

Over time, a company can formalize this process further.

Among other data-focused policies, Chevron has implemented a companywide mandate that every project proposal above a certain dollar threshold include specific types of analysis.

"You have to prove that you included ranges of uncertainties in the data or the economic assumptions," says Amy Absher, general manager of strategy planning, service, and control at Chevron. "Funding is dependent on this insight."

> "Most people see a lot of data in their daily jobs."
>
> — Margery Connor

For large projects at Chevron, Connor says, "An independent group comprising representatives from across the company comes in and looks at the decisions we're making, the alternatives we considered, and the uncertainties to determine whether we did a broad assessment prior to moving forward."

Chevron also performs a formal project review to compare predictions with results.

"The fact that senior management expects that analytical work to be done and asks these questions sets the expectation and the tone throughout Chevron," Absher says.

Most businesses today are gradually building a data-centric culture, Borne notes. To those just beginning, he advises simply finding an analytics project with a good chance of payoff.

A financial-services firm Borne consulted, for example, decided to search its Web analytics logs for an indication that a customer was thinking about defecting to a competitor.

"They found a signal in their data," Borne recalls, involving "a lot of price comparisons. They applied a very soft 'customer services' response, inviting people to look at some of the company's new products or offering to answer questions." After a quarter, "they estimated they saved the company more than $100 million in potential lost customers."

A smaller company can experience the same impact with smaller-scale returns. And as analytics projects deliver value, the culture and processes to support more data work can be built step by step.

SENSING URBAN CHANGE

Cities turn to the Internet of Things

As urban populations grow, more cities are relying on emerging analytics and connected devices to keep up. Intel works closely with dozens of major metropolitan governments worldwide; here are IoT lessons from some of the leaders in the Smart City movement.

TAKEAWAYS

1 London and Dublin are deploying networked sensors as part of an effort to make themselves more livable and better prepared for the future

2 These cities are some of the first places to test how to efficiently deploy IoT systems city-wide

3 Smart cities projects point the way for understanding sensor applications and the power of IoT, as well as the need for data accuracy, organizational cooperation, and more

Four days a week, Duncan Wilson rides his white Ducati 899 motorcycle through 8 miles of bumper-to-bonnet London traffic to his office at Imperial College.

As he passes Hyde Park, a sensor mounted on a utility box just inside Victoria Gate detects levels of nitrogen oxides, sulfur oxides, and particulate matter. It's one of nearly 80 such devices recently deployed throughout London to help the city identify its most polluted areas, or black spots, and better grasp how to combat smog.

"If we think about the black spots, they're all based around major intersections and traffic routes," says Wilson, an Intel research director leading the project, called Sensing London. "We're monitoring parks to make the case for preserving this green space."

Sensing London, a collaboration between the Intel Collaborative Research Institute for Sustainable Connected Cities, Imperial College, University College, the Future Cities Catapult, and members of London city council, is one of many efforts around the globe using the Internet of Things to deal with issues like climate change and stretched resources. These projects range from grassroots efforts to huge undertakings between government, corporate, academic, and civic groups.

In Barcelona, sensors in waste bins alert trash collection services when they're full. At the Port of San Diego, engineers have deployed sensors in an HVAC system to help reduce energy use and prepare for tightening state regulations.

The common thread among smart-city projects is the principle that data—and insights from data—can lead to better ideas, decisions, and results.

"There's only one starting point, and that's the faith that knowledge is a good thing, and it can help solve problems," says Alan Sitkin, a councilor for London's Enfield borough, part of the Sensing London project. Information can "raise people's awareness, again in the hope that it will begin to change behavior."

London air

London, which sprawls across 607 square miles, has nearly 2.5 million cars and trucks on its roads. More than 30 percent of these vehicles are fueled by diesel, which releases far more nitrogen dioxide and particulate matter than vehicles running on unleaded gasoline. These pollutants, some of which were recently measured at higher levels in London than in Beijing, were linked to 9,500 premature deaths in London in 2010.

Monitoring London's air in its entirety is, for the time being, impossible. So Sensing London decided to focus on three strategic locations in addition to Hyde Park: Tower Bridge, where cars idle for several minutes three times a day, as the bridge is raised for ships to pass; Elephant and Castle, where researchers are studying a nitrogen oxides-absorbing paint; and the Northern borough of Enfield, which is sandwiched between two old and overwhelmed highways.

For the Enfield portion, "In one month, we had a program set up," Wilson says.

The sensors capture data that an on-site system-on-a-chip gateway processes in real time. The gateway then sends the data to the cloud, which provides flexible and scalable processing infrastructure for applications that transform numbers into meaningful, actionable information.

It's not without complications, however. The placement of the sensor and its casing, temperature, humidity, and wind are among many things that can lead to inaccurate data.

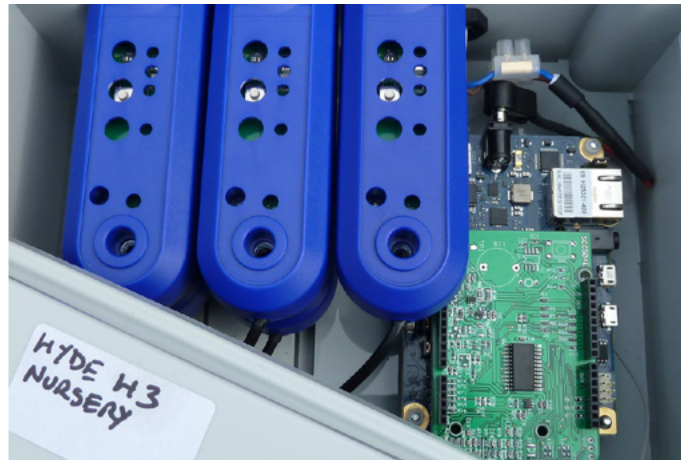
"In one month, we had a program set up."

— Duncan Wilson

To help compensate, the ICRI team calibrated its sensors with London's three high-fidelity air quality stations. Algorithms added to those gateways helped align the numbers.

"We've been learning a lot about the performance of electrochemical sensors themselves, and we've been updating the algorithms used to process the data," Wilson says. "While one approach to IoT is to send data right to the cloud, we're also investigating at-the-edge processing, where we send the transformed data to the cloud."

The advantage of cutting the noise from the captured data, Wilson says, is that the cloud doesn't get filled with



Interior of Air Pollution monitoring box

meaningless data. It's filtered out before it leaves the gateway processor.

Dublin flooding

Almost 300 miles northwest of London, Intel last April signed an agreement with the Dublin City Council to create a citywide network of gateway sensors and get a similar IoT collaboration across industry, academia, and government up and running.

"We initially spent a lot of time internally with operational staff, asking, 'what are the priority areas that we could look at?'" recalls Jamie Cudden, Dublin City Council's Smart Cities coordinator. "We came down to the issue of flooding."

The warm North Atlantic Current has always kept Ireland temperate. One Victorian-era poet described the rain there as "warm as an Irish welcome, and soft as an Irish smile."

Climate change means the Irish air now holds approximately 4 percent more water than it did in 1890. This has contributed to "monster rains" that wreak havoc across Eastern Ireland.

"Pluvial flooding can turn streets into rivers," according to Dr. David Prendergast, Intel anthropologist and project lead for the Dublin IoT Demonstrator. It can also "cause the overtopping of rivers, the overtaxing of drain systems, and the flooding of basement flats."

Between December 2012 and January 2013 alone, flooding cost Dublin 61 million euros, or $85 million, in damage, according to the country's Directorate for Fire and Emergency Management.

Through its Smart Cities Program, Dublin is working to batten down the hatches.

"We found that a lot of [city] engineers were already very proactive in innovating—finding and trialing new solutions, in terms of rain and river level-monitoring sensors," Cudden says. Yet these efforts were siloed. "The guys managing the rivers and the guys managing the drains," he says, had very limited means of collaborating.

Over the next year, Intel will help Dublin build an IoT system

Flooding in Dublin forces traffic on the N2 at Glasnevin Cemetery onto the opposite side of the road. Credit: Shane Lewis

that gathers rain, river, and drain data from as many sensors as possible.

Dr. Prendergast says that in strategic locations around the city, Intel will deploy rain gauges and weather stations, as well as experimental sensors such as river-monitoring buoys and low-power ultrasonic water level sensors, to measure tidal surges or provide real-time information about rivers and streams that are sensitive to monster rains.

"All of these data streams are no good, unless you can respond effectively."

— Jamie Cudden

Experts ranging from engineers to data managers are partnering with local colleges and businesses to build sensing equipment, synthesize gathered data, and ultimately develop and implement flood-monitoring plans.

"All these data streams are no good, unless you can respond effectively," says Cudden, who envisions a sensor-based system that will alert workers to clear sewer drains, move cars, and notify basement apartment dwellers before a big storm.

The future of smart cities

While large-scale initiatives in Dublin and London are making strides in building an IoT infrastructure, Anthony Townsend, senior research fellow at New York University's Rudin Center for Transportation Policy and Management and author of Smart Cities, sees top-down advancements as just the beginning.

He anticipates urban infrastructure eventually supporting citizen-led projects. He imagines access to government representation, education, and child care services being "rewritten by these smart little devices that are in our pocket and in the walls," Townsend says.

Why not? As smartphones become ubiquitous, the power of edge processing is in people's hands.

"It's getting more complicated, and that's a good thing. It's not just all about making the world easier," he says. "What about more sociable? More fun?"

INTRACTABLE PROBLEMS

By Bob Rogers

How to get value from almost any analytics project

There are ways to get meaningful results even when a problem seems like it can't be solved. Bob Rogers, Intel's chief data scientist, explains how.

TAKEAWAYS

1 Data scientists regularly run up against problems that are (or seem) unsolvable by big data.

2 Asking questions correctly and choosing the best algorithms for your problems are critical for the success of analytics solutions.

3 Troubleshooting techniques can help analysts get a meaningful answer, even if it's for a somewhat different question.

I spent more than a decade forecasting futures as the manager of a hedge fund. We had tick-by-tick data going back decades, but there was a huge random component to this data that made automated prediction beyond a certain accuracy impossible. All the motives people have for buying and selling at a particular moment, combined with the sheer number of people trading, meant that no matter what we did, we'd never perfectly pluck signals from the noise.

In data science, we call these intractable problems, and, past a certain point, analytics and big data may simply never make progress.

The good news is that many problems that at first seem intractable can be addressed by tweaking your approach or your inputs.

Knowing when problems that seem intractable can be solved with some affordable changes will position a business—and a project sponsor—for ongoing success. Conversely, being able to recognize problems that are defined at an unrealistic scale will prevent squandering time and money that you could profitably apply to a more focused question.

Here are four troubleshooting methods that could improve your results. By iteratively applying one or more of them, you could exchange banging your head against a wall for increasing the chances of finding value in your analytics work.

1. Ask a more focused question.

Often, the best way forward is to try to solve for a subpart of your original question and extrapolate lessons. Trying to determine the likelihood that any given social-media user will be interested in a car model you're designing is likely intractable. Even with lots of good data, you might have too many variables to arrive at a model with real predictive value.

But you might be able to predict an increase or decrease in sales to a specific demographic. From there, you could determine whether a change such as a boxier design would boost sales to soccer moms more than they would hurt sales to single twentysomethings. That's a more manageable problem scope that still delivers real value to your business.

The same approach can help you isolate variables that are throwing off your algorithm. Instead of trying to predict hospital readmission rates for all patients, for example, you might divide a patient set into two groups—perhaps one of patients with multiple significant conditions and the other of patients with only a single condition, such as heart failure.

If the quality of prediction in one group shows a meaningful improvement or decline, that would indicate that your algorithm works for a data set that is not just smaller, but specifically clear of a particular confounding variable present in the larger pool.

2. Improve your algorithm.

In data science, algorithms not only define the sequence of operations that your analytics system will perform against the data set, they also reflect how you think about, or "model", potential relationships within the data.

Sometimes creating the right algorithm, or modifying an available algorithm for your specific new purpose, requires many iterations. (Machine learning offers promise for automating the improvement of algorithms; that's a discipline to watch.)

One sign that your algorithm isn't working is if you've scaled your compute power by, say, a factor of five, but are seeing a much smaller improvement in processing time.

> Sometimes when you add a new set of data, skies open up, and you find new predictive power.

Another test is to slightly tweak your algorithm parameters. Slightly different algorithms should produce only slightly different answers. If they produce drastically different answers, chances are that something is off, and you need a different algorithm.

And perhaps you've chosen the wrong type of algorithm altogether. Model selection is often based on assumptions

about the data, such as expecting a linear progression between two elements when their relationship could be more accurately represented by a decision tree.

There are many libraries of publicly available, open-source algorithms. You rarely have to start from scratch.

3. Clean up your data.

This is an age-old challenge for IT. Garbage in, garbage out. Ideally, this is something you will have tackled prior to starting any analytics project, but problems with data sets often aren't clear until you begin your analysis.

4. Use different data.

This is a slightly trickier variation of the previous step. To get more data, you might just need to update your metadata. You might need to change some processes to capture the data you need.

Most businesses have already squeezed as much value as possible out of the data they store in traditional data warehouses. Sometimes when you add a new set of data—especially unstructured data, such as text progress notes written by doctors or documented interactions between call center employees and customers—skies open up, and you find new predictive power.

As a general rule, more data should help produce better answers. As you test an analytics project, add data in sequence to see how they change the answers. So long as your answers keep getting better, you most likely haven't hit the point of intractability.

When your progress slows, take stock of the cost of possible approaches versus the potential payoff. And it doesn't hurt to keep this in mind: Trying to predicting human behavior too accurately might be the root of all intractability.

THE PRIORITIZATION PROBLEM

How to identify the best analytics projects to pursue

Once your early analytics work starts to show business value, you may find yourself with more project proposals than you can reasonably take on. You'll need a project prioritization process. Here's how Intel does it.

TAKEAWAYS

1 Analytics projects can be prioritized based on measurable business value

2 ITDMs can start with smaller projects and work against common-sense goals

3 Each successful project builds capabilities that can be re-used to make subsequent work more efficient

In 2011, a small team within Intel's IT department started looking for ways to more accurately predict the computing demand required to test chip designs.

By knowing when certain resource-intensive steps would take place and how much memory each test would require, Intel could make more efficient use of its large computing server grid, scheduling more concurrent tests without running out of memory or processing power.

The stakes were high: For every 1 percent of efficiency gained, Intel would save $1 million a year. The team of five spent six months applying advanced analytic techniques to data from test simulations, finding design process improvements worth more than $10 million a year.

The team, now known as the Advanced Analytics group, has been swamped with project requests ever since—even as the group has grown to approximately 100 members. To prioritize proposed projects, it analyzes the strength of business unit support and collaboration, the quantity of high-quality data, and the size and extent of the potential impact, among other factors.

Today, projects the team deems worth pursuing could last a year and yield returns much greater than those of the chip design simulation project.

Demanding high returns on projects "makes very clear that this is the order of magnitude of value that we are expecting," says Moty Fania, Intel IT's principal engineer for big-data analytics.

The team's work still includes plenty of exploration and experimentation, though. "Almost always you start looking at the data with a hypothesis or assumption in mind," he says. "But sometimes you find a different signal in the data that you weren't expecting, and that becomes the important thing."

High demand, high stakes

In 2014, Intel's inside-sales team made the Advanced Analytics group aware of a big problem: It had a list of 10,000 customer prospects for its representatives to contact but little insight into which ones to address first. Some might lead to significant new accounts; others might need more handholding. There was an opportunity to use data to identify how the sales team should apply its limited resources to generate the highest return.

Here are the steps the analytics team followed to assess the project:

1. Executive sponsorship. Projects need the backing of the relevant business leaders. In addition to justifying the investment the project requires, the analytics data often indicate that a meaningful business process change is required—which someone needs to be willing to implement.

Project won't succeed "if we don't have a strong endorsement from upper management and also from the people at the end of the day who are doing the work itself," says Chen Admati, an Intel IT Advanced Analytics manager.

The leaders of the inside-sales organization were on board.

2. The right problem. The question the analytics team is seeking to answer must matter to the business. Addressing this problem also requires understanding how any solution fits into the organization's existing business processes, systems, and staffing arrangements.

In this case, the problem aligned perfectly with the organization's mandate to increase sales. Top managers pledged to carry out business process changes, if necessary.

3. Data. The analytics team has to determine if there's enough high-quality data to make an analytics project feasible and worthwhile.

The team carried out a pilot project for one geographical area of the inside-sales organization that managers knew had a high-quality data set. Successful tests with this data showed that the analytics project held promise, which justified work to upgrade the data quality in other areas.

4. Resources. The analytics team, working with the sponsor organization, has to evaluate what individuals, skills, tools, and processing power are needed for the project to succeed, and whether they're available.

The analytics group had a team in place working with the inside-sales function. And sales leaders made people available to the analytics team.

5. Time. The analytics team should assess whether it can achieve results for the project within a desired time frame. You want to quickly demonstrate value to the organization.

For the inside-sales example, a pilot project yielded strong results in short order. This justified more work on a broader data set.

Chen Admati, Advanced Analytics manager

6. Projected benefits. The analytics group evaluates the degree to which the project can benefit the business unit and the company at large. "It's not one-size-fits-all," Admati says. "It's about the specifics based on the business you're supporting." Calculate expected return on investment, whether quantified in dollars or resolution of a strategic problem.

The inside-sales organization and analytics group projected that they could significantly increase revenues.

Record the results, and remain engaged

During meetings, Admati says, sales managers and analytics experts devised a plan. A machine-learning algorithm, they predicted, could help them narrow in on the best prospects and cater their chip sales messages to those businesses' needs, ultimately adding to Intel's bottom line.

The Advanced Analytics group developed an algorithm that identified the 1,000 best prospects out of a pool of 10,000 resellers.

It also helped the team develop specific topics for sales representatives to discuss, individualized for each prospect based on its needs. One might be quite receptive to chips for servers, while another might need networking hardware.

Armed with prioritized lists of customers to call and individualized talking points for each call, the online-sales team increased its 2014 sales by more than $76 million. (Intel's finance department tracks and validates results attributed to analytics projects, which provides accountability and improves the accuracy of ROI projections for future projects.)

Although the project is technically complete, Admati says the Advanced Analytics group regularly checks in with the online-sales team.

"In many cases, the conversation starts in one place, and we end up tweaking it, and we end up with a different solution," Admati adds. "You're learning about the business unit you're supporting, and how decisions are made, and how the people are needing to be influenced. Your project prioritizations will change based on that."

PRIORITIZATION 1-2-3

A guide to comparing analytics proposals

Intel prioritizes projects based in part on quality of data and expected return. Here's a simple framework to guide your decisions. For each question, give your project 1, 2 or 3 points based on the accompanying statement. You can weight the criteria that matter most to your company. Add up the total. Make the highest-scoring project your top priority.

PROJECT NAME: ______________________ BUSINESS SPONSOR: ______________________

CRITERIA	1	2	3	SCORE
Will sponsor manage organizational changes the project deems worthwhile?	Sponsor lacks authority to make process changes	Yes, but change is likely to be big	Yes, and change is minimal or welcome	
How closely aligned is the project to the business mission and goals?	Project is tangential	Critical to the business unit	Critical to the overall business	
How simple is this project to execute?	Significant difficulty	Medium difficulty	Straightforward	
Do we already have the necessary tools in place?	Requires significant tool updates	Most, not all, tools in place	All tools in place	
How much work is required to capture and prepare data?	Significant data preparation effort	Moderate data preparation effort	Project uses well-formatted or pre-integrated data	
How fast can we begin to deliver business benefit?	10 or more months	7 to 9 months	0 to 6 months	
How will the outcome improve our capability to perform future analytics projects?	Will not build significant reusable capabilities	Modest applicability across other projects	Builds significant analytics capabilities	
What is the expected business value of this project in dollars?	Low (define range)	Medium (define range)	High (define range)	

TOTAL SCORE: ______________________

LESSONS FROM THE FIELD

How precision agriculture is pioneering analytics

Farms are using analytics to solve unlikely problems. Intel is part of a broad network of researchers and developers working with tools such as sensors and drones to help them. These emerging agricultural applications may trigger some new thinking about one of your business problems.

TAKEAWAYS

1 Farming is one of the first industries to be totally transformed by data.

2 Hardware and analytics advances together are allowing farmers to adopt more ecologically friendly practices and produce more flavorful food.

3 Farmers are realizing the sorts of goals that experts foresee reaching in almost every other industry.

When Nathan Stein hears the term "Internet of Things," he thinks of corn and soybeans. These plants' second-by-second adaptation to weather and soil conditions produces a nonstop stream of data that help him better run his family's Iowa farm.

Using analytics software developed for farmers, he can simulate the impact of water, fertilizer, and pesticide adjustments.

"I can basically virtualize the entire crop," Stein says.

Stein is among a growing number of farmers using real-time data collection and computer-based analysis. Thanks to farmers like Stein—as well as researchers and companies developing technology for them—agriculture, the oldest of human industries, is becoming a prime testing ground for sensors, drones, and big-data analytics.

These methods are helping farmers increase yields, margins, and efficiencies on a massive scale—goals of every industry.

Something that works "in the context of large-scale farms could allow for that application into other domains," says Vin Sharma, director of strategy, product, and marketing for Big Data Solutions at Intel.

For example, a retailer could use a single-function foot traffic sensor to replace video analytics in measuring and improving the effectiveness of in-store displays. A fulfillment center manager could embed a sensor on a general-purpose drone to check inventory. And across many other industries, CIOs could implement sensor-derived data analytics to precisely control corporate resources ranging from raw materials to computing power. Targeted control promises efficiencies not only within the company, but potentially all along the supply chain.

"The data collection, the model development and deployment, the analytic capabilities—we see that all of those

Daniel Robinson of the AggieAir team adjusts the Minion 2 aircraft in Cache Junction, Utah. Credit: AggieAir, Utah State University

activities are very similar across industries," says Sharma. In response, Intel has built a Trusted Analytics Platform cloud service for building analytics applications, and also open-sourced its components and the glue that makes them work together.

"Companies don't have to reinvent the wheel every time; they can re-use common elements and then stitch their own domain expertise" to create industry-specific commercial or proprietary applications, he says.

Back on the farm, innovators of precision agriculture, as the field is known, are working with data analytics to resolve farming's biggest challenges.

> "I can basically virtualize the entire crop."
>
> — Nathan Stein

The Utah Water Research Laboratory at Utah State University, for example, is testing a drone called the AggieAir that director Mac McKee developed with funding from the U.S. Department of Agriculture.

A one-hour flight with a typical payload of three cameras—RGB, near-infrared, red-edge, or thermal-infrared—produces about 200 gigabytes of image files that, combined with proprietary software, could help a winemaker estimate the amount of water each grapevine on his vineyard needs, making watering more efficient and saving them money. McKee says flights of the next-generation AggieAir will produce about a terabyte of data.

And at the University of California at Davis, where Intel has funded research, Shrinivasa Upadhyaya is developing an in-field leaf monitor that uses a thermal-infrared sensor to detect a plant's transpirational cooling.

The sensors, which account for environmental factors like ambient temperature, relative humidity, radiation, and wind speed, send data in real time to desktop computers and mobile devices, which in turn analyze the data using software Upadhaya and his students developed to determine which areas of a field need more or less water at a given time. Such site-specific data can help farmers use just the right amount of water needed, illustrating analytics' ability to allocate resources and reduce waste.

"The availability of increasingly less expensive and more powerful sensor and analytics technologies is helping farmers to watch over their terrain more effectively," says Chris Seifert, director of data science at San Francisco startup Granular, which makes cloud-based software for managing farms. "Farmers can have a much more thorough understanding of what's happening on their land without having to go out and visit each acre each day."

Granular uses APIs to upload sensor data from things like irrigation systems, tractors, and farm implements to Amazon Web Services. The biological systems of densely planted fields of corn, Seifert notes, "represent some of the ultimate things to connect and to be able to monitor remotely."

Intel's Sharma points out that this volume and variety of data types is "several orders of magnitude greater" than in traditional farming, when data collection meant a thumb in the soil. That's why advanced analytics are necessary to put the 'precision' in precision agriculture.

"The goal is essentially yield management. You want to figure out which variables are most responsible for increasing the output of the farm," he says. "Human beings aren't going to sit and build a statistical model with a thousand dimensions. You're going to have to use machine learning to do that effectively."

Since 2010, Iowa farmer Stein has used aerial imagery from satellites and planes to detect information such as elevation, temperature, soil moisture, and chlorophyll levels. He exports images and data to mapping and analysis software from senseFly—he works for the Swiss company as liaison between the company's customers and engineers—to identify unhealthy areas of his crops.

"We anticipate that the data center and the edge devices are going to evolve together."

— Vin Sharma

One thing Stein has observed through the process of collecting and analyzing data is the extent to which conditions on his farm can change throughout a day. As the sun's angle changes, and heat builds up in the ground, "You see a shift in thermal data, and you see the transpiration of the plants pick up and drop off," he says.

The data Stein derived from the overhead imagery of his family's farm "quickly showed us in the spring just how much damage not [installing] more drainage...was costing our corn field"—almost 40 bushels an acre.

"This fact alone triggered us to spend thousands of dollars to put in a new main and laterals, to adequately drain water-logged soils," says Stein.

Soon, he plans to soon use the senseFly drones and software to optimize fertilizer distribution on his farm. Using senseFly's software and a post-flight drone map, he could program a self-driving tractor to distribute a prescribed amount of fertilizer throughout the field.

Minion 2 aircraft flies over a field. Credit: AggieAir, Utah State University

Devices such as smart drones and autonomous tractors raise the question of where the intelligence lies and where data processing will happen: in equipment at the edge of the farm, or in a cloud-based data center? Sharma says the answer is both.

"There is a somewhat specious either/or argument in some parts of the industry," he says. "We anticipate that the data center and the edge devices are going to evolve together."

Sharma gives the human nervous system as an apt metaphor. You want enough reflexive intelligence at the edge to pull your hand off a hot stove without having to "think" about it. But central intelligence of the brain can help improve or override actions to create higher-level value. Future farms will pair smart semi-autonomous devices with cloud-based central command system that benefits from analyzing data across many locations.

Stein echoes that point. "Agriculture data's very timely, and it has to be captured at a very precise moment, and it has to work every time," he says. On today's farm, he adds, a farmer is a "connoisseur of data."

ARCHITECTING FOR ANALYTICS

Before building an analytics system, IT departments must consider these key issues

To get the payoff from Big Data, users have to make a lot of decisions. Intel has seen many approaches to building an analytics 'stack' and their architectural implications. Here are some factors that can make your project a success.

TAKEAWAYS

1. Storing data close to processing can save time and transmission cost
2. Real-time analysis creates a different set of demands that require different tools
3. Access controls should be matched to the sensitivity of the data involved

When Intel's IT department moved an advanced data query from a traditional database to an in-memory database, it cut the processing time from four hours down to 10 minutes.

It's an example, Aziz Safa, vice president of information technology at Intel, says, of how, with analytics, the right architectural decisions can make a big difference.

There's no one-size-fits-all solution, of course. Instead, IT executives need to make a series of decisions—each with its own advantages, tradeoffs, and architectural implications. Chief among them: Where to store and process data; what kinds of databases to use; and how to make sure only the right people access the data.

Luckily, some IT leaders are already well down this path. Here's some of their guidance.

Where to store and process data

IT executives need to decide how far data should travel before it's streamlined and analyzed. The two most practical choices each have strengths and drawbacks.

IT executives need to decide how far data should travel before it's streamlined and analyzed. The two most practical choices each have strengths and drawbacks.

On the other hand, having to sift through raw data can slow down analysis, and data lakes inevitably store data that ultimately isn't needed.

For Patricia Florissi, global chief technology officer for sales and distinguished engineer at EMC, the pros outweigh the cons.

"You should be able to do analytics without moving the data," she says.

In its data lake solutions, EMC stores raw data from different sources in multiple formats. The approach means that analysts have access to more information and can discover things that might get lost if data was cleaned first or some was thrown away.

Florissi adds that big analytics efforts might require multiple data lakes.

Media conglomerate AOL also uses data lakes, says James LaPlaine, the company's chief information officer. The company engages in billions of transactions per day, and "the time it takes to copy huge data sets is a problem," he says. Leaving data in native formats and moving it from the point of capture directly to public cloud avoids the cost of copying it over the internal network.

However, for certain types of data, AOL prioritizes having a single, "gold standard" data set. For analytics projects in which accuracy and consistency are paramount, such as ones involving transactional data, AOL has aggressively moved data from localized data warehouses or data lakes to a centralized data set that is cleaned, standardized, and normalized before going to public cloud storage.

"We want all of our rich data in one place so we can have a single source of truth across the company," says Mike Bojdak, the senior technology director at AOL, which is now a Verizon Communications subsidiary. This approach applies particularly to analytics and reporting on corporate performance.

"As we develop this architecture, we want to foster data exploration in new and interesting ways," Bojdak says.

What kind of database to use

It's important to choose the right database for an analytics project, with factors like data quantity, formatting and latency all playing a role.

The project where Intel switched databases involved an advanced query "using data from a bunch of noncorrelated sources," Safa says. Running on an SQL database, the query took four hours. On an in-memory database, the same query took 10 minutes. But he notes that doesn't make in-memory the right choice for every application. It always comes back to the business goals for the task at hand.

> "We want all of our rich data in one place so we can have a single source of truth across the company."
>
> — Mike Bojdak

As a starting point, Safa says, consider whether a project is looking for patterns or requires pinpoint accuracy.

Distributed databases like Hadoop that store data in different formats work well for projects focused on finding trends, he says. In these cases, a few inaccurate data points won't materially change the result.

On the other hand, he says, "If you are trying to determine where specific materials are at a given moment in your manufacturing process, then you need 100 percent accuracy with no latency."

That requires a database with more structure or controls, tuned for real-time results. Depending on its specific needs, a company might choose an in-memory data-processing framework or a performance-focused NoSQL database. Although many analytics database types have overlapping capabilities, their features are materially different.

Database choices also have to balance cost against business demands, Safa says. It might sound ideal to preprocess, structure, and clean all your data, "then put it on a disk sitting right next to the user" in a high-performance database, he says. But that's the most costly approach and no IT department has unlimited resources.

"So the question is," Safa says, "What is the value you will get from that data?"

In considering these trade-offs for each project, EMC has embraced a best-of-breed approach, picking the right database for each project rather than focusing on standardization. It's betting that the payoff across all projects will outweigh the costs and complications of using multiple database types.

Florissi doesn't think it makes sense to limit the number of databases a company uses, in part because there are so many emerging use cases for analytics that it's hard to predict what kind of data might prove useful in the future.

"A couple of years ago, companies couldn't care less what

people were Tweeting, or about information on blogs or social media," Florissi says. "Today, entire organizations are looking at those data sources to identify analytics opportunities."

AOL agrees that every company should figure out the most effective database type it could use for each business need, Bojdak says. "If it's real-time data capture, then it's in-memory. If it's a more traditional need, then OLAP [online analytical processing] works."

> "Data classification is...labor-intensive, but it's an important thing to get right."
> — James LaPlaine

However, he's trying to standardize the company on just a few database technologies, to help keep the company's IT management needs and costs in check.

How to control access

In securing big data, IT departments face a familiar trade-off between preventing inappropriate access and providing adequate access.

Brian Hopkins, vice president and principal analyst at Forrester Research, recommends controlling access via standard perimeter authentication and authorization mechanisms, such as passwords or multi-factor authentication. But companies should also encrypt data and restrict the sharing of data via tokenization, he says.

Other ways to keep data secure are to keep in place the access privileges from the system the data came from, and to restrict access to data that's been analyzed to the person or team that performed the analysis.

Although AOL is aiming to put all of its rich data in a centralized cloud, it has access controls in place at multiple levels.

An analyst manually reviews data and sets a level of access based on its sensitivity; an authentication system ensures that only people granted that level of access can view the data.

AOL constantly reviews data to make sure it has the correct access classifications for the authentication system, LaPlaine says. "Data classification is a manual process," says LaPlaine. "It's labor-intensive, but it's an important thing to get right," he says.

"We're trying to balance meeting the needs of analysts and making sure the data is completely secure," adds Bojdak.

10 QUESTIONS TO ASK YOURSELF ABOUT SECURING BIG DATA

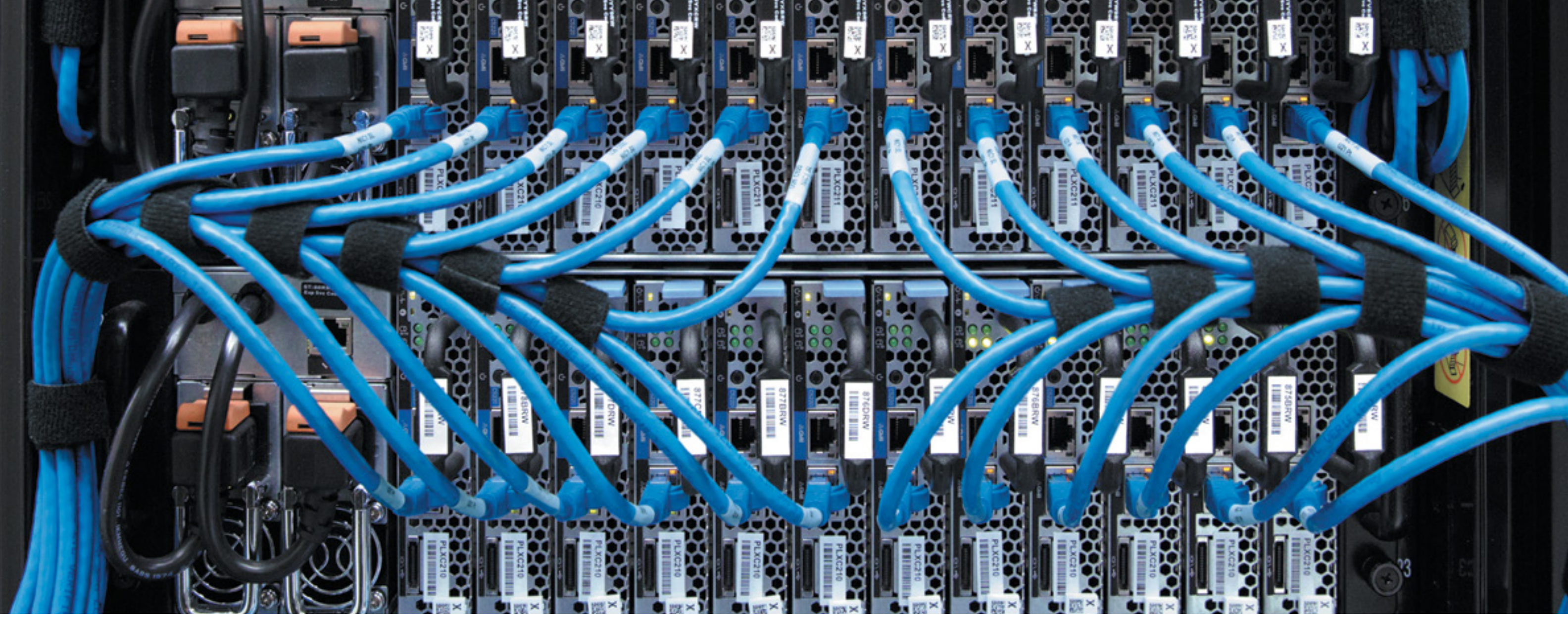
Big data introduces new wrinkles for managing data volume, workloads, and tools. Securing increasingly large amounts of data begins with a good governance model across the information life cycle. From there, you may need specific controls to address various vulnerabilities. Here are a set of questions, compiled with input from experts across Intel's product and internal security groups, to help ensure that you have everything covered.

TAKEAWAYS

1 Security concerns should be addressed at the outset of data collection, rather than later.

2 A solid information governance model is the foundational element of a strong security program.

3 Critical big data security controls include encryption, authentication, and access control.



1. What is your high-risk and high-value data?

AOL CIO James LaPlaine says "data classification is labor-intensive, but you have to do it."

It just makes sense: the most valuable or sensitive data requires the highest levels of security. Line-of-business teams have to collaborate with legal and security personnel to get this right.

A well-defined classification system should be paired with determination of data stewardship. If everybody owns the data, nobody is really accountable for its care and appropriate use, and it will be more difficult to apply information lifecycle policies.

2. What is your policy for data retention and deletion?

Every company needs clear directions on what data is kept, and for how long. Like any good policy, it needs to be clear (so everyone can follow it) and it needs to be enforced (so they will).

More data means more opportunity, but it can also mean more risk. The first step to reducing that risk is to get rid of what you don't need. This is a classic tenet of information lifecycle management.

"If data doesn't have a purpose, it's a liability," says Paula Greve, senior director of data science at McAfee Labs.

One idea for reducing that liability in regards to privacy is to apply de-identification techniques before storing data. That way you can still look for trends, but the data can't be linked to any individual. De-identification might not be appropriate for any given business need, but it can be a useful approach to have in your toolbox.

3. How do you track who accesses which data?

"How are you going to track the data, and who has access to the data—that's really a foundational element," says Greve. "You want to make sure the tools and storage mechanisms have that tracking capability built in from the beginning."

Greve points out that as your analytics programs are successful, you are likely to be exposed to more sensitive data.

"If you don't have the right tracking tools put in place, it's hard to add them after the fact," she says.

4. Are users creating copies of your corporate data?

Of course they are. Data tends to be copied. A department might want a local copy of a database for faster analysis. A single user might decide to put some data in an Excel spreadsheet. And so on.

So the next question to ask yourself, says Jim Greene, security technology lead at Intel's Datacenter Group, is "What is the governance model for this process—how are policies for control passed through to the new copy and the maintainer of this resource?"

Articulating a clear answer for your company will help prevent sensitive data from leaking out by gradually passing into less-secure repositories.

> "If data doesn't have a purpose, it's a liability."
>
> — Paula Greve

5. What types of encryption and data integrity mechanisms are required?

Beyond technical issues of cryptographic strength, hashing and salting and so on, here are sometimes-overlooked questions to address.

Is your encryption setup truly end-to-end, or is there a window of vulnerability between data capture and encryption, or at the point when data is decrypted for analysis? A number

of famous data breaches have occurred when hackers grabbed data at the point of capture.

Does your encryption method work seamlessly across all databases in your environment?

Do you store and manage your encryption keys securely, and who has access to those keys?

Ramnath Venugopalan, Chief Architect for Global Threat Intelligence with McAfee Labs, points out that encryption protects data from theft, but does not guarantee its integrity.

Separate data integrity mechanisms are required for some use cases, and become increasingly important as data volumes grow and more data sources are incorporated.

For example, to mitigate the risk of data poisoning or pollution, a company can implement automatic checks flagging incoming data which doesn't match the expected volume, file size or other pattern.

6. If your algorithms or data analysis methods are proprietary, how do you protect them?

Protecting proprietary discoveries? That's old hat. What's easier to miss is the way you arrive at those discoveries. In a competitive industry, a killer algorithm can be a valuable piece of intellectual property.

The data and systems get most of the glory, but analysis methods may deserve just as much protection, with both technical and legal safeguards.

Have you vetted and published a plan for securely handling this type of information?

7. How do you validate the security posture of all physical and virtual nodes in your analysis computing cluster?

Big data analysis often relies on the power of distributed computing. A rogue or infected node can cause your cluster to spring a data leak. Hardware-based controls deserve consideration. (Yes, this is an area of focus for Intel).

8. Are you working with data generated by Internet of Things sensors?

The key with IoT is to ensure that data is consistently secured from the edge to the data center, with a particular eye on privacy-related data.

IoT sensors may present their own security challenges. Are all gateways or other edge devices adequately protected? Industrial devices can be difficult to patch or have a less mature vulnerability management process.

9. What role does the cloud play in your analytics program?

You'll want to review the contractual obligations and internal policies of those hosting your data or processing. It's important to know which physical locations they will use, and whether all those facilities have consistent physical (not just logical) security controls.

And of course, the geographic locations may impact your regulatory compliance programs.

10. Which individuals in your IT organization are developing security skills and knowledge specific to your big-data tool set?

Over time, your project list, data sets, and toolbox are likely to grow. The more in-house knowledge you develop, the better your own security questions will be.

Special thanks to Paula Greve, senior director of data science at McAfee Labs; Jim Greene, security technology lead at Intel's Datacenter Group; Catherine Huang, research scientist at Intel; and Ramnath Venugopalan, Chief Architect, Global Threat Intelligence, McAfee Labs.

THANK YOU FOR READING THE ANALYTICS GUIDE

To access more resources on analytics and related topics, visit our data center resource hub: www.intel.com/datacenter. The hub is part of the IT Center, a resource collection designed to assist enterprise IT leaders evaluate emerging technologies and research trends and implementation strategies on a wide range of IT topics.

You can also follow IT Center on Twitter (http://bit.ly/1JrYE5K) and LinkedIn (http://bit.ly/1ZnvRZi).