Top 10 Data and Analytics Technology Trends That Will Change Your Business

Published 11 April 2019 - ID G00379563 - 65 min read

By Analysts Rita Sallam, Donald Feinberg, Mark Beyer, W. Roy Schulte, Alexander Linden, Joseph Unsworth, Svetlana Sicular, Nick Heudecker, Ehtisham Zaidi, Adam Ronthal, Erick Brethenoux, Pieter den Hamer, Alys Woodward

These data and analytics technology trends will have significant disruptive potential over the next three to five years. Data and analytics leaders must examine their business impacts and adjust their operating, business and strategy models accordingly.

Overview

Key Findings

- The expanded and strategic role of data and analytics in digital transformation is increasing the complexity of data, the number of variables to analyze, and the types of analyses required for success. This is pushing the limits of current capabilities and approaches.
- Virtually every aspect of data management, analytics content, application development and sharing of insights is using machine learning (ML) and artificial intelligence (AI) techniques to automate or augment manual tasks, analytic processes and human insight to action.
- Intelligent capabilities that enable emergent and agile data fabrics and explainable, transparent insights and AI at scale are necessary to meet the new demands and expand adoption.

Recommendations

For your data and analytics strategies:

- Educate, ideate and engage with business leaders about your strategic priorities and where data and analytics can automate or augment human activities.
- Put in place formal mechanisms to identify technology trends and prioritize those that can be incorporated into your strategy and roadmap with the biggest potential impact on the business.
- Take action over the next three to five years to proactively monitor, experiment with or exploit key trends. Don’t just react to trends as they mature.

- Identify the gaps in your data, analytics and organizational capabilities that are preventing you from exploiting the disruptive trends.
- Use success metrics and incentives that emphasize learning and reward innovation.
- Invest in nontechnology trends — such as data literacy, AI governance, data engineering, data storytelling and privacy and ethics — as these are critical key success enablers.

Strategic Planning Assumptions

By 2020, augmented analytics will be a dominant driver of new purchases of analytics and business intelligence as well as data science and machine learning platforms, and of embedded analytics.

Through 2022, data management manual tasks will be reduced by 45% through the addition of machine learning and automated service-level management.

By 2020, 50% of analytical queries will be generated via search, natural language processing or voice, or will be automatically generated.

By 2021, natural language processing and conversational analytics will boost analytics and business intelligence adoption from 35% of employees, to over 50%, including new classes of users, particularly front-office workers.

The application of graph processing and graph databases will grow at 100% annually through 2022 to continuously accelerate data preparation and enable more-complex and adaptive data science.

By 2022, 75% of new end-user solutions leveraging AI and ML techniques will be built with commercial instead of open-source platforms.

By 2022, cloud-based ML services from the hyperscale cloud providers (Amazon, Google and Microsoft) will achieve the digital tipping point of 20% share in the data science platform market.

By 2022, every personalized interaction between users and applications or devices will be adaptive.

Through 2022, custom-made data fabric designs will be deployed primarily as a static infrastructure, forcing organizations into a new wave of “cost to complete” redesigns for more-dynamic data mesh approaches.

By 2023, over 75% of large organizations will hire artificial intelligence specialists in behavior forensic, privacy and customer trust to reduce brand and reputation risk.

By 2021, most permissioned blockchain uses will be replaced by ledger DBMS products.

By 2022, more than half of major new business systems will incorporate continuous intelligence that uses real-time context data to improve decisions.

By 2021, persistent memory will represent over 10% of in-memory computing memory GB consumption.

Analysis

What Is Driving the Top 10 Data and Analytics Technology Trends?

The expanded and strategic role of data and analytics in digital transformation is increasing the complexity of data, the number of variables to be analyzed, the types of analysis, and the speed of analysis required for success. With this increasing complexity comes ever more subtle and potentially damaging risks and challenges, such as the potential for bias and the need for transparency and trust in analytics and in ML and AI models.

The size, complexity and distributed nature of data needed for increasingly closer to real-time and optimized decision making means that rigid architectures and tools are breaking down. This complexity is pushing the limits of current approaches, and is leading to unprecedented cycles of rapid innovation in data and analytics to meet the new requirements.

At the same time, to have an impact, data and analytics must be pervasive and scale across the enterprise and beyond to customers, partners and to the products themselves.

The strategic technologies covered in this research (and summarized in Figure 1) represent trends that you cannot afford to ignore. They have the potential to transform your business and will accelerate in their adoption over the next three to five years.

Many of the trends are interrelated as they are enabled by many of the same technology disruptions, but have an impact on different parts of the data and analytics technology stack.

All have three attributes in common — they support **intelligent, emergent** data and analytics and are **scalable** for pervasive AI- and ML-driven insights and agile data-centric architectures:

- **Intelligent:** Advanced analytics — including AI and ML techniques — are at the core of future platforms, solutions and applications. We see the green shoots of this today. Virtually every aspect of data management, analytic content and application development, and sharing of insights incorporates ML and AI techniques. These are used to automate or augment manual tasks, analytic processes and human insight to action for a range of user roles. Enhanced intelligence in all data and analytics platform components will democratize the skills required to leverage these capabilities and scale adoption across the enterprise to unprecedented levels.

- **Emergent:** Data and analytics is becoming more inductive in nature, assisted by AI and ML, as data models and analytics models are increasingly created by autogenerated models rather than by code. Multiple diverse structures and insights emerge *from* the data, rather than a single structure being imposed *on* the data. Intelligent capabilities that enable an emergent, agile and transparent data infrastructure are necessary to meet new demands and expand adoption. Success with data and analytics will depend on building a foundation of trust, accountability, governance and security that respects privacy and promotes digital ethics.

- **Scalable:** The very challenges created by digital disruption — too much data — have also created an unprecedented opportunity. Vast amounts of data are being coupled with increasingly powerful cloud processing capabilities (for both data management and data science) and the emerging capabilities enabled by a number of the top 10 trends. This makes it possible to train and execute algorithms at the scale necessary to realize the full potential of AI. Realizing this potential not only requires scale of processing, but also broader adoption of advanced analytics.

The compressed speed at which disruption is occurring requires data and analytics leaders to have formal mechanisms to identify technology trends, and prioritize those with the biggest potential impact on their competitive advantage. They should be added to your strategic planning, or if part of current plans, looked at in a fresh way based on the extent to which the trends enable your top business priorities.

Data and analytics leaders should proactively manage how they monitor, experiment with, or deploy emerging trends. Implementing success metrics and incentives that put an emphasis on learning and reward innovation when experimenting will further contribute to success.

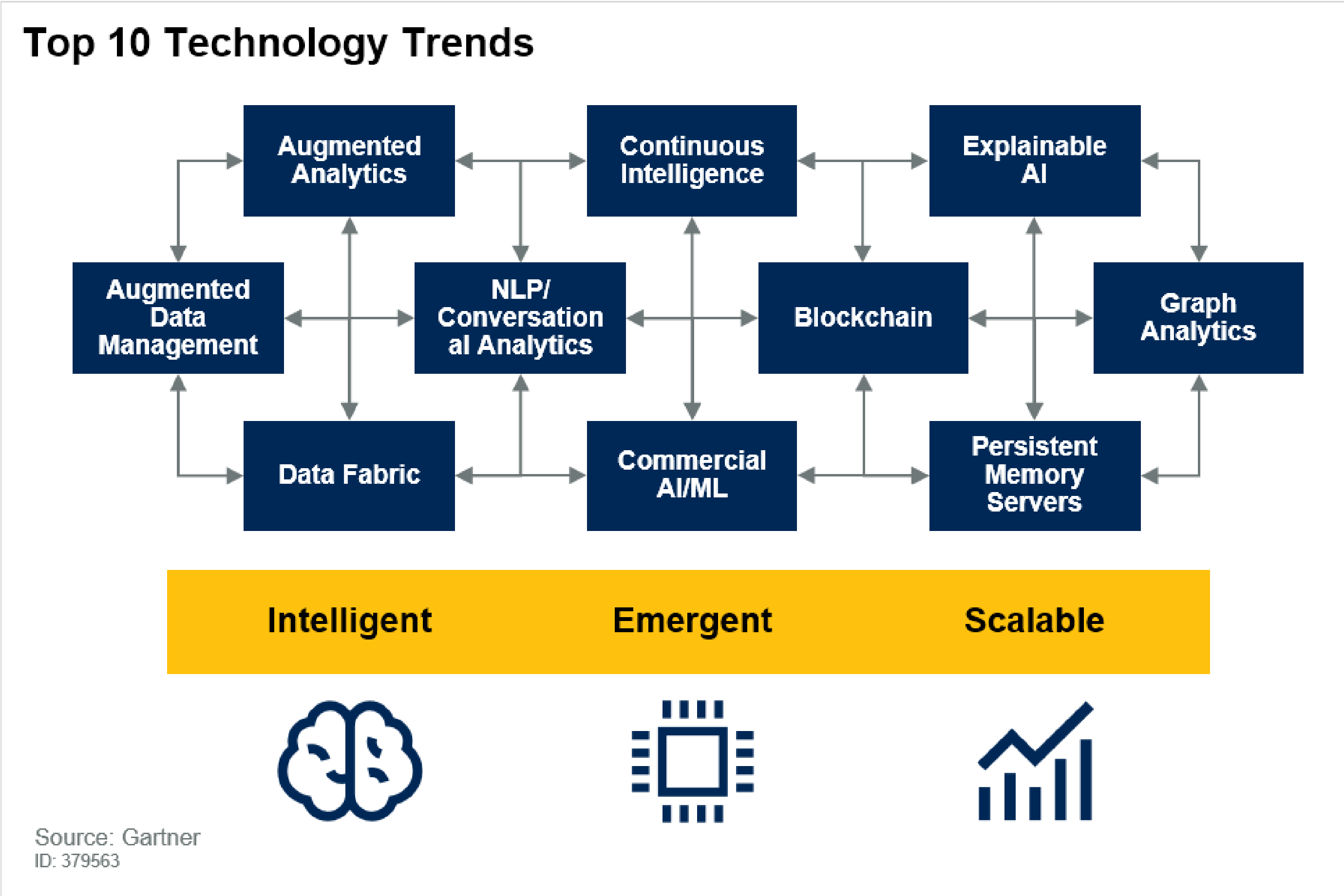
Note that, although we highlight some key nontechnology-related data and analytics trends in Figure 2, this report does not cover them in detail. It is, however, important to place an equal effort and investment on nontechnology trends (such as data literacy, AI governance, data engineering, data storytelling, privacy and ethics) as these are critical key success enablers.

This report also does not cover current trends that are less than three years away from mainstream adoption (such as cloud business intelligence or self-service) or trends that may be more than five years out (such as quantum computing).

You can find deeper insights into these trends in "The Future of Data and Analytics: Tales and Trends From the Center to the Edge."

For a current-period view of trends and actionable advice, see "Leadership Vision for 2019: Data and Analytics Leader."

Figure 1. Top 10 Technology Trends That Will Change Your Business



AI = artificial intelligence; ML = machine learning; NLP = natural language processing

Trend 1: Augmented Analytics

inline image

> **By 2020, augmented analytics will be a dominant driver of new purchases of analytics and business intelligence as well as data science and machine learning platforms, and of embedded analytics.**

Augmented analytics enables machine learning and AI assisted data preparation, insight generation, and insight explanation to augment how business people, analysts, explore and analyze data in analytics and BI platforms. It also augments the expert and citizen data scientists by automating many aspects of data science and ML model development, management and deployment.

Why Now?

- As data complexity increases, business people and decision makers are awash with data, risk, complexity and confusion. They struggle to identify what is most important, and what best actions to take (or avoid). Larger and more-varied dataset combinations also mean more variables and relationships to analyze, explore, model, test and govern. These tasks are increasingly difficult — if not impossible — to do using current manual approaches without introducing bias.

- Across the analytics stack, tools have become easier to use and more agile, enabling greater access and self-service. However, many processes remain largely manual, and are prone to bias. These include managing data, preparing data for analysis, analyzing data, building data science and ML/AI models, interpreting results, telling stories with data, and making insights actionable. A fundamental component of all of these activities is that hypotheses about relationships in data must be known in advance. Using current approaches, it is not possible for users to explore every possible combination and pattern, let alone determine whether their findings are the most relevant, significant and actionable out of all possible options.

- Relying on business users to find patterns, and data scientists to build and manage models manually, may result in them exploring and proving some relationship in their own biased hypotheses. As a result, they may miss key findings, and draw their own incorrect or incomplete conclusions. This adversely affects decisions, actions and outcomes.

What Does It Enable?

- Augmented analytics automates aspects of finding and surfacing the most important insights or changes in the business (in a user's context) to optimize decision making. It does this in a fraction of the time, with less data science and ML skills, and without prior knowledge of the relationships in data that is required when using manual approaches.

- Augmented analytics uses ML/AI techniques to automate key aspects of data science and ML/AI modeling, such as feature engineering and model selection (autoML), as well as model operationalization, model explanation and, ultimately, model tuning and management. Consequently, highly skilled data scientists have more time to focus on creative tasks, and on building and operationalizing the most-relevant models. The number of models that can be tested increases substantially, and the time and cost to iterate and test models goes way down.

- Many autogenerated and human-augmented ML models created through augmented analytics are being embedded in enterprise applications — for example, those of the HR, finance, sales, marketing, customer service, procurement and asset management departments. This helps to optimize the elusive "last mile" of data and analytics — the decisions and actions of all employees, not just those of analysts and data scientists.

- Augmented analytics capabilities are advancing rapidly to mainstream adoption, as a key feature of data preparation, broader data management, modern analytics and BI as well as data science and ML platforms.

- Augmented analytics can also be deployed with NLP and conversational interfaces (also a top 10 trend), to allow more people across the organization to interact with, make predictions and get actionable from data and insights without requiring advanced skills. It will make deep insights available to people who do not have the skills or access to ask their own questions from analytics and BI platforms.

How Does This Impact Your Organization and Skills?

- Augmented analytics will democratize insights from analytics (including AI) to all business roles. While this will reduce the reliance on expert analytics, data science and ML skills, this trend will require an increased focus on data literacy across the organization.
- Augmented analytics reduces the requirement for specialized data science and ML skills to generate, operationalize and manage an advanced analytics model. It also opens up data science and ML content creation to “citizen data scientists” (including business analysts and application developers who must embed ML/AI into applications). This makes expert data scientists more productive and collaborative, freeing them for high-value tasks. Putting in place processes to promote collaboration across roles leveraging augmented analytics capabilities will be critical to success.

Use Cases

Use cases span all industries and domains where data and the variables being analyzed have grown in complexity past the point of exploring completely and accurately using current approaches.

Some examples include:

- **Banking:** Before augmented analytics, banks targeted older customers for wealth management services. With augmented analytics, they found that younger clients (between the ages of 20 and 35) are actually more likely to transition into wealth management.
- **Agriculture:** Before augmented analytics, data scientists took months to build models to find the best handful of hybrid seed combinations out of thousands to sell to farmers. With augmented analytics, domain specialist geneticists took over the process and reduced process duration to days.
- **Healthcare:** Before augmented analytics, U.S. healthcare insurers tracked patient sickness measures as a key driver of transportation (ambulance) costs. With augmented analytics, they found that the main cost driver was under 12-year-olds. Investigation found that journeys were charged per person, and included parents accompanying sick children.

Recommendations

- Explore opportunities to complement existing data and analytics initiatives by piloting augmented analytics for high-value business problems currently requiring time-consuming, manual analysis.

- Build trust in machine-assisted models by fostering collaboration between expert and citizen data scientists to back-test and prove value. Understand the limitations of machine-assisted models, which work best with proven algorithms versus cutting-edge techniques.
- Monitor the augmented analytics capabilities and roadmaps of established data and analytics providers, enterprise application vendors and startups.
- Assess upfront setup, data preparation, openness and explainability of models, the number of variables supported, the range of algorithms provided, and the accuracy of the models.
- Develop a strategy to evolve roles, responsibilities and skills, and increase investments in data literacy.

Recommended Reading

"Pursue Citizen Data Science to Expand Analytics Use Cases"

"Augmented Analytics Is the Future of Data and Analytics"

"How Citizen Data Science Can Maximize Self-Service Analytics and Extend Data Science"

"Rebalance Your Integration Effort With a Mix of Human and Artificial Intelligence"

Trend 2: Augmented Data Management

inline image

> **Through 2022, data management manual tasks will be reduced by 45% through the addition of machine learning and automated service-level management.**

Similar to how ML and AI capabilities are transforming analytics, business intelligence and data science, across data management categories, vendors are adding ML capabilities and AI engines to make self-configuring and self-tuning processes pervasive. These processes are automating many of the manual tasks and allowing users with less technical skills to be more autonomous when using data. By doing so, highly skilled technical resources can focus on higher-value tasks. This trend is impacting all enterprise data management categories including data quality, metadata management, master data management, data integration and databases.

Why Now?

Data is growing in volume and diversity of source. At the same time, it is being acquired and generated from a combination of trusted and untrusted data sources. With technical skills in short supply, the need to automate data management tasks that respond to governance policies, rules and processes is increasing at a geometric rate. Even previously mundane (and easily accomplished) tasks are simply too numerous in a distributed environment to meet the critical demands of scaling at the speed of digital business.

As customers move their data assets to the cloud, more usage statistics are available, and across a standardized base of configurations. With each provider growing more confident in recognizing data management patterns, it is becoming possible to build models that allow cloud providers to use and execute automated data management strategies.

These patterns — and the ability to recognize the veracity of each approach — make it easier to "clone" the approaches and deployment options that work best. For example, "I need another sandbox configured like this one" is a matter of a few clicks (or even machine-issued instructions). The result is to dramatically enhance deployment agility.

What Does It Enable?

- Emergent metadata can be inferred from data utilization, users and use cases rather than descriptive metadata that is often no longer synchronized with actual data capture/write and subsequent usage.
- Active utilization of all metadata types can be used to develop an established data fabric through processes that leverage the inventorying, cataloging, automatic discovery of semantics, taxonomy and ontology that is crucial to data management. Organizations need to easily know what data they have, what it means, how it delivers value, and whether it can be trusted. Utilizing the statistics of existing systems — and their available capacity, known resources, cost parameters of using each resource, and the communications capability among infrastructure components — policy-level instructions will determine where data operations will take place. They can even manage their production deployment to the point of reconfiguring that deployment when necessary.
- Data fabric approaches that are composed of microservices, APIs, data pipelines and data quality/governance services, will be deployed as part of the new environment in a more customized fashion when consistency of performance and persistence of operations is required.
- Data fusion can track which assets are used by use case, and form a knowledge and utilization graph. When new data assets are encountered, fusion engines will analyze the similarity to other well-known data assets. They will determine their affinity to existing data/use cases, then alert other automated systems that new data is available, and is a valid candidate for inclusion.
- Dynamic data identification allows new and existing data assets to be evaluated "in stream," and cumulative information about them to be used to develop related event models. Over time, the use cases will form processing requirements that indicate which data needs to be

provided for operational and/or analytics use cases — and the rate or frequency of providing it.

How Does This Impact Your Organization and Skills?

> **By 2023, AI-enabled automation in data management will reduce the need for IT specialists by 20%.**

Impact on the data and analytics organization and skills:

- Augmenting the data engineer or automating some data engineering tasks.
- Alerting data engineers to potential errors, issues, and alternative interpretations of the data.
- Creating automated system responses to errors, issues and alternative interpretation of data.
- Increasing the capability to use publicly available data, partner data, open data and other assets that are currently difficult to determine as appropriate for utilization.
- Automating data interrogation that mimics data discovery and even evaluates the "confidence" that new assets conform to known or existing models.

Impact on distributed data management:

- Continuous monitoring of capacity and utilization for data management environments for potential redistribution of resource planning, even across multicloud and on-premises implementations.
- Optimization and performance management that eliminates the manual (or human) determination of when to create intermediate, temporary or permanent copies of data to enhance operational or analytical performance.
- Policy-based decision engines that capture regulatory requirements as metadata configurations and then manage the hot/warm data use cases, cooler data usage, archiving expectations and even data purge or data shredding to maintain legal and audit compliance.

Use Cases

Augmented data management leverages machine learning and AI techniques for:

- **Data quality:** To extend profiling, cleansing, linking, identifying and semantically reconciling master data in different data sources.
- **MDM:** For ML-driven configuration and optimization of record-matching and merging algorithms.
- **Data integration:** To simplify the integration development process, by recommending or even automating repetitive integration flows.
- **DBMS:** For automated management of storage, indexes, partitions, database tuning, patching, upgrading, security and configuration.
- **Metadata management:** Well beyond data lineage, ML can be used to evaluate data rules, populate semantic frameworks, and aid metadata discovery and ingestion across increasingly diverse sources.

Recommendations

- Plan on fewer skilled personnel and staff with expertise in balancing physical infrastructure with logical data management requirements. Begin by exploring capacity planning tools that include dynamic hardware/infrastructure provisioning.
- Develop a strategy for isolating data assets that will be required in multicloud and on-premises scenarios for potential replication/copy instances while simultaneously considering regulatory and privacy constraints on their physical location.
- Build data literacy into the organization. Although automation will reduce the skills barrier to using data, it will still require users to understand data and make proper uses of it. This will increase the demand for expanded metadata access and utilization.

Recommended Reading

"Toolkit: Map Your Data Management Landscape With the Data and Analytics Infrastructure Model"

"Toolkit: Comparative Total Cost of Ownership Calculator for Cloud and On-Premises DBMS Deployments"

"Data Hubs, Data Lakes and Data Warehouses: Choosing the Core of Your Digital Platform"

"Data Management Strategies: Navigating Diverse Roles, Use Cases and Markets"

Trend 3: NLP and Conversational Analytics

inline image

By 2020, 50% of analytical queries will be generated via search, natural language processing or voice, or will be automatically generated.

By 2021, natural language processing and conversational analytics will boost analytics and business intelligence adoption from 35% of employees, to over 50%, including new classes of users, particularly front-office workers.

NLP provides any user with an easier way to ask questions about data, as well as receive an explanation of rendered insights. Just as search interfaces like Google made the internet accessible to everyday users and consumers, search and natural language query (NLQ) approaches to asking questions of data enable data and analytics access to mainstream business users. This includes those that don't have the skills or access to current point-and-click, visual-based analytics and BI systems.

Currently, vendors take a wide range of approaches to enabling such interfaces. Some use simple keyword matching (the familiar Google-style "search"). Others use robust natural language queries with support for more-sophisticated queries using industry-specific language. Early efforts, for example, may allow a user to ask for "sales by product" that generates a basic bar chart. More-robust models would support questions such as, "show me top-selling products within a 50 mile radius, and comparing this year with last year." Although "top selling" is not a keyword, the platform understands that it must apply a rank function, perform geographic analytics, and then do a subquery to compare year over year.

Natural language query (NLQ) is being combined with natural language generation (NLG) to interpret and explain the results presented with automatically generated text. This text may describe visualizations on the screen, or it may describe related patterns in the data that were not specifically requested, and suggest actions in natural language.

Poor data literacy is currently hindering the impact of analytics and BI initiatives. By improving consistent interpretation of insights in data, regardless of user skill, NLG (by itself and in combination with NLQ and augmented analytics) has the potential to improve the level of data literacy across an organization. Here too, the textual descriptions may be in the form of written text, or voice-generated, or both. The ability to control aspects of language such as verbosity and tone — and how much of the text is form-based, template-driven or automated — varies between products.

Conversational analytics is still emerging, but takes the concept of NLQ and NLG a step further, first by allowing such questions to be posed verbally. This voice interaction could be through a digital assistant (such as Amazon Alexa), on a mobile phone, or using other devices. In addition to voice support for queries, these capabilities are emerging to be conversational in the form of a virtual AI assistant or bots. NLP/conversational interfaces will also leverage insights and embedded ML/AI models generated from augmented analytics. The virtual assistant may answer the initial question posed, but then elaborate with autogenerated insights from augmented analytics, and suggestions in natural language such as, "there was a spike in sales for purple products in week 30. Would you like more insights about what happened and why?" It will also make predictions and prescribe actions.

Why Now?

As organizations transform into digital businesses, analytics becomes a critical enabler. Expanding access to insights from analytics to all workers will be key to driving transformative business impact. However, access to analytics content from BI and analytics and data science platforms has mostly been limited to power users, business analysts and specialist data scientists with varying degrees of analytical and technical skills.

Despite major improvements to the ease of use of analytics and BI platforms, adoption is still low, with a high emphasis on empowering power users over mainstream business users.

Early attempts at search-based BI required significant administrative set up, supported simple questions, and worked only on small sets of data. More recent solutions enable more complex questions and require little advanced configuration with automatic indexing and fast performance on large datasets.

Massively parallel processing, broader use of complex data models (including structured and unstructured data) often based on knowledge graphs, and the use of GPUs and cloud platforms to crunch data, are all making NLP at scale possible and enabling broader adoption.

Voice-based digital assistants, and the open APIs that give free access to these, enable vendors to enhance their platforms with voice, without necessarily having to develop their own technology. The same is true with chatbots.

What Does It Enable?

- Most analytics and BI tools currently require users to choose data elements and place them on a page to create queries and visual analysis. NLP/conversational analytics brings ease of use to a new level, and allows a query to be as easy as a Google-like search or a conversation digital assistants, such as Alexa.

- Any user can ask questions using text or voice with increasingly complex questions and responses. NLP is increasingly an interface to querying and interacting with autogenerated insights from augmented analytics.

- The combination of NLP with augmented analytics — including automatic insight generation — allows users to rapidly find the proverbial needle in a haystack, presenting the most-important and actionable insights via conversational analytics with natural language generation.
- More-robust NLP interfaces include industry or domain-specific taxonomies and linguistics (such as HR, finance, healthcare or financial services) so that phrases are more correctly interpreted.
- NLP and conversational interfaces are being embedded in analytics and BI platforms, digital personal assistants, chatbots and applications.
- NLP provides intuitive forms of communication between humans and systems. NLP includes computational linguistic techniques (both traditional and machine learning) aimed at recognizing, parsing, interpreting and generating natural languages.
- NLP and underlying knowledge graphs are being extended to analyze unstructured and other data types, and used as a foundation for data science as well as ML/AI models.

How Does This Impact Your Organization and Skills?

Conversational analytics can dramatically improve the adoption of analytics by every employee, rather than by predominant power users and business analysts, resulting in higher business impact. However, as access to more powerful insights permeates the organization at all skill levels, there is a need for a formal and enterprisewide focus on improving the data literacy of all users.

Use Cases

Instead of logging into a complicated dashboard, any user — from the C-suite to analysts to operational workers — can interact verbally with a virtual personal assistant or their mobile phone, and ask for an analysis that is relevant to them.

In combination with augmented analytics capabilities, a salesperson might, for example, ask for an analysis of sales or a pipeline. Or the system may have learned that the sales manager also looks at this information. Based on that person's role and/or behavior, they will be served up an explanation or narrative — in text or voice — of statistically important drivers of change. They could then be sent visualizations (via a device) that show important trends, patterns or outliers based on their role. Predictions and prescriptive recommendations could also be communicated to the sales manager. Conversational analytics will also be embedded in the workflow of applications that every employee uses. The starting point of the analysis might be an autogenerated visualization or insight from augmented analytics where any user can explore further by asking additional questions in text or voice.

Recommendations

- Evaluate the capabilities, roadmaps and partnerships of your analytics and BI platform and enterprise application vendors, as well as those from startups.
- Assess the maturity and scalability of solutions, particularly in terms of integration and ease of use, upfront setup/configuration requirements, languages supported and types of analysis.
- Invest in data literacy as a critical element of success.

Recommended Reading

"Magic Quadrant for Analytics and Business Intelligence Platforms"

Trend 4: Graph Analytics

inline image

> **The application of graph processing and graph databases will grow at 100% annually through 2022 to continuously accelerate data preparation and enable more-complex and adaptive data science.**

Graph analytics is a set of analytic techniques that allows for the exploration of relationships between entities of interest such as organizations, people and transactions. One application of graph analytics — graph-enabled semantic knowledge graphs — forms the foundation of many NLP/conversational interfaces and data fabrics, and enriches and accelerates data preparation and data science.

Why Now?

Business people are asking increasingly complex questions across structured and unstructured data. Complex analysis often requires blending data from multiple sources, multiple business units, and increasingly external data. Concepts such as weather, economics, production, personnel, regulations, performance metrics, alternative assets versus primary assets — and even something as basic as time-zone shifts — are complicated when it comes to assuring that content and media are relevant. Analyzing this at scale is not practical, or in some cases possible, using traditional query tools, or the use of query languages (SQL is most commonly used in most tools or platforms).

Many use cases have graphs as their enabling technology. Graph analytics are typically portrayed via a visualization for business-user consumption. In graph visualization, it is possible to drag one metric inside of another, or remove part of a metric to recombine it with other datasets. Graph analytics consists of models that determine the “connectedness” across data points.

Improved, scalable, and lower-cost processing options such as the cloud and GPUs are making graph analytics and databases prime candidates for accelerated adoption.

What Does It Enable?

Graph analysis shows how tightly several trends or data points are related to each other. For example, an automated lighting system could be set to come on at dusk, but it might also come on during a severe weather situation, and time-of-day and seasonal variations may be combined with “waking hours” evaluation models. All of these inputs may or may not be related to each other, and graph analytics shows these as compact (in terms of their degree of relatedness) “nodes” of data. It also reveals whether those nodes are tightly compacted or just loose connections. When conditions change, the demarcation points between nodes may change, and a whole new set of nodes may dynamically emerge. Graph analysis also shows if something is only a correlation, or if the two (or more) things are related as precedents with dependency, or just have “causal” relationships. This means the nodes are connected explicitly or implicitly, indicating the levels of influence, frequency of interaction, or probability.

Graph models determine “connectedness” across data points and create clusters based on levels of influence, frequency of interaction and probability. Once highly complex models are developed and trained, the output is easier to store because of the expanded capabilities, computational power and adoption of graph databases. The user can interact directly with the graph elements to find insights, and the analytic results, and output can also be stored for repeated use in a graph database.

Graph databases therefore present an ideal framework for storing, manipulating and analyzing graph models, although some graph analytics vendors have their own graph analytics engines that do not require a separate graph database.

Generating a dynamic graph about how different entities of interest — people, places and things — are related, instead of more-static relational schemes, enables deeper insights that are closer to human knowledge representation. For example, it can easily combine, relate and find dynamic insights from data from exercise apps, diet planners, medical records and health news feeds.

Graph technology underpins the creation of richer semantic graphs or knowledge bases that can enhance augmented analytics models, as well as the richness of conversational analytics. Graph analytics also supports the creation of emergent metadata management and data catalogs. It does so by capturing all of the knowledge about what data you have, where it resides, how it is all related, who uses it, why, when and how. That insight can be leveraged by ML models to make recommendations to provide more-personalized, automated and properly governed insights to the business and its applications.

How Does This Impact Your Organization and Skills?

- New skills will be required. These include graph-specific standards, graph databases, technologies and languages such as Resource Description Framework (RDF), SPARQL Protocol and RDF Query Language (SPARQL); and emerging languages, such as Apache TinkerPop or the recently open-sourced Cypher. The scarcity of these skills has been an inhibitor to graph adoption.
- Commercialization of graph analytics is still at an early stage, with a small number of emerging players.
- While current technologies still require specialist skills, SQL-to-graph interpreters are emerging. These interpreters convert graph-based procedures into disaggregated procedural SQL (and back again). These capabilities are helping to make graph technologies compatible with existing datasets.

Use Cases

The number of use cases is increasing with the need for complex analysis. They range from fraud detection to customer influencer networks, through to social networks and semantic knowledge graphs. Conversational analytics, health advisors, financial crimes and risk detection can also leverage graph capabilities.

Other specialized applications include:

- Route optimization between waypoints for transportation, distribution, and even foot traffic.
- Market basket analysis to show products that have a coincidental or dependency relationship.
- Fraud detection to identify clusters of activity around connected groups of “actors.”
- Social network analysis to determine the presence of influencers, “canaries,” decision makers, and dissuaders.
- CRM optimization to determine probabilities of success for things like “next best offer.”
- Location intelligence to determine probable routes or locations from known data points when location is otherwise indeterminate.
- Load balancing in any type of network system like communications or utilities, but also something more concise like computer networks.
- Special forms of workforce analytics, such as enterprise social graphs and digital workplace graphs.
- Recency, frequency, monetary analysis of related networks of objects, assets and conditions to help designate the best resources at the best utilization point.

- Law enforcement investigation to isolate missing or unknown risks and even identities (such as third-party relationships in child endangerment cases).
- Epidemiology for analyzing the intersection of environment, overall condition of a patient, diet, exercise, acute/chronic disease analysis, interaction of therapeutic and chemical/pharmaceutical treatment regimens.
- Genome research for gene interaction to determine the potential for and treatment of existing hereditary disease conditions to assist with targeted medical practices.
- Detection of money laundering to unearth relationships among "actors" to identify malignant versus benign actions.

Recommendations

Data and analytics leaders should:

- Evaluate opportunities to incorporate graph analytics into their analytics portfolios and strategies for both analytic applications, and as part of the underlying data fabric in the form of semantic knowledge graphs.
- Assess the products available from both incumbent and new vendors, but recognize that most solutions are maturing, and are often focused on specific domains and verticals.
- Explore the use of new graph-to-SQL interpreters, but also invest in developing the necessary unique skills and competencies.

Recommended Reading

"An Introduction to Graph Data Stores and Applicable Use Cases"

"The Future of Data and Analytics: Tales and Trends From the Center to the Edge"

"Making Big Data Normal With Graph Analysis for the Masses"

Trend 5: Commercial AI/ML Will Dominate the Market Over Open Source

inline image

By 2022, 75% of new end-user solutions leveraging AI and ML techniques will be built with commercial instead of open-source platforms

Why Now?

The current dominance of open-source platforms (including Python, R, Apache Spark, H2O.ai, Anaconda and TensorFlow) has been a sign of market immaturity. Continued fast-track innovation in both algorithms and development environments over the past five years has mostly occurred in open-source options. However, commercial vendors are now building connectors into the open-source ecosystem. Importantly, they also provide the kind of enterprise features necessary to scale AI and ML — features that open-source technologies lack. These features include project and model management, reuse, transparency, data lineage, platform cohesiveness and integration.

IT megavendors were slow to respond to open-source innovation and only started to deliver viable offerings in 2017. In 2018, Amazon announced its current ML platform (SageMaker) and Gartner predicts, that Google will most likely opt to put most of its ML innovations into the Kubernetes framework. IBM and SAS have been de-emphasizing their older SPSS and Enterprise Miner platforms for a number of years, concentrating instead on their more-modern platforms (IBM Watson, SAS Visual Data Mining and Machine Learning). SAP and Oracle have also significantly revamped their current ML strategies and offerings.

By 2022, cloud-based ML services from the hyperscale cloud providers (Amazon, Google and Microsoft) will achieve the digital tipping point of 20% share in the data science platform market.

Open-source libraries and development environments have provided a much-needed democratization of the data science and machine learning fields, as well as innovation and flexibility. But, as was the case in the past with predictive analytics techniques, flexibility and power can come at the expense of disciplined operationalization mechanisms. Many models have been created, but their business value has not been fully realized as they have not been put into production at scale.

Commercial platforms offer a much more rigorous and disciplined approach and capabilities including:

- Rapid innovation in big data, data science and AI has started to cool off. As commercial vendors build connectors into the open-source ecosystem, users have found that they can have both the cool and innovative open-source components, and the increasingly enterprise-ready commercial ML/AI platforms.
- Innovation in the open-source market was never evenly distributed. Most of it was happening around algorithms and developer-oriented machine learning platforms. Much less innovation occurred in the areas of collaboration, data and user access rights, operationalization,

metadata/model management, and a seamless user-experience (especially for less-skilled end users like citizen data scientists).

By 2022, every personalized interaction between users and applications or devices will be adaptive.

Data science and machine learning teams are now starting to be measured on business results rather than production metrics (the number of models produced, or projects started, for example). Consequently, the required disciplined approach brought about by commercial platforms is becoming a required condition to achieving business value and data science team sustainability.

What Does It Enable?

- **Better productivity:** The assembly and retro-fitting of diverse OS tools requires lots of skills and manual labor, which has become the focal point of commercial attention, along with a clearer path toward business value.
- **Democratization of AI:** Commercial providers will increasingly “smooth” the rough edges often associated with open-source projects by orchestrating the user experience and “connecting the dots.” This provides much more integrated capabilities through orchestration (see “Embracing Competition to Evolve and Enable Data and Analytics Product Offerings”). This also makes cutting-edge AI/ML development much more available to the broader skill set found in most enterprises.
- **Better AI/ML planning and roadmaps:** Current AI strategies are full of uncertainty, conflict and vagueness. The comeback of capable commercial platforms will increase plannability and IT roadmaps by providing concrete anchor points into the IT software infrastructure.

How Does This Impact Your Organization and Skills?

- Increased use of commercial data science and machine learning will help to narrow the current skills gap in these areas. It will also facilitate greater collaboration among AI/ML developers with varying levels of skill.

Use Cases

- Marketplaces for algorithms and pretrained models from Amazon SageMaker, Google AI Hub, Microsoft, KNIME, RapidMiner will be accelerators of commercial data science and ML platforms. This will further extenuate the need for building a core competency around collecting data, managing it, curating it and making it accessible in an agile way.

Recommendations

- Start or continue to upskill existing professional staff (from business and technology departments, for example) into a cast of citizen data scientists, while progressively moving to a commercial production environment.
- Emphasize easy-to-use augmented data science/ML solutions and upcoming marketplaces as catalysts for simplified ML adoption, or to bootstrap ML efforts.
- As your data science, ML and AI capabilities mature and scale, instigate success metrics for the models you have deployed into production, and the business impact of those models, rather than the number of models created.
- Create an open-source audit process aimed at validating open-source models before integrating them into a commercial production environment.
- Focus on data management competency because, as algorithms commoditize, data will be the critical determinant of AI/ML success.

Recommended Reading

"Embracing Competition to Evolve and Enable Data and Analytics Product Offerings"

"How to Operationalize Machine Learning and Data Science Projects"

"Six Pitfalls to Avoid When Executing Data Science and Machine Learning Projects"

Trend 6: Data Fabric

inline image

> **Through 2022, custom-made data fabric designs will be deployed primarily as a static infrastructure, forcing organizations into a new wave of "cost to complete" redesigns for more-dynamic data mesh approaches.**

Deriving value from analytics investments depends on having an agile and trusted data fabric. A data fabric is generally a custom-made design that provides reusable data services, pipelines, semantic tiers or APIs via combination of data integration approaches (bulk/batch, message queue, virtualized, streams, events, replication or synchronization) in an orchestrated fashion. Data fabrics can be improved by adding dynamic schema recognition, or even cost-based optimization approaches (and other augmented data management capabilities). As a data fabric becomes increasingly dynamic, or even introduces ML capabilities, it evolves from a data fabric into a data mesh network.

Why Now?

Organizations are successfully deploying a greater number of larger and more-complex data and analytics implementations than in any previous period. The steady growth of the data management and analytics markets has demonstrated the value of consistent, semantically harmonized and governed information assets.

Increasingly, dynamic recognition of assets of critical importance to business outcomes exists for the organization ecosystem. The convergence of data management technologies will enable the rise of common platforms for new and differentiated managed data services exchanges. This is being driven by the need for consumption, modeling and effective visualization of a growing and varied source of information assets. All of this needs to be carried out in a consistent and semantically harmonized way, and should be enabled by layers of active metadata utilization.

What Does It Enable?

The best approach to understanding the intersection of data integration strategies is to start from traditional practices, and then simultaneously consider new or emerging practices in architecture and infrastructure design. This will allow you to compare and contrast these practices with your organization's aspirational design goals.

Data integration approaches now (and in the future) must effectively address the same three issues:

1. Metadata drives the overall comprehension and performance optimization of any data asset in the organization.

2. Perpetual connections are not yet above 80% reliability when it comes to periodically connected devices or sensors at the edge, and lack the physical infrastructure needed for reliable communications.

3. Processing is always a mobile consideration and, as a logical requirement, can be relocated in such a way that the process can be brought to the data or the data to the process. This results in the negation of most considerations for selecting the deployment platform (including cloud, on-premises, multicloud and hybrid, among other combinations).

Current and traditional practices are focused on forms of physical consolidation combined with semantic interpretation:

- Data warehouses, data lakes and operational data stores remain prominent as data repositories that represent specific persistence and performance expectations.
- More-traditional data integration approaches still prefer to identify and design specific targets. For comprehension and performance reasons, the data warehouse, base operational data store systems, and even the logical data warehouse use a combination of the physical data lake and the physical data warehouse alongside some type of unifying semantic access tier.
- Data hubs represent a next step in traditional design. They combine the idea of distributed data management for geographic or domain “regions” with a data services back-plane that exchanges data in predispositions developed using microservices, APIs or P2P data services in the form of desktop as a service (DaaS).

If the metadata, physical management and processing design are considered as the primary “poles” for deployment, the infrastructure design has two complementary approaches:

- **Data fabric** is more of a designed approach, mostly tending toward use cases and locations on either “end” of a thread. The threads may cross and do handoffs in the middle, or even reuse their parts, but they are not built up dynamically. They are merely highly reusable, normalized services.
- **Data mesh** is a fully metadata-driven approach. Statistics in the form of metadata accumulation are kept relating to the rate of data access; platform, user and use case access; the physical capacity of the system; and the utilization of the infrastructure components. Other data points include the reliability of the infrastructure, the trending of data usage by domain and use case, and the qualification, enrichment and integrity (both declared and implied) of the data.

How Does This Impact Your Organization and Skills?

Specific roles will undergo new workflows and definitive changes in their responsibilities and tasks.

- **Data engineer:** This role is composed of tasks that determine the accessibility, qualification, delivery and processing of data. It requires an understanding of how data is used in a broad spectrum of business processes, the variable semantic and schema-based approaches to using the same data concepts, and the widely varied data quality presented when data is combined. The human capabilities of data engineers will be augmented by AI/ML processes that identify nearly all of the initial pain points for data refactoring, modeling, schema production and data quality recognition. These processes might even provide expert advice regarding infrastructure decisions — if not providing actual, dynamic resource allocation and provisioning.

- **Data scientist:** This role will benefit from data fusion outputs that create alerts about expanding data assets. These alerts will be specifically tuned for the current project — but, can also begin to recognize data that a given scientist utilizes in terms of data patterns.
- **Data modeler:** Data modelers, data integration developers and database administrators responsible for data modeling will model less, and verify more. Data models are always present, some formal and others "not so much." In less-structured data, models can be collaborative.
- **Information architect:** Information architects determine the alignment of intended functions of information collection and management with the appropriate form needed for acquiring, managing, accessing and utilizing data. Information architects working with data fabric will need to focus on identifying the required functionality of a data asset and imputing it as metadata.

Use Cases

- **Dynamic data engineering:** The data fabric architecture can be investigated and adopted for more-dynamic, reusable and optimized data engineering pipelines. This can be designed once, and can then work in "auto" execution mode, with data engineers acting as adjudicators and facilitators rather than developers of integrated data flows. They need to be flexible to changing data environments (cloud, multicloud and hybrid cloud) and even the users and use cases involved in this data.
- **Governed/trusted data science:** Data scientists now demand end-to-end lineage and understanding of their data models and algorithms for efficiency and compliance, and the data fabric design helps them with data fusion outputs that alert them to expanding data assets. This gives them more visibility into their active metadata (including performance optimization, data quality, design and lineage, among several others) and allows them to be in more control of their projects.
- **Logical data warehouse architecture:** The data fabric and mesh concepts will liberate data management from having to confine data to a physical consolidation combined with static semantic interpretation. This limits reuse of integrated data across heterogeneous applications needing specific data models and semantic interpretations.

Recommendations

- Begin tracking the origins and sources of data as well as types of reporting and analysis based on the characteristics of all of the use cases.
- Introduce data fabric designs when the primary data management and integration approach is focused on *connecting* (as opposed to *collecting*) data. Many of the tools used to build a data fabric are also referred to as "semantic tools," and these tools can also move data to collections as write-out caching tiers.

- Apply a three-tiered data fabric design that isolates industrywide data for one design and one common conceptual model, the organization's prevalent (but not universal) data in a second tier, and transactional values as an almost "attached data only" third tier.
- Build a pilot data asset resource utilization engine that uses machine learning-enabled data catalogs. Start by collecting the required metadata and performing analysis over this metadata to determine the specifications for a simple "alert" function that tells users that new data is available — possibly from a data lake.

Recommended Reading

- "Maverick* Research: Revolutionizing Data Management and Integration With Data Mesh Networks"
- "Toolkit: Identify Data Management Issues Before Moving Data to or From the Cloud"
- "Critical Capabilities for Data Integration Tools"
- "Magic Quadrant for Data Integration Tools"

Trend 7: Explainable AI

inline image

> **By 2023, over 75% of large organizations will hire artificial intelligence specialists in behavior forensic, privacy and customer trust to reduce brand and reputation risk.**

As we move toward more augmented analytics including autogenerated insights and models, the explainability of these insights and models will become critical to trust, to regulatory compliance, and to brand reputation management. Explainable AI is the set of capabilities that describes a model, highlights its strengths and weaknesses, predicts its likely behavior, and identifies any potential biases. It has the ability to articulate the decisions of a descriptive, predictive or prescriptive model to enable accuracy, fairness, accountability, stability and transparency in algorithmic decision making.

AI governance is the process of assigning and assuring organizational accountability, decision rights, risks, policies and investment decisions for applying AI, predictive models and algorithms. Explainable AI provides a technical foundation to support AI governance.

Why Now?

The proliferation of autogenerated insights and models from augmented analytics, and the creation of complex black-box machine learning and AI solutions, combined with biases that are unavoidable in AI, are all contributing to the rise in explainable AI.

Black-box approaches can inhibit trust and acceptance of AI. Opaque algorithms (DNNs, for example) incorporate many implicit, highly variable interactions into their predictions that can be difficult to interpret.

Such an approach can also lead to a misappropriation of investment. You might conclude that you need to invest in AI to improve or automate the context development aspect of making decisions, when in fact there is more value in augmenting the use of the analytics to select a decision option. Understanding how decisions are made and modelled can help you identify where the technology is most needed.

Although augmented data science and ML bring productivity enhancements to the data science world by helping with skills shortages and long development times for custom AI solutions, the need for explainability becomes more pronounced as their adoption expands.

Augmented data science and machine learning brings automation to the tasks of algorithm selection, feature selection and hyperparameter tuning. Augmented analytics solutions with explainable AI features are not only showing data scientists the input and output of a model, but are also explaining why the system selected particular models, and the techniques applied by augmented data science and ML.

Bias in AI raises concerns of accountability and fairness. Consequently, the AI community and enterprise leaders are concerned with detecting and explaining the consequences of bias that might endanger society and the business. For example, bias in AI causes polarization of political opinions, persistence of discredited beliefs, and false associations between business moments.

Without acceptable explanation, autogenerated insights and models, or black boxes combined with AI bias, can cause concerns about regulation, reputation, accountability and fairness, and lead to distrust in AI solutions.

What Does It Enable?

Explainable AI enables a better adoption of AI by increasing the transparency and trustworthiness of AI solutions and outcomes. Explainable AI also reduces the risks associated with regulatory and reputational accountability for safety and fairness.

Increasingly, these solutions are not only showing data scientists the input and the output of a model, but are also explaining the reasons the system selected particular models and the techniques applied by augmented data science and ML.

Bias has been a long-standing risk in training AI models. Bias could be based on race, gender, age or location. There is also temporal bias, bias toward a specific structure of data, or even bias in selecting a problem to solve. Explainable AI solutions are beginning to identify these and other potential sources of bias.

Explainable AI technologies may also identify privacy violation risk, with options for privacy-aware machine learning (PAML), multiparty computation, and variants of homomorphic encryption to identify privacy violation risks.

How Does This Impact Your Organization and Skills?

Diversity is a critical foundation for explainable AI because:

- Diversity of data is necessary to have an objective view of the problem and deliver trusted outcomes.
- Diversity of algorithms is necessary to make trade-offs, especially to resolve the dilemma of accuracy versus explainability.
- Diversity of people is necessary to build a successful AI team and an AI ethics board to minimize reputational and business risks, as well as to ensure AI safety.

Data and analytics leaders should invest in training and education to develop the skills needed to mitigate risks in black-box models.

This should include:

- How to make data science and ML models interpretable by design, and how to select the right model transparency from a range of models, from least to most transparent (taking into account their implications for accuracy).
- How to select the right model accuracy when required, and methods of validating and explaining these models "post hoc," such as model-agnostic or model-specific explainability; global (across the entire model) or local (a specific output) explainability.
- Various methods, such as generative explainability and combining simple, but explainable models, with more complex, but less explainable ones.
- Exploring the latest explainability techniques, such as the ones that are tracked by DARPA, or that are coming from commercial vendors.
- Visualization approaches for seeing and understanding the data in the context of training and interpreting machine learning algorithms (for example, visualization of correlations or of outliers in the data).
- Techniques for understanding and validating the most-complex types of predictive models, such as sensitivity analysis, surrogate models and "leave one covariate out" (LOCO).

- Communication and empathy skills for data scientists to detect the users' attitude and needs for explainability and successful AI adoption.
- Establish AI ethics boards and other groups that are responsible for AI safety, fairness and ethics. These boards should include internal and external individuals known for their high reputation and integrity.

Use Cases

The commercial sector delivers explainable AI to analytics and BI platforms and in data science and ML platforms in various ways.

Commercial offerings include:

- Data science platforms such as DataRobot Labs and H2O.ai, which automatically generate model explanations in natural language.
- tazi.ai provides pattern visualizations (such as profit/loss prediction patterns) for nontechnical business users and enables them to investigate possible patterns with explanations in an interactive fashion.
- DarwinAI offers Generative Synthesis technology, which provides a tool for granular insights into neural network performance.
- ZestFinance specializes in the financial industry, and provides accurate and transparent AutoML models for lending.
- Salesforce Einstein Discovery explains model findings and will alert users to potential bias in data.

In addition to vendor technologies that enhance model explainability, many initiatives have emerged to tackle the challenge of AI explainability.

- The U.S. Defense Advanced Research Projects Agency (DARPA) has determined that the effectiveness of AI systems is limited by the machine's current inability to explain its decisions and actions to human users. DARPA is playing a key role in increasing explainability by running an explainable AI program to create a suite of ML techniques that:
  - Produce more explainable models, while maintaining a high level of learning performance (prediction accuracy).
  - Enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners.

Others initiatives include:

- Institutions, such as UC Berkeley, UCLA, MIT, Oregon State, Rutgers, SRI institute, PARC and others, are introducing explainability solutions that examine certain aspects of deep neural nets (DNNs), to interpret DNN representations, and to experiment with the learned model. [1]
- LIME (or Locally Interpretable Model-Agnostic Explanations) is an open-source technology that analyzes the input of data samples and observes how the predictions change. [2]
- SHAP (or SHapley Additive exPlanations) [3] is an open-source technology unites six different methods for local explainability. [4]
- Many research papers on explainability are published on arXiv. [5]

Recommendations

Data and analytics leaders should assign responsibility and focus on the following aspects of AI explainability:

- Include an assessment of AI explainability features when assessing analytics, business intelligence, data science and machine learning platforms.
- Use the features provided by selected vendors to build trust and expand your adoption of autogenerated models and insights.
- Assess the trade-off between accuracy and explainability on a use-case-by-use-case basis.
- Develop governance approaches that decide when explainability is necessary, and guidelines that assess the trade-offs.
- Understand that not all AI solutions require explainability, but those that affect user adoption, regulation and risk probably will.
- Establish accountability for determining and implementing the levels of trust and transparency of data, algorithms and output for each use case.

Recommended Reading

"Build Trust With Business Users by Moving Toward Explainable AI"

"Build AI-Specific Governance on Three Cornerstones: Trust, Transparency and Diversity"

"Seek Diversity of People, Data and Algorithms to Keep AI Honest"

"Predicts 2019: Digital Ethics, Policy and Governance Are Key to Success With Artificial Intelligence"

Trend 8: Blockchain in Data and Analytics

inline image

By 2021, most permissioned blockchain uses will be replaced by ledger DBMS products.

The promise of blockchain is substantial. It offers cryptographically supported data immutability, shared across a network of participants. But this comes with massive complexity. Your internal business processes effectively become shared across your network. And your blockchain-based system won't be your system of record, meaning a huge integration effort involving data, applications and business processes. Lastly, the technology hasn't matured to real-world, production-level scalability yet.

Why Now?

The core value proposition of blockchain, and distributed ledger technologies is providing decentralized trust across a network of largely untrusted participants. This trust is founded on the notion that all transactions, or state changes, are publicly documented, immutable and verifiable. This means that information asymmetry does not exist in a blockchain network because all network participants have perfect information. The potential ramifications for analytics use cases are significant, especially those relying on participant relationships and interactions.

What Does It Enable?

Blockchain technologies address two challenges in data and analytics. First, blockchain provides the full lineage of assets and transactions. Second, blockchain provides transparency for complex networks of participants.

A shared ledger exposes an entire participant network as a graph of interacting, connected nodes.

This makes it possible to identify:

- Which participants are interacting with each other.
- How a first interaction potentially triggers second or third interactions.
- How interactions change over time, and whether they are persistent or transient.
- Who are the most important (and possibly most influential) network participants.

This graph can be analyzed to understand the flow of goods and data through a network and its sensitivity to changing conditions (see "Making Big Data Normal With Graph Analysis for the Masses"). In a supply-chain context, optimization modeling and simulation may also be enriched. While these methods aren't new to analytics, blockchain-based networks greatly simplify data access and consistency across all participants.

Early work applying graph analysis capabilities to a distributed, trusted blockchain network has already been done in areas around auditing, product lineage and fraud analysis. Data sharing and collaboration is an untested — but promising — potential use of the technology.

There are a number of inhibitors to blockchain adoption in data and analytics, including:

- **Inadequate data management capabilities:** Blockchain is not a stand-alone data store. Due to block size limitations (how much data a block can carry) and scaling limitations in other types of distributed ledger, large datasets must be stored "off chain" in some other data store. The block representing the data holds both the checksum and the uniform resource identifier (URI) to the data asset. The checksum guarantees the data hasn't been tampered with, while the URI is the pointer to the actual dataset.

- **Data life cycle management is rudimentary:** The pointers to data can be revoked or updated, and the off-chain files updated, but the actual data assets cannot be revoked once downloaded or otherwise claimed. In other words, there is no digital rights management integrated with a blockchain implementation. At least one firm, BigchainDB, is currently working on this challenge.

- **Limited support for analytics beyond Bitcoin:** While blockchain is a promising technology for complementing data and analytics, this work is still in its infancy. Current blockchain-based analytics efforts focus on detecting fraud in the Bitcoin blockchain, primarily money laundering.

- **Technology maturity and interoperability:** It will be several years before four or five major blockchain technologies become dominant. Until that happens, technology end users will be forced to integrate with the blockchain technologies and standards dictated by their dominant customers or networks. This includes integration with existing data and analytics infrastructures. The costs of these integrations may outweigh any potential benefit from blockchain.

How Does This Impact Your Organization and Skills?

Organizations adopting blockchain will need to recast their existing centralized business processes in the context of a distributed computing environment. Traditional skills around data, application and process integration will be essential to successful blockchain evolution. Enterprises will also need to develop blockchain-specific application development skills, including smart contracts. Complicating this work are challenges around the lack of interoperability between competing blockchain implementations. Until the blockchain technology landscape becomes less volatile, organizations should expect to support multiple blockchain implementations.

Use Cases

Auditing and product lineage: An irrevocable dataset shared across all supply chain participants also extends visibility throughout the network. This increases the amount of surface area that can be monitored, improving traceability and understanding of product provenance. Early work in this area is being pioneered by firms like Everledger, which traces the global movement of diamonds and other expensive assets, and Provenance, which targets the supply of things like fish from Indonesia to Japan.

Fraud analytics: A public distributed ledger potentially enhances fraud analytics and risk management by providing all parties with the same set of data. Fraudulent activity detected by one party could be quickly flagged and distributed throughout the network, notifying other participants before they become victims. Any industry involving multiparty transactions or coordination, like healthcare or insurance, will likely benefit from the fraud analytics possibilities available from blockchains or distributed ledgers.

Data sharing and collaboration: When external data is brought into an internal process, it must be treated as completely unverified and untrusted data. This is a reasonable (if somewhat paranoid) approach when data originates from a foreign governance model. The content and context of the data is outside your control, impacting how data is described, shared and used (see “Apply Gartner’s Information Capabilities Framework to Achieve Algorithmic Business Success”).

The guarantees of irrevocability and consistency provided by blockchain implementations, and supported by a shared or self-describing data model, allow you to treat external data as if it originated internally. Data creation and modification can be tracked to its source, offering guarantees on data lineage and provenance.

Recommendations

- Position blockchain technologies as supplementary to your existing data management infrastructure by highlighting the capabilities mismatch between data management infrastructure and blockchain technologies.
- Explore unique use cases in partnership with business leaders by forming a pilot team to exploit the strengths of blockchain in areas of immutability, fault tolerance and transactional transparency.
- Design for flexibility. Given the early, volatile state of blockchain technologies, limit your long-term commitments by introducing abstraction layers around vendor-specific platforms when possible.

Recommended Reading

“Building Blockchain Into Your Data and Analytics Program”

"Debunking the Top 3 Blockchain Myths for Data Management"

"Amazon QLDB Challenges Permissioned Blockchains"

Trend 9: Continuous Intelligence

inline image

> **By 2022, more than half of major new business systems will incorporate continuous intelligence that uses real-time context data to improve decisions.**

Continuous intelligence combines data and analytics with transactional business processes and other real-time interactions. It leverages augmented analytics, event stream processing, optimization, business rule management and ML.

Why Now?

Organizations have long sought real-time intelligence, and they have been able to acquire systems that provide this for a limited set of tasks. However, it is finally practical to implement these systems — what Gartner calls continuous intelligence — on a much broader scale because of these technology advances:

- Increasingly available augmented analytics, including ML, AI and stream analytics; decision management software; and time-series DBMS software.
- Low-cost, high performance, CPUs, GPUs, memory, storage, networks, cloud computing and mobile devices.
- Inexpensive sensor data from ubiquitous Internet of Things (IoT) sources.

What Does It Enable?

Continuous intelligence is the combination of three things:

1. **Situation awareness** is real-time or near-real-time, based on continuous (always on) ingestion of streaming data.
2. **Proactive behavior** means that the system can push alerts, update dashboards, or trigger automatic responses when it detects a situation that requires attention (it's not just reactive, waiting for a person or application to inquire or pull information).

3. **Prescribing behavior** means that the system tells you what to do. It provides decision support for a human decisions or decision automation for lights-out processes (in other words, it's not just notifying you of what's happening and making you figure out an appropriate response).

Continuous intelligence improves the quality and precision of a wide variety of operational decisions because it incorporates more kinds of trusted data into the algorithms that are used to compute decisions. It is relevant to real-time and near-real-time decisions where there is a benefit to having an understanding of the current situation (or events in the past few seconds or minutes).

Systems are able to process high volumes of data quickly, shielding people from overload. They are able to apply rules and optimization logic to evaluate far more options than a person could consider in the available time.

How Does This Impact Your Organization and Skills?

Continuous intelligence has a bigger impact on data and analytics teams than most of the other trends outlined here because it directly affects transaction processing, customer-facing, logistics, B2B and other operational systems. Data and analytics teams must work hand in hand with application architects, application developers, business process management analysts, and business analysts to design, build, deploy and maintain the systems.

Use Cases

Rudimentary, stand-alone versions of continuous intelligence are already ubiquitous. The navigation system in your mobile device is a continuous intelligence system that provides real-time advice on what route to take.

In general, there are two types of continuous intelligence systems:

- **Proactive push systems:** People are overloaded with information, so they need management by exception. Always-on, continuous intelligence (monitoring) systems run all day, listening to events as they occur. When they detect a threat or opportunity situation that requires a response, they update dashboards, send alerts, or trigger an automated response. These are proactive — a push — because the system initiates the response. They might kick off a process or send a control signal to a machine. Real-time dashboards provide situation awareness.

- **On-demand:** All continuous intelligence systems are proactive in some of their processes by definition. However, virtually all are also reactive (on-demand) for some other processes. On-demand, real-time analytics are invoked when a business process reaches a decision point, or when a person chooses to trigger a process or decision (such as a loan approval or a next best offer).

Major business benefits are emerging from corporate usage scenarios of continuous intelligence including:

- Turning 360-degree views of customers into real-time 360-degree views for more precise and effective customer offers and improved customer support.
- Implementing condition-based, predictive maintenance to extend the life of machines while reducing wasteful, premature replacement of parts.
- Providing enterprise nervous systems for airlines, railroads, trucking companies and shipping fleets that monitor and optimize resource scheduling decisions while improving customer satisfaction.

Recommendations

Data and analytics leaders should:

- Work collaboratively with application architects, application developers, business process management analysts, and business analysts to develop a centralized plan and design, build, implement and maintain the systems.

Recommended Reading

“Innovation Insight for Continuous Intelligence”

“Building Your Continuous Intelligence Capability for Digital Transformation”

“Make Your Customer Engagement Hub Real Time With Continuous Intelligence”

Trend 10: Persistent Memory Servers

inline image

> **By 2021, persistent memory will represent over 10% of in-memory computing memory GB consumption.**

Why Now?

Today, most database management systems make use of in-memory database structures, both row-based and/or column-based. In most cases, memory size is restrictive, and all data cannot be located in memory. Either the DBMS software decides what is kept in memory, or it is userdefined. Some servers do exist with large memory (up to 64 terabytes) however, these require multiple processors and therefore are very expensive.

Intel Optane DC Persistent Memory (based on the 3D XPoint memory technology) will become available to the public as directly addressable memory in mid-2019, after considerable delay. Although it does not replace the need for DRAM, it is fully addressable memory in large capacities (currently up to 512GB) in a single NVDIMM. As with all new memory technologies, it will take several years for the price of the hardware to achieve scale, and will go through several iterations to reduce costs further. It will also take several years for DBMS vendors to modify their software to take advantage.

Many DBMS vendors are experimenting with persistent memory and several (including Aerospike and SAP HANA) have already made the necessary modifications. Although the software is available today, it is only usable on test systems from selected hardware vendors and on the Google Cloud Platform.

What Does It enable?

The amount of data available and required for modern systems (across all areas of IT) is growing rapidly and the urgency of transforming data into value in real time is growing at an equally rapid pace. New server workloads are demanding not just faster CPU performance, but massive memory and faster storage. Historically, DRAM has been the costly but reliable byte-addressable memory solution, but it still lacks the economics of the much cheaper and denser (but slower) nonvolatile NAND flash memory used as block-addressable storage. This new persistent NVDIMM does not replace DRAM, but is used together to optimize software environments.

Intel DV Optane persistent memory has two modes of operation:

1. **Memory mode:** Plug and play as a larger memory pool.
2. **Application direct mode:** Persistent memory pool, requiring application modification and optimization.

In memory mode, large memory sizes are available with no changes necessary to the software. However, memory mode does not include persistence. Although far easier to use (as no modifications are necessary) it loses the benefits of persistence for high availability (HA), as it remains volatile.

In application direct mode, large memory sizes are also available and are persistent, realizing the true benefits of persistent memory. This mode requires vendors to modify their software to take advantage of persistent memory. This will have a profound effect on HA, disaster recovery, and how a system restarts after a failure.

How Does This Impact Your Organization and Skills?

There is little or no change to the skills required to use persistent memory as necessary changes are within the DBMS software. Over time, as the cost drops, and as servers with persistent memory become widely available, the use of in-memory DBMS will grow. The impact on the organization will be in areas of consolidation, more-efficient HA for in-memory systems, higher performance, and the enabling of new architectures, such as hybrid transactional/analytical processing (HTAP).

Use Cases

The use cases for persistent memory are many and varied, including:

- **Virtualization:** Most software virtualization (such as VMware) requires memory. Persistent memory will allow larger, more-efficient virtualization environments. This will be true not only for server virtualization, but also for desktop virtualization systems.
- **DBMSs and data grids:** Data grids are already in-memory architectures, and DBMSs are increasingly using in-memory techniques. Persistent memory will allow most or all data to be in-memory, not only boosting performance but also simplifying HA while reducing the restart implications for HA systems.
- **Analytics:** Many analytics vendors today make use of in-memory structures to increase performance. With persistent memory, these systems will be able to increase the amount of data to keep in memory, and therefore increase the performance even more. This will be especially true of augmented analytics using ML algorithms where massive amounts of data are required.

Recommendations

Data and analytics leaders interested in data management solutions should:

- Ask your software providers to identify what they are doing to take advantage of persistent memory, and when it will be available.
- Evaluate the use of persistent memory servers (as they become available) in your infrastructure for simplification of HA architecture, increased use of virtualization, server consolidation and potential cost savings.

Recommended Reading

"Determining the Data Center Opportunity Created for 3D XPoint Persistent Memory"

Nontechnology Trends That Are Critical to Success

Data and analytics are no longer peripheral to the business. They are core to how organizations serve their customers, and optimize business processes, and are the foundation of new transformational business models and revenue streams. To achieve broad business impact, however, data and analytics leaders must extend these investments beyond individual departments and projects to empower everyone in the enterprise and beyond.

Focused, business-driven leadership, cultivating a data-driven culture, and finding the right organizational model will all play pivotal roles. This means tighter collaboration than ever with teams of people across the organization and beyond your borders. Growing the size and the reach of your data and analytics teams, bringing more of the right skills into those teams, and engaging more roles in a distributed way across the business, will be a necessary condition of success.

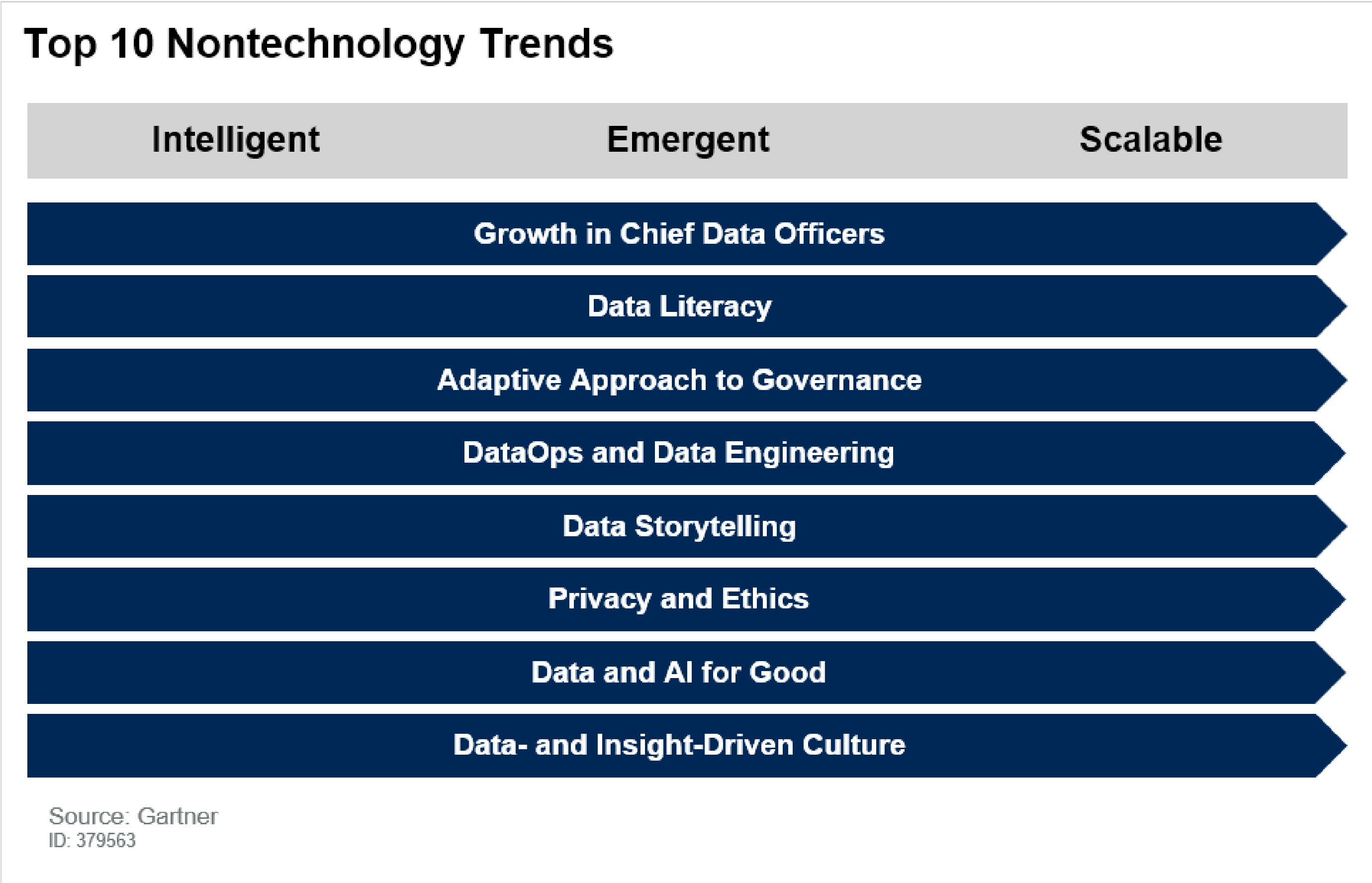
Specific investments must be made to create new roles and responsibilities (such as the chief data officer) to link data and analytics investments with strategic business outcomes and value. Building new skills competencies in data science, machine learning, artificial intelligence, data engineering — as well as data literacy and fluency for everyone in the organization — is the new cultural imperative and organizational success factor in the digital era. It will be necessary to establish new ways of working and new data-driven approaches that exploit complex and diverse data assets, as well as the kind of diverse thinking that sparks creativity and innovation.

Importantly, making analytics pervasive means that you will have to apply best practices from initial pockets of success in data and analytics initiatives more broadly to all parts of the business. Investing in data and analytics as a strategic priority aligned to line-of-business priorities and outcomes, and pushing it out to all corners of the business, is an imperative. Delivering on most digital business goals and objectives will depend on it.

Data and analytics leaders should focus and invest equally in nontechnology trends as they do in technology-focused trends.

Figure 2 shows nontechnology trends that will be critical to successful data and analytics program success.

Figure 2. Nontechnology Trends That Are Critical to Data and Analytics Program Success



Recommended Reading:

"The Future of Data and Analytics: Tales and Trends From the Center to the Edge"

"Leadership Vision for 2019: Data and Analytics Leader."

"Getting Started With Data Literacy and Information as a Second Language: A Gartner Trend Insight Report"

"Fostering Data Literacy and Information as a Second Language: A Gartner Trend Insight Report"

"Toolkit: Enabling Data Literacy and Information as a Second Language"

"Survey Analysis: Third Gartner CDO Survey — How Chief Data Officers Are Driving Business Impact"

"How Data and Analytics Leaders Can Create Effective Influence and Communication Strategies"

Evidence

[1] "Explainable Artificial Intelligence (XAI)," (https://www.darpa.mil/attachments/XAIProgramUpdate.pdf) DARPA (pdf)

[2] "LIME," (https://github.com/marcotcr/lime) GitHub

[3] “SHAP,” (https://github.com/slundberg/shap) GitHub

[4] “A Unified Approach to Interpreting Model Predictions,” (http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions) NIPS Proceedings

[5] “Explaining Explanations to Society,” (https://arxiv.org/pdf/1901.06560.pdf) “Interaction Design for Explainable AI,” (https://arxiv.org/ftp/arxiv/papers/1812/1812.08597.pdf) “Automated Rationale Generation: A Technique for Explainable AI and its Effects on Human Perceptions” (https://arxiv.org/pdf/1901.03729.pdf)