NUCLEUS
RESEARCH

SIX TRENDS IN DATA AND ANALYTICS FOR 2022

ANALYST
Daniel Elman, Alexander H. Wurm

THE BOTTOM LINE

Reflecting on vendor briefings and customer interactions over the year, Nucleus has identified several key factors driving vendor investment and user adoption in data and analytics technologies. On the vendor side, we see consolidation as companies look to control as much of the analytics pipeline as possible, particularly with data integration, preparation, governance, and management technologies to complement core analytical functionality. Users are driving demand for easy-to-use interfaces such as graphical and natural language to control high complexity analyses and other workflows, as well as increased centralized control over data within organizations and lower-latency options for data processing both at the edge and within data repositories.

OVERVIEW

Nucleus has identified several key trends that stand out in the data and analytics world as we look ahead towards 2022. The past two years were distinguished by large-scale digital transformation projects where companies of all sizes looked to modernize IT systems to become more resilient, flexible, distributed, and data driven. The rapid uptick in adoption has moved the space forward significantly but also has pointed out some challenges that need to be addressed for customers to maximize the value and success of their analytics projects. For example, the challenge around data management and the appropriate permissioning of data to allow for exploration and value-add analysis without becoming overwhelming to the end-user or creating a security risk became increasingly significant once analytics became adopted at a large scale.

> **In the analytics gold rush, supporting technologies like data integration and governance are the pickaxes and sieves.**

The challenge to the general business population on analytical techniques and best practices persists. Still, significant progress is beginning to bear fruit as we see large-scale analytics adoption outside of traditional data science/IT/analyst roles and data becoming more embedded in everyday user experiences. Vendors are aiding in this with educational materials and informational content, but also with increasing investment in usability enhancement such as low/no-code capabilities, explainability enhancements to help users understand outputs, more natural user interfaces such as graphical and natural language, and other improvements to self-service. This is significant as self-service analytics has been evangelized for years, but in practice, analytical duties were carried out by the same people with technical expertise in analytics. We see this beginning to change with analytics as working with data becomes a fixture in job roles across departments and functions.

The business world has embraced analytics and data-driven decision-making as best practice, so the momentum toward continued adoption and increased investment in the area does not look set to slow down any time soon. Even with the economy becoming more uncertain, which often leads to a slowdown in business spending, the value of organizing and making sense of data to drive businesses through challenging times is becoming recognized outside of the IT and financial wings. We expect large-scale adoption and modernization initiatives to continue, driven by increased demand forself-service/usability-focused capabilities, embedded analytics capabilities (including the capabilities needed for managing the integrations and other backend aspects), and enabling more specialized analytics to optimize total costs and workload priority level.

TRENDS

Nucleus has identified the following six trends in data and analytics going into 2022: increasing vendor ability to deliver self-service analytics; direct correlation between the growth of embedded analytics, data integration, and data governance technologies; the continued obsolescence of department-level data organization and analytics; increased proliferation of edge and real-time analytics outside of the manufacturing sector; expanded no/low-code capabilities for higher-level analytics tasks; and increased investment in data explainability capabilities.

INCREASING VENDOR ABILITY TO DELIVER SELF-SERVICE ANALYTICS

Vendors have been offering self-service analytics tools for years without seeing the long-anticipated wave of adoption. That is now changing as vendors' offerings are reaching a level of maturity and high usability that is driving a marked uptick in adoption. Graphical and natural language interfaces are rapidly superseding code and spreadsheet-based interfaces as the industry standard in analytics technology. This may result in a slight trade-off in customization capabilities for high-level users, but it also enables a more significant improvement in ease-of-use for the majority of users that lack technical statistics/analytics expertise. Additionally, vendor campaigns to educate end-users with targeted content, free training materials, and programs and university partnerships are beginning to bear fruit as more businesses embrace analytics to drive decision making and reveal value-add insights. We are beginning to see widespread adoption of natural language interfaces for searching and querying data; this is technology that significantly improves usability and is finally reaching enterprise-level maturity after several years since the first iterations went to market. Vendors that are successful in delivering additional usability-enhancing tools will be differentiated as self-service takes on greater value to business decision-makers.

DIRECT CORRELATION BETWEEN GROWTH OF EMBEDDED ANALYTICS, DATA INTEGRATION, AND DATA GOVERNANCE TOOLS

It's old news now that embedded analytics is a major trend to watch, but to serve the embedded analytics "gold rush," there are supporting technologies needed to play the roles of "pickaxe" and "sieve." Data integration and governance tools are those technologies, as modern data management, integration, and governance strategy and infrastructure are necessary to enable embedded analytics at scale. The vendors that provide these technologies to orchestrate and prepare the analytics pipeline will continue to show strong growth. Considering analytical capabilities added to an application, to produce value-add analytics to the business, application data is often needed to be merged and analyzed together with data from other areas of the business (i.e. aligning annual sales from

CRM system with invoiced revenue from ERP system). Tools are necessary to facilitate and secure data sharing and transfer between disparate sources; vendors that serve these spaces are showing rapid growth that should continue as analytics adoption is nowhere near its peak yet. Increasingly we are seeing analytics vendors looking to build out comprehensive platforms that include data governance and integration capabilities to unify the analytics pipeline on a single platform. This serves to reduce complexity and often also total cost for customers once the technologies are successfully integrated. but there may be some early hiccups as vendors who add these capabilities by acquisition work to merge technologies and operations.

CONTINUED OBSOLESCENCE OF DEPARTMENT-LEVEL DATA MANAGEMENT AND ANALYTICS

It is not uncommon for large corporations to have three, five, or even more different analytics tools serving different departments or areas of the business. Of course, each of those tools would generally be used to analyze different data, and cross-functional analytics would require merging data from disparate databases or data warehouses before conducting the analysis for the desired report. Experimentation and exploration were severely limited, as there were no resources to support that many requests for data from different systems in a timely fashion. In addition to the limiting effect on the value of analytics for high-level decision making, supporting so many disparate systems and their associated data warehouses, security, and prep tools is costly. It creates silos of institutional knowledge within companies for different systems. As the cost of large-scale data management and integration decreases, companies are coming to understand the necessity of a modern data strategy to drive decisions, and eliminating department-level systems. By breaking down silos and housing data in centralized warehouses with more modern governance strategies allows business analysts to experiment with and explore permissioned data without being restricted by departmental procedures. Different analytics and visualization tools will likely still be employed to support various use cases or experience levels among end-users., But the underlying infrastructure will be consolidated to reduce costs, eliminate data silos, and enable more seamless analytics at a company level rather than the department level.

PROLIFERATION OF REAL-TIME AND EDGE ANALYTICS OUTSIDE OF THE MANUFACTURING SECTOR

 In this context, Edge Analytics is defined as analytics performed on devices or endpoints where data is collected. It has been adopted most significantly in the manufacturing space where industrial hardware is controlled by on-site infrastructure, often with a single specialized server or a computer running a machine or row of machines. Data analysis is

performed on the device immediately as it is collected, eliminating the latency inherent when sending the data back to a centralized repository for bulk analysis. This type of application is gaining adoption with increasing levels of automation used to operate and monitor factories and industrial sites. The ultimate goal is to completely automate these facilities, and we anticipate organizations outside of this space to leverage the best practices and existing infrastructure to enable increased edge analytics adoption in industries such as telecommunications, oil and gas, and agriculture.

Keeping with the goal to minimize or eliminating latency, in-database analytics and query accelerators are seeing strong growth. As companies mature in their analytics strategies, most develop various priority levels for different analytical workloads as a means of optimizing costs and resource consumption, with the highest requiring immediate analysis and response and the lowest requiring very infrequent batched analytics or reporting. With real-time and ultra-low latency analytics being among the most expensive and complex workflows to implement and manage, all but the most well-funded and technologically savvy organizations had generally looked to outsource or avoid large-scale investment in such processes. We anticipate this will change as new capabilities and services from vendors that lower the cost and complexity of real-time/low-latency analytics become available.

EXPANDED LOW/NO-CODE CAPABILITIES TO DEMOCRATIZE HIGHER-ORDER ANALYTICS

As organizations seek to apply higher-order analytics to a wider range of departments and business practices, organizations have experienced increasing challenges resulting from a lack of technical savvy and data science knowledge. With every step in AI/ML-based analytics initiatives becoming prohibitively complex to citizen data scientists and business users, analytics vendors will focus on building out no-code tools capable of demystifying complex models and automating time-consuming data science processes from data preparation and model refinement to visualization building. This focus on no/low-code analytics allows users of all skill levels to discover data-driven insights and drive efficiencies throughout their organization by easily translating operational data into digestible dashboards and visualizations. This focus allows customers to harvest greater value from their data by limiting technical bottlenecks to higher-order analytics, driving more widespread analytics utilization. As this trend matures, analytics vendors can deliver more functionally complex models and analytics while maintaining an intuitive ease-of-use critical to democratizing analytics.

INCREASED INVESTMENT IN DATA EXPLAINABILITY CAPABILITIES

Although prescriptive and predictive capabilities continue to reach new heights in terms of relevance and accuracy, without an understanding of what processes drive observed predictions and insights, organizations experience difficulties using these capabilities to optimize their business strategy. To eliminate this pain point, many analytics vendors are developing tools and capabilities to enhance their explainability. One common innovation involves natural-language data storytelling, which automates the final stage of the analytics process: communicating findings with digestible and technically-accessible delivery. Another focus of development involves machine learning-powered root cause analysis, which highlights the most likely causes of key insights or changes using a probability statistics-based approach. By leveraging this enhanced explainability, organizations can limit potential sources of oversight common to many higher-order analytics, such as implicit bias or overreliance on certain metrics, which can limit model flexibility and effectiveness. As vendors continue developing richer explainability alongside their analytic content, organizations can expect to attain greater value from their data via improved reliability and strategic actionability.