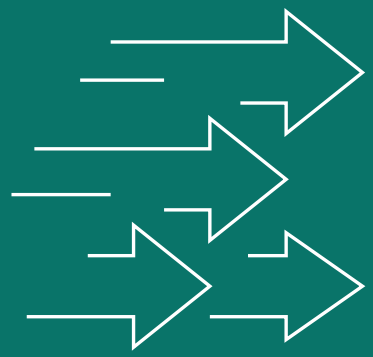
EBOOK

Exasol
The analytics database



THE STATE OF PLAY - PREDICTIONS FOR 2020

WHAT WILL BE THE BIG ISSUES IN 2020?'

Customer demands will change. Products may offer different functionality. New use cases for technologies will be uncovered. A lot will happen over the course of 12 months, but there is one thing that remains constant - the need to stay on top of what is happening in the market. That is why we have compiled our insights into the most important issues of 2020.

From data strategy to the cloud, right through to data science and AI, this paper covers the market trends that we will be keeping a close eye on over the next year and what they mean for your business. One thing is for certain, it's going to be another busy year. We're looking forward to taking it on with you.

Cloud

On-premises investment by the public cloud providers cements hybrid cloud's dominance

In 2020, hybrid cloud deployments will become the primary focus of the hyperscale public cloud providers.

In the last year alone, the three largest public cloud providers have made investments in on-premises capabilities. AWS, for instance, announced an on-premises version of its IaaS and PaaS services called AWS Outposts. It is scheduled to ship in late 2019. Microsoft announced an expanded catalogue of on-premises Azure Stack services, including software-only and converged hardware packages. Similarly, Google added the ability to centrally maintain and manage Kubernetes clusters as a primary feature of its Anthos microservices platform.

This move to hybrid recognizes that the journey to the cloud hasn't been as seamless as previously expected for certain organizations, especially since not every workload can or should run in the public cloud.

For enterprises transitioning to the cloud, a hybrid deployment is a must because it enables a smoother migration and allows the enterprise to set its own pace. The hybrid model keeps risks and costs under control while still providing scalability, infrastructure provisioning and efficiency and effectiveness benefits.

Data warehouse modernization projects utilizing containers will expand rapidly

To date, cloud has primarily been used to build new apps and rehost infrastructure. In 2020, we expect enterprises will increasingly leverage cloud to modernize existing business apps, processes and data environments.

We expect more data warehouse modernization programs to be deployed in a containerized hybrid or multi-cloud environment, helping organizations become more agile and deliver a more frictionless deployment and management experience. This investment will be driven by the need to speed up data accessibility, improve the timeliness of insights, lessen support and maintenance costs, and future proof existing data environments.

A container-based approach allows organizations to get the benefits of "cloud-native" as quickly as possible in the enterprise. Containers can help these companies manage data in a hybrid cloud setup in ways that other approaches cannot. Moving data to the public cloud can be extremely risky and expensive.

One of the main reasons for this is data gravity: a tendency to leave data where it currently resides. Containers often equate to agility, but they also increase portability. By building out services and data stores within containers, businesses can more easily move them all—or some of them at a time as part of a migration strategy—to the public cloud. Containers also provide flexibility in terms of maintaining a similar architecture across all on-premises and cloud applications, but with the ability to customize rollouts in geographical regions.

Cost optimization for cloud data warehouses becomes a growing priority

As organizations continue their journey to the cloud, CIOs are finding that a growing portion of their tech budgets are going to cloud subscriptions. While cloud brings organizations many benefits in terms of business agility and on-demand computing services, it also creates new problems not least the possibility of wasted or inefficient cloud spend. Enterprises are finding that there is a lot of waste in cloud budgets, including those for data and analytics, as well as ineffective resource utilization. This is why they are looking to better plan, budget and forecast spending requirements for cloud consumption.

In the case of a cloud data warehouses, costs are often based on resource usage patterns, including how much data is stored, queried and inserted. In cloud terms, storage is relatively cheap, but compute costs can start to become prohibitive if the complexity of analytics functions increases. In particular, operations such as aggregating data or raw fact-based queries push up compute resource and costs considerably. Some providers also have pricing models that cap costs at a fixed level, and therefore they cap the amount of resources employed to process workloads.

This means that organizations may need to avoid scanning raw data and limit expensive operations, such as joins, in order to maintain costs within budget.

In 2020, we expect that cost efficiency will be a critical piece of successfully migrating from legacy on-premises data platforms to cloud data warehouses. Organizations will be looking for cost predictability and assurance to ensure they will not be penalized for becoming a more data-driven organization.

Data Strategy

"Data literacy" will take on the buzz of agile methodologies

In 2001, the agile methodology began its advance from the world of software development to a widespread project management initiative, touted as the way

to cope with continuous change. It combines disciplined execution with continuous innovation in ways that energize employees involved.

One of the tenants of agile is the ability to break down organizational siloes. But, in a time where innovation is (or should be) highly technical and data driven, without the proper language in which to communicate, those from business focused functions will not contribute to the conversation in the same way as the data savvy.

In 2020, data literacy will emerge as mission critical for organizations looking to constantly innovate around the growing volumes of data collected. Diversity of thought is accepted as a key to high innovation capabilities, hence the emergence of tools that empower everyone across organizations with the ability to ideate and group around innovations they support. Data literacy will be seen as the facilitator of this dream state. Non-technical employees will be able to describe their proposals to the data scientists and understand barriers to the success of their ideas.

People will stop talking about big data, but enterprise data strategy will still be a top priority for enterprises, proven by the growth of the CDO role

86% of the participants in Exasol's 2019 Cloud Survey report having a CDO, proving and solidifying the status of data strategy as a mission critical initiative. But, in 2020, we will see the term 'big data' drift away as companies mature beyond this buzzwordy lexicon. Instead, they will have use-case specific terms to frame their data analytics efforts. For example, instead of saying "We do big data", they will say "We're working with customer demographics, credit card statements, transactions and point of sale data, online and mobile transfers and payments, and credit bureau data to discover similarities to define tens of thousands of micro-segmentations in the customer base. We then build 'next product to purchase' models that increase sales and customer retention."

The proportion of CDOs in financial services institutions will surpass other industries as the role transforms rapidly

Appointments of CDOs have risen dramatically in the last two years, but none more so than in Financial Service (FSI) organizations. In 2020, we expect CDO appointments to be even more prevalent in FSI over other industries as they formalize and commit to implementing an office of the CDO. This is broadly driven by a number of factors including FSI companies' acceptance and broader understanding of the role; the ability to demonstrate impact and value; continuing and growing budgets; and increasing positive engagement across the C-suite.

As one of the most analytically advanced sectors, the CDO role in FSI firms is transforming, moving from its original technical roots to encompass a broad agenda that spans data management, analytics, data science, ethics, and digital transformation. More importantly, CDOs are using their high profile and pivotal role to act as change agents

for the business, moving from a focus on risk mitigation toward business impact and value realization. The CDO's primary responsibilities have focused on regulatory compliance and operationalizing regulatory mandates. Yet, leading FSI companies are using the position as an enabler of business insights, strategies, and innovation such as developing value-adding data services that are enabled by the new foundational processes and policies.

Retail Customer Experience

Retailers will increase investments in personalization to improve conversion rates and customer lifetime value

According to McKinsey, personalization will be the prime driver of retail success within five years. This will be driven by advances in data and analytics that personalize critical touch points in a way that best drives value for the customer as well as the retailer.

Personalization doesn't work if businesses don't have the means to understand the needs of high-value customers on a continuous basis. This requires organizations to continue to invest in data and analytics that can pool and analyze structured and unstructured data, algorithms that can identify behavioral patterns, intent signals and customer propensity, and analysis capabilities to deliver recommendations or personalized content to systems or front-line staff. Core to any personalization effort will be garnering a 360 view of the consumer and/or customer.

While most personalization efforts to date have focused on digital channels, in 2020 we expect leading retailers to incorporate more in-store personalization. For example, leading retailers will rely on insights from advanced analytics to generate personalized product recommendations for specific customers automatically in-store. Others will use AI-enabled tools to make real-time customer recommendations and some will use facial recognition, location recognition, and biometric sensors to make the shopping experience faster and easier for consumers

One of the biggest challenges retailers face is how to deliver a higher level of personalization without breaking a customer's trust and privacy. In 2020, we expect to see retailers future proof their customer privacy tactics by giving customers more meaningful choices about the data they share through preference management and being more proactive in illustrating the value this may bring in return.

Data Science & AI

Widespread/general adoption of artificial intelligence will only appear in the most advanced firms

In 2020 we will continue to see AI investments gather speed, but for most companies this will only be in narrow use-cases that allow them to pick off the low hanging fruit in their industries. For example, CPG (Consumer Packaged Goods) firms are more likely to invest in physical robotics for the factory floor, and telcos will invest in customer-facing virtual agents.

The top performers will look to use AI to generate value more broadly across business lines and functions. For example, sentiment analysis can be used not only to gain a deep understanding of customer complaints, but also to inform marketing content and micro segmentation for sophisticated sales strategies. Shared sentiment around an issue will stand alongside spending patterns to determine next-to-buy models and deep marketing personalization.

A barrier in place to broad adoption of AI is a lack of training data. For large tech firms like Google, Apple, and Amazon, gathering data is not an arduous task in comparison to most companies. Because of the breadth and depth of their products and services, they have a near-endless supply of diverse data streams, creating the perfect environment for their data scientists to train their algorithms. For smaller companies, access to comparable datasets is limited or simply too expensive.

Growing availability of synthetic datasets will satisfy the demand for data

In 2020, this will allow less advanced or smaller companies to make meaningful strides in their AI journey. Synthetic data is data that is generated programmatically. For example, realistic images of objects in arbitrary scenes rendered using video game engines or audio generated by a speech synthesis model from known text. The two most common strategies for synthetic data usage we will see are:

1. Taking observations from real statistic distributions and reproducing fake data according to these patterns.
2. A model is created to explain observed behavior, and then creates random data using this model. It aids in the understanding of the effects of interactions between distinct agents that are had on the system as a whole.

Companies who considered their data storage capacities to be minimal will come to the realization that they need a sophisticated solution to house their synthetic data if they are to compete on the hard-hitting elements of machine learning.

For example, The Pistoia Alliance, a global non-profit that works to lower barriers to innovation in life sciences R&D, announced survey results showing that access to data (52 percent) and lack of skills (44 percent) are the biggest barriers to the adoption of AI and machine learning (ML).

Data science teams will increase use of GPUs to handle massively parallel analytical workloads

It is early days for deep learning and AI. However deep learning offers opportunities for enterprises that need to build applications for recognizing objects in images, identifying voices and creating predictive models from vast sums of data. Deep learning can use regular CPUs, but for serious enterprise projects, we expect more data science teams in 2020 to explore the use of specialized chips such as GPUs that can handle massively parallel workloads to more quickly train and retrain models. Without these specialized chips deep learning would not be practical or economical since the discipline requires significant amounts of compute capacity to process high volumes of data.

GPUs are typically used in two different types of data science workloads:

1. **To train data as part of the model building process where AI engineers and data scientists use deep learning frameworks to analyze historical data.**
2. **Inferencing; this is where models are used in production applications to make predictions on live data.**

Both are suitable candidates for GPU use and help data science teams become more productive by giving the most performant-possible data infrastructure.

Blockchain

Blockchain in conjunction with IoT and AI will deliver value in logistics use cases

Today, supply chain transactions are coordinated across multiple parties, and spread across multiple geographies and legal jurisdictions. Some of these parties will be organizations; some will be individuals; and some may be automated, network-connected devices. Solutions that leverage blockchain in conjunction with IoT, and potentially AI, will cut supply chain costs and boost CX through seamless logistics.

The IoT device tracks and sensors the products, providing answers to questions such as: Where is the product? Was it too hot or cold (particularly useful when transporting food)? Is it genuine? Each connected device acts as a node on the blockchain network, authenticating the transaction and providing an immutable record. The AI is the learning system; it looks at the behavior of the component over time across the ecosystem, correlates with other data points (weather, market data), then makes prescriptive recommendations OR automates decisions. Smart contracts built into the blockchain give the ability to execute code autonomously, which is ideal in an ecosystem environment, when many parties interact but have a limited relationship (and therefore trust) with one another.

However, the utopian vision of enterprises across industries migrating all their processes onto shared blockchains, and interacting on a trusted ecosystem, still seems far-fetched. Blockchain projects that are not narrow in focus continuously run into problems with scaling. Many businesses are still starting with a desire for blockchain, rather than the business problem, which is often solvable without the need for distributed ledger technology.

LONGER TERM PREDICTIONS

Persistent Memory

Hype around persistent memory will ramp up but adoption will remain low

Persistent memory (PM) has the potential to address the growing gap between storage and memory technologies. Specifically, system memory such as DRAM may not be big enough to process and analyze all data in real time and storage, such as solid-state drives, although cheaper can't compete on performance.

In 2020, we expect to see a ramp up in awareness building and marketing around emerging PM technologies such as Intel Optane and ReRam (resistive random-access memory). As a result, data intensive enterprises will investigate and experiment with PM to decide if it offers a significant shift in data management processing and economics. Over the next 12 months however, most initiatives will focus on specialist workloads that require ultra-low response times such as those for IoT and high frequency trading applications.

Two factors will contribute to this decision making. Firstly, there are economic considerations. PM chips such as Optane are slower and cheaper than DRAM chips, but dramatically faster and costlier than SSD chips. Organizations must consider the cost of using PM for data centric workloads in order to determine if it is economically viable. Second, there are architectural considerations as organizations may need to re-evaluate the design of applications to assess what data and in what scenarios it makes sense to take advantage of PM.

Quantum

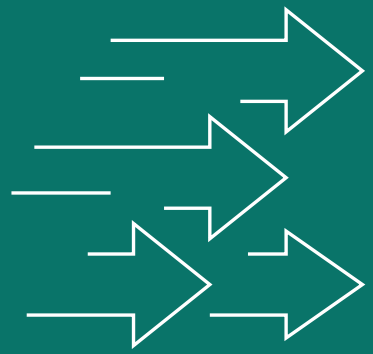
Quantum Computing investment will continue to grow, but won't move out of the research phase for some time

The fundamental function of computers is the ability to store and manipulate information. Current computers manipulate individual bits, which store information as binary 0 and 1 states. Quantum computers leverage quantum mechanical phenomena to manage information. To do this, they rely on quantum bits, or qubits.

The reason for this is that challenges currently exist that we are able to articulate but not solve with today's systems. For problems above a certain size and complexity, we don't have enough binary computational power on Earth to tackle them.

IBM, Rigetti, Google and IonQ all provide public access with open-source tools to real quantum computing hardware. In September, Google's quantum computer succeeded in performing a task that a classical computer could not conceivably pull off, solving an algorithmic challenge in 200 seconds instead of the 10,000 years it would have taken a powerful supercomputer. That being said, current QCs are extremely temperature-sensitive, error-prone and not yet powerful enough to break RSA 2048-bit keys, and probably won't be for another 10 to 20 years. This key has been the gold standard for encryption for a number of years, but quantum computers would render it obsolete.

Exasol
The analytics database



FIND OUT MORE

If you want to read more of our latest industry insights, check out our Insights Blog.