State of the Operational DBMS Market, 2017

Published 25 July 2017 - ID G00325613 - 21 min read

The operational DBMS market has demonstrated little innovation over the last two years. Instead, vendors are largely focusing on competitive catch-up and execution across product lines. Data and analytics leaders need to be aware of these trends as part of their evolving data management strategy.

Overview

Key Findings

- The megavendor database management systems (DBMSs) continue to dominate the market. Although nonrelational DBMS is increasing its share, in 2016 it was only about 3% of the total DBMS market.
- In contrast to the single multimodel product approach gaining traction for on-premises deployment, some vendors offer more than one style of dbPaaS favoring a polyglot model, where services are consumed as point solutions from a wider portfolio.
- The term "NoSQL" has lost its distinction in the market as NoSQL DBMS vendors add early SQL capabilities; the term "nonrelational" is more appropriate.
- Hybrid transactional/analytical processing (HTAP) enabled by in-memory computing (IMC) is still in its infancy, but the benefits of HTAP are already driving adoption.

Recommendations

Data and analytics leaders concerned with information infrastructure modernization and evolving their DBMS strategy should:

- Evaluate the modern data management capabilities (i.e., multimodel, distributed, scale-out, eventual consistency) in existing and well-understood products, especially for distributed use cases, before adopting new DBMS products.
- Analyze a range of dbPaaS offerings from providers — not just those from major vendors and your standard vendor. There are many specialty dbPaaS offerings that bring unique functionality and flexibility.

- Use in-memory DBMS (IMDBMS) capabilities of incumbent DBMS technologies to enable HTAP-capable applications; for example, when retrofitting for HTAP-established relational database management solutions (RDBMS) based applications.
- Lead a nonrelational technology assessment, and ensure that DBAs, system architects and the CTO are involved in vetting and deciding the DBMS that best meet business requirements.

Strategic Planning Assumptions

- By 2018, more than 70% of new DBMS deployments will leverage cloud for at least one use case.
- Through 2020, relational technology will continue to be used for at least 70% of new applications and projects.
- By 2018, less than five independent, nonrelational vendors will be able to grow to sufficient size (more than $200 million) to remain competitive.
- By 2019, the separation of storage and compute designed for cloud DBMS architectures will become the dominant dbPaaS model and also begin to appear on-premises.
- By 2018, at least 75% of HTAP projects will adopt the less disruptive "point of decision" instead of the more powerful "in-process" approach.

Analysis

The operational database management system (OPDBMS) market (see Note 1) returned to revenue growth in 2016 and continues to spawn new entrants. Total DBMS market revenue [1] was $34.4 billion in 2016, representing 7.7% growth year over year. As the OPDBMS market continues to grow, we see the five largest vendors (Oracle, Microsoft, IBM, SAP and Amazon Web Services [AWS]) holding the majority of the revenues at 87.7%. In 2016, AWS moved into the top five with almost 5% of the market and 107% growth year over year. The nonrelational vendors (including Hadoop distributions) account for almost $1.5 billion or about 4.4% of the total revenue. Many of the new entrants and nonrelational vendors offer open-source versions of their products and, therefore, have less revenue than the larger, more mature vendors in this market.

In this research, we examine the major trends in OPDBMS, including the accelerating move to cloud computing and continued growth of open-source products, areas of innovation, and pressures of consolidation taking place in 2017. This research will be of interest to data and analytics leaders, architects and implementers. We cite as evidence the Gartner DBMS Market numbers, [1] Gartner's inquiry service [2] and the Gartner 2016 OPDBMS Magic Quadrant Survey, [3] among other sources.

Major Trends in the OPDBMS market

The explosion of features, and the vendors emerging to implement them, has slowed. The features that initiated the expansion, such as storing new data types,

geographically distributed storage, cloud and flexible data consistency models, have become common.

Lack of Innovation

The OPDBMS market has shifted from a phase of rapid innovation to a phase of maturing products and capabilities. Today, nearly every established or emerging DBMS vendor supports storing new data types, geographically distributed storage, flexible data consistency and cloud. Emerging vendors, having captured their respective early adopter markets, have begun back-filling capabilities around management, security and operational excellence. Also, some of these vendors have begun offering their products as cloud-hosted, managed database platform as a service (dbPaaS). dbPaaS offerings allow these vendors to counter customer concerns around operational skills, as well as to address smaller market segments with a lower cost of sale.

The 2016 OPDBMS Magic Quadrant (see "Magic Quadrant for Operational Database Management Systems") demonstrates this lack of innovation, as the entire set of vendors has moved left since the 2015 Magic Quadrant, leaving a large white space in the Visionaries quadrant. This implies that the OPDBMS market had little innovation and future vision while most vendors have been busy working on execution. Although we believe it will begin to increase in 2017, we do not believe there will be a major change until 2018. This trend is also evident in the "Magic Quadrant for Data Management Solutions for Analytics" and "Magic Quadrant for Data Integration Tools."

Recommendation:

Data and analytics leaders (including database and application managers) should work with CIOs and CTOs to:

- Evaluate the modern data management capabilities (i.e., multimodel, distributed, scale-out, eventual consistency) in existing and well-understood products, especially for distributed use cases, before adopting new DBMS products.

The Nonrelational Opportunity

Nonrelational (formerly NoSQL) DBMSs (see Note 2), including Apache HBase and other offerings shipped in Hadoop distributions (although used more often for DMSA), emerged as a response to changing application requirements. As mobile applications proliferated, there was an increasing need for cost-effective horizontal scale-out, versus vertical scale-up, capabilities. Mobile applications, particularly those that weren't mission-critical, made consistency less important and allowed eventually consistent implementations to be an acceptable option to achieve desired levels of application throughput and availability. Finally, the development overhead required to keep a relational schema in sync with the changing requirements of an application under rapid development became undesirable to developers. This began the shift from schema on write to schema on read.

The nonrelational DBMS vendors had first-mover advantage in this new application space. Established vendors have reacted quickly. Several have added new data types, such as JavaScript Object Notation (JSON), and graph capabilities to existing

products (e.g., EnterpriseDB, IBM, Microsoft, Oracle and SAP). Others introduced new products, acquired capabilities (e.g., Azure Cosmos DB and IBM's acquisition of Cloudant) or simply implemented popular protocols and APIs already in use. Some vendors, such as Oracle, have even added a sharding capability for horizontal scalability.

As nonrelational DBMS vendors pushed into the enterprise, they identified the need for analytics, security, information governance and administration capabilities. Over the past two years, multiple nonrelational DBMS vendors have begun introducing SQL, or SQL-like, capabilities to facilitate business intelligence use cases (e.g., Couchbase, DataStax and MongoDB). Several have introduced, or are introducing, schema management capabilities for more control and governance over data quality and schema evolution. All vendors are actively working on their security as well as administration and management capabilities.

What began as bifurcated market segments are merging into a single DBMS market segment as the different vendor cohorts trade features in response to customer demands. Feature sets are converging rapidly. Hence, we use the term nonrelational instead of "NoSQL" to describe these products and portion of the DBMS market.

We believe that the nonrelational DBMS products will survive in the market for many years to come. Obviously, we cannot predict a lifetime as long as the RDBMS has enjoyed. RDBMS technology today is affordable, scalable, flexible and enterprise-hardened, with widespread availability of skills, tools and products. Those are qualities not to be discarded lightly. In addition, RDBMS technology addresses some issues, such as online transaction processing (OLTP) at scale, that have been decades in development — capabilities not easily duplicated in the short term. We also expect further consolidation in the nonrelational space, as we have seen by several recent acquisitions. [4] This is normal when there are so many products available. Also, as the multimodel, nonrelational DBMS becomes more of a reality, we believe this will also lead to further consolidation as well as further exits from the market. Therefore, data and analytic leaders, architects and implementers should tread carefully when using nonrelational technology today for select applications that require the specific functionality brought by these DBMS engines.

Recommendation:

Data and analytics leaders (including database and application managers) should work with CIOs and CTOs to:

- Lead a nonrelational technology assessment and ensure that DBAs, system architects and the CTO are involved in vetting and choosing the DBMS options that best meet business requirements.

Multimodel DBMSs

Multimodel DBMSs support several types of data model, such as relational, XML and the common nonrelational types (e.g., document, key-value, search and graph), in a single DBMS. They are intended to reduce complexity in the data access layer by using a single language to access different persistent types, although multiple languages can also be used.

Multimodel DBMSs can reduce the complexity of existing portfolios of production systems. They can often more consistently apply auditing, concurrency controls, versioning, distributed data complexity management, points of governance and security. They offer a potential solution to vendor proliferation and complexity, but also create new skills requirements and potentially competing alternatives.

Vendors are adding capabilities to products, because DBMS solutions based on a single persistence type can be ineffective at supporting business use cases that require the convergence of multiple data formats (such as tabular, document, graph, key-value and temporal). Sometimes, it will be sufficient to use the capabilities of a multimodel product; in other cases, a "best fit" approach will be much better.

Recommendation:

Data and analytics leaders (including database and application managers) should work with CIOs and CTOs to:

- Choose vendors whose stability and track record are demonstrable and whose roadmaps are consistent with your planned use cases. Also, consider the traditional DBMS vendors, as they have been adding multimodel capabilities.

Open-Source DBMS maturity

As open-source DBMS (OSDBMS) has matured, its impact on the market has accelerated. The five megavendors whose annual DBMS software revenue exceeds $1 billion are followed by 24 whose annual revenue exceeds $10 million. Of those, 15 base their product on open-source projects, typically marketing a "pure," often free, community edition and an extended version with extended proprietary features for enterprise-class use. Since 2014, the collective revenues of these OSDBMS vendors have nearly doubled. Average growth across the cohort for 2016 was 45.8%. Although, several vendors grew much more, and revenue of Actian, whose Ingres is the oldest offering, declined. Note that a substantial additional amount of OSDBMS revenue is not broken out from the portfolios of megavendors. Amazon Relational Database Service, Microsoft Azure HDInsight and Oracle MySQL, among others, generate revenue not counted separately here.

Nonrelational offerings are the majority in this group — eight of these vendors are based either on nonrelational engines or Hadoop. Relational engines continue to be strong, however, with multiple emerging Postgres products competing with EnterpriseDB, and a similar phenomenon occurring in the "MySQL-compatible" space as cloud offerings and MariaDB compete with Oracle's original. They are likely to do what Oracle will not by adding PL/SQL compatibility and other features that will make their offerings more competitive with the Oracle Database product.

Recommendation:

Data and analytics leaders (including database and application managers) should work with CIOs and CTOs to:

- Add an open-source DBMS to your organization's DBMS standards for new uses and to replace a commercial RDBMS. The functionality of the OSDBMS may be sufficient for existing applications.

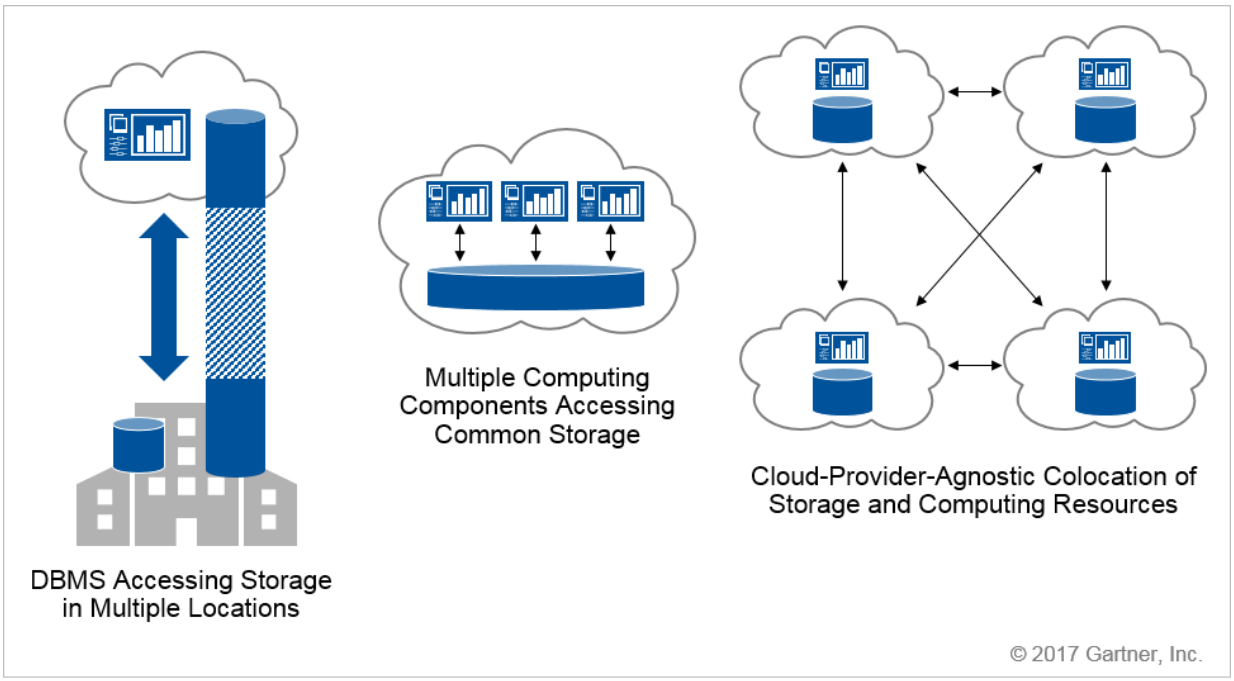
Database Platform as a Service (dbPaaS)

Gartner forecasts show dbPaaS (see Note 3) tripling in revenue over the next five years, with a 27.7% compound annual growth rate (CAGR) from 2015 through 2020 (see "Forecast: Public Cloud Services, Worldwide, 2014-2020, 2Q16 Update"). Further evidence of this growth is AWS's rapid movement to fifth position in overall DBMS revenues with greater than 5% of the market. [1] dbPaaS growth is 42% faster than the overall public cloud growth, which is doubling at 16.1% CAGR, over the same period. With the recent growth of vendors and offerings in the dbPaaS market — especially with the entry of IBM, Microsoft, Oracle and SAP — interest in dbPaaS has increased sharply, as evidenced by the number of Gartner client interactions on the subject. Historically, dbPaaS has often been of interest as an experimental platform and one used for small, third-party software vendors. The increase in maturity and numbers now makes dbPaaS of interest across all sizes of organizations. We believe dbPaaS now has attracted the interest of CIOs, CTOs, database managers and application development managers — regardless of whether there is a desire to locate production applications in the cloud or not.

Gartner believes the number of dbPaaS offerings will continue to increase over the next few years, creating even more choice, especially for specific use cases (such as operational, data warehousing and Internet of Things [IoT]). As the dbPaaS offerings mature in scalability, reliability, security and governance, the uses of dbPaaS in production workloads and, specifically, in mission-critical workloads will increase. As we stated previously, we are already seeing an increase in regional availability addressing some of the data privacy laws in different countries and regions. Additionally, most application platform as a service (aPaaS) offerings include bundled or integrated dbPaaS, providing a powerful growing channel for dbPaaS and also accelerating maturity. Finally, as the major vendors release dbPaaS offerings that are increasingly compatible with their on-premises DBMS products, the ability to use skills, tools and applications across on-premises and cloud-based systems will become a major distinguishing factor among the vendors.

An emerging trend is the separation of compute and storage (see Figure 1) offered in AWS and Microsoft Azure, as these vendors and others create versions of their products that use the vendor-managed storage layer (e.g., Amazon Simple Storage Service [Amazon S3] or Microsoft Azure Storage or Azure Data Lake Store). This allows greater flexibility in pricing and better elastic scaling of resources (see "Separating Cloud Resources for Data Management Increases Flexibility and Helps Prevent Lock-In").

Figure 1. Cloud Deployment Models for Storage and Computing Resources
Source: Gartner (July 2017)



Recommendations:

Data and analytics leaders (including database and application managers) should work with CIOs and CTOs to:

- Establish a policy for the mixed adoption of traditional DBMS on-premises and public dbPaaS. Each may have a role to play in your organization over the next three to five years.
- Analyze a range of dbPaaS offerings from providers — not just those from major vendors and your standard vendor. There are many specialty dbPaaS offerings that bring unique functionality and flexibility.

In-Memory DBMS

An IMDBMS is a DBMS that stores the entire database structure, including the applications, "in memory" and accesses all the data directly. This occurs without the use of input/output instructions to store and retrieve data from disks. This should not be confused with a caching mechanism, which stores and manages disk blocks in a RAM memory cache for speed. IMDBMSs are available in both row-store and column-store models, or a combination of both.

Over the past five years, almost every DBMS product has added some in-memory capabilities, ranging from completely in-memory to an in-memory column store for analytics. The primary exceptions are several of the open-source DBMS products, such as PostgreSQL and MySQL. MySQL has an in-memory storage engine but no internal process changes to support it. Today, it is pervasive throughout the DBMS products and no longer a distinguishing characteristic of the leaders. It is now expected for a DBMS to support some form of in-memory technology. We believe that, over the next several years, most DBMS products will support full in-memory capabilities. Another way to look at in-memory is that it has moved from a visionary innovation to an execution standard.

Recommendation:

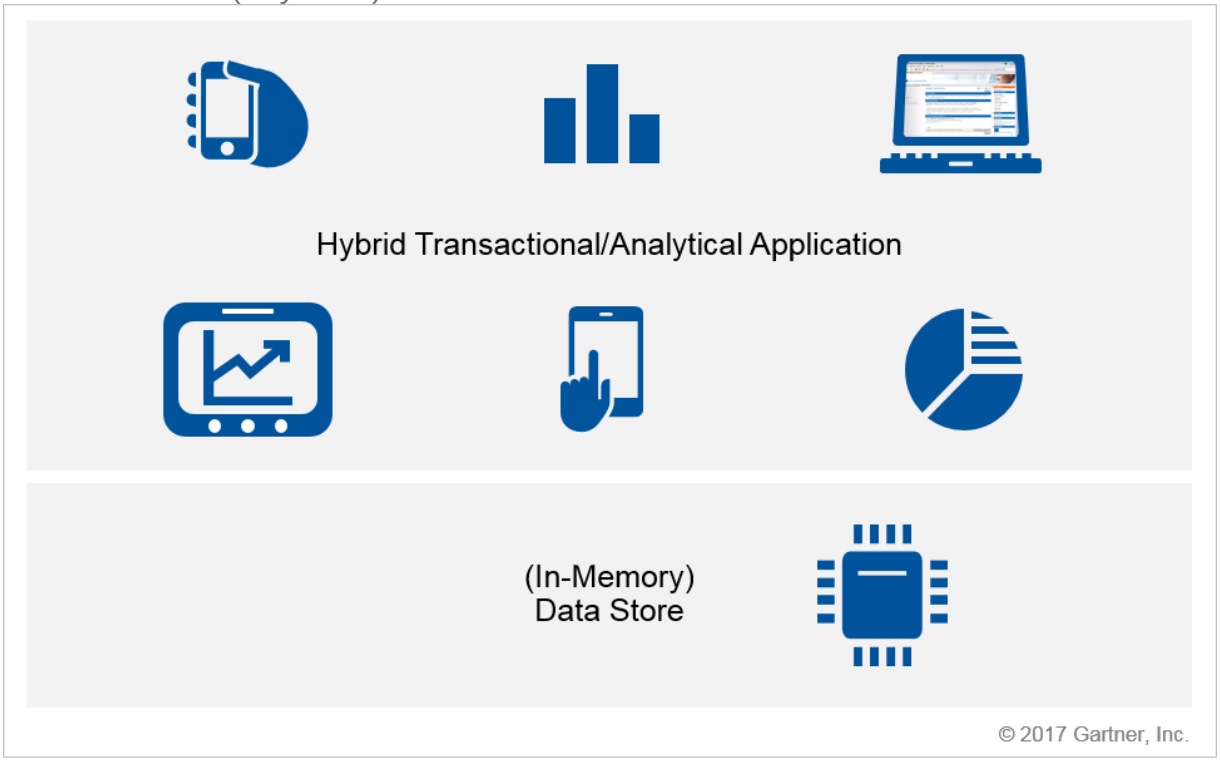
Data and analytics leaders (including database and application managers) should work with CIOs and CTOs to:

- Evaluate building applications on operational IMDBMS for both business value and the possibilities of implementing systems of innovation.

HTAP

A hybrid transactional/analytical processing (HTAP) architecture (see Figure 2) is best-enabled by IMC techniques and technologies to enable analytical processing on the same (in-memory) data store that is used to perform transaction processing. By removing the latency associated with moving data from operational databases to data warehouses and data marts for analytical processing, this architecture enables real-time analytics and situation awareness on live transaction data as opposed to after-the-fact analysis on stale data (as in traditional approaches). HTAP is one of the primary use cases evaluated in "Critical Capabilities for Operational Database Management Systems."

Figure 2. Hybrid Transactional/Analytical Processing
Source: Gartner (July 2017)



The number of survey responses [3] using HTAP applications has remained at about 20% over the past several years. We believe this was due to a lack of HTAP capabilities (especially the analytic capabilities) in all but the most mature DBMS products. More products are beginning to include these capabilities, examples of which include built-in analytics and in-memory functionality. Also, HTAP transactions must be created; simply migrating applications to IMC does not make them HTAP. This requires a degree of modification to existing applications or completely new

applications using the HTAP architecture (see "Market Guide for HTAP-Enabling In-Memory Computing Technologies").

Recommendation:

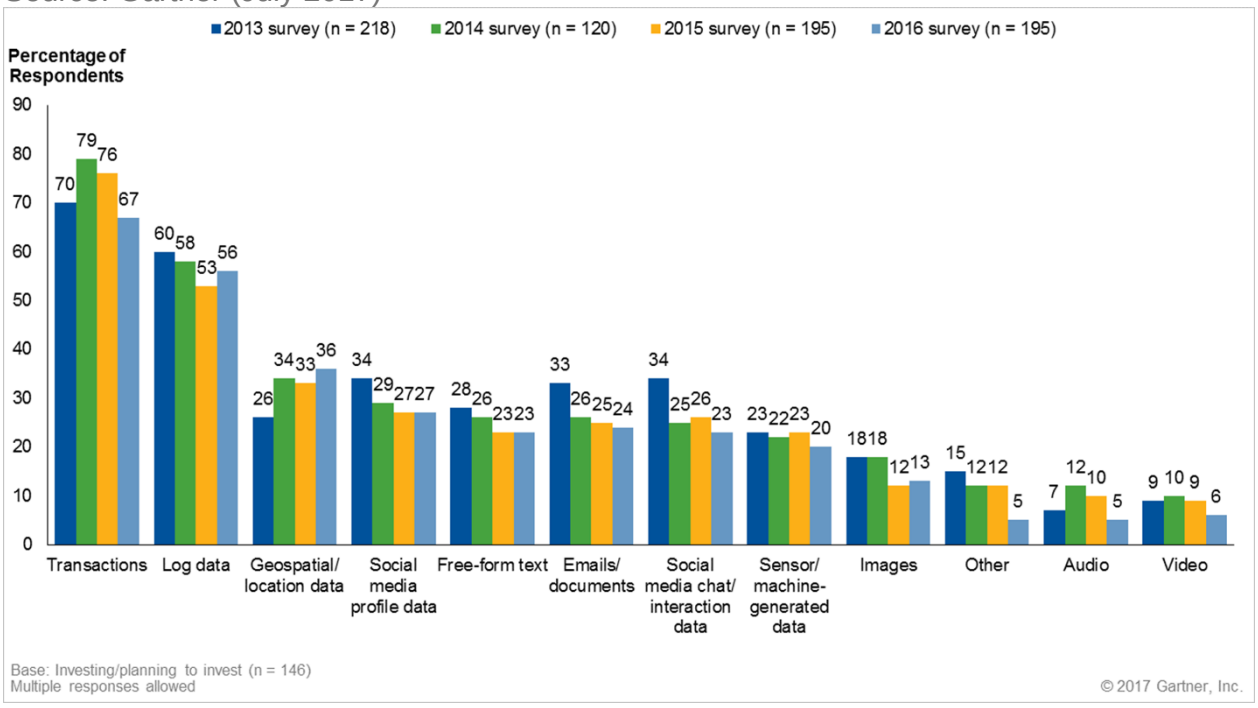
Data and analytics leaders (including database and application managers) should work with CIOs and CTOs to:

- Use IMDBMS capabilities of incumbent DBMS technologies to enable HTAP-capable applications ; for example, when retrofitting for HTAP-established RDBMS-based applications.

IoT and Stream Processing

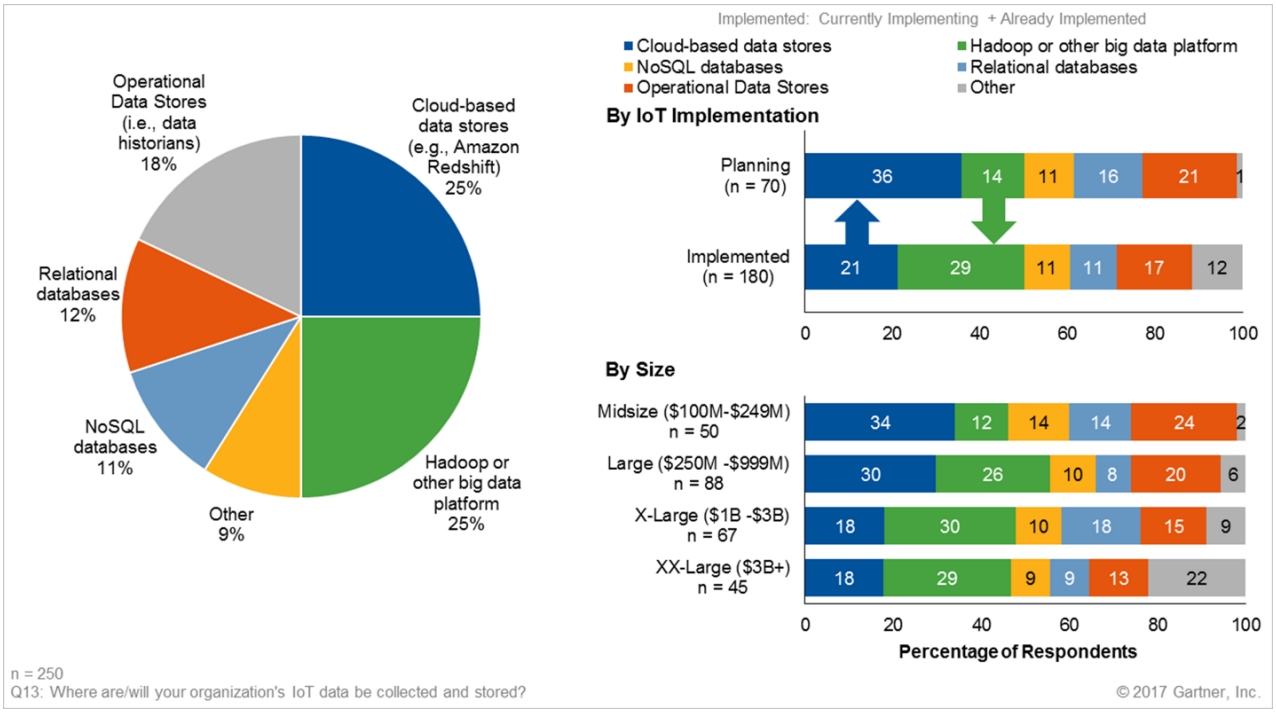
The types of data emitted by IoT ecosystems are becoming more important to organizations, particularly those investing in big data solutions. Based on our 2016 survey for big data adoption (see "Survey Analysis: Big Data Investments Begin Tapering in 2016"), 20% of organizations are using sensor and machine-generated data as part of their big data efforts, while 36% of organizations are using geospatial or location data (see Figure 3).

Figure 3. Data Types
Source: Gartner (July 2017)



Organizations are storing data in a variety of places. Based on a 2016 Gartner Research Circle survey [5] on IoT implementation trends, half of organizations with IoT implementations are using analytical or batch processing stores, with 22% using either relational or nonrelational DBMSs (11% each) (see Figure 4). Additionally, 39% of organizations with IoT implementations store data in operational DBMSs or data historians. [6]

Figure 4. Storage Platforms for IoT Data
Source: Gartner (July 2017)



Most DBMSs support storage of time-stamped data in some way — usually by indexing the relevant field. However, a rudimentary level of time series support — by storing rows of time-stamped data — may be insufficient. DBMSs with time series support have a number of optimizations. These products optimize time series storage by removing duplicate information and grouping data by the producing device. DBMSs with time series support also include time-specific functions, such as range aggregations, time-based intersections and unions, time series filters, and smoothers, among others.

Recommendation:

Data and analytics leaders (including database and application managers) should work with CIOs and CTOs to:

- Consider the use of data streaming products for operational use cases, with the DBMS as a "downstream component" of the stack, not the heart of the architectural model.

Machine-Learning Tools for DBMS Tuning and Maintenance

It can be argued that virtually all DBMS optimizers have been using forms of machine learning (ML) for years through cost-based optimization. However, most other tuning and maintenance operations have been left to the discretion of the database administrator. This is in the form of advisors for indexing, partitioning, reorganization of database spaces and much more. Over the past 12 months, we have begun to hear from vendors that they are adding ML and artificial intelligence (AI) techniques for many of the day-to-day DBMS maintenance and optimization tasks.

With the current industry focus on developing applications with ML, we now believe that this increased attention to ML will lead to its broader use in DBMS optimization and maintenance tasks. This is clearly one of the new innovative areas within the OPDBMS market.

Recommendation:

Data and analytics leaders (including database and application managers) should work with CIOs and CTOs to:

- Ask DBMS vendors for specific roadmaps using AI and ML for DBMS optimization and maintenance tasks, and add ML to all RFPs for DBMS selection.

Evidence

[1] Gartner estimates for 2016 as published in "Market Share: All Software Markets, Worldwide, 2016."

[2] Gartner's client inquiry service accounting for several thousand inquires over the past 12 months.

[3] Survey of vendors' reference customers: As part of the Magic Quadrant research process, we sought the views of vendors' reference customers (details of whom were supplied by the vendors) via an online survey conducted during May 2016. The survey included requests for feedback about vendors' maturity (for example, typical use cases, provision of innovation, responsiveness to new requests, total cost of ownership and pricing) and product capabilities (for example, high availability/disaster recovery [HA/DR] capabilities, support for high-speed ingestion of data, performance, support for multiple data types and problems encountered with the products). More than 500 organizations, representing all the featured vendors' customers, responded to the survey. The average was 26.7 per vendor.

[4] List of several acquisitions (e.g., FoundationDB, Orchestrate, WiredTiger and Tokutek) and several closing their doors (e.g., Basho).

[5] Gartner Research Circle study, "Information Infrastructure Modernization," September 2016.

[6] Gartner study, "IoT Solutions: Implementation Trends," July 2016.

Note 1 Operational DBMS Market Definition

The operational database management system (OPDBMS) market is concerned with relational and nonrelational DBMS (see Note 2) products suitable for a broad range of enterprise-level transactional applications, and DBMS products supporting interactions and observations as alternative types of transactions. These include purchased business applications, such as those for ERP, CRM, the Internet of Things, catalog management and security event management, plus custom transactional systems built by organizations' own development teams.

Note 2 Nonrelational DBMS Types

There are four types of nonrelational DBMSs. They are grouped by the type of data model supported:

- Document-store DBMSs hold data in a hierarchical, tree-like format. Document stores do not enforce an externally defined schema, making them an ideal choice when combining data from multiple sources. Document stores commonly describe data using web-centric interchange formats like JSON or XML. These formats allow for easy mapping to web applications, making document-store DBMSs popular for rapid application development.
- Graph DBMSs store information in a structure that records the direct relationship between any two adjacent elements. Nodes have properties and connect to other nodes at "edges." This inherent structure makes graph DBMSs ideal for storing and analyzing connected data for use in relationship analysis, route planning and optimization, as well as identity and access control (among others).
- Key-value DBMSs store both the key and value as binary objects. Dating back to indexed sequential access method (ISAM) in the 1970s, key-value is actually the oldest nonrelational model. Key-value DBMSs evolved to support rapid scaling for simple data collections by automating the process of distributing data across several nodes. Their attributes allow key-value DBMSs to support data access patterns, largely driven by key lookups and requiring consistent access times. They are commonly used in scenarios requiring a constant stream of small reads and writes, such as web cookies or session tokens.
- Wide-column DBMSs store rows of data in tables, making them the most similar in concept to the relational model. However, table-style DBMSs do not have relationships between rows and they support flexible schema definitions. These traits make table-style DBMSs popular for storing semistructured data, such as log or clickstream data.

Note 3 Database Platform as a Service Definition

A database platform as a service (dbPaaS) is any database management system (DBMS) or data store engineered as a scalable, elastic, multitenant subscription service with a degree of self-service. It is offered and supported by a cloud service provider (CSP) or a third-party software vendor on CSP infrastructure. Direct access to system services such as the operating system and storage software are restricted. We do not define cloud infrastructure as a service (IaaS) with services running DBMS licenses as dbPaaS. We track the offerings to be sold and supported by the provider, even if a third party owns the underlying software.