erwin

by Quest

Taking Control of NoSQL Databases

Data-driven businesses must
adopt a new data modeling mindset

Learn more at **erwin.com**

Section 1

▶ NoSQL Is Vital to Data-Driven Organizations

The world is moving faster than ever before, and the enterprise needs to keep pace by becoming a data-driven organization. Companies striving to achieve this goal aim to provide fluid, on-demand access to data that can inform and impact critical decisions within context to drive better business results.

Accelerating the retrieval and analysis of Big Data—so much of it unstructured—is vital to becoming a digital business that can effectively respond in real time to customers, partners, suppliers and other parties, and profit from these efforts.

What do those gains look like? Expert assessments show that businesses that use data-driven decision-making are **5% more productive and profitable** than their peers. Those who effectively leverage analytics realize a **$13 return for every $1 contributed** to these investments.

Becoming a data-driven organization, however, will test enterprises that increasingly must leverage NoSQL solutions to put unstructured data to work to fine-tune insights and feed algorithms that can initiate the most appropriate actions.

Indeed, the velocity with which businesses are able to harness and query large volumes of unstructured, structured and semi-structured data in NoSQL databases makes them a critical asset for supporting modern cloud applications and their scale, speed and agile development demands.

This is the case for Couchbase, Cassandra and MongoDB, three of the most popular NoSQL database platforms.

Unstructured data makes up **80%** and more of enterprise data, and is growing at the rate of **55%** and **65%** per year.

▶ **Source: Datamation**

The global NoSQL database market was valued at **USD $3.5 billion** in 2019 and is anticipated to reach **USD $21 billion** by 2026, expanding at a CAGR of **31.4%** during the forecast period 2020-2026.

The growth of the market is attributed to the increasing implementation of Big Data among various industry verticals.

Moreover, the growing awareness of NOSQL benefits for various web applications is stimulating market growth.

▶ **Source: Dataintelo**

▶ NoSQL's Rise Changes the Modeling Game

Structured data and relational databases, of course, continue to have a role in the data-driven enterprise. A company can correlate structured data about the poor sales of a product that it's made available in a new color, for example, with unstructured social media commentary to discover why the shade isn't finding favor with buyers.

While the SQL and NoSQL worlds can complement each other in today's Big Data ecosystem, most enterprises need to focus on building expertise and processes for the latter format. After all, they've already had decades of practice designing and managing SQL databases that emphasize storage efficiency and referential integrity rather than fast data access, which is so important to building cloud applications that deliver real-time value to staff, customers and other parties. Query-optimized modeling is the new watchword when it comes to supporting today's fast delivery, iterative and real-time applications *(see sidebar New Approach for Speeding Access to Data).*

Modeling takes on a completely new dimension in NoSQL platforms, such as Couchbase, Cassandra and MongoDB, whose more flexible nature offers benefits for speed, agile development, horizontal scalability, simpler designs and higher availability—all with more control and at lower cost. At the same time, NoSQL has emphasized the role of developers over data modelers, and it's time to bring more balance to the equation.

New Approach for Speeding Data Access

Query-Optimized Modeling™ helps data professionals and developers migrate from traditional relational structures to cloud-based NoSQL collection structures, effectively building schemas that support new application data access requirements. The patent-pending technology from erwin for moving from a storage-optimized model to a query-optimized model targets the easiest and fastest route to data, subject to application and developer requirements.

Those who leverage this capability can determine to conduct transformations using normalization, which keeps the model as close to relational data models as possible and focuses on distinct querying and cross-referencing of top-level collections. Or they can take advantage of de-normalization for duplicating data everywhere it makes sense. That approach results in fewer joins between collections, which is typically better for access speeds.

The use cases for adopting a new mindset oriented to NoSQL data modeling include:

• Transforming existing relational database models to NoSQL models using native, non-relational modeling support as part of refreshing applications for cloud environments;

• Reverse-engineering existing NoSQL implementations, which are characterized by structures implemented within the application code itself, for review and knowledge acquisition purposes; and

• Building NoSQL models from scratch for new applications.

▶ Accommodating a New Age

It's easier to adjust to new principles when modelers and developers can rely on tools that build in guidance for designing or redesigning data stores, so that they feature a more dynamic schema for faster querying of NoSQL-optimized constructs in a proper and rigorous way.

Best practices or Q&A guidance instilled in a tool can lead database modelers down the path of turning SQL tables and the relationships within and across them into appropriate NoSQL collections, based on expectations of how data will be queried. For example, selectively de-normalizing traditional relational database structures as they transition to embedded components in a NoSQL collection can eliminate long queries and slow performance for real-time solutions.

Additionally, providing native graphical visualizations of NoSQL database structures via reverse-engineering improves the ability to understand the current state of existing NoSQL databases. It makes it possible to track what is actually being used in those estates, and to create documentation relationships (depicting data relationships that still exist but are no longer explicitly deployed in a NoSQL environment) in support of institutional knowledge.

When available as a web-based tool with URL access to proposed NoSQL models, enterprises also may gain collaborative benefits so that database managers, developers, enterprise architects, stewards and even business analysts can have conversations around a new model or an existing one as it evolves over time.

New Approach for Speeding Data Access

(Continued)

Taking advantage of custom methods lets users make their own choices of how to leverage referencing (normalization) or embedding (de-normalization) approaches in dynamic data structures. Generally speaking, it's better to embed when dealing with a smaller number of objects or entities in a collection—depending, of course, on the complexity of queries. When the number of embedded entities grows too large, queries can become heavier and speed advantages can be reduced.

Need for a Stronger Grounding in Governance

A few years ago, **Gartner Research VP Nick Heudecker** informally surveyed NoSQL adopters and discovered that database administrators represented only 5.5% of respondents. Commenting on that in a **blog**, he noted that "application developers may be getting what they want from NoSQL now, but cutting out the primary data stewards will result in long-term data quality and information governance challenges for the larger enterprise."

It wasn't the first time Gartner raised concerns about governance in the NoSQL space. In its report, **Does Your NoSQL DBMS Result in Information Governance Debt?**, Gartner noted that it was important for information management leaders to understand the ways in which NoSQL deployments could create some issues.

▶ Bringing Governance to the NoSQL Space

Still, it can be a challenge to leverage the latest NoSQL database technologies while also attempting to maintain the integrity, quality and governance of underlying data.

Logical data models, where an entity relationship diagram (ERD) provides a picture of different data categories and how they relate to one another, offer a strong underpinning for data governance. Database administrators, data architects and data stewards know the importance that logical ERD has for cross-domain modeling and management.

But logical modeling has not been widely embraced by NoSQL developers, who favor physical models, so it's hard to seamlessly integrate NoSQL data in its native form into the larger world of data. There's a need for change on this front, to garner support for logical modeling to reference many data sources and ultimately map back to a single set of enterprise data elements with appropriate governance pieces for security, privacy and other requirements.

With an appropriate data governance tool, it will be possible to link logical to physical modeling and representations of NoSQL databases. For instance, a company may use such a tool to force NoSQL developers to select from an existing data glossary of objects drawn from a logical ERD before they create a new collection or physical object, and in that way, achieve more holistic governance.

Other aspects of governance to consider are the benefits that come from providing collaboration on data models across business and technology roles. That way, individuals can bring different perspectives to the table, potentially reducing risks in a model that those serving other functions might not see and thereby mitigating them as part of the process.

A summary of the report discussed that:

- "Flexible schema databases support storage of structured and unstructured data falling outside of established information governance practices.

- NoSQL databases complicate information governance by moving business logic from the data store to the application tier.

- NoSQL DBMSs are additive in most environments, increasing data movement and the number of storage locations."

NoSQL data governance has been hampered on a couple of other fronts. In fast-moving organizations where agile development of web- and cloud-based apps rule, governance processes and leadership haven't always been able to keep pace. There's also been a dearth of tools to map disparate data sources into a single, logical enterprise model for improved governance and collaboration on model representations.

Industry leaders know this state of affairs can't continue, which is why the most forward-thinking among them are taking steps to address these issues.

▶ A Look at erwin Data Modeler by Quest

As NoSQL gains a bigger foothold in every business transitioning to become a data-driven enterprise, the industry must respond to companies' requirements to transition relational models, review and present the structures of existing NoSQL databases, and propel new projects—all with better adherence to stronger governance practices.

To that end, data modeling industry leader erwin has extended its erwin Data Modeler software to include NoSQL modeling to enable any data from anywhere to be visualized for greater understanding and better decision-making.

In a nutshell, erwin Data Modeler is the first enterprise-class modeling solution with native NoSQL database support for Couchbase, Cassandra and MongoDB. Additional NoSQL database support will be added in the future.

Building onto erwin Data Modeler's Firm Foundation

For seven years in a row erwin Data Modeler has been named No. 1 by readers of DBTA Magazine.

Businesses now can support the three primary use cases for NoSQL data modeling, with multiple advanced features. Specifically, with erwin Data Modeler:

• Users will be able to import both physical and logical relational models and migrate them to Couchbase, Cassandra and MongoDB, leveraging graphic and tree visualizations as well as built-in guided editors and query-optimized modeling to automatically embed, reference or customize their cloud-based collection transformation options based on application and query requirements.

• Users will be able to reverse-engineer existing Couchbase, Cassandra and MongoDB databases to infer their structures and gain visualizations of collections and embeds within the database.

• Users will be able to support agile development approaches by creating Couchbase, Cassandra and MongoDB models from scratch as well as fully updating them with new collections and data to service new applications.

• Teams including all classes of data professionals, developers and business colleagues can collaborate on models, creating scenarios and working out application code changes.

• Post-review models can be forward-engineered and scripts can be pushed out to a Couchbase, Cassandra or MongoDB instance for validation of top-level collections as part of dynamic structures.

• Logical data models will be supported, enabling governance of objects, entities, tables and collections and integration of NoSQL and other data.

• Institutional knowledge will be preserved through relationship documentations.

• Users ultimately will be able to model data in the context of enterprise architectures and business processes, further enabling a well-plotted, data-driven enterprise.

Section 4 *(continued)*

▶ Designed for the Way We Work Today

Complementing Couchbase, Cassandra and MongoDB and designed to suit a new generation of technologists "born in the cloud," erwin also offers erwin Data Modeler as a hosted solution, making it possible for companies to start using it on demand with no overhead.

erwin's acknowledgment of the critical importance of collaboration in NoSQL modeling processes is exhibited by its licensing model. Every developer license includes the ability to invite five other participants to view and comment on a model so there is a constant feedback loop from those served by Couchbase, Cassandra, JSON, AVRO and MongoDB collections, services and applications in agile environments. This drives a more finessed process to build applications atop the databases, leverage the data for business requirements, and ensure that no governance requirements are breached in models.

Existing erwin Data Modeler customers who want to create better connections between their relational models and Couchbase, Cassandra and MongoDB deployments—as well as others who are just looking to more easily discover and understand their NoSQL data—are prime beneficiaries of the capabilities embodied in erwin Data Modeler.

▶ **Take erwin Data Modeler for a free spin.**

erwin
by Quest

About erwin by Quest

erwin is a leader in enterprise modeling and data intelligence software. The erwin EDGE platform creates an enterprise data governance experience for IT and business collaboration, driving meaningful insights, agile innovation, risk management and business transformation. Integrated data modeling, data governance, enterprise architecture and business process modeling capabilities help guide smart decisions. With erwin, organizations of all types across the globe can maximize the security, quality and value of their assets to control data chaos and prepare for the next IT challenge.

Visit **erwin.com** to learn more