Leveraging the right Data Model towards building a robust foundation for Enterprises

A good data model is like having a building plan for an architect and provides the blueprint for any modern analytics platform or data transformation initiative. Getting the right data model in place is a key step in the database design process. A robust data model enables you to get the best performance out of your dashboards and storage, helps reduce maintenance time, and makes all the difference to your daily operations and end user experience.

Data Modelling approaches to develop a modern analytics platform

There are several modelling approaches and the those are usually defined as per the standards of the organization. To dig deeper two of the most widely used approaches currently in the industry are dimensional modelling and data vault modelling.

Dimensional modeling (top down or bottom up process) includes a set of concepts, methods, and techniques that help in designing a data warehouse. The focus is to identify, prioritize and model crucial business processes.

Case in Point: For one of the world's largest restaurant chains, we identified a need for data transformation to better leverage at a global scale with over 50,000 restaurants in over 150 countries and territories. We helped them develop a common language and approach as a foundational step in realizing value from their data assets. We enabled consolidation of terabytes of data (approx. 500+ tables) with integration of source data model with new target data model including data dictionaries and KPI standardization which aided in building a real-time Sales and Customer behavior analytics across their global Markets. This in turn resolved organizational bottlenecks with easy access to customer data and enabled high customer engagement and retention.

Alternatively, **Data Vault Modeling** is an approach which is designed to support long-term storage of historical data for the long-term scenario where the data comes from disparate sources. It deals with concerns regarding audit, lineage, load speed and resilience to change. This means that every row in the data vault needs to have the record source and date attributes, which enable an auditor with data tracing.

Case in point: In Financial Services, an industry growing at a fast pace with increase in exposure to fraud and credit risks, it is essential to foresee and minimize these risks through effective data models at scale. Hence, for a European financial services customer, we leveraged the data vault approach to build a data set which was agile enough to scale according to the requirements. It allowed for parallelization because the modeling approach had fewer points where data needed to be synchronized. This resulted in 50% faster data loading processes, a key benefit especially for the client as they were dealing with dataset in terabytes (200+ tables with delta in GBs) and handling real-time transactions or near real-time data inserts.

Case in point: For one of the largest banks in the US, we formulated unified data platform that enables them to discover new insights and create business opportunities in a secure manner. We built an analytical layer which uses data vault modelling that resulted in 55% reduction in data ingestion time and increase the scalability with 95% than their legacy system. The complexity lied at maintaining the data up to date when there is frequent delta sizing in giga bytes. The new platform facilitates data consumers with autonomy to discover and share new insights at a faster rate.

The need for optimization in data modeling

Over time, it is likely that data sources change may result in significant rework of the traditional models. It becomes difficult for enterprises to verify that their data is being efficiently and fully utilized to enhance business if no standards exist to check the basic accuracy, coverage, extensibility, and interpretability of data. Data-driven decisions may not bear fruit in absence of a trusted process to maintain data quality.

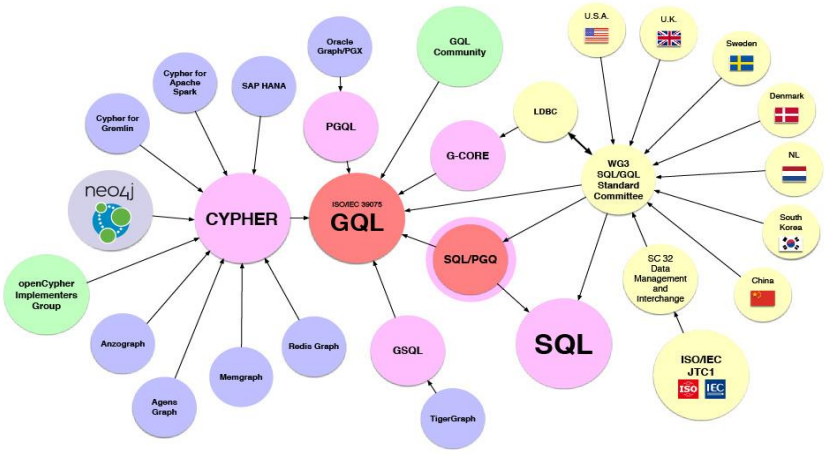
Advancements in technology that enable data modeling optimization

For businesses to stay competitive in the VUCA world, data modeling optimization will have to be conceptualized and implemented faster and more seamlessly. More and more architects and product owners will rely on emerging technologies such as graph database to aid use case development, agile data modeling for real-time manipulation and move towards realizing a universal data model. Some key trends that will dictate data modeling optimization include:

- **Data management on cloud**: Emerging cloud-based quantum computing promises to increase computing strength resulting in faster interconnects. The low cost and scalability will foster large

applications that disseminate more data through the hardware. Development and access to data models in the cloud will become imperative.

- **Graph Database:** It provides easy and quick visuals of business cases. This database type contains nodes and edges, grouping data sets, their descriptions, and relationships. They enable evaluation of group data and a logical view of various business rules. Database administrators can scale high data values and create usable data models.



- **Agile Data Model:** It provides data model which is just in time using a minimum requisite of design for pertinent circumstances and deals well with a mix of a variety of data types such as relational, unstructured, dimensional, and master data. This enables business users to create their own models and eliminates the need for human data engineers to provision the data, speeding up the data modeling process. Successful agile data modeling, however, requires a thorough understanding of

- **Universal Data Model (UDM):** It enables quicker design and deployment with improved performance and facilitates ease of maintenance and integration at a reduced cost, in an enterprise environment. It streamlines internal communication, increases consistency of documentation, simplifies, and increases the applicability of data modeling for a multitude of use cases.

Focus on Universal data Model

Organizations cannot afford to rely on labor and time intensive data models that are built from scratch. Universal data models provide data modelers with the building blocks required to develop the enterprise conceptual, logical, and physical data model with minor customizations. Each industry has its own set of business issues, with differences in personas and use cases that apply to every function. However, subject data around customers, distributors, agents, suppliers, internal organizations, and people within the organization that span across sales, marketing, customer service, purchasing, shipping, invoicing, budgeting, and human resources is identical and critical to track.

Pre-defined templates with mainstream relationships and interactions between functions or a UDM can provide an effective solution to maintain this high value information and ensure the data integrity issues are not ignored. A UDM provides an accurate view of unified data for business decision making

responsible for transforming data management to an enabler of a data-centric enterprise by empowering data modelers with a framework leading to high quality designs in less time to market.

Some of the key approaches in the industry driving towards the concept of UDM are:

- **Cross Enterprise model:** Exchanging data between organizations is being increasingly adopted for strategic alliances, mergers & acquisitions etc. Federated MDM approach is well known for real-time enablement of these capabilities.

- **Industry model:** They predefine the data in terms of the business objects, business entities, business metrics, and you map the data to those business metrics. These models are pre-built and reduce management effort and lead to swift implementation and ease of use.

These approaches widen the value of data modeling over singleton use cases to facilitate mainstream methods like mapping, schema, time-series analysis, and terminology standards across and between enterprises.

How advances in technology enable Universal data model

Entity Modeling: A basic requisite for spanning data models across the enterprise, industry-specific deployments, or between organizations is to center them on entities that are a business' primary concern. Coupling event schema with individual entities offers advantages such as simplicity, feature generation for machine learning, and tracing customer journey.

Federated Master Data Management: Data models expand their value via a federated approach between multiple organizations for data supported truths by offering exactness of exchangeable data between organizations. They enable real-time responses to evolving business conditions and offer rich predictive analytics to enable collaboration for problem solving.

Terminology Standards: Universal data models must standardize the terminology describing business concepts, especially across different data types. This will drive adoption throughout the organization.

Mapping Automation: The source mapping is separate compared to the business rules running on top. This characteristic is imperative to the reusability of common data models for the long-term since the business rules don't change with changes in the data source(s).

Time Series Analysis: UDM reduces the time spent engineering data and enable time sensitive requests with low latency. Event-based schema exemplifies these temporal benefits with applicability throughout the enterprise. For example, events include start/stop times of client contact center interaction and sub-events such as thoughts about products, service cancellations etc.

Case in point: We partnered with a global Bio Pharma company to help them realize their vision of accelerating digital adoption into becoming a cloud first customer centric organization. We built a solution capable of leveraging industrial grade models with analytical layer having terabytes of data augmented by automated data-driven supply chain engine to gather insights for decision making. It helped in 4X faster processing of clinical data leading to faster outcomes for scientists.

Sticking to the best data modeling practices is essential to prevent inefficiencies and delays. We delve into the best practices for optimization of data models in the next section.

Best Practices for Data Model Optimization

- Always try to go for the most granular level of data that is as per the analytics use cases for visualizations and create custom datasets.
- Ingest only the data you need, as unnecessary chunks are attributes to slowed data processing. Also, using only the relevant data helps in efficient storage.
- Use appropriate data types and data size. Use numerical values whenever you can. Using text consumes more storage space and is slower when evaluated.
- The data model needs to be accommodative of the change data capture.
- While deciding on the model, the focus should be also on the data size, the frequency of data update and size of data inserts.
- Optimize time frames and activity days. Enable users to refresh data whenever it's really needed. This offers a better personalized experience.