speedb

Overcoming IoT Metadata Sprawl Challenges in Smart Cities

Introduction

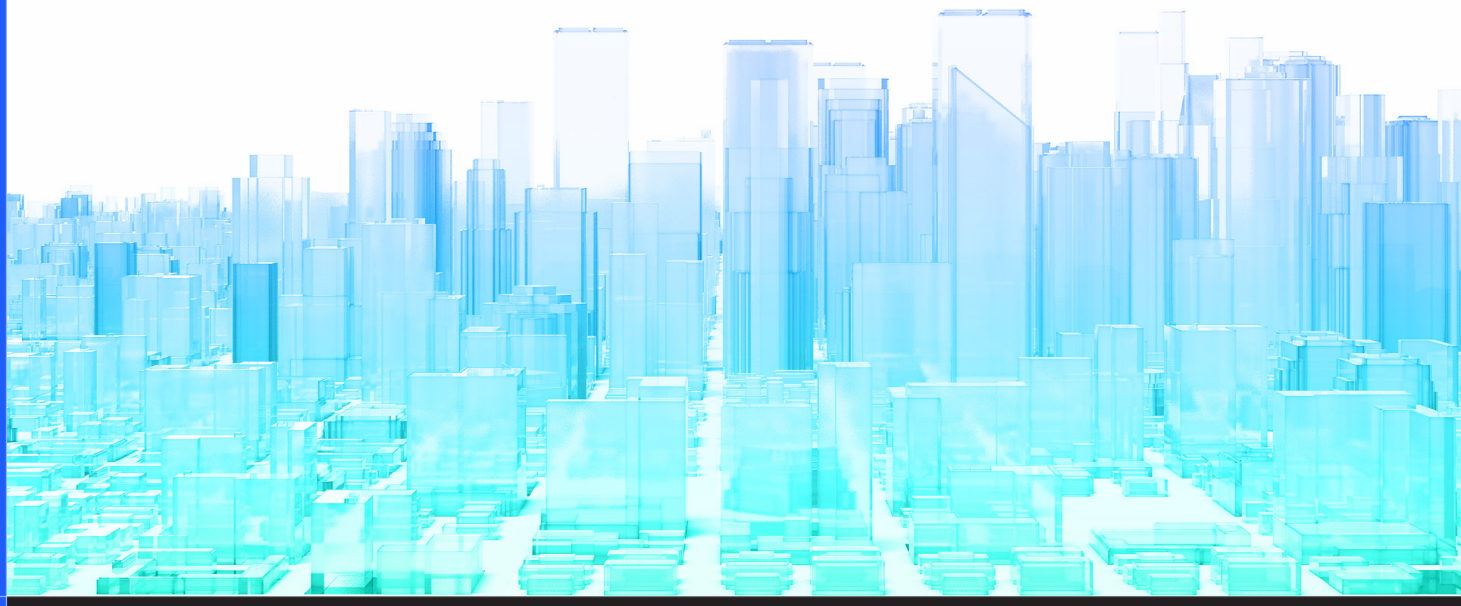
Over the past five years, businesses around the world have begun facing a new challenge brought on by the digital age: an explosion of data. While many enterprises prepared for this growth, few analysts could have predicted the scale and pace at which stockpiles of data would overwhelm existing storage infrastructure.

Unfortunately, despite expert predictions and "future-proof" products, businesses today find themselves buried beneath an avalanche of data. Much like a growing city, businesses must now seek out the best way to provide their services to a wider customer base without compromising the quality of their product. For most this entails providing services globally—not a small feat.

However, to provide services on such a scale requires data about each user, each region, each service, and so much more. This form of data about data is known as metadata. While data enables businesses to provide services across the world, metadata enables them to simplify data discovery, integration, interoperability, governance, analysis and management, ensure data quality, and much more.

When an application is used, how it is used, and what it is used for are all bits of information provided by metadata. Metadata can range from simple functions like user location to the smallest details like second by second activity on an eCommerce website. They enable businesses to track fraudulent activities, identify popular products, and even see where a user is clicking on a mobile application.

These are just a few examples of how metadata enriches the data stored by businesses. However, it also dramatically increases the amount of data businesses must store as an increasingly large number of objects (e.g. video, images, audio and sensor data) that may only be a few bytes large may now be holding metadata of about the same size, and sometimes even more.



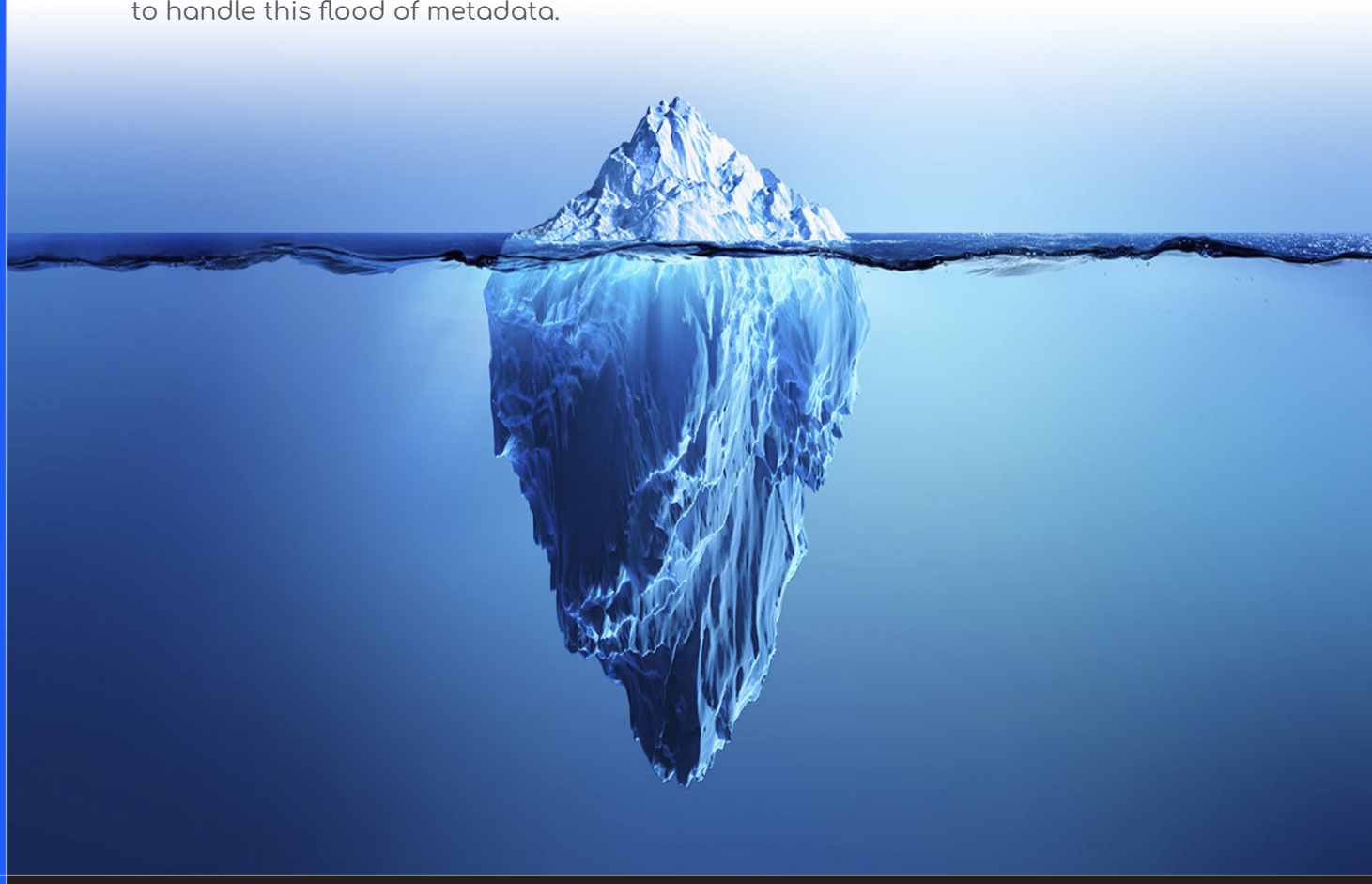
IoT and Metadata

Internet of Things (IoT) devices are largely responsible for much of the metadata growth impacting modern businesses. While the data transmitted by these devices might be small, say less than 2 bytes, the metadata associated with that data (device name, location, date, time, operating system, etc.) is significantly larger.

When thousands and millions of these devices all transmit chunks of metadata every second it is nigh impossible for businesses to not be overwhelmed. According to the DataSphere Forecast report by research firm IDC, by 2025 there will be 55.7 billion connected devices worldwide that will generate 73.1 ZB of data. In the coming years, metadata is expected to account for an increasingly larger share of this total. Furthermore, a growing number of use cases are based on the metadata collected from IoT devices. Collected sensor data is now enriched with metadata (e.g. device type, location, time, etc.) before it is stored. This data is then processed via advanced analytics through older processes, like running queries, or newer tools, like machine learning.

Once the raw data has been processed businesses are left with the useful bytes that represent use cases relevant to their industry. This data provides key insights into metrics like performance, usage statistics, location hot zones, and so much more. With this data they can enrich their applications, services, and products for unique customer bases without compromising their larger offerings.

However, this comes with one key problem: existing data architectures were not designed to handle this flood of metadata.



Outdated Infrastructure

The basic operations of storage management are commonly executed by the data engine (aka storage engine). Installed as a software layer between the application and the storage layers, a data engine is an embedded key value store (KVS) that sorts and indexes data.

In addition, KVS is being increasingly implemented as a software layer within the application to execute different on-the-fly activities on live data while in transit. This type of deployment is often aimed at managing metadata-intensive workloads and preventing metadata access bottlenecks that may lead to performance issues.

Data engines often use a Log-Structured Merge (LSM) tree-based KVS to keep metadata in memory. While offering more flexibility and speed compared with traditional relational databases, an LSM-based KVS has limited capacity and high CPU utilization and memory consumption due to high write amplification, which refers to the ratio of actual writes to storage compared to writes requested from the database.

As the number of write operations grows, an LSM tree-based KVS tends to suffer from degraded as well as unpredictable performance, resulting in slower access to the underlying media, and consequently—application performance degradation.

Application performance is the largest factor in customer satisfaction in the digital age. A business that suffers from application performance degradation, no matter how minimal, will quickly find customers abandoning their services in droves. Applications and digital storefronts must now represent the entire business and a failure in either can have far-reaching consequences. Unfortunately, due to the limitations of traditional data engines, many businesses are finding themselves under scrutiny for these afore-mentioned performance stutters.

Smart City Metadata Challenges

Much like the industrial revolution, steam engines, the printing press, and other advancements have changed how people live, the digital age has introduced smart cities. These cities flourish because of the digital footprint created by its inhabitants. The name comes from the thousands of smart devices that make up the web of IoT devices woven throughout the homes, businesses, and streets of the city.

These cities are perhaps the best example of the challenges metadata presents. smart cities collect IoT data from multiple sensors and devices for various purposes, e.g., traffic management, security and surveillance, smart lighting, waste management, and more. This data is then processed, stored, and analyzed along with the accompanying metadata, which is essential for extracting the meaning of data collected from a myriad of sources, and making smart correlations between data elements to support advanced smart cities use cases.

Implementing an appropriate data infrastructure to store and manage IoT data is critical for the success of smart city initiatives. This infrastructure should be scalable enough to handle the ever-increasing influx of metadata without sacrificing performance. This is particularly important for smart cities applications and use cases that are highly sensitive to response time and latency, e.g., traffic optimization, smart parking, and many others.

Speedb's Metadata-Optimized Data Engine

Speedb has developed a new data engine designed from scratch to support metadata-intensive applications such as smart cities. Speedb completely redesigned the basic components of RocksDB, the popular open source KVS, in favor of a new, scalable data engine. Speedb features a new compaction method that dramatically reduces write amplification for large scale LSM, a new flow control mechanism to eliminate spikes in user latency, a revolutionary indexing method, and more. The result is a modern data engine, which enables businesses to fully utilize their existing infrastructure and avoid the problems presented by an overload of metadata.

Furthermore, Speedb reduces management overhead for businesses by enabling applications to run on a single database. IT teams are no longer beholden to continuous sharding or other time-consuming maintenance tasks, freeing them up to focus on providing better services for smart cities and their global end-users. Hyperscale data operations can now provide the perfect digital experience demanded by their end-users without worrying about a failure in the back end.

Speedb eliminates the challenges created by metadata. By providing a scalable database without compromising on performance and cost-efficiency, Speedb enables businesses to optimize hyperscale data operations so users don't feel the impact of data congestion in their daily lives. Metadata is here to stay so it's time that businesses stopped relying on outdated infrastructure to handle a modern problem.