Power LSEG Tick History with a data-centric cloud-based service

By moving your tick history data to the cloud rather than storing it on site, you could reduce your costs by 90%. There are other powerful reasons for firms to shift this dataset to the cloud, too. This Expert Talk explores today's challenges around on-premise tick history data management, as well as the substantial benefits of adopting a cloud-oriented, data-centric market data strategy for tick history.

Today, many financial services firms are proactively rethinking the way they engage with tick history datasets. Certainly, the COVID-19 pandemic spurred a general shift to the cloud – data, analytics and technology are much easier to access when they are in the cloud, no matter where an individual is working from. However, there are multiple reasons why firms are moving to the cloud.

Facing enterprise challenges

Cost is one overarching theme. According to a recent survey by LSEG Data & Analytics, for every US$1 spent on financial market data, a further US$8 is spent on processing, storing and transforming that data before it can be analysed. Some LSEG Data & Analytics clients have reduced the total cost of their tick history data ownership by more than 90% — simply by moving to the cloud.

Another theme is firms' need to rethink data strategies in light of issues such as increasing data volumes, the need for more operational resilience, and evolving use cases for the data. If we break these down into more specific day-to-day challenges, we find that firms need to make adjustments in three areas: data, cost and innovation.

Data

– **Manage exponential data growth** – It's growing increasingly difficult for financial firms to build and maintain their own tick history dataset. One significant issue is the large volume of data – for example, LSEG Tick History can add up to 2TB+ of data per week. These data volumes only grow larger over time because the markets never take a break.

LSEG DATA & ANALYTICS

- **Break down data silos and access a single source of truth** – Often, a firm's front, middle and back offices each have their own data silos, resulting in multiple sets of the same data rather than all working from a single data source. This increases cost and even tiny variations between datasets can result in errors and questions about data integrity. Data needs to remain consistent and integrity must be maintained throughout the organisation.
- **Improve data governance** – The first step in data governance is often to create a single "golden" source of data, which can then be monitored for data quality and tracked for data lineage across the whole data life cycle.
- **Create data lakes** – Financial services firms are deploying data lakes — vast pools of data that provide insight for both pre-determined and new queries or requirements across business functions — in the cloud.

Cost

- **Reduce or redeploy hardware** – Firms want to remain agile and reduce or redeploy their hardware and infrastructure. They also want to be able to scale their data estate more easily, lower their costly on-premises data management and storage footprint and reallocate resources to revenue-generating activities.
- **Shrink the costs of cleaning and normalising data** – Firms hire expensive quant experts to perform analytics and modelling and build data pipelines. Consistency across multiple markets and asset classes is key — the data must be able to satisfy multiple use cases and be prepared for use in the most efficient manner across the organisation.

Innovation

- **Adopt cloud-based or cloud-first solutions** – Most financial firms today have a cloud-first strategy. Storing tick data on-premise creates a heavy burden for firms trying to engage in this approach. Rather than reinvent the wheel and move tick history data to the cloud via lift-and-shift, LSEG Data & Analytics offers a ready-made solution to overcome this challenge.
- **Boost operational resilience** – COVID-19 highlighted just how essential operational resilience is to any market data strategy. The pandemic simply accelerated what was already an emerging regulatory concern about data and technology resilience.
- **Become more agile** – Firms want to reduce the time it takes for those who work with data to access it and process it for standard workflows. Firms are also keen to leverage cloud compute to optimise loads and benefit from its elastic capabilities to perform analytics in the cloud. In addition, firms are very keen to reduce the time to insight and improve the quality of decision-making. Lastly, firms want to be nimble enough to build and adopt new artificial intelligence and machine learning solutions.

These challenges are significant and often interconnected. At their root is the fact that a systems-based design underpins many firms' overall market data strategy. This means that:

- Individual datasets are tied to specific software or services
- The software or services may not be able to connect through APIs or other connectivity interfaces to transfer data in or out
- Datasets may be geographically dispersed within on-premises hardware
- The data usually sits under a variety of information security structures
- Data is often formatted in several different ways
- Data governance over these sets is complex and thus often substandard

Clearly, to meet the challenges that lie ahead, firms need to rethink their approach to data acquisition, management and use across the front, middle and back offices. A systems-based approach to data management simply won't enable firms to meet the data challenges they face today — or those they'll face in the future.

"Firms can drive digital transformation by taking a data-centric approach to their strategy for tick history data," says Tim Anderson, Global Head of Tick History at LSEG. *"While improving efficiency and reducing costs, firms can also democratise data across their entire organisations."*

Adopting a data-centric approach for tick history

Firms are recognising that they can solve these challenges in the months ahead by shifting from a systems based design to a data-centric design. A data-centric design puts data, not systems, at the heart of a firm's market data strategy. One instance of cloud-based data is created and maintained within a time series database, with individual compute power for millions of users. Individuals and technology in the front, middle and back office work with this single data source in the cloud. Not only does this approach dramatically reduce costs, but it also addresses the other challenges discussed above, including data quality, agility and resilience.

"Firms can drive digital transformation by taking a data-centric approach to their strategy for tick history data," says Tim Anderson, Global Head of Tick History at LSEG. "While improving efficiency and reducing costs, firms can also democratise data across their entire organisations."

A useful case study which explores how this works in practice is the adoption of LSEG Tick History — Query, powered by Google® BigQuery. LSEG Tick History is a historical archive of real-time pricing data, covering OTC and exchange-traded instruments, from more than 500 trading venues and third-party contributors. Built on the LSEG Data & Analytics universe of more than 70 million active and retired securities, with data coverage going back to 1996, it contains more than 8PB (compressed) of historical data. LSEG Tick History is used across the front, middle and back offices of firms to:

– Perform quantitative research and analytics

– Use precision-based pattern matching within the markets and decision tree analytics

– Test real-time algo trading strategies

– Build and back-test trading strategies

– Perform post-trade TCA, including order book recreation or market microstructure analysis

– Fulfil regulatory requirements that require the use of tick data, such as best execution or trade surveillance

Tick Google BigQuery Compute: performance and cost

– **Fully-managed time series database**
LSEG Data & Analytics, in partnership with Google Cloud Platform (GCP), provides a fully managed time series database with a compute engine. No additional capacity planning needed

– **Enormous depth and breadth – 8PB+/25 years**
500 global venues and third-party contributors back to 1996

– **Level 1, 2, 3 market depth and order book with Corax and exchange qualifiers**
Level 1, order book, corporate actions and specific trade qualifiers available for querying

– **High performance – query millions of rows of data**
Query millions (or trillions) of rows of data in seconds using GCP Big Query – no need to break data down into chunks or bins

– **Thousands of users – no batch jobs, no queuing**
Query the same dataset concurrently with no impact on performance (no job queuing)

– **Enterprise ready – snap into existing architecture**
Use a single API to connect users or legacy apps to the new Tick Store. Snap into existing data flows or data warehouses such as Snowflake

– **Sublinear scaling – the more data the better**
Sublinear scaling on all queries of tick data, so no need to pre-prepare data into smaller batches. The more rows processed, the better the overall performance

– **Full graphing and visualization**
Tools such as Looker provided in browser blocks or full-page graphing create client interactive experiences

– **Interactive analytics and functions**
Best-of-breed multi-line and multi-content queries, including build decision trees and select statements

– **Interface independent**
Run any interface to query, compute and review/extract output (Python, R, SQL etc.)

– **End-to-end encryption**
For all data at rest and in transit, end-to-end encryption is provided

– **Time series auto-partitioning**
Time series partitioning with in-memory analytics for super-fast returns with SQL Windowing

– **PermID® ready**
Link multiple datasets through a common identifier – apply intelligent tagging across many datasets

The LSEG Tick History – Query is a managed service database that contains the full depth and breadth of LSEG Tick History data. Firms no longer need to clean, normalise and manage their tick history data – LSEG Data & Analytics does that for them. Also, firms can query this database directly using the Google BigQuery engine, either via the Google-provided GUI or via API from the firm's preferred environment (e.g., Python, Java, SQL) and query language, taking advantage of the large data store and cloud computing. After running analytics, firms can leave the results in the cloud, move them to their other clouds or into business intelligence system that may use: query in the cloud, analyse in the cloud.

The Google BigQuery engine provides exceptional querying and analytics capabilities on very large data sets such as LSEG Tick History. This all happens in the cloud, without the need to perform Extract, Transform, and Load (ETL) or to support the data in other ways. Analytics that would normally take hours to compute now take seconds. For example, if a business analyst needed to run a VWAP on an entire day of NYSE Tick History from a lockdown location with limited broadband, it would normally take hours to wrangle the data and then download it. With LSEG Tick History – Query and the Google BigQuery engine, this would take a mere 1.7 seconds to do, regardless of locality.

The LSEG Tick History data in Google Cloud BigQuery is delivered via the LSEG Data Platform, which offers access to global market data and analytics as a service in the cloud. Firms can power and deploy their applications with reduced time to market by leveraging the LSEG Data & Analytics consistent data model, permissioning and entitlement capabilities and cloud-based tools such as the Google BigQuery engine.

Overall, LSEG Data Platform contains more than 100 million instruments and datasets, with 40 billion market data updates delivered every day. Clients who subscribe to LSEG Tick History data in Google Cloud Platform can also subscribe to LSEG Real-Time – Optimized, a new all-trade, quote-conflated solution delivered via the LSEG Data Platform, so that the whole firm is working with consistent datasets for both real-time and tick history.

In summary, LSEG Tick History – Query, coupled with the Google BigQuery engine, is able to address many of the significant challenges firms face today by helping them reduce costs, improving data quality and access and dramatically increasing performance. Looking more deeply into how these services can reduce costs provides a more clear example of what is possible.

Translating data-centric into reduced costs

Cost pressures around data management are a significant challenge for financial services firms – and particularly those expenses associated with managing and maintaining tick history data stored on premises. The LSEG Tick History team is able to perform in-depth analysis of the expenditure associated with an on-premises approach to tick history. These expenses include:

– **Compute server footprint** – Including the cost of the servers for processing the data, along with the housing of the hardware in a building. Other costs include cooling, racking, networking connectivity and other supporting infrastructure such as built-in redundancy solutions

– **Storage hardware** – The cost of providing and storing database solutions, along with the ability to retrieve the data for various market-related use cases. This also includes costs for the cooling, racking, networking connectivity and other supporting infrastructure, including built-in redundancy solutions

– **Off-premises storage back-up** – For business continuity plans, firms will need to have a copy of their tick history database running in another location. This is often a regulatory requirement

– **Relational time series database technology fees** – Charges paid to the database software provider (KDB, Microsoft, Oracle, etc.)

– **Human capital** – The cost of the employees who manage databases, network infrastructure and capacity planning for the hardware, database maintenance and storage

These expenses add up quickly and are expected to grow over time. In total, firms that want to continue storing their tick history data on premises will pay more than US$110 per gigabyte of stored data. When working with LSEG Tick History – Query, the cost can be as low as US$4.

Even migrating to LSEG Tick History – Query is inexpensive. Since the data is in the cloud, there is no need for the costs usually associated with migrating from one database to another. For example, firms can be given access to an API to test with and spend 100 days testing for just a couple of thousand dollars.

To help foster a better understanding of why shifting to the cloud could make sense, the LSEG Tick History team can perform analysis on a firm to demonstrate the scale of the existing costs associated with on-premises tick history data as opposed to adopting LSEG Tick History – Query.

The team can also discuss the other significant benefits of adopting a data-centric approach to tick history data with stakeholders and provide a free trial. Working hands-on with the data can help firms understand how hands-on apply it to their own use cases across the front, middle and back offices.

Performance analysis is one example of the kind of insight that the team can bring. A common misconception is that moving to the cloud will reduce the performance of an on-premise local database but, due to the way Google and LSEG implemented Tick History within BigQuery, firms can obtain the best possible performance at a lower compute cost.

Performance statistics:

| Algorithm/test	Description	Result	Est. compute cost*
VWAP	VWAP = (Σ(Price * Volume))/(Σ Volume) on NYSE, one day of tick data (microsecond updates) – 165 million rows computed	1.7 seconds	US$0.00072
OHLC (bar)	Open-High-Low-Close interrogated all of NYSE for 15 days at 75 sec intervals – 4 million+ rows of data computed	27 seconds	US$0.02
Latency/ performance (VWAP/TWAP)	US desktop queried EU tick data: 150 million rows; ran two concurrent algorithms on same dataset; displayed results on iPhone in US over LTE	2.7 seconds to iPhone displaying 50,000+ rows of data	US$0.0014
Local use	Two algorithms ran concurrently, three-year date range on 155 million rows in the EU, porting 5,000 rows of results to local MS Excel instance in the U.S.	5 seconds	US$0.00062
Intraday summaries	LSE at 10-second intervals; single day of data – (Trades & Bid/Ask Quotes (difference)) + volume; 10 GB analysed and ordered in sequence 1.8 million records produced	23 seconds	US$0.02
	LSE at 10-second intervals – one month of data – (Trades & Bid/Ask Quotes (difference)) + volume; 255 GB analysed and ordered in sequence 4.75 million records produced	59 seconds	US$0.05
	LSE at 10-second intervals; one year of data – (Trades & Bid/Ask Quotes (difference)) + volume; 2 TB analysed and ordered in sequence; 475 million records produced	2 minutes 07 seconds	US$10.00
Real time	Last hour VWAP on entire NYSE < 10 sec delay vs. real time	1.9 seconds	US$0.0025

Around the globe, financial services firms are struggling to cope with a rapidly-evolving market data landscape which is creating a range of challenges. The key issue is that both data volumes and on-premises costs are rising and are expected to accelerate further. Other substantial issues include the need for greater agility and resilience, as well as the need to be able to use data in evolving use cases.

As a result, firms should consider shifting their market data strategy for tick history data from a system-centric to a data-centric approach. An important step is beginning to engage with tick history data and analytics in the cloud. Taking this approach enables:

– Data centre rationalisation

– Hardware decommissioning

– Resource and infrastructure reallocation

– Leveraging of new cloud capabilities for data science and analytics

– Migrating workflows to the cloud

As a result of these benefits and others, firms' ability to adopt data-centric cloud-based solutions for data, analytics and workflow will become an important source of competitive advantage.

LSEG Tick Data in Google BigQuery via CDN and LSEG Data Platform
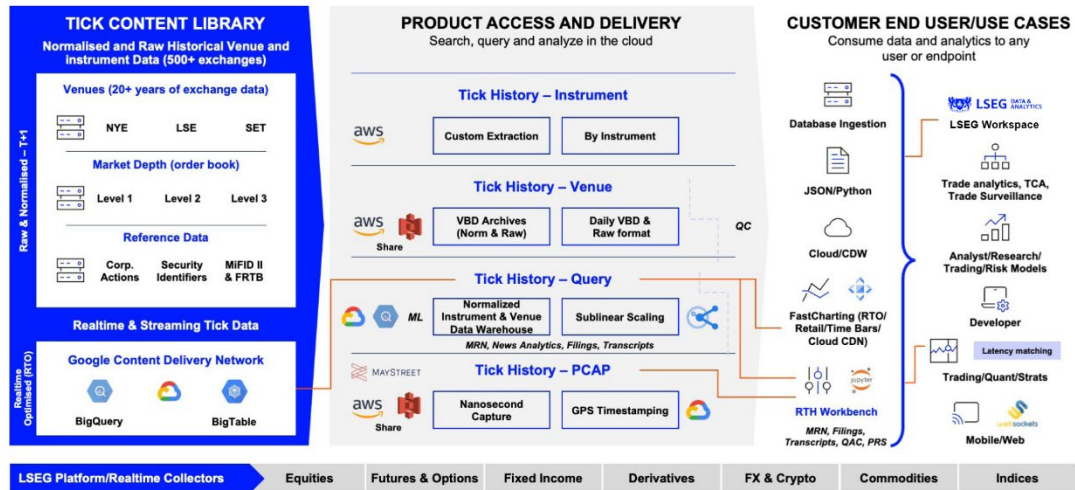
"LSEG Tick History – Query, powered by Google BigQuery, provides superfast querying and analytics, with ML-ready data output that can quickly be deployed into ML models or used to create and train new ones. The BigQuery engine is so powerful and flexible that this, coupled with the full history, Level 2 order book and trade qualifiers of normalised exchange data from LSEG Data & Analytics, makes it very easy to discover trading patterns and create decision treebased trading strategies. There is no need to extract, batch or break down data into sizeable compute chunks for the ML models – all queries and analytics can be run straight from the source data in the BigQuery engine, across days, weeks, months or even years of exchange data, for your ML workflows to learn and provide near-instantaneous insights."

Tim Anderson, Global Head of Tick History LSEG

Visit LSEG.com | @LSEG LSEG

LSEG DATA & ANALYTICS