TRANSFORM TICK HISTORY BY IMPLEMENTING A DATA-CENTRIC, CLOUD-BASED DESIGN

Moving tick history data to the cloud could result in significant cost savings of up to 90% versus storing data on premises. There are other powerful reasons for firms to shift this data set to the cloud, too. This Expert Talk explores today's challenges around on premises 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 approach engaging with tick history data sets. Certainly, the COVID-19 pandemic has spurred on 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 a host of other reasons why firms are moving to the cloud, too.

Facing Enterprise Challenges

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

Another theme is the need to rethink firms' data strategies in light of issues such as increasing data volumes, the need for more operational resilience, and evolving use cases for the data. Breaking these down into more specific day-to-day challenges, firms are finding they need to:

Data

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

- **Break down data silos and access a single source of truth** – Often in firms, the front, middle and back offices will have their own data silos, with multiple sets of the same data for their own use, rather than working from a single data source. This adds to cost, and can also create errors and questions about data integrity if the data sets have differences. Data consistency and integrity needs to be maintained across the front, middle and back offices

REFINITIV®

- **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 lifecycle
- **Create data lakes** – Financial services firms are deploying data lakes in the cloud: cast pools of data that can easily provide insight for both pre-determined and new queries or requirements across business functions

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, modelling and build data pipelines. Consistency across multiple markets and asset classes is key to ensure the data is prepared to satisfy multiple use cases and can be utilised 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, and the storage of tick data on premise creates a heavy burden when trying to engage in this approach. Rather than 'reinvent the wheel' and move tick history data to the cloud with a mere lift and shift, Refinitiv offers a ready-made solution for this challenge
- **Boost operational resilience** – COVID-19 has highlighted the need for operational resilience within market data strategies. The pandemic is accelerating what had already been 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 elastic cloud computing capabilities to perform analytics in the cloud. In addition, firms are very keen to reduce the time to insight, to 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 many firms have a systems-based design underpinning their overall market data strategy. With a systems-based strategy:

- Individual data sets 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
- Data sets 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 so often it is 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, and will face in the future.

"Taking a data-centric approach to the market data strategy for tick history is a big step towards true digital transformation for firms," says Catalina Vazquez, Head of Tick History Customer Proposition at Refinitiv. *"While dramatically reducing costs and enhancing efficiency, it also democratises the data across the whole organisation."*

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 at the heart of a firm's market data strategy, instead of the systems. 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 one 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.

Taking a data-centric approach to the market data strategy for tick history is a big step towards true digital transformation for firms," says Catalina Vazquez, head of tick history customer proposition at Refinitiv. "While dramatically reducing costs and enhancing efficiency, it also democratises the data across the whole organisation.

A good case study to explore how this works in practice is the adoption of Refinitiv Tick History in Google® Cloud Platform. Refinitiv 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. It's built on the Refinitiv universe of more than 70 million active and retired securities, with data coverage that goes back to 1996, and it contains more than compressed 8PB of historical data. Refinitiv 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**
  Refinitiv, in partnership with Google GCP, provide a fully managed time series database with a compute engine. No more 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 data set concurrently with no impact to 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 visualisation**
  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 data sets through a common identifier – apply intelligent tagging across many data sets.
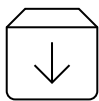
The Refinitiv Tick History in Google Cloud Platform is a managed service database that contains the full depth and breadth of Refinitiv Tick History data. Firms no longer need to clean, normalise and manage the tick history data – Refinitiv does that for them. Also, firms can query this database directly 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 other clouds that the firm has, 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 Refinitiv 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 Refinitiv Tick History in Google Cloud Platform and the Google BigQuery engine, it would take a mere 1.7 seconds to perform, regardless of locality.

The Refinitiv Tick History data in Google Cloud BigQuery is delivered via the Refinitiv Data Platform. The Refinitiv Data Platform 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 Refinitiv consistent data model, permissioning and entitlement capabilities, as well as cloud-based tools such as the Google BigQuery engine.

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

In summary, Refinitiv Tick History in Google Cloud Platform, coupled with the Google BigQuery engine, is able to address many of the significant challenges firms face today, by helping them reduce costs, improve data quality and access and dramatically increasing performance. Improving data quality and dramatically increasing performance. Looking deeper into how these services could reduce costs can provide a more in-depth example of what is possible.

Translating data-centric into reduced costs

As mentioned above, cost pressures around data management are a significant challenge for financial services firms – and particularly for the expenses associated with managing and maintaining tick history data stored on premises. The Refinitiv 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 the cooling, racking, networking connectivity and other supporting infrastructure such as built-in redundancy solutions
- **Storage hardware** – The cost of providing and storing database solutions for the data, along with the ability to retrieve the data for the 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. It is often a regulatory requirement to have this
- **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, for firms that want to continue storing their tick history data on premises going forward, the cost is more than $110 per gigabyte of stored data. When working with Refinitiv Tick History in Google Cloud Platform, the cost can be as low as $4.

Even migrating to Refinitiv Tick History in Google Cloud Platform 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 just a couple thousand dollars testing for 100 days.

To help foster a better understanding of why shifting to the cloud could make sense, the Refinitiv Tick History team is able to perform analysis on a firm that will uncover the scale of the existing costs associated with on premises tick history data, versus adopting Refinitiv Tick History data in Google Cloud Platform.

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

One example of the kind of insight that the team can bring is analysis of performance. It is often thought that a move to the cloud reduces the performance of an on premise local database. Due to the way Google and Refinitiv implemented Tick History within BigQuery, firms are able to obtain the best possible performance at a lower compute cost. See the chart for illustrative performance statistics.

See the table for performance statistics:

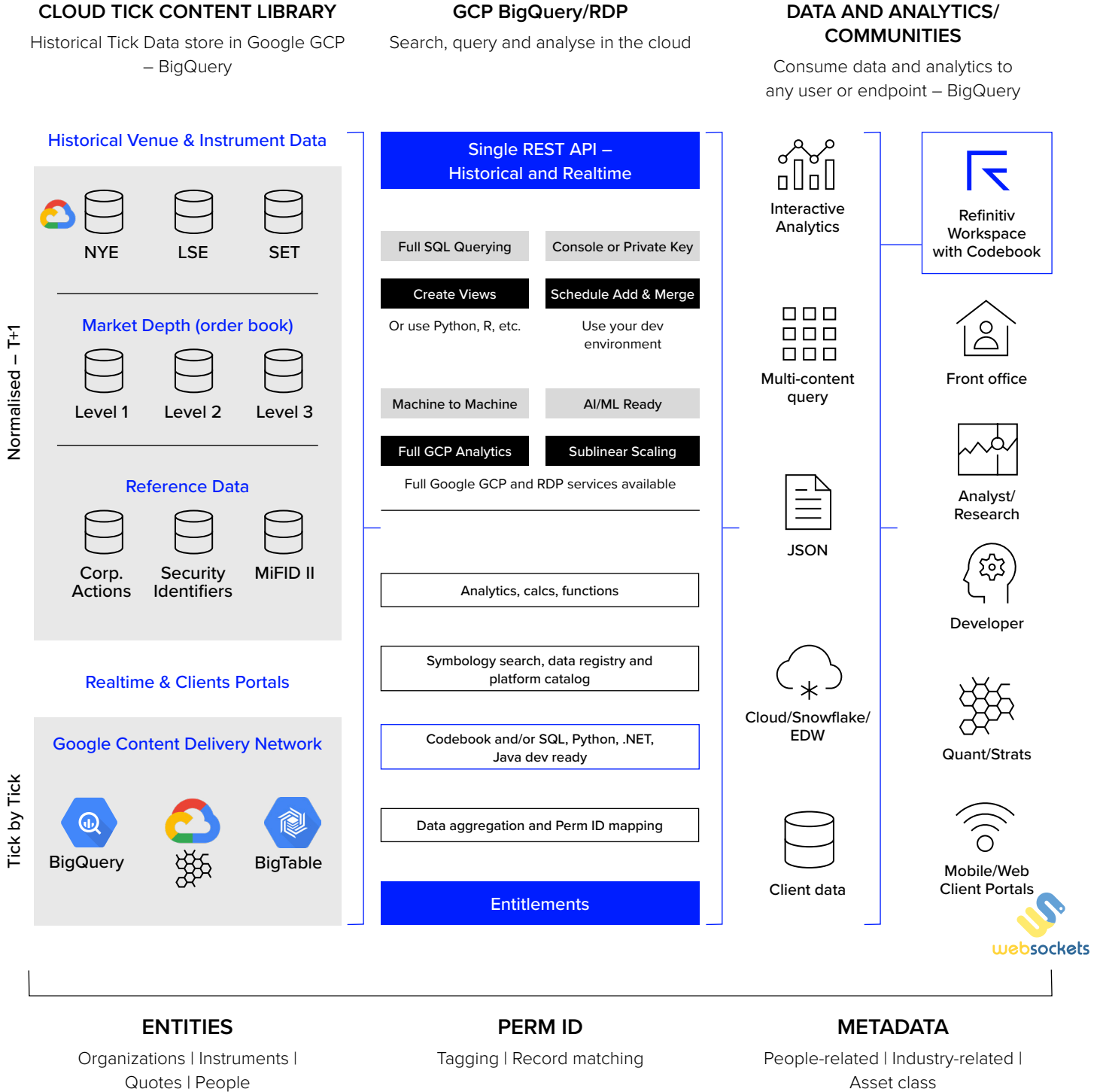| Algorithm/test	Description	Result	Est. compute cost*
VWAP	VWAP = (Σ(Price * Volume))/(Σ Volume) on NYSE, one day of tick data (microsecond updates) – 165 million rows computed	1.7 seconds	$0.00072
OHLC (bar)	Open-High-Low-Close interrogated all of NYSE for 15 days at 75 sec intervals – four million+ rows of data computed	27 seconds	$0.02
Latency/ performance (VWAP/TWAP)	U.S. desktop queried EU tick data: 150 million rows; ran two concurrent algorithms on same data set; displayed results on iPhone in U.S. over LTE	2.7 seconds to iPhone displaying 50,000+ rows of data	$0.0014
Local use	Two algorithms ran concurrently, 3-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	$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	$0.02
	LSE at 10 sec 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	$0.05
	LSE at 10 sec 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	$10.00
Real time	Last hour VWAP on entire NYSE < 10 sec delay vs. real time	1.9 seconds	$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 more 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 to begin 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, the ability of firms to adopt data-centric cloud-based solutions for data, analytics and workflow will become an important source of competitive advantage.

Refinitiv Tick Data in Google BigQuery via CDN and Refinitiv Data Platform

CLOUD TICK CONTENT LIBRARY
Historical Tick Data store in Google GCP – BigQuery

GCP BigQuery/RDP
Search, query and analyse in the cloud

DATA AND ANALYTICS/ COMMUNITIES
Consume data and analytics to any user or endpoint – BigQuery

Normalised – T+1

Historical Venue & Instrument Data

NYE | LSE | SET

Market Depth (order book)

Level 1 | Level 2 | Level 3

Reference Data

Corp. Actions | Security Identifiers | MiFID II

Realtime & Clients Portals

Tick by Tick

Google Content Delivery Network

BigQuery | BigTable

Single REST API – Historical and Realtime

Full SQL Querying | Console or Private Key

Create Views | Schedule Add & Merge

Or use Python, R, etc. | Use your dev environment

Machine to Machine | AI/ML Ready

Full GCP Analytics | Sublinear Scaling

Full Google GCP and RDP services available

Analytics, calcs, functions

Symbology search, data registry and platform catalog

Codebook and/or SQL, Python, .NET, Java dev ready

Data aggregation and Perm ID mapping

Entitlements

Interactive Analytics

Multi-content query

JSON

Cloud/Snowflake/ EDW

Client data

Refinitiv Workspace with Codebook

Front office

Analyst/ Research

Developer

Quant/Strats

Mobile/Web Client Portals

websockets

ENTITIES
Organizations | Instruments | Quotes | People

PERM ID
Tagging | Record matching

METADATA
People-related | Industry-related | Asset class

"Refinitiv Tick History on the Google Cloud Platform 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 Refinitiv, makes it very easy to discover trading patterns and create decision tree-based 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 close to instantaneous insights."

Tim Anderson – Solutions Business Director – Refinitiv

Visit **refinitiv.com** | @Refinitiv   Refinitiv

REFINITIV®

An LSEG Business