

Build versus buy SaaS platforms challenge old-school thinking on technology

Phil Anderson,
Head of Product Management
& Strategy

Background

After assessing the total cost of internal resources required to address data quality and transformation to ensure the data is fit for purpose, a global hedge fund aimed to reduce the total cost of ownership for key historical market data. The initial target was to save at least \$2m after optimizing the data operations and support team.

The financial target would be achieved by effectively managing multiple sources of historical market data and integrating alternate data more efficiently and accurately. This would not include the productivity benefit of the quantitative analytic team spending far less time on data wrangling and more time assessing current trading strategies and creating innovative new strategies to ensure consistent growth in trading returns.

Client situation

The hedge fund has multiple offices worldwide using a mix of overlapping data feeds. Proprietary feed handlers collect historical tick data from major exchanges for primary analysis, with licensed tick data as a secondary source to monitor the performance of existing trading strategies and models.

The fund had two key challenges:

1. Developing, backtesting and simulating new trading strategies in emerging markets and asset classes the fund did not actively trade.
2. The integration of multiple sources of historical tick data and the increasing number of alternative data for existing and new markets.

While each office had differing data requirements based on trading in different geographies and asset classes, there was a common need to integrate multiple tick data sources with alternate data across all analytical teams. These teams used a variety of cloud-based platforms and analytic environments.

The fund had licensed data from multiple market data vendors and, in many cases, had duplicate data feeds into multiple platforms.

A recent licensing of tick data had caused significant issues with data access, curation and overall quality. Senior analytical and data operations resources were forced to complete extensive data wrangling to organize the data and make it fit for purpose, reducing the time available for value creation activity.

Solution overview

RoZetta's solution to this challenge is DataHex, a cloud-based SaaS market data management platform that is data source agnostic. DataHex integrates multiple sources of trading data into a single platform with a single normalized data architecture. These were the major issues resolved once the client leveraged the DataHex SaaS platform which included:

Managing multiple data sources

Ingestion pipelines in DataHex were applied to provide a standard data architecture for historical tick data. This single source of cleansed and enhanced data eliminated a significant amount of data wrangling by senior resources.

It was an automatic productivity gain for the fund.

Symbolology

Instrument code changes are only sometimes available and when available, are often inaccurate, leading to requests for data returning an incorrect mapping.

All the symbolology issues were resolved as DataHex built and updated a Security Master as part of the ingestion process. This ensured an accurate symbol history and optimized the matching across data sets using the entity identifiers from each data feed.

Entity resolution and matching

The integration of the historical market data increased the entity matching rate across the data sources.

Implementing and maintaining a single Entity Master to complement the Security Master on all data in the DataHex platform provided a higher match rate and greater confidence in the match across multiple data sets.

Data enhancements

DataHex's ingestion pipelines enhance the historical market data during the load process by establishing and maintaining a Security Master, implementing a single set of normalization rules, and calculating market metrics, like One-Minute Time Bars. Users have access to critical metrics immediately after data loading is completed.

In this case, the client required additional fields appended to the standard One-Minute Timer Bar format. The data arrived analytics-ready.

Option chains

Option chain expansions lead to arbitrary instruments included and additional quota charges; some underlying instruments were excluded during some time periods.

Options chains were validated, and incorrect and unnecessary instruments were removed. Chains were reconciled, and values were recalculated.

Additional value add fields

The non-availability of fields like the Bid Trade Value and Ask Trade Value were unavailable in intraday reports.

These fields were computed from source data and incorporated into the ingestion pipeline.

Corporate actions

A requirement to identify the price impact of corporate action from the dividend records to avoid double counting the dividend.

Corporate Action references were validated and updated where necessary.

Reference data

The non-availability of instrument expiry dates, multiple expiry dates for some instruments, asset category not included, and incorrect field lot units for some instruments.

Values were either refreshed from source data or replaced by the data vendor after contact.

UNLOCK THE FUTURE

The client outcome

Implementation of the DataHex solution would exceed the targeted savings by more than 20%. It is estimated that a productivity gain of 30% - 40% for senior analysts will boost the capacity of the fund to develop new trading strategies.

Further, numerous case studies show that passive management of data vendors leads to duplicated data licensing, further compounded by repeatedly fixing core data issues by many users. These data fixes are typically not socialized across a common platform or fed back to data vendors to remediate at the source. DataHex's ability to address data quality issues and transform data extracts to multiple cloud platforms and analytical tools formats was a real game changer for the client.

About RoZetta Technology

RoZetta Technology's core belief is that fusing data science, technology, and data management is the path that amplifies human experience and knowledge. RoZetta's DataHex cloud platform enables organizations to accelerate speed to market insights and then create value by encouraging data-driven decision-making.

RoZetta brings proven capability, experience, and a mindset to develop products and systems that resolve these challenges to create value. Blending data science, data engineering, and domain knowledge, RoZetta has a proven track record of exceeding clients' expectations.

Your productivity challenge

RoZetta has developed a model that estimates the value of moving to our cloud-based data management model. With very few inputs RoZetta can provide an estimate of the financial gain based on moderating operational costs, vendor management, increased productivity and increasing revenue growth opportunities.

[Contact us today](#) to learn how we can enhance your productivity and overall data management efficiency.