This paper illustrates how Xenomorph’s real-time analytics and data management system, TimeScape, enables extremely rapid and extensible analysis of tick and intraday timeseries data delivering competitive advantage in pre- and post-trade decision support.
# Table of Contents

A Time of Change and Opportunity  ..................................................................................................... 3  
Data Transparency – Increased Productivity for All ................................................................. 4  
Tick Capture – Flexibility and Ease of Use Together .............................................................. 6  
Tick Storage – Real-Time, Intraday and History All in One ...................................................... 7  
Data Access - for Both Traders and Technologists ............................................................ 8  
  Loading Intraday Data with Functions ........................................................... 8  
  Loading Intraday Data with TimeScape QL+ ....................................................... 9  
Data Validation – Why Waste Half Your Time? ............................................................ 10  
  Data Frequency and Time Snapping .......................................................... 10  
  Data Filling, Aligning and Loading Rules .................................................... 10  
  Real or Interpolated Data? ................................................................. 11  
Data Analysis – From Backtesting to Best Execution .................................................. 12  
  Chaining Analytical Functions ............................................................... 12  
  Intraday Time Period Analysis .............................................................. 13  
  Beyond the Spreadsheet ................................................................. 15  
Event Processing – Real-Time Analysis Automation ................................................ 16  
Your Proprietary Advantage ................................................................. 17  
Summary ............................................................................................................. 18
A Time of Change and Opportunity

Data management in financial markets are being driven through a period of fundamental change. Trade volumes are increasing exponentially as electronic execution delivers faster trading with ever-tighter margins. The proprietary algorithms used in algorithmic and statistical arbitrage trading are becoming more complex. Developments in areas such as credit theory are establishing market relationships that motivate more complex cross-asset trading strategies. Regulations such as MiFID and Regulation NMS are pushing the whole industry towards better and more transparent execution, but are also fundamental drivers behind both huge business change and dramatically increased data volumes. All of these factors are combining to provide both profit and cost incentives to move away from single asset class data silos.

Looking at data management from a trader’s perspective, then a decade ago many practitioners were content with analysing end of day historic data for strategy back-testing and instrument pricing purposes. It should be said that they quite possibly had no choice from a technological perspective; the capture, storage and analysis of intraday data volumes even then was challenging, especially at a time when the relational database was still a relatively new technology. Given however that derivative pricing margins were wider and statistical arbitrage was profitable using end of day prices, then there was also little incentive to store and analyse intraday tick and high frequency data. Much tighter trading margins, cross-asset trading and improved technology have changed traders’ perceptions of what is required and what kind of analysis is possible with high frequency intraday data.

Risk management has previously not been greatly concerned with intraday data. As many risk managers will confirm, obtaining clean data for end of day risk measurement is challenging enough. Risk measurement techniques such as monte-carlo or historical simulation VaR require large amounts of historical data and are calculation intensive. Large data universes or poor implementation may mean that it is challenging to attempt to run these techniques as an overnight batch, let alone perform the calculations in real or near real-time. Increased intraday trading exposure, better understanding of intraday market behaviour and recent regulatory requirements concerning data transparency and data quality are driving risk managers towards more analysis of tick by tick and intraday data.

The changes above require systems that can adapt to the pace of change, delivering high performance analysis even when the quantity of intraday data being analysed is massive. This paper describes how Xenomorph’s real-time analytics and data management system, TimeScape, has been designed to meet these current challenges and to deliver competitive advantage in pre- and post-trade decision support.
Data Transparency – Increased Productivity for All

One of the first design criteria for TimeScape was to make data accessible by both business users and technologists alike. As a result, much of the data presented in TimeScape is designed to directly represent business-friendly financial objects such as instruments, curves, indices and so on, and the system does not require end users to understand database technology or table-based data models. The same approach is taken when viewing tick data, as is shown in Figures 1 and 2 below:

Figure 1 shows the stock Imperial Chemical Industries (ICI) being viewed via the TimeScape WorkBench. The view shows intraday data captured from the London Stock Exchange, and in particular the number of ticks per day for intraday trade prices captured for the stock. Once a particular date has been selected, then ticks for that day can be observed and charted in the WorkBench as shown overleaf in Figure 2.
Figure 2 – Viewing Tick History for One Day

The left hand side of Figure 2 shows how firstly the stock ICI is selected and how the properties available for ICI are expanded for further browsing by the user. Upon selecting the trade price ("TradePrice") property for ICI, all of the available data sources for this property are displayed. Whilst there could be multiple data sources, in this case there is only one, that of “LSE”. Underneath the “LSE” data source each day’s history is displayed. The right-hand side of the application is split into two, one listing the tick times and values for the day selected on the left (in this case the 12th of January 2006), the other charting the ticks for this same day. This way of hierarchically browsing around both instruments and data is used throughout TimeScape in order to increase the productivity with which business users can view, verify and manipulate data.
Tick Capture – Flexibility and Ease of Use Together

TimeScape has been designed to make the configuration of real-time data capture very intuitive. Rather than forcing users to directly base data capture on scripting, TimeScape also offers a Tick Capture front end within the TimeScape Workbench. It allows multiple data attributes to be selected for real-time capture and monitoring, and provides easy configuration of how the download is to be monitored, filtered and logged. It is illustrated below in Figure 3, previewing a Reuters data universe defined by the user.

The Tick Capture front end allows instrument universes to be modified and monitored in real-time, and offers the following methods of defining the instruments universe to be captured:

- Manual selection of instruments to be added
- Instrument selection driven by the instruments universe in an existing database
- Instrument universe defined by a dynamic query

Using the query option, a tick capture process could be set up that only captures data for stocks within market capitalisation between certain levels, or that are part of an index and that are of a certain level of volatility. This query can also be configured to automatically refresh itself so that no new instruments are ever excluded from the capture process.
Tick Storage – Real-Time, Intraday and History All in One

Storage of intraday data is becoming increasingly challenging due to the data volumes involved. A highly traded stock may have tens of thousands of price events per day, quickly resulting in a storage requirement to store Gigabytes of data per day and Terabytes of data per year for any reasonable sized instrument universe.

TimeScape’s primary database engine (TimeScape XDB) delivers this high volume data storage capability, having the capability of storing up to 3.2 Petabytes (3,200 Terabytes) of intraday data in one database. Assuming an average tick rate of 2,000 ticks per day per instrument, this maximum database size equates to being able to store over 270 million years of trading data in one database. Figure 4 below illustrates how TimeScape XDB fits within the TimeScape architecture as a whole:

Figure 4 – TimeScape XDB and System Architecture

This storage capability of TimeScape XDB is coupled with high performance data retrieval at speeds beyond that possible with mainstream relational database technology. This means that large calculations with intraday and time series data that ordinarily run as a batch job can be run in real-time and near real-time, again delivering competitive advantage in implementing trading strategies, back-testing of trading strategies, risk management and transaction cost analysis.
Data Access - for Both Traders and Technologists

Loading Intraday Data with Functions

Loading tick data into Excel or any of TimeScape’s programming interfaces is extremely simple. TimeScape comes with an extensive range of functions for loading, saving and analysing intraday data. One of the simplest is the XLoadNSeries() function illustrated loading intraday data into Excel as in Figure 5 below.

The example shown in Figure 5 illustrates the loading of a month’s worth of tick data for a number of banking stocks, sourced from the London Stock Exchange. The data has been aligned with the data series on the left, Barclays, and where unavailable data for the other stocks has been interpolated using log linear interpolation. This function aligns, interpolates and returns around 230,000 data points in a fraction of a second into Excel, and is available in equivalent form across all of TimeScape’s APIs, enabling functionality developed as Excel prototypes to be implemented within enterprise-class thin client applications.
Loading Intraday Data with TimeScape QL+

TimeScape comes with several hundred functions within its Excel and programming interfaces, building upon the functionality of core functions such as XLoadNSeries() as shown in Figure 5. Whilst these functions offer a very compelling compromise between ease of use and power for the business user, in terms of flexibility and extensibility, many business and technical users require and expect more.

In order to address this need for an even more powerful way to access intraday data, the system also contains an object oriented query language, called TimeScape QL+, which has been designed specifically for financial markets and offers the following features:

- User-friendly representation of financial objects and data
- Extensive frequency conversion facilities for historic data
- Vector arithmetic of historic data for use in custom index/basket/spread calculations
- Extensive rules for data preferences, interpolation, alignment and proxying
- Support for complex data types such as arrays, curves and matrices.
- Self-describing, context sensitive query building
- Easily extensible through the addition of user-defined analytics and objects

A simple example of a TimeScape QL+ query is shown below, where the trade price and trade size intraday tick series are multiplied together to give a resultant series containing a series of the value of each transaction over time:

```
Item("InstXDB", "BARC.L").TradePrice * Item("InstXDB", "BARC.L").TradeSize
```

The above query example shows two series (TradePrice and TradeSize) being loaded for a single instrument (BARC.L). If we wished to extend the example to deal with multiple instruments we could use a query such as the one below:

```
Database("InstXDB").Items("Equity").Where(SharesIssued > 10,000,000)
 .Values(Code, Description, (TradePrice*TradeSize).Average)
```

This example takes all of the equities in the “InstXDB” database, only including those that have more than 10 million shares issued and returns the values of each equity instrument’s code (e.g. Reuters RIC), description and finally calculates the average transaction value through time.

The query examples shown above are relatively simple in nature, however they emphasise that in order to do quite powerful analysis the user does not need to be aware of any underlying database table structure or complicated query syntax. Again this re-emphasises a fundamental premise of the design of TimeScape is to bridge the gap between business and technological groups, allowing both to deliver more business value more rapidly.
Data Validation – Why Waste Half Your Time?

One of the key issues in analysing market data, and in particular intraday data, is that many traders, analysts and risk managers spend far too much of their time checking, validating and interpolating data and not enough time on making decisions with it. For many kinds of analysis, it is essential that the raw database data stored from the market be adjusted in some way prior to applying a calculation. For example, the data may need rebasing in frequency, aligning across multiple series or filling in some way.

In standard SQL databases this kind of manipulation is very difficult to do, requiring much post-processing of queried data. This has the net result of reducing the usefulness of SQL, increasing the complexity of the queries written and limiting the analysis that can be done by non-technologists. In order to address this issue, TimeScape QL+ has been designed to have “data rules” that set a data context for the data to be loaded in a query or calculation. Some examples of data rules are shown below, the key point being that by separating out the data rules, the queries themselves can be kept simple and hence more productive for both technologists and business users alike.

Data Frequency and Time Snapping

TimeScape QL+ makes it very easy to change from tick-by-tick (irregular time frequency) to some other more regular basis as shown using the “.Data” operator below:

```
Item("InstXDB","JPY=").MidRate.Data('10-Sep-05 14:00', '20-Sep-05 16:30', '5 min')
```

The example above illustrates how Japanese Yen to US Dollar intraday mid rate data can be converted from irregular frequency to a frequency of 5 minutes from 2pm on the 10th of September 2005 to 4:30pm on the 20th of September 2005.

Data Filling, Aligning and Loading Rules

In the data frequency example above, it cannot be guaranteed and indeed it is unlikely that there will be a traded price observed at every 5-minute frequency point. Should the user require that the latest, next or an interpolated price is used at each point, then TimeScape provides extensive data rules to automate this. For example, the user could run the above query in the context of the following data rule:

```
FILL = [MISSING_RULE = TAKE_PREV WITHIN 5 min; CALC = LINEAR_INTERP]
```

If no data was found for a particular 5-minute time snap, then the rule above specifies that linear interpolation should be used to calculate any missing data based on previous value stored in the database, so long as it is no longer than 5 minutes old. Building upon these “fill” rules outlined above, TimeScape also provides rules for data alignment when multiple series are being loaded. The example below aligns dates to the left most series in the query, using a form of the “FILL” rule above to generate data for any missing data in the alignment:

```
ALIGN = [MISSING_RULE = TAKE_PREV; DATES = LEFT]
```
Another data management issue that can arise is where multiple data sources are available, and the user has some kind of preference for which sources should be used first. Running any query in the context of the data rule below would mean that Bloomberg data would be used where available in preference to Reuters:

\[
\text{LOAD} = [\text{DATA\_SOURCES} = \text{Bloomberg : Reuters; STATUS\_FILTER = Approved}]
\]

Additionally, for “official” end-of-day data used for risk and compliance purposes, TimeScape allows time series to have a status level associated with each data point. In the data rule above, only data that has an “Approved” status level set is returned to the calculation.

Real or Interpolated Data?

The data rules briefly outlined above are extremely powerful, but as a result a key requirement for anyone using them is to quickly understand the effect of applying them to the data loaded in a query. An example query to do this is shown in Figure 6 below:

![Example Query Excel Screenshot](image)

**Figure 6 – Example Query Explaining Effects of Data Rules Applied**

This example query converts tick data to an hourly frequency for Barclays, retrieving date/time, value, the data status and a data status explanation. It can be seen that none of the data has been officially validated as shown in the data status column. The data point explanation column tells the user how the data value was arrived at i.e. is the data value stored or has it been interpolated? This transparency of understanding around data rules and their effects is vital to the validity of any calculation or report implemented.
Data Analysis – From Backtesting to Best Execution

Data only becomes valuable when data analysis translates it into information. Why store vast amounts of market data if you don’t have the capability to analyse it? The current competitive challenge in the market is to apply ever more complex analysis to ever-increasing volumes of data.

Backtesting for automated and statistical trading strategies is growing in importance as a pre-trade decision support function. Within risk management, intraday and cross-market trading is driving the need to analyse intraday data more fully and faster. Regulation is now also driving data analysis in areas such as historic verification of best execution policies.

In order to deal with these expanding data analysis requirements, both traders and technologists need the tools to analyse more data, more quickly. It is no longer sufficient to solely rely on technologists to translate business requirement into business analytic. Such a process makes technology staff the business bottleneck in responding to trading requirements. It also means that the technology staff are hit with daily tactical requirements from trading that inhibit the timely delivery of strategic systems and processes.

What is needed is a strategic way of alleviating this tactical pressure point, providing a transparent, easy to use and powerful framework for data analysis across all asset classes. TimeScape QL+ provides such an analytical framework, as we shall illustrate in the examples below.

Chaining Analytical Functions

TimeScape QL+ allows calculations to be chained together to perform more complex analysis. Building upon one of the previous examples to simply load some data, we could extend it in the following manner:

\[
\text{Item("InstXDB","HBOS.L").TradePrice.Volatility(20,365.25,","T").Average}
\]

This query takes all of the trade price data for HBOS, calculates a rolling twenty point volatility (adjusting automatically for irregular data frequency) and then calculates the average of these volatilities. When you combine the analysis capabilities shown above with some of the data rules and frequency examples explained previously, it is possible to achieve some quite complex pieces of analysis without having to write complex queries.
An example of these powerful analysis capabilities is shown below in Figure 7:

![Figure 7 – Volatility Measured as a Market Closes]

In this example, the `LoadNSeries()` function described previously is combined with TimeScape QL+ to convert historic tick data series into daily time series at different times of day, just as the market is closing. It can be seen how the volatility calculated at market close is much lower than that observed just a few minutes earlier. This kind of intraday measurement of volatility and other measures such as correlation is proving increasing important to the pricing and risk management of derivatives, and in the formulation of new trading ideas.

**Intraday Time Period Analysis**

The example in Figure 7 shows how the behaviour of a market can apparently be very different dependent upon what time during the day you observe it. Looking at this issue in a more granular way, TimeScape QL+ has the ability to split a trading day into any number of time buckets and return values and calculations during each bucket:

```
Item("InstXDB", "AAPL.OQ" ).TradePrice
 .PAnalysis("10min",,,"Open", "High", "Low", "Close", "Count", "Average")
```

The above example would return values for the open, high, low and close prices plus the number of points and average price for each 10-minute period bucket applied to the intraday price history of Barclays PLC.
Using the PAnalysis() function in a slightly different way, we can calculate a time series of VWAP (Volume Weighted Average Price) values as shown in the interactive query environment of Figure 8 below:

In the query above, intraday trade price and associated trade volume of Barclays are used to calculate a time series of VWAP calculations using a time window of 15 minutes. Firstly, within each 15-minute window, trade price is multiplied by trade size and the PAnalysis() function is used to sum this time series of multiplied values. This resultant time-series is then divided by another 15-minutely time series produced by using PAnalysis to sum the sizes traded within each 15-minute window.

The result is a time-series containing the VWAP for each 15-minute window, which if desired can be analysed further with TimeScape QL+ or easily customised to a particular users needs. The flexibility of QL+ allows many and varied ways of calculating VWAP and other forms of transaction cost analysis (TCA), and these calculations can then be designed to automatically feed through to processes designed to backtest and validate a client’s best execution policy against real-time and historic market data.
Beyond the Spreadsheet

The spreadsheet is the tool of choice for many market participants wanting to analyse market data. Whilst spreadsheets are an extremely efficient tool for ad-hoc market analysis, they can be limited in the volumes of data that they can be used to analysed. In order to accommodate the productivity of spreadsheet analysis without the limitations on the amount of intraday data that can be analysed, TimeScape combines the best of spreadsheet design with database transparency within its SpreadSheet Inside technology.

This technology allows spreadsheet-like interaction with market data to define new analytics and reports, together with centralised, transparent storage and execution of the calculations defined. Taking the VWAP query example from Figure 8, SpreadSheet Inside technology allows the user to define a “virtual” property that is able to execute powerful spreadsheet like logic as shown in Figure 9 below:

Additionally, and of particular importance for intraday data, this spreadsheet environment allows very large time series to be placed within single cells. In Figure 9 above, cell “A1” has been set to be “=TradePrice” telling the system to populate this single cell with hundreds of thousands of prices, and cell “A3” contains the time series result of multiplying together trade price and trade size series. This is just one example of how new analytics can be designed, locked down and centrally deployed to users almost instantly, enabling TimeScape to lead the way in real-time deployment of real-time data analytics.
Event Processing – Real-Time Analysis Automation

Analysing high frequency data can reveal new trading opportunities within the market, however seeing what these opportunities were is no use if you are never in a position to capture them as they happen. Put another way, understanding history may help you understand what could happen in the future, but the trading opportunity is now.

The need to process and analyse market events in real-time has driven the development of TimeScape Event Services, an asynchronous event processing engine that runs in parallel with the Tick Capture capabilities described earlier in this paper. Figure 10 below shows how Event Services maintains low latency for event propagation and high frequency data storage:

![Figure 10 – TimeScape Event and Tick Capture Services Architecture](image)

Real-time streaming data from different market datafeeds are passed into TimeScape Event Services. These feeds are connected into TimeScape Event Services through a common API, a thin normalisation layer that enables downstream client applications and processes to be insulated away from the complexities of each vendor’s individual technology and interfacing preferences.

Client applications can optionally subscribe to listen to events within TimeScape Event Services, and such clients can then be connected to trade execution engines to automate trading strategies. TimeScape Tick Capture Services is also a client of Event Services, however a key design consideration was that Tick Capture should run in parallel with client subscriptions, and not affect the latency of the events being notified to the clients.

Once the data of interest is being processed within Tick Capture, the user can configure the type of buffering and validation desired as the data is persisted to the TimeScape database. As TimeScape deals with real-time, intraday and historic tick data in this unified and consistent manner, more complex analysis can be implemented more quickly.
Your Proprietary Advantage

Whilst TimeScape has a rich library of functionality for intraday data analysis, it is the ease with which TimeScape can be integrated with a client’s proprietary analytics and systems that sets it apart from the competition.

Figure 11 below shows the result of using TimeScape’s context sensitive help facility. The “Item("InstXDB", "BARC" ).TradePrice” is a simple example query, requesting that TimeScape simply load the intraday trade price series for Barclays. Adding “?” to the end of this query causes TimeScape to display all functions and properties that are valid in the context of analysing an intraday time series such as TradePrice:

![Figure 11 – Extendable Analytic and Function Library](image)

Of particular note in the example above is that Xenomorph has used this concept to automatically link TimeScape QL+ with any Excel XLL addin library, outside of the Excel environment. This is illustrated with the functions prefixed with “aa” in the list above, which are part of the FinCAD (www.fincad.com) analytics library.

Even though TimeScape comes with over 100 analytical functions that apply to intraday time series, the key aspect is that it is possible to quickly write addin extensions to TimeScape QL+ to integrate your proprietary analytical libraries into the system. Once integrated and deployed centrally, your analytics become natively available to all users of the system, and can be applied quickly and easily to the huge datasets available with TimeScape XDB.
Summary

So what does high frequency data analysis require? Fundamentally, the current needs of the market translate into capturing, cleansing, storing and analysing ever more data, ever more quickly.

Building on the above fundamental market requirement, Xenomorph’s additional take on this field is that more analytical power needs to be put in the hands of the people who make trading and risk management decisions. The productivity gains from such an approach are many fold, particular when it is based upon a foundation that delivers automated data cleansing, centralisation and transparency to IT and compliance departments.

The speed with which new trading ideas, new financial products and new risk management techniques can be brought to market is a key competitive issue within financial markets. Xenomorph believe that TimeScape is the enabling platform to deliver this competitive advantage.
About Xenomorph

Xenomorph delivers Analytics and Data Management (ADM) solutions to the financial markets. Our TimeScape technology leverages our clients’ proprietary expertise, enabling them to analyse and manage more data with greater control and transparency.

Our focus is to make our clients more successful by closing the productivity gaps between high performance database technology, data management and end-user data analysis. Through unified and transparent access to data and data analysis, our clients achieve even higher levels of financial innovation, business process efficiency and regulatory compliance.

Trading, research, risk, product control, IT and back-office staff use Xenomorph’s TimeScape data management platform at investment banks, hedge funds and asset management institutions across the world’s main financial centres.