



it's about time



# Time Series Analytics for Big Fast Data

---

Jonathan Guy  
June 2018

- Founded in 1993
- Subsidiary of First Derivatives Plc (LON:FDP)
- 2,000+ employees
- Global presence

# Customers & Partners



THOMSON REUTERS



J.P.Morgan

Morgan Stanley



# Other Verticals



## Health & Life Sciences

*Genomics Data Processing*  
*Connected Health*  
*Patient Record Analytics*



## Automotive

*Telemetry Analytics*  
*Anomaly Detection*  
*Performance Analytics*



## Utilities

*Sensor Analytics*  
*Smart Meters Data*  
*Predictive Analytics*



## Space and Telco

*Earth Observation*  
*Geospatial Data Analytics*  
*Anomaly Detection*



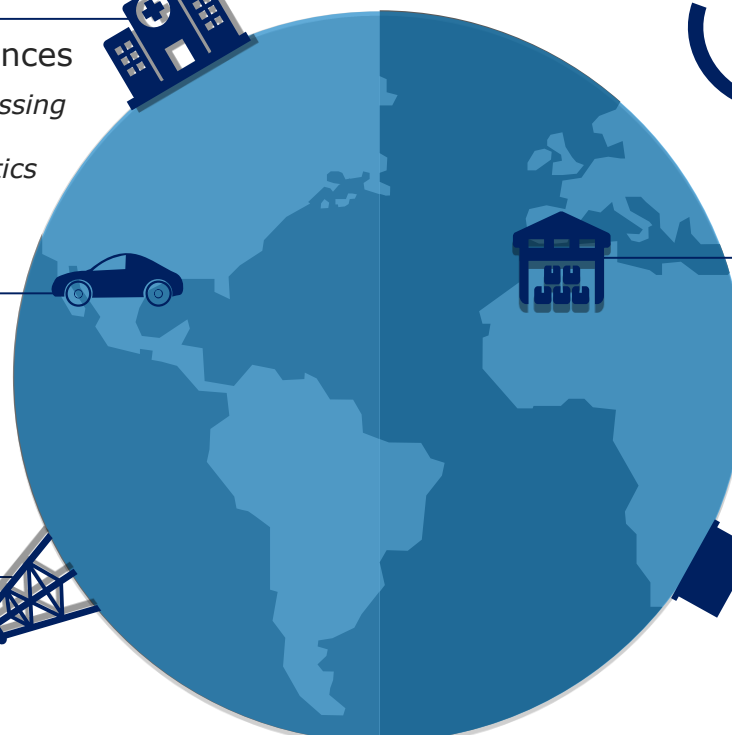
## Retail

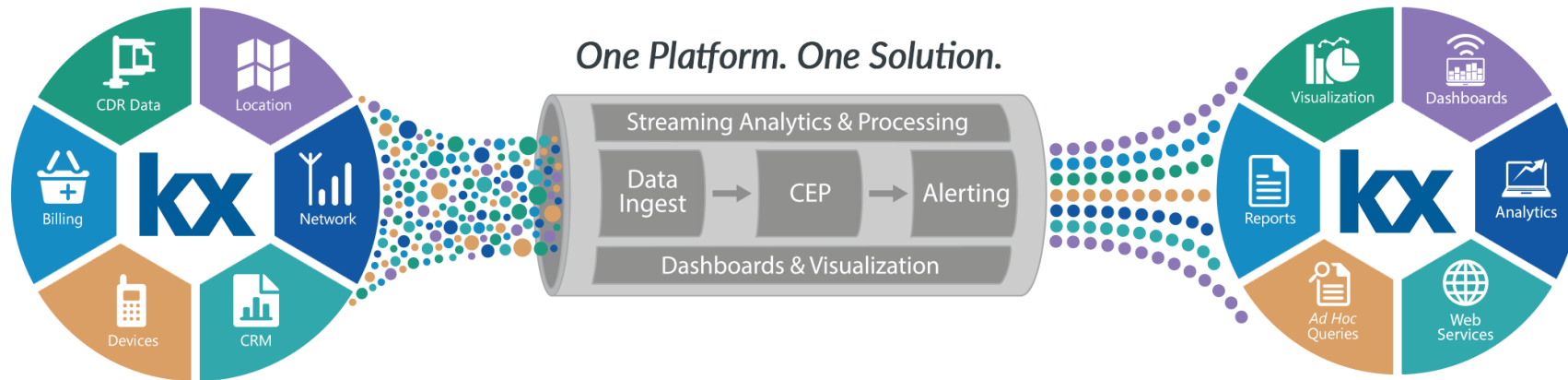
*Retail Analytics*  
*Marketing Optimization*  
*Customer Journey*



## Manufacturing

*Edge Computing*  
*Multivariate Analysis*  
*Fault Detection*





## *Limitless New & Existing Sources*



Network  
Data



Real-time &  
Historical Data



Fully Managed  
Data Quality

## *Real-time Actionable Insights*



Extreme  
Performance



In-memory,  
Relational &  
Columnar

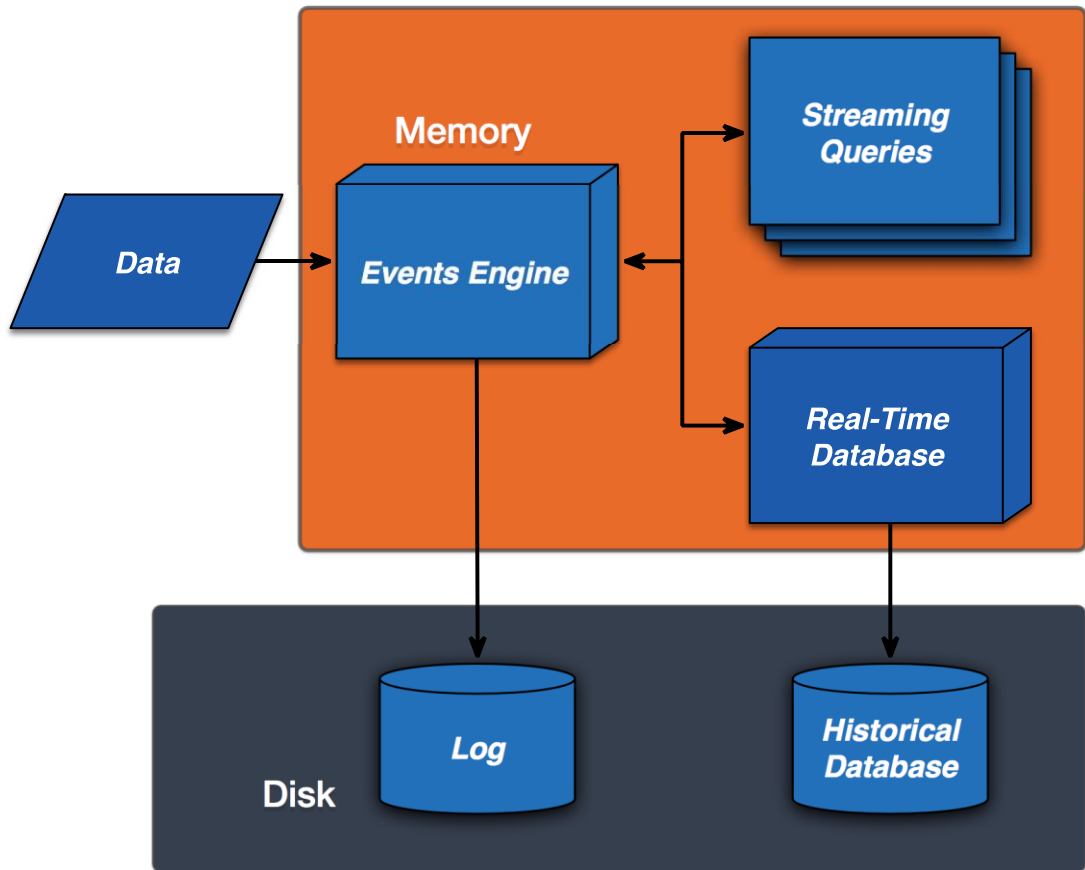


Powerful  
Query Language

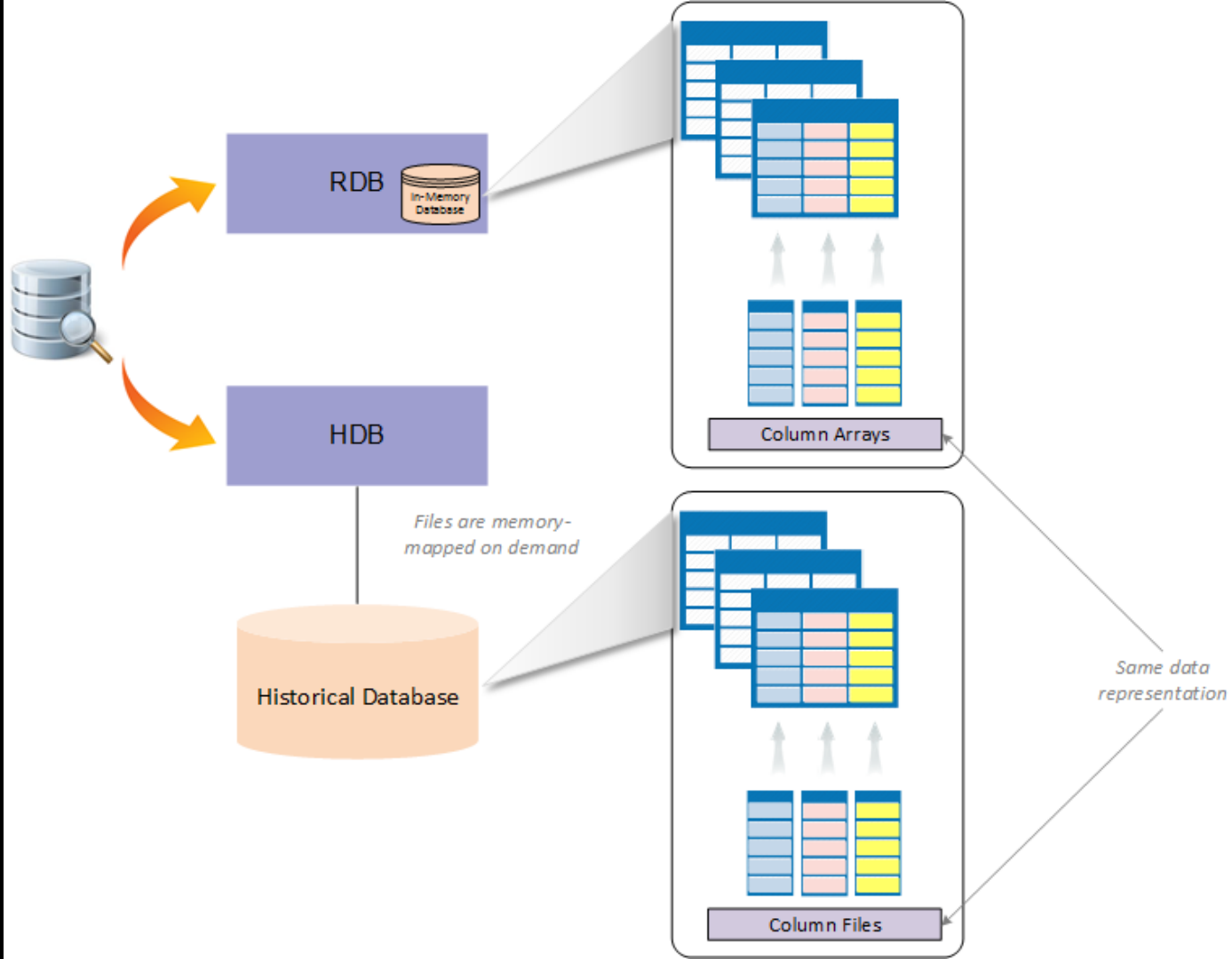
- Unified columnar database & programming system
- Lambda architecture
  - Streaming
  - Real-Time (In-Memory)
  - Historical (On-Disk)
- In-database analytics
- Support for joins
- 500 kB binary



# Typical kdb+ Architecture









## Database Data Types

- Boolean
- Byte
- Integer (short/int/long)
- Floating Point (real/float)
- Character
- Symbol (interned character)
- GUID
- Enumeration
- Dictionaries
- (Keyed) Tables
- Functions

## Time Series Data Types

- Date
- Time
- Minute
- Second
- Month
- Datetime
- Timespan (ns)
- Timestamp (ns)

Executions								
Time	Instrument	Side	Price	EBBO		ExecutionLag	ExecutionFees	BBOShortfall
				BestBid	BestAsk			
2016-01-12 17:01:43	VOD.L	S	220.3000	220.3500	220.2500	00:00:00.649	21.0464	2.2701
2016-01-12 17:05:26	VOD.L	S	220.6000	220.6000	220.6000	00:00:00.772	35.7372	0.0000
2016-01-12 17:08:19	VOD.L	S	220.4500	220.4500	220.4500	00:00:00.627	14.4141	0.0000
2016-01-12 17:10:11	VOD.L	B	220.7500	220.8000	220.7500	00:00:00.173	18.3675	-2.2645
2016-01-12 17:13:03	VOD.L	B	220.5500	220.6000	220.5500	00:00:00.620	39.6990	-2.2665
2016-01-12 17:15:08	VOD.L	S	220.7500	220.8000	220.7500	00:00:00.883	149.0063	0.0000
2016-01-12 17:18:02	VOD.L	B	221.2000	221.2500	221.2000	00:00:00.420	43.7976	-2.2599
2016-01-12 17:20:44	VOD.L	B	221.1500	221.1500	221.1000	00:00:00.042	40.7823	0.0000
Showing all 395 rows								
Daily Stats								
Instrument	AveragePrice	VWAP	Low	High	Open	Close	TotalVolume	
VOD.L	221.79	221.24	218	223.25	220.45	221.88	93358379	
VOD.LBS	221.84	221.88	219.55	223.00	220.30	220.44	6698765	
VOD.LCHI	221.75	221.74	219.55	223.00	220.15	221.15	14983028	
VOD.LTQ	221.76	221.76	219.55	223.00	220.15	221.20	13823647	

- Functional
- Array/Vector
- Query
- Interpreted

- Time Series Functions/Joins
  - xbar
  - Bi-Temporal
    - aj (As-of join)
    - wj (Window join)
- Temporal Arithmetic





OMG I CAN'T BREATHE 🤔😭



## Parallelism

- Vertical Scaling
  - Multi-threading
  - No serialization
  - Automatically distributes queries across CPU cores
- Horizontal Scaling
  - Multi-processing
  - Automatically distributes queries across machines

## Compression

- WebSocket compression
- In-flight compression between hosts
- On-disk compression
  - kdb+ algorithm
  - gzip
  - Google Snappy
  - lz4c



# Kx Performance Snippets

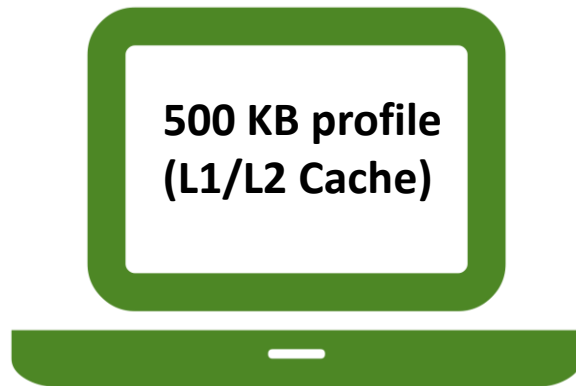
Process & store  
**4.5 million**  
events/second/core

Ingest data at  
**10 million**  
records/  
second/core

Search in-  
memory  
tables at **4**  
**billion**  
records/  
second/core



Trusted by 19/20  
World's Top  
Investment Banks

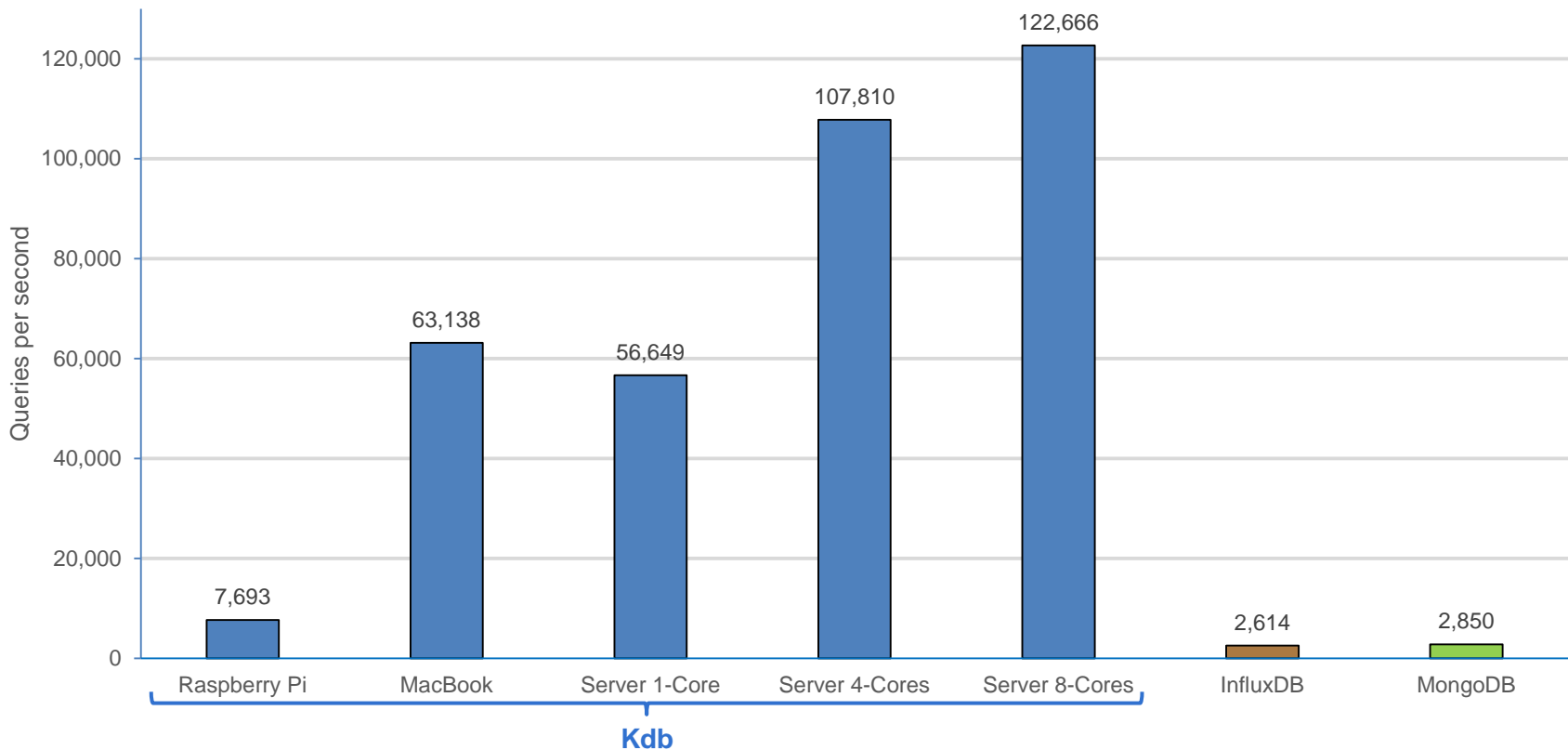


**Streaming**

**1.6 TB**

**of Data Daily**

# Query Rate: Kdb+ vs InfluxDB vs MongoDB



On similar servers, kdb+ is over **40 times faster** than InfluxDB and MongoDB





A diagram showing eight blue circles arranged in a circular pattern, each containing text representing an enterprise interface. The circles are arranged in two rows of four. The top row contains 'R & Matlab', 'C/C++', 'Python & Perl', and 'Web Services'. The bottom row contains 'C#/.NET', 'CDC', 'TCP Sockets & Web Sockets', and 'ODBC/JDBC'. The 'Java/Scala' circle is positioned between 'R & Matlab' and 'C#/.NET'.

R & Matlab

C/C++

Python &  
Perl

Java/Scala

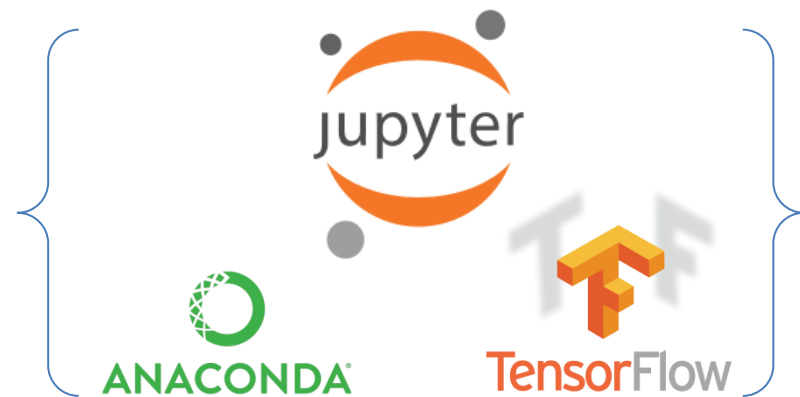
ODBC/  
JDBC

Web  
Services

C#/.NET

CDC

TCP Sockets  
& Web  
Sockets



← **[embedPy]** →



## PyQ

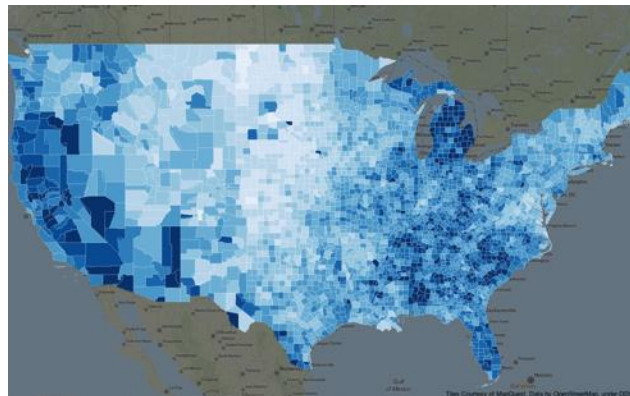
- In PyQ, Python and q objects live in the same memory space and share the same data
- This is achieved by bringing the Python and q interpreters into the same process, so that code written in either of the languages operates on the same data

## embedPy

- The interface allows execution of Python code directly in a q console or from a script. embedPy, embeds python functions/data/modules directly into a q process

## jupyterQ

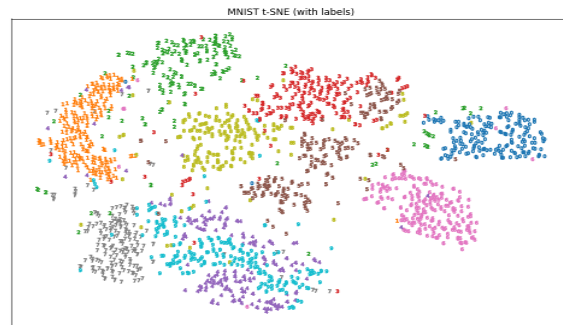
- jupyterQ enables users to utilise the power of Jupyter Notebooks for visual inspection and interpretation of Q and Python. Documents are both human-readable documents containing the analysis description and the results (figures, tables, etc.) as well as executable documents which can be run to perform data analysis



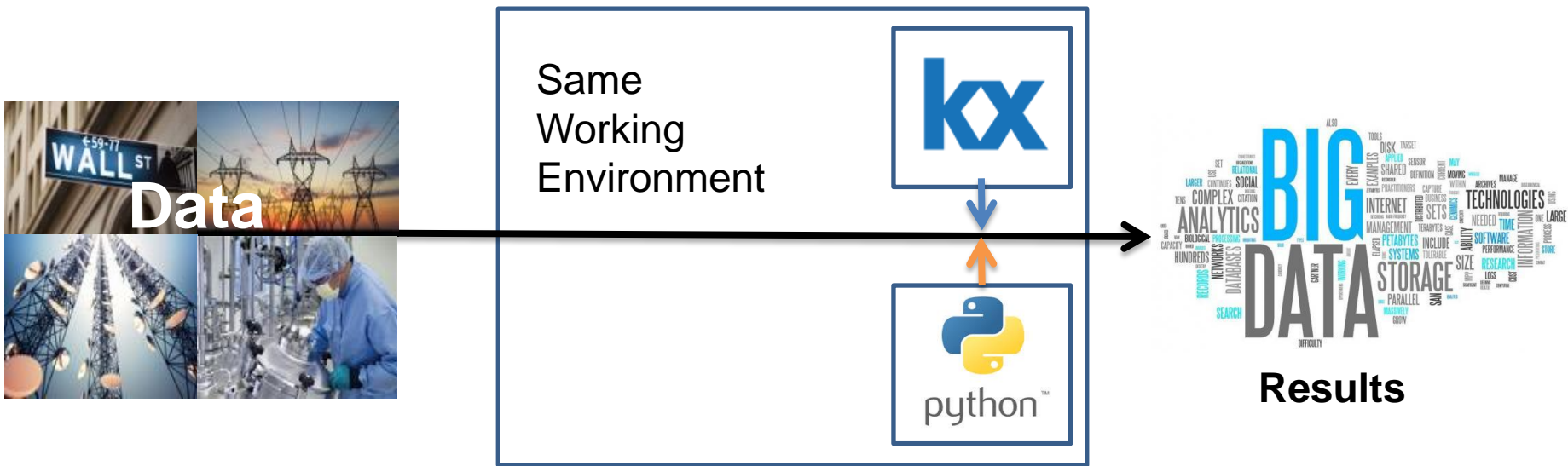
Visualise t-SNE (checking with labels)

```
in [16]: subplots=plt.subplots(1,1)
fig=plt.figure(figsize=(10,10))
fig.set_figwidth(10)
ax=fig.subplots(1,1)
ax.set_title('MNIST t-SNE (with labels)')
ax.scatter('restSNE[1,0].restSNE[1,1]'; s=pykw 25; c=pykw "C",string x; marker=pykw "S",string x)
ax.set_xlabel('restSNE[1,0].restSNE[1,1]')
ax.set_ylabel('restSNE[1,0].restSNE[1,1]')
plt.show()
```

Out[16]:



- Python/q **share the same memory space.**
- Makes q more accessible to others.
- Code can be developed in either python or q.
- Both languages can operate on the same data.



```
>>> q()
```

```
q)trade:([ ]date:();sym:();qty:())
```

```
q)\
```

```
>>> q.insert('trade', (date(2006,10,6), 'IBM', 200))
k('0')
>>> q.insert('trade', (date(2006,10,6), 'MSFT', 100))
k('1')
```

```
>>> q.trade.show()
date      sym  qty
-----
2006.10.06 IBM   200
2006.10.06 MSFT  100
```

Define a function in q:

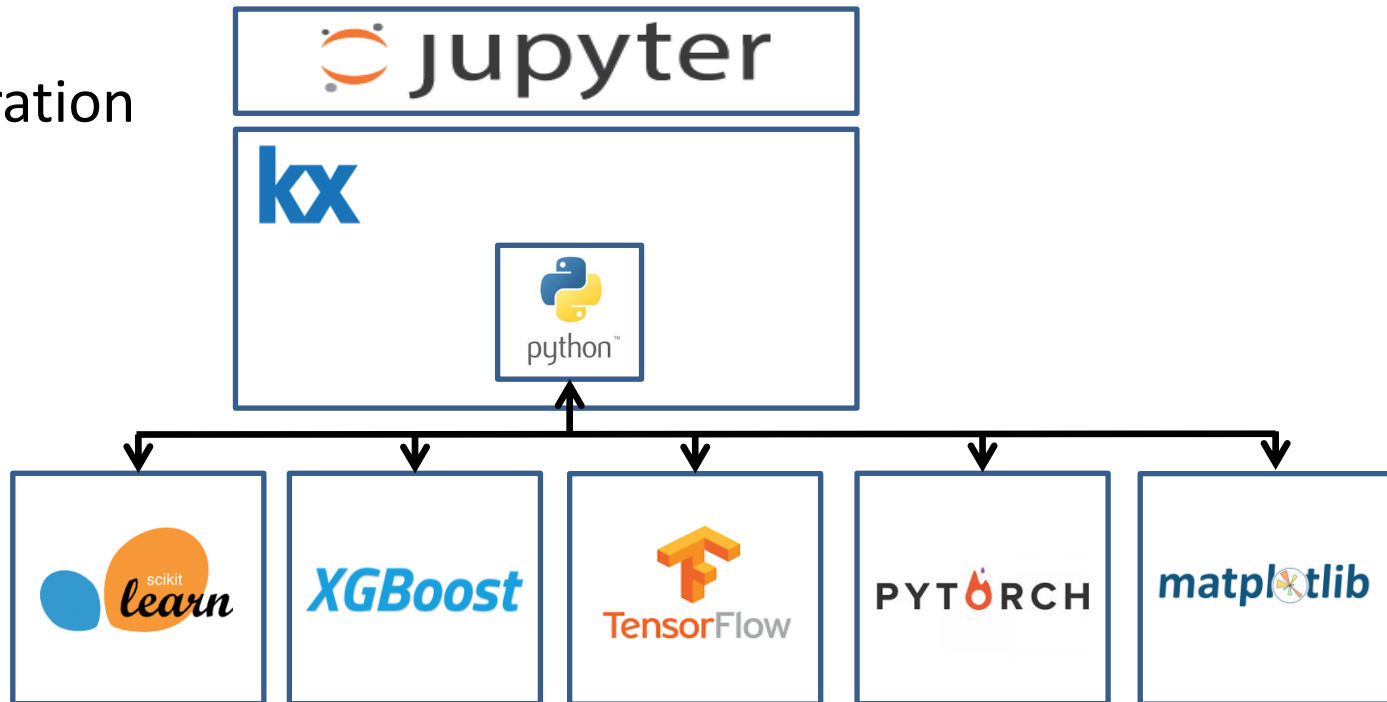
```
q)f:{[s;d]select from trade where sym=s,date=d}
```

Call the `q` function from Python and pretty-print the result:

```
>>> x = q.f('IBM', date(2006,10,6))
>>> x.show()
date      sym  qty
-----
2006.10.06 IBM  200
```



- Python *inside* q
- Access to world of Machine Learning
- Visualizations
- Jupyter integration



**Precision Medicine & Genetic Testing are relatively untapped resources of massive potential – Machine Learning can change this.**

### The Challenge:

- Cancer tumours can have thousands of genetic mutations.
- Distinguishing the mutations that contribute to tumour growth is a real challenge!
- **The mutations are categorized from one to nine by expert oncologists.**
- Very Laborious.
- The ability to automate this process could revolutionize the field of oncology.
- **The application of Kx Technology could change lives.**



## Case Study: embedPy and Machine Learning – The Data



The data is presented as massive text files of clinical records.  
This is very laborious and time consuming work...

ID | 1290i

Text| `We report a 10-year-old girl presenting with severe neonatal hypertrophic cardiomyopathy (HCM), feeding difficulties, mildly abnormal facial features, and progressive skeletal muscle symptoms but with normal cognitive development. Targeted oligonucleotide-selective sequencing of 101 cardiomyopathy genes revealed the genetic diagnosis, and the mutation was verified by Sanger sequencing in the patient and her parents. To offer insights into the potential mechanism of patient mutation, protein structural analysis was performed using the resolved structure of human activated HRAS protein with bound GTP analogue (PDB id 5P21) in Discovery Studio 4.5 (Dassault Systèmes Biovia, San Diego, CA). The patient with hypertrophic cardiomyopathy and normal cognitive development was diagnosed with a rare mutation in the highly conserved amino acid, T58I, which is responsive to GTPase activating protein (GAP). This mutation is likely affects binding of GAP to HRAS, leading to constitutive activation. Patients with other mutations causing Costello syndrome. We report the first case in Europe and the fourth case in the world with this mutation. The patient is currently being treated with p.T58I mutation.

The training and testing database  
have 3321 and 5668 samples  
respectively.

**The power of kdb+ and embedPy can change this!**

## 3. The Algorithm - Random Forest Classifier

It is important to note before diving into the Random Forest classifier that there are a plethora of methods to produce accurate results. Gradient Boosting Machine (GBM), Generalized Linear Model (GLM) & Deep Learning (DL) have all been used to great effect in this challenge thus far. An equally accurate algorithm that is easy to follow is the random forest.

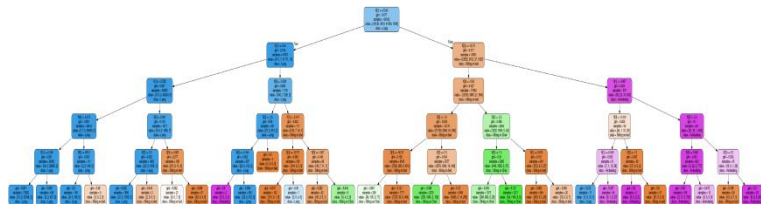
```
[28]: //The line below imports from the python code required to create a pipeline, allowing us to perform multiple tasks sequentially
pipeline:.p.import['sklearn.pipeline';:Pipeline]

//countvectorizer is imported from python again here
countvectorizer:.p.import['sklearn.feature_extraction.text';:CountVectorizer]

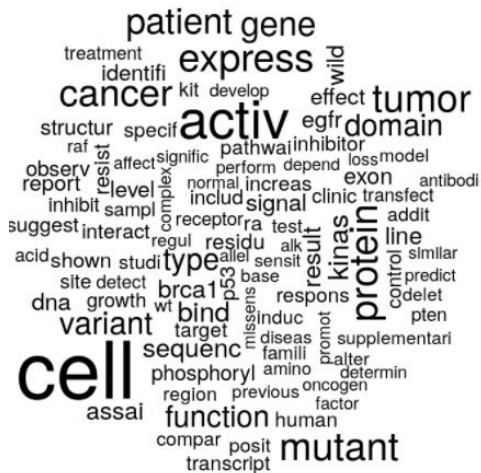
//The Random Forest Classifier code is imported (once again from the SKLearn Toolkit) here
randomforest:.p.import['sklearn.ensemble';:RandomForestClassifier]

//Define rf to be the Random Forest without any parameters yet
rf:randomforest[]

//Define the pipeline
classifier:pipeline[(('vect,countvectorizer[]');('clf,randomforest[]'))]
```



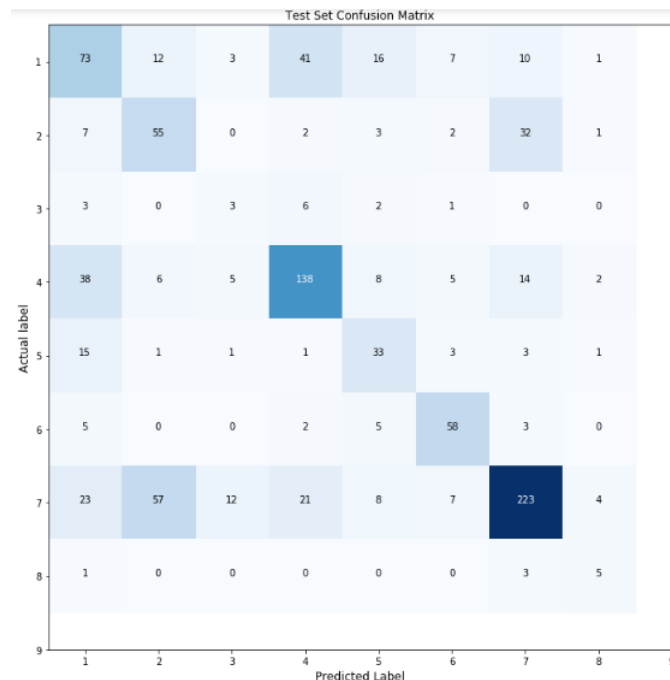
The **text** holds  
the *key* to the  
classification  
problem



Using Kx Technology, a Random Forrest ML algorithm produced:

- Predictions for *every record* in **30 seconds**.
- Had an overall accuracy of 63%

ID	Gene	Variation	Pred	Text
0	ACSL4	R570S	7	2. This mutation resulted in a myeloproliferat..
1	NAGLU	P521L	1	Abstract The Large Tumor Suppressor 1 (LATS1) ..
2	PAH	L333F	7	Vascular endothelial growth factor receptor (V..
3	ING1	A148D	7	Inflammatory myofibroblastic tumor (IMT) is a ..
4	TMEM216	G77A	1	Abstract Retinoblastoma is a pediatric retinal..
5	CD40LG	A123E	4	The accurate determination of perfluoroalkyl s..
6	KLF11	T220M	7	Aberrations in the mTOR (mechanistic target of..
7	SGCB	T151R	4	Oncogenic mutations in the monomeric Casitas B..
8	CLCF1	R197L	2	NPM1 gene at chromosome 5q35 is involved in re..
9	SDHAF1	R55P	7	Introduction The epidermal growth factor recep..
10	SPTLC2	I504F	7	The protein kinase B-RAF is mutated in approxi..
11	SUMF1	A348P	7	Genes of the Raf family encode kinases that ar..
12	TET2	Y1902A	6	TET proteins oxidize 5-methylcytosine (5mC) on..
13	G6PD	D312H	7	The p110δ subunit of phosphoinositide 3-kinas..
14	SNCB	P123H	7	Abstract Purpose: Positron emission tomograph..





it's about time



# Additional Resources







TM

## 1.1 Billion Taxi Rides on kdb+/q & 4 Xeon Phi CPUs

Q is a programming language with a built-in, column-oriented, in-memory and on-disk database called kdb+. Q both includes and extends SQL. Q is native to the database engine so, unlike most databases, there is no shipping of data between the client and the server.

Q and kdb+ are shipped as a single binary that has a small memory footprint and is capable of running in the L2 and L3 caches of modern CPUs making the system very performant. Its tables can be stored on a local disk or distributed but will nonetheless appear as a single table. Table data can be partitioned and segmented into memory mapped files which helps remove I/O bottlenecks.

## DOWNLOAD KDB+

Kx offers a 32-bit version for free so you can try it for yourself and see how powerful and fast it is. This version has all the functionality of the full 64-bit version and is for non-commercial use only. This software may not be used in production systems. Kx does not provide technical support.

32-bit can only be used in a commercial environment for proof of concept applications, development, and in cases where permission is explicitly granted by Kx. Please contact us at [sales@kx.com](mailto:sales@kx.com) if you are a current customer who would like to use 32-bit Kx in a production application.

**No Fee, Non-Commercial Use, 32 Bit Kdb+ Software License Agreement**  
(revised 13 Sep 2015)

CAREFULLY READ THE FOLLOWING TERMS AND CONDITIONS. BY DOWNLOADING THE 32 Bit Kdb+ SOFTWARE, YOU ARE AGREEING TO BE BOUND BY THESE TERMS AND CONDITIONS. IF YOU DO NOT AGREE TO THESE TERMS AND CONDITIONS, DO NOT DOWNLOAD THE SOFTWARE.

This No Fee, Non-Commercial Use, 32 Bit Kdb+ Software License Agreement ("Agreement") is made between Kx Systems, Inc. ("Kx") and you, the User, with

For Installation Instructions [GO HERE](#).

For getting started resources [GO HERE](#).

For control libraries for Raspberry Pi [GO HERE](#).

☒ I agree to abide by the "Download Kx Software License Agreement".

### Select operating system

- ☒ Windows
- ☐ Linux (x86)
- ☐ Linux (ARM / Raspberry Pi)
- ☐ MacOSX

Email address\*

☒ I would like to receive communications from Kx, including those regarding software updates.

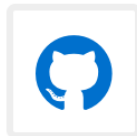
**PROCEED TO DOWNLOAD >**




[Home](#)[code.kx.com](#)[Home](#)[Get started](#)[Reference card](#)[Reference](#) ▾[Tutorials](#) ▾[Interfaces](#)[Interfaces](#) ▾[Cookbook](#) ▾[Developer](#)[GitHub](#)[Q for Mortals 3](#)[Archive](#)[About](#) ▾


## Get going with kdb+


Kdb+, from [Kx Systems](#), is a high-performance historical time-series columnar database, an in-memory compute engine, and a real-time streaming processor – unified with an expressive query and programming language called q.

[Download kdb+](#)[Reference](#)[Get started](#)[Whitepapers](#)[Q for Mortals](#)[FAQ](#)[Share code](#)[Connect](#)




Search for topics



aideen@kx.com


Groups


NEW TOPIC



Mark all as read

Filters





My groups

Home

Starred

▼ Favorites

Click on a group's star icon to add it to your favorites

▼ Recently viewed

Kdb+ Personal De...

Privacy - Terms of Service


Kdb+ Personal Developers

Shared publicly

30 of 1372 topics (99+ unread) ★


Manage · About

A moderated group for q and kdb+ questions from users of the 32bit edition of kdb+. As this is a public forum please do not publish proprietary code. Licensed users of kdb+ should use the k4 listbox.

Function overloading with variable number of arguments (3)


By Yan Yan - 3 posts - 26 views

4:35 AM

2 `p# columns (4)


By Yan Yan - 4 posts - 19 views

Apr 7

new kdb+ book, Nick Psaris' Q Tips (9)


By Simon Garland - 9 posts - 156 views

Apr 6

Why is .Q.dpft making `p#sym the 2nd col? (2)


By Yan Yan - 2 posts - 16 views

Apr 6

running kdb from c compiling issues (8)


By Varun Tomar - 8 posts - 29 views

Apr 4

reading output of `ls` (3)

By effhiae - 3 posts - 16 views

Apr 4

it's about time

INTERFACES   APPLICATIONS   DEVELOPMENT TOOLS   EXAMPLES   FRAMEWORKS   MATHEMATICS   UTILITIES   WEB

## Kx REPOSITORIES ON GITHUB

Kx Technology is an integrated platform – kdb+ – which includes a high performance historical time-series column-store database, an in-memory compute engine, and a real-time event processor all with a unifying expressive query and programming language. Designed from the start for extreme scale, and running on industry standard servers, Kx Technology has been proven to solve complex problems faster than any of its competitors.

### FEATURED

#### hypertree

Recursive aggregating treetable and 3-d pivot table for hypergrid. (Stevan Apter)

#### Kdb-stuff

ServerChecker: how to execute commands on a remote box via ssh from within a Kdb process + parse linux system info (cpuinfo/meminfo/df). (Mohammad Noor)

#### kdb-wc


Kdb+ web components. (Andrey Kozyrev)


#### studio


A rapid-development environment for kdb+. (Charles Skelton)


Please contact [librarian@kx.com](mailto:librarian@kx.com) with additions or changes.


### Contributors


Aaron Davies  USA



Alex B  Canada


Alex Whitney  UK


Andrej Grinev 


Andrew King  UK


Andrey Kozyrev (5) 


Andrey Zholos (2)   Ukraine


Anton Dovzhenko  Singapore


AquaQ Analytics (6)  UK


Benjamin Conlan (2) 


Brandon Yang 


brogar (2) 


Charles Skelton (4)  Switzerland


chuchunt 


Daniel Nugent (8) 


Darren Armstrong  UK


dbyu 


DEVnet (7)  Germany


Dmitry Marienko  Ukraine


Enlightenment Research  USA


Etay Schnapp  USA

Eugene Kononov (2)  USA

fadefy 

Felix Lungu (2)  Romania

First Derivatives plc  UK

Freddie Wu  Singapore

...

### INTERFACES

Adobe Flex	kdb	Alex B
...	...	...

### PLUGINS TO EDITORS

Atom	language-kdb-q	Andrey Kozyrev
...	...	...



Find out what's happening in Kx Meetup groups around the world and start meeting up with the ones near you.

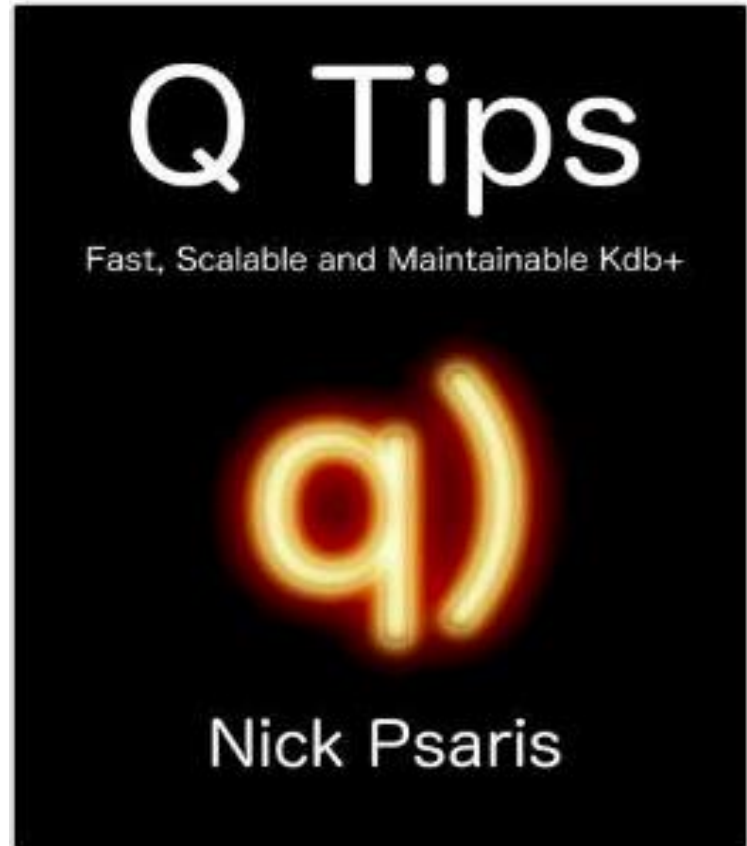
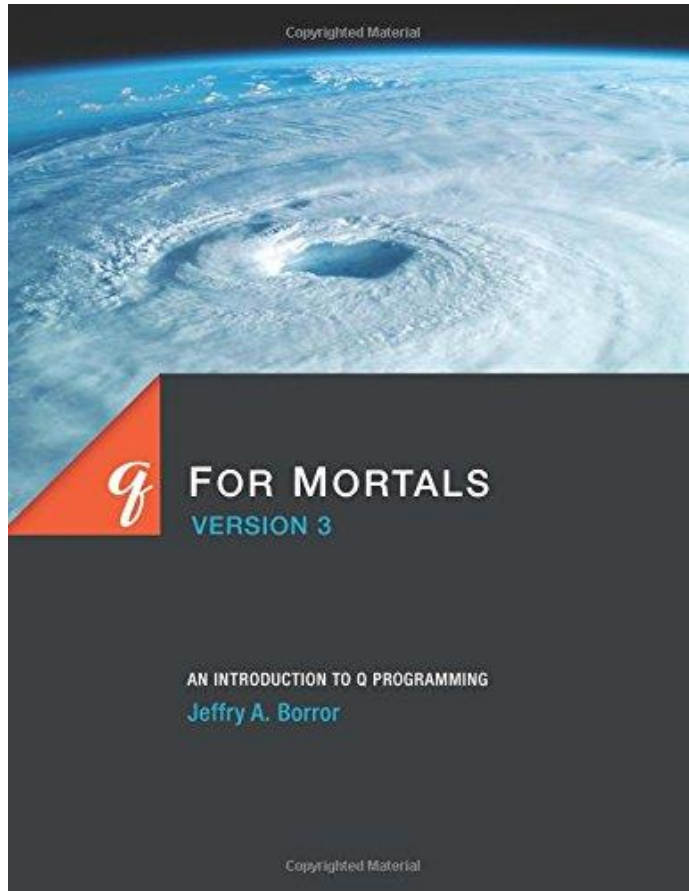
**12,833**  
members

**131**  
Meetups

[Join Kx Meetups](#)

**Related topics:** [kdb+](#) · [Big Data](#) · [Data Visualization](#) · [Big Data Analytics](#) · [Time Series Database](#) · [Data Analytics](#) · [Internet of Things](#) · [Predictive Analytics](#) · [NoSQL](#) · [Data Science](#)







# Kx TECHNOLOGY FUND



Kx Tech  
Access



£10 Mil  
Funding



Int. Office  
Access



Sales &  
Marketing  
Support



Logistical  
Support



Learn more at [www.kx.com/kx-technology-fund/](http://www.kx.com/kx-technology-fund/)

- Opportunities to travel and live internationally
- Paid for city accommodation
- Competitive salary and expenses package
- Health and travel insurance programme
- Generous retirement savings programme

**Interested in joining us or know someone who might be?**

**Talk to me at the end or visit**

[www.firstderivatives.com/careers](http://www.firstderivatives.com/careers)

# Thank You

---



**Jonathan Guy**

Email: [jguy@kx.com](mailto:jguy@kx.com)

<https://kx.com/>

<https://kx.com/download/>