Conclusion

Michael Enudi

Journey through the world of databases and data engineering

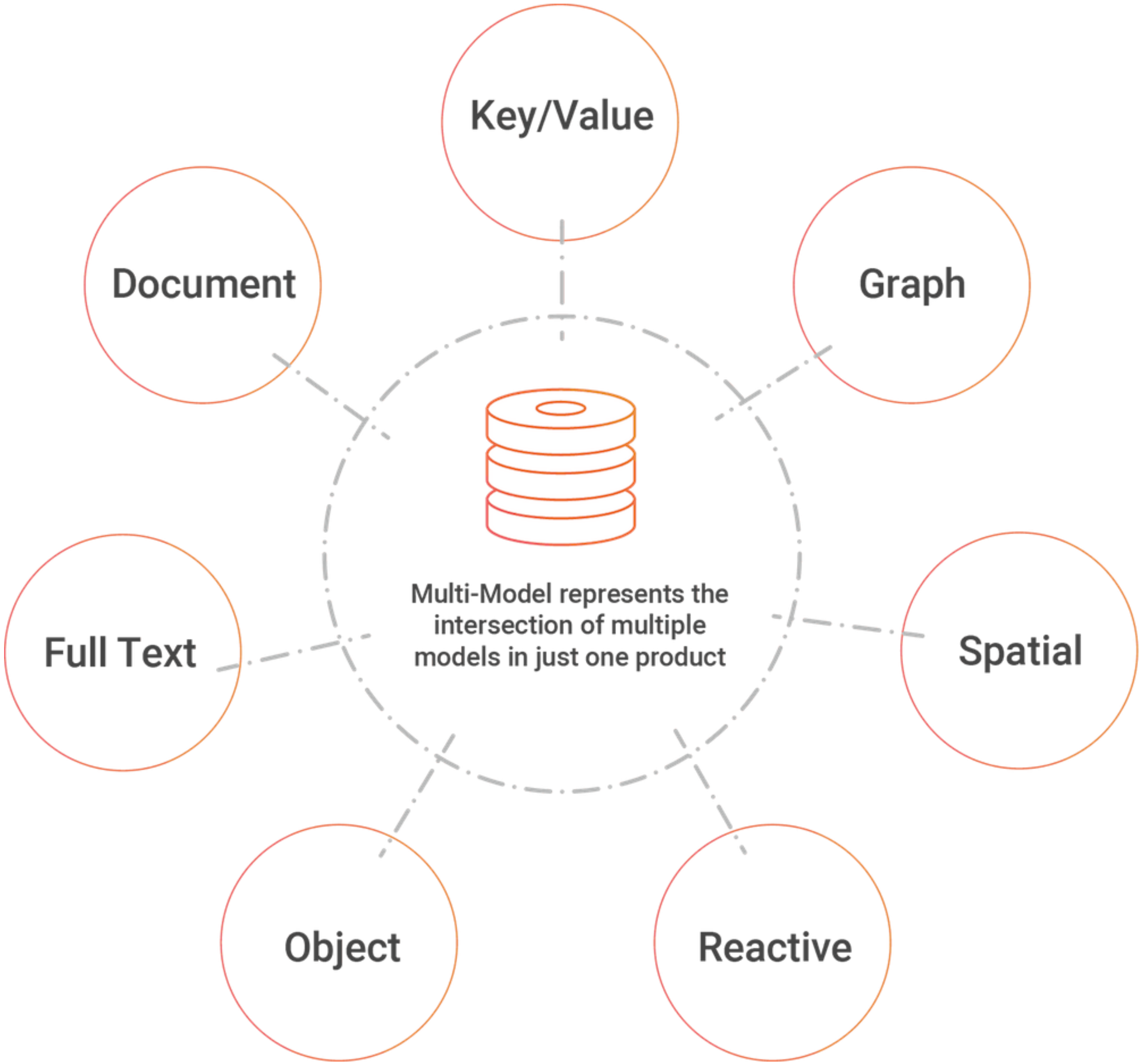
Database Polyglot

Ability to right choose and use from different classes of database or data storage for different need in an organization

Database Skill Projection

Multi-model Database

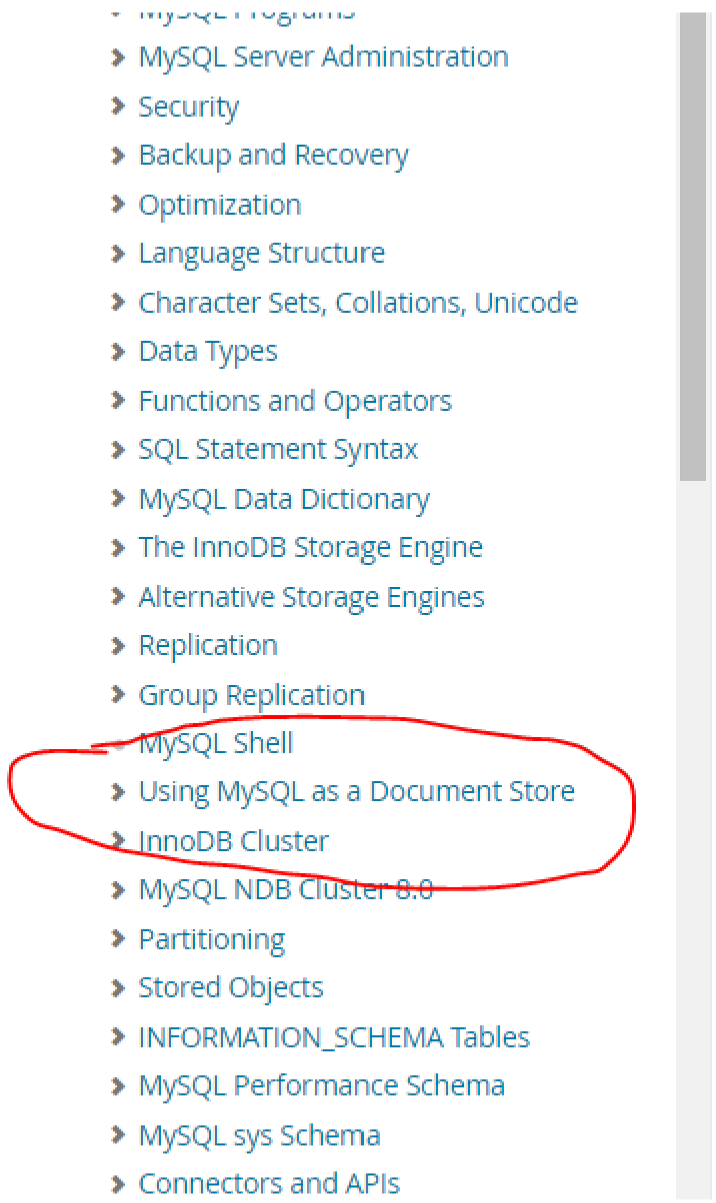
- ✓ MySQL
- ✓ MariaDB
- ✓ Redis
- ✓ PostgreSQL
- ✓ OrientDB
- ✓ AWS DynamoDB
- ✓ Microsoft CosmoDB
- ✓ SAP HANA
- ✓ Datastax Enterprise (on Cassandra)



Multi-Model represents the intersection of multiple models in just one product

Multi-model

- ✓ MySQL 8.0+
- ✓ MariaDB
- ✓ Redis
- ✓ PostgreSQL
- ✓ OrientDB
- ✓ AWS DynamoDB
- ✓ Microsoft CosmoDB
- ✓ SAP HANA
- ✓ Datastax Enterprise (on Cassandra)

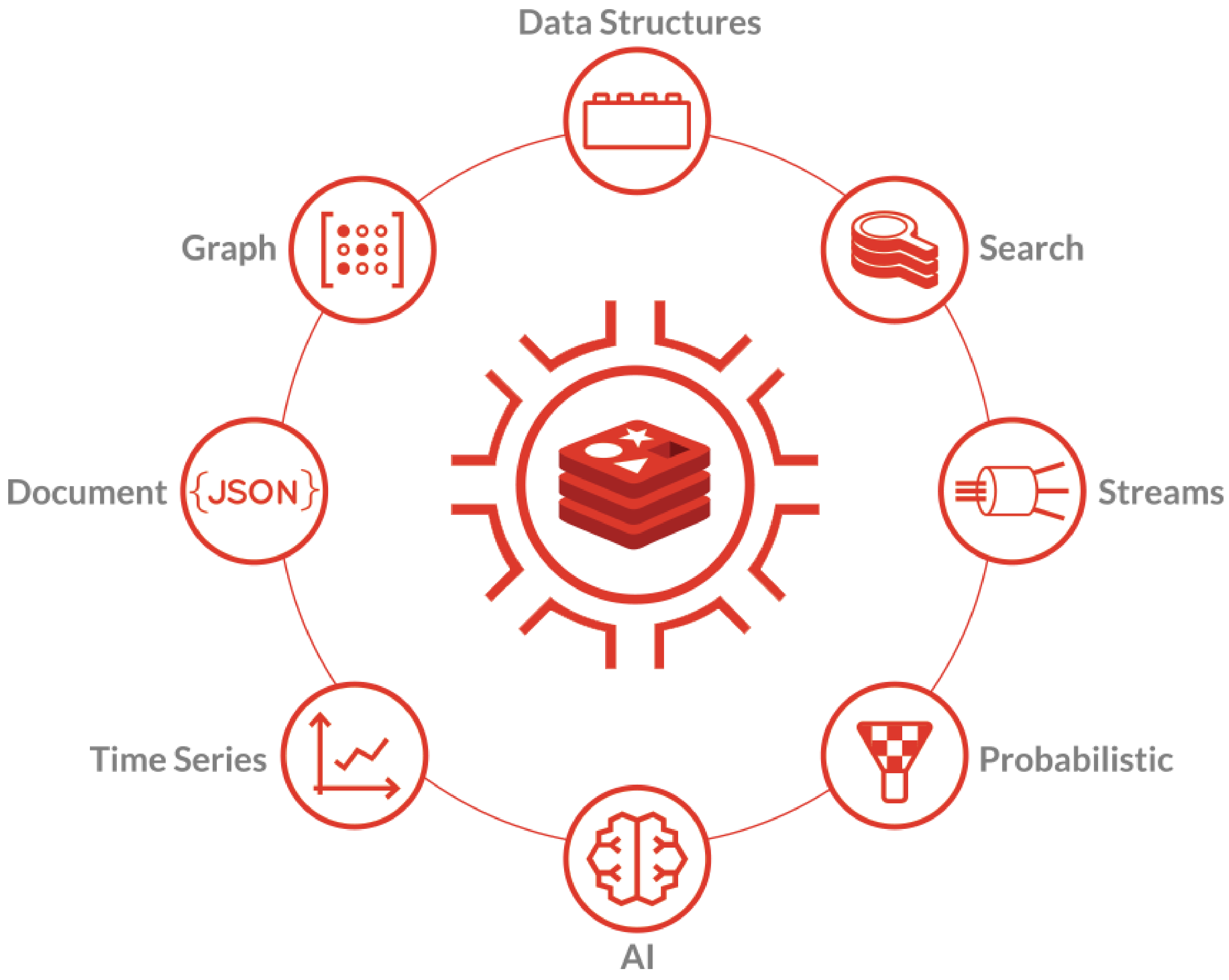


- › MySQL Programs
- › MySQL Server Administration
- › Security
- › Backup and Recovery
- › Optimization
- › Language Structure
- › Character Sets, Collations, Unicode
- › Data Types
- › Functions and Operators
- › SQL Statement Syntax
- › MySQL Data Dictionary
- › The InnoDB Storage Engine
- › Alternative Storage Engines
- › Replication
- › Group Replication
- › MySQL Shell
- › Using MySQL as a Document Store
- › InnoDB Cluster
- › MySQL NDB Cluster 8.0
- › Partitioning
- › Stored Objects
- › INFORMATION_SCHEMA Tables
- › MySQL Performance Schema
- › MySQL sys Schema
- › Connectors and APIs

https://dev.mysql.com/doc/refman/8.0/en/document-store.html

Multi-model

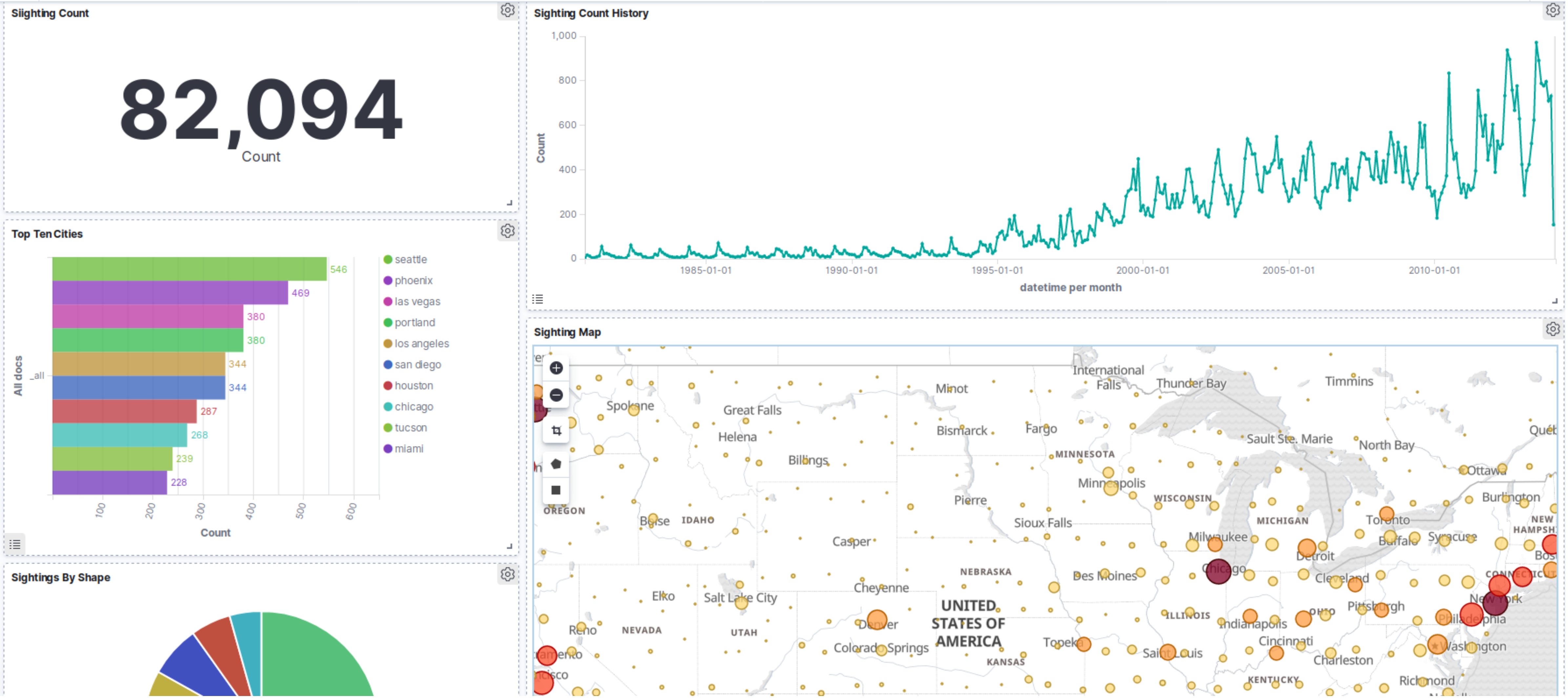
- ✓ MySQL
- ✓ MariaDB
- ✓ Redis
- ✓ PostgreSQL
- ✓ OrientDB
- ✓ AWS DynamoDB
- ✓ Microsoft CosmoDB
- ✓ SAP HANA
- ✓ Datastax Enterprise (on Cassandra)



A New Database

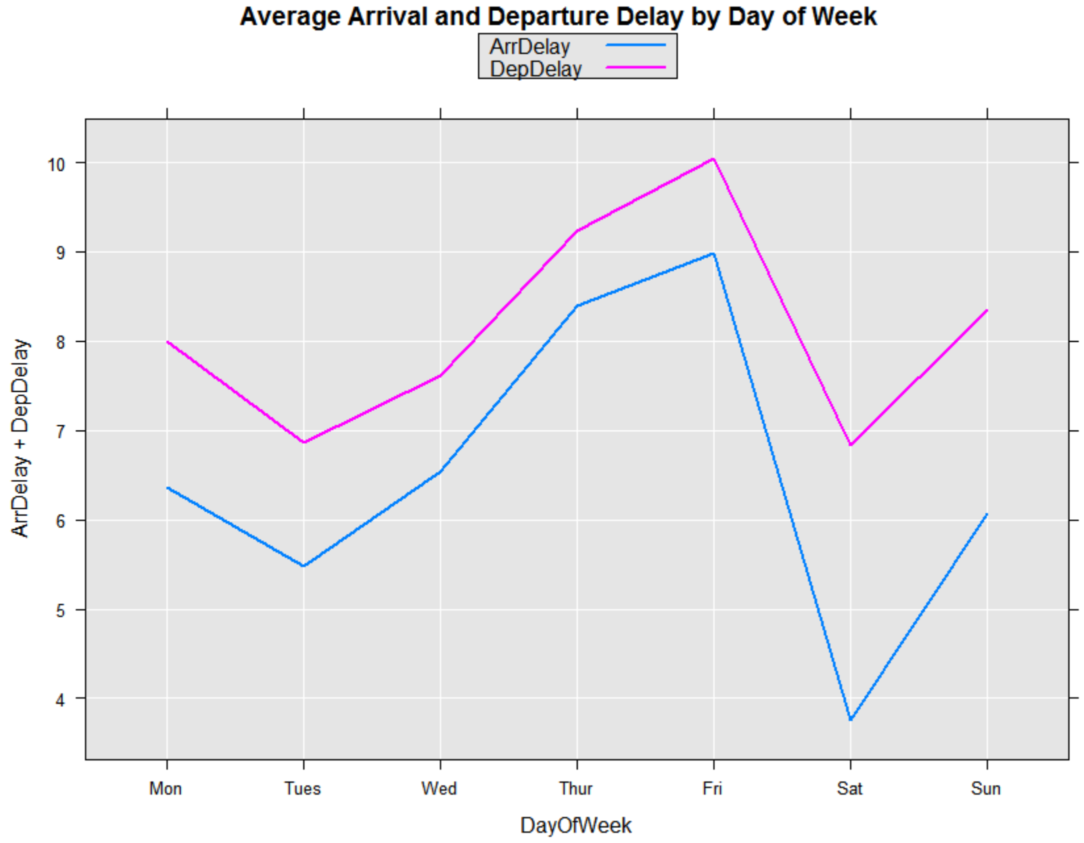
Data Visualization

Visualization Examples

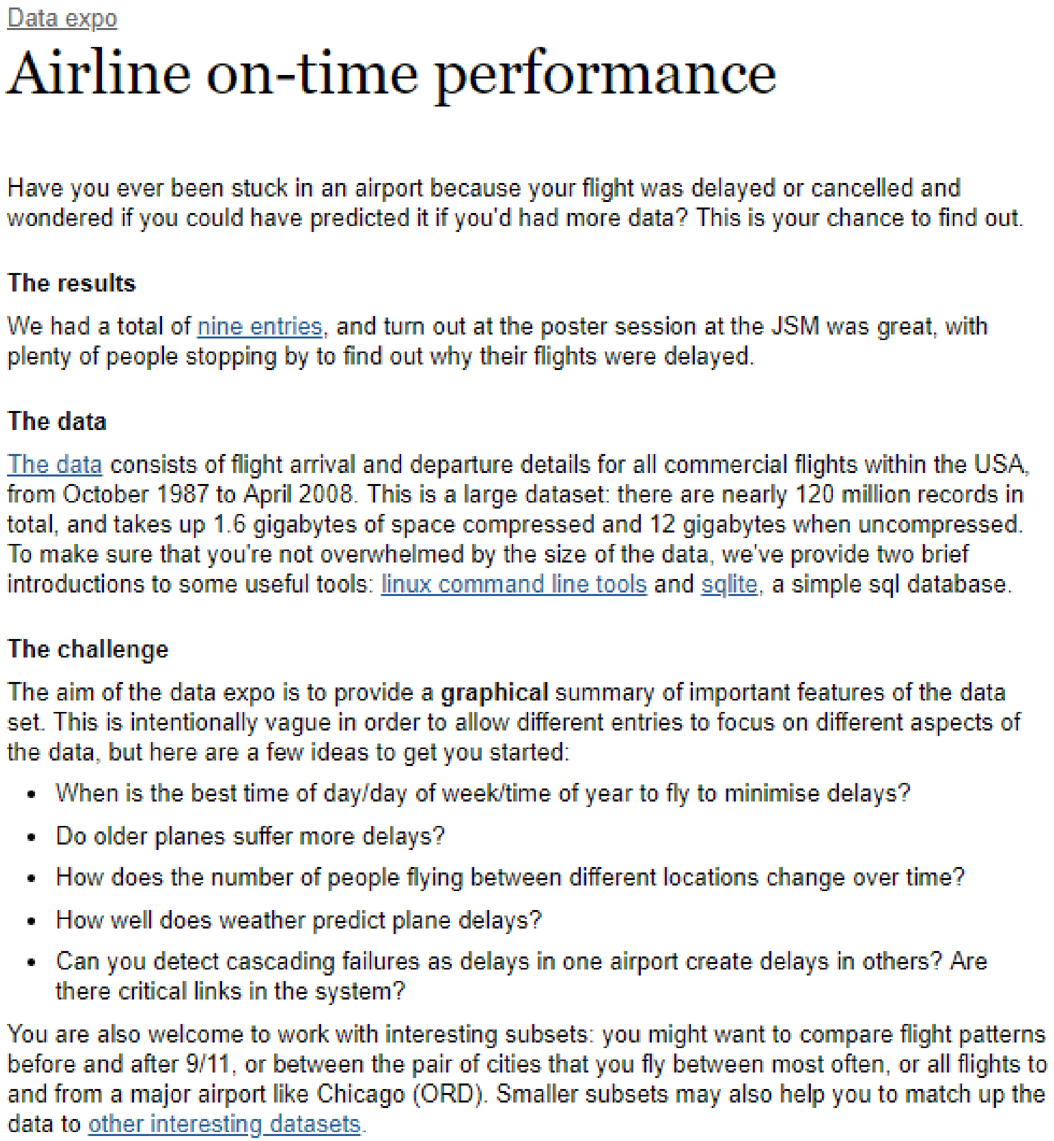


UFO sighting dashboard

Visualization Examples



Identifying the days of the week with most flight delays using the airline on-time performance dataset.



Data expo

Airline on-time performance

Have you ever been stuck in an airport because your flight was delayed or cancelled and wondered if you could have predicted it if you'd had more data? This is your chance to find out.

The results

We had a total of nine entries, and turn out at the poster session at the JSM was great, with plenty of people stopping by to find out why their flights were delayed.

The data

The data consists of flight arrival and departure details for all commercial flights within the USA, from October 1987 to April 2008. This is a large dataset: there are nearly 120 million records in total, and takes up 1.6 gigabytes of space compressed and 12 gigabytes when uncompressed. To make sure that you're not overwhelmed by the size of the data, we've provide two brief introductions to some useful tools: linux command line tools and sqlite, a simple sql database.

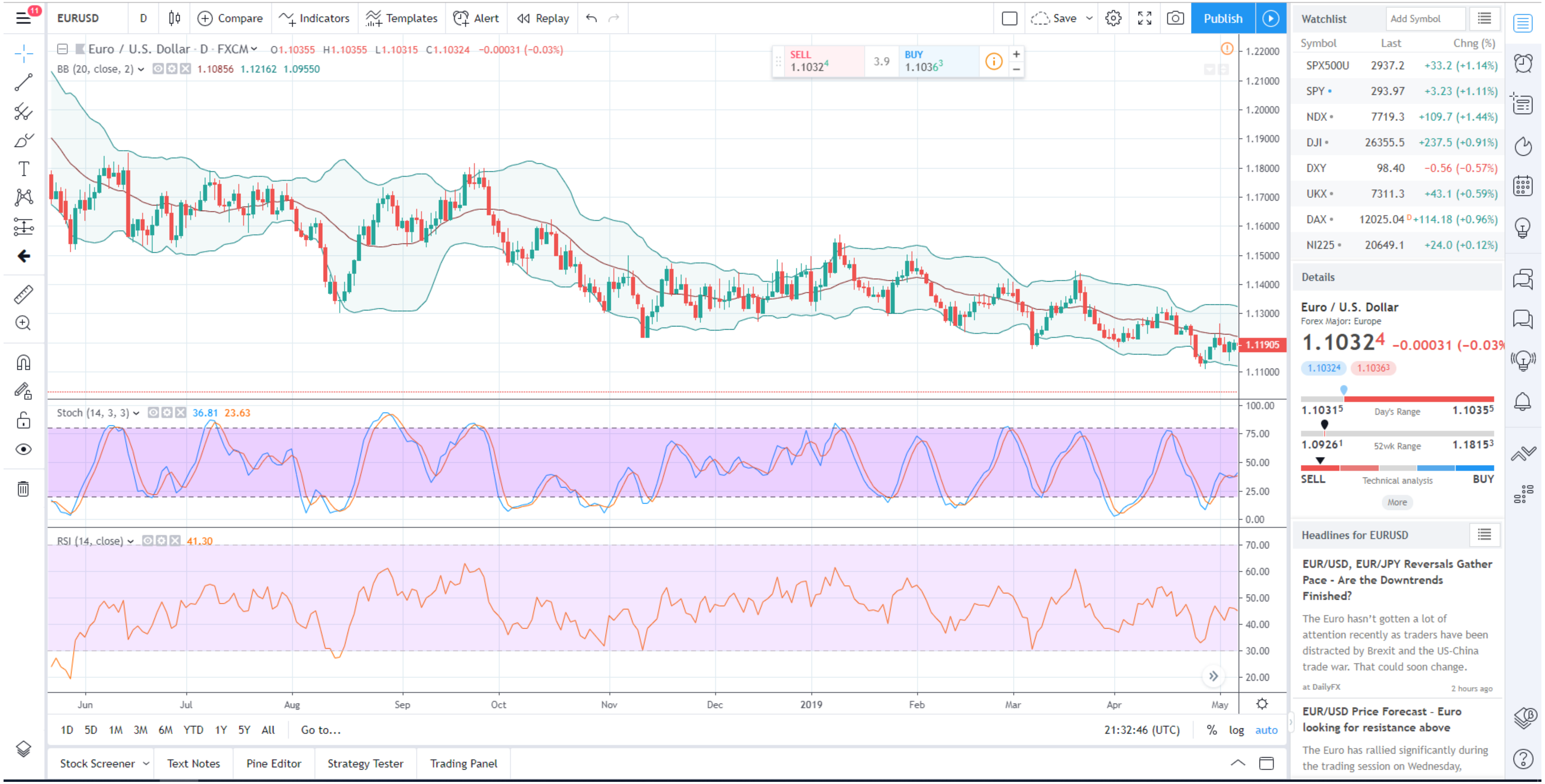
The challenge

The aim of the data expo is to provide a **graphical** summary of important features of the data set. This is intentionally vague in order to allow different entries to focus on different aspects of the data, but here are a few ideas to get you started:

- When is the best time of day/day of week/time of year to fly to minimise delays?
- Do older planes suffer more delays?
- How does the number of people flying between different locations change over time?
- How well does weather predict plane delays?
- Can you detect cascading failures as delays in one airport create delays in others? Are there critical links in the system?

You are also welcome to work with interesting subsets: you might want to compare flight patterns before and after 9/11, or between the pair of cities that you fly between most often, or all flights to and from a major airport like Chicago (ORD). Smaller subsets may also help you to match up the data to other interesting datasets.

Visualization Examples



Symbol	Last	Chng (%)
SPX500U	2937.2	+33.2 (+1.14%)
SPY •	293.97	+3.23 (+1.11%)
NDX •	7719.3	+109.7 (+1.44%)
DJI •	26355.5	+237.5 (+0.91%)
DXY	98.40	−0.56 (−0.57%)
UKX •	7311.3	+43.1 (+0.59%)
DAX •	12025.04 D	+114.18 (+0.96%)
NI225 •	20649.1	+24.0 (+0.12%)

Euro / U.S. Dollar
Forex Major: Europe
1.1032⁴ −0.00031 (−0.03%
1.10324 1.10363

1.10315	Day's Range	1.10355
1.09261	52wk Range	1.18153
SELL	Technical analysis	BUY

Headlines for EURUSD

EUR/USD, EUR/JPY Reversals Gather Pace - Are the Downtrends Finished?

The Euro hasn't gotten a lot of attention recently as traders have been distracted by Brexit and the US-China trade war. That could soon change.

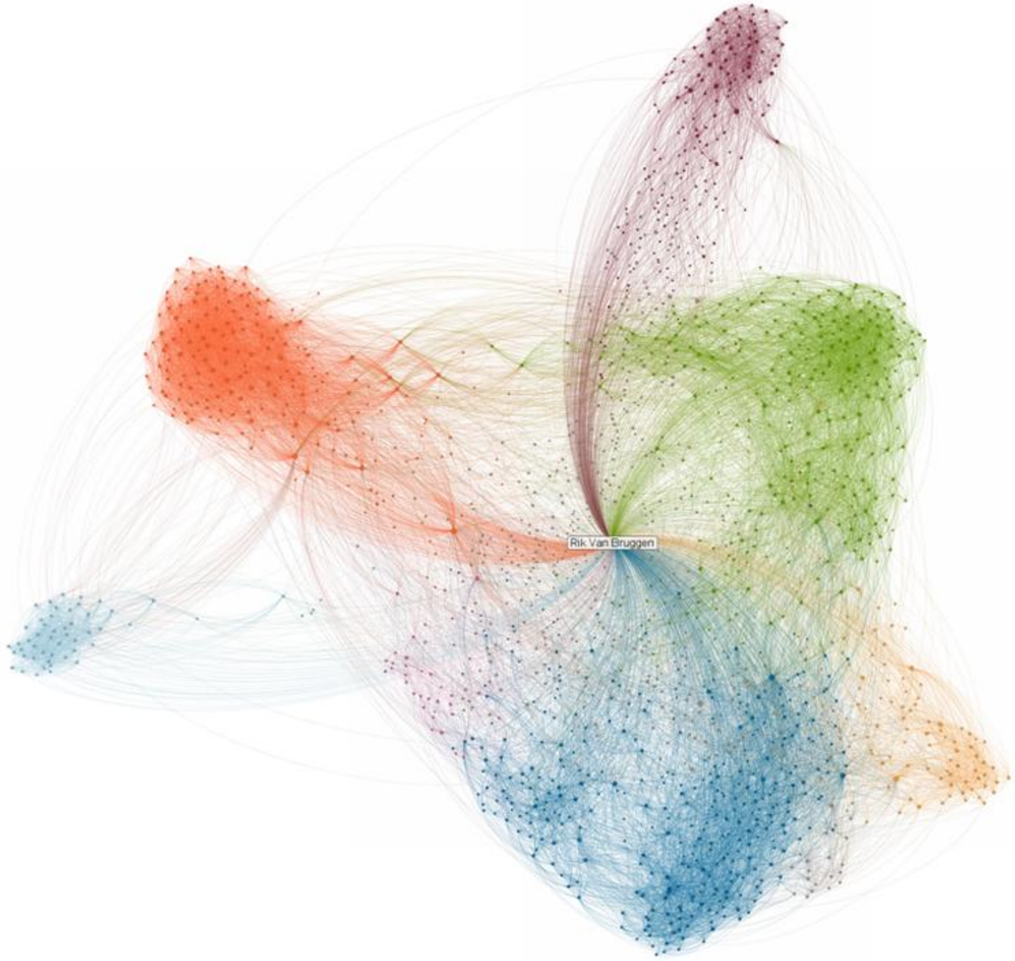
at DailyFX 2 hours ago

EUR/USD Price Forecast - Euro looking for resistance above

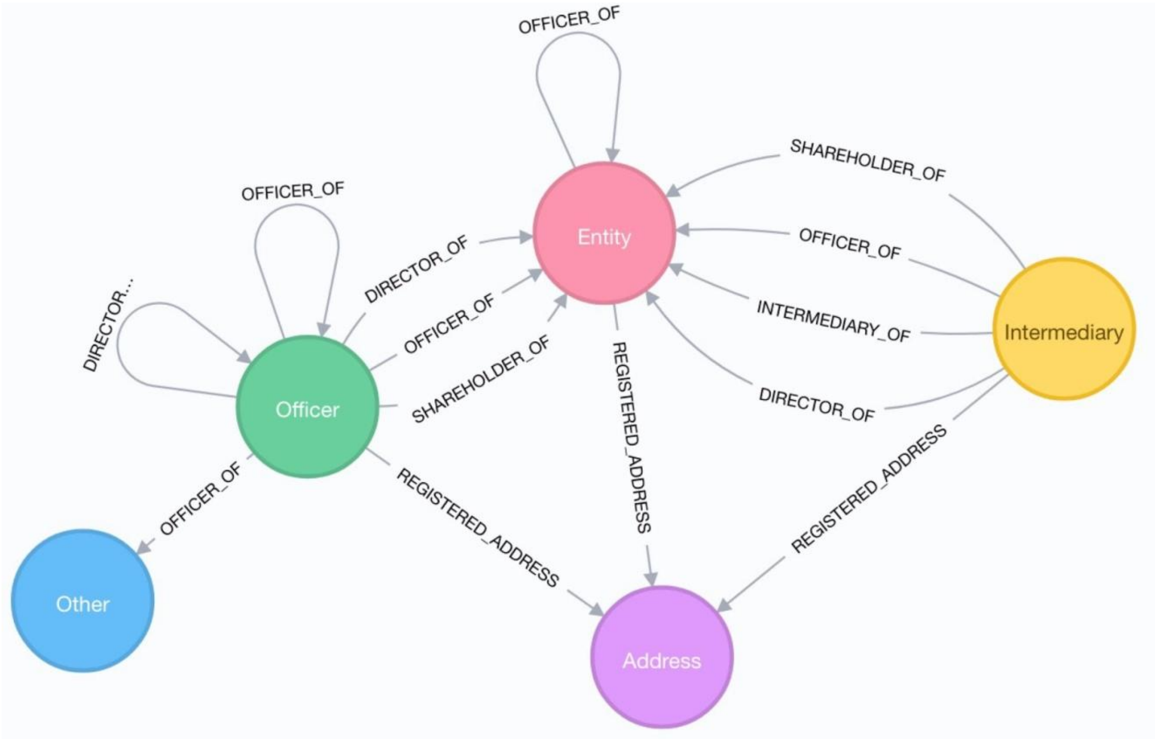
The Euro has rallied significantly during the trading session on Wednesday,

Identifying trading opportunity by plotting visual objects in a price chart.

Visualization Examples



Observing connections of people and businesses from the leaked panama papers



Panama paper graph schema

Data Visualization

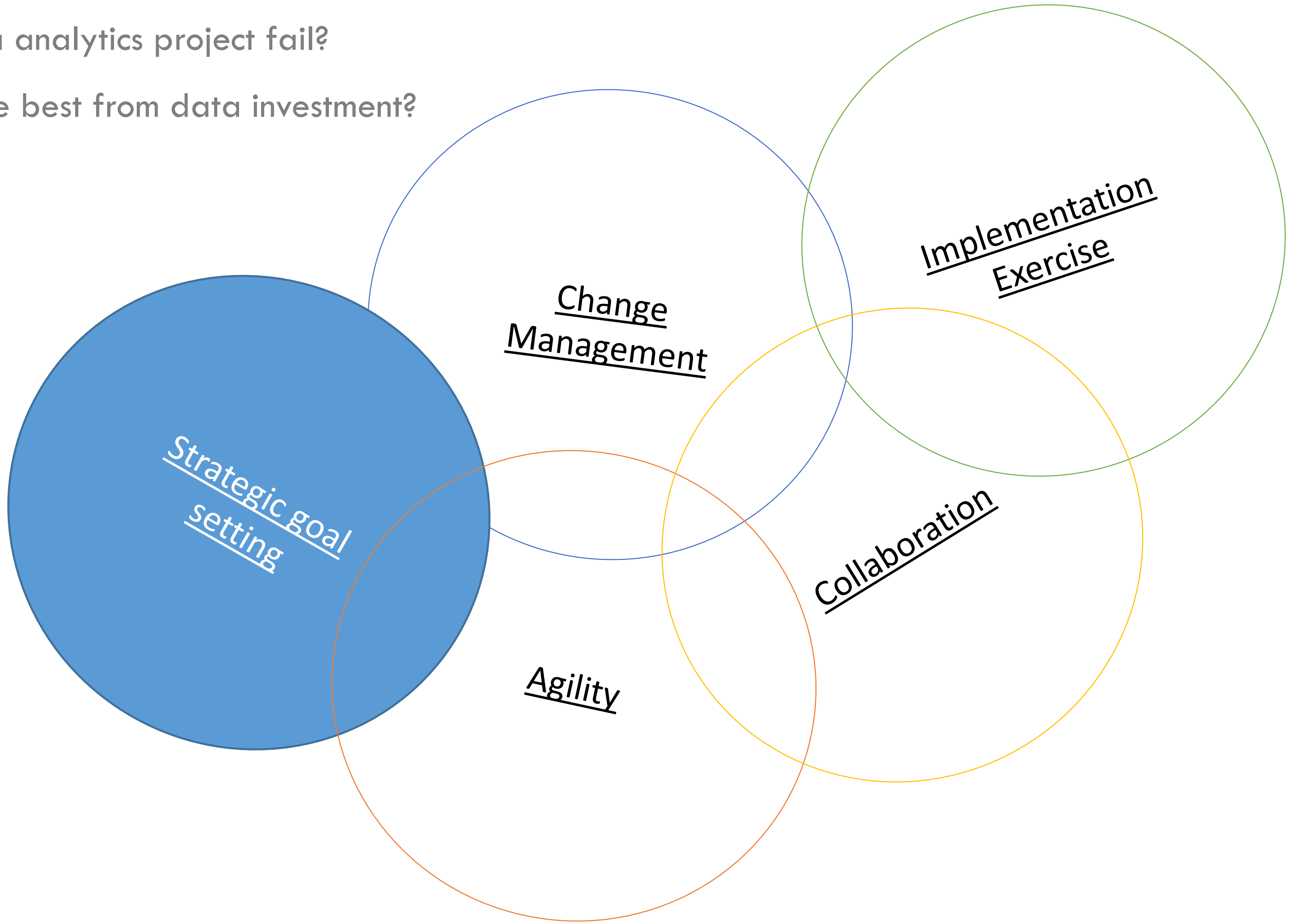
- Explanation
- Exploration

Data management in the Cloud

Building a Data-driven Organization - Conclusion

Becoming Data-Driven

- Why do majority of big data analytics project fail?
- how can organizations get the best from data investment?



Strategic Goal Setting

Data engineering is a tool. Those who intend to use it, must be clear what they want to do with it.

- ✓ Clear business goal
- ✓ Disseminate low-level goals to each department or unit.
- ✓ Create indicators to gauge the success of each strategy implementation.

Implementation Expertise

When you know what to do and how to do it, the journey completes before its started

- ❑ Ability to start small and scale along.
- ❑ Ability to implement a working data-quality infrastructure.
- ❑ Able to enable every other units with data infrastructure and the necessary training for them to use it.
- ❑ Those implementing data science - machine learning, artificial learning and those awesome stuff must be a critical part of a rich communication chain.

Change Management

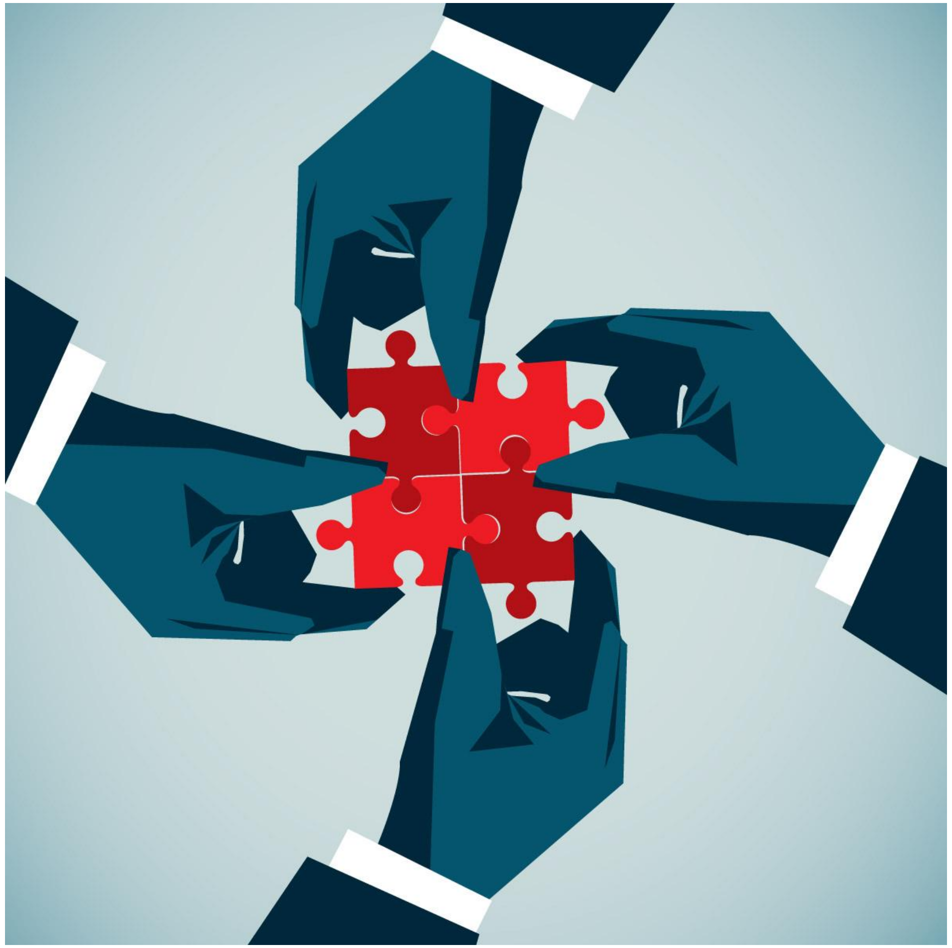
We could change the way we perceive the tool for it to bring different results.

- ❑ How to use data
- ❑ How to collaborate
- ❑ How to integrate
- ❑ How to communicate
- ❑ How to ask questions

Agility

We could change the way we perceive the tool for it to bring different results.

- ✓ Start small and grow
- ✓ Failing fast
- ✓ Integrate



Collaboration

"It is literally true that we can succeed best and quickest by helping others to succeed." – Napolean Hill

THANK YOU ☺

THANK YOU ☺

Michael Enudi

- Lives and works in Johannesburg, South Africa
- Senior Big Data Engineer with over 14 years of working experience writing enterprise java applications, architecting data solutions.
- Over 3 years of training in fields of databases, NoSQL and Hadoop tech.
- EMC Data Science Certified Associate
- Cloudera Certified Spark and Hadoop Dev.
- Oracle Certified SQL Expert
- Oracle Certified Java Master
- Sun Certified Java Business Component Dev.
- Sun Certified Java Programmer
- Big data enthusiast