TERADATA

Duncan Irving

Teradata (Oil & Gas Consulting Practice EMEA/APAC)

Data Science in E&P: “and”, not “or”

Dr Duncan Irving

Getting started with Data Science

Agenda

- Data Management: New v. Old
- Where Data Science fits in
- Data Science in Upstream:

  6 case studies
- What we have learnt from our recent data science projects in terms of data management

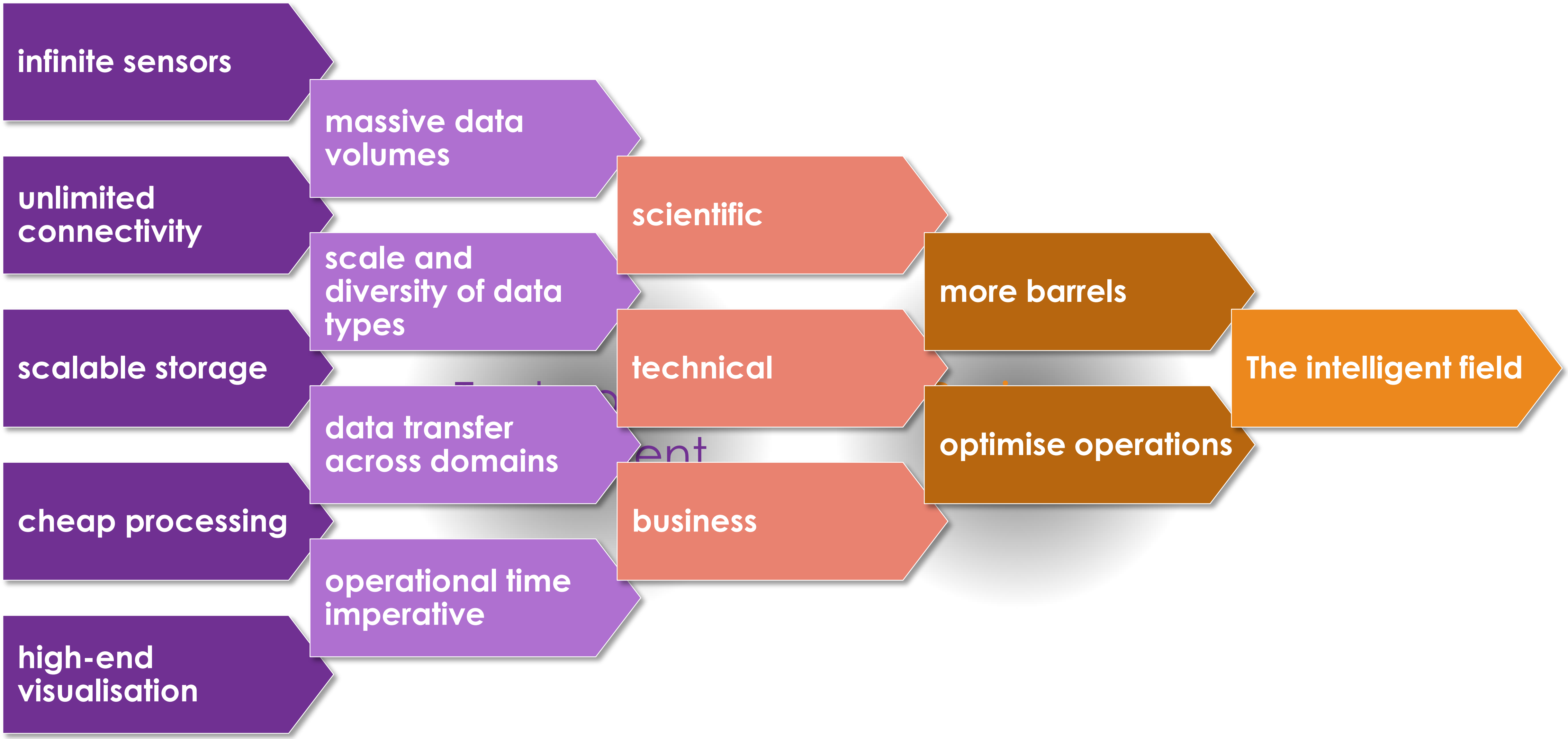
How we understand and interact with each other

How organisations understand and interact with us

How we interact with technology and services

How we exploit knowledge…
at scale and pace

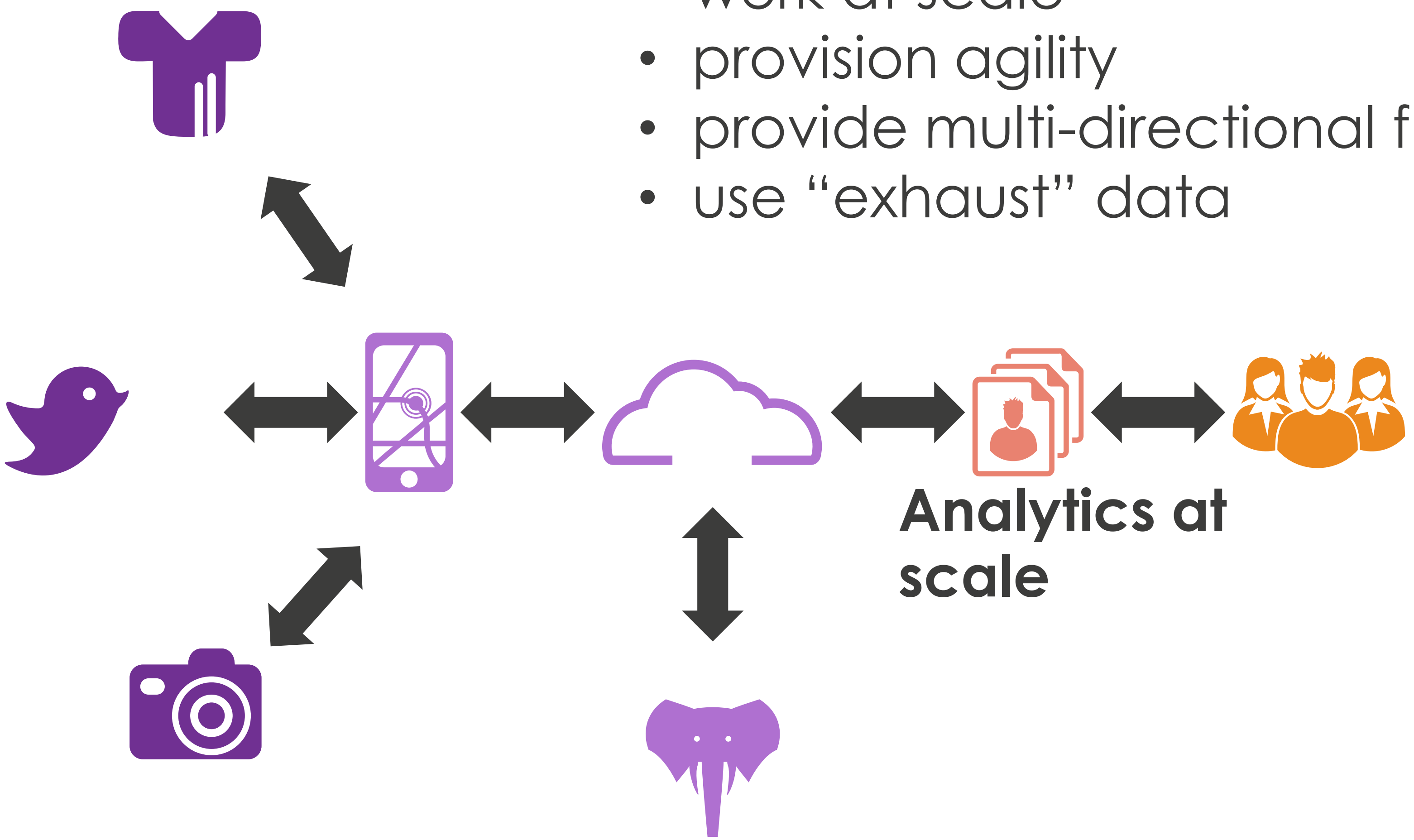
Disruption is upon us – how do we exploit all the new data?



Data-driven analytical architectures

Designed to:

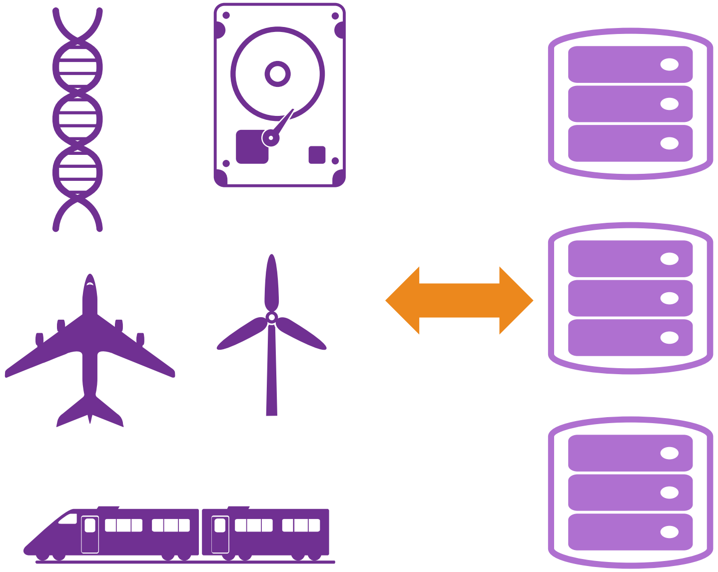
- work at scale
- provision agility
- provide multi-directional flow
- use “exhaust” data

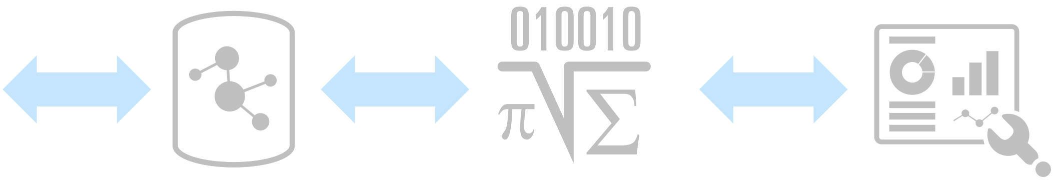


Heavy industry analytical architectures

In E&P:

- *Raw data is too siloed*
- *Sensors are for operational control, not business value*
- *Acquisition is deliberate and often costly*



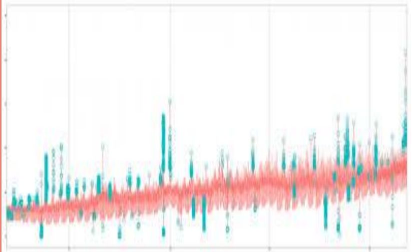
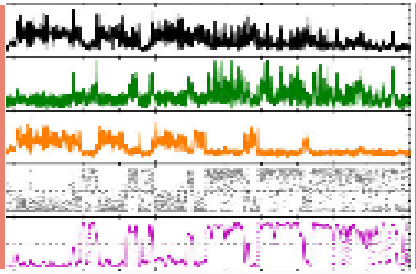
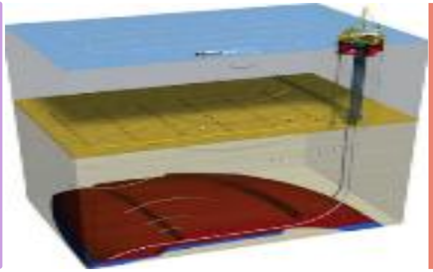
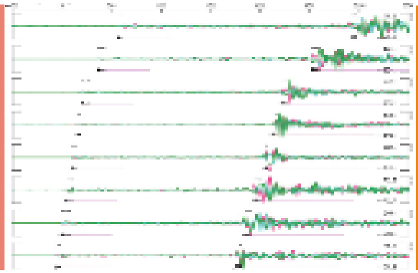


No Analytics at scale

But…

Our workflows haven’t really changed much since the first data started coming back to shore with the oil…

“New data” comes in three flavours

It comes from		It can contain…		It has impact
Fleets: from lots and lots of similar things		**Outliers:** Which of my things are behaving differently?		“Fleet-wide” 24/7 for holistic management
Systems: across the same big “thing”		**Emergent behaviour:** Is my system changing to a new state?		High-level KPIs at business units and facilities level
Collectors: “big models” or monitoring		**Events:** are there hidden signals?		Performed at sub-second level and data kept for decades

…but that looks a lot like the old data!

Yes, but the KPIs are different

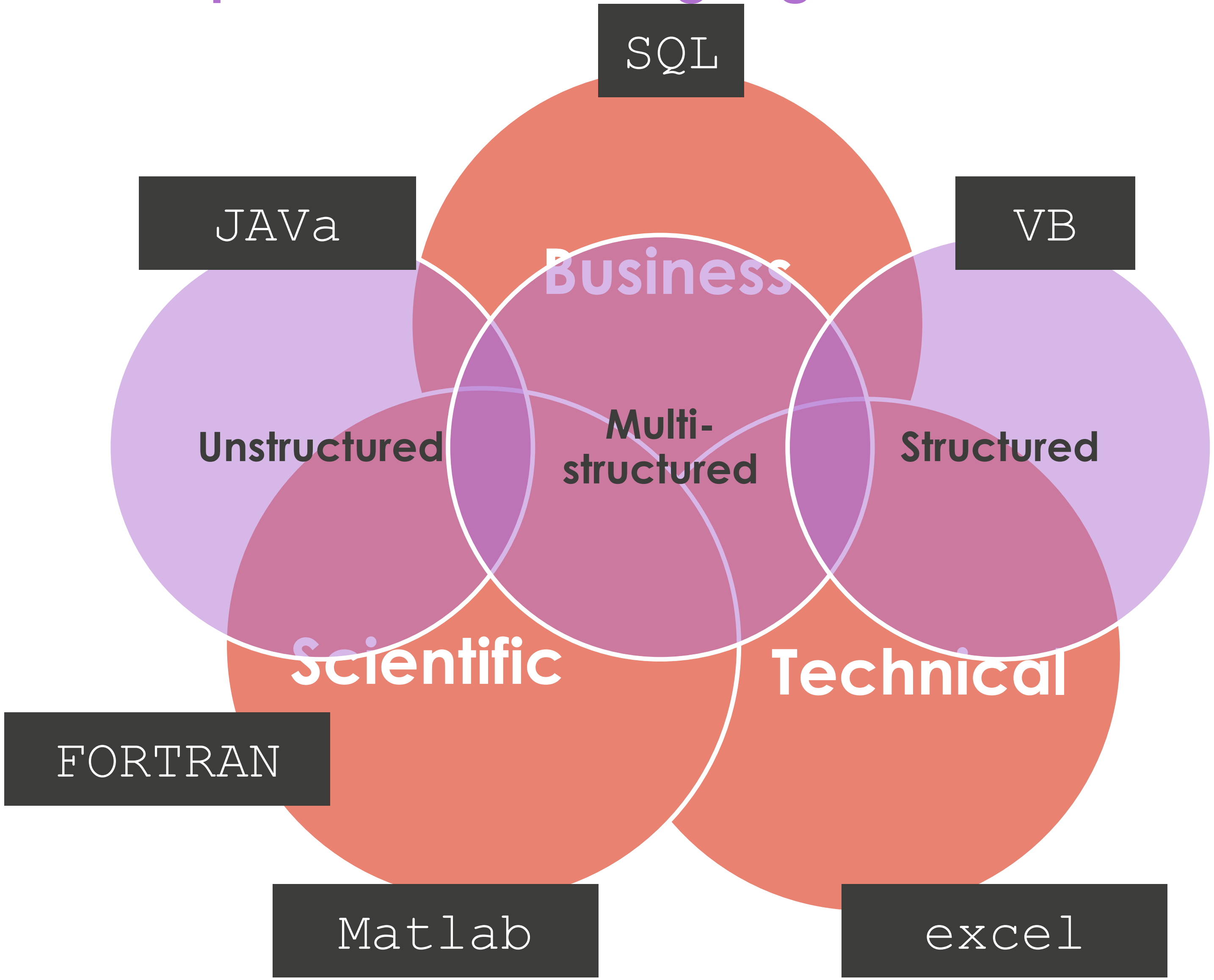
- Business related
- Business budgets, not IT (Low Capex / spend from Opex)
- Show business value – early, and continuously

Our data managers are **highly skilled “librarians”**

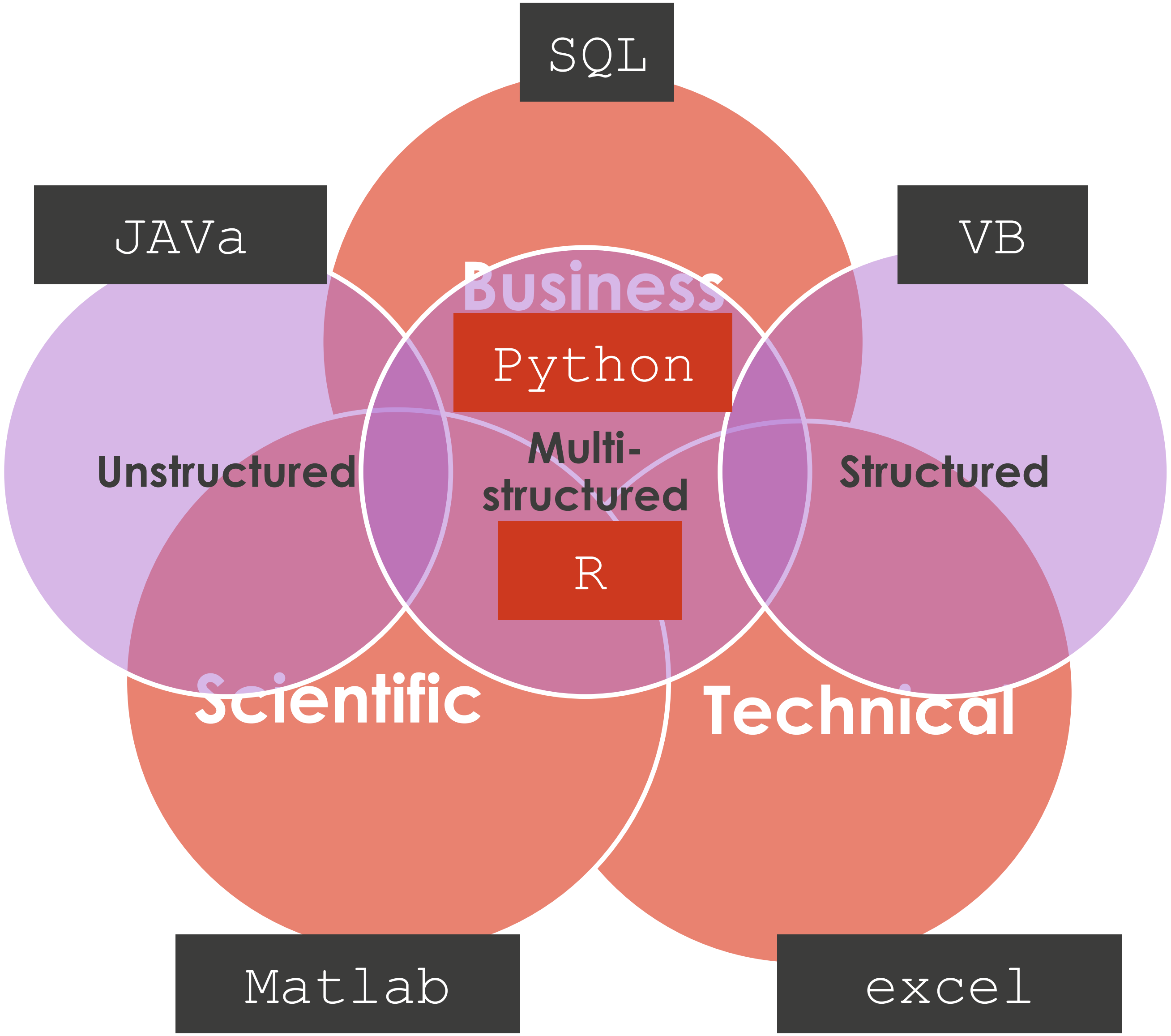
- curate measurement data
- Ad hoc management of interp
- “work to spec”

…but want to deploy their domain expertise much more!

Our different tribes speak different languages

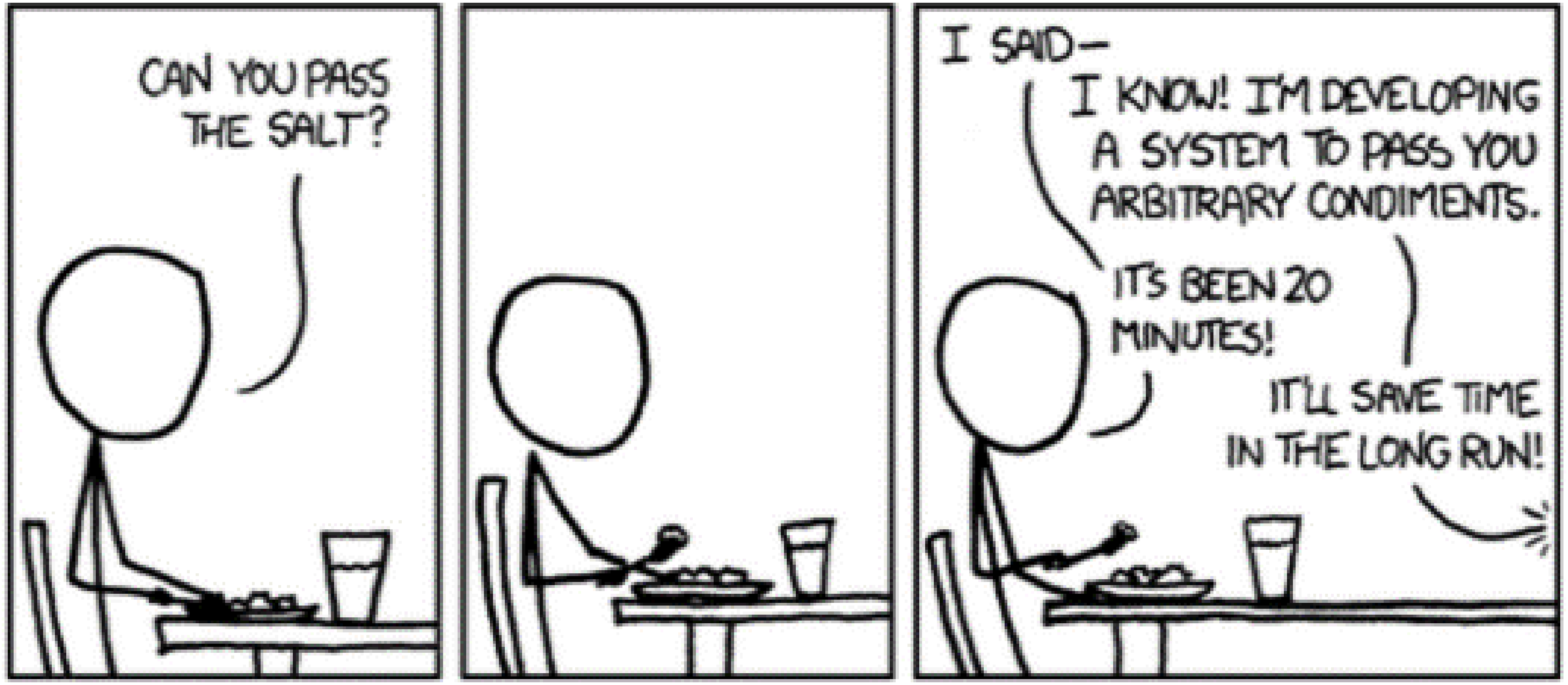


How do other industries deal with this?



Shine a light – see what's in your data

So how do I get started?



Source: xkcd.com

"With a small project, an open mind and a big vision"

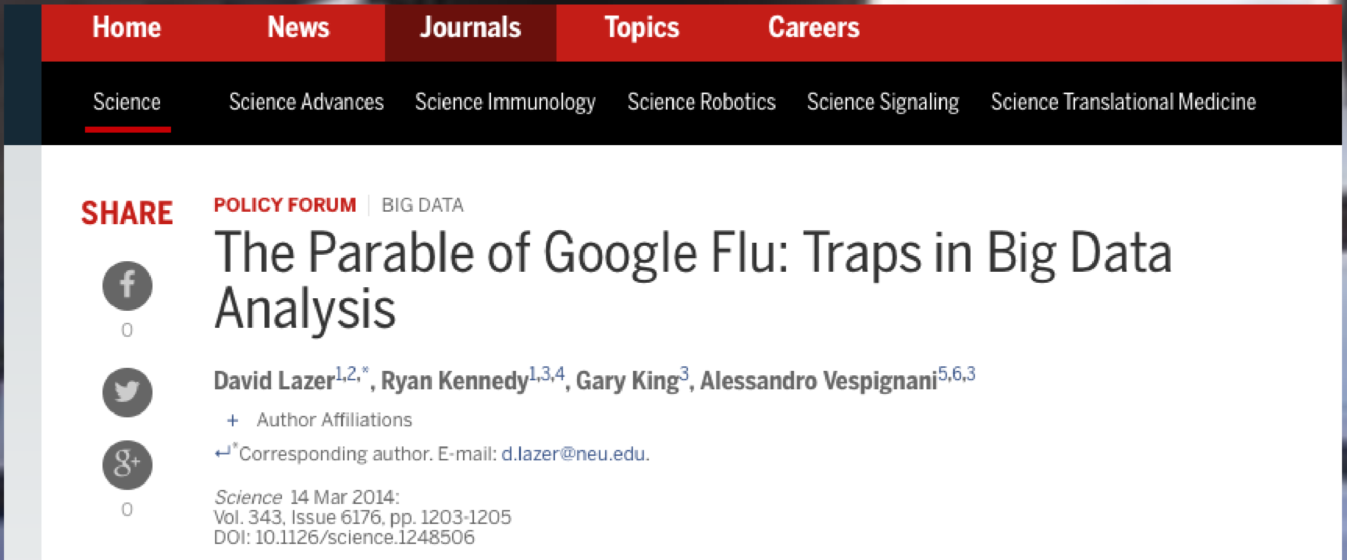
We've heard about the data. So where's the science?

We're still not really sure what use Geostatistics is

You're going to revolutionize E&P with a scripting language, some stats packages and some random data?

Google flu trends let us down

…and where's your data governance?


Home News Journals Topics Careers

Science Science Advances Science Immunology Science Robotics Science Signaling Science Translational Medicine

SHARE

0

0

POLICY FORUM | BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

David Lazer[1,2,*], Ryan Kennedy[1,3,4], Gary King[3], Alessandro Vespignani[5,6,3]

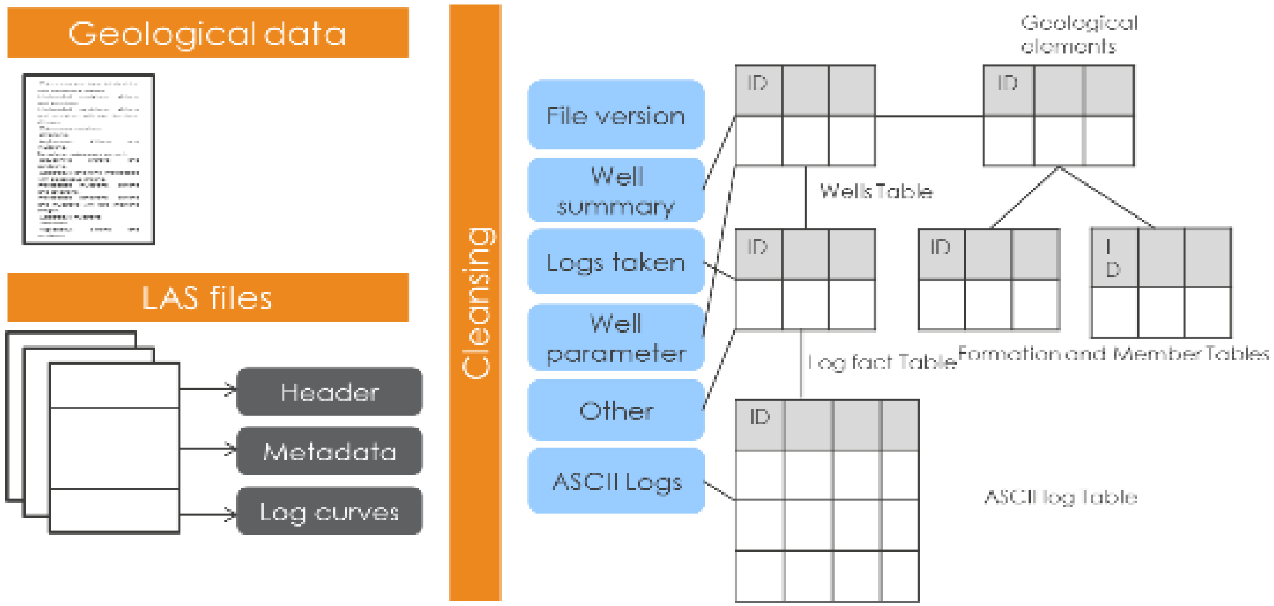
+ Author Affiliations

*Corresponding author. E-mail: d.lazer@neu.edu.

Science 14 Mar 2014:
Vol. 343, Issue 6176, pp. 1203-1205
DOI: 10.1126/science.1248506

Case Study #1
Basin-scale prospectivity analytics

Pragmatic data model from:
- LAS files
- Well headers
- Mud logs
- Well summary
- Completion Report
- A well constrained vocabulary was fundamental to enabling numerical analysis



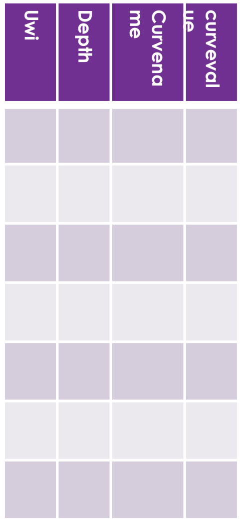
- 6 week MSc project at University of Manchester with New Zealand public data
- 3 weeks spent on data prep and engineering

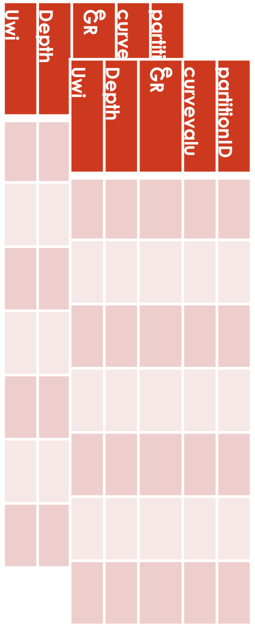
Formation Name	Member Name	New Formation Name	New Member Name
Moki Formation	Moki	Moki Formation	Moki A Sandstone
	Moki A		Moki B Sandstone
	Moki A Sandstone		
	Moki A SS		
	Moki B		
	Moki B Sandstone		
	Moki B Sandstone interval		
	Moki B Equivalent		
	Moki Equivalent		

Case Study #1
Basin-scale prospectivity analytics

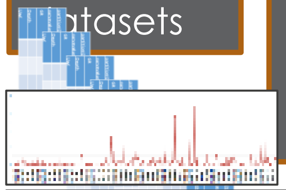
Workflow to classify interbedded sandstone/mudstone and sandstone/siltstone facies:

base dataset → **analytical dataset** → **nPath** → **Rebuild and pivot** → **Apply time warping** → **locate**

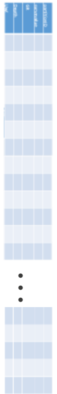




nPath: create multiple windows (paths) along datasets

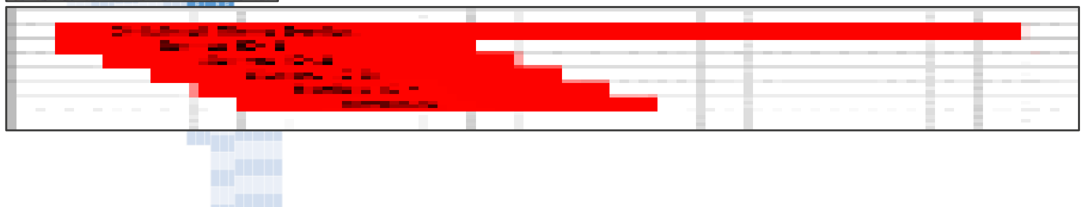


Rebuild and pivot: Flip paths of points into separate key-value pairs and rebuild as table



Apply time warping: Create *template* and map onto partitions in pre-built table then calculate goodness of fit using DTW

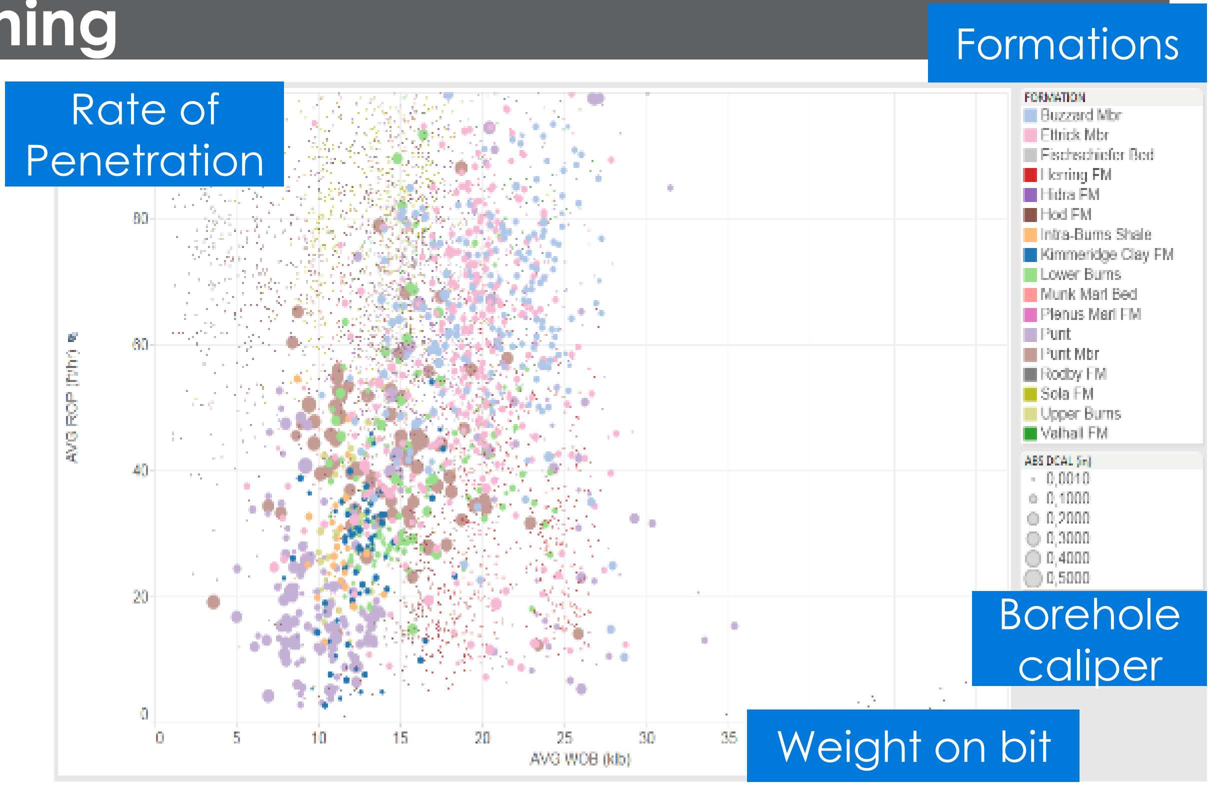
locate: Take top centile of matches and validate



- A much clearer, simpler reservoir model with 62 members in 17 formations
- An open-ended model to incorporate other data **(e.g. production histories)**
- Ask any question of the data with spatial, chronological and logical relationships – **at scale**
- Identified overlooked pay features (hot shales) and re-classified others (interbedded facies)

Case Study #2
Drilling and Well analytics: Planning

- Data analytics across Drilling & Wells is not typically performed due to silos and limitations of existing solutions
- Modern D&W activities already generate a large number of parameters and will generate even more in the near future



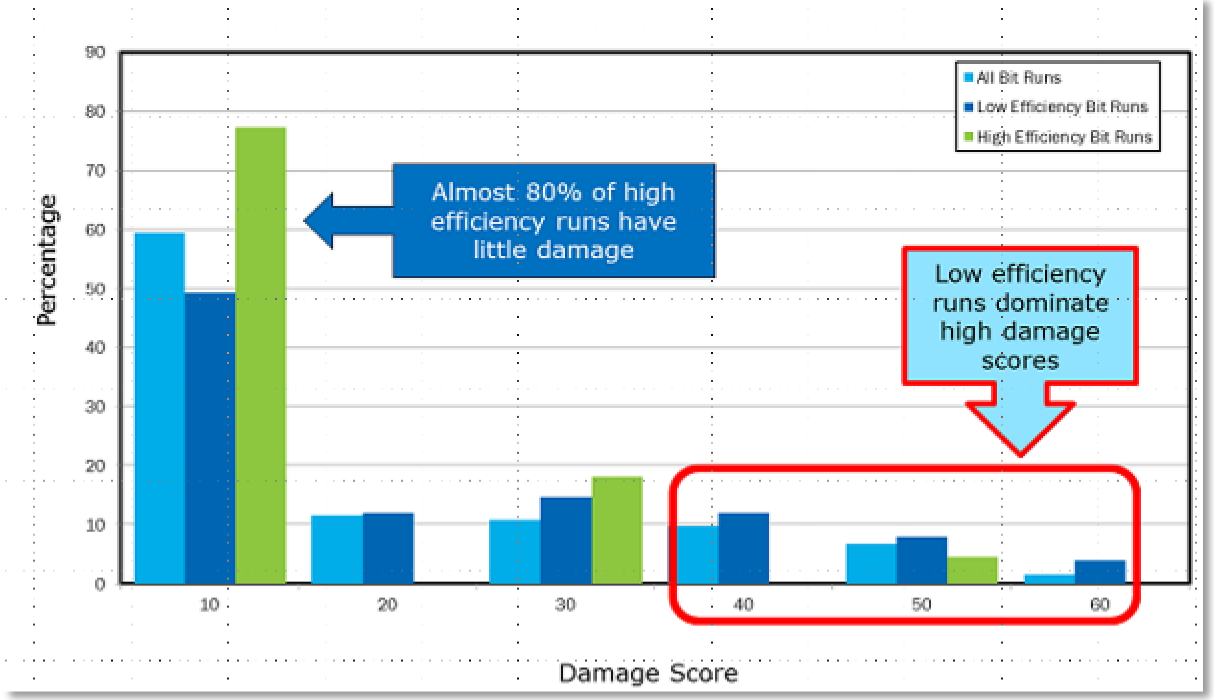
- How will oil and gas operators ensure safe, accurate, efficient and economical D&W operations?
- CGG has access to geology, petrophysics, wells, and drilling data
- Teradata provides analytical platform to run complex data analyses

- We can identify trends, patterns, and risks in D&W domains and suggest optimal parameters for D&W planning and operations

Source: CGG-Teradata

Case Study #3
Drilling and Wells analytics: Operations

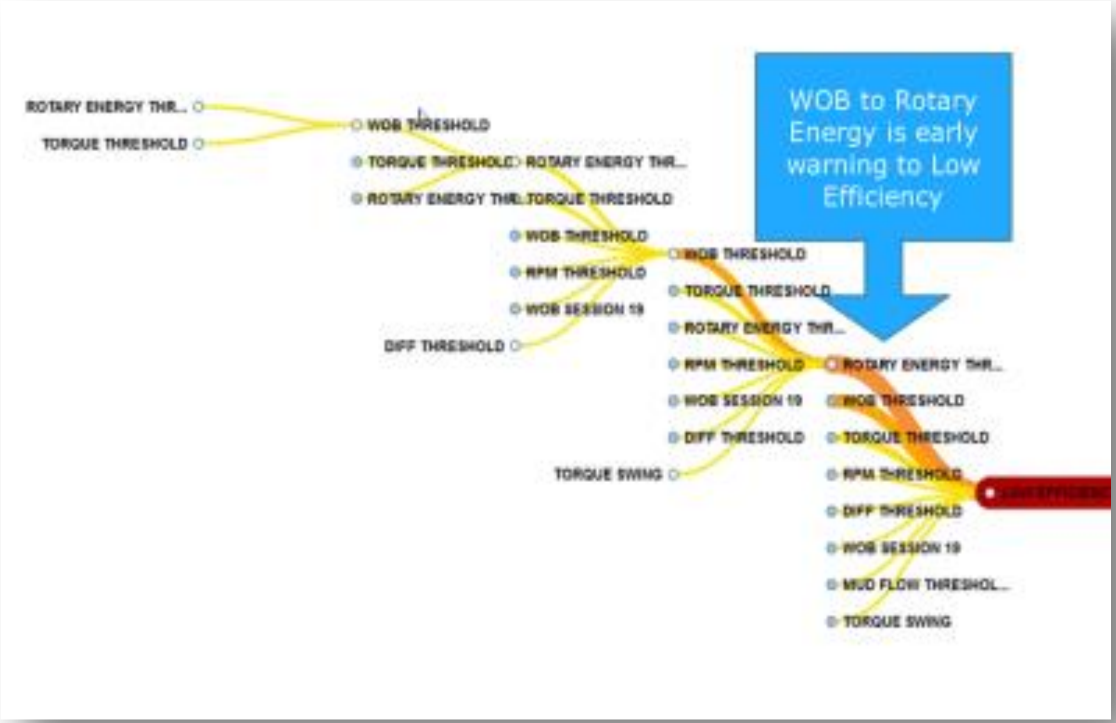
- surface and downhole
- metadata relating to well and drill string
- bit damage severity and profile
- well position and trajectory
- petrophysical information



- "It's just hard formation – that's the way it is".
  Unpredictable and repeated failures occur. Some single-trip sections achieved, but success/failure criteria not understood

- look for patterns to that will inform better operational decisions: increase drilling efficiency to avoid catastrophic bit damage
- An 8-week Data Science study across scientific and operational datasets identified $17M of savings in drilling practice

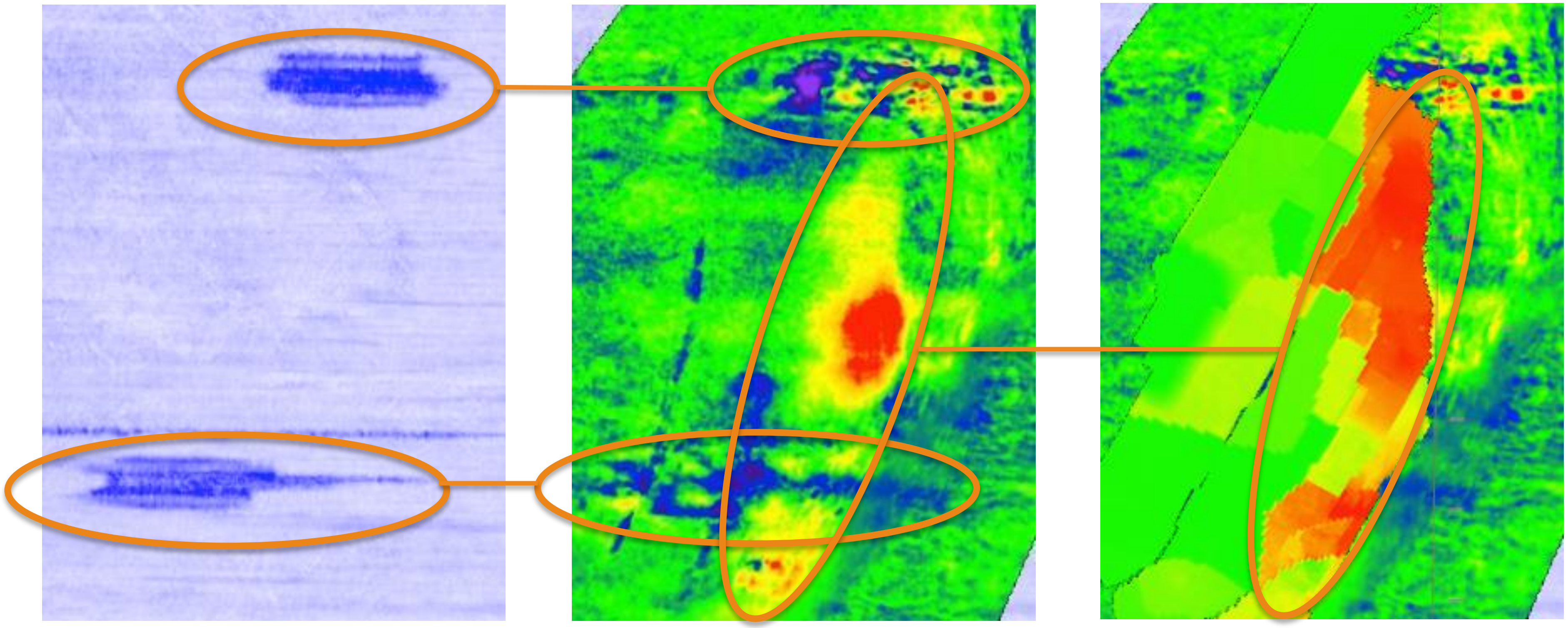
Case Study #3
Drilling and Wells analytics: Operations



- Find combinations of a wide range of drilling parameters likely to avoid bit failure and model alarms to ensure efficient drilling
- Create rules for best practice during operations based on ever-growing knowledge base

- Consistently drill horizontal section in a single trip in hard formations

Case Study #4
4D seismic effects



Repeatability (NRMS)

Time shift

Pressure difference

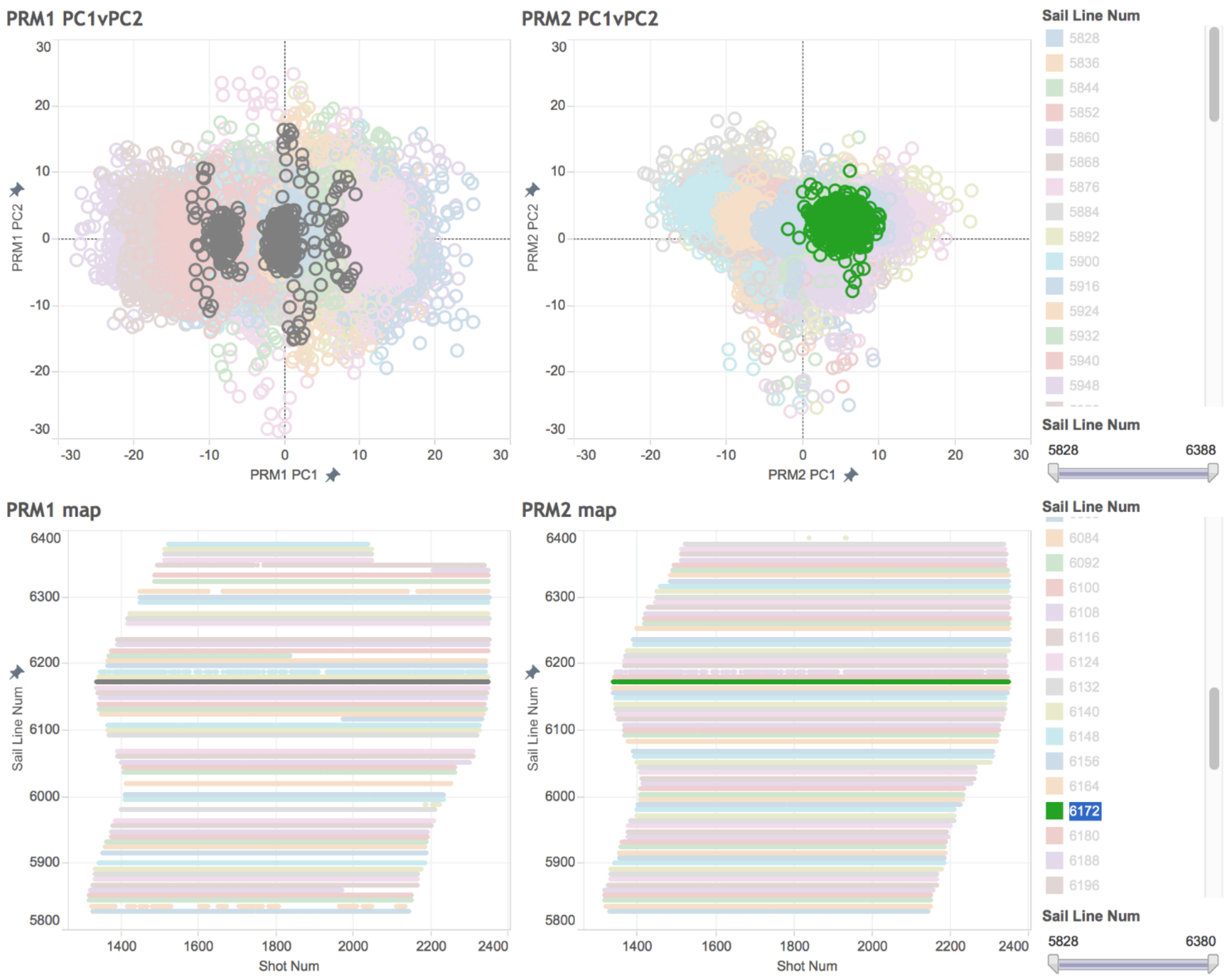
- See PNEC2014 or EAGE papers!

Case Study #5
4D Seismic acquisition analytics

- Navigation, gun array, Met/Ocean and seismic trace data from 4D surveying
- How can data be integrated for analysis and possible operationalization?

- What is there of value in the multitude of file formats?
- What are the analytical questions?
- What approaches?
- Lots of science v. lots of stats!

- What value in the answers?
- One-off insight or should it be operationalised?



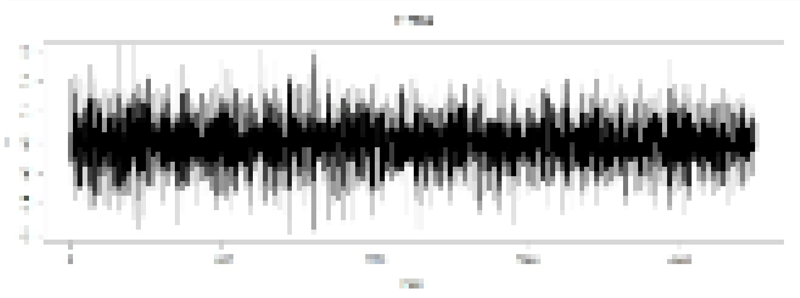
Case Study #6
Passive Seismic Monitoring analytics

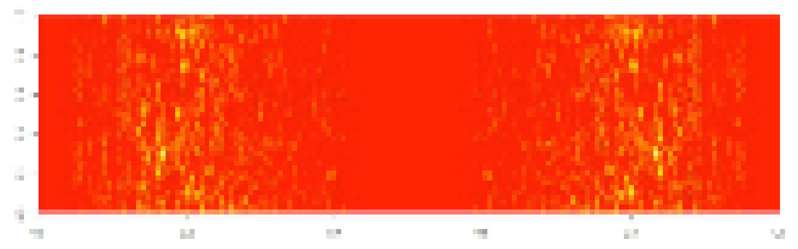
Internal 48 hr Hackathon to test Teradata capabilities with passive seismic reservoir monitoring data

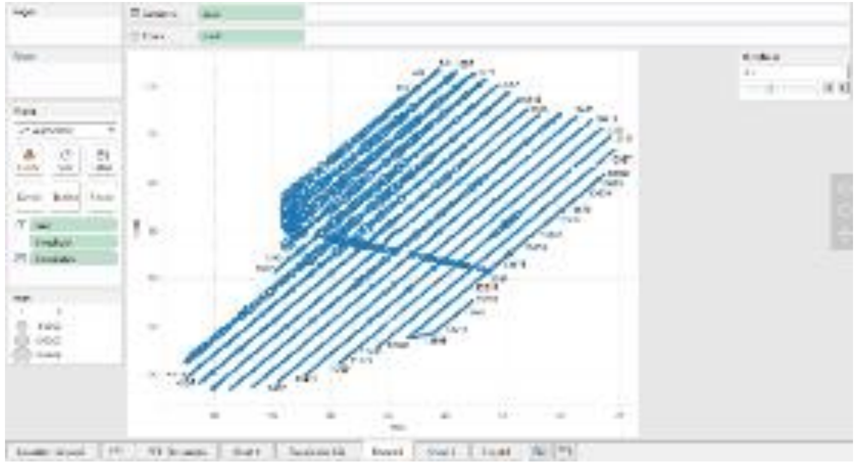
- 2 hrs of data from seafloor array containing seismic event from casing collapse
- What are data management considerations for these analytical workloads

- What are the analytical components and strategies?
- Can the event be located in time and space?
- Can we define signatures for operationalization?
- Can precursors be extracted to provision early warning?

- Could borehole collapse event be identified using simple statistical filters?

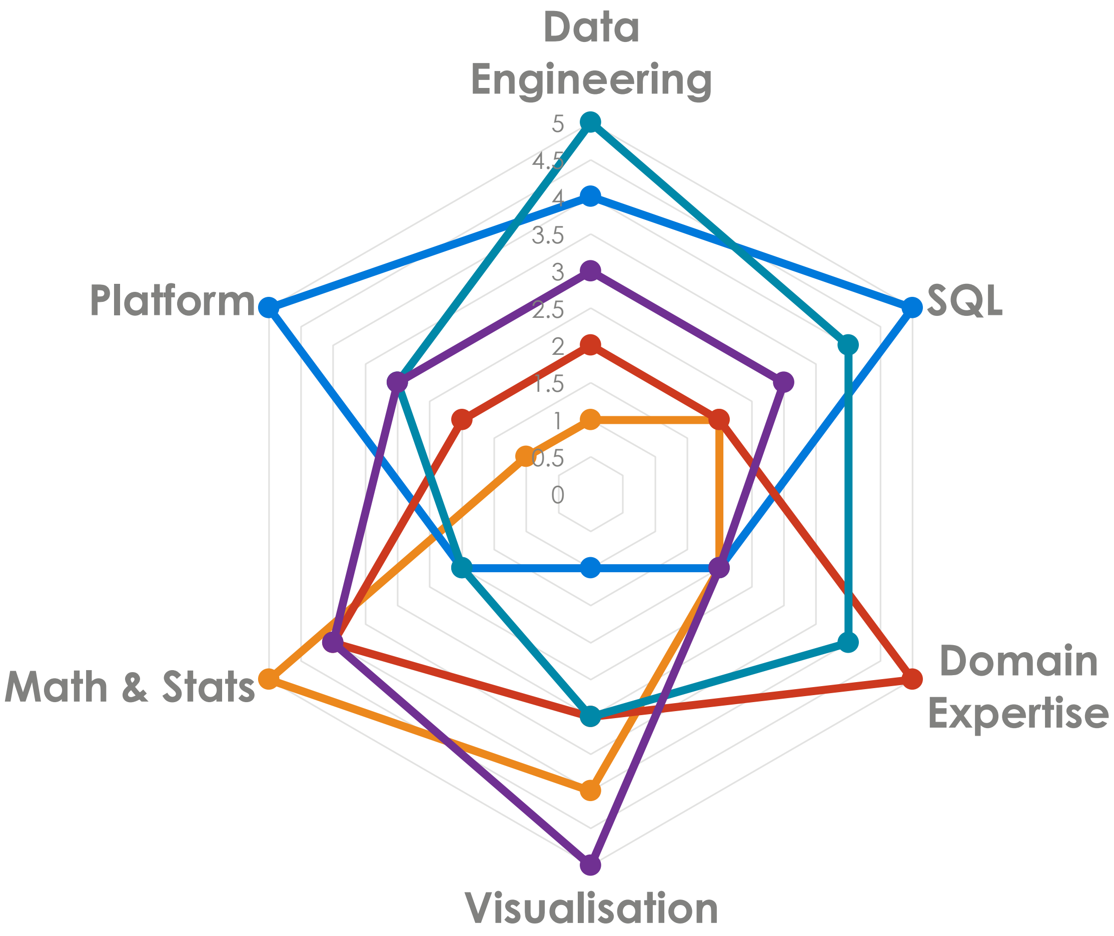






What should a data science team look like?

- No such thing as a perfect data scientist
- For deployment you need platform expertise
- You need outstanding data management and data engineering skills (and culture)



Data Management Learnings

- Loading into granular form
- Single view of data for whole team (cloud, or on premise)
- No up-front modelling
- Clear documentation and audit trail
- Keep loaders in a repository so they can be reused –not bound to application import functionality
- Data Lineage – reproducibility
- Data Quality – profiling what numerical values make sense?

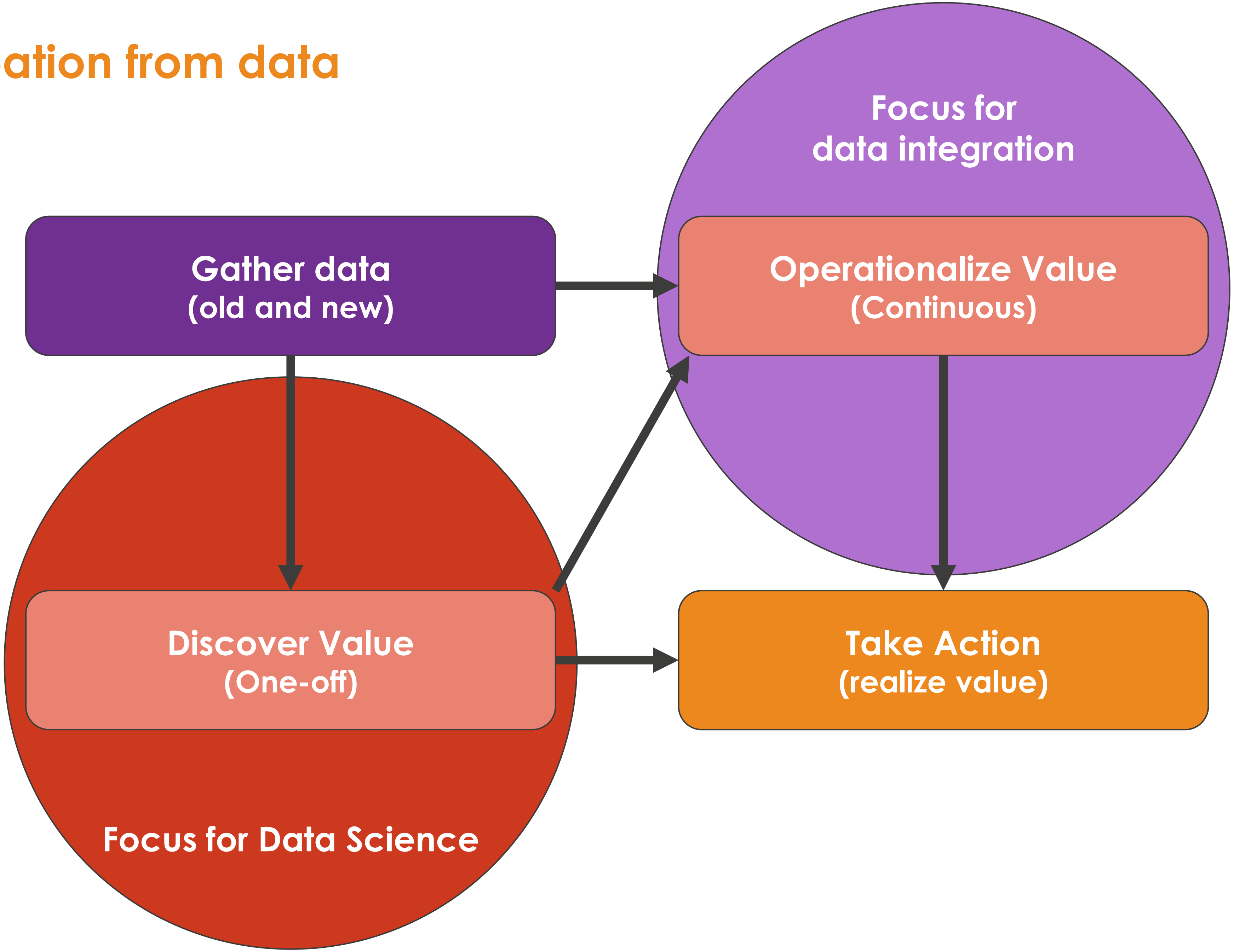
Data Engineering Learnings

- How should data be stored?
  - Granular
  - Profiles of activity – e.g. regular frequency profile instead of storing complete time series
  - Profiles of valuable patterns
- Use a scalable platform (MPP)
- Use a language that is as universal as possible e.g. python
  - Data Analysis – sciPy, NumPy require scientific and numerical prowess
  - APIs into other domains e.g. HPC, filesystem, visualisation

Data Mining Learnings

- **Keep data online and accessible** – one-off studies may lead to a more operationalised event processing usage
- **Profile incoming data** regularly (e.g. production time series every few minutes across a reservoir) – keep profiles as descriptions of system states
- **Store well-understood patterns** of behaviour for repeatable mining (i.e. where have I seen this before?)
- **Document activity continuously** – people and skills are fluid through the life time of data. What has worked, what hasn't worked, what approaches were considered but never picked up?

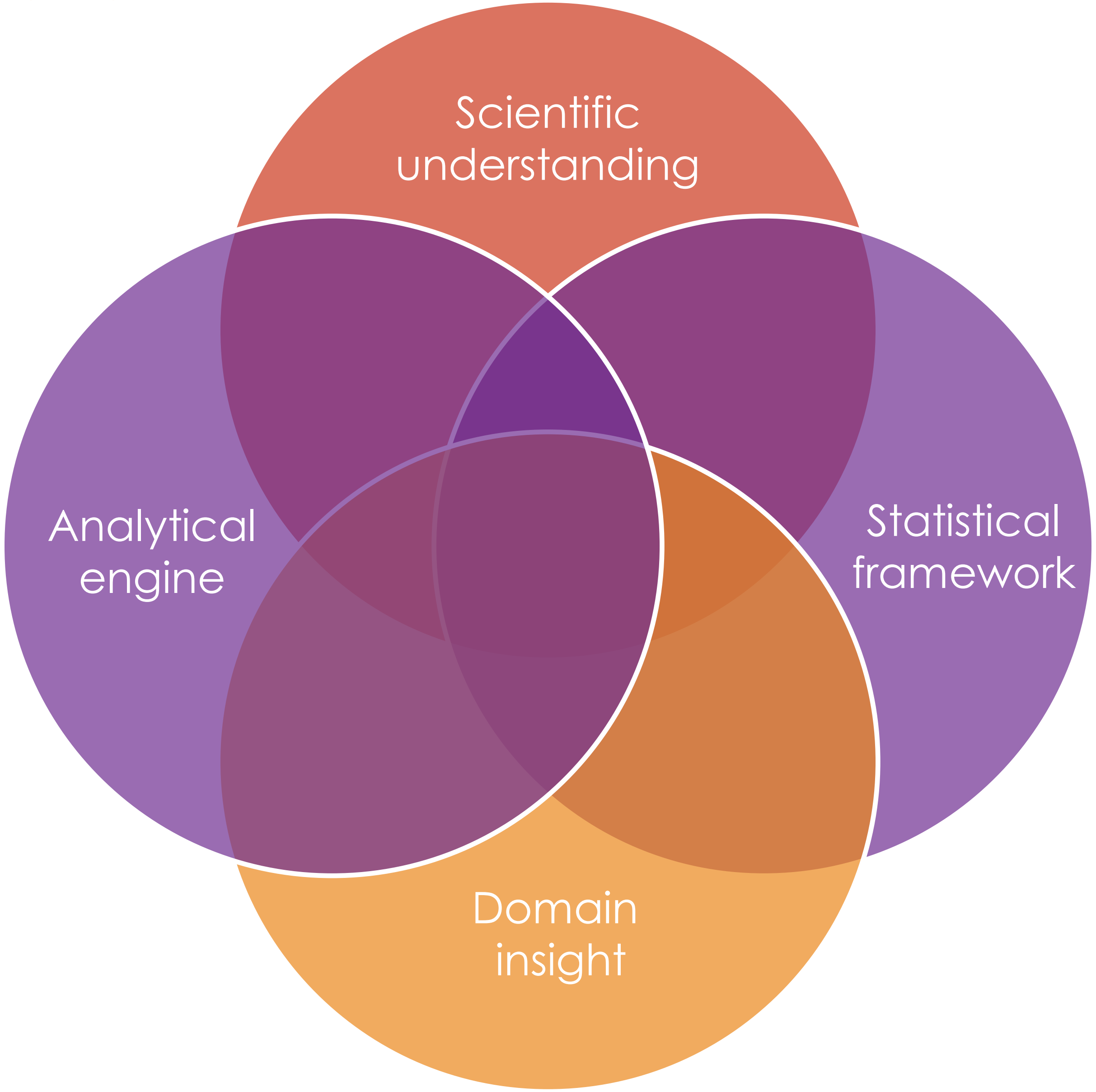
Value creation from data



Business Impact Learnings

- Domain understanding is vital
- Have a well-scoped value proposition
- Work in agile mode with regular, well-managed sprints (no fixed agenda, no free-for-all)
- Have good visualisations
- How will you deploy and operationalise your insights?

The magic ingredients



Duncan Irving
duncan.irving@teradata.com
@duncanirving