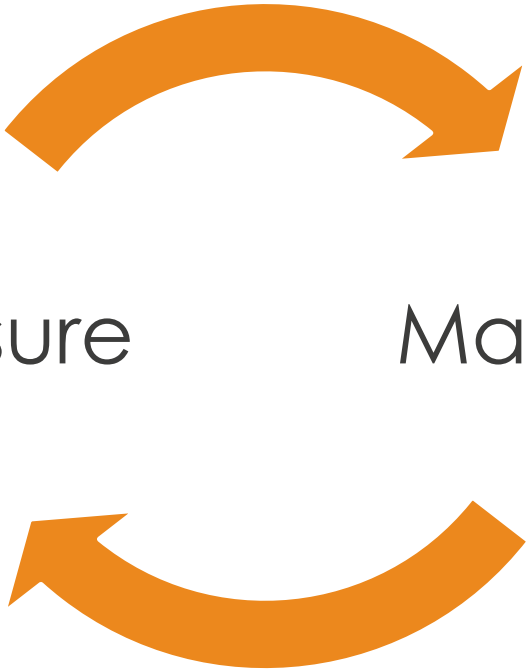
Doing DIGITALISATION **right** (***with machine learning***)

Duncan Irving

Finding Petroleum, November 2017

Digitalisation? …not another buzzword, please



Measure Manage

"It is wrong to suppose that if you can't measure it, you can't manage it – a costly myth."

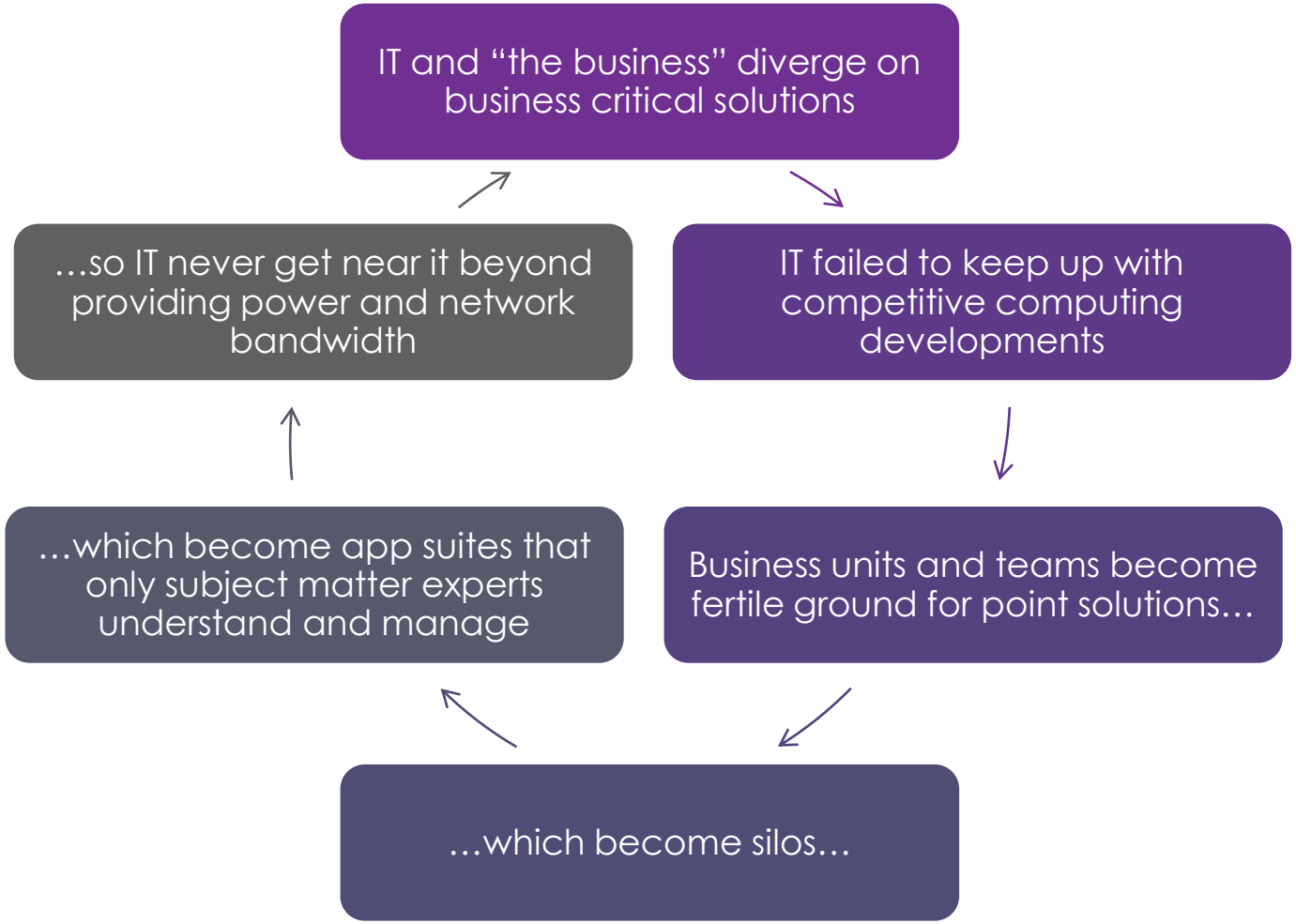
W Edwards Deming

But didn't we do the Digital Oilfield a decade ago?

THINKBIG ANALYTICS
A TERADATA COMPANY

"No, the Digital Oilfield is a bunch of operational technologies that operate in an effectively closed loop.

They never linked the oilfield to the supply chain depots or finance and planning functions."

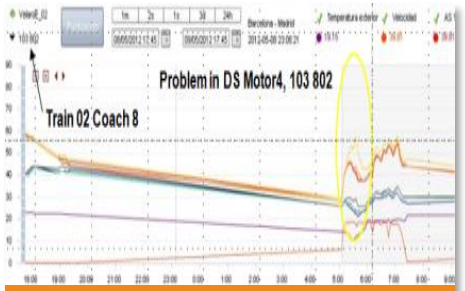


Data-driven maintenance powers business-critical capabilities

THINKBIG ANALYTICS
A TERADATA COMPANY

SIEMENS

- "Performance-based maintenance" make Siemens the operator of the **most reliable high-speed train** on RENFE network
- One in 2,300 rides noticeably delayed: key criterion for business success: **passengers reimbursed when delay >15 mins**



Machine Learning to identify failure onset

- Sensors measure constantly **key parameters** e.g. traction motor bearings
- Analytics enables a **stable incident prediction** based on strong understanding of patterns
- Abnormal patterns trigger inspection to prevent failure on track

MAERSK LINE

Piston ring failures are infrequent events (and hence little data), but can immobilize a vessel at sea. Data integration exposed statistical relationships in behavior that was not seen in conventional data silos.

- Delivered **>10-day lead time** on predictions for component failure
- **75% of failures** (target: 60-80%) captured by sensor-based analytics
- **63% accuracy** in prediction (target: 50-65%) of piston condition

This new understanding of failure of fundamental components such as piston rings could be acted on for productionised business impact.

Integrating HSSE data across other business functions for wider impact

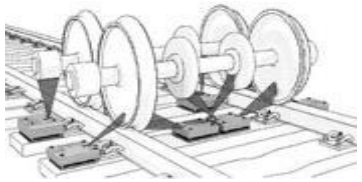
THINKBIG ANALYTICS
A TERADATA COMPANY

Analysis of track side data collected for safety improvements:

- Bearing-related derailments reduced by 75% - hundreds of minor incidents per year, plus some major ones
- 20 million daily pattern matches of sensor spot 1500 issues/day with problem wheels

3350 trains per day (hundreds of axles each)

45,000 km of 150 yr. old track



Infrared sensors every 30 km gather wheel temperature and send to operations centre.

The drive for improved safety led to new service offerings when integrated with operational, resourcing and planning data:

- More competitive pricing structures
- Improved revenue collection
- Better equipment utilization
- Millions saved in payroll audit / time keeping process
- Analysis provides impact of having added 141 locomotives to fleet

"Estimated benefits from the utilization of Union Pacific's enterprise data warehouse is in excess of $100 Million each year." (Union Pacific)

H2S (hydrogen sulphide) can be deadly even at low concentrations. Wells are properly equipped for safety monitoring. Sulphide deposits also impact flow rates, and subsequent corrosion compromises asset integrity. Data was stored in HSE, ops and maintenance silos.

Data mining and real-time insight operationalisation led to:

- Understanding of how to mitigate sulphide deposits
- More effective well planning and production engineering
- Ensure correct equipment and people are assigned to drilling and interventions based on H2S shows

H2S dashboard drives decision support workflows for artificial lift

Strong safety regime was paramount – but more well plans completed on time, and higher production assurance achieved as a result of real-time data integration.

Five Key Observations from Upstream studies

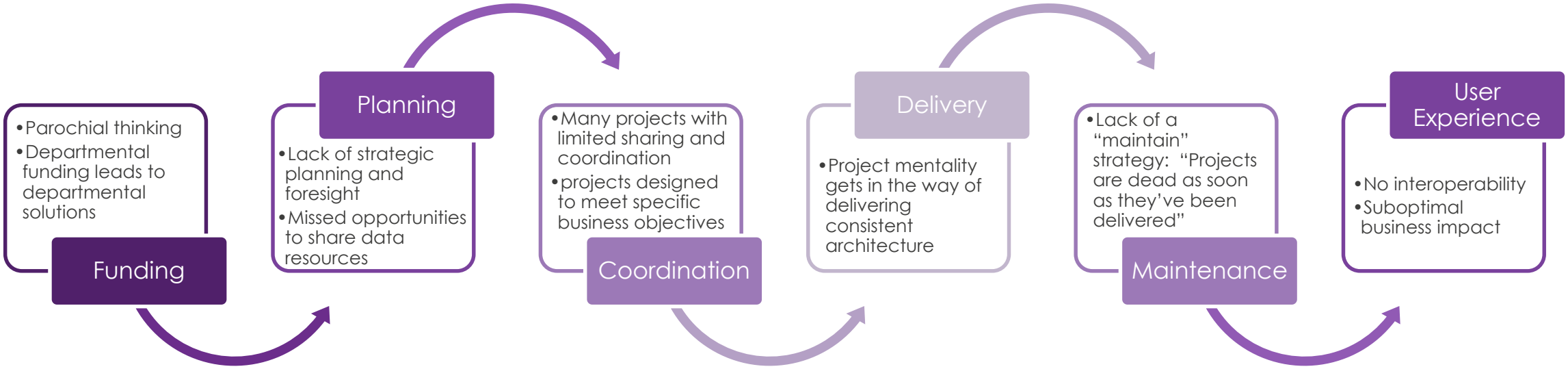
THINKBIG ANALYTICS
A TERADATA COMPANY

1	**Organisational** constructs, culture and a **project mentality** make value delivery difficult
2	**Business Requirements** gathering and prioritisation needs improving
3	Key common **Data Management capabilities** not in place across upstream
4	**Definitions** – key terms mean different things to different people –communication fraught with misunderstanding
5	**Future Data Platform** IT architecture often **lacks integration capabilities**

Organisational and Cultural Challenges

THINKBIG ANALYTICS
A TERADATA COMPANY

Issues arising from a project-centric approach



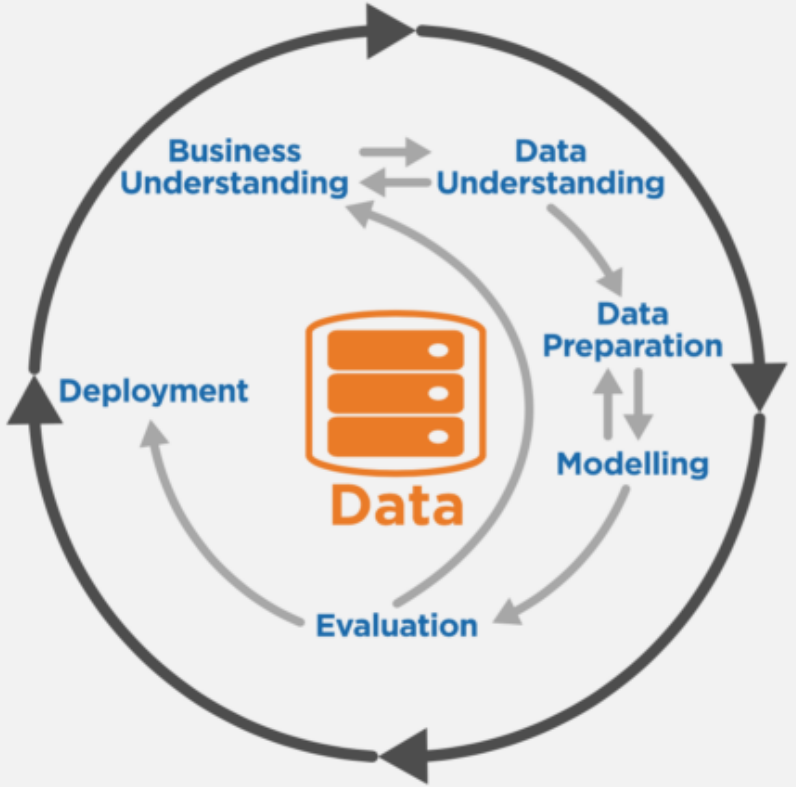
Governance is a key driver to a coherent a data strategy but is not addressed strategically

Data Science or "Data R&D" is concerned with the *discovery* of hidden patterns and insights

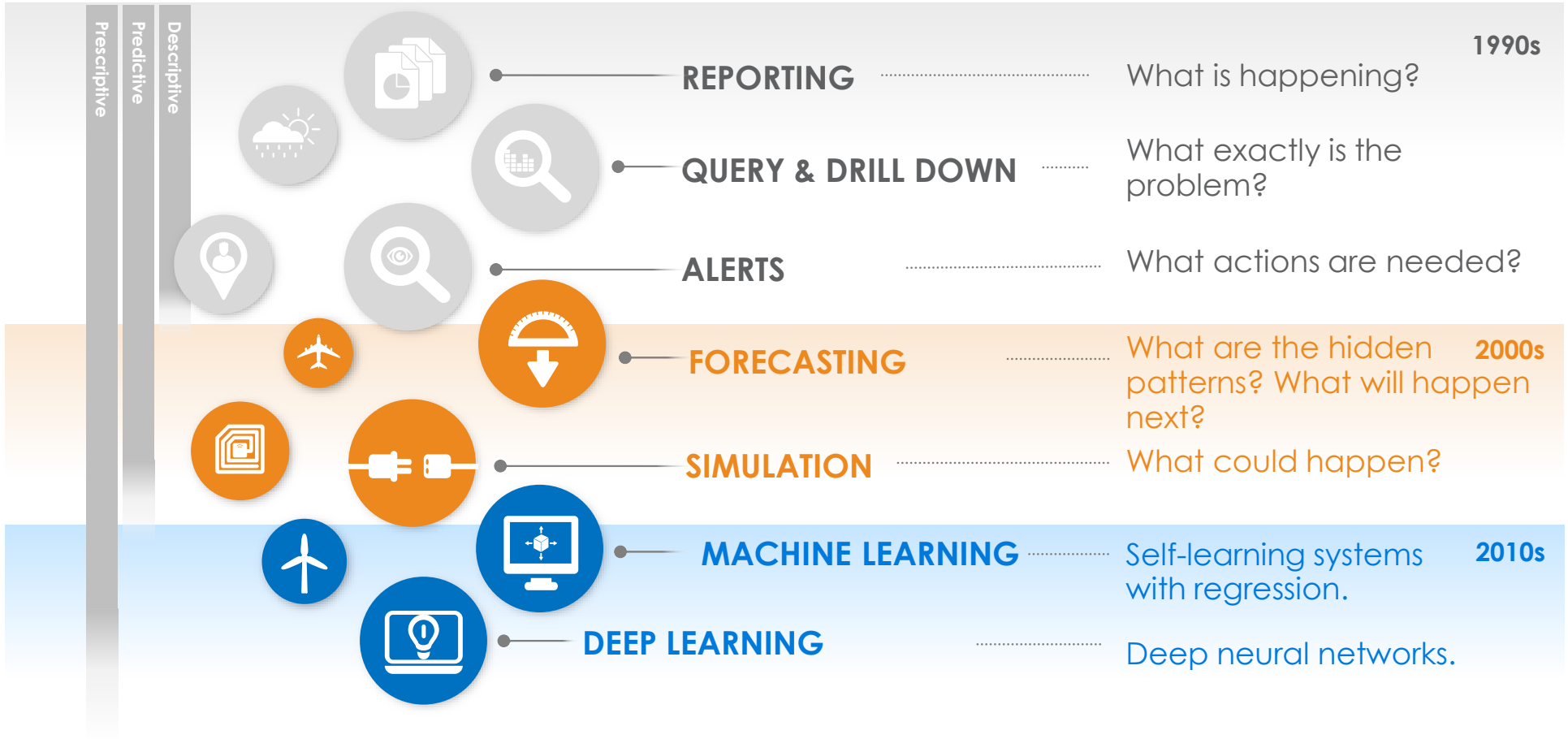
THINKBIG ANALYTICS
A TERADATA COMPANY

"...data mining is an **exploratory** undertaking closer to research and development than it is to engineering. The CRISP cycle is based around **exploration**; it iterates on approaches and strategy... outcomes are far less certain, ***and the results of a given step may change the fundamental understanding of th***

Data Science For Business,
Fostor Provost & Tom Fawcet, my emphases



Multi-genre Analytics Evolution



Apply Common Dimensions of Data Quality

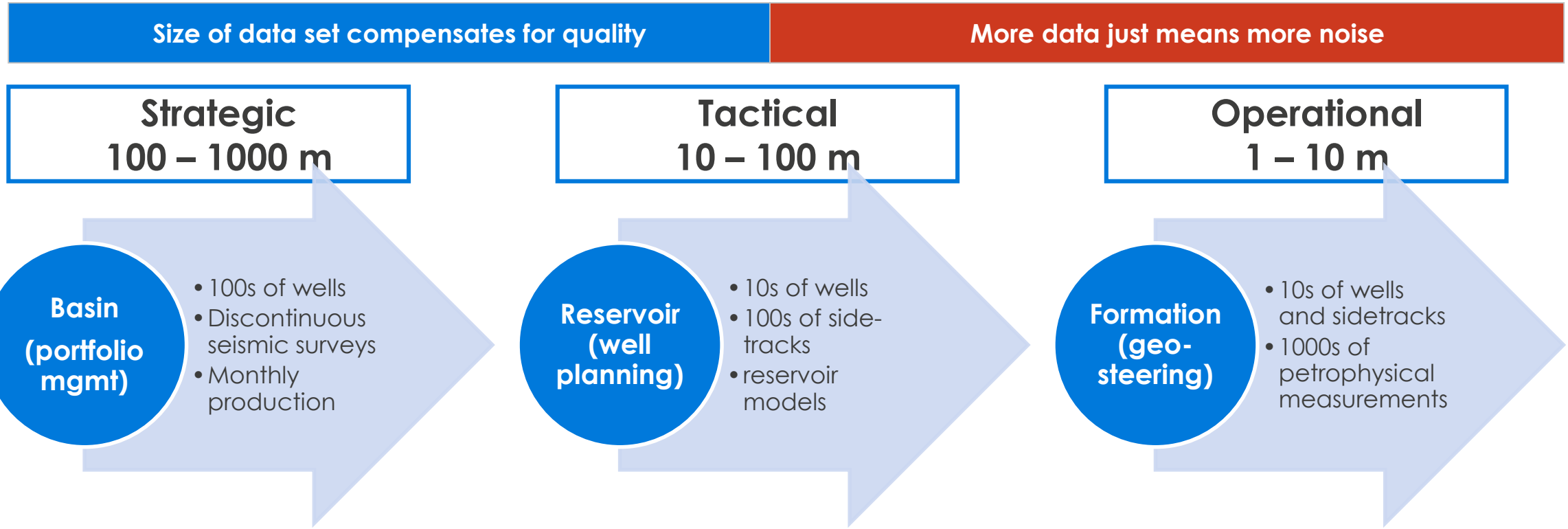
Quality should be consistently defined and monitored to drive performance measuring

THINKBIG ANALYTICS
A TERADATA COMPANY

Dimension	Description	Conformance
Accessibility	A measure of information access / availability	Data is available for all users
Accuracy	A measure of information correctness	A flow rate of 12.7 m^3s^{-1} really is stored as a flow rate of 12.7 m^3s^{-1}
Consistency	A measure of the conflicts with redundant data	A pressure of 1200 bar in the ABC system is also stored 1200 bar in the XYZ system.
Entirety	A measure of the quantities of entities created	All well interventions made were recorded and stored.
Breadth	A measure of the amount of information captured about an object or event	All information about a specific injection regime is captured including duration, start and stop time, injected material, etc.
Completeness	A measure of information captured within a specific entity	UWI, status, field, well name are all stored for each well
Uniqueness	A measure of unnecessary information replication	Well Header is stored once for each well.
Interpretability	A measure of semantic standards being applied	A date is stored as 11 June 2002
Freshness	A measure of how current a record is	Well status represents current status
Precision	A measure of exactness	The timestamp for Historian data is precise (e.g. 13:42:00.015627)
Depth	A measure of the amount of history that is retained	A complete history of interventions, modifications, shut-ins, simulations, well tests is stored for each well
Integrity	A measure of validity relative to another item	Measured depth should be greater or equal to total vertical depth

Quality drives business application

THINKBIG ANALYTICS
A TERADATA COMPANY

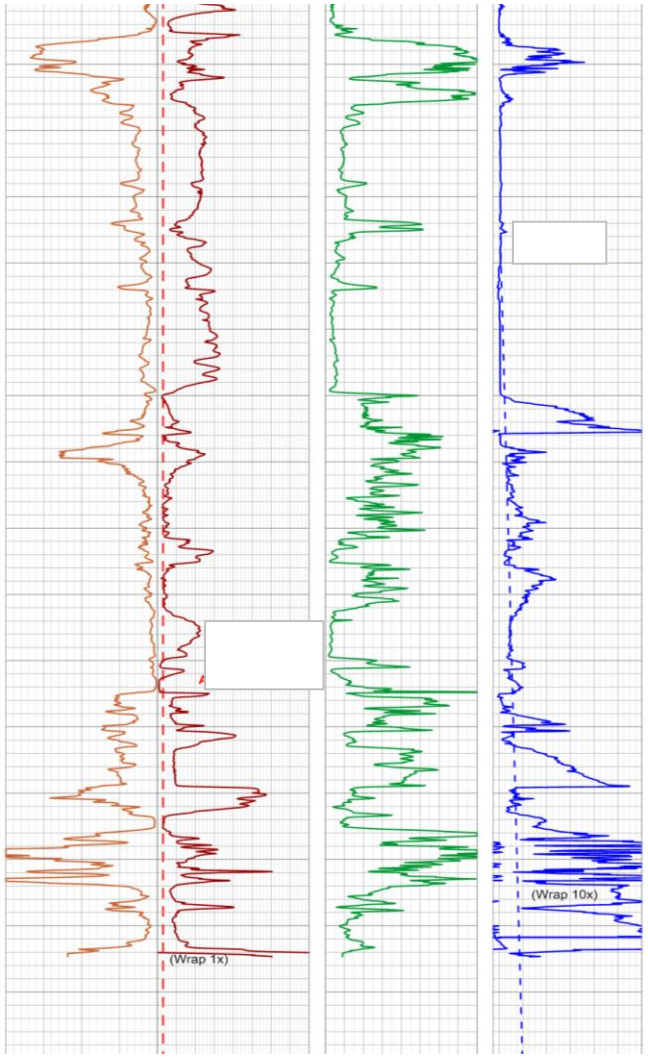


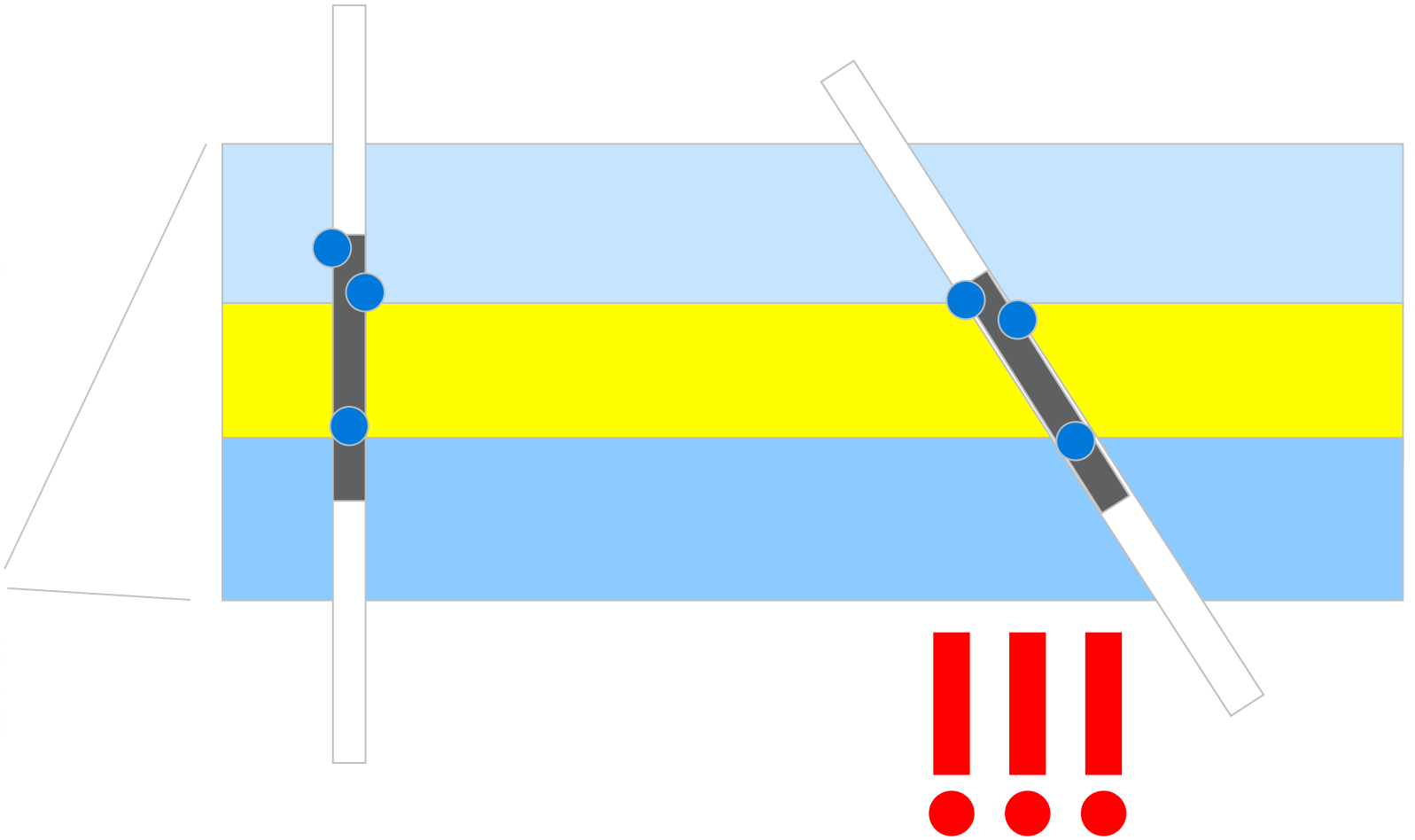
Basic CRS and projection checks. Units? Duplicates? Nomenclature?

Reduced standard vocabulary, numerical validation at file level

Knowledge of tool string, muds, borehole reports, petrophysical database

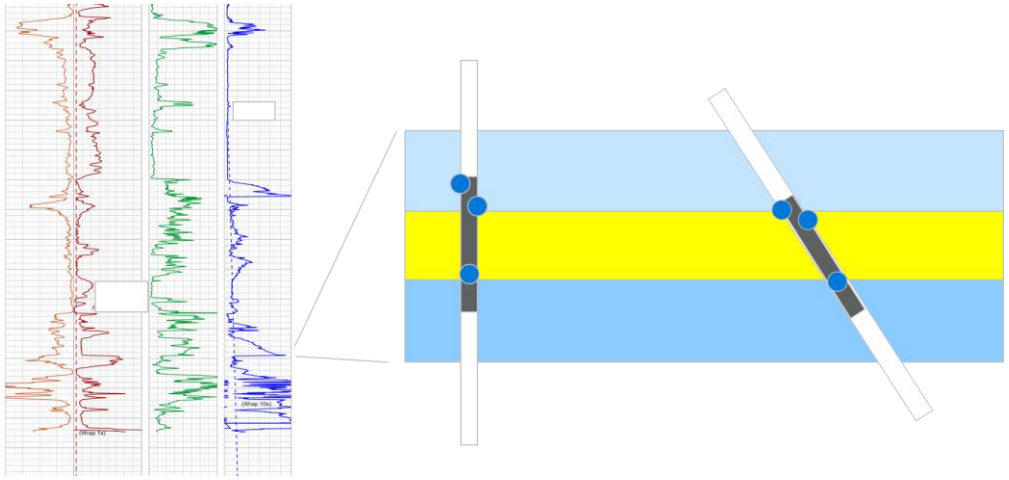
Let's take a close look at your well logs…





Can we fix it? (Yes we can)

- Better understanding of the problem and hence the business requirements
- Stronger Master Data Management – ensure data is indexed to the granular level
- Better Metadata Management – build cross-domain catalogues
- Quality and uncertainty metrics can be at the measurement level – storage is cheap



Data should be corrected for:

- Pull rate
- Vertical location
- Angle
- Numerical range
- **Missing data**

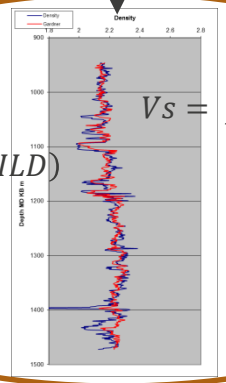
Compensating for missing data at the measurement level

THINKBIG ANALYTICS
A TERADATA COMPANY

LAS Tops

Well	Vp	Vs	ρ	Rt	Vsh
W1	✓		✓	✓	✓
W2	✓	✓		✓	✓
W3		✓	✓	✓	✓

$Vp = f(depth, ILD)$



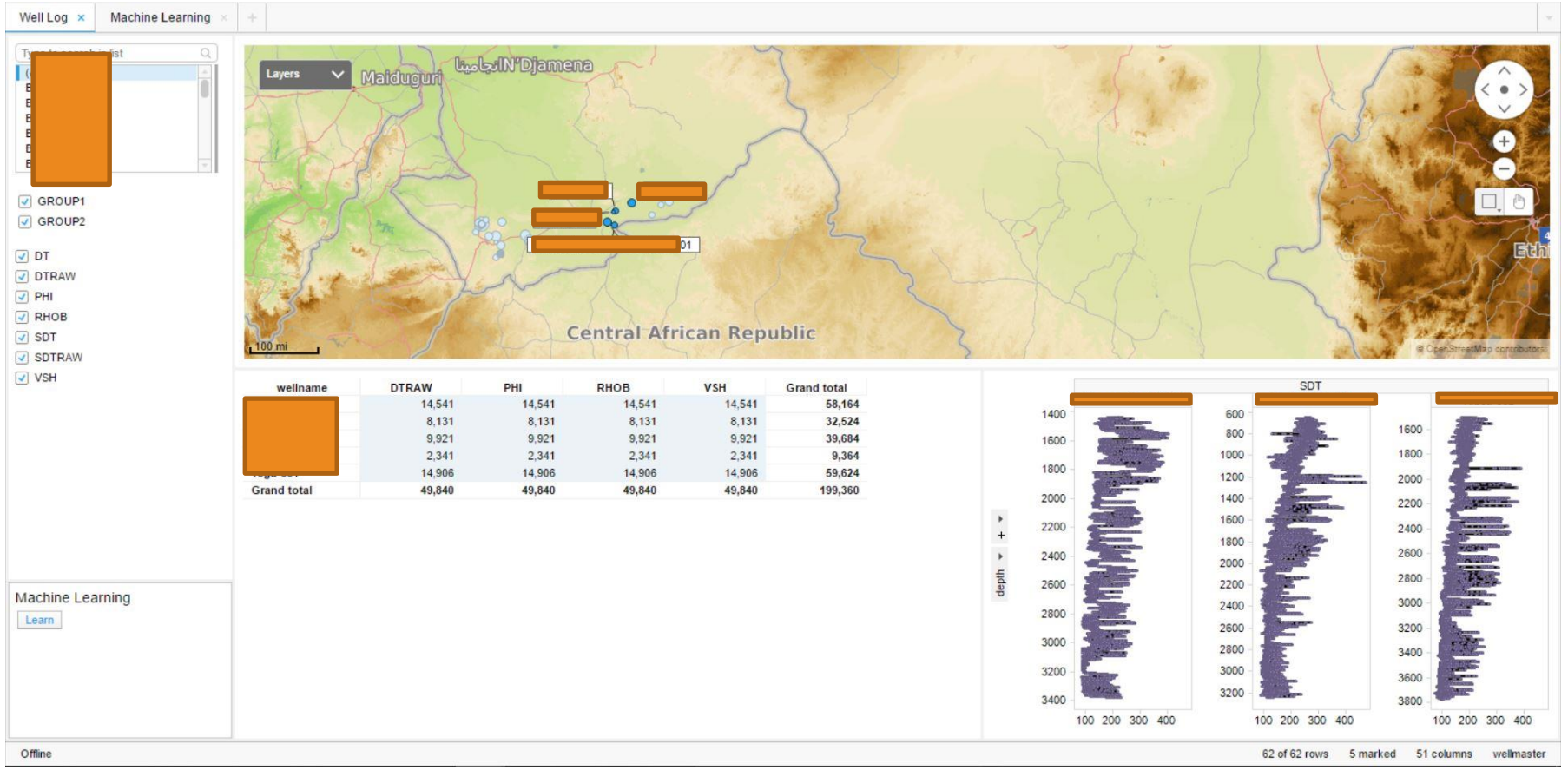
$Vs = f(Vp, Vsh)$

$\rho = f(Vp, Vsh)$

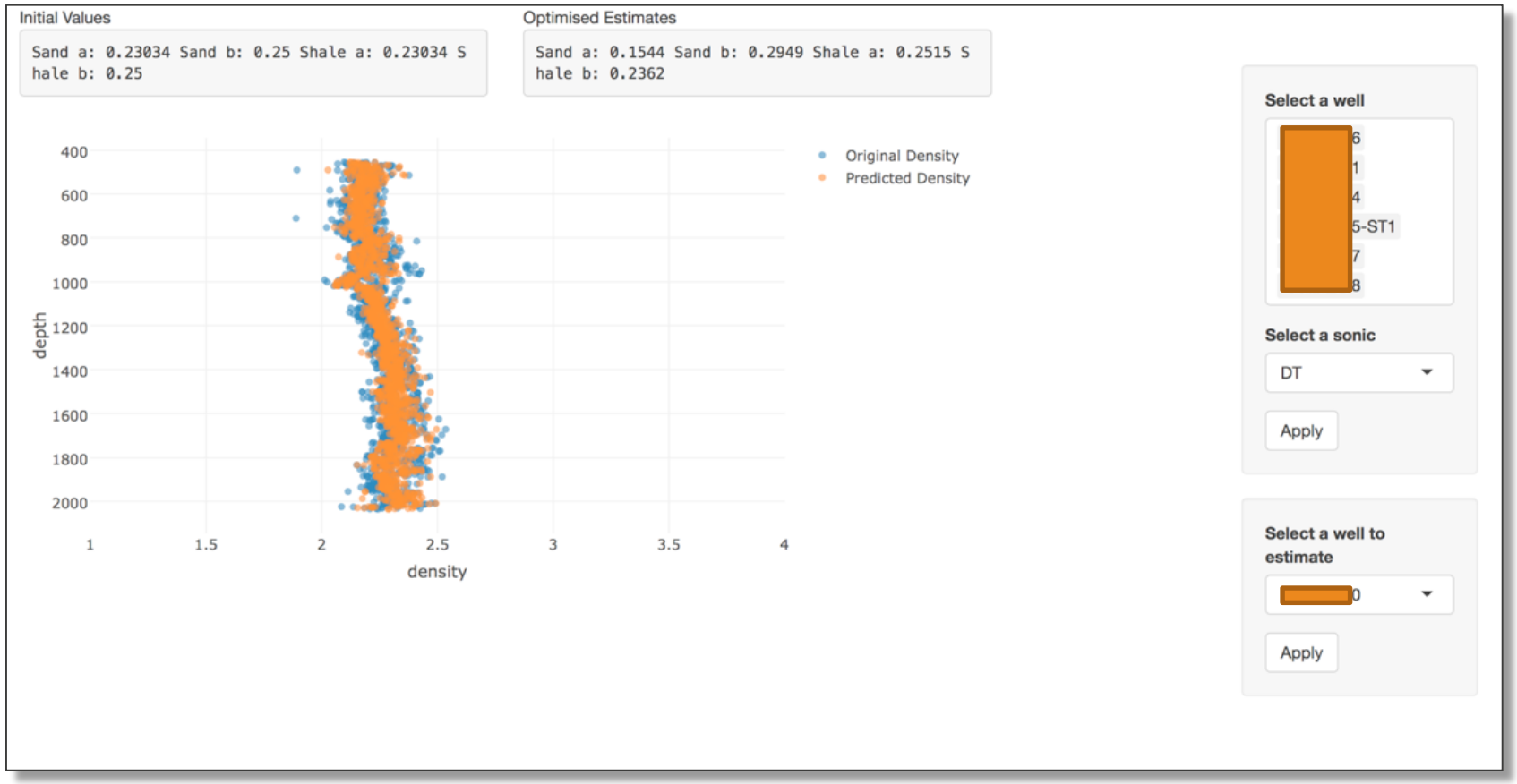
Well	Vp	Vs	ρ	Rt	Vsh
W1	✓	✓	✓	✓	✓
W2	✓	✓	✓	✓	✓
W3	✓	✓	✓	✓	✓

LAS
Rock Property Zone Averages

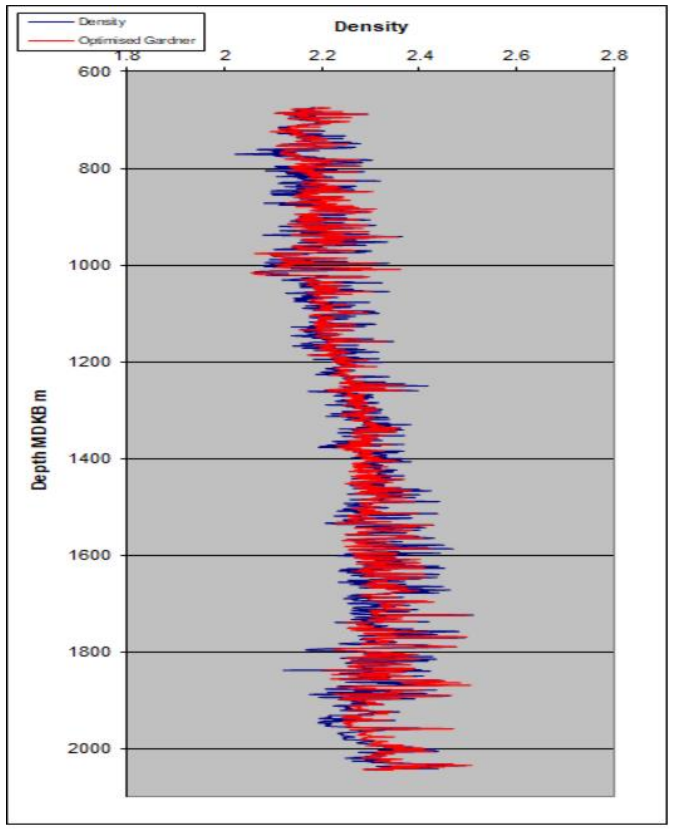
Basic user interface with standard BI tool
(a few hours' work for proof of principle!)

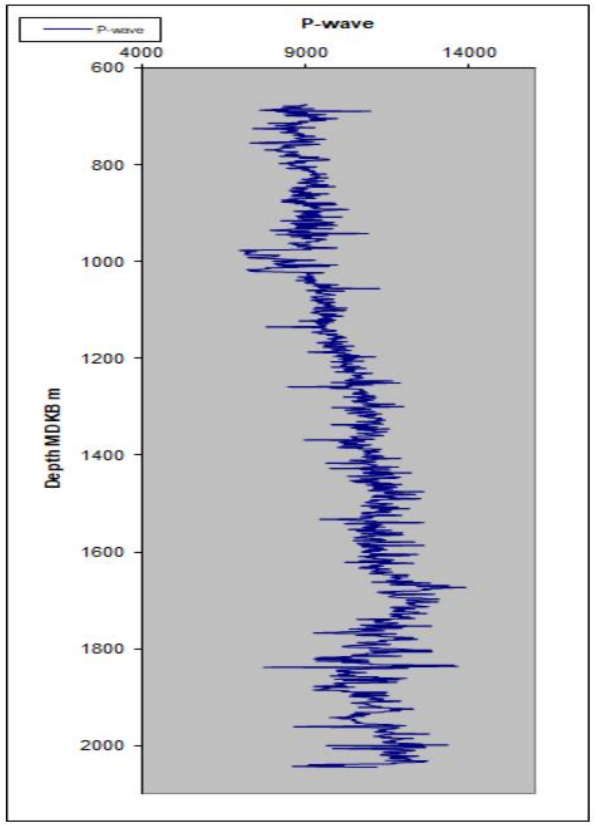


Real-time well log triage running ML algorithm



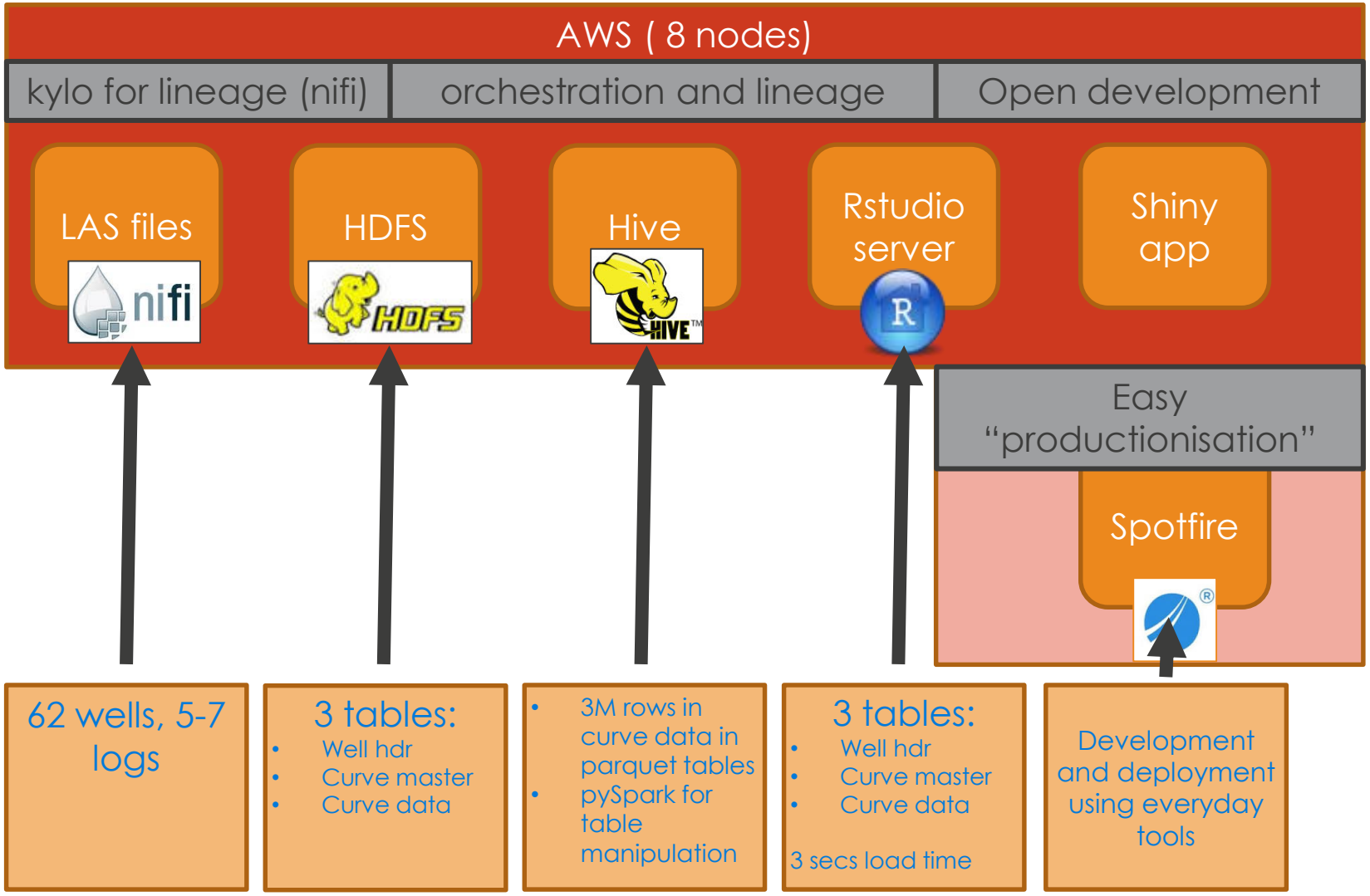
Does the ML work?





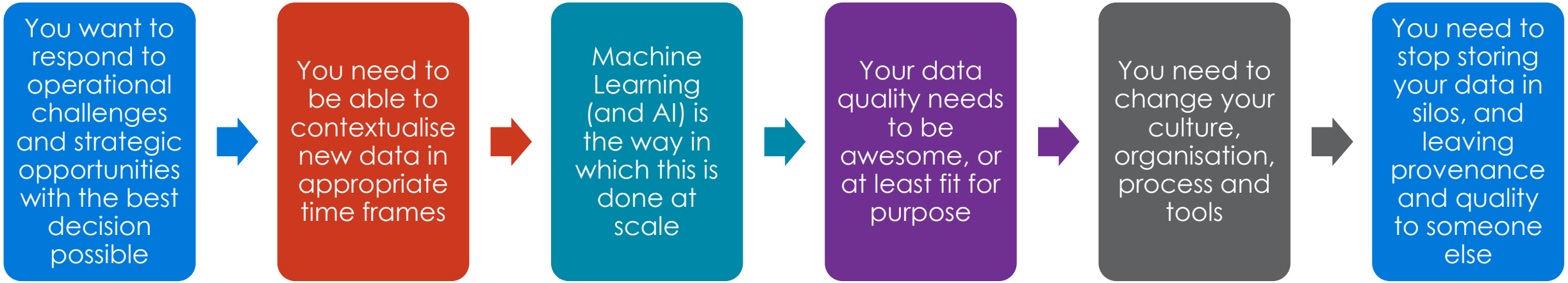
Yes! Data triage applied, with quality metrics at the measurement level

Build and populate a data lake architecture from a fresh AWS account

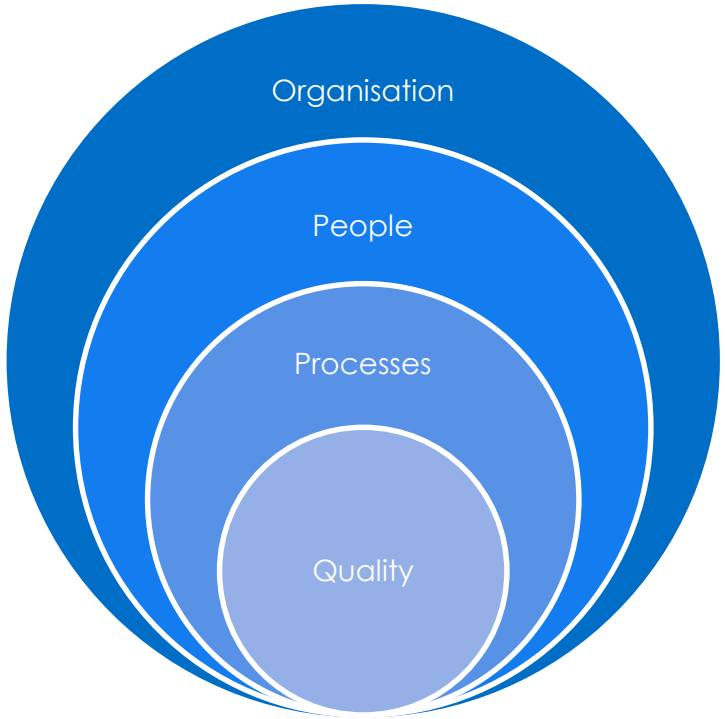


Putting ML and Quality back into the Digitalisation context

THINKBIG ANALYTICS
A TERADATA COMPANY



Delivering on a digitalisation vision



Organisation: Define and establish an Upstream Analytics CoE with a Data Governance board and Technical Design Authority to orchestrate and prioritise business-aligned data management. This should evolve into a Chief Data Office.

People: Train people in Agile methodologies to provide domain data leadership. Educate data domain and IT teams on how data management works in industries further down the digitalisation road. Align on concepts and vocabulary.

Processes: Data ingest and contextualisation needs close alignment with business requirements. Ensure strong governance of data management and utilisation capabilities

Quality: Drive data quality consistently across all petrotechnical domains to ensure a consistent framework of quality measurement, publication and continuous improvement. This will drive trust in data across an organisation and deliver on a digital vision.

Any questions?

THINKBIG ANALYTICS
A TERADATA COMPANY

You could buy a copy of this awesome book…