Big Data Analytics for the Upstream Domain
The art of the possible
Dr Duncan Irving, Principal Consultant, Teradata Oil and Gas Team
4th February, 2015
Key facts about Teradata

- **1 focus:** Value from Data!
  - Analytic data platforms, applications and services
- 30+ years of growth and innovation
- **1 patent** per week since 2009
- **1,500+ customers** in 12 industries
- **10,000+ employees** in 70 countries
- **$2.6 billion revenue in 2013**
- ~$7-8 billion market cap Q3 2014
- Constituent of the S&P 500 index

Leading companies trust Teradata for data management and analytics

Activities within predictive maintenance
Teradata's business model – key demarcations

**We do...**

- Integrate and analyze any type of data on our market leading HW/SW platforms
- Apply a horizontal view of data to release the value of breaking down silos
- Build analytical solutions in perspective of data reuse
- IT/BI service to enable our customers

**We do not...**

- **Sell** or advise on choice of Oil & Gas equipment
- Offer “silver bullet” pieces of software targeted for solving single business problems
- Offer packaged program solution “in a box”
Dealing with data in motion
Data Lake...
...or more like a reservoir
Many Lakes
What happens when your infrastructure needs a refresh?
Operational Data Store
A proliferation of data marts
Information supergarden
So what does Teradata do?

Put simply, our unified architecture advocates storage, discovery, operational decision support and event processing components.

Our clients use this conceptual architecture in part to understand how insight and value can be extracted from data assets across as many data and user domains as possible.

Each activity has a purpose and an envelope of cost-effectiveness; this extends to ease-of-integration.

Teradata can provide architectural consulting with the wealth of experience from our wide client base – and most of our clients operate very mixed architectures indeed.

Big Data are plural – and managing and exploiting them effectively is about AND, not OR.
How do we achieve insight today?

Analysis
How do we achieve insight today?

## Analysis

<table>
<thead>
<tr>
<th>Tag</th>
<th>Status Date</th>
<th>Year</th>
<th>Comments</th>
<th>Entered By</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-7001-A</td>
<td>9/30/2014</td>
<td>2014</td>
<td>2014 diesel x over valve us</td>
<td>James Eaton</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>9/30/2014</td>
<td>2014</td>
<td>2014 x over valve does not work only running on one diesel filter</td>
<td>Phil Roberts</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>9/27/2014</td>
<td>2014</td>
<td>2014 Diesel filter changed out</td>
<td>Phil Roberts</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>9/27/2014</td>
<td>2014</td>
<td>2014 x over valve does not work only running on one diesel filter</td>
<td>Phil Roberts</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>9/27/2014</td>
<td>2014</td>
<td>2014 x over valve does not work only running on one diesel filter</td>
<td>Phil Roberts</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>9/26/2014</td>
<td>2014</td>
<td>2014 Changed over to diesel</td>
<td>James Eaton</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>9/26/2014</td>
<td>2014</td>
<td>2014 x over valve does not work only running on one diesel filter</td>
<td>James Eaton</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>9/25/2014</td>
<td>2014</td>
<td>2014 Changed over to gas.</td>
<td>Phil Roberts</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>9/25/2014</td>
<td>2014</td>
<td>2014 x over valve does not work only running on one diesel filter</td>
<td>Phil Roberts</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>9/25/2014</td>
<td>2014</td>
<td>2014 x over valve does not work only running on one diesel filter</td>
<td>James Eaton</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>9/25/2014</td>
<td>2014</td>
<td>2014 failed to change over to fuel gas</td>
<td>James Eaton</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>9/24/2014</td>
<td>2014</td>
<td>2014 High DP on filter. looking into changing it over</td>
<td>James Eaton</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>9/22/2014</td>
<td>2014</td>
<td>2014 Genny manually changed over to diesel</td>
<td>Phil Roberts</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>9/1/2014</td>
<td>2014</td>
<td>2014</td>
<td>James Eaton</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>8/19/2014</td>
<td>2014</td>
<td>2014 changed over to fuel gas</td>
<td>James Eaton</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>8/19/2014</td>
<td>2014</td>
<td>2014 changed over to diesel</td>
<td>James Eaton</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>8/18/2014</td>
<td>2014</td>
<td>2014 Changed over to gas.</td>
<td>Phil Roberts</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>8/18/2014</td>
<td>2014</td>
<td>2014 Changed over to gas.</td>
<td>Phil Roberts</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>8/16/2014</td>
<td>2014</td>
<td>2014 Changed over to gas.</td>
<td>Phil Roberts</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>8/16/2014</td>
<td>2014</td>
<td>2014 Changed over to gas, not on load.</td>
<td>Phil Roberts</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>8/14/2014</td>
<td>2014</td>
<td>2014 Started on diesel 1st time</td>
<td>Phil Roberts</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>8/14/2014</td>
<td>2014</td>
<td>2014 Started on diesel 1st time</td>
<td>Phil Roberts</td>
</tr>
<tr>
<td>X-7001-A</td>
<td>8/14/2014</td>
<td>2014</td>
<td>2014 Started on diesel 1st time</td>
<td>Phil Roberts</td>
</tr>
</tbody>
</table>

© 2014 Teradata
How do we achieve insight today?
The answer: Data Integration

Cartoonist Hugh MacLeod nailed it with this cartoon:

There is a world of difference in the workplace between knowing facts, and knowing how those facts fit together. Even more important is knowing what to do about it.¹

Source: ¹ Marc Cenedella, Founder The Ladders
The Power of Integration: No Integration...

1+1=2

Limited Business Value

- Each data mart can provide answers to subject-specific questions
- With each new data mart, IT repeats its development efforts
- This includes sourcing data that already exists in another environment

Business Value
IT Development
The Power of Integration: With Data Integration...

1+1=3

More Business Value

- Combining environments requires less new work for each new subject
- Enables new insights that can’t be achieved with separate systems
The Power of Integration: More Data Integration...

1+1+1=6
Value to effort ratio increases significantly!
Data-centric approach to answer strategic questions in Field Monitoring operations

Selection of questions across 12 domains and 120+ data sources in ConocoPhillips

- What is the optimum well spacing to maximize reservoir drainage?
- What is the **optimum** number of frac stages per well, to maximize production at a reasonable cost?
- Can we **predict** production decrease based on the predictive maintenance plan and potential asset failures?
- Can we make **automated** adjustments to the wells, to optimize production, lower costs, and increase asset profitability?

“Business insight is extended”

- Understand drilling practices
- Validate history matching
- Modelled Reservoir Data
- Drilling and Wells Data
- Production Data
- Improve recovery factor
- Increase production
- Rapid and accurate well performance analysis

Understand drilling practices

Validate history matching

Subsurface Monitoring

Modelled Reservoir Data

Drilling and Wells Data

Production Data

Improve recovery factor

Increase production

Rapid and accurate well performance analysis

Business insight is extended
What would you like to see?
Trends, patterns, and risks in D&W domains and suggest optimal parameters for planning and operations
The power of becoming data-driven
Analytics at scale (descriptive to predictive)

The journey to becoming data-driven

Turning the discovery of potential relationships in your data into validated insight requires careful contextualisation – the joint assessment of operational data with expected behaviours from interpreted and modelled data sets (e.g. seismic, flow models).

Once an understanding of an underlying process is achieved, it can be built into operational decision support. Exceptions and advance signatures can be spotted, driving early warnings to catch issues before they become significant, allowing corrective action to be advised and scheduled.

To perform at this level of operationalization, real trust is paramount. This is driven not only through improved data quality, but also the clarity with which the relationship and its context is communicated.
Upstream Data Mining and Discovery Analytics

“The quickest way to find a needle in a haystack...is to burn the haystack” crunch all the data

Gary Class, Head of Digital Analytics, Wells Fargo
Statoil’s new **Big Data** problem

Permanent Reservoir Monitoring investment on the Snorre and Aasgard Fields

- $800M in seafloor cable
- 38 wells, 2 platforms
- LOF: 1992-2040
- 5th largest field and 3% of PRs
The Business Perspective

+442081505292

+44
4D Seismic – seeing what happened
Traditional Marine Seismic Surveys

Planning 1 year
Acquisition 1 month
Processing 6+ months
Interpretation 3-6 months

Average 2 years between “snapshots” of the reservoir
Permanent Reservoir Monitoring (PRM)

“Operationalizing” the workflow
With a fixed seabed receiver array:
• Simpler source vessel
  • Cheaper per survey
  • More weather independent
• Receiver geometry the same
  • Surveys are more repeatable
  • Faster processing turnaround
• Can do more frequent surveys
PRM: Shortens the time frame

- New survey at least every 6 months
- Decision making on the timescale of interventions
- Need a much more streamlined process for receiving new data and interpreting it
And brings so much data...

So many surveys, time shifts, differences between pairs, attributes to look at. And the governance..?
Our Solution
Learning from other industries...
We made a Reservoir Data Warehouse!

We store detailed subsurface data in an MPP Analytical database.

We integrate it in space and time as well as logical relationships.

And users can visualise detailed data and analysis, calculated on-the-fly.
to bring the disciplines together
We brought Analytics to the Subsurface

• 4D workflows not fully supported by today's tools
• Explaining 4D “effects” requires other data
  – Identifying artefacts of processing or acquisition
  – Identifying events that correlate
Example 1: Repeatability Analysis

- Good image of 4D effect requires seismic image taken ‘from the same position’.
- NRMS is measure for survey ‘repeatability’.

Good image of a 4D effect

Bad image of a 4D effect
Example 1: Repeatability Analysis

- **Bad repeatability**
  - Not great repeatability
  - OK repeatability

- **Source/receiver positions repeated well**
  - Distance repeated well, angle not
  - Source/receiver (base)
  - Source/receiver (monitor)

- **Source/receiver angle repeated well, distance not**
- **Offset x Azimuth deviation**
- **Base receiver station nr**

- **Offset deviation (m)**
  - Colored by base-monitor azimuth deviation (degree)
  - Colored by base-monitor source + receiver deviation (meters)

- **NRMS**
  - Good
  - Bad

Classification: Internal
2013-02-27
Example 2: Subsurface Analytics

- **Repeatability (NRMS)**
- **Time shift**
- **Pressure difference**
Example 2: Subsurface Analytics

The query is as follows:

```sql
SELECT
    when corr_type_nm = 'Time shift around horizon VS Reservoir model property between strat layers (on reservoir model grid)' AS Corr_Resolution,
    'TS_Volume' || TRIM(Timeshift_Dataset_id) AS Timeshift_Dataset,
    'ReservoirModel' || TRIM(Res_Model_ID) AS Res_Model_Nm, Res_Model_Cell_Prop_Desc AS Res_Model_Prop_Nm,
    TRIM(Extract(YEAR FROM Res_Model_Period_D1)) AS Year_Base, TRIM(Extract(YEAR FROM Res_Model_Period_D2)) AS Year_Monitor,
    'StratUnit' || TRIM(Strat_Unit_id1) AS Top_Strat_Unit_Nm, 'StratUnit' || TRIM(Strat_Unit_id2) AS Base_Strat_Unit_Nm,
    TimeShiftVsProp_Corr_Meas AS Correlation_TimeShiftVsProp,
FROM DD_AL_Corr_TS_VS_RM_DYNAMIC_PROP
ORDER BY Correlation_TimeShiftVsProp DESC
```

The SQL query and result set are shown, with the best correlation on top.

Further down the list...

**Time Shift vs Pressure**

**Time Shift vs Water Saturation**
Example 2: Subsurface Analytics

Clear correlation

Weak correlation
Yes, you can put detailed subsurface data into a relational database and do analytics

- If you use a High Performance Analytical Database
- If you **model** and integrate the data in time, space and logical relationships

Yes, new analytical workflows can change how we do business
Drilling Effectiveness Case Study
Drilling Efficiency/Safety

- Stuck Pipe = NPT = cost

- Why stuck?
  - Geology (link) e.g. swelling shales
  - Rock properties e.g. weak rocks
  - Deviation/deviated wells
  - Bit type
  - Mud type WBM vs OBM
  - Other

- If we can analyse the conditions causing stuck pipe we can reduce the risk/cost

- Pilot for Big Data Analytics partnership between Teradata and CGG
Bad Hole Example – Single Well

The completion log gives no clues to the problems encountered.
It is widely recognised that data quantities have ballooned and continue to do so:

**O&G Data is:**
- Seismic
- Well logs
- Formations tops
- Checkshot surveys
- Pressures
- Drilling data
- Core data
- Well test data
- Completions
- Production data
- Fluid data

UKCS Data
One Terabyte
200,000+ files
A lot of these connections are routine, check shots and seismic, fluids and pressures. Some of this data is used in combination in reservoir studies, seismic, well logs, formation tops, pressures, fluids, core data. However these are single instances, single wells or a field study.
Data Not Linked

- A lot of data types are not-linked or only linked occasionally. Why?
- Are all links equal or are some ridiculous?

- Sometimes new techniques are found by linking diverse data types for example
  - Seismic to Fluids is AVO
  - Seismic to pressures is overpressured zones
Data Visualisation Multi Well

- Visualisation of a number of parameters simultaneously.
Data Analysis

Analysis of the data gives correlations and probabilities.
Drilling NPT Case Study: integrating geomechanics and engineering data
More efficient development drilling
Fewer bit failures | Fewer Trips | Reduced Opex

**Goal:** consistently drill horizontal section in a single trip in hard formations

**As-is:** “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.

**To-be:** find combinations of a wide range of drilling parameters likely to avoid bit failure and model alarms to ensue efficient drilling insights.

**How?** look for patterns to that will inform better operational decisions: increase drilling efficiency to avoid catastrophic bit damage
What data was used?
Source data sets and derived properties

- surface and downhole drilling data
  - MWD/LWD time series
  - Logging notes
- metadata relating to well and drill string configuration
  - Wellview schema
  - CSD
- bit damage severity and profile
  - Synthetic scoring from IADC codes
- well position and trajectory
  - LAS, DLS and x,y,z trajectories
- petrophysical information
  - Formation strength, density, elastic moduli
- Operations data
  - Project logging (time allocations and costing) from ERP
**BIT DAMAGE TO ROCK HARDNESS**

- Calculating both a Bit Damage Score and Rock Hardness allows us to compare them looking for insights.

  Example, compare bit damage score to ‘total rock hardness’ of the bit run

<table>
<thead>
<tr>
<th>propno</th>
<th>bitrun</th>
<th>damage_score</th>
<th>total_rock_hardness</th>
</tr>
</thead>
<tbody>
<tr>
<td>649954</td>
<td>9</td>
<td>60</td>
<td>204</td>
</tr>
<tr>
<td>646195</td>
<td>5</td>
<td>60</td>
<td>339.6</td>
</tr>
<tr>
<td>648407</td>
<td>7</td>
<td>52.5</td>
<td>763</td>
</tr>
<tr>
<td>649954</td>
<td>5</td>
<td>50</td>
<td>120</td>
</tr>
<tr>
<td>649954</td>
<td>6</td>
<td>50</td>
<td>56</td>
</tr>
<tr>
<td>649556</td>
<td>4</td>
<td>50</td>
<td>490.8</td>
</tr>
<tr>
<td>649026</td>
<td>5</td>
<td>50</td>
<td>442.4</td>
</tr>
<tr>
<td>645305</td>
<td>5</td>
<td>50</td>
<td>240</td>
</tr>
<tr>
<td>645305</td>
<td>4</td>
<td>50</td>
<td>140</td>
</tr>
<tr>
<td>651595</td>
<td>6</td>
<td>48.75</td>
<td>580</td>
</tr>
<tr>
<td>647070</td>
<td>7</td>
<td>45</td>
<td>108</td>
</tr>
<tr>
<td>647070</td>
<td>8</td>
<td>45</td>
<td>216</td>
</tr>
<tr>
<td>649947</td>
<td>5</td>
<td>45</td>
<td>354</td>
</tr>
<tr>
<td>645986</td>
<td>7</td>
<td>45</td>
<td>104</td>
</tr>
<tr>
<td>642068</td>
<td>4</td>
<td>45</td>
<td>358</td>
</tr>
</tbody>
</table>

Some high bit damage scores are with low rock hardness scores.
BIT DAMAGE SCORES AND EFFICIENCY

Almost 80% of high efficiency runs have little damage.

Low efficiency runs dominate high damage scores.
### Event Correlations to Bit Damage Score

<table>
<thead>
<tr>
<th>Bit Damage Score &lt;= 25</th>
<th>Bit Damage Score &gt; 25</th>
<th>Bit Damage Score &gt; 45</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cor</strong></td>
<td><strong>value</strong></td>
<td><strong>cor</strong></td>
</tr>
<tr>
<td>damage_score:efficiency_session</td>
<td>11.373</td>
<td>damage_score:efficiency_session</td>
</tr>
<tr>
<td>damage_score:rpm_session_80</td>
<td>8.174</td>
<td>damage_score:rotary_energy</td>
</tr>
<tr>
<td>damage_score:weight_energy</td>
<td>7.406</td>
<td>damage_score:efficiency</td>
</tr>
<tr>
<td>damage_score:rpm</td>
<td>6.087</td>
<td>damage_score:total_energy_mjoules</td>
</tr>
<tr>
<td>damage_score:weight_swing</td>
<td>2.879</td>
<td>damage_score:drilling_seconds</td>
</tr>
<tr>
<td>damage_score:rotary_energy</td>
<td>2.731</td>
<td>damage_score:diff_session_1000</td>
</tr>
<tr>
<td>damage_score:torque</td>
<td>1.551</td>
<td>damage_score:torque</td>
</tr>
<tr>
<td>damage_score:wob</td>
<td>15.192</td>
<td>damage_score:drilling_seconds</td>
</tr>
<tr>
<td>damage_score:wob_session_19</td>
<td>13.857</td>
<td>damage_score:torque_session_12000</td>
</tr>
<tr>
<td>damage_score:torque_session_15000</td>
<td>11.357</td>
<td>damage_score:torque_session_15000</td>
</tr>
<tr>
<td>damage_score:mud_flow</td>
<td>9.151</td>
<td></td>
</tr>
<tr>
<td>damage_score:torque_session_17000</td>
<td>8.09</td>
<td></td>
</tr>
<tr>
<td>damage_score:wob_session_40</td>
<td>7.526</td>
<td></td>
</tr>
<tr>
<td>damage_score:rpm</td>
<td>6.489</td>
<td></td>
</tr>
<tr>
<td>damage_score:total_rock_strength</td>
<td>4.557</td>
<td></td>
</tr>
</tbody>
</table>

**Energy efficiency is a leading factor to bit damage**

**Total rock strength as a factor is almost zero**
TREE DIAGRAM SHOWING DOMINANT PATHS

WOB to Rotary Energy is early warning to Low Efficiency
EVENTS LEADING TO LOW EFFICIENCY

Another view of Event Clusters - values represent number of connections between events.
Putting insights into operation: integrating across all the domains
3,000+ Tag Types
487,000 data points
650 wells, 2,000 tanks
60 central facilities

Frequency: 5-30 min

INTEGRATED DATA WAREHOUSE

- new wells
- production decline
- completion status
- history matching
- shut ins, alarms
- H2S, PVT, FBHP

Growth: 40 GB/day
Size: 25 TB
Pre/Post Activity Tracker

• Analysis of production and pressure curves prior to and after a well event
  > e.g. well stimulation, equipment modification, scheduled maintenance, etc.
Well Scout Analysis

- View all integrated data for a single well
- Includes data from all sources
- Enables users to view
  - Well Header
  - Well Completion
  - Directional Survey
  - Pressure & Production Profiles
  - Well Activity Notes
  - Production Losses & Reasons
Infill Drilling Analysis

Identify wells within a certain vicinity of wells that will be fracked
Value

• 6.600 BOE/d production increase (5%)
• Unnecessary downtime avoided
• Increase reliability of downhole equip.

Enablers

• Identified optimal well spacing
• Near real time intervention & adjustment
• Accelerated improvement in reservoir management