Why is it so difficult to connect users to data?

Alan H Smith
Luchelan Limited
Agenda

• Introduction
• Changing technology
• Example - MultiClient seismic data
• Enabling technology
• Where next?
Agenda

• Introduction
• Changing technology
• Example - MultiClient seismic data
• Enabling technology
• Where next?
The author...

L. C. (Lee) Lawyer is chief geophysicist of Chevron Corp. A 1955 graduate of Oklahoma University, he joined Standard Oil Co. of Texas, forerunner of Chevron, as a client representative on a seismic crew. He has been division geophysicist for the Midcontinent and Alaskan divisions, chief geophysicist of Chevron Overseas, and vice-president of Chevron Geosciences Co.

After Tonstad, 2002
Where Does the Time Go?
The amount of time spent on various tasks by surveyed nonmanagers in data-science positions

- **Basic exploratory data analysis**
  - Less than 1 hour a week: 12%
  - 1 to 4 hours a week: 46%
  - 1 to 3 hours a day: 31%
  - 4 or more hours a day: 29%

- **Data cleaning**
  - Less than 1 hour a week: 7%
  - 1 to 4 hours a week: 31%
  - 1 to 3 hours a day: 19%
  - 4 or more hours a day: 42%

- **Machine learning, statistics**
  - Less than 1 hour a week: 10%
  - 1 to 4 hours a week: 34%
  - 1 to 3 hours a day: 27%
  - 4 or more hours a day: 29%

- **Creating visualizations**
  - Less than 1 hour a week: 7%
  - 1 to 4 hours a week: 23%
  - 1 to 3 hours a day: 29%
  - 4 or more hours a day: 41%

- **Presenting analysis**
  - Less than 1 hour a week: 6%
  - 1 to 4 hours a week: 20%
  - 1 to 3 hours a day: 27%
  - 4 or more hours a day: 47%

- **Extract, transform, load**
  - Less than 1 hour a week: 5%
  - 1 to 4 hours a week: 20%
  - 1 to 3 hours a day: 32%
  - 4 or more hours a day: 43%

---

1 Correcting or removing faulty data
2 Creating computer models
3 Also known as ETL — moving information to a data warehouse

Source: O'Reilly Media Inc. online survey of more than 600 data science professionals, conducted from November 2014 to July 2015

THE WALL STREET JOURNAL.
Agenda

- Introduction
- **Changing technology**
- Example - MultiClient seismic data
- Enabling technology
- Where next?
Technology in the 90s
Data delivery in 1990’s
Technology about 2005
Data delivery in 2000’s
Technology from ~2015
Agenda

• Introduction
• Changing technology
• Example - MultiClient seismic data
• Enabling technology
• Where next?
The statistics – PGS MultiClient

- Important Balance Sheet item
- Significant proportion of vessel time
- Significant revenues
- Huge data volumes acquired
- Long shelf life
- Shelf life “reset” with reprocessing etc
- Pre- and Post-Stack and Ancillary products all equally important for exploration purposes

Data from PGS Annual reports
PGS MultiClient Data delivery

• Heritage system developed in 1990s
  – Outsourced service provision
  – Slow & restricted functionality by current standards
  – Only really handled post stack
  – Barrier to exit very high

• New system
  – Still outsourced
  – Trace handling, not processing (*sensu stricto*)
  – Handles pre and post stack data efficiently
  – Modern database integrated with IT infrastructure and other enterprise software systems
Loading & QC

Compresses ‘snapshot’ format

Extracts

Loads SEG-Y attributes to Exploration Archives

Generates
Delivery
Multiclient complications

What are they getting?
- Prestack (options)
- Stack
- Migration
- Velocities
- ...

All need cutting to correct coordinates

Historic
Manual handling
Manual intervention

Now
Automatic
Parallel processing
The impact

Then

COOPS!

Faster – Cheaper – Better
Agenda

• Introduction
• Changing technology
• Example - MultiClient seismic data
• Enabling technology
• Where next?
The management software
Changing technology
Ensuring availability

Client → PGS Houston → Ovation Houston → PGS London → Ovation London → Client

10Gb connections:
- Client to PGS Houston
- PGS Houston to Ovation Houston
- Ovation Houston to PGS London
- PGS London to Ovation London
- Ovation London to Client
The technology

- Intel
- XEON inside
- EMC
- Isilon X
- IBM
- Palo Alto Networks
- Aspera
Agenda

• Introduction
• Changing technology
• Example - MultiClient seismic data
• Enabling technology
• Where next?
Leave the data where it is

What about latency?

What about bandwidth?

What about security?

Format “A”

Format “B”

Format “C”
Conclusions

• Large volumes of data can be “live” on the internet
• Still issues to tackle if it is to become a reality
• Some interpretation will be automated (rules based)
• Analytics will play a bigger part in interpretation
• It will come!
Acknowledgements
Thanks for your attention

Alan Smith
Luchelan Limited
alan.smith@luchelan.com
+44 7768 063042