



PETRONAS

Production Data Quality Management – Best Practices

Impiana Hotel KLCC

8 Oct 2013

Daljit Singh Dhaliwal

Senior Manager, Petroleum Engineering Data, Technical Data, PMU, PETRONAS

© 2013 PETROLIAM NASIONAL BERHAD (PETRONAS)

All rights reserved. No part of this document may be reproduced, stored in a retrieval system or transmitted in any form or by any means (electronic, mechanical, photocopying, recording or otherwise) without the permission of the copyright owner.



PETRONAS

Daljit Singh Dhaliwal

© PETRONAS 2013

Presentation Outline

1. Part 1

1. What is data quality and DQM?
2. Why data quality is so important
3. Impact of poor data quality
4. Examples of disaster due to data quality
5. Common production data errors
6. Reasons for poor production data quality
7. Challenges in production data quality management

2. Part 2

1. Production Data Quality Best Practices
2. In a nutshell

What is Data Quality ?

- Data are of high quality “if they are fit for their intended uses in operations, decision making and planning”
 - *J M Juran*
- Alternatively, the data are deemed as high quality if they correctly represent the real-world construct to which they refer.
- Data are of high quality “if the data is comprehensive, consistent, conforms to requirements, fit for purpose and timely”

- *Daljit Singh Dhaliwal*

What is Data Quality Management(DQM)?

- Data Quality Management entails the establishment and deployment of roles, responsibilities, policies and procedures concerning the acquisition, maintenance, dissemination and disposal of data.

“25% of Fortune 1000 companies are working with poor quality data.”

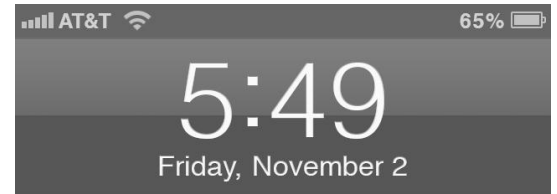
- *Garther*

Why data quality is so important

- “Almost everyone is impacted by poor data quality”
 - Agree or Disagree?

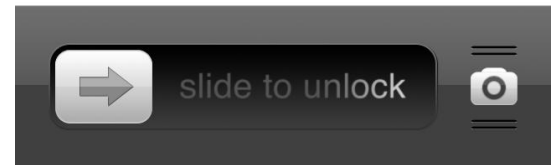
A husband, while he is on a business trip to a hill station sends an sms to his wife

“Bad Data are like viruses ... no way of knowing where they will turn up or the damage they cause.”



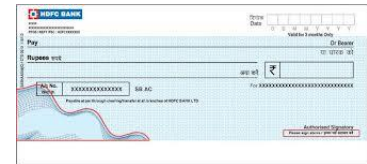
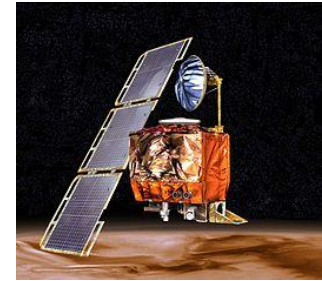
Having the most amazing and wonderful time, I wish you were her.

What!



Examples of disaster due to data quality ...

- Sept 1999, NASA Mars Climate Orbiter(MCO) broke up in the Martian atmosphere
Due to **errors in units used**, one team(NASA) used newton's while another team(Lockheed) used pound, its height above planet surface not the planned 150km but 60km.
- Scottish Widows
Sent out 1000 letters offering health cover – to people it knew had died, for example “Mrs Mary Dale (Deceased)”. She had died of cancer a year earlier, but the letter asked “If you become too ill to work, how would you cope financially?” Enclosed mock cheque also made out to “**Mrs Mary Dale (Deceased)**”.
- March 2013, UK National Health Service closed down a children heart surgery unit,
Data submitted shows twice as much children and baby died in the unit compared to anywhere else in UK. 11 days later the unit was reopened. Due to data malaise. **35% of expected data submitted was missing completely.**
- Potential cause of delays in FFR, FDP and simulation studies. Most simulation model become obsolete when first recommended well drilled.



GIGO → “Garbage in, Garbage out”

Impact of poor data quality ...

- Data is not fit for purpose it was intended
- Decrease in end user satisfaction. Loss of confidence.
- Impedes decision making
- Incorrect results
- Database unusable (excels popup)
- For a organization
 - Damage to reputation
 - Decrease in profits
 - (opportunity lost, cost escalate, sued, etc)
 - Cause business to cease

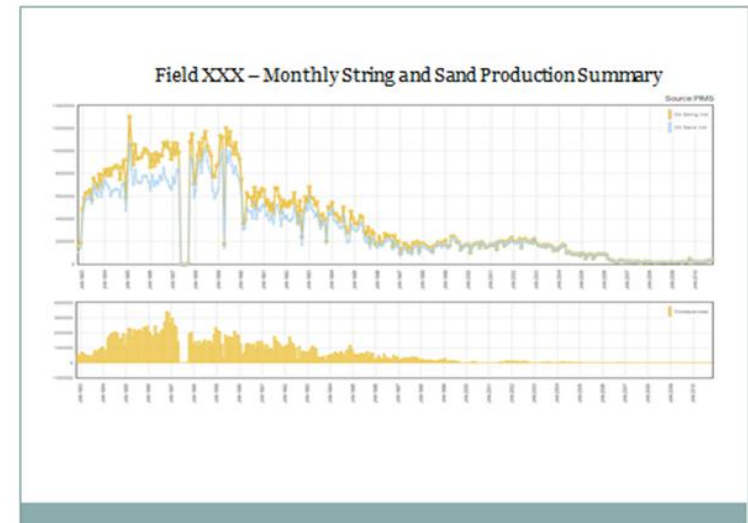
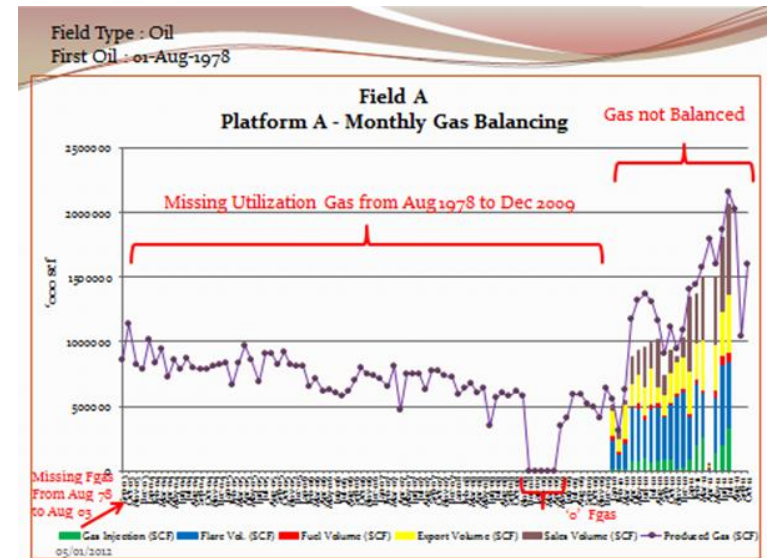


“Data Quality problems cost U.S. businesses \$600 billion each year.”




The Data Warehouse Institute(TDWI)

Common production data errors ...

- Data not fit for purpose it was intended
 - Split ratio, well status, etc not update regularly
 - Invalid strings exist
 - New strings not updated in Production database
- Data Completeness (data gaps)
 - Gas utilization data missing
 - Production volumes gaps
- Duplicate data
 - Same data exist in multiple databases
- Data Accuracy(data discrepancies)
 - Discrepancies between reconciled, string and sand production volumes
 - Discrepancies between OFM and Corporate Production database
 - Reconcile factors(RF) out of range
- Multiple Units
 - Prod volumes in multiple units



Challenges in production data quality management

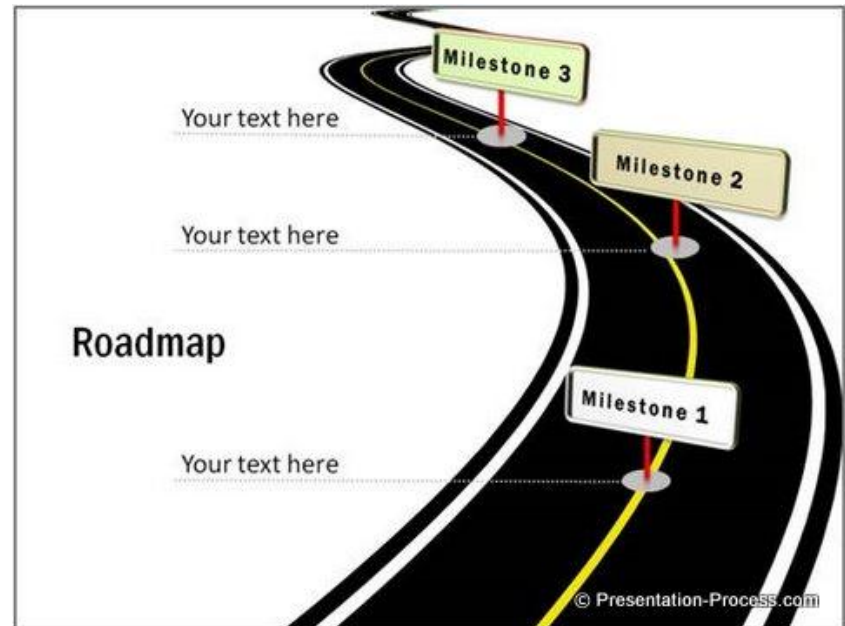
 Tools	 Process	 People
Multiple data silos exist within an organization. No data Integration.	Historical data migration from previous legacy system done without proper QC	Lack of awareness on data quality.
No data visualisation tool	No data audit conducted previously. Data errors detect late during FDP, FFR studies stage.	No dedicated data management team or limited resources responsible on DQM
Databases No Live link between Corporate DB with OFM.	No or poor enforcement of Data Ownership	User preference to use excel sheets resulting many data types in hard disks
No data up loader(ETL) tool available.	Unclear Roles & Responsibility	Management do not recognize significance of data quality. Data not treated as a business asset.
No data audit tool available	Inconsistent work process across regions	Requires discipline
	No Data Cleanup Initiative. User unwilling to participate.	Data quality is an IT Problem. (misconception)
		No cross-functional cooperation. Business users too busy with daily operations

Part 2

Production Data Quality Best Practices

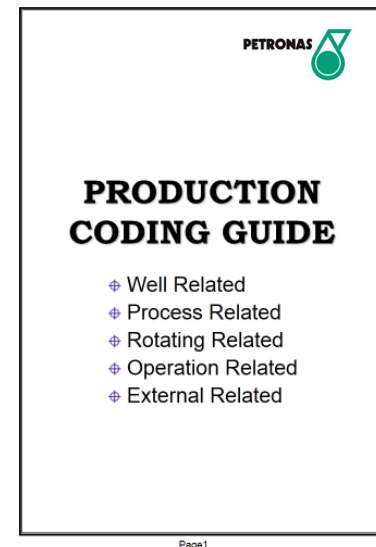
A Data Quality Strategy(Road Map) established

- PETRONAS's way of data quality management:



Policy, procedure and standards in place

- Policy and Procedure are in-place and practiced.
- Data Quality Standard properly defined and accepted by the data owners
- Data Standards need to be maintained across business process during the full data life cycle



Clearly defines Data Roles & Responsibility

- Data Ownership roles & responsibility need to be clearly defined.
- All data have an identified set of data owner in the business
- Data owners are accountable for all the data elements of the data type
- Effective DQM requires every organization to adopt a data stewardship approach
- Employ Data Stewardship matrix ("CRUD" matrices)

CRUD Matrix				
Entity \ Function	Reservation			
Make Booking	C			
Confirm Booking		RU		
Accept Cancellation			D	
Issue Tickets				RU

Create

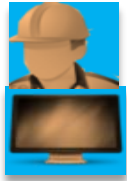
Read

Uppdate

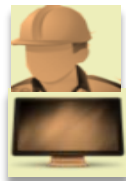
Delele

Dedicated Data Quality Team exist

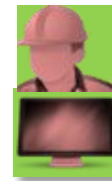
- A data quality team(s) is necessary to:
 - To ensure data audit is conducted regularly
 - Internal and External Audit conducted
 - Add value to data being collected
- A data quality team should be trained and have the right skills and experience
- Data Management skill group is essential for career progression. This will ensure long term sustainability



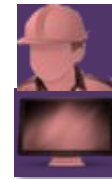
Data Technician



Data Analyst



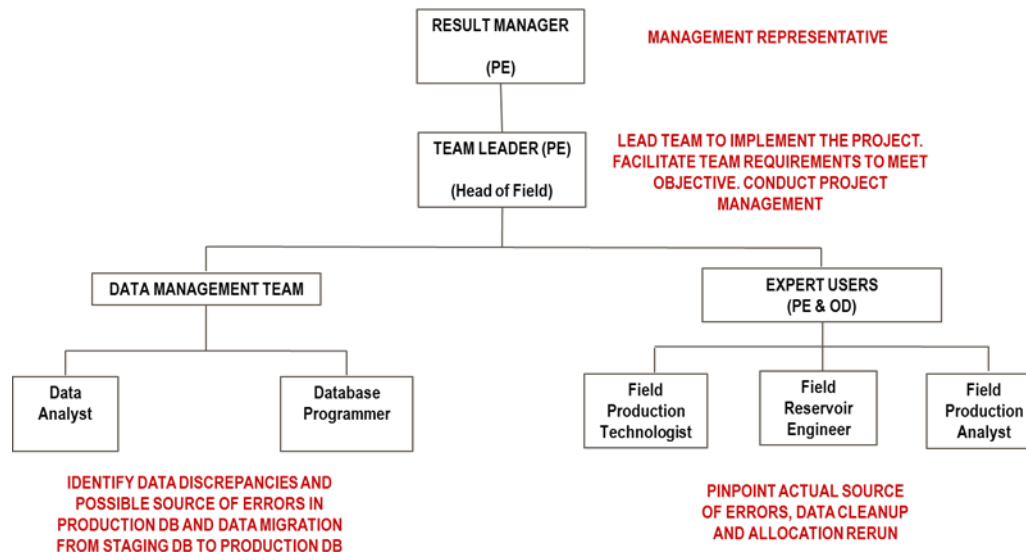
Data Scientist



Programmer


Partnership between data management and business unit is critical

- A partnership between business and data management is essential for any data quality management efforts to succeed
- Representative from different units must cooperate to resolve data issues
- Data owners involvement is high critical as they are responsible to assure correctness and accuracy of data.
- Business is also responsible to establish business rules for the data they own



Data quality tools to reduce effort and time

- Every organization ensure they have the right data quality tools to reduce manual and tedious process of data profiling and data cleansing.
- The data quality tool can be either:
 - In-house developed
 - PCSB Data Doctor
 - Purchased Off-the shelf product (Rule based toolkit)
 - IQM (Information Quality Metrics)
 - InnerLogix
 - DataVera
- An organization need to investigate the these available tool to decide most appropriate tool
- The data quality tool allows:
 - To easily conduct data quality audit
 - Visualization, which aids data quality
 - To conduct data cleanup
 - Considerable time saving

Summary Results by Query Set 

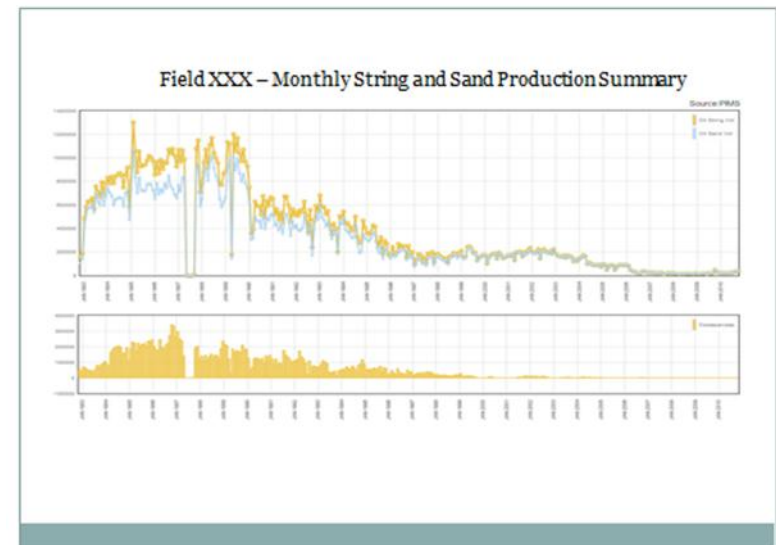
All Org Units

Query Set Quality %

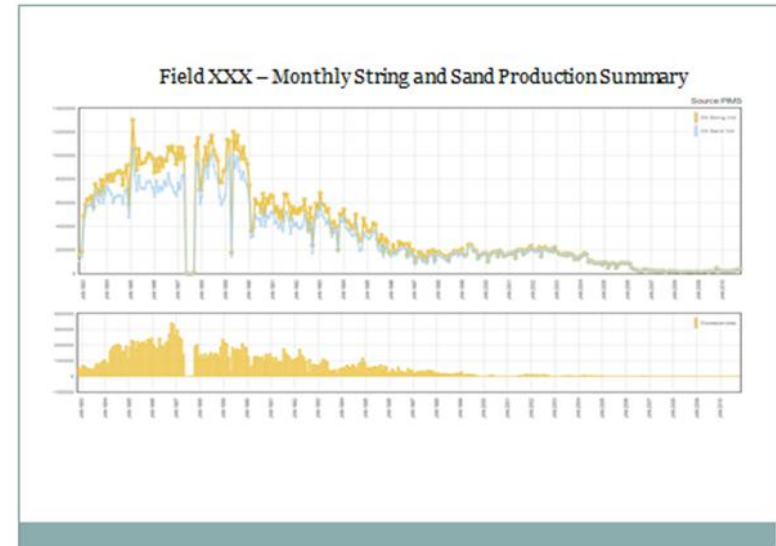
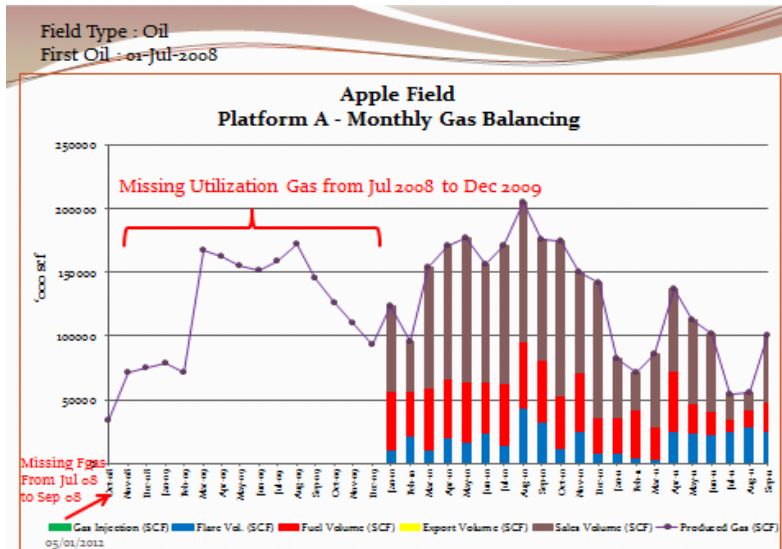
Type Results

The results columns show Total Objects, Total Errors, Quality %, Trends (Week, Month, Quarter). Trends are calculated back from the latest summary date 20-Mar-2008.

Query Set Name	Type	Lower Norm	Upper Norm	CNS	IRISH	NNS	SNS	WOS
→ Coordinate queries	discipline	80.0	90.0	192.0 53.0 72.4	15.0 6.0 60.0	12.0 3.0 75.0	206.0 56.0 72.8	8806.0 317.0 96.4
→ Geological	discipline	90.0	95.0	600.0 4.0 99.3		288.0 140.0 51.4	704.0 5.0 99.3	26300.8 4787.8 81.8
→ Petrophysical queries	discipline	80.0	90.0	1560.0 360.0 76.9	144.0 23.0 84.0	1616.0 363.0 77.5	89128.0 14696.0 83.5	
→ Seismic/horizon/fault queries	discipline	90.0	95.0	2840.0 418.0 85.3				
→ Well headers	discipline	90.0	95.0	307.0 72.0 76.5	18.0 3.0 83.3	333.0 81.0 75.7	19395.0 4098.0 78.9	
Totals				5499.0 907.0 83.5	15.0 6.0 60.0	462.0 169.0 63.4	2859.0 505.0 82.3	143629.0 23898.0 83.4



Data Visualisation aids data quality



BAYAN Field - Monthly Production Summary till 1 MAY 2011 - Sorted by Month

Data Audited by TIM, PCSB using PCSB Data Doctor(PIMS Module) - Ver. 1.0

Status Date: 24-JUN-2011 10:06:59

No	Field	Month	Oil Str Vol.	Oil Sand Vol.	Oil Vol Diff.	Status	Gas Str Vol.	Gas Sand Vol.	Gas Vol Diff.	Status	Condy Str Vol.	Condy Sand Vol.	Condy Vol Diff.	Status	Water Str Vol.	Water Sand Vol.	Water Vol Diff.	Status
1	BYAN	01-JUN-84	21197	18837	2360	X	26185000	23718000	2467000	X	0	0	0	O.K.	0	0	0	O.K.
2	BYAN	01-JUL-84	199156	182306	16850	X	130610000	119459000	11151000	X	0	0	0	O.K.	0	0	0	O.K.
3	BYAN	01-AUG-84	351275	308550	42725	X	226581000	201766000	24815000	X	0	0	0	O.K.	0	0	0	O.K.
4	BYAN	01-SEP-84	450067	379842	70225	X	302236000	263950000	38286000	X	0	0	125	O.K.	125	0	0	O.K.
5	BYAN	01-OCT-84	360275	301577	58698	X	203193000	171782000	31411000	X	0	0	417	O.K.	417	0	0	O.K.
6	BYAN	01-NOV-84	534956	444611	90345	X	320253000	263815000	56438000	X	0	0	0	O.K.	3430	3430	0	O.K.
7	BYAN	01-DEC-84	656050	536797	119253	X	366182000	296731000	69451000	X	0	0	0	O.K.	8864	8864	0	O.K.
8	BYAN	01-JAN-85	626463	510282	116181	X	331796000	264934000	66862000	X	0	0	0	O.K.	8064	8064	0	O.K.
9	BYAN	01-FEB-85	496073	407133	88940	X	282118000	230276000	51842000	X	0	0	0	O.K.	6908	6908	0	O.K.
10	BYAN	01-MAR-85	656138	545590	110548	X	366651000	312634000	54017000	X	0	0	0	O.K.	9270	9270	0	O.K.
11	BYAN	01-APR-85	633823	519891	113932	X	399740000	321968000	77772000	X	0	0	0	O.K.	6270	6270	0	O.K.
12	BYAN	01-MAY-85	597719	479261	118458	X	443642000	355090000	88552000	X	0	0	0	O.K.	8640	6899	1741	X
13	BYAN	01-JUN-85	684794	563298	121496	X	482426000	409047000	73379000	X	0	0	0	O.K.	16104	8183	7921	X
14	BYAN	01-JUL-85	592448	482512	109936	X	393300000	334204000	59096000	X	0	0	0	O.K.	10657	4799	5858	X
15	BYAN	01-AUG-85	552835	447104	105731	X	364025000	301095000	62930000	X	0	0	0	O.K.	0	0	0	O.K.
16	BYAN	01-SEP-85	479259	421005	58254	X	280347000	246982000	33365000	X	0	0	0	O.K.	4653	3887	766	X
17	BYAN	01-OCT-85	656243	532043	124200	X	370282000	308927000	61355000	X	0	0	0	O.K.	7682	7129	553	X

Eliminate data entry errors

- Data Validation rules:
 - Reduce data entry errors esp. during manual input
 - Automatically detect erroneous/questionable data
 - Alert users of potential low quality data
 - Easy navigation to such data
- Drop down menu:
- Use excel Unloader's:
 - Reduce data entry errors esp. during manual input

The screenshot shows the 'Energy Components' software interface. The 'Daily Oil Stream Status' window is open, displaying a table of Measured Figures and Derived Figures. The 'Production Operations' tree on the left shows a red 'X' icon next to 'Daily Oil Stream Status', indicating an error. A callout box highlights the 'Production Operations' and 'Daily Oil Stream Status' sections. Below the main window, a detailed view of the error message is shown, including the severity, date, message, and validation time.

Stream Name	Gra Vol [Sm³]	BS&W [%]	Net Vol [Sm³]
FPSO Export		0.00	
Process Train 1	33,009.0	1.00	32,678.9
Process Train 2	18,432.0	1.00	18,247.7

Stream Name	Gra Vol [Sm³]	Net Vol [Sm³]
Planned Oil To ONS-A Storage		
MonthNetOilProd		
FPSO Oil Inventory Difference in the label to this		
FacNetOilProd (daily)		

Severity	Daytime	Message	Validation time
ERROR	2003-01-09T00:00:00	Stream FPSO Export has negative or missing gross volume	2005-02-10T18:04:39

Conduct data audit(monitoring) regularly

“What can be measured, can be improved.”

- Data Quality Matrix(DQM) or Dashboard established
- A data audit is conducted to ensure data:
 - Completeness
 - Conformity
 - Consistency
 - Correctness (Accuracy)
 - Duplication
- A data audit can be conducted:
 - Internally
 - External
- A data audit conducted on regular basis:
 - Daily
 - Weekly
 - Monthly
 - Quarterly

[illegible]

The screenshot displays the Raven QI application window. The top menu bar includes File, Tools, View, Window, and Help. Below the menu is a toolbar with icons for Browser, Job Edit, Attribute Edit, Data Edit, Rename, Logpage, Logpage Plot, All Checks, Metadata, Completeness, Bulk Data, Child Quality, and Refresh.

The main interface is divided into several panes:

- Primary Keys:** Lists key information for the project:
 - Real Project Name: CANADA_STAGING
 - Field Name: FERPER
 - Well Name: PC FERPER 11-35-45-10
 - Well Api Number: 201123041200000
 - Log Name: HOB/DZL
 - Log Service: AC
 - Log Category: RAW
 - Log Title: PETRO-CANADA
 - Log Source: MAN1
 - Log Activity: 1.00 NO
 - Log Run: QRA
 - Log Version: 1
 - Curve Name: GAMMA
 - Curve Type: Curve Set 100
 - Curve Version: 1
- Attributes:** A table showing various attributes and their values:

Attribute	Value
Bottom Depth	2377.70
Business Value	HQH
Curve Units	GAPI
Depth Increment	10
Depth Limit	M
Core GSA Measurement	20.181369019
Maximum Value	9999.0
Top Depth	262.70
- Checks:** A table showing the status of various data quality checks:

Check	Meta	Complete	Children	OCB	Online
ECA CHAMBERS 1-33-40-11	Pass	Pass	Pass	Pass	Pass
DEVENUE/NEW/NEW/NEW	Pass	Pass	Pass	Pass	Pass
GOOD_FILES_LAS	Pass	Pass	Pass	Pass	Pass
PC FERPER 11-35-45-10	Pass	Pass	Pass	Pass	Pass
PC FERPER 11-35-45-10	Pass	Pass	Pass	Pass	Pass
AC	Pass	Pass	Pass	Pass	Pass
C13A	Pass	Pass	Pass	Pass	Pass
C13B	Pass	Pass	Pass	Pass	Pass
C13C	Pass	Pass	Pass	Pass	Pass
C24H	Pass	Pass	Pass	Pass	Pass
C24E	Pass	Pass	Pass	Pass	Pass
GAPI	Pass	Pass	Pass	Pass	Pass
CALH	Pass	Pass	Pass	Pass	Pass
CALZ	Pass	Pass	Pass	Pass	Pass
CH7A	Pass	Pass	Pass	Pass	Pass
CH7H	Pass	Pass	Pass	Pass	Pass
CH7Z	Pass	Pass	Pass	Pass	Pass
CN	Pass	Pass	Pass	Pass	Pass
CHC3	Pass	Pass	Pass	Pass	Pass
GR	Pass	Pass	Pass	Pass	Pass
GGA	Pass	Pass	Pass	Pass	Pass
GRH	Pass	Pass	Pass	Pass	Pass
GRZ	Pass	Pass	Pass	Pass	Pass
MGR1	Pass	Pass	Pass	Pass	Pass
MGR2	Pass	Pass	Pass	Pass	Pass
MGR3	Pass	Pass	Pass	Pass	Pass
MGR6	Pass	Pass	Pass	Pass	Pass
MGR9	Pass	Pass	Pass	Pass	Pass
MGR1	Pass	Pass	Pass	Pass	Pass
PE	Pass	Pass	Pass	Pass	Pass
POK2	Pass	Pass	Pass	Pass	Pass
SPA21	Pass	Pass	Pass	Pass	Pass
SPA22	Pass	Pass	Pass	Pass	Pass
- Metadata:** A table showing metadata information:

Test	Impact	Result
CK Meta Date	2010/04/21:09:53:14	
CK Meta File	HQH	
CK Meta Quid	0	
CK Meta Total	HQH/OCB	
CK Meta Curve Units	Minor	PASS
CK Meta Bottom Depth	Severe	ERROR: Target attribute missing: TO LOGGER
CK Meta Curve Units	Severe	PASS
CK Meta Top Depth	Severe	PASS
CK Meta Depth Units	Severe	PASS
CK Meta Depth Units	Severe	PASS
- Details:** A table showing details for the 'CK Meta Bad Null' test:

Key	Value
Test Name	SAD Null
Description	Check for legal null values
Result Attribute	CK Meta BAD NULL
Impact	FATAL
Result	FATAL
Message	Legal nulls: 9999
Minimum Value	9999.0

Production Data Cleanup is a continuous process

“Data Quality is a habit of Continuous Improvement”

- **Proactive Component → For Current Data**
 - Continuous Data Quality sustainability effort



2 PRONG
APPROACH

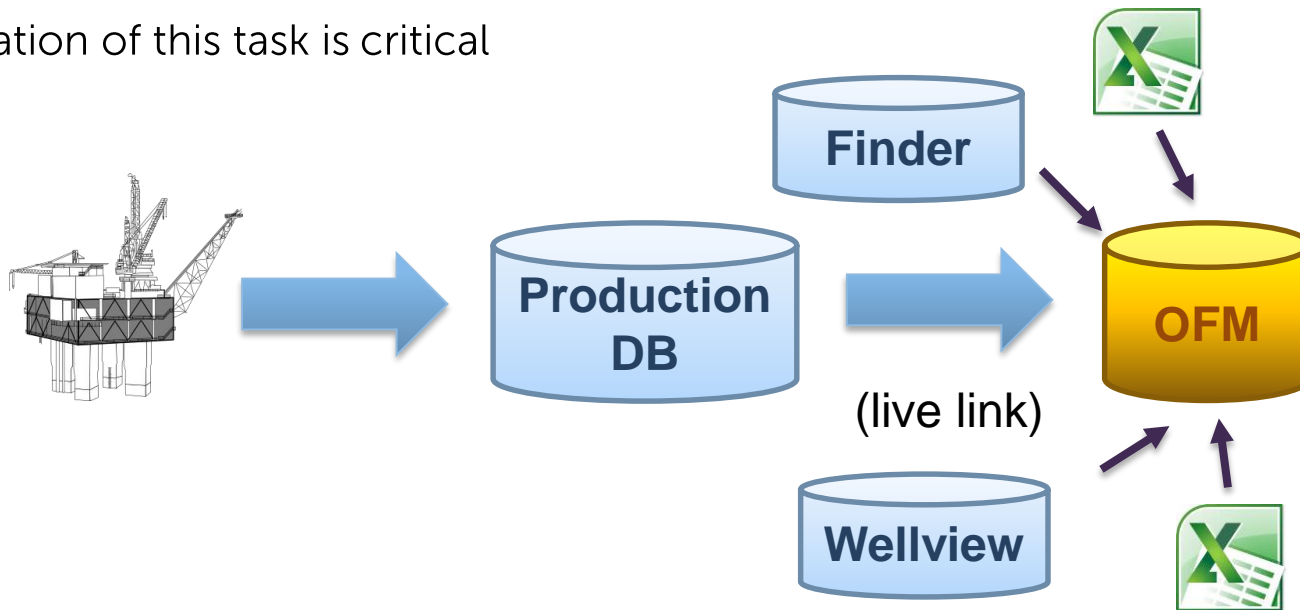


- **Reactive Component → Historical(Legacy) Data**
 - Conduct Historical Production Data Cleanup initiative from time to time
 - Dedicated team project team required
 - Business data owner involvement critical to conduct data cleanup



Data Integration for Better decisions

- Production data resides in many systems:
 - Excel based
 - In silo databases
- A single view of data(Production Dashboard) is recommended to ensure organization make better decisions
- Automation of this task is critical



Data Augmentation to add value(in-sight)

- Integration of production data with other data and 3rd party data for opportunity to augment the value of data so your organization can add more reserves or increase production
- “BIG Data” wave.



Production Data Quality Management – Best Practices

1. **Data Quality Strategy (Roadmap) established**
2. **Policy, Procedure and Standards in place**
3. **Clearly define data roles & responsibility**
4. **Partnership between data management and business unit is critical**
5. **Eliminate data entry error**
6. **Conduct Data Audit regularly**
7. **Data Quality tool reduce effort and time**
8. **Dedicated Data Quality team exist**
9. **Data visualization aids data quality**
10. **Production Data Cleanup Initiatives**
11. **Data Integration for better decision**
12. **Data Augmentation to add value (BIG Data concept)**

“As our organization, and the business uses for our data, continues to evolve, so must our data quality practice.” - Data Quality is a Best Habit.

Thank you.

For comments and feedback:

Daljit Singh Dhaliwal at dsukhdev@petronas.com.my

