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.
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.”
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 an 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

<table>
<thead>
<tr>
<th>Tools</th>
<th>Process</th>
<th>People</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple data silos exist within an organization. No data Integration.</td>
<td>Historical data migration from previous legacy system done without proper QC</td>
<td>Lack of awareness on data quality.</td>
</tr>
<tr>
<td>No data visualisation tool</td>
<td>No data audit conducted previously. Data errors detect late during FDP, FFR studies stage.</td>
<td>No dedicated data management team or limited resources responsible on DQM</td>
</tr>
<tr>
<td>Databases No Live link between Corporate DB with OFM.</td>
<td>No or poor enforcement of Data Ownership</td>
<td>User preference to use excel sheets resulting many data types in hard disks</td>
</tr>
<tr>
<td>No data up loader(ETL) tool available.</td>
<td>Unclear Roles &amp; Responsibility</td>
<td>Management do not recognize significance of data quality. Data not treated as a business asset.</td>
</tr>
<tr>
<td>No data audit tool available</td>
<td>Inconsistent work process across regions</td>
<td>Requires discipline</td>
</tr>
<tr>
<td>No Data Cleanup Initiative. User unwilling to participate.</td>
<td></td>
<td>Data quality is an IT Problem. (misconception)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No cross-functional cooperation. Business users too busy with daily operations</td>
</tr>
</tbody>
</table>
Part 2

Production Data Quality Best Practices
A Data Quality Strategy (Road Map) established

- PETRONAS’s way of data quality management:
  
  Data Profiling ➔ Data Quality Improvement ➔ Data Integration ➔ Data Augmentation

(Data Assessment) ➔ (Planning & Cleansing)
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)

```
<table>
<thead>
<tr>
<th>Entity</th>
<th>Reservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
<td></td>
</tr>
<tr>
<td>Make Booking</td>
<td>C</td>
</tr>
<tr>
<td>Confirm Booking</td>
<td>RU</td>
</tr>
<tr>
<td>Accept Cancellation</td>
<td>D</td>
</tr>
<tr>
<td>Issue Tickets</td>
<td>RU</td>
</tr>
</tbody>
</table>
```

© PETRONAS 2013
Dedicated Data Quality Team exist

- A data quality team(s) is necessary to:
  - To ensure data audit in 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
Data Visualisation aids data quality
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
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
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

• **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