ETL Verification: Do Your ETL Processes Do What You Think They Do?

Best Practices for Data Integration / ETL Testing Series

David Loshin, Industry Analyst

Praveen Radhakrishnan, Cognizant

Ash Parikh, Informatica

30 **Minutes with Industry Experts**

INFORMATICA®

1

Best Practices for Data Integration / ETL Testing – Expert Insight Series

With Industry Expert
David Loshin

- **Part 1 - Instituting Good Practices for Data Testing**
  David and Senthil Kanakarajan from Wells Fargo contrast the matured software quality and testing practices with the existing, mostly manual approaches to data integration testing.

- **Part 2 - Using Data Integration Testing for Reconciling Production Data Assets**
  David and Rob Myers from HealthNow NY explore how defined business rules and automated data validation can be used to identify unexpected production data results.

- **Part 3 - ETL Verification: Do Your ETL Processes Do What You Think They Do?**
  David and guest will review reasons for ETL process validity and reflect on the need for continuous monitoring of existing ETL activities to ensure there is no degradation in ETL reliability.

- **Part 4 - Managing Consistency and Trust During Data Migrations**
  David and guest will look at introducing automated methods for managing consistency of migrated data.

Got Unexpected Results?

Visa card surprise:
$23,148,855,308,184,500

Company cites 'temporary programming error' for 17-digit charges

Below: Text Discuss Related

NBC NEWS

A Texas man is accidentally charged $23 quadrillion for dinner. KXAS reports Ellen Goldberg explains.

Jon R. Seale

($23,148,85...

($23,148,855...

XXXX...

---

18 January 2012 Last updated at 08:06 ET

Indian teacher's shock at his $9.8bn bank balance

By Amitabha Bhattasali
BBC News, Calcutta

An Indian teacher got the shock of his life when he checked his bank account online and found 490bn rupees ($9.8bn).

Parijat Saha expected his balance to be $200 and could not believe his eyes - but a check at an ATM confirmed his billionaire status, on paper at least.

The honest Mr Saha rang the State Bank of India (SBI) to point out the error.

Red-faced bank officials would not comment, but sources say the funds were "uncleared" and he could not have withdrawn the money if he had tried.

Jan 16, 2012 2:57am

Man in Stable Condition After $44 Million Hospital Bill

Like 504   Tweet 166   +1 6   14   Text

ABC News' Ben Waldron reports:

A Bronx man says he nearly had an asthma attack after opening his mail to find a $44 million medical bill from a local hospital.

Unemployed doorman Alexis Rodriguez, 28, received the astronomical bill after receiving successful treatment for pneumonia at Bronx-Lebanon Hospital last week, The New York Daily News reports.

"I almost had an asthma attack," said Rodriguez, who apparently is not the only patient to receive an outsized bill. The firm responsible for the botched billing, PHY Services, was reportedly inundated with complaints and has since apologized.

"If you are calling with respect the billing statem... Bronx-Lebanon Hospital, please disregar... ...ncerned callers, "you will be...

This is Just the "TIP" of the Iceberg…

Why ETL Processes Fail…

1. **Bad upstream data in the environment**

2. **Failed jobs that were not rerun properly**

3. **Logic errors in integration system**

4. **Logic errors in related processing systems**

5. **"Rogue" updates bypass change control process**

6. **Data and environments are dynamic**

It Matters to Monitor ETL Activities!!!

- *Incorrect data flowing inside / outside your company*

- *Poor or incorrect decisions affecting the business*

- *Lost business opportunities in a competitive market*

- *Money wasted on inappropriate initiatives and projects*

- *Customer dissatisfaction and degrading experience*

- *Exposure to regulatory non-compliance risks*

The Fix?

- ☐ **SQL scripts**

- ☐ **Spreadsheets**

- ☐ **Hand coding**

- ☐ **Testing tool**

- ☐ **Partial effort**

The Challenges with Data Testing…

Data validation testing takes about 30% of time and effort spent on Data Integration projects

Currently testing is done manually with SQL scripts and Spreadsheets and partially

There typically is not time to validate data when doing nightly production updates

Solution

A way to identify **ERRORS** in data sets that have been **MOVED** or **TRANSFORMED** to ensure they are **COMPLETE** and **ACCURATE** and meet **EXPECTATIONS** or **REQUIREMENTS**.

What is the Ideal Solution?

✓**AUTOMATION**

✓**REPEATABILITY**

✓**AUDITABILITY**

How Can Informatica Help?

Informatica's Data Validation Solution (DVO)
Ensures the integrity of data as it moves through the IT environment...

Development & Test	Production Reconciliation
Provide automation for unit and regression testing of integration logic.	Protect the integrity of data that is loaded into production systems.
Ensure that data produced by DI code **meets requirements** and expectations	Erroneous data due to failed loads, faulty logic or operational issues is caught in a **proactive automated** manner and can be addressed as needed

INFORMATICA®

11

Benefits with Informatica

- **High project success/low project risk**

- **More cost savings/faster time to market**
  - *50% source-to-target testing*
  - *80-90% regression / upgrade testing*

- **Test all data, not just a small sample**

- **Test in heterogeneous environments**

- **No need to write SQL**

- **Complete audit trail of all testing**

- **No additional server technology**

Summary

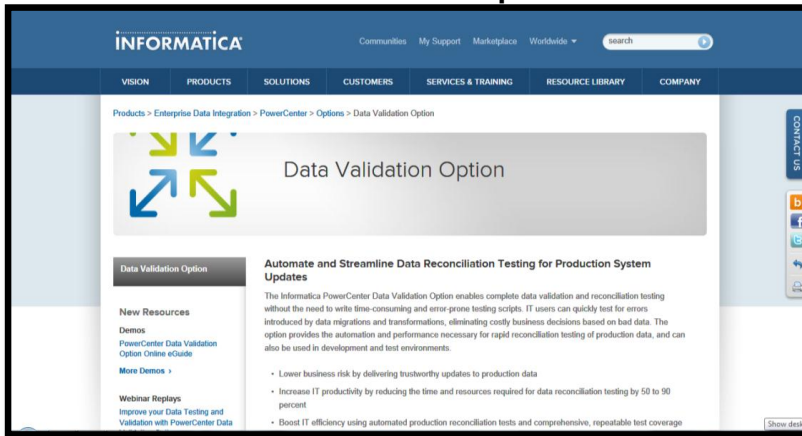| REQUIREMENTS	MANUAL TESTING	INFORMATICA
Test large volumes of data, not just selected rows/columns (i.e. increased test coverage)	🔴	🟢
Easily test heterogeneous data	🔴	🟢
No need to write SQL	🔴	🟢
Easily deal with data definition changes	🔴	🟢
Easily reconcile production data	🔴	🟢
Full audit trail of your testing activity	🟡	🟢
Increased likelihood of success. Reduced project risk	🟡	🟢

*"We used **Informatica** to **compare 14 tables & appx. 30 million rows in less than 5 hours**. The largest of the tables was 94 columns. When I asked our QA people how long it would take them to run the scripts and test this amount of data, they mentioned **months**…"*
- Customer

INFORMATICA®

Next Steps…

Informatica.com > Products > Enterprise Data Integration > PowerCenter > Options

**On-Demand
Product Desk Webinar:**
Best Practices for DI/ETL
Testing Presentation & Demo
Click on link in browser

NEW

**PowerCenter
Productivity Pack***

**Proactive Monitoring
for PC Governance**

**Proactive Monitoring
For PC Operations**

**PowerCenter Data Validation
Option (DVO)**

***Single Package** at an **Attractive Price**

NEXT UP IN THE SERIES:
Managing Consistency and Trust During Data
Migrations
PART 4

nextgendi@informatica.com