



# DAMA MICHIGAN Bits & Bytes



Michigan Chapter of DAMA International

Fall 2006

## Critical Data Quality Controls

By Denise Jeffries, Enterprise Information Management Team, R.L. Polk

Separate sets of processing checks must be used to find and control critical data quality issues. First, check your work stream jobs in-line and at the job level. These checks are for the most crucial issues, they are checks that are handled directly as you process the data so they take up time within the data processing cycle. Catching problems with the data here will be of most benefit to the organization. These controls are to keep you from publishing erroneous/incomplete data or having to back-out and re-run. Secondly, use audit controls and a monitoring and control process that are run after the data is processed. This can be done in batch or pseudo-batch mode. Finally, do overview data reporting, use KPI and scoring mechanisms to publish data quality information for consumption by the business.

### In-line Job controls:

Start with checking your source data. Check at the file level. This information should be collected during your data capture/receipt function and does not involve having to open up the file. Did you receive all the source files you were expecting? Did you receive too many files? (You may want to flag to check for potential duplicates) Are the file sizes within their normal byte range?

Check for file type (What is the file type? comma delimited, spreadsheet, plain text, XML, etc.). Now we can go ahead and open the file. Within a source did you receive the amount of data you were expecting?

*(Continued on page 3)*

### The Presidents Corner By Sid Stoffer

As data people, some would say that at times we deliberate way too much. At this time of year, I marvel at the beautiful colors and other fall activities such as the harvesting of crops such as apples or sugar beets. All of these activities



relate to data collection, data analysis and/or data deliberation. Questions come to mind such as how many leaves are there in Michigan? What percentage are fire engine red? What triggers the color change in a particular tree on someone's lawn or field? None of this is particularly useful. But that never stopped any of our brains from asking

*(Continued on page 2)*

*We hope by now we have attracted your attention. In the following pages, we will be providing details about the next event, the location, speaker's bios, and more.*

*We hope you will attend. It is a wonderful change to listen to experts in our industry who are making significant contributions. See you there!*

*Editor*

## Corporate Sponsors



(Continued from page 1)

the questions. The same brains also ask questions like is the apple crop as good in Michigan as last year? How much change is there? What is the economic impact? What is the average size of a sugar beet from the "Thumb" area? How does the sugar content compare with sugar cane from Alabama? How many tons of sugar beets are processed in Michigan? There is some time to reflect on the little details and more time to enjoy the season with our families. Take time to do both this fall.

## Fall Event Announcement

The Michigan Chapter of DAMA International, [www.dama-michigan.org](http://www.dama-michigan.org), is pleased to announce our Fall 2006 event. The event will be held on November 14, 2006. The speaker will be Jennie Dulac, RN, MS, Senior Director, Quality Improvement from Spectrum Health, [www.spectrum-health.org](http://www.spectrum-health.org), about their experiences with Data Quality.

We are looking forward to seeing you at the Fairfield Inn, 2335 Woodlake Dr., Okemos, MI 48864. Registration begins at 12:30pm. For more information and to Pre-Register please visit our web site.

Sincerely,

The Michigan Chapter

## DAMA Day Report

By Guest Columnist, Denise Jeffries, Enterprise Information Management Team, R.L. Polk

Dear Reader,

We had an exciting and successful DAMA Day on June 22, 2006. The campus of Lawrence Technological University (LTU) in Southfield Michigan was a great spot to host our bi-annual feature event. Their facilities provided a comfortable, professional environment for our meeting with over 135 participants setting a new attendance record. Our sponsors Informatica, LTU, IBM and Jaros were on hand with product information and representatives. Moreover, their sponsorship provided the means for DAMA to bring in two of the pre-eminent data experts of our day, Bill Inmon, Inmon & Associates and John Zachman, Zachman International. They both enlivened the event with their perspectives on Advanced Analytics. Their expertise was intertwined with discussions with local experts Dr. Michael Cavaretta, Ford Motor Company, Tommy Drummond, Informatica, Clifford Hodges, General Motors, Paul Scheibal, Jaros Technologies, and Bob Zurek, IBM.

## *Bit & Bytes* is Looking for Contributing Writers

If you have something to say and a talent for saying it, please contact the editor of Bits & Bytes. In our continuing effort to provide our readers with informative, educational and entertaining material we would welcome your contribution to our publication. The work doesn't pay much but we'll put your name in 10 point bold at the beginning of your article.

Looking forward to hearing from you.  
*Ed*

Bill Inmon's Corporate Information Factory (CIF) framework and his new DW 2.0 blueprint discussion challenged us to think about information architecture in innovative ways while applying tried and true methodologies for full understanding, we can explore his concepts further at <http://www.inmongif.com/>. John Zachman challenged and informed us with the notion that the Framework provides the critical interrogatives to conduct the

(Continued on page 3)

## Check Out the DMA Michigan Web Site Links to:

*Events*

*Data Management*

*Feature Articles*

*Membership*

*Contact Us*

*Members only Section*

*DAMA International*

*Bits & Bytes Newsletter*

<http://www.dama-michigan.org>

**For Web Site Information  
Contact:**

Vice President -  
Online Communications  
Jerry E. Stembridge  
EDS (248) 823-5937  
[jerry.stembridge@eds.com](mailto:jerry.stembridge@eds.com)

Advanced Analytical work; follow up information is available at <http://www.zachmaninternational.com/Default.htm>. Dr. Cavaretta capped off discussion with the importance of Advanced Analytics in automotive research, while Clifford Hodges's charismatic presentation included stating the need for vendors to provide consistent information throughout all GM departmental units. Important points on data representation for business understanding were made by all.

Vernon Hoffner  
Ph.D., CCP,  
CDMP

hoffner@ltu.edu



Dr. Hoffner is absent this issue. He will return in future releases.

Did you receive duplicate files? Is the file formatted as expected? Look at the important "\*" fields within each file. Did you receive the information with the expected values (In a State field do you have state values: AL, MI, CA, etc.? Are the numeric fields really numeric, the date fields dates, etc.)? Are there entire fields not populated? Are they null or blank? Check to see if there are important values missing (i.e. if you are expecting the State field to include all 50 states – are some missing?). Do the fields conform to expected values?

**Job level controls:**

Check to make sure the number of records received is the number of records processed for each file. Check for how long it takes to process each file. Next take measurements. These measurements occur as the data is being processed and are set based on business rules and Service Level Agreements (SLA) for data acceptance. Thresholds are set which allow the data to automatically process unless the thresholds are exceeded, at which point manual inspection and analysis occurs. This information is collected and stored as metadata and is available for inspection by Audit Controls, the next step in the process.

**Audit controls:**

Quality checks allow an organization to look at data after completion of the data processing cycle. Data can be inspected in combination with data previously published. Trending, anomalies and factual information about published data is made available to the organization through this process. The audit provides a check against the input data and the processing of the data and can uncover source data issues as well as processing data issues. Audit data also makes information available to the organization about the data as a whole which can lead to efforts to improve

## Useful Web Links

### Data Warehouse Institute

[www.dw-institute.com](http://www.dw-institute.com)

### DM Review

[www.dmreview.com](http://www.dmreview.com)

### Bill Inmon

[www.inmoneif.com](http://www.inmoneif.com)

### Intelligent Enterprise

[www.intelligententerprise.com](http://www.intelligententerprise.com)

### Ralph Kimball

[www.ralphkimball.com](http://www.ralphkimball.com)

### The Data Administration Newsletter

[www.tdan.com](http://www.tdan.com)

### Zachman Framework

[www.zifa.com](http://www.zifa.com)



source data and show the need for new data sourcing. Audit data makes great Key Performance Indicator (KPI) information for the organization.

### **Monitoring and Control:**

You monitor to control, there is no point in monitoring just for the sake of monitoring. Focus on capturing and dealing with the collected metadata from the in-stream and audit information. Monitor and control should be the next steps implemented by an organization as it matures in its data quality efforts. The emphasis here is on process improvement. At each of the points where data quality metrics are collected monitoring and control processes should be implemented with checks against the data.

**Data Profiling Checks:** Remember you inline job controls, use the rules you used to put these in place for your data profiling checks. This is about measurement and data quality control based on the input sources previously discussed. Since these are the raw inputs to the data production process, it is vital that the data quality of input sources be as high as possible.

**Data Processing Defined Checks:** Receiving and converting the input sources needs processing checks. Look at your steps of extraction, transformation, staging and loading operations typical in ETL efforts, and don't forget the staging for areas to place process checks. During any one of these operations, defects and errors can be introduced due to dropped records, interrupted or aborted processing and logic exceptions. Process checks within a set of rules and thresholds to find and report these types of errors.

**Business Defined Checks:** Look to your data users to endorse the data. They can give you some of the best measures for data quality by giving you business rules to look at in defining checks. Users know what the business needs from the data, rely on user and business defined metrics to measure the

quality of the data. Look at the transformation steps in your ETL efforts and you will typically find areas to asking about user-defined checks. The steps where data is transformed to form new data elements are important areas of consideration, you didn't implement these transform for nothing.

### **Overview Data Reporting**

Report and publish interesting information about the corporation's data. This can lead to new knowledge regarding corporate and customer information. This also includes overview reporting of metadata, which can be beneficial to

show anomalies and trends in data quality.

Do scoring. Get an analysts help to decide what type of information is pertinent to score – we all know about credit scores, it works the same way on other data points. This is about measuring other pertinent categories of information to assess the quality of the data. This information can be made available as supplemental KPI information as the organization matures. This is a more advanced practice and is typically implemented only as knowledge of the data grows by measuring and analyzing collected metadata. You



***Written and edited by Matt Roush, Technology Editor  
WWJ News radio 950, Detroit  
mnroush@cbs.com***

**Start your day with the most informative,  
easy to read News Report on  
IT happenings in Michigan.  
Direct to your desktop five days a week.**

**Free subscriptions available for a limited time.**

**Comments? News tips?  
Want a free subscription for a friend?  
Email mnroush@cbs.com.**

**Marketing or advertising queries contact  
Wendy Baca, wjbaca@cbs.com**

may need to tap into business analytical teams to help you build scoring algorithms to implement within the data quality process.

Another way to look at this is to *analyze and act* on the information you have collected. Again you do not analyze in isolation, you analyze to act so you can control your processes. Steps in approaching how to analyze and act are straightforward.

We can see that implementing data quality control processes needs to be well thought out and considered along with the other efforts surrounding our data. We easily acknowledge the ETL processes; we need to implemented

data quality control processes along with the typical data processing being done within the organization. Quality metrics allow an organization to prove the value of its data, to the end data consumers and report creators internal to the organization as well as to the external customers of the organization. The value of reporting out on your crucial data quality controls will directly effect the corporate bottom line by helping retain your current customers and offer them new solutions as well as enhancing your corporate reputation which will allow you to bring in new customers. Crucial data quality controls are for the enterprise as a whole, not just IT.

## Point of View

By W. Thomas Hamlin , EIS

On occasion when I get a chance to leave the "project trenches" and get a breath of fresh air someone will ask me why I like some models over others. At my present assignment, this is really a political question about taking sides.

Well, the side I am on is pragmatism. My view of models supports my pragmatism. That is, I like simple and robust models. They are durable, apply in my situation, and provide meaningful results without the necessity of a Ph.D. That is the model output is understandable by the intended audience. Audience understanding is a precursor to acceptance. Acceptance leads to change. Examples are  $E=MC^2$  and the Zachman Framework for Enterprise Architecture. I would encourage everyone to judge their models by these standards.

## Membership Report

### To the prospective member,

For more information about becoming a corporate sponsor, corporate member, or individual member, please refer to our membership web page via our chapter site at: <http://www.dama-michigan.org>.

Sincerely,

Dave Rath  
VP of Membership

## 2006 Board of Directors

### President, Sid Stoffer,

EDS,  
Phone: (248) 696-3118,  
email: [sid.stoffer@eds.com](mailto:sid.stoffer@eds.com)

### Admin VP, Vernon Hoffner, Ph.D.,

Lawrence Technological University,  
Phone: (248) 204-3058,  
email: [hoffner@ltu.edu](mailto:hoffner@ltu.edu)

### VP of Programs, Nanette Yuzon,

EDS  
Phone: (248) 265-6426,  
email: [nanette.yuzon@eds.com](mailto:nanette.yuzon@eds.com)

### Treasurer, John Mieczkowski,

EDS,  
Phone: (586) 575-4588,  
email: [john.mieczkowski@eds.com](mailto:john.mieczkowski@eds.com)

### VP Marketing, W. Thomas Hamlin,

Extensible Information Systems (EIS)  
Phone: (517) 887-3087,  
email: [thomas@thomahamlin.com](mailto:thomas@thomahamlin.com)

### VP of Logistics, Teresa Sun,

Ford Motor Company,  
email: [teresatsun@yahoo.com](mailto:teresatsun@yahoo.com)

### VP of Membership, Dave Rath

EDS,  
Phone: (248) 952-2781,  
Email: [david.rath@eds.com](mailto:david.rath@eds.com)

### VP of On-Line Services,

**Jerry Stemberidge,**  
EDS,  
Phone: (248) 823-5937,  
email: [jerry.stemberidge@eds.com](mailto:jerry.stemberidge@eds.com)