Unprecedented Performance and Scalability Demonstrated For Meter Data Management:

Ten Million Meters
Scalable to One Hundred Million Meters
For Five Billion Daily Meter Readings

Performance testing results of AMT-SYBEX Affinity Meterflow application using IBM Informix TimeSeries software

IBM and AMT-SYBEX testing has demonstrated the capability of the IBM® Informix® TimeSeries software to enable the Affinity Meterflow™ application to offer linear scalability up to 100 million meters to load and process meter data at 30-minute intervals in less than 8 hours. The time required for 10 million meters took less than 36 minutes.

In August 2011, IBM and AMT-SYBEX, an IBM Business Partner and provider of high performance energy data processing solutions for energy and utilities companies across Europe, collaborated to perform a benchmark test of Affinity Meterflow (formerly Smart DTS), AMT-SYBEX's solution for Meter Data Management (MDM). The Affinity Meterflow application uses IBM Informix TimeSeries software to manage smart meter data. The benchmark was performed at the IBM Power Systems Benchmark Center in Montpellier, France, on a single IBM POWER7 System, utilizing 16 cores. *

Understanding the significance of this benchmark may help energy and utilities organizations and MDM solution providers to:

- Deploy best-in-class meter data management solutions
- Dramatically reduce storage and system costs
- Accelerate time-to-value gained from smart meter data
- Reduce risk of smart meter deployments
- Help protect growing smart meter investments

September 2011

IBM

TABLE OF CONTENTS

1. Executive Summary

The results of this benchmark demonstrate the substantial business value that use of Informix TimeSeries combined with a proven MDM application can deliver to the Energy and Utilities Industry. We invite the reader to compare these results with other published benchmark results.

- *Faster processing of meter data means faster business processes that use that data!*
    - Achieved daily billing calculations of 21,000 bills per minute while concurrently running the data load.

- *Significant reduction in storage requirements.*
    - Total storage required for 1 month of interval and register data for 100M meters was less than 4 terabytes.
    - Average throughput exceeded 420,000 records per second in meter data load using standard storage disks.

- *Consistent, scalable performance helps yield highly predictable costs.*
    - Daily end-to-end processing times remained constant for 100M meters over a 31 day period, irrespective of the amount of data stored.
    - Storage requirements remained linear over time.

These results are enabled by the unique technology of the Informix TimeSeries software used together with a proven MDM solution such as Affinity Meterflow. Unlike traditional relational databases, Informix TimeSeries technology consolidates and organizes time-stamped (interval) data in a way that requires less storage space and significantly improves data load and query times.

In summary, the benchmark results show that Informix TimeSeries software is able to break the bottleneck of massive meter data management and deliver significant business benefits to utility organizations, as well as to ISVs offering packaged MDM solutions, such as Affinity Meterflow from AMT-SYBEX.

The Value of this Benchmark

Data is increasingly the lifeblood of every industry. A deluge of digitized information can create huge disruptions and risk if not anticipated and effectively managed. Consider the communications industry forced to accelerate huge infrastructure investments as their systems and networks strained under the data pressure created by an avalanche of smart devices and new data services.

Enter the smart meter. As the cornerstone of Smart Grid initiatives and Advanced Metering Infrastructure (AMI), smart meters promise to deliver significant benefits to energy and utilities organizations and their consumers. Smart meters fundamentally change how and when energy usage data is collected. The result is a tsunami of data that must be processed, transferred, stored and analysed, creating unforeseen technical challenges and business risks for which many utility organizations are not prepared.

As AMI initiatives increase, how does a buyer compare one MDM solution to another? Armed with the knowledge that MDM will create data storage and performance challenges, how does a buyer know the amount of systems and storage that will be needed or whether service level agreement goals have been met? Answers to these questions are determined by many factors including the database system that sits at the core of all MDM solutions, whether packaged or home-grown. Does it matter which data management solution is used?

The answer is decidedly *yes*.

Providers of MDM solutions understand that faster processing of meter data enables faster business processes that utilize that data (such as meter-to-cash) and faster data analysis. With a database system as the enabling technology, performance is affected by the speed, efficiency, and scalability of that database system.

IBM has conducted several Proof of Concepts (PoCs) with a number of companies using IBM Informix TimeSeries as the enabling technology used in the customer's own MDM applications. In each PoC, the IBM Informix TimeSeries solution consistently outperformed the customer's existing solutions, at times enabling up to 50 to 70 times faster processing of meter data, while requiring as little as 30 percent of the storage.

This benchmark was conceived and conducted to illustrate these benefits in tangible ways. It tested the preparation, loading, validation, estimation, and editing (VEE) of meter data for a 10 million meter utility, as well as a "day-in-the-life" scenario for a 100 million meter utility. If you compare the performance of Affinity Meterflow with other MDM solution offerings that do not use IBM Informix TimeSeries software, we believe you will conclude that these performance and scalability results are exceptional.

2. The IT Challenge of Smart Meters

The avalanche of data generated by smart meters creates new IT challenges for Energy and Utilities organizations. There are new storage requirements for massive amounts of data, as well as performance requirements that may be tied to service level agreements (SLA) which create a need to measure performance, such as data load times and query processing times.

Most Meter Data Management (MDM) solutions, whether packaged or "home grown", utilize a traditional relational database. What is frequently not understood is that, with such a choice, implicit trade-offs are made: One must either optimize how fast data is loaded, OR optimize how fast data is queried. Both benefits are not possible with a traditional relational database. This trade-off occurs because in order to get good query performance, traditional relational databases must store data multiple times using complex partitioning schemes. Storing data multiple times increases storage costs and, inevitably, load times. Further, queries take increasingly longer to complete as the database grows. The result is that load times increase, queries take longer, and storage costs can increase exponentially.

IBM Informix data management software offers a unique and unmatched approach to managing interval data, thanks to its native TimeSeries capability. While maintaining all the standard relational features and functions of other database software, Informix also has specialized access methods that are built for managing time-stamped data, commonly referred to as interval data or time series data. Today, Informix TimeSeries technology is unique because it stores the time-stamped meter data grouped by meter ID and sorted in time-stamp order, rather than storing the data as a collection of rows.

The advantages of using the Informix TimeSeries technology to manage meter data include:

- Dramatically less storage space required
- Much faster data loading
- Faster queries against meter data
- Simplified SQL function generation

3. Benchmark Setup

The MDM application that uses Informix Time Series software for this benchmark was Affinity Meterflow by AMT-SYBEX. The benchmark was performed at the IBM Benchmark Center in Montpellier, France, conducted by performance engineers from both IBM and AMT-SYBEX.
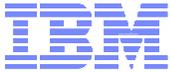
Hardware and Software Configurations

The commercially available hardware consisted of a single IBM Power7 server and an attached standard storage device:

CPU
- 1 - IBM Power® P750 32 cores (3.5GHz) running AIX® 7.1 (using 16 cores)
- 1 - 1GB LAN fiber adapter (dual port) using 1 port
- 2 - 8GB FC adapters (dual port) using 4 ports

Storage

- 1 - IBM XIV Storage System
- 15 – 2 TB data modules (only 4 TB of storage used for one month of data) consisting of a total of 180 physical disks.
- 6 - interface modules with 6 FC connections at 4 GB

A benchmark system configuration diagram is shown in **Figure 1**.
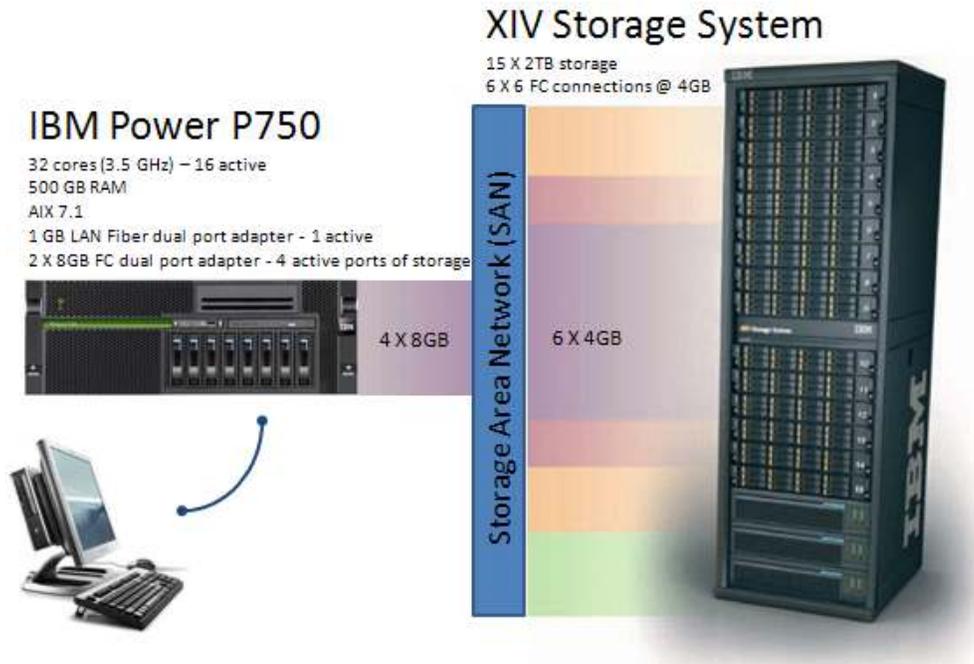
Figure 1 – Benchmark system configuration

The off-the-shelf software consists of an IBM Informix database server and the AMT-SYBEX MDM software:

- IBM Informix V11.70.xC3
- AMT-SYBEX Affinity Meterflow

4. Benchmark Measurements and Testing

Benchmark Measurements

1. Measure processing times for data collected from a "typical" 10 million meter utility at 30-minute intervals.

2. Measure processing times for data collected from 100 million meters at 30-minute intervals in a 24 hour time period.

3. Measure processing times and storage requirements for 31 days of data gathered from 100 million meters.

4. Measure one day's billing cycle while simultaneously processing and loading meter data for 100 million meters.

The benchmark tested the following processes:

1. Generate and perform technical verification of meter data.
2. Load verified data into an Informix data server.
3. Perform validation, estimation, and editing (VEE) processes on the data.
4. Run billing-type queries and calculations for 6% of the one 100 million meters using one month of data, concurrent with the data load process.

Data Generation and Verification

Test data was generated by the AMT-SYBEX Affinity Meterflow application, which is capable of generating different percentages and types of errors. 5.94% of the records contained simulated erroneous data at an error rate that, as we understand it, is approximately 3 times the common average of 1% - 2%. This error rate was chosen so that a comprehensive set of validation functions would be exercised during the VEE phase of the benchmark test.

 Two types of data were generated: Interval data consisting of 48 records per day, and register data consisting of one record per register per day with a mix of single and multi-register meters in the dataset. Interval data contained a time stamp, an interval flag, and an interval reading. The register data contained a register flag field, a register reading field, and a register ID field. The interval and register data were combined into one record.

The final step was to verify that the meter data and metadata were translated into standard formats. The result of the technical verification was a collection of scrubbed data files. The verification step split the single file containing input data into sixty verified output files for parallel processing.

Data Load

After the data was generated and verified, it was loaded with 16 parallel load processes into multiple physical disks

Validation, Estimation, and Editing (VEE)

VEE processing performs comprehensive checks on the data. These checks relate to the context in which the data was received. The comprehensive validation checks on the meter data, of which 5.94% was simulated to contain erroneous data, include, but are not limited to, the following checks:
- Check for spike, sum, consumption gap and overlap and missing values
- Consumption profile – high, low usage limits
- Check for : Test mode, pulse overflow, time change, meter diagnostic, reverse energy check, data collection, data estimation, AMI error
- Check for valid, energized meter
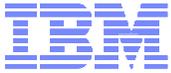
Any data that fails validation is marked for estimation and is passed to an estimation routine. The estimation routine attempts to populate the missing or invalid interval data. For the interval data, Affinity Meterflow uses estimation rules and estimation logic to determine the estimated value. For register data, estimation is based on device's historical data, historic profiles for the device category or by interpolation and following the estimation phase, Affinity Meterflow implements the changes identified, while also retaining the original data for a full audit trail.

Billing Determinant Calculation

Affinity Meterflow has the ability to create complex billing determinant calculations based on the specific requirements of the utility. During validation this processing is carried out against all meters where more than one time of use (TOU) period has been configured and based on local requirements this data can be presented for billing.

Figure 2 displays the processing flow and software architecture that occurs on a single IBM Power7 system for 100 million smart meters at 30 minute intervals.
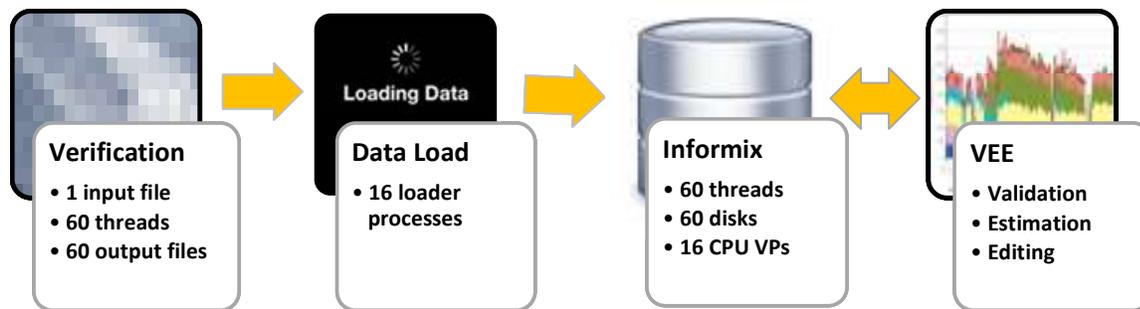

Figure 2 - Software architecture and flow for processing meter data

The input file contained one day of meter interval data read at 30-minute intervals, and daily register reads for each register for 100 million meters, for a total of 48 interval records and one record for each register on the meter per day per meter. The verification phase ran sixty threads in parallel to each other to scrub the input data. Sixteen load processes ran in parallel, one for each physical CPU, to load the data into the Informix data server. Upon completion, VEE processing runs 60 threads in parallel to validate and estimate the loaded data. The billing queries (not shown in **Figure 2**) ran in parallel with both the data load and VEE processes.

5. Benchmark Results

The benchmark testing was designed and carried out to accomplish all the benchmark measurements specified in section 4. The initial test was conducted for 10 million meters for one day's meter data.

10 Million Meters: Single Test Run

Table 1 shows the results for the test run using data for 10 million meters.

Process	Total Elapsed Time	Throughput Rate
Preparation and Technical Verification	10 min 02 sec	797,342 records/sec
Data Load	13 min 56 sec	457,162 records/sec
Validation, Estimation, and Editing	11 min 18 sec	707,964 records/sec

Table 1 –Total elapsed time for processing data from 10 million meters

The total end to end processing time for 10 million meters was measured to be less than 36 minutes.

100 Million Meters: Results for One Day (Day-In-The-Life)

Table 2 shows the average results from the benchmark testing. The daily data for 100 million meters at 30-minute intervals was loaded for one month (31 days) and the average load and VEE time was calculated.

Process	Avg. Elapsed Time	Avg. Throughput Rate
Preparation and Technical Verification	2 hrs 10 min	628,205 records/sec
Data Load	3 hrs 14 min	420,962 records/sec
Validation, Estimation, and Editing	2 hrs 11 min	623,409 records/sec

Table 2 – Average elapsed time for processing data from 100 million meters

As shown in **Figure 3**, the processing rate remained steady for the duration of the data load process irrespective of the amount of data loaded. During the data load the number of interval and register records that were written to the physical disk was measured.
The testing resulted in a load rate that ranged from 300,000 records per second to 539,000 records per second, with an average load rate of 420,000 records per second The Y-axis represents the number of meters processed based on an average load rate of 420,000 records per second.
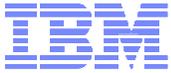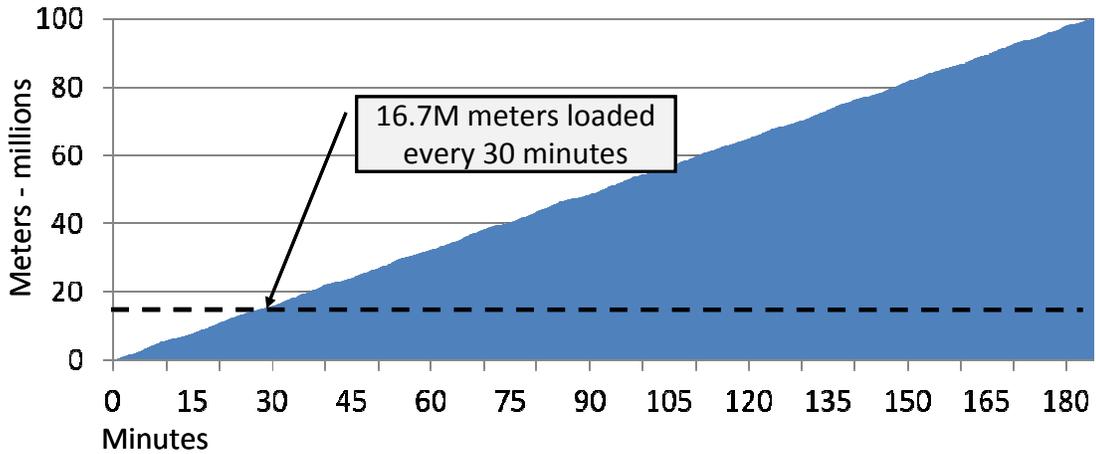
Cumulative Load Rate for One Day

16.7M meters loaded every 30 minutes

Figure 3 –Data load results for 100 million meters for one day

Figure 4 shows the VEE processing performance for one day of data from 100 million meters. The testing resulted in a VEE processing rate that ranged from 609,000 records per second to 736,000 records per second. The average VEE rate was 623, 000 per second. The Y-axis represents the calculated number of meters processed based on an average VEE processing rate of 623,000 records per second. The VEE processing rate remained steady irrespective of the amount of meter data processed.

VEE Processing Rate for One Day

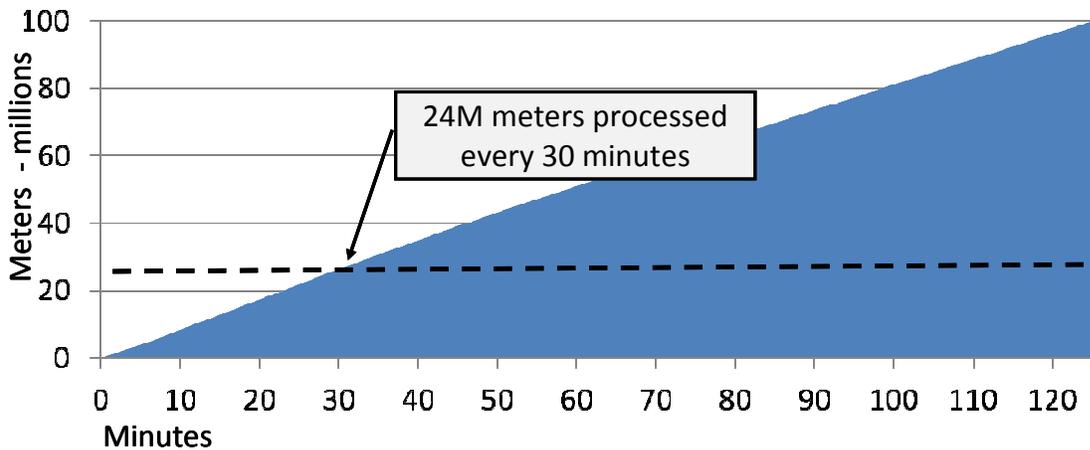24M meters processed every 30 minutes

Figure 4 – VEE results for 100 million meters for one day

Billing calculation was run each day on 6 million meters (6% of the 100 million meters). This calculation was performed in parallel to the data load and finished in less than 5 hours with a throughput of 21,000 billing-type calculations per minute.

100 Million Meters: Results for 31 Days

The benchmark measured the performance and storage requirements for one month of interval and register data for 100 million meters. The total number of records processed for 31 days of data was almost 152 billion records. The results are presented in **Figure 5**.
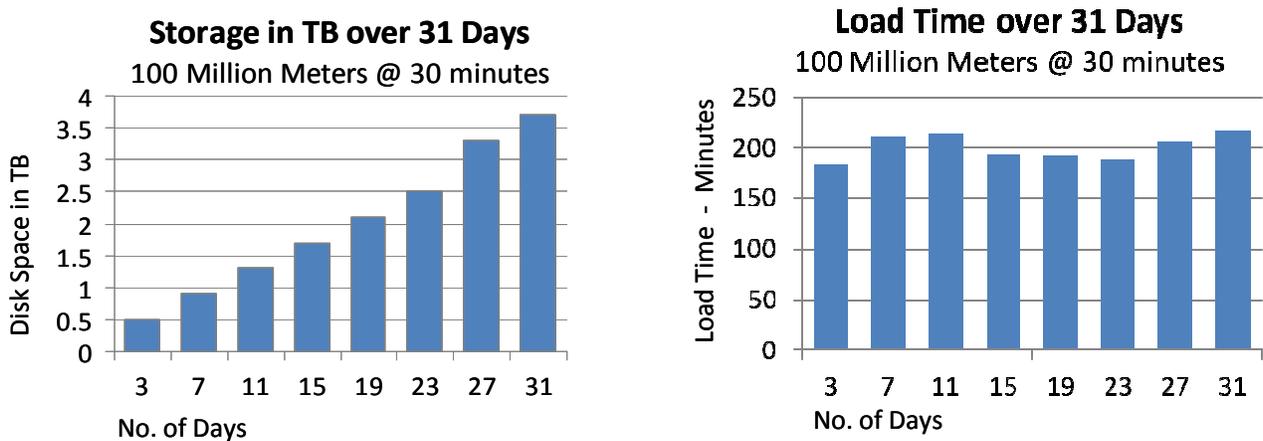

Figure 5 - Disk space consumption and load times over 31 days

As shown on the left, data storage requirements over 31 days is directly proportional to the amount of meter data. This linear relationship makes it easy to estimate the disk requirements for any size of implementation. The total storage used was 3.7 TB.

As shown on the right, data load times over the one month period remained relatively consistent, irrespective of the amount of historical data that was stored.

Total Processing Time: 100 Million Meters over 31 Days (One Month)

End-to-end processing times were measured for each day of the month for 31 days, consistent with the goal to process data for 100 million meters within the stipulated 8 hour window, even as the amount of data grew. As shown in **Figure 6**, the processing time remained fairly steady regardless of the amount of historical data that was stored. The total average daily processing time was 7 hours and 35 minutes.
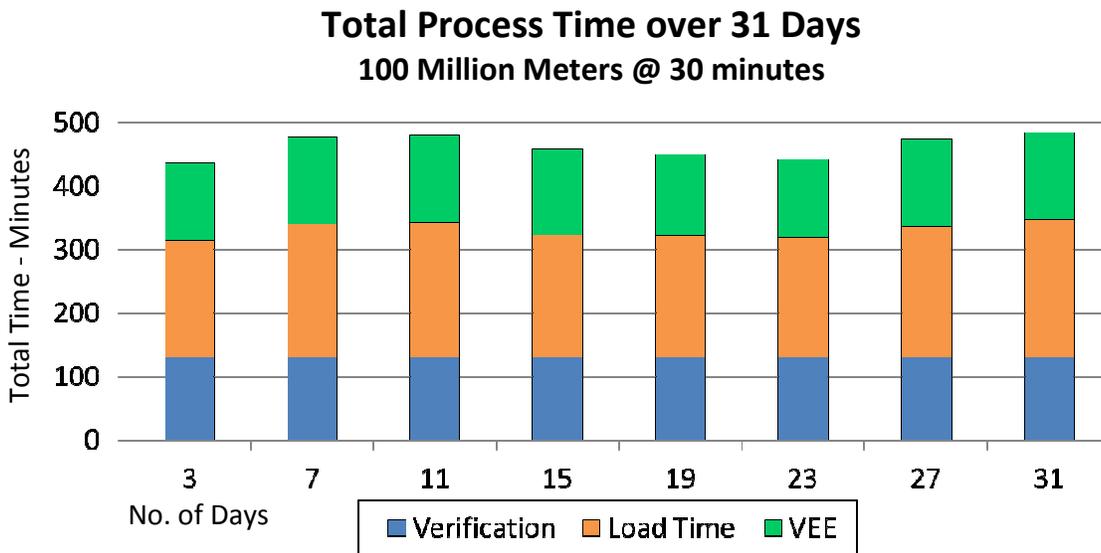

Figure 6 – End to end processing time for 100 million meters over one month

6. Summary and Conclusions

This benchmark demonstrates that IBM Informix TimeSeries and AMT-SYBEX Affinity Meterflow solution can perform and scale well to process data generated by 10 million smart meters, at 30-minute intervals, in under 36 minutes.

This performance benchmark also demonstrates that the same solution can perform and scale well to process data generated by 100 million smart meters, at 30-minute intervals, within 8 hours, and that a billing cycle can be completed on the same system simultaneously.

The data processed for 100 million meters required less than 4 TB of standard disk storage for one month of data. If you compare, we believe that you will find that this is but a fraction of the storage required for many competitive solutions.

The benchmark processed 4.9 billion records for one day and 152 billion records over the 31 day period.
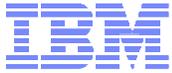
The benchmark demonstrates that all the processing can be done using a highly affordable combination of commercially available hardware, storage, and software.

This single-server solution is architecturally simple, easily managed and maintained, and can greatly reduce your resource consumption, leaving any additional resources available for further processing of the meter data or for other tasks.

7. Reference Information

About AMT-SYBEX:
http://www.amtsybex.com

About IBM Informix:
http://www.ibm.com/informix/

About the IBM Informix TimeSeries solution:
Managing Time Series Data with Informix, article in July 2011 issue of Database Trends and Applications:
http://bit.ly/ManagingTSdataWithInformix

What's new in TimeSeries data for Informix 11.7:
http://bit.ly/IfxTimeSeries1170Docs

About IBM Power7 systems:
http://www.ibm.com/systems/power/