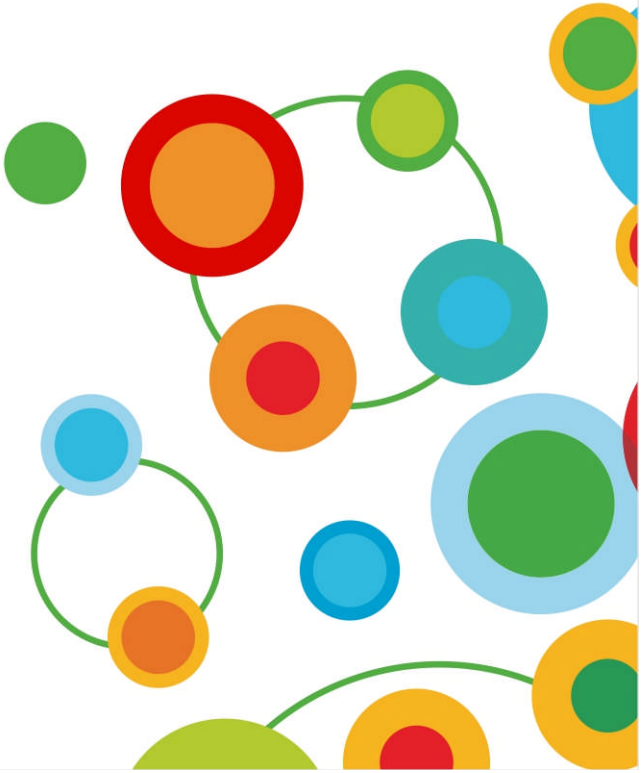
IBM Software

Information Management & Analytics Forum 2013

Return on Information: The New ROI

Time Series Data Management with unprecedented performance



Agenda

- What is Time Series
- Traditional Relational Approach
- Challenges
- IBM Informix TimeSeries
- Using Informix TimeSeries Schema
- Virtual Table Interface – Relational View
- Real Customer Comparision TimeSeries Vs. Relational
- Benchmark results
- Reference Links

What is Time Series

Time series data is a sequence of data points, measured typically at successive times spaced at uniform time intervals. Basically, it is a set of data where each reading element or row is time-stamped.

Who Generates Time Series Data?

- **Capital Markets**
  Arbitrage opportunities, breakout signals, risk/return optimization, portfolio management, Value at Risk (VaR) calculations, simulations, back testing
- **Energy & Utilities**
  Smart metering
- **Telecommunications:**
  Network monitoring, load prediction, blocked calls from load, phone usage, fraud detection and analysis...
- **Manufacturing:**
  Machinery going out of spec; process sampling and analysis
- **Logistics:**
  Location of a fleet (e.g. GPS); route analysis
- **Scientific research:**
  Temperature over time

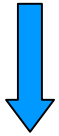
… And Many Many More …

Id	Timestamp	value 1	value 2
1	2012-01-01 01:00:00.0000	100	110
2	2012-01-01 01:00:00.0000	110	100
3	2012-01-01 01:00:00.0000	90	100
.			
.			
1	2012-01-01 01:10:00.0000	100	110
2	2012-01-01 01:10:00.0000	90	100
3	2012-01-01 01:10:00.0000	100	80

Traditional Relational Approach

Composite Index

Table Grows



Id	Timestamp	value1	value2
1	2012-01-01 01:**00**:00.0000	100	110
2	2012-01-01 01:**00**:00.0000	110	100
3	2012-01-01 01:**00**:00.0000	90	100
.			
.			
1	2012-01-01 01:**10**:00.0000	100	110
2	2012-01-01 01:**10**:00.0000	90	100
3	2012-01-01 01:**10**:00.0000	100	80

- Each row contains exactly one record = billions of rows
- Additional indexes are required for efficient lookups
- Data is appended to the end of the table as it arrives
- Id's stored in every record
- No concept of a missing row

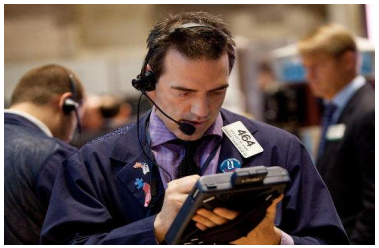
- Multiple tags
- Frequent intervals
- Huge index size
- Data dispersed on disc
- Performance drop

Humongous Data

- Roll-up information
- Running average
- Last few elements
- Calendar based Queries
- …..may more

TCO grows with time

CHALLENGES



Complex analysis

- Disc space requirement
- High memory/CPU
- High Administration cost

- Missing entries
- Interpolation

Tracking- Resolving glitches and errors

IBM Informix TimeSeries

- **Types**
  - Regular and Irregular
- **Performance**
  - Extremely fast data access
    - Data clustered on disk to reduce I/O
  - Handles operations hard or impossible to do in standard SQL
- **Space Savings**
  - Can be over 50% space savings over standard relational layout
- **Toolkit approach**
  - Allows users to develop their own algorithms
  - Algorithms run in the database to leverage buffer pool for speed
- **Easier**
  - Support built into the database
  - Conceptually closer to how users think of time series

Relational & TimeSeries Schema

```
CREATE TABLE intervals
(
meter_id CHAR(16),
timestamp DATETIME,
value1 NUMERIC(5,2),
value2 NUMERIC(5,2)
);
CREATE UNIQUE INDEX pk_intervals ON intervals
(meter_id, timestamp);
CREATE TABLE meters
(
meter_id CHAR(16),
meter_type SMALLINT,
location ST_Point,
[ … ]
);
CREATE UNIQUE INDEX pk_meters ON meters
    (meter_id);
```

```
CREATE ROW TYPE values
(
timestamp DATETIME YEAR TO FRACTION(5),
value1 NUMERIC(5,2),
value2 NUMERIC(5,2)
);
CREATE TABLE meters
(
meter_id CHAR(16),
meter_type SMALLINT,
location ST_Point,
intervals Timeseries(values)
[ … ]
);
CREATE UNIQUE INDEX pk_meters ON meters
    (meter_id);
```

Using Informix TimeSeries Schema

Id	Series
1	{2012-01-01 01:**00**:00.00000,100,110},{2012-01-01 01:**10**:00.00000,100,110},…..
2	{2012-01-01 01:**00**:00.0000,110,100}, {2012-01-01 01:**10**:00.0000,90,100},…..
3	{2012-01-01 01:**00**:00.0000,90,100}, {2012-01-01 01:**00**:00.0000,100,80},….
.	
.	

Table grows 

- Each row contains a growing set of records = one row per meter
- Data append to a row rather than to the end of the table
- Meter Ids not stored in individual records
- Data is clustered by meter id and sorted by time on disk
- Missing values take no disk space, missing interval reads take 4 bytes

Getting Started

- **Define Calendar**

```
insert into calendarpatterns values ( 'min15','{1 on, 14 off },minute' );


insert into CalendarTable ( c_name,c_calendar )
values ('cal15min','startdate(2012-02-01 00:00:00.00000), pattstart(2012-02-01
   00:00:00.0000), pattname(min15)');
```

- **Create Container**

```
execute procedure tscontainercreate('cont1', 'datadbs','values',1024,51200);
execute procedure tscontainercreate('cont2', 'datadbs','values',1024,51200);
```

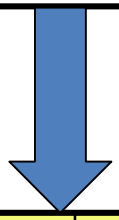
- **Initialize TimeSeries**

```
INSERT INTO meterdata VALUES('4727354321000111', 'type1','{lat, long}' ,
TSCreate('cal15min', '2012-02-01 01:00:00.00000', 0, 0, 0, 'cont1'));

INSERT INTO meterdata VALUES('4727354321000112', 'type1','{lat, long}' ,
TSCreate('cal15min', '2012-02-01 01:00:00.00000', 0, 0, 0, 'cont2'));
```

Virtual Table Interface – Relational View

Id	Series
1	{2012-01-01 01:**00**:00.00000,100,110},{2012-01-01 01:**10**:00.00000,100,110},…..
2	{2012-01-01 01:**00**:00.0000,110,100}, {2012-01-01 01:**10**:00.0000,90,100},…..
3	{2012-01-01 01:**00**:00.0000,90,100}, {2012-01-01 01:**00**:00.0000,100,80},….
.	
.	



Execute procedure tscreatevirtualtable ('SM_vt', 'Smart_meter');

Id	Timestamp	value1	value2
1	2012-01-01 01:**00**:00.0000	100	110
2	2012-01-01 01:**00**:00.0000	110	100
3	2012-01-01 01:**00**:00.0000	90	100
.			
.			
1	2012-01-01 01:**10**:00.0000	100	110
2	2012-01-01 01:**10**:00.0000	90	100
3	2012-01-01 01:**10**:00.0000	100	80

Accessing TimeSeries

- **Access through standard tabular view**
  - Virtual Table Interface (VTI)
  - Makes TimeSeries look like a standard relational table
- **SQL Interface**
  - 100+ functions
- **Customized functions**
  - Written in Stored Procedure Language (SPL), “C”, Java
  - 65+ “C” functions
  - Fast Loader APIs

Real Customer Comparison TimeSeries vs. Relational(Major Texas Power Distributor)

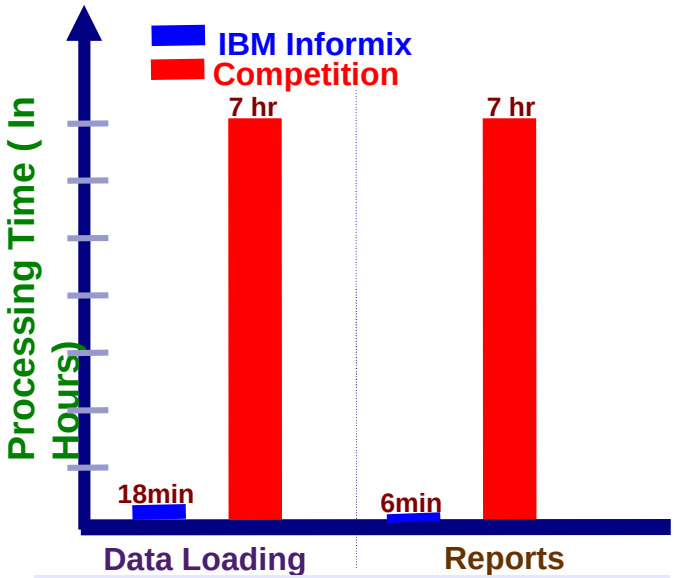
- Simulated Environment
  - 1 million meters
  - 90 days worth of meter data
  - 15 minutes intervals
  - 200 locations
  - 500 feeders
  - 34 substations
- Hardware/OS Used
  - Power7 with 2 sockets each with 8 cores
  - 64 bit SUSE Linux v11
  - 128 GB of memory
    - Memory actually needed = 44GB
  - 6 disks dedicated to the database
    - Only 350GB disk space used by database
    - 2 additional disks for OS and input file staging
- Software Used
  - Informix Ultimate edition
  - Informix TimeSeries

Why Did they Contact Us?

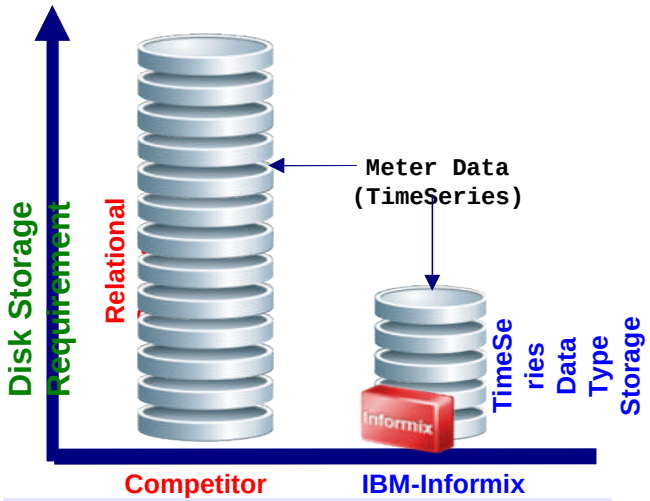
- Competitor was barely able to load and process the data in 24 hours
  - It was taking the competitor about **5-7 hours** to load and validate the data for **1 million** meters
  - The *ERCOT compliance reports were each taking many hours to run
  - They were having problems running reports while data was being loaded
- Customer was looking for ways to cut their costs
  - Competitor used about **1.3 TB** to store **90 days** worth of data
    - Estimate for competitor to handle **3 years** of data for **3.5 million** meters was about **55 TB**
  - Customer wanted to do more with the hardware they had
    - Process the data more quickly to allow other applications to run against the data
- Customer was worried about their future
  - In the next 1 to 2 years they expect to expand to manage 3.5 million meters
    - Their current 73 billion records stored would become 220 billion records stored
  - Even if the competitor scaled linearly there would not be enough time in the day to read and process all this data

*Electric Reliability Council of Texas

Results of the POC with Informix TimeSeries



Performance Comparison for Data Loads and Reports for 1 Million Meters



Data Storage Comparison for 1 million meters

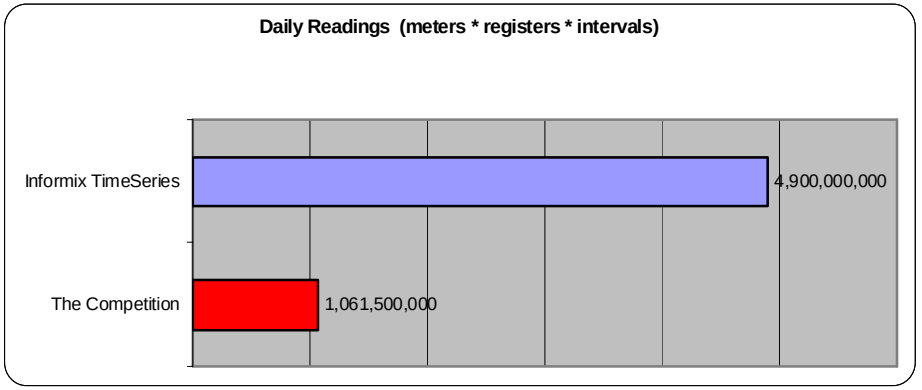
*"We are committed to improving energy efficiency by adopting the latest digital technologies providing Smart Metering solutions to our customers. Our major challenges are the significant storage space and performance required to manage the data collected by the smart meters. With Informix and its native support for time series data, we believe we will meet both of these challenges. Our initial testing has shown that with an Informix based solution we are able to reduce the storage requirements by **two thirds** and speed the query **performance up to 60 times compared to the existing solution**."*

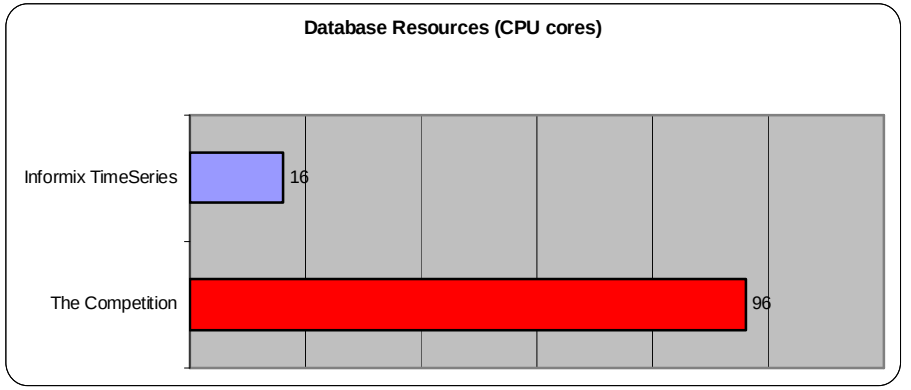
Director Technology, Strategy, & Architecture

Informix TimeSeries vs Relational Database Systems

Comparison of Published Benchmarks for Meter Data Management

	Meters	Daily Reads	Total Cores	Total RAM		DB cores	App cores	DB RAM	App RAM
Informix TimeSeries	100M	4.9B	16	500		16	(shared)	500	(shared)
The Competition *	5.5M	1.06B	456	3668		96	360	768	2900





5 times the performance

< 1/5 the resources

… with significantly simpler management using a single node system

* Based on published Oracle benchmark
http://www.oracle.com/us/industries/utilities/performance-benchmark-exadata-wp-161572.pdf

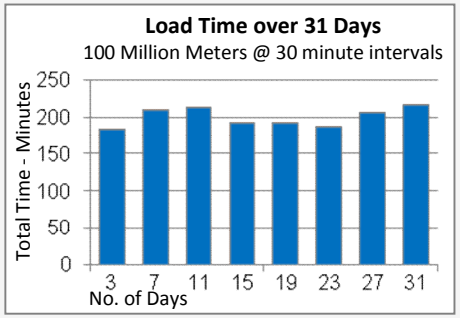
Informix TimeSeries – AMT-SYBEX Benchmark

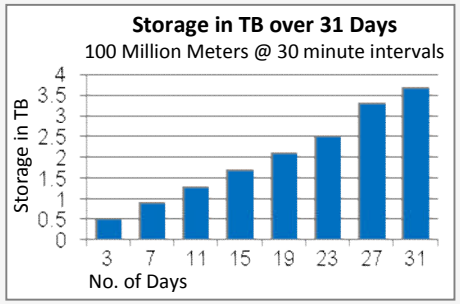
10 Million Meter Benchmark Results (36 minutes – total time)

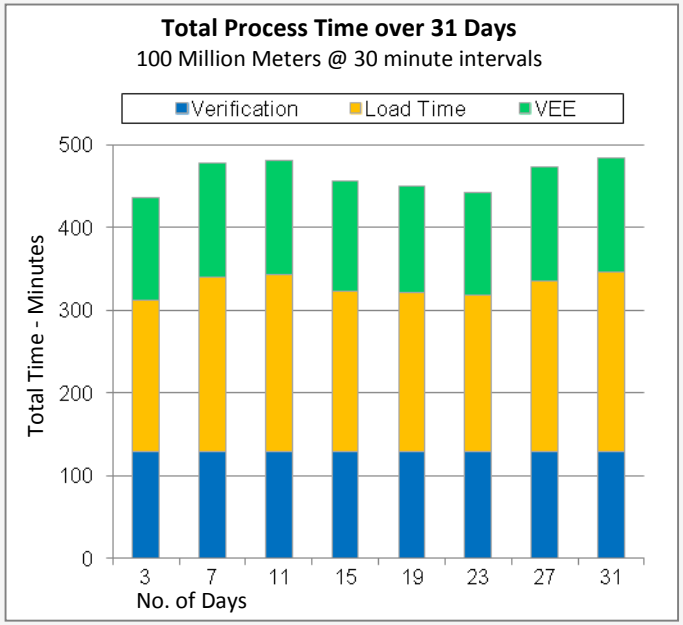
Process	Elapsed Time	Avg. Throughput Rate
Preparation and Technical Verification	10 min 02 sec	797,342 records/sec
Data Load	13 min 56 sec	457,162 records/sec
Validation, Estimation, and Editing (VEE)	11 min 18 sec	707,964 records/sec

100 Million Meter Benchmark Results (7 hrs 35 minutes – total time)

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 (VEE)	2 hrs 11 min	623,409 records/sec







Benchmark Results

- End to end processing time of 36 minutes for **10** Million meters at 30 minute intervals.
- End to end processing time of < 8 hours for **100** Million meters including billing queries.
- Processed 31 days of data using < 4TB of XIV storage on a single 16 core Power 750 machine.
- Overall processing time remained constant with increased historical data.
- Storage used remained **linear** with increasing data.

Benefits of Informix TimeSeries

- Only DBMS with native time series capability.
- Extreme performance – up to 50 times compared to relational.
- High level of compression – up to 70% storage savings.
- Lowers hardware, storage and CPU operating expenses.
- Benchmark processed 100M+ meters in 1 business day.
- Enterprise advanced analytics delivered by Netezza.
- Incorporates tool kit for custom analytic functions.
- Leverages IBM hardware platforms.

Outstanding Business Value with Amazing Performance!

Data Management needs for IUN

- Reduce hardware and storage costs as readings grow.
- Performance throughput and real-time access to data.
- Data aggregation and advanced analytic capabilities.
- Configuration and performance management tools.
- Achieve high error processing rates.
- Ensure high levels of data accuracy & consistency.

Informix 11.7 – Call to Action

- **Test drive Informix 11.7**
  - Download the Developer Edition
  - Download the Innovator-C Edition
  - Save operational costs by upgrading to Informix 11.7
- **Check out the latest edition of IBM Data Management Magazine for detailed coverage of Informix 11.7**
  - Informix 11.7 Hands-on Look; by Lester Knutsen and Art Kagel
  - IIUG User View - Why Do You Love Informix?; by Stuart Litel
  - More Musings on Informix 11.7; by Lester Knutsen and Art Kagel
  - DHL Mexico Follows the Money; by Merv Adrian, Gartner
  - Contact Cameron Crotty to be featured in upcoming issues Cameron@tdagroup.com

- See what industry analysts are saying about the value of Informix
  - ITG report: Cost/Benefit Case for IBM Informix compared to Microsoft SQL Server
  - Forrester Consulting report: Informix Total Economic Impact
- Sign up for a no charge on-site technical assessment
- Join the IIUG www.iiug.org & http://informixindia.in/ and register for 2013 IIUG Conference
- See Redbook on **Solving Business Problems with Informix TimeSeries** http://www.redbooks.ibm.com/abstracts/sg248021.html
- Contact your IBM representative for additional informationZ

http://www.ibm.com/informix

Vaibhav S Dantale: vdantale@in.ibm.com