

Informix Time Series Acceleration. **La analítica para sensores inteligentes.**

José Manuel Ruiz Gallud.
Especialista de IBM en Soluciones de Gestión de la Información



Datos de Sensores traen consigo nuevas aplicaciones

Utilidades (Electricidad, Agua, etc)

- Analisis de impacto del clima en la generación de electricidad.
- Monitorización de transmisión.
- Manejo de redes de electricidad inteligentes (Smart Grid)

Finanzas

- Detección de fraudes
- Manejo de riesgos
- Análisis de stock en tiempo real

Manufactura

- Localización y rendimiento de recursos
- Manejo y respuesta de eventos.



Salud

- Monitorización remota de salud
- Sistemas de Aviso temprano de epidemias
- Monitorización en Unidades de Cuidados Intensivos (UCIs)

Logística y Transporte

- Gestión de flotas y entregas
- Impacto del tiempo y el tráfico en la logística y consumo

Militar y Policia

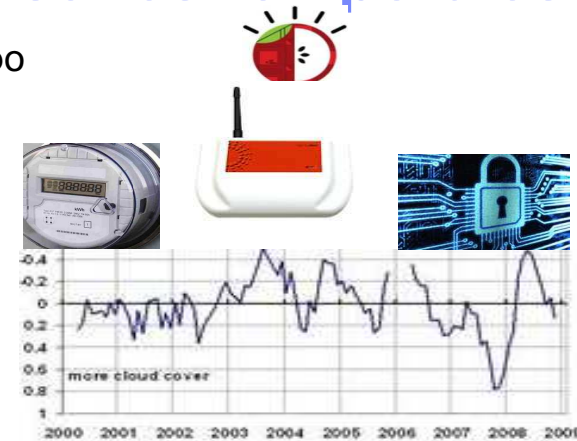
- Vigilancia y movimiento de blancos en tiempo real
- Situational awareness
- Cyber security detection



La naturaleza 'Big-data' de los datos de series-temporales

- Los instrumentos generan un gran volumen de datos basados en tiempo
 - Los datos se basan y se serializan en el tiempo
 - Mercado continuo, smart meters, dispositivos en red, sensores industriales, sensores de tráfico, defensa, etc.

- Características de los datos de series-temporales
 - Los datos están ordenados por tiempo, pe. 15 min por lectura
 - Los valores de los datos varían con el tiempo



Ejemplos de datos de series-temporales

ID, Values, Kwh	ID, Values, Kwh	ID, Values, Kwh	ID, Values, Kwh
ID101, 220, 0.5	ID101, 220, 0.8	ID101, 215, 0.6	ID101, 218, 0.5
ID102, 219, 0.6	ID102, 220, 0.6	ID102, 222, 0.8	ID102, 219, 0.7

01:00:00	01:15:00	01:30:00	01:45:00	Timeline
----------	----------	----------	----------	----------



Típico Esquema Relacional para Datos de Smart Meters

Índice

Tabla Smart_Meters

¿1000 * billones de filas?

La Tabla Crece ↓

Meter_id	Time	KWH	Voltage	CoIN
1	1-1-11 12:00	Value 1	Value 2	Value N
2	1-1-11 12:00	Value 1	Value 2	Value N
3	1-1-11 12:00	Value 1	Value 2	Value N
...
1	1-1-11 12:15	Value 1	Value 2	Value N
2	1-1-11 12:15	Value 1	Value 2	Value N
3	1-1-11 12:15	Value 1	Value 2	Value N
...

- Cada fila contiene un registro = billones de registros en la tabla
- Índices adicionales son necesarios para búsquedas efectivas
- Los datos se van añadiendo al final de la tabla tal y como llegan
- Cada registro incorpora el identificador del contador 'Meter ID's'



La misma tabla utilizando Informix TimeSeries (vista lógica)

10M cuentas/filas *
105120 por cuenta

Tabla Smart_Meters

Meter_id	Series
1	[(1-1-11 12:00, value 1, value 2, ..., value N), (1-1-11 12:15, value 1, value 2, ..., value N), ...]
2	[(1-1-11 12:00, value 1, value 2, ..., value N), (1-1-11 12:15, value 1, value 2, ..., value N), ...]
3	[(1-1-11 12:00, value 1, value 2, ..., value N), (1-1-11 12:15, value 1, value 2, ..., value N), ...]
4	[(1-1-11 12:00, value 1, value 2, ..., value N), (1-1-11 12:15, value 1, value 2, ..., value N), ...]
...	...

La tabla crece →

- Cada fila contiene un conjunto creciente de registros = una fila por contador
- Los datos se añaden al final de la fila, no al final de la tabla
- El indentificador 'Meter IDs' se almacena sólo una vez
- Los datos están clusterizados en disco por meter id y ordenados en el tiempo
- Los valores que faltan no usan espacio, cada intervalo que falta utiliza 2 bytes



Informix TimeSeries para series temporales - Big Data

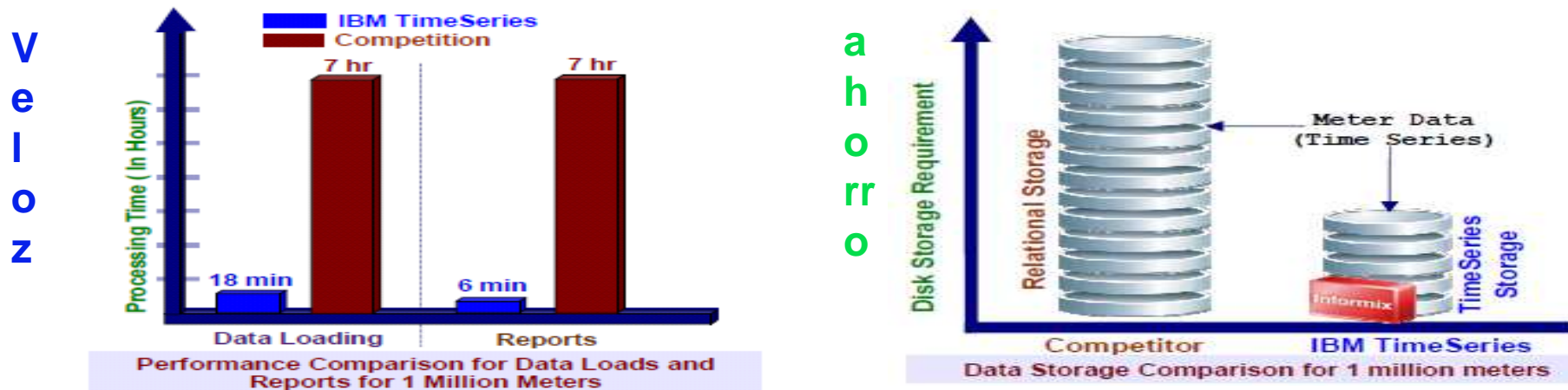
- **La Tabla Frontal – table de la base de datos relacional**
 - Esta tabla mantiene la clave primaria, más efectiva
 - Ejemplo: 10M de cuentas de smart meters

- **La Tabla Trasera – tabla TimeSeries**
 - Los datos para cada contador se almacenan ordenados por tiempo

No other RDBMS has native time series support



Comparativa: gestión de datos Relacional vs TimeSeries



- **Comparativa entre Relacional y TimeSeries**
 - El motor de TimeSeries es órdenes de magnitud más rápido que el sistema relacional
 - El sistema con TimeSeries requiere la mitad del espacio en disco
- **El gestor Informix con su capacidad relacional + TimeSeries cumple con las necesidades y variedad de negocio**
 - Sistema de facturación (base de datos relacional)
 - Sistema analítico en tiempo real (TimeSeries)



Comparativa con datos de Smart-Meters

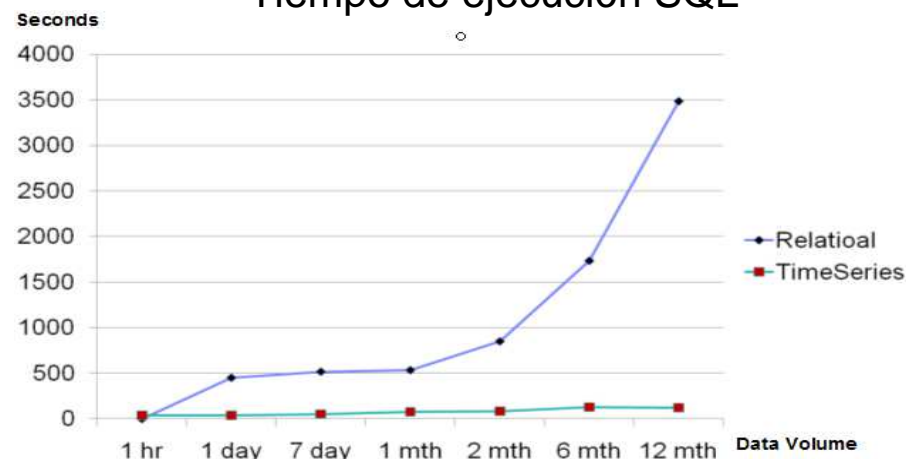
- El volumen de datos
 - Número de meters: 10.000.000
 - Intervalo: 1 hora
 - Días registrados: 365 días
 - Total registros: $10.000.000 * 24 * 365 = 87.600$ millones

Tiempo medio de las consultas

Interval	Relational	TimeSeries
1 hour	0.491s	36.622
1 day	453.295	37.115
7 day	519.494	52.707
1 month	536.028	76.161
2 month	852.244	79.430
6 month	1735.238	127.071
12 month	3484.112	118.851



Tiempo de ejecución SQL



- En el entorno relacional las consultas degeneran su rendimiento exponencialmente a medida que crecen los datos a tratar
- El rendimiento con TimeSeries se estabiliza aún aumentando el volumen de datos

Virtual Table Interface: Time Series como datos Relacionales

Tabla TimeSeries
Smart_meter

mtr_id (int)	Series timeseries(mtr_data)
1	[(Mon, v1, ...)(Tue,v1...)]
2	[(Mon, v1, ...)(Tue,v1...)]
3	[(Mon, v1, ...)(Tue,v1...)]
4	[(Mon, v1, ...)(Tue,v1...)]
5	[(Mon, v1, ...)(Tue,v1...)]
6	[(Mon, v1, ...)(Tue,v1...)]
7	[(Mon, v1, ...)(Tue,v1...)]
8	[(Mon, v1, ...)(Tue,v1...)]

Tabla Virtual TimeSeries
SM_vt

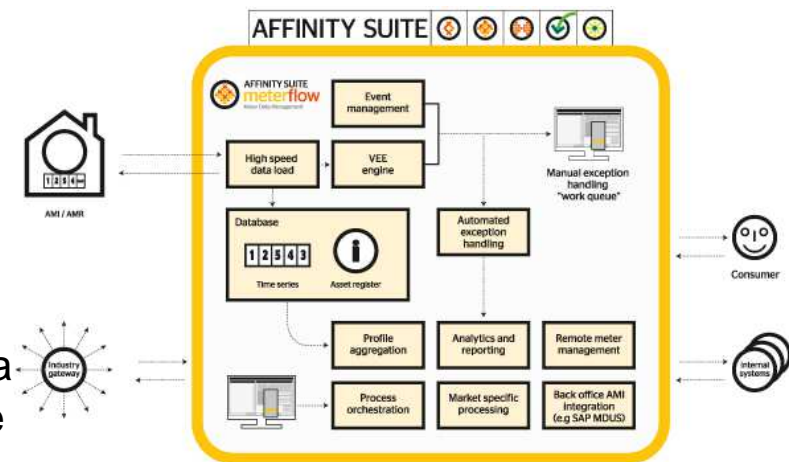
mtr_id	date	col_1	col_2	
1	Mon	Value 1	Value 2	...
1	Tue	Value 1	Value 2	...
1	Wed	Value 1	Value 2	...
...
3	Mon	Value 1	Value 2	...
3	Tue	Value 1	Value 2	...
3	Wed	Value 1	Value 2	...
...

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

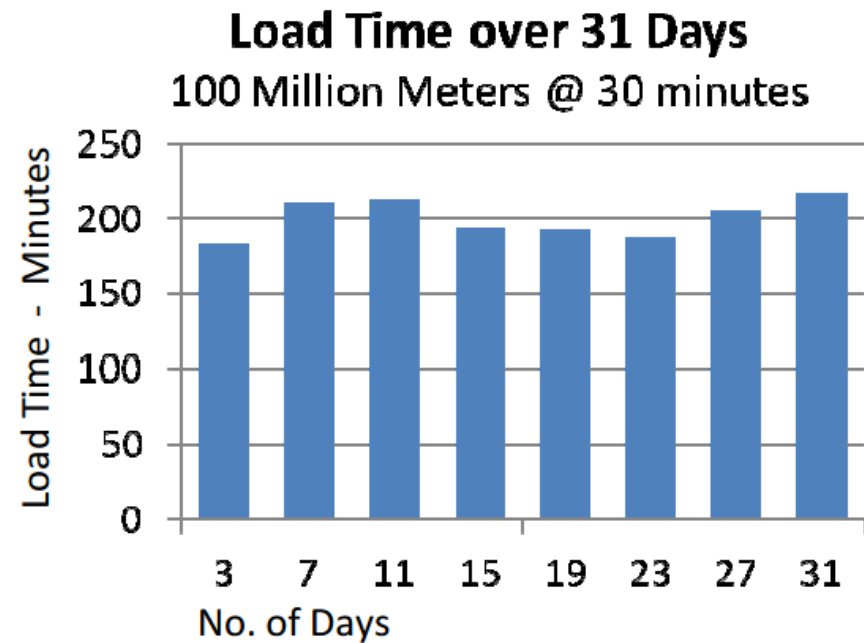
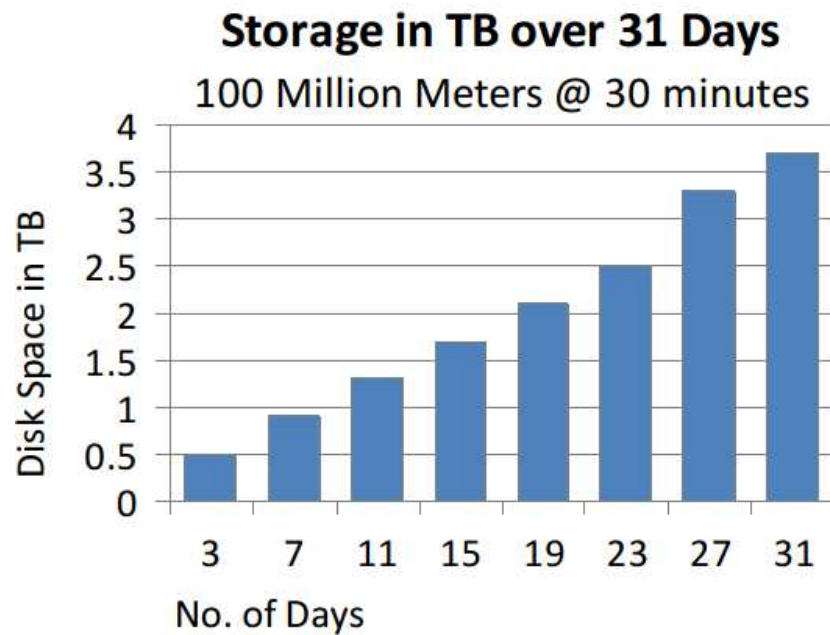


Soluciones Meter Data Management (MDM) con Informix

- AMT-Sybex incorpora Informix como motor de datos
- AMT-Sybex e IBM realizaron un benchmark con 100M de smart-meters en IBM-Montpellier
 - IBM p750, 16 cores power7, AIX 7.1, cabina XIV
 - 31 días con intervalos de 30 min= 148.800M filas
 - Menos de 4 Tb de espacio en disco consumido
 - Cálculos de 21.000 facturas por segundo mientras se cargan datos
 - 420.000 registros por segundo de carga de media
 - Tiempo de carga, y espacio utilizado se mantiene lineal con el tiempo
 - Sistema capaz de cargar datos y calcular facturas diarias en una ventana de 8 horas para 100M de meters y en 36 minutos para 10M de meters.



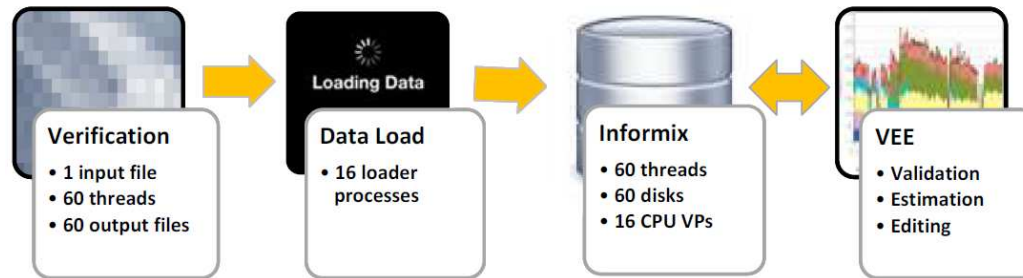
Soluciones Meter Data Management (MDM) con Informix



Disk space consumption and load times over 31 days



Soluciones Meter Data Management (MDM) con Informix



Total Process Time over 31 Days
100 Million Meters @ 30 minutes

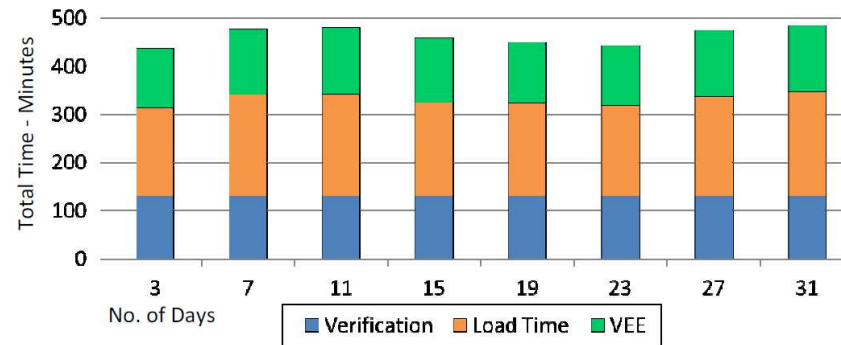


Figure 6 – End to end processing time for 100 million meters over one month



La Compañía Eléctrica más grande de Europa resuelve el reto de 'Big Data' de sus Smart Meters con Informix TimeSeries y el Acelerador IWA

- E.ON Metering (EMTG) es el centro de excelencia para el desarrollo y comercialización de soluciones de energía inteligentes y tecnologías en la propia E.ON, compañía de electricidad y gas más grande de Europa
- EMTG opera una sofisticada infraestructura de datos de Smart Meters basada en tecnología IBM Informix TimeSeries en combinación con el acelerador Informix In-Memory Data Warehouse Accelerator
- Los productos de IBM utilizados en la solución:
 - *Informix TimeSeries con Informix Warehouse Edition 11.70*
 - *Cognos Enterprise 10*



IBM Informix TimeSeries: Fortalezas



- ¿Qué es Time Series?
 - Una serie de registros interconectados de manera lógica y ordenados por tiempo
- Rendimiento con Informix
 - Las consultas sobre time series van ordenes de magnitud (x60) más rápido que sobre entornos relacionales
 - Realiza operaciones complejas o imposibles con SQL estándar
 - Con cargadores continuos en tiempo real o en batch
- Ahorro de Espacio con Informix
 - Al menos ahorra un 50% de espacio en disco
- Flexibilidad con Informix
 - Algoritmos que corren en el propio gestor de base de datos
 - Combina datos de series temporales y datos relacionales en un único producto
- Facilidad de uso con Informix
 - Fácil integración con aplicaciones vía ODBC/JDBC
 - Es conceptualmente más atractivo para aquellos usuarios habituados a trabajar con series temporales.

**No other RDBMS
has native time
series support**



Gracias



Big data, integración y gobierno: www.ibm.com/software/es/info/rte/bdig/



[@ibmanalytics_es](https://twitter.com/ibmanalytics_es)