Solving the Big Data Challenge of Sensor Data

Bernie Spang
Director, Strategy and Marketing, IBM Database Software and Systems

Kevin Brown
Chief Architect, IBM Information Management

IBM

Big Data is the next Natural Resource

"We have, for the first time, an economy
based on a key resource (Information)
that is not only renewable, but self-generating.

Running out of it is not a problem, but drowning in it is."

— John Naisbitt

Big Data is All Data and All Paradigms

Transactional & Application Data	**Enterprise Content**	**Social Data**	**Sensor Data**

• Volume	• Variety	• Variety	• Velocity
• Structured	• Unstructured	• Unstructured	• Structured
• Throughput	• Volume	• Veracity	• Ingestion

New Opportunity is Fueled by Disruptive Technology Factors

Internet of Things

Billions of smart devices instrument our world today

33 billion
RFID tags embedded into our world

24 billion
Devices connected to the internet by 2020

The world is becoming more interconnected and more complex ...

534 billion
Dollars in mobile transactions by 2015

50%
Of businesses plan to adopt more collaborative sourcing models

Interconnecting these smart devices creates a **Central Nervous System**

5

Sensor Data Brings New Applications Across Industries

Utilities
- Weather impact analysis on power generation
- Transmission monitoring
- Smart grid management

Financial Services
- Fraud detection based on biometrics
- Risk management

Manufacturing
- Asset location and performance
- Event management and response

Health & Life Sciences
- Remote healthcare monitoring
- Epidemic early warning system
- ICU monitoring

Logistics & Transportation
- Fleet management and dispatch
- Weather and traffic impact on logistics and fuel consumption

Military and Law Enforcement
- Real-time surveillance and movement detection
- Situational awareness
- Cyber security detection

Healthcare Industry

Healthcare industry uses sensors to remotely monitor vital signs of patients, track critical equipment, under treatment. Monitoring of these vitals are utmost important in providing the right care at the right time to ensure patients are treated well especially in critical situations

Example Data Points

Vital signs (heart rate, temperature, blood pressure, etc.)

EKGs, ECGs

Asset/device location

Potential Benefits

Faster clinician insight into real-time data

Faster response and diagnoses

Proactive intervention

Medical device tracking

Applicability

Neo-natal care

Critical/Intensive care

Emergency response

Home/remote healthcare

Intelligent Transportation/Fleet Management

A vehicle tracking system uses sensors installed in vehicles to collect several data points over time for business intelligence.
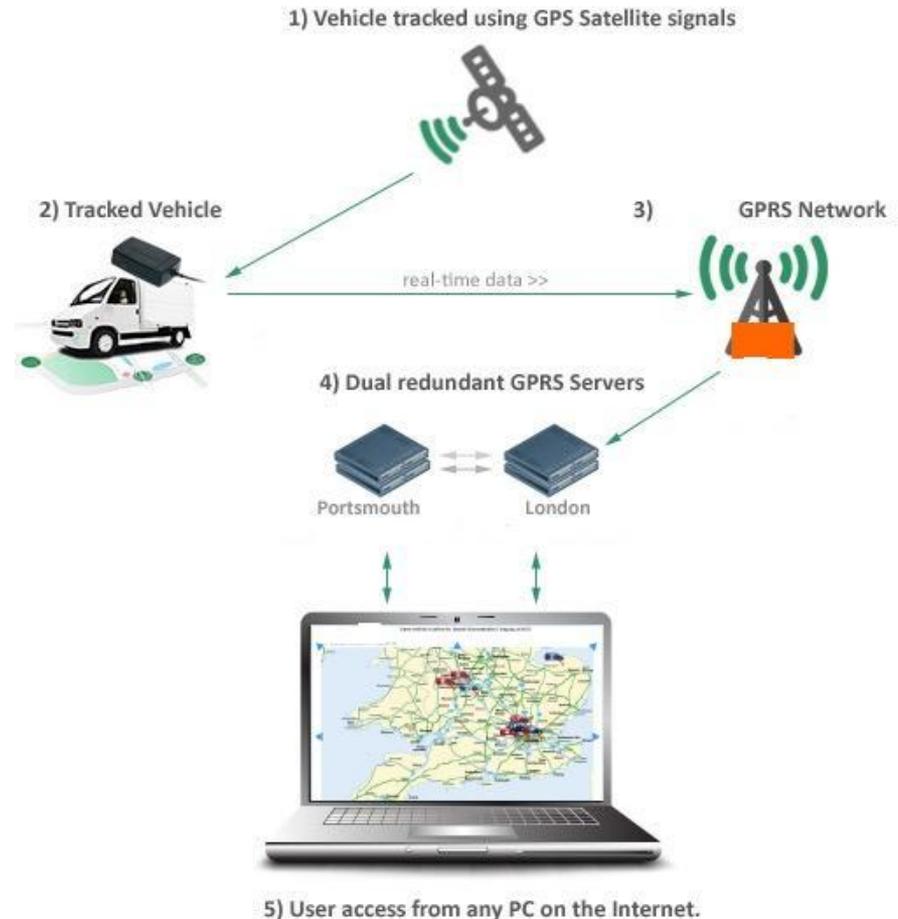
Example Data Points

Vehicle location

Rate of acceleration

Average speed

Idle time

Seat belt use

Potential Benefits

Optimize driver route/time

Predict vehicle shortages

Improve driver safety

Proactive vehicle maintenance

Reduced environmental impact

Applicability

Insurance (usage-based rates, theft prevention)

Fleet management (rental cars, moving vans, delivery trucks, etc.)

Urban transit – Bus and taxi management

1) Vehicle tracked using GPS Satellite signals

2) Tracked Vehicle

3) GPRS Network

real-time data >>

4) Dual redundant GPRS Servers

Portsmouth   London

5) User access from any PC on the Internet.

8

Process-Driven Manufacturing Plants

Process-driven businesses such as power plants, gas, oil & water utilities, and machinery manufacturing are under continual pressure to optimize business processes to increase quality and reduce costs. Sensor-based data generated throughout the manufacturing process provides critical information to deliver business benefits

Example Data Points

Asset location
Environmental data
Equipment condition and operations data
Defect Data

Potential Benefits

Optimize asset management
Reduce cost of plant operation
Proactive maintenance
Minimize defective product
Reduce environmental impact

Applicability

Energy & Utilities
Chemicals and Petroleum
Machinery Manufacturing (cars, farm equipment, military equipment, etc.)
Mining & Metallurgy

Energy & Utilities Smart Metering

Utilities companies are deploying smart meters to consumers' homes to automatically record utility usage at frequent intervals. They capitalize on the smart meter data to optimize energy purchases, resolve service issues quickly and improve power reliability while also helping customers reduce consumption and cut costs.

Example Data Points

Asset location

Environmental data

Equipment condition and operations data

Defect Data

Potential Benefits

Optimize asset management

Reduce cost of plant operation

Proactive maintenance

Minimize defective product

Reduce environmental impact

Applicability

Energy & Utilities

Chemicals and Petroleum

IBM

Why is Sensor Data Unique?

Massive sets of

Streaming data

Unpredictable data arrival

Data persistence not required

Analyze data-in-motion continuously

Predefined queries

Time series data

Time-sequenced data arrival with geo-spatial location

Data always persisted to disk

Analyze stored data for trending and patterns

Predefined and ad hoc queries

IT Challenges of Sensor Data Applications

- Sensors output data repeatedly over specified time intervals
  - *Data system must be able to receive and process data at extremely high input velocities*

- Data output is time-sequenced resulting in a different structure
  - *Data system must be able to store large volumes of data in a scalable fashion and allow near real-time access to it*

- Data naturally includes geospatial information
  - *Data system must be able to efficiently store, correlate, and analyze data with repect to time and location*

- There is a higher need to deal with communication failure
  - *Data system must be extremely reliable and continue to operate in case of hardware failures*

- Queries may need to access sensor data along with relational data
  - Data system must provide near real-time analysis of the collected data
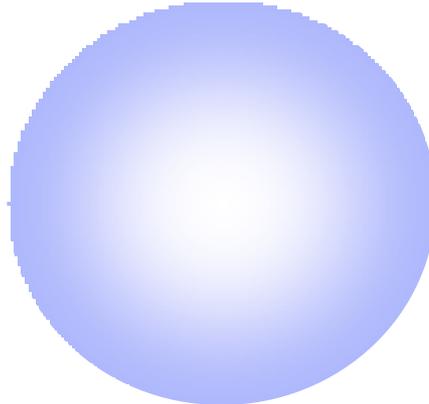
IBM Big Data Landscape

InfoSphere BigInsights
Hadoop-based low latency analytics for variety and volume

Hadoop

Information Integration

Stream Computing

InfoSphere Information Server
High volume data integration and transformation

IBM InfoSphere Streams
Low Latency Analytics for Streaming Data

Structured Database

IBM DB2 and Informix Warehouse Software
Large volume structured data analytics

IBM Informix TimeSeries
Machine Data Mgmt & Operational Analytics

IBM Smart Analytics System
Operational Analytics on Structured Data

IBM PureSystems
Expert integrated systems

TerraEchos combines InfoSphere Streams + Informix
to support covert intelligence and surveillance systems

Need

- Provide covert surveillance capabilities to detect, classify, locate and act on threats to a highly sensitive national lab

Benefits

- Faster, more intelligent response to any threat with real-time analysis of 42 TB streaming data daily

- Tracks all intrusion events in spatial context for unprecedented insight and response to potential threats

- Embedded data management provides high data availability to critical applications and eliminates need for DBAs

"We are capturing and reducing a tremendous amount of digital acoustic data and running intensive computational statistical analysis, all in one-fourteenth of a second."
~ *Dr. Alex Philp, President and CEO, TerraEchos, Inc.*

Time Series Data is BIG DATA

- Large amounts of data cause problems in 2 areas:
  - It's expensive and difficult to manage
    - Massive storage space required
    - Sophisticated partitioning schemes needed
    - Frequent reorganization leads to downtime

  - Poor query/analytic performance affects quality of service
    - Compliance reports must be completed before the end of each day
    - Customer portal queries must be handled in a timely manner
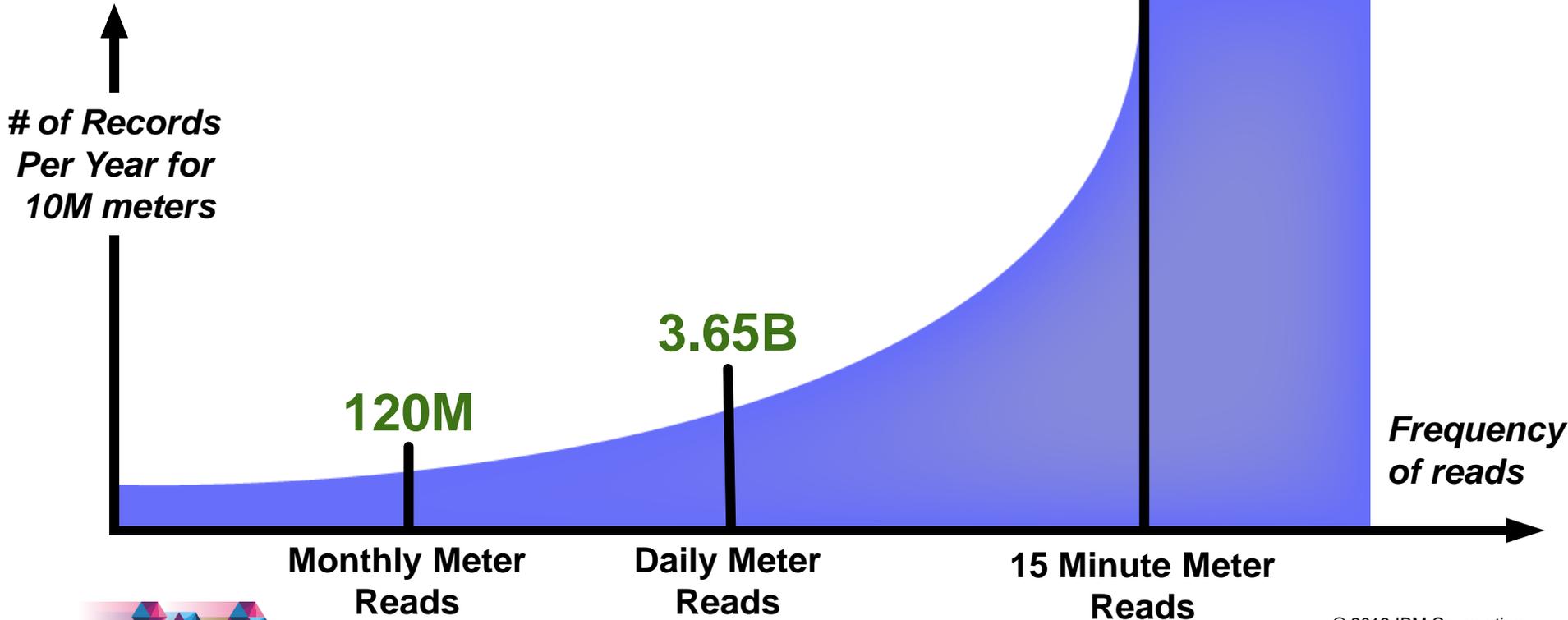   - Customer billing must be completed each day

Shifting Paradigm – Growing Volumes of Time Series Data

Smart Meter Example: Changing Requirements for 10 Million Smart Meters

Historical	Each meter read once per month
In process	Each meter read once every 15 minutes
Regulations	Need to keep data online for years

Data for a Utility In California

350.4 Billion Records per Year

# of Records Per Year for 10M meters

350.4 Billion
3.65B
120M

Frequency of reads

Monthly Meter Reads   **Daily Meter Reads**   **15 Minute Meter Reads**

Traditional Time Series Storage VS Informix TimeSeries

Traditional Table Approach

Meter_id	Time	KWH	Voltage		ColN
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
...	...	...	...	....	...

Informix Time Series Approach

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), …]
...	...

Accessing TimeSeries Data with Informix

- **Access through standard tabular view**
  - Makes TimeSeries look like a standard relational table

- **SQL Interface**
  - 100+ functions

- **Extensive library of API functions**
  - Write your own functions in Stored Procedure Language (SPL), "C", Java

Informix TimeSeries for Smarter Energy & Utilities

Oncor changes data management platform for Meter Data Management to deliver benefits throughout their value chain, from suppliers to consumers.

SMART METER TEXAS™

ONCOR. WE DELIVER.

A very **smart way** for Texans to **manage** electricity!

powered by Informix TimeSeries

Smarter Business Outcomes
- Provide customers better insight and control over how they use energy
- Automated outage management
- Accelerate time-to-value from smart meter data
- Minimize system and storage costs

Smarter Big Data Solution
- 500% faster processing of meter data
- 80% less storage space required
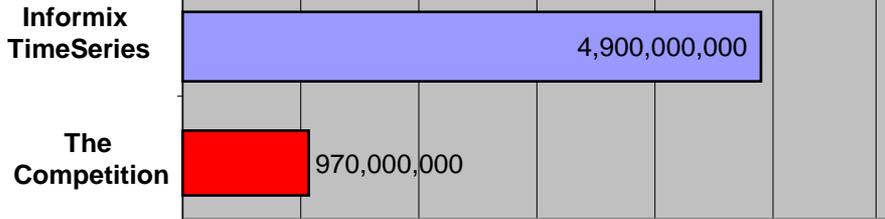- Consistent, scalable performance yields highly predictable costs

Informix TimeSeries vs Relational Database Systems
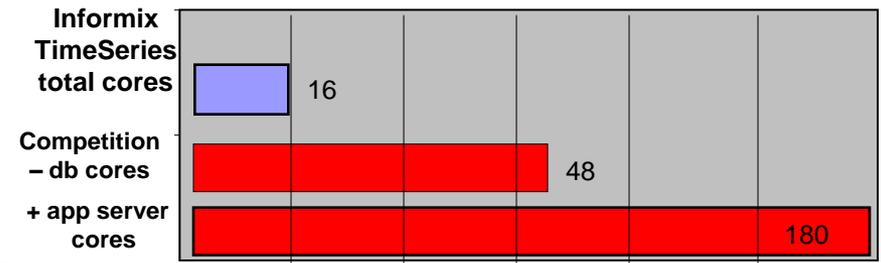*Comparison of Published Benchmarks for Meter Data Management*

5 times the performance

Daily Readings
(meters * registers * intervals)

Informix TimeSeries	4,900,000,000
The Competition	970,000,000

< 1/5 the resources

CPU resources (cores)

Informix TimeSeries total cores	16
Competition – db cores	48
+ app server cores	180

… with significantly simpler management using a single node system

No additional installations or license fees required for time series and spatial capabilities!

Comparison based on published Oracle benchmark
http://www.oracle.com/us/industries/utilities/utilities-meter-data-exadata-wp-486477.pdf
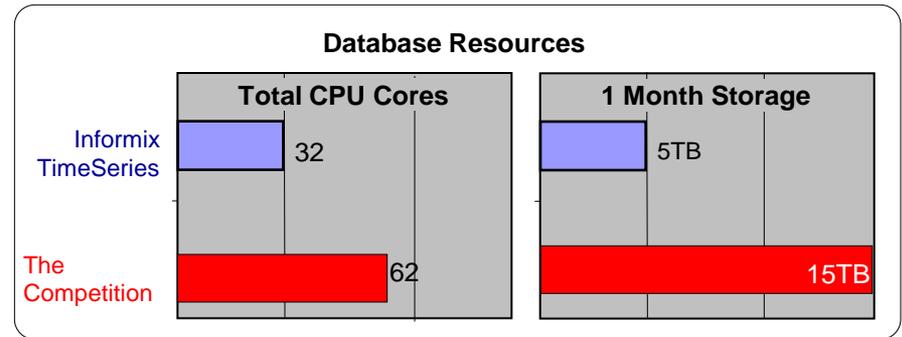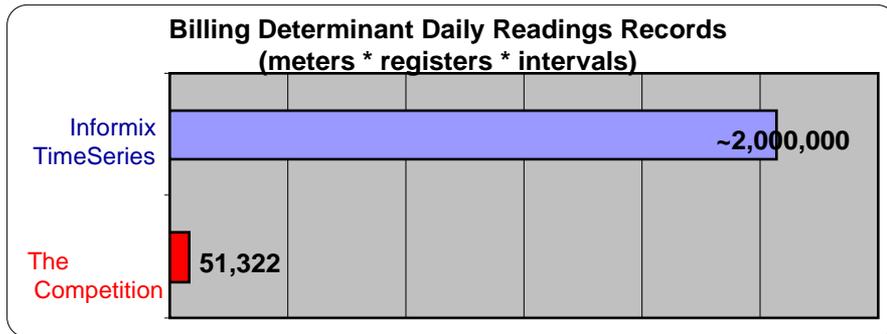
Solving the Big Data Challenge of Sensor Data

Landys + Gyr GridStream MDM Benchmark for Meter Data Management

	Meters	Daily Reads	RAM Used	Total Cores	DB Size / Month Of Data	# Of Process	Grid S/W	Daily Processing Time	Data Read Performance			
									Register Data	Interval Data	Register & Interval VEE	Billing Report
Informix TimeSeries	30M	3B	192 GB	32	5TB	90	N/A	5h 50m	33K	320K	266K	2000K
The Competition	30M	3B	192 GB	62	15TB	30000	Yes	11h	13K	255K	129K	51K

Billing Determinant Daily Readings Records (meters * registers * intervals)

- Informix TimeSeries: ~2,000,000
- The Competition: 51,322

Database Resources

Total CPU Cores
- Informix TimeSeries: 32
- The Competition: 62

1 Month Storage
- Informix TimeSeries: 5TB
- The Competition: 15TB

- ➢ 38 times reporting performance
- ➢ 4 times daily processing performance

… with significantly less resources

Delivers unbeatable low Total Cost of Ownership (TCO) for Utilities

Solving the Big Data Challenge of Sensor Data

The Defense Advanced Research Projects Agency (DARPA) enhances safety of US soldiers with devices that monitor exposure to explosive blasts

Indicator Lights

Blast Gauge

Sensor Dome

Impact-Resistant Casing

Micro-USB Port
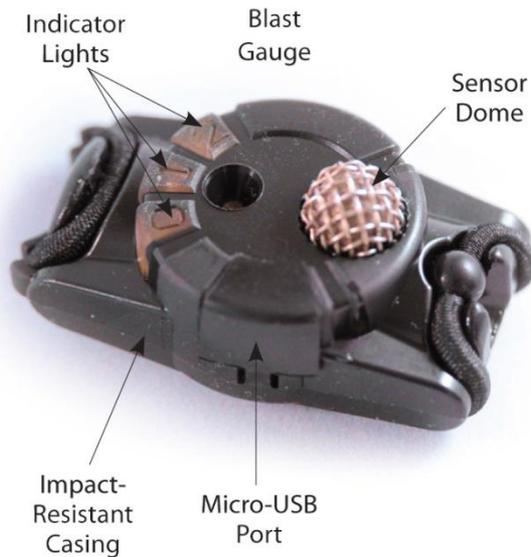
The data-logging Blast Gauge measures pressure, head acceleration, and time, to help correlate blast events with injuries. The compact device weighs less than one ounce, making it easy for soldiers to wear.

What's smart?
- Soldiers wearing small devices that capture thousands of data points per second about the severity of explosions
- Medics having instant access to device data to determine blast severity and provide appropriate level of medical care
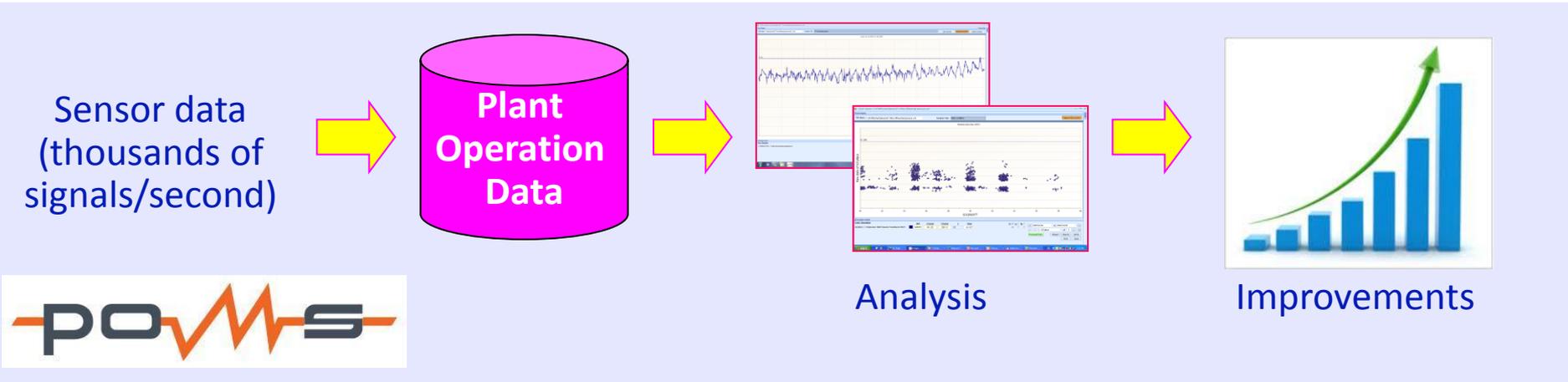
Smarter Outcomes
- Better diagnosis and treatment of brain injuries will save billions of dollars in health care costs
- Soldiers will suffer fewer incidents of traumatic brain injury

powered by Informix TimeSeries

Chronos POMS – Plant Operation Management System

Sensor data (thousands of signals/second)

Plant Operation Data

Analysis

Improvements

Specially Designed for:

- Power Plants
- Water Utilities
- Petrol Chemical Process Plants

Powerful Analytical Tools for:

- Condition Monitoring – critical facilities
- Performance Monitoring – overall plant activities
- Preventive Maintenance – all equipment etc.

powered by Informix TimeSeries

Informix TimeSeries Summary of Key Strengths

- Informix TimeSeries feature solves the big data challenge of time series data
  - Up to 66% less storage required
  - Analytical workloads run orders of magnitude faster using Informix TimeSeries
  - Supports relational, time series, and spatial data *all in the same query*
  - Delivers enterprise-class quality of service

- Little-to-no time spent on normal database maintenance tasks
  - Self-managing capabilities automatically sense and respond to changes
  - On average, Informix requires less than 60 minutes of 1 DBA time/week
  - Upgrades done without downtime

Learn more by visiting www.ibm.com/informix