IEEE Bid Data Initiative Workshop

BDIW October 2014

What's Next for Big Data Analytics

Mahmoud Daneshmand, PhD
Industry Professor, Stevens Institute of Technology
mahmoud.daneshmand@stevens.edu

The Challenge

Billions of “things” such as sensors/RFID, mobility devices & smart phones are generating mountains of data on all aspects of the human life: health, environment, transportations, security, shopping, smart cities, smart grids, home, etc.

The Biggest Challenge of the current Big Data Era is management and mining of ever-increasing streams of data generated by these devices.

“Without stream processing, there’s no Big Data and no Internet of Things”, (Dana Sandu, March 19, 2014).

White House
The Big Data Research and Development Initiative

On March 29, 2012, Obama Administration announced the "Big Data Research and Development Initiative." By improving our ability to extract knowledge and insights from large and complex collections of digital data, the initiative promises to help accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning.

To launch the initiative, six Federal departments and agencies will announce more than $200 million in new commitments that, together, promise to greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data. Learn more about ongoing Federal government programs that address the challenges of, and tap the opportunities afforded by, the big data revolution in our Big Data Fact Sheet.

Internet of Things (IoT) Promise

IoT promises to be the most disruptive technological revolution since the advent of the Internet. Projections indicate that up to 100 billion objects will be connected to the Internet by 2020. IoT covers all types of sensors, communication protocols, computational tools, techniques, devices, processors, embedded systems, data warehousing, big data, cloud computing, server farms, grid computing etc.

IEEE TENSYMP 2015

Internet of Things (IoT) Promise

No matter which vendor companies choose, the Internet of things holds growth projections to make everyone's head spin:

- **Gartner predicts that the Internet of things and personal computing will unearth more than $1.9 trillion in revenue before 2020,**
- **Cisco thinks there will be upwards of 50 billion connected devices by 2020,**
- **IDC estimates technology and services revenue will grow worldwide to $7.3 trillion by 2017 (up from $4.8 trillion in 2012).**

Internet of Things (IoT) Promise

Intel, IBM, AT&T, Cisco and GE found the Industrial Internet Consortium:

Intel, IBM, AT&T, Cisco and GE announced the formation of the Industrial Internet Consortium this week, a membership group of telcos, research institutes and technology manufacturers focused on developing interoperability standards and common architectures to bridge smart devices, machines, mobile devices and the data they create.

March 28, 2014 Jonathan Brandon

Size

Morgan Stanley: 75 Billion Devices Will Be Connected To The Internet Of Things By 2020.

There are 200 unique consumer devices or equipment that could be connected to the Internet that have not yet done so.

Business Insider, Oct 2, 2013

Intel: Intel creates a new ‘Internet of Things Solutions Group’ reporting directly to Chief Executive Brian Krzanich.

The Santa Clara, California chipmaker and other technology companies are betting that what they call the 'Internet of Things' - a trend toward connecting everything from bathroom scales to skyscraper ventilation Systems via the Internet - will create massive demand for new electronics and software.

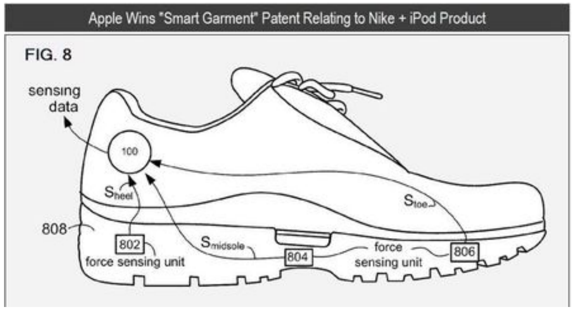
Reuters, Nov 5, 2013

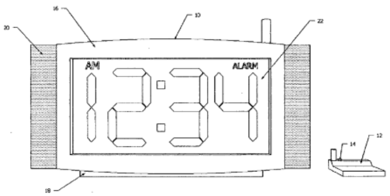
“Without Stream Data Analytics, there’s no Big Data, no Internet of Things, and No Revenue”

Example of Connected Things

Past: E-Zpass (Barcode, etc)

Today: Google Glass





Future: Apple's Smart Garment
(e.g. shoe)
-> automatic clothing suggestion for everyday based on weather, your business schedule, your body condition, etc

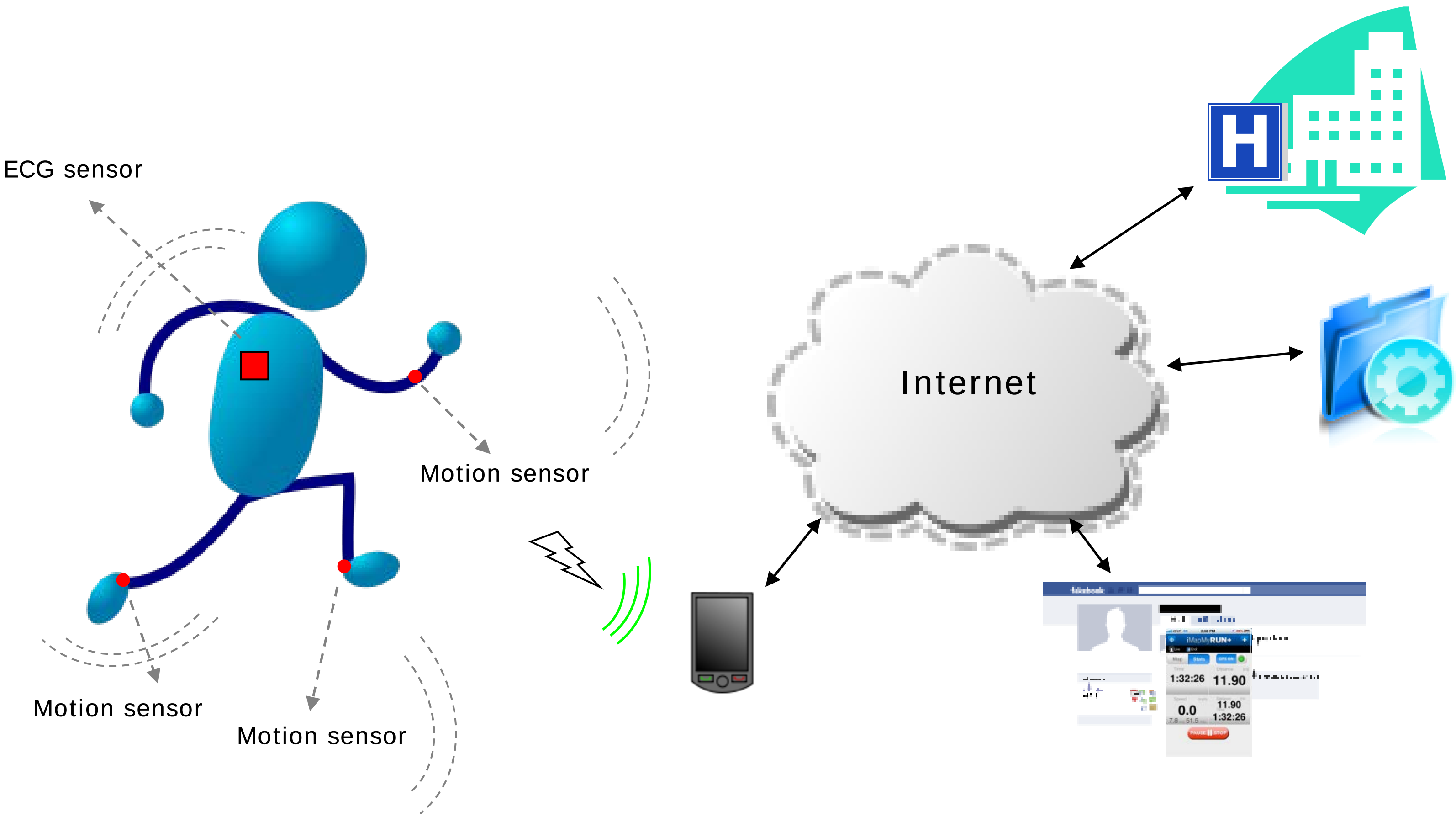
Future: Tooth Tattoo
-> put on cow's teeth to detect disease wirelessly
-> from Princeton University (http://www.princeton.edu/main/news/archive/S33/79/62E42/index.xml?section=topstories)

Future: Smart Clock
-> integrated with Bedpost Pressure Sensor
-> alarming only when the bed is occupied
-> Patent App 20110085423

People Connecting to Things

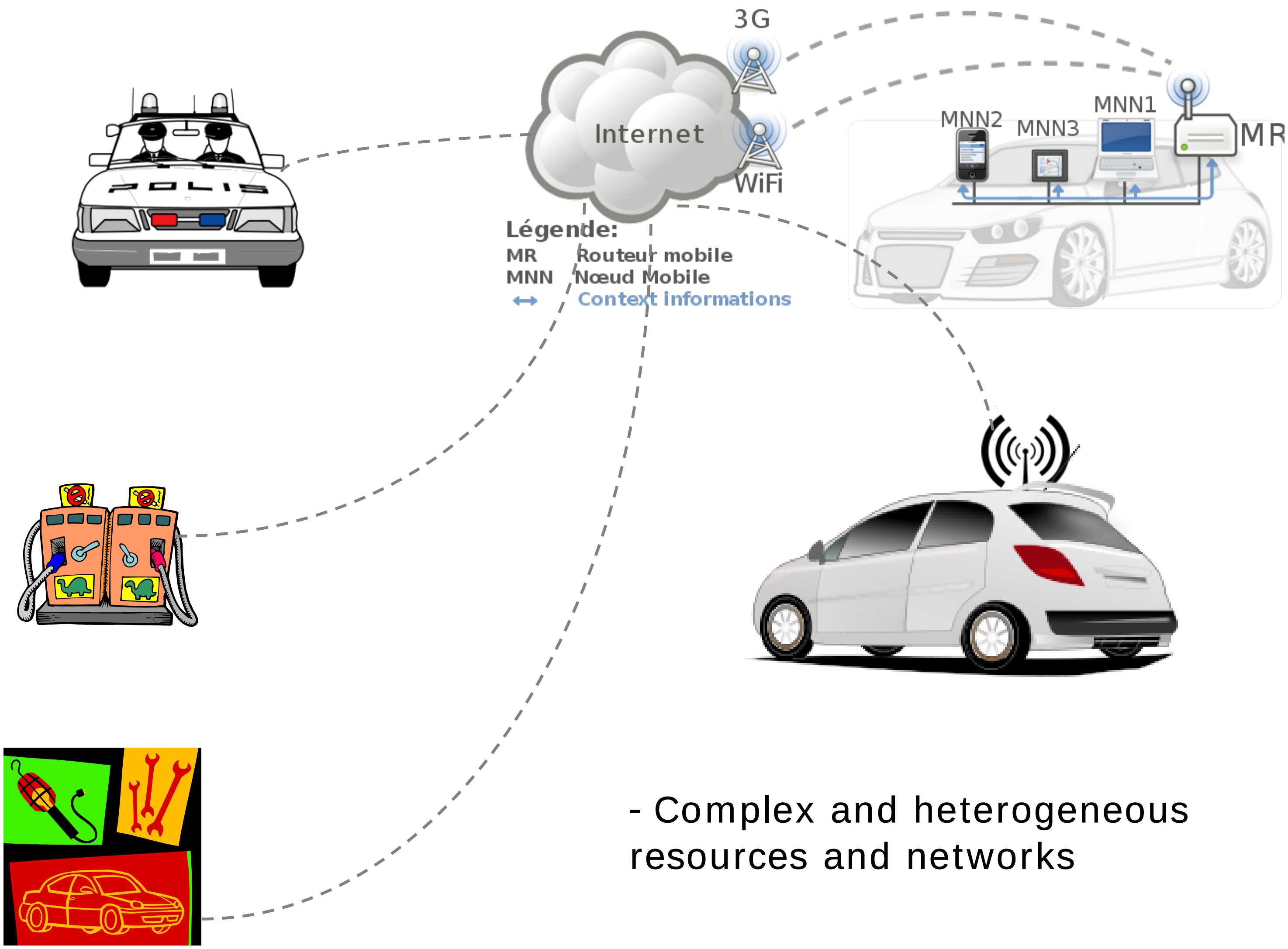
Source: Wei Wang, Cory Henson, Payam Barnaghi

UNIVERSITY OF SURREY



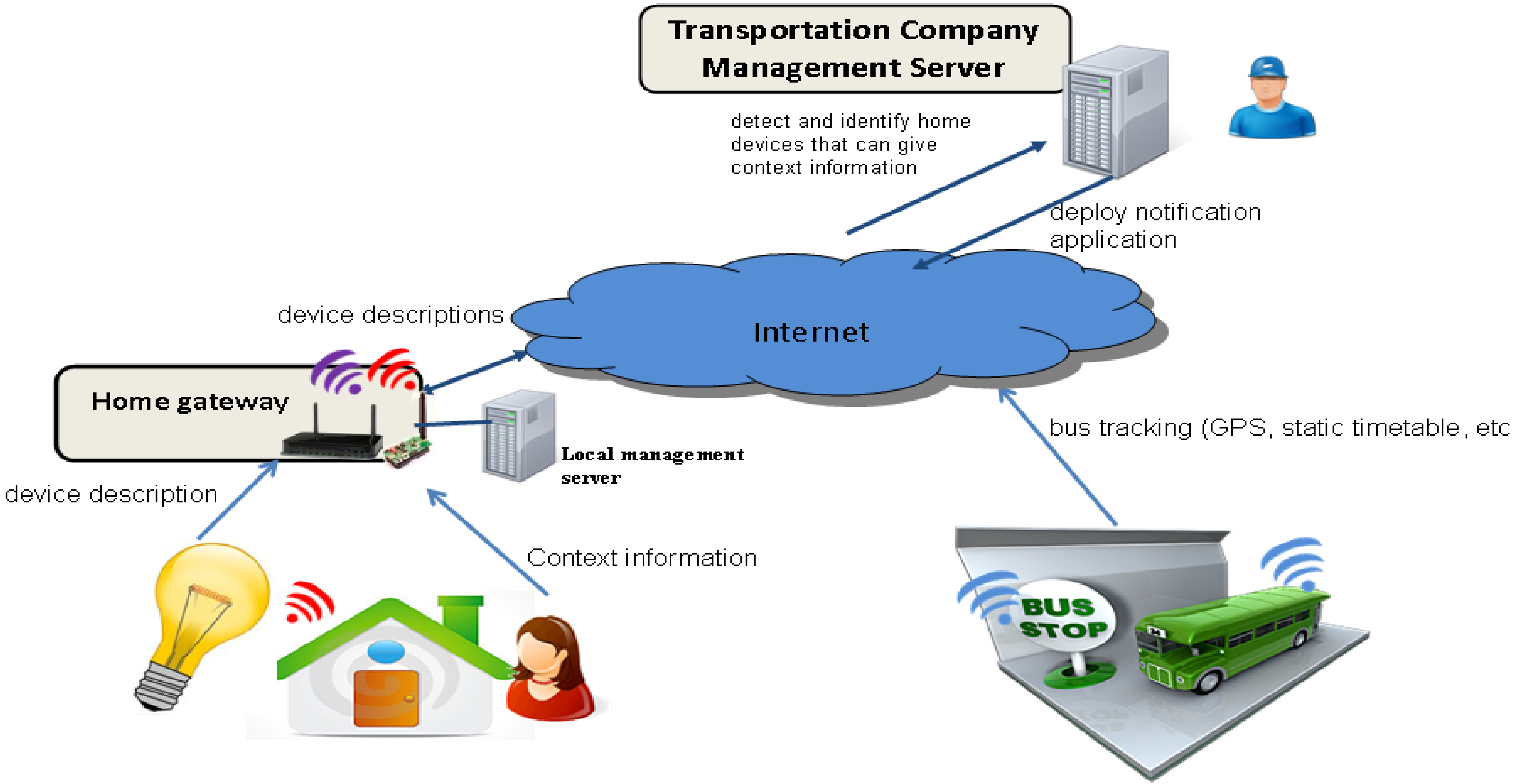
Things Connecting to Things

Source: Wei Wang, Cory Henson, Payam Barnaghi



- Complex and heterogeneous resources and networks

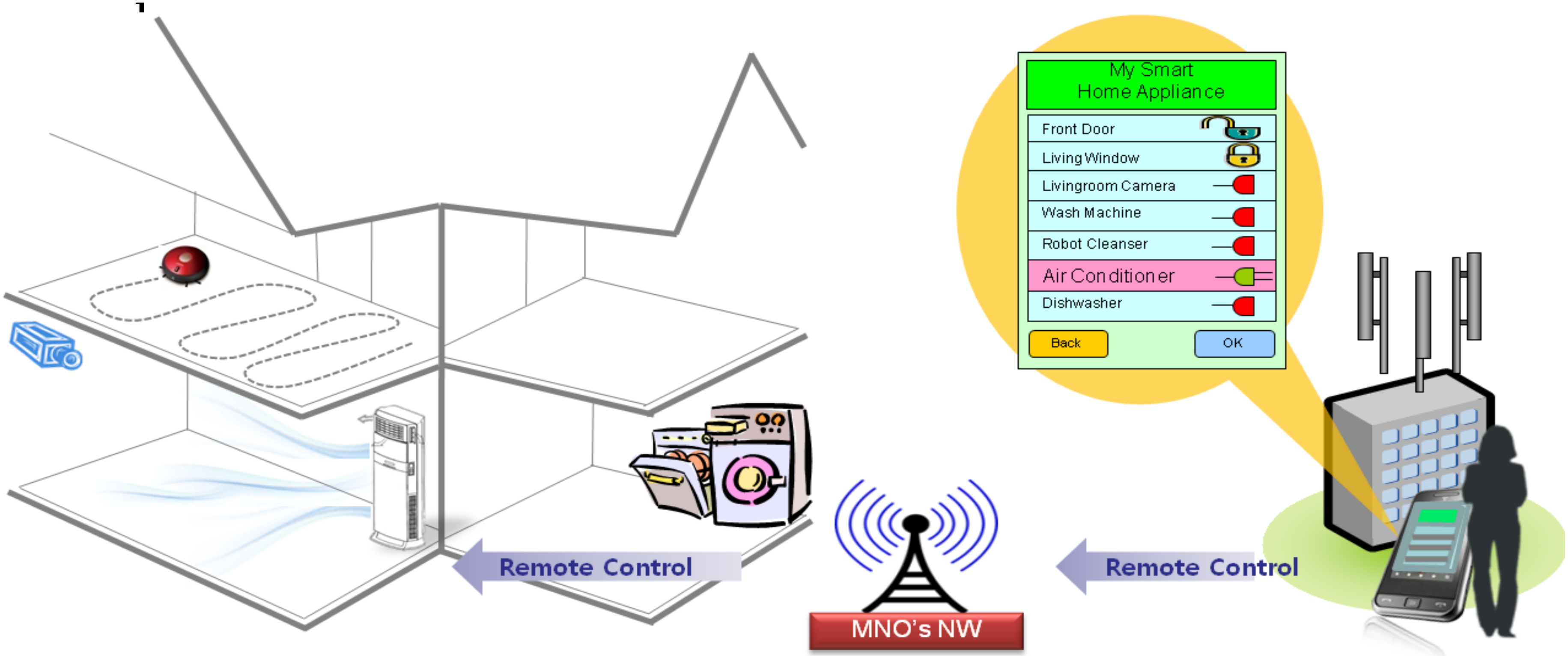
Smart Transportation



Source: EU IoT Project BUTLER

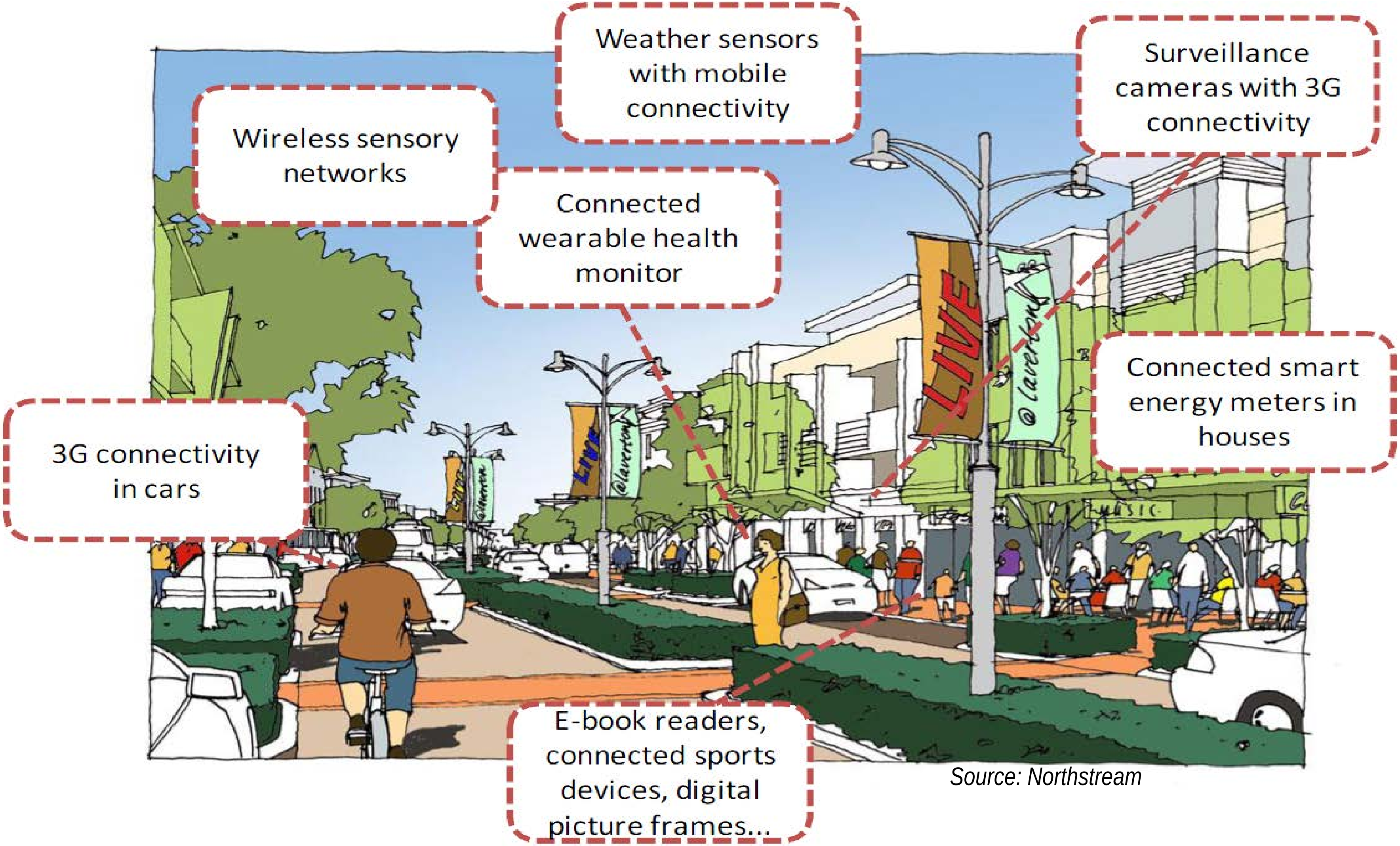
Smart Home

- A user requests a video streaming data from the surveillance camera has taken his/her puppies during he/she is not at



Source: ETSI TR 102 857 V0.3.0

Smart Cities



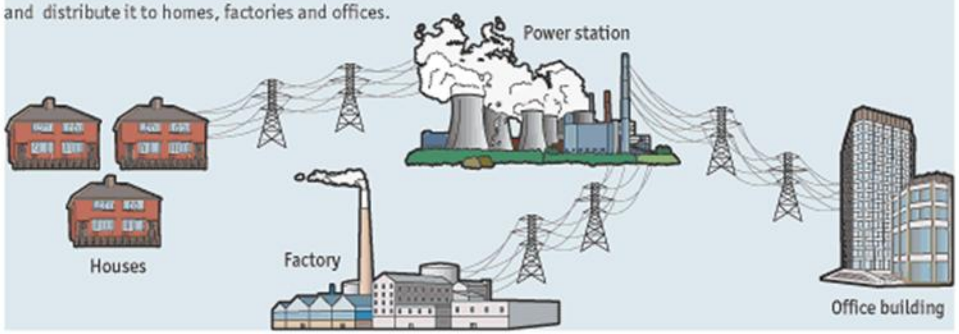
Source: Northstream

Smart Grid

The shape of grids to come?

Conventional electrical grid

Centralised power stations generate electricity and distribute it to homes, factories and offices.

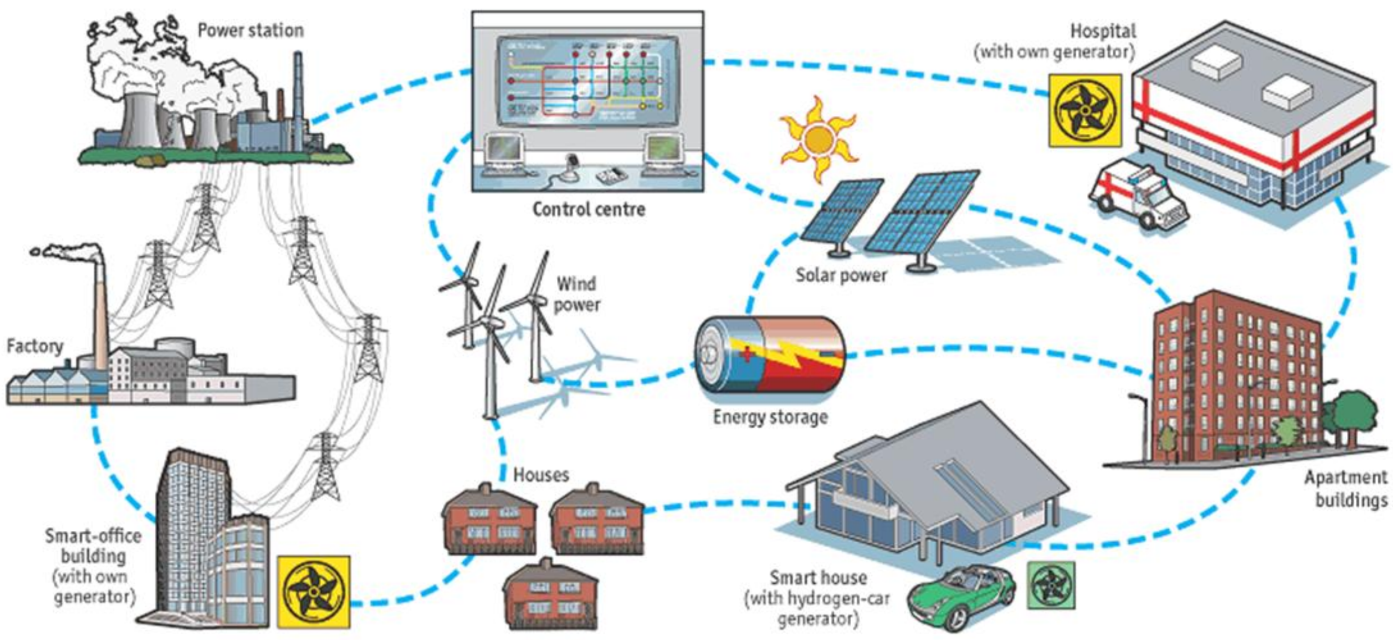


Energy internet

Many small generating facilities, including those based on alternative energy sources such as wind and solar power, are orchestrated using real-time monitoring and control systems.

Offices or hospitals generate their own power and sell the excess back to the grid. Hydrogen-powered cars can act as generators when not in use. Energy-storage technologies smooth out fluctuations in supply from wind and solar power.

Distributing power generation in this way reduces transmission losses, operating costs and the environmental impact of overhead power lines.



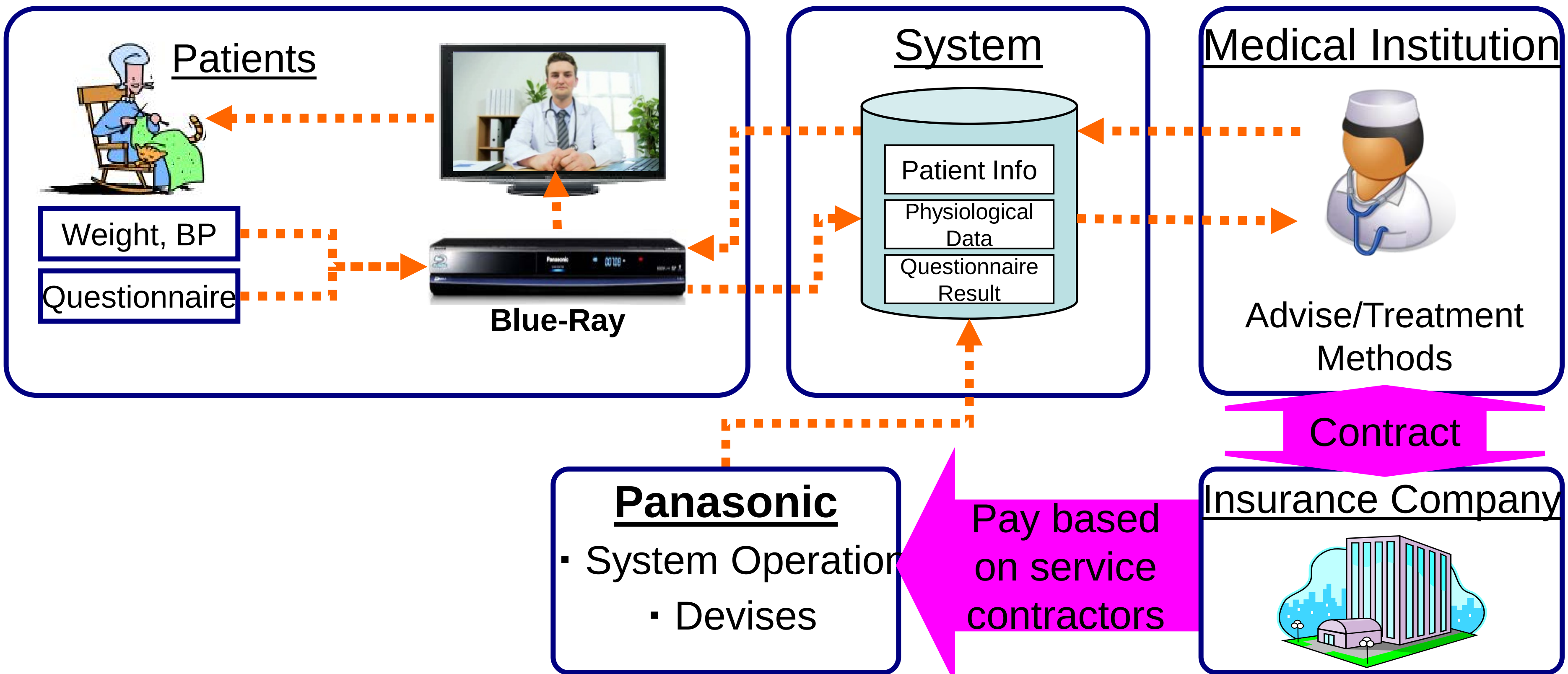
Sources: *The Economist*; ABB

Source: Economist, http://www.economist.com/node/13725843

“In-Home Chronic Disease/Wellness Management”

- Patients’ physiological data is captured and sent via Internet connected Blue-ray player and monitored by Medical institution
- Medical institution give advises or treatment methods remotely based on the data and delivered to in-home patients

“Win-Win-Win” : Panasonic, Insurance Companies and Patients

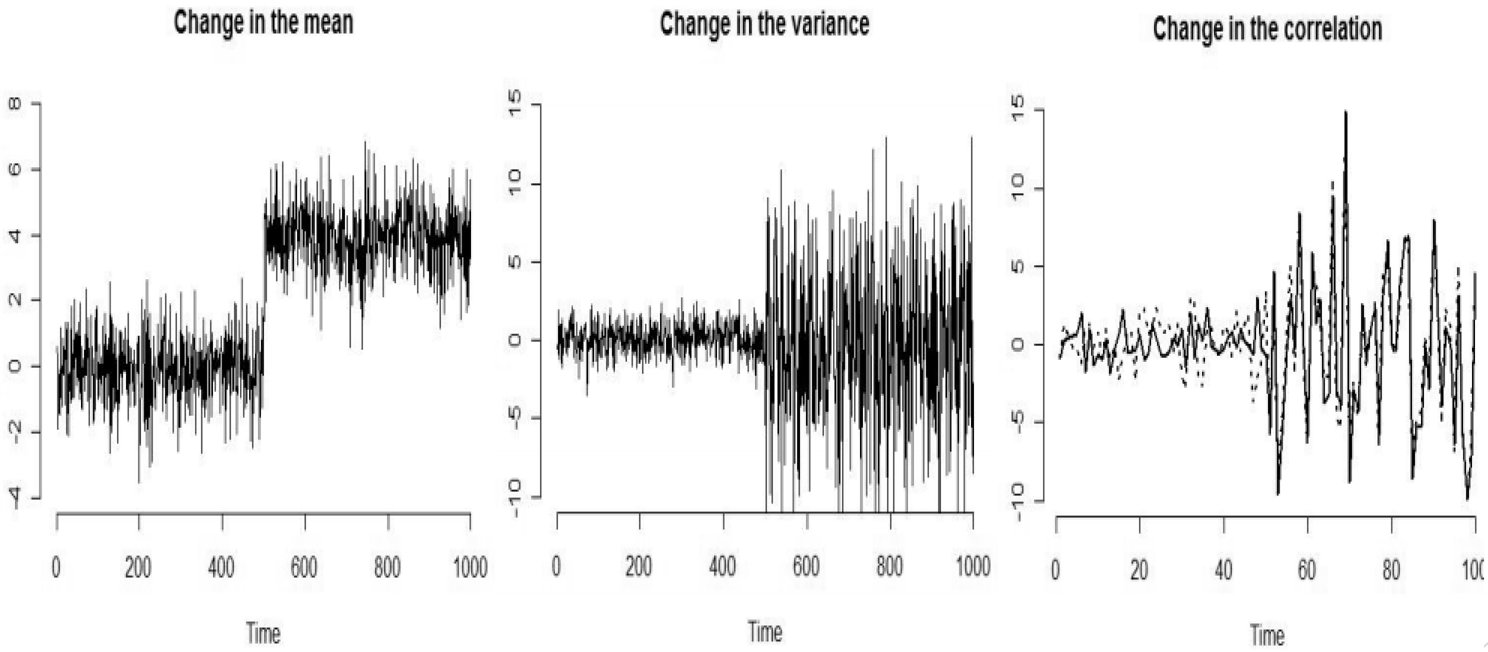


The Age of Observation: Smart Sensing, Reasoning and Decision

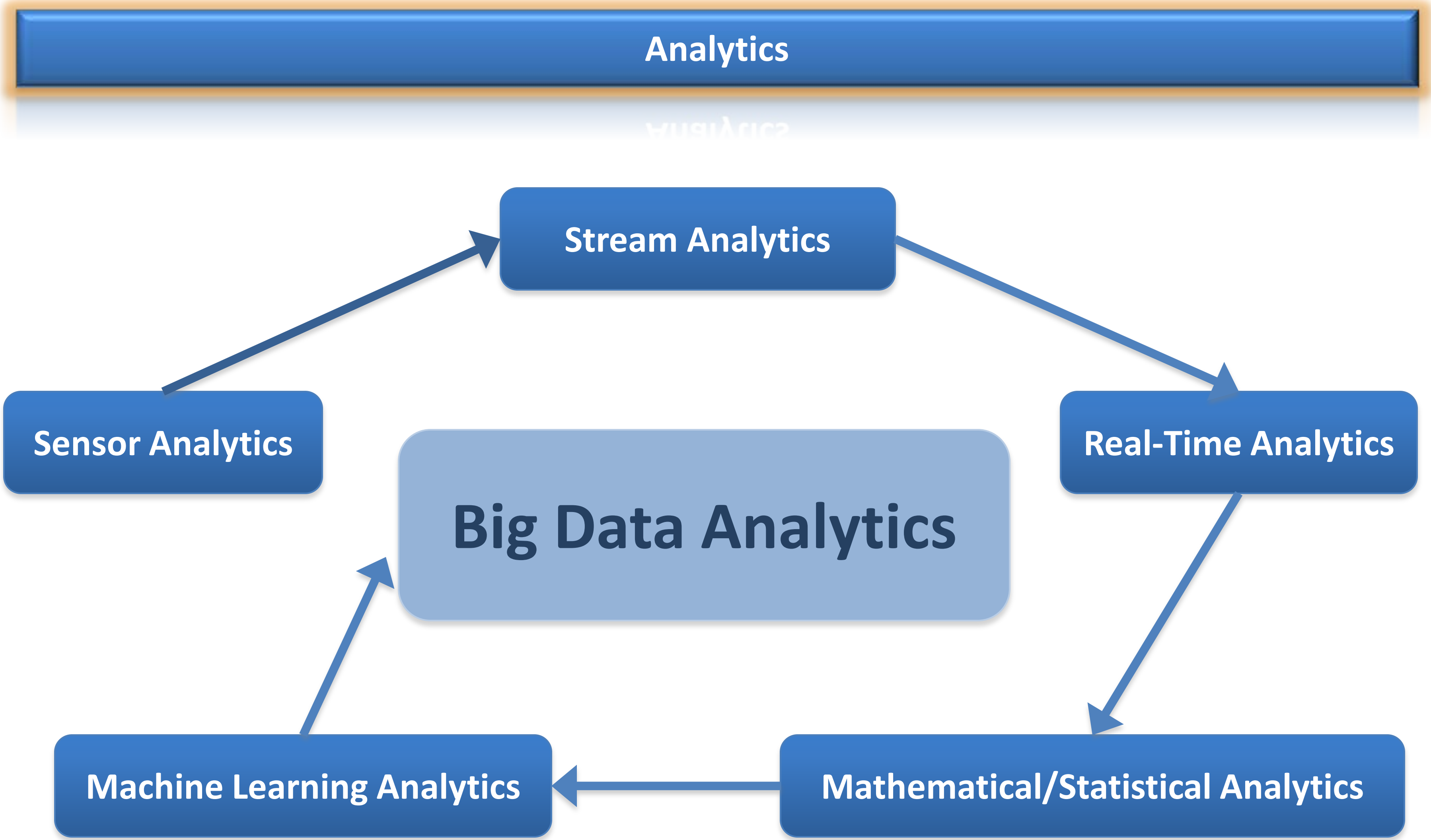
Source: Sajal Das, Keith Marzullo

Credit: Image courtesy of University of Florida

Change Detection Examples



The Technologies of IoT



Analytics:

Static Data

Streaming Data

Discovery with Data: Leveraging Statistics with Computer Science to Transform Science and Society

The most productive approach for turning data into knowledge will involve multidisciplinary teams with statistical, computational, mathematical, and scientific domain expertise

ASA Team from: MIT, Duke, Harvard, CMU, NCSU, Hopkins, Berkeley, PSU, U Washington
July 2, 2014

Tribute to Sir Ronald Aylmer Fisher
R. A. Fisher

R. A. Fisher

Statistician

Evolutionary Biologist

Geneticist

The Iris flower data set or Fisher's Iris data set

Sepal Length, Sepal Width, Petal L, Petal W, Species

The morphologic variation of Iris flowers of three related species.

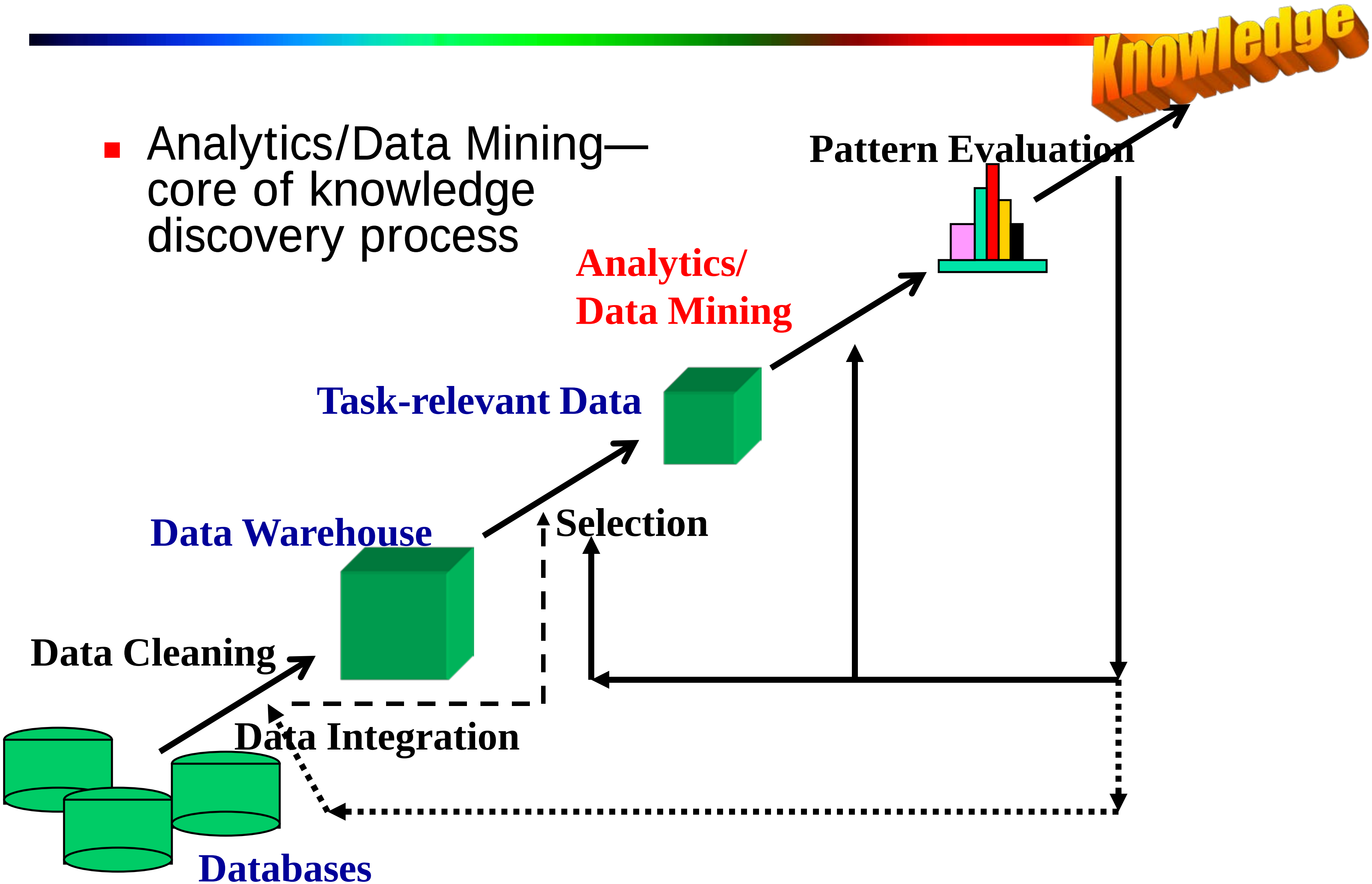
The Fisher's *Iris* Flower Data Set

Sir Ronald Fisher (1936) as an example of discriminant analysis

	Sepal Length	Sepal Width	Petal Length	Petal Width	Class
	X1	*X2*	*X3*	*X4*	*X5*
x1	6.2	3.6	4.4	1.5	Iris-Versicolor
x2	6.9	1.0	1.1	1.6	Iris-Versicolor
x3	3.3	1.6	1.4	1.5	Iris-Versicolor
x4	4.0	2.9	3.2	0.4	Iris-Setosa
x5	4.5	1.5	1.0	1.0	Iris-Versicolor
x6	4.4	2.9	5.4	0.3	Iris-Setosa
x7	4.7	1.7	5.0	0.4	Iris-Virginica
…	…	…	…	…	…
x148	5.7	3.4	1.2	1.1	Iris-Setosa
x149	3.0	3.7	1.1	1.7	Iris-Virginica
x150	6.3	3.2	5.6	1.3	Iris-Virginica

Static/Batch Data Analytics: A KDD Process

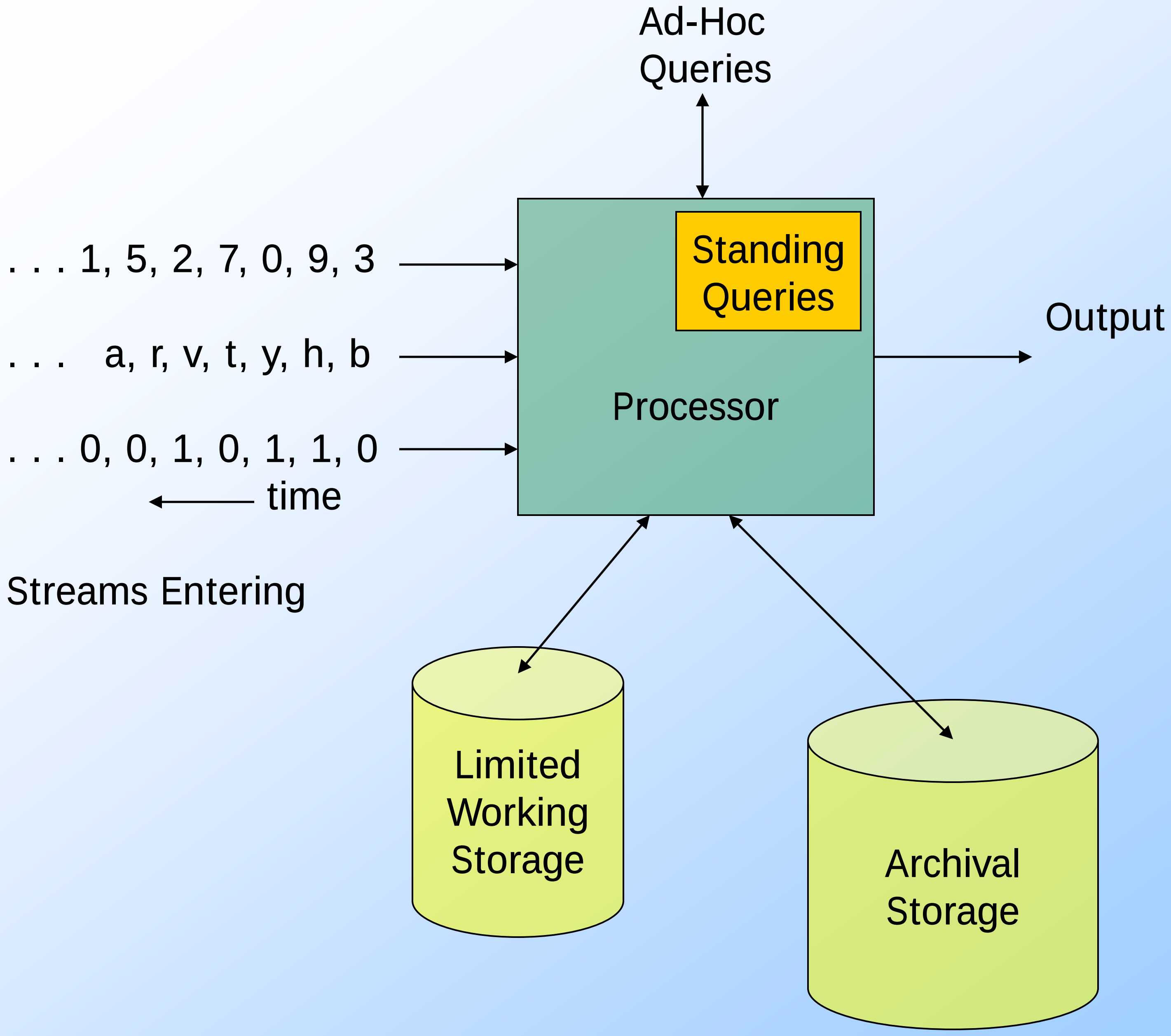
- Analytics/Data Mining—core of knowledge discovery process



Streaming Data / Data in Flight

Data arrive in streams continuously and so rapidly that it is not feasible or useful to store it in a conventional database and then analyze it at the time of our choosing, ***if it is not processed immediately, then it is lost forever*****.**

The Nature of analytics has changed


Ad-Hoc
Queries
Standing
Queries
Processor
Output
. . . 1, 5, 2, 7, 0, 9, 3
. . . a, r, v, t, y, h, b
. . . 0, 0, 1, 0, 1, 1, 0
time
Streams Entering
Limited
Working
Storage
Archival
Storage

The Stream Data Model

- Any number of streams can enter the system,
- Streams have their own schedule,
- Streams do not have same data rates or data types,
- No uniform arrival, …
- No control over the rate at which data is read from the disk,
- Standing queries: Ocean-Surface-Temperature Sensors (output an alert when ocean temperature exceeds 25 centigrade),
- Ad-hoc queries: Search Engine Query Streams (what fraction of a specified user's search queries were repeated over past 24 hours)

Data Streams

- Data arrive in a stream or streams, and if not processed immediately, then lost for ever. Data arrives so rapidly that it is not feasible to store it all in active storage and analyze later at the time of our choosing

Examples of Stream Data

- **Sensor Data:** one million sensors with GPS units deployed on surface of an ocean sending surface temperature and height every tenth of a second produces 3.5 terabytes of data every day.
- **Image Data:** satellites send down to the earth streams consisting many terabytes of data per day (London is said to have six million surveillance camera)

Data Streams

Examples of Stream Data

- **Internet and web Traffic:**

  - **a switching node receives streams of IP packets; the switch has the additional ability to detect denial-of-service or reroute packets based on information about congestion in the network**

  - Google receives several hundred million search quires per day

  - Yahoo accepts billions of "clicks" per day on its various sites

  - an increase in queries like "soar throat" enables us to track the spread of a virus

  - a sudden increase in the click rate for a link indicate some news related to that site, or link is broken needs to be repaired

Applications – (1)

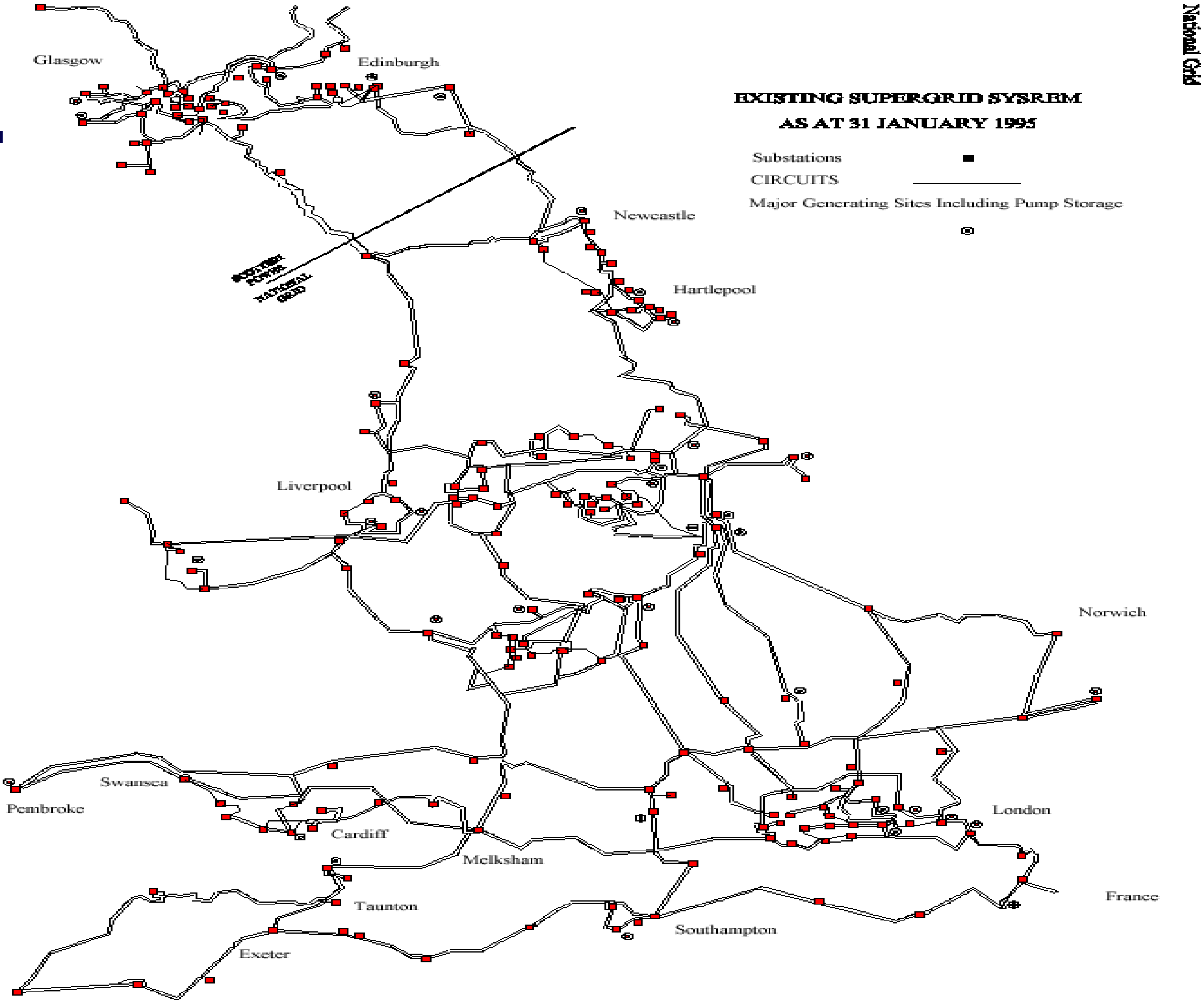
- Mining query streams.
  - Google wants to know what queries are more frequent today than yesterday.
- Mining click streams.
  - Yahoo wants to know which of its pages are getting an unusual number of hits in the past hour.

Applications – (2)

- Sensors of all kinds need monitoring, especially when there are many sensors of the same type, feeding into a central controller.
- Telephone call records are summarized into customer bills.

Applications – (3)

- IP packets can be monitored at a switch.
  - Gather information for optimal routing.
  - Detect denial-of-service attacks.


National Grid
EXISTING SUPERGRID SYSREM
AS AT 31 JANUARY 1995
Substations
CIRCUITS
Major Generating Sites Including Pump Storage
Glasgow
Edinburgh
Newcastle
Hartlepool
Liverpool
Norwich
Swansea
Pembroke
Cardiff
Melksham
London
Taunton
Southampton
France
Exeter

Electrical Grid

Sensors continuously produce streams of data at high-speed

- **Cluster Analysis (urban, rural, industrial, etc.)**
- **Predictive Analysis (each sensor, each time horizon, peek on demand, …)**
- **Change Detection (failures, abnormal activities)**
- **Extreme Values: Anomaly, Outliers, ..**

Cyber Security

When a fast scanning worm propagates through the Internet, the propagation activity look like this: a smaller number of infected hosts tries to find other hosts to be infected by attempting to connect to them in a *random* fashion. This traffic is different from *normal Internet traffic.*

Wagner and Platter , 2005

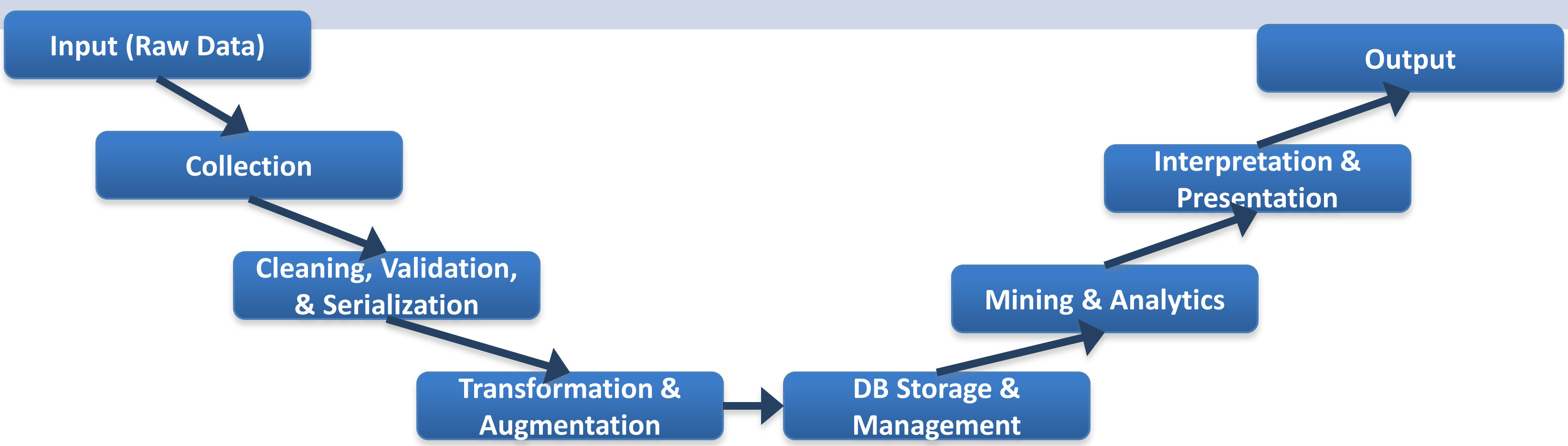
- **Changes in the entropy indicate a massive network event**
- **Source IP address fields will contain less entropy per address than normal traffic**
- **Target IP address will have more entropy than normal traffic**

Analytics
Static Data vs. Streaming Data

Static Data	Streaming Data/Data in Flight
Multiple Passes	Single Pass
Persistent / Relatively Stable Data	Inherently Temporal/Comtinouly Eveloving
Offline Analytics	Online as well as Offline Analytics
Analytics Based on All the Data	Analytics Based on a Subset of Data
Only the current state is relevant	Consideration of the order of the input
Relatively low update rate	Potentially extremely high update rate
Little or no time requirements	Real-time requirements
Assumes exact data	Assumes outdated/inaccurate data
Plannable query processing	Variable data arrival and data characteristics

Big Data Challenges, & Data Life Cycle

Big Data Challenges
• Sensor data brings numerous challenges with it in the context of **data collection**, **storage** and **processing**. This is because sensor data processing often requires **efficient in-network and real-time data stream processing** from **massive volumes** of possibly **uncertain data** from **various sources**. The data generated from these sensors arrives in the form of streams.
• At every phase of the big data life cycle, there are research issues (i.e. challenges) along each step in the data life cycle .
• To handle these streaming sensor data **model-based techniques** are employed, such as: statistical, signal processing, regression-based, machine learning, probabilistic, or time series.



Static Processing vs. Stream Processing

	Static	Stream
Number of Passes	Multiple	Single
Processing Time	Unlimited	Restricted
Memory Usage	Unlimited	Restricted
Type of Result	Accurate	Approximate
Distributed	No	Yes

Management DBMS vs. DSMS

DBMS (Database Management System)	DSMS (Data Stream Management System)
Persistent relational data	Volatile transient data streams
Random access	Sequential access
One-time queries	Continuous queries
Unlimited secondary storage (theoretically)	Limited main memory
Only the current state is relevant	Consideration of the order of the input
Relatively low update rate	Potentially extremely high update rate
Little or no time requirements	Real-time requirements
Assumes exact data	Assumes outdated/inaccurate data
Standing queries	Ad-hoc queries

Big Data Analytics: Sensors & Data Streams

Sensors & Data

Let us denote a sensor network as $S = \{s_j \mid 1 \le j \le m\} =$ *Sensor network consisting of sensors* s_j, *where* $j = (1, \ldots, m)$.

$s_j =$ *Sensor identifier for a sensor in* S.

$v_{ij} =$ *Sensor value observed by the sensor* s_j *at time* t_i, *such that* $v_{ij} \in R$, *the real numbers.*

$v_i =$ *Row vector of all sensor values observed at time* t_i, *such that* $v_i \in R^m$.

$V_{ij} =$ *Random variable associated with the sensor value* v_{ij}.

Data Streams

Let us denote a data stream as $D_i = \{(t_i, v_{ij})\}$, *an ordered sequence of data tuples, where* v_{ij} *is the sensor value at time* t_i.

A data stream is a structured tuple composed of time (implicit *or* explicit) *and sensor values.*

v_{ij}, *defined above, is a data stream element.*

v_i, *defined above, is a data stream.*

Big Data Analytics: Sensors & Data Streams

Data Streams

- The general data stream model can be defined as, an infinite tuple of time and values.

- In the example here the stream is in tabular form for easy readability, time is defined implicitly by *index i and* explicitly by time t_i , sensor identifiers are either the sensor ID s_j or the sensor spatial coordinates x_j *(i.e. longitude) and* y_j(i.e. latitude), and the sensor reading values v_{ij}.

i	t_i	s_j	x_j	y_j	v_{ij}
1	09:00	1	1.5	4.0	2.3
1	09:00	2	2.0	3.0	0.1
1	09:00	3	4.5	0.5	2.7
1	09:00	4	3.5	3.5	3.1
2	09:15	1	1.5	4.0	2.5
2	09:15	2	2.0	3.0	7.2
…	…	…	…	…	…

Streaming Data Analytics

- **Many existing data mining methods cannot be applied directly to data streams**
- **Popular existing algorithms, designed for and tested with conventional static data, have to be adapted/modified for streaming data analytics**
- **New emerging techniques are being innovated for management and mining of streaming data**

Streaming Data Analytics

Popular Existing Algorithms

- **Unsupervised: Clustering**

  Micro-Clustering Algorithm

  - Step 1: on-line, Micro–Clusters (utilizes Cluster Feature, CF)
  - Step 2; off-line, Macro-Clusters

  Very efficient, enables tracking cluster evolution

- **Supervised: Classification/Decision Tree Algorithms**

  C4.5 (Quinlan), CART (Breiman, et al.) algorithms

  - **VFDT Algorithm: Very Fast Decision Tree (Hoeffding Tree)**

    Waits for overall statistical evidence in favor of a split (not all data in a given node)

  - **OLIN Info-Fuzzy Network Algorithm**

    IFN algorithm applied to sliding window (most recent training cases), dynamically adjusts window size

Streaming Data Analytics

New Emerging Techniques

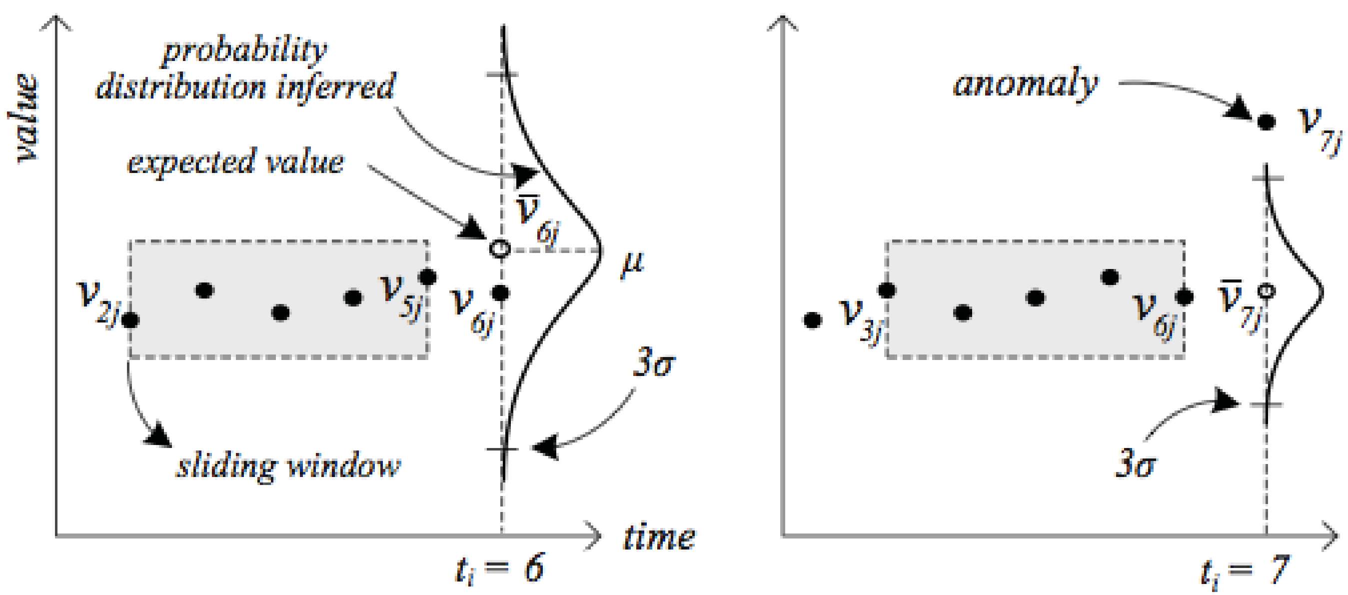
- Sampling
- Filtering
- Wavelets
- Synopsis and Histograms
- Entropy of a Stream Computation
- Sliding Window
- Change Detection/Anomaly Detection
- Outliers
- Moments Estimation
- Bounds and Confidence Limits

Big Data Analytics & Techniques

Example: Model-Based Techniques – The Kalman Filter

Probabilistic Models: In sensor data cleaning, inferring sensor values is perhaps the most important task, since systems can then detect and clean dirty sensor values by comparing raw sensor values with the corresponding inferred sensor values.

- The **Kalman filter** is perhaps one of the most common probabilistic models to compute inferred values corresponding to raw sensor values.



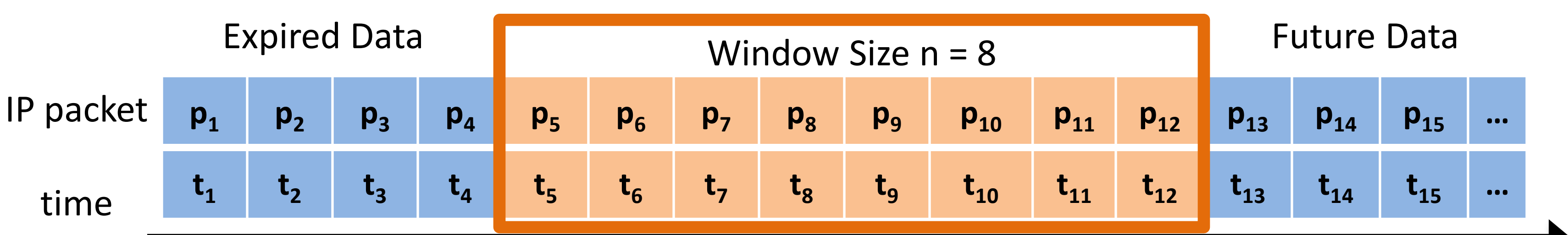
The Data Stream & Sliding Window Models

The Data Stream Model
In the data stream model, some or all of the input data that are to be operated on are not available for random access from disk or memory, but rather arrive as one or more *continuous data streams*. Data streams differ from the conventional stored relation model in several ways:
• The data elements in the stream arrive online.
• The system has no control over the order in which data elements arrive to be processed, either within a data stream or across data streams.
• Data streams are potentially unbounded in size.
• Once an element from a data stream has been processed it is discarded or archived — it cannot be retrieved easily unless it is explicitly stored in memory, which typically is small relative to the size of the data streams.
Example:

IP packet	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	p_{11}	p_{12}	p_{13}	p_{14}	p_{15}	…
time	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	t_{15}	…

The Data Stream & Sliding Window Models

The Sliding Window Model
In the sliding window model, only the *recent* past is the objective concern of stream processing. The fundamental sliding windows are of fixed size, which are similar to first-in, first-out data structures.
• The input is still a stream of data values or elements.
• A data value/element arrives at each time instant; it later *expires* after a number of time stamps equal to the *window size n*.
• The *current window* at any time instant is the set of data elements that have not yet expired.
Example: Fixed Size Sliding Window.



So, in 2014, What are the Streaming Data Analytics Problems?

- Number of Data Points: Potentially Unlimited

  ➡ Analysis Cannot be Based on all Data Points

- Memory Usage: Limited
- Processing Time: Limited, Near-Real-Time
- Accuracy: No Need for Exact Accurate Result; Good Approximation/Estimation Works

So, in 1930s, What Were the Data Scientists (Fisher, Neyman, Pearson, ...) Problems?

- **Number of Data Points: impossible to observe, measure, and collect data for all units/elements of interest - *complete enumeration* not possible**
- **Computation & Memory Capabilities: Very Limited, impossible to analyze and process all the data**
- **Processing Time: Very Long**
- **Accuracy: ???**

So, What Did the
Data Scientists of the Last Century Do?

- **Sampling Theory:**
  - **Units, Measurements, ..**
  - **Time, Sliding Window**
  - **Many Sampling Designs**
- **Design of Experiments**
- **Sufficient Statistics**
- **Estimation Theory and Confidence Limits**
- **Test of Hypothesis**
- **…**

Sampling Accuracy* often better than *Complete Enumeration

All these Techniques are Applicable to the 2014 Streaming Data Analytics

Discovery with Data: Leveraging Statistics with Computer Science to Transform Science and Society

The Nature of analytics has changed

The most productive approach for turning data into knowledge will involve multidisciplinary teams with statistical, computational, mathematical, and scientific domain expertise (and COMMINICATIONS)

ASA Team from: MIT, Duke, Harvard, CMU, NCSU, Hopkins, Berkeley, PSU, U Washington (July 2, 2014)

Summary and Conclusion

Research Opportunities

The Stream Data Analytics domain is wide open. Great research opportunities:

- **Sampling Designs**
- **Sampling for Time Series**
- **Statistical Inference and Estimations**
- **Sufficient Statistics**
- **Outliers theory for Change & Anomaly Detection**
- **........**
- **The first Textbook on Stream Data Analytics!**

Questions?

References

Resources

Massive Data Analysis
http://dimacs.rutgers.edu/~graham/

Distributed Data Mining Maintained by Hillol Kargupta

UCR Time-Series Data Sets
Maintained by Eamonn Keogh, UCR, US
http://www.cs.ucr.edu/~eamonn/time_series_data

Mining Data Streams Bibliography Maintained by Mohamed Gaber
http://www.csse.monash.edu.au/~mgaber/WResources.html

Data Stream Management Systems

Niagara (OGI/Wisconsin) – Internet XML databases

Aurora (Brown/MIT) – sensor monitoring, dataflow
http://www-db.stanford.edu/sdt

Stream (Stanford) – general-purpose DSMS
http://www-db.stanford.edu/stream/index.html

COUGAR (Cornell)

GigaScope and Hancock (At&T) Medusa (Brown University)

Master References

J. Gama, *Knowledge Discovery from Data Streams*, CRC Press, 2010.

S. Muthukrishnan *Data Streams: Algorithms and Applications*, Foundations & Trends in Theoretical Computer Science, 2005.

B.Babcock, S. Babu, M. Datar, R. Motwani, and J. Widom. Models and Issues in Data Stream Systems, in Proc. PODS, 2002.

Gaber, M, M., Zaslavsky, A., and Krishnaswamy, S., Mining Data Streams: A Review, in ACM SIGMOD Record, Vol. 34, No. 1, 2005.

C.Aggarwal, *Data Streams: Models and Algorithms*, Ed. Charu Aggarwal, Springer, 2007

J. Gama, M. Gaber (Eds), *Learning from Data Streams – Processing Techniques in Sensor Networks*, Springer, 2007.

Sampling and Synopsis

Cormode & Muthukrishnan. *An improved data stream summary: The count-min sketch and its applications*. Journal of Algorithms, 2005.

A. Arasu, G. Manku, *Approximate Counts and Quantiles over Sliding Windows*, in PODS 2004.

M. Datar, A. Gionis, P. Indyk, R. Motwani; *Maintaining Stream Statistics over Sliding Windows*, in the ACM-SIAM Symposium on Discrete Algorithms (SODA) 2002.

J. Vitter.*Random sampling with a reservoir*, ACM Transactions on Mathematical Software, 1985.

A. Broder, M. Charikar, A. Frieze, and M. Mitzenmacher; *Min-wise independent permutations*, Journal of Computer and System Sciences, 2000

A. Chakrabarti and G. Cormode and A. McGregor, *A Near-Optimal Algorithm for Computing the Entropy of a Stream*, SIAM, 2007

J. Gama, C.Pinto, *Discretization from data streams: applications to histograms and data mining*. SAC 2006.

Bibliography on Frequent Item's

Probabilistic Counting Algorithms for DataBase Applications, Flajolet and Martin; JCSS, 1983

Finding repeated elements, J. Misra and D. Gries. Science of Computer Programming, 1982.

What's Hot and What's Not: Tracking Most Frequent Items Dynamically, by G. Cormode, S. Muthukrishnan, PODS 2003.

Dynamically Maintaining Frequent Items Over A Data Stream, by C. Jin, W. Qian, C. Sha, J. Yu, A. Zhou; CIKM 2003.

Processing Frequent Itemset Discovery Queries by Division and Set Containment Join Operators, by R. Rantzau, DMKD 2003.

Approximate Frequency Counts over Data Streams, by G. Singh Manku, R. Motawani, VLDB 2002.

Finding Hierarchical Heavy Hitters in Data Streams, by G. Cormode, F. Korn, S. Muthukrishnan, D. Srivastava, VLDB 2003.

A Geometric Approach to Monitoring Distributed DataStreams, I. Sharfman, A. Schuster, D. Keren, SIGMOD 2006

References

1. D. Belanger, "It's About the Data: A Decade+ Experiment in "Big Data", April 2012, http://www.ieee-noms.org/keynotes.html
2. M. Daneshmand, "Intelligent Network Operations and Management – it's about the Data", July 2011, http://www.ieee-iscc.org/2011/speakers.htm
3. F. Jahanian, "Innovating for Society: Realizing the Promise of Computing and Communications, National Science Foundation", December 2011, http://www.ieee-globecom.org/speakers.html
4. R.C. Johnson, "Dialing in the value of Big Data: The intangible assets of the Internet of Things" , EE Times, issue 1620, April 2012, WWW.EETIMES.COM
5. S. Lohr, "The Age of Big Data", February 11, 2012, The New York Times
6. J. Manyika, etal "Big Data: the Next Frontier of innovation, Competition, and Productivity", McKinsey Global Institute, May 2011.
7. C. Volinsky, "Recommender Systems for Fun and Profit", April 2009, Stevens Institute of Technology
8. Press Release, the White House Office of Science and Technology Policy (OSTP), March ,29, 2012, whitehouse.gov/ostp
9. KDnuggets, June 2012

References: Timeline

Year	Event
1834	An electromagnetic telegraph was created by Baron Schilling in Russia, and in 1833 Carl Friedrich Gauss and Wilhelm Weber invented their own code to communicate over a distance of 1200m within Göttingen, Germany.
1844	Samuel Morse sends the first morse code public telegraph message "What hath God wrought?" from Washington, D.C. to Baltimore.
1926	Nikola Tesla in an interview with Colliers magazine: "When wireless* is perfectly applied the whole earth will be converted into a huge brain, which in fact it is, all things being particles of a real and rhythmic whole.........and the instruments through which we shall be able to do this will be amazingly simple compared with our present telephone. A man will be able to carry one in his vest pocket." *not the 802.11 version :)
1950	1950: Alan Turing in his article Computing Machinery and Intelligence in the Oxford Mind Journal (Via @Kevin_Ashton) "...It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English. This process could follow the normal teaching of a child."
1964	In Understanding Media Marshall McLuhan stated: "....by means of electric media, we set up a dynamic by which all previous technologies -- including cities -- will be translated into information systems"

References: Timeline

Year	Event
1966	Karl Steinbuch a German computer science pioneer said "In a few decades time, computers will be interwoven into almost every industrial product".
1969	Arpanet
1974	Beginnings of TCP/IP
1984	Domain Name System is introduced
1989	Tim Berners-Lee proposes the World Wide Web
1990	John Romkey created the first Internet ‘device’, a toaster that could be turned on and off over the Internet. At the October '89 INTEROP conference, Dan Lynch, President of Interop promised Romkey that, if Romkey was able to "bring up his toaster on the Net," the appliance would be given star placement in the floor-wide exhibitors at the conference. The toaster was connected to a computer with TCP/IP networking. It then used an information base (SNMP MIB) to turn the power on. (See also: Xerox PARC networked coke machine)
1991	The first web page was created by Tim Berners-Lee
1991	Mark Weiser's Scientific American article on ubiquitous computing called ‘The Computer for the 21st Century’ is written. “The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it”.

References: Timeline

Year	Event
1993	Created by Quentin Stafford-Fraser and Paul Jardetzky the Trojan Room Coffee Pot was located in the 'Trojan Room' within the Computer Laboratory of the University of Cambridge and was used to monitor the pot levels with an image being updated about 3x a minute and sent to the buildings server. It was later put online for viewing once browsers could display images.
1994	Steve Mann creates WearCam.
1995	The Internet goes commercial with Amazon and Echobay (Ebay).
1997	Paul Saffo's prescient article "Sensors: The Next Wave of Infotech Innovation"
1998	Google is incorporated
1998	inTouch a project at MIT was developed by Scott Brave, Andrew Dahley, and Professor Hiroshi Ishii (Via: @mrosenblatt) "....We then present inTouch, which applies Synchronized Distributed Physical Objects to create a "tangible telephone" for long distance haptic communication." - Original paper (PDF), Video
1998	A year before losing his battle to cancer Mark Weiser continues his explorations into the topic and constructed a water fountain outside his office whose flow and height mimicked the volume and price trends of the stock market. "Ubiquitous computing is roughly the opposite of virtual reality," Weiser wrote "Where virtual reality puts people inside a computer-generated world, ubiquitous computing forces the computer to live out here in the world with people."

References: Timeline

Year	Event
1999	A big year for the IoT and MIT The Internet of Things term is coined by Kevin Ashton executive director of the Auto-ID Center: "I could be wrong, but I'm fairly sure the phrase "Internet of Things" started life as the title of a presentation I made at Procter & Gamble (P&G) in 1999. Linking the new idea of RFID in P&G's supply chain to the then-red-hot topic of the Internet was more than just a good way to get executive attention. It summed up an important insight which is stil often misunderstood." - Full article
1999	Neil Gershenfeld was speaking about similar things from the MIT Media Lab in his book When Things Start to Think and when establishing the Center for Bits and Atoms in 2001 “in retrospect it looks like the rapid growth of the World Wide Web may have been just the trigger charge that is now setting off the real explosion, as things start to use the Net.”
1999	Auto-ID Labs opens which is the research-oriented successor to the MIT Auto-ID Center, originally founded by Kevin Ashton, David Brock and Sanjay Sarma. They helped develop the Electronic Product Code or EPC, a global RFID-based item identification system intended to replace the UPC bar code.
1999	Neil Gross in Business Week. "In the next century, planet earth will don an electronic skin. It will use the Internet as a scaffold to support and transmit its sensations. This skin is already being stitched together. It consists of millions of embedded electronic measuring devices: thermostats, pressure gauges, pollution detectors, cameras, microphones, glucose sensors, EKGs, electroencephalographs. These will probe and monitor cities and endangered species, the atmosphere, our ships, highways and fleets of trucks, our conversations, our bodies--even our dreams." - Full Article
2000	Starting off what is now becoming a meme, LG announces it's first Internet refrigerator plans.

References: Timeline

Year	Event
2002	The Ambient Orb created by David Rose and others in a spin-off from the MIT Media Lab is released into the wild (and is still on the market) with NY Times Magazine naming it as one of the Ideas of the Year. The Orb monitors the Dow Jones, personal portfolios, weather and other data sources and changes its color based on the dynamic parameters.
2003	The term is mentioned in main-stream publications like The Guardian, Scientific American and the Boston Globe. - Projects like Cooltown, Internet0, and the Disappearing Computer initiative seek to implement some of the ideas, and the Internet of Things term starts to appear in book titles for the first time. - RFID is deployed on a massive scale by the US Department of Defense in their Savi program and Walmart in the commercial world.
2005	The IoT hit another level when the UN's International Telecommunications Union ITU published its first report on the topic. "A new dimension has been added to the world of information and communication technologies (ICTs): from anytime, any place connectivity for anyone, we will now have connectivity for anything. Connections will multiply and create an entirely new dynamic network of networks – an Internet of Things"
2005	Ahead of its time, the Nabaztag (Now a part of Aldebaran Robotics) was originally manufactured by the company Violet and created by Rafi Haladjian and Olivier Mével. The little WiFi enbabled rabbit was able to alert and speak to you about stock market reports, news headlines, alarm clock, RSS-Feeds, etc as well as connect to each other (see: Nabaztag opera). The statement was "if you can even connect rabbits, then you can connect anything" (via @inakivazquez)
2006	Recognition by the EU, and the First European IOT conference is held

References: Timeline

Year	Event
2008	A group of companies launched the IPSO Alliance to promote the use of Internet Protocol (IP) in networks of "smart objects" and to enable the Internet of Things. The IPSO alliance now boasts over 50 member companies, including Bosch, Cisco, Ericsson, Intel, SAP, Sun, Google and Fujitsu.
2008	The FCC voted 5-0 to approve opening the use of the 'white space' spectrum.
2008	2008-2009: The Internet of Things was "Born". According to Cisco Internet Business Solutions Group (IBSG), the Internet of Things was born in between 2008 and 2009 at simply the point in time when more "things or objects" were connected to the Internet than people.
2008	U.S. National Intelligence Council listed the Internet of Things as one of the 6 "Disruptive Civil Technologies" with potential impacts on US interests out to 2025.
2010	Chinese Premier Wen Jiabao calls the IOT a key industry for China and has plans to make major investments in it.
2011	IPV6 public launch - The new protocol allows for 2128 (approximately 340 undecillion or 340,282,366,920,938,463,463,374,607,431,768,211,456) addresses or as Steven Leibson put it, "we could assign an IPV6 address to every atom on the surface of the earth, and still have enough addresses left to do another 100+ earths." - Cisco, IBM, Ericsson produce large educational and marketing initiatives on the topic. - Arduino and other hardware platforms mature and make the IoT accessible to DIY'ers taking interest in the topic. - Acquisitions and VC investment in the space including the IoT platform Pachube being aquired, IoT security company Mocano raising a round of funding and other VC's taking notice of the industry.