Big Data Why and Where Big Data will Matter

IOT WEEK 2017 – GENEVA JUNE 6-9

7th international IoT Week conference OVER 200 IoT SESSIONS AND ACTIVITIES

Day 2 Session: 16:15-17:15 Room 2

Dr. Martin Serrano
Principal Investigator and Data Scientist
Insight Centre for Data Analytics, Ireland
<martin.serrano@nuigalway.ie>
Dr. Martin Serrano
IoT & Stream Processing Unit Head
Chair IEEE ComSoC IoT Experimentation
OASC Board Member, Ireland

IEEE ComSoC
Emerging Technologies Chapter
Sub-Committee Internet of Things IoT Experimentation

siliconrepublic

2016

25 key people influencing the internet of things
by John Kennedy

Irish and Ireland-based leaders, scientists and
technologists are putting the country on the global map
in terms of the internet of things (IoT) revolution.

NIST GCTC Smart Cities Project, Technical Coordinator, USA
R+D+I Advisor, Dew Mobility, Fremont, Ca USA
Santa Clara University Lecturer, Silicon Valley, USA

2015

Research Excellence
President's Award Nominee SFI-NUIG, Ireland
California State University Lecturer, San Luis Obispo (CalPoly), USA
MIT-IoT Hackaton
IoT Best Industry Solution
IoT Media Lab, Cambridge, Ma. U.S.A

2014

Industry

IoT Scientific Director, Galway, Ireland
NUIG-National University of Ireland

2013

Industry

Irish Software Association
Software Industry Awards outstanding Academic Achievement Nominee, Ireland

NATIONAL Panasonic
Kumamoto, Japan

Design Engineer Supervisor, AKME-BC

© Copyright 2017 Insight Centre for Data Analytics Galway. All rights reserved.
The Internet of Things and Unit

Insight Centre for Data Analytics

Smart Cities, Crowd Sensing, LSD Services and Applications

Semantic Web

Big Data

Interoperability

Data Streams

Technology

NATIONA UNIVERSITY OF IRELAND
GALWAY CITY, IRELAND

Insight

Internet

Cloud Computing

Systems

Stream Processing

Applications

Devices
Big Data and IoT

...Big Data

How big is Big Data?

Production of Big Data

IoT Big Data for Healthcare...

Conclusions
Big Data

Big Data is not 'just' data, there are a few new considerations

**Volume**
- Data at Rest
  - Terabytes to exabytes of existing data to process

**Velocity**
- Data in Motion
  - Streaming data, milliseconds to seconds to respond

**Variety**
- Data in Many Forms
  - Structured, unstructured, text, multimedia

**Veracity**
- Data in Doubt
  - Uncertainty due to data inconsistency & incompleteness, ambiguities, late voy, deception, model approximations

‘Big data’ is defined by IBM as any data that cannot be captured, managed and/or processed using traditional data management components and techniques.

© 2014 IBM Corporation
IoT: One Paradigm, Different Visions

Diagram adapted from L. Atzori et al., 2010, “the Internet of Things: a Survey”
Internet usage growth vs Data Production

Device Data production in the world
(Number of Sensor/Devices in Millions)

How Big is Big Data?

The units for measuring Big Data

“The amount of data stored in the U.S. in 2010... about 3,500 Petabytes is equal to the big data storage for all the rest of the world combined.” (A petabyte is 1 million gigabytes)

Sensing as a Service and Big Data,” Arkady Zaslavsky, Charith Perera, and Dimitrios Georgakopoulos, Research School of Computer Science at The Australian National University.

Information from the Internet of Things:
We have gone beyond the decimal system

Today data scientist uses Yottabytes to describe how much government data the NSA or FBI have on people altogether.

In the near future, Brontobyte will be the measurement to describe the type of sensor data that will be generated from the IoT (Internet of Things).

1 Petabyte = 1 000 000 000 000 000 bytes
1 Terabyte = 1 000 000 000 000 bytes
1 Gigabyte = 1 000 000 000 bytes
1 Megabyte = 1 000 000 bytes
# Big Data Readers

## 1 TERA byte = 1,000 gigabytes

### Word Document Pages

85,899,345 pages of Word documents would fill one terabyte

<table>
<thead>
<tr>
<th>Book title</th>
<th>Author</th>
<th>Edition/publisher</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>volumes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Page size</th>
<th>Word count</th>
<th>Language</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.0 inches (17.8 cm) x 5.0 inches (12.7 cm)</td>
<td>1,267,069</td>
<td>French</td>
<td>Guinness World Record Longest Novel</td>
</tr>
</tbody>
</table>

85,899,345 pages / 3031 pages = 28340.26 Guinness novels can be stored

A long book of 1214 pages which means it could archive about 70,757 similar size books.
It can read in Two months, So the entire library may take 141,514 months or 11,792 years.

"If a human read one book per week during 100 years would end reading only 5200 books"
Music Files
Assuming that an average song takes up five megabytes, one terabyte could fit approximately 200,000 songs or 17,000 hours of music.

60min / 3 min average song = 20 Songs / hour then
24 hours = 480 songs a day,
480 songs x 365 Days = 175,200 Songs

“Assuming a human non-stop listening songs today 1T = 1+ year Continuously”

Movies
You could fit approximately 500 hours worth of movies on one terabyte. Assuming each movie is roughly 100 minutes long, that would be about 300 movies.

“If a human watch a movie/day a terabyte can store almost the video library of a year”
Big Data Image Addicts (Selfies)

1 TERA byte = 1,000 gigabytes

Pictures
Depends on the compression format in one terabyte:

<table>
<thead>
<tr>
<th>Compression Format</th>
<th>5MP</th>
<th>22MP</th>
<th>Photos (Approximately)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compressed JPEG</td>
<td>1.5 megabytes</td>
<td>6.6 megabytes</td>
<td>500 000 photos</td>
</tr>
<tr>
<td>Uncompressed RAW</td>
<td>15.0 megabytes</td>
<td>66.0 megabytes</td>
<td>60 000 photos</td>
</tr>
</tbody>
</table>

“If a human takes a selfie at every awake hour (16 Selfies / day) at 5MP a terabyte can store 85 years of selfies”

“If a human takes a selfie everyday at 5MP a terabyte can store 164 years of selfies”
Data Production in the Internet

- PayPal processes over $315M in payments per day.
- Facebook's Hadoop cluster contains over 100 PB of data.
- Netflix streams over 1 billion hours of video monthly.
- Amazon owns the 3 largest Linux DBs.
- PayPal center to end user traffic will be 24.7 TB.
- Data center to end user traffic will be 561 EB in 2013.
- 400,000,000 tweets per day on Twitter.
- Facebook processes over 350 million images uploaded daily.
- eBay handles over 1 billion transactions daily.
- YouTube handles 500,000 transactions per second.
- Amazon cloud handles 72 hours of video are uploaded to YouTube every minute.
Who is producing the Big Data?

- 900 MILLION posts to Facebook per day
- $17 BILLION credit card transactions per day
- 90% OF THE DATA in the world created in last two years
- 7 BILLION Mobile devices in use
- $3.7 TRILLION transactions per day
- 2.5 QUINTILLION bytes data created per day
- 75 MILLION Posts to Instagram per day
- 11 BILLION internet-connected things

90% of the data in the world today was created within the last two years.

Growth in the Big Data Market

- In 2010: $3.2 BILLION
- In 2015: $16.9 BILLION

Big Data Is Only Getting Bigger

International Data Corporation Forecast
Big Data is data that is too large, complex and dynamic for any conventional data tools to capture, store, manage and analyze. The right use of Big Data allows analysts to spot trends and drive richer insights that help create value and innovation much faster than conventional methods.

### Value

<table>
<thead>
<tr>
<th>Sector</th>
<th>Productivity Increase</th>
<th>Value Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retail</td>
<td>46%</td>
<td>$9.6B</td>
</tr>
<tr>
<td>Consulting</td>
<td>30%</td>
<td>$3.08B</td>
</tr>
<tr>
<td>Air Transportation</td>
<td>121%</td>
<td>$4.3B</td>
</tr>
<tr>
<td>Construction</td>
<td>20%</td>
<td>$5.8B</td>
</tr>
<tr>
<td>Food Products</td>
<td>20%</td>
<td>$3.4B</td>
</tr>
<tr>
<td>Steel</td>
<td>20%</td>
<td>$3.4B</td>
</tr>
<tr>
<td>Automobile</td>
<td>19%</td>
<td>$1.3B</td>
</tr>
<tr>
<td>Industrial Instruments</td>
<td>16%</td>
<td>$1.2B</td>
</tr>
<tr>
<td>Publishing</td>
<td>16%</td>
<td>$0.8B</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>17%</td>
<td>$5.4B</td>
</tr>
</tbody>
</table>

The estimated size of the digital universe in 2013 was 1.8 zettabytes. It is predicted that between 2009 and 2020, this will grow 44 fold to 35 zettabytes per year. A well defined data management strategy is essential to successfully utilize Big Data.
ACTIVAGE CONTEXT

ACTivating InnoVative IoT smart living Environments for AGEing well

“Activity is Medicine”

SILVER SOCIETY

(SILVER ECONOMY)

ACTIVE and HEALTHY AGEING (AHA)
ACTIVAGE Project in FIGURES

9 Deployment Sites

46 Partners → 1,642 Targeting Efforts PMs

7,200 users

6,000 Elderly Users → 1,200 Carers

*Number of partners in the local ecosystem
ACTIVAGE USE CASES

4. Emergency Trigger

8. Safety, Comfort & Safety at Home

7. Prevention of Social Isolation

2. Integrated Care

Disease management

Social engagement

Assistive technologies

Housing

Transport & mobility

Behaviours

Environmental Factors

Individual factors

5. Activity Promotion

6. Cognitive stimulation

3. Monitoring Assisted Persons outside home

1. Daily Activity Monitoring

9. Support for Transportation & Mobility
# ACTIVAGE Living Lab Solution

## IoT-enabled Vital Signs Monitoring Services (Example 1)

<table>
<thead>
<tr>
<th>Data Project Initiation</th>
<th>Data Governance Process</th>
<th>Data Quality Process</th>
<th>Self-Audit Process</th>
<th>Data Project Completion</th>
<th>Data Project Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
</tr>
<tr>
<td>Activity Track</td>
<td>Activity Track</td>
<td>Activity Track</td>
<td>Activity Track</td>
<td>Activity Track</td>
<td>Activity Track</td>
</tr>
<tr>
<td>Sleep Quality</td>
<td>Sleep Quality</td>
<td>Sleep Quality</td>
<td>Sleep Quality</td>
<td>Sleep Quality</td>
<td>Sleep Quality</td>
</tr>
<tr>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
</tr>
<tr>
<td>Blood Pressure</td>
<td>Blood Pressure</td>
<td>Blood Pressure</td>
<td>Blood Pressure</td>
<td>Blood Pressure</td>
<td>Blood Pressure</td>
</tr>
<tr>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
</tr>
<tr>
<td>Oxigen Sensor</td>
<td>Oxigen Sensor</td>
<td>Oxigen Sensor</td>
<td>Oxigen Sensor</td>
<td>Oxigen Sensor</td>
<td>Oxigen Sensor</td>
</tr>
<tr>
<td>Weight Control</td>
<td>Weight Control</td>
<td>Weight Control</td>
<td>Weight Control</td>
<td>Weight Control</td>
<td>Weight Control</td>
</tr>
<tr>
<td>Balance</td>
<td>Balance</td>
<td>Balance</td>
<td>Balance</td>
<td>Balance</td>
<td>Balance</td>
</tr>
<tr>
<td>Chestband</td>
<td>Chestband</td>
<td>Chestband</td>
<td>Chestband</td>
<td>Chestband</td>
<td>Chestband</td>
</tr>
<tr>
<td>BPM</td>
<td>BPM</td>
<td>BPM</td>
<td>BPM</td>
<td>BPM</td>
<td>BPM</td>
</tr>
<tr>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
<td>Wristband</td>
</tr>
<tr>
<td>GPS Tracker</td>
<td>GPS Tracker</td>
<td>GPS Tracker</td>
<td>GPS Tracker</td>
<td>GPS Tracker</td>
<td>GPS Tracker</td>
</tr>
<tr>
<td>UV Ray Indicator</td>
<td>UV Ray Indicator</td>
<td>UV Ray Indicator</td>
<td>UV Ray Indicator</td>
<td>UV Ray Indicator</td>
<td>UV Ray Indicator</td>
</tr>
</tbody>
</table>

Delivered by: [VITAL](www.vital-iot.eu)
Supported by: [ACTIVAGE AIO TES](www.openiot.eu)
Powered by: [OPENIoT](www.openiot.eu)
ACTIVAGE KIOSK Solution

Devices, Application and Technology (Example 2)

A FAMILY OF PRODUCTS TO HELP YOU STAY HEALTHY
Empower yourself with our ecosystem to reach your health goals and live a better life.

Wristband
Activity Track
Sleep Quality

Wristband
Blood Pressure

Wristband
Oxigen Sensor

Weight Control
Balance

Insight
ACTIVAGE IoT Data Technology Challenges

• To have a “Dynamic Knowledge Network”, based on:
  – Data Exchange between Deployment Sites.
  – IoT Data-enabled communication between Stakeholders.
  – Real Time Big Data Analytics “Stream Processing”

• To provide IoT Data insights to the main stakeholders of Silver Society and AHA communities with appropriate data.

• To define new business models based on IoT Data usability and with real impact on Innovating European AHA and IoT data available services.
Data / Cloud / Stream / IoT Data Research Timeline

- **1980's**: Computing
- **2000**: Virtualization
- **2010**: Cloud Storage
- **2015**: Cloud Processing
- **2020**: Edge Processing
- **2025**: Edge Processing
- **2030**: Future Research

**Key Technologies and Trends**:
- **1980's**: Computing
- **2000**: Virtualization
- **2010**: Cloud Storage
- **2015**: Cloud Processing
- **2020**: Edge Processing
- **2025**: Edge Processing
- **2030**: Future Research

**Research Areas**:
- **Digital**: Established Research
- **Network Services**: Traffic Data
- **Cloud Virtualization**: IoT Insights
- **Applications Economy**: IoT Systems
- **Edge Computing**: IoT Platforms
- **Multi-Domain Transition**: IoT OS

**Technologies**:
- **Internet of Things (IoT)**
- **Networks**: Social Networks
- **Cloud Storage**: Query Data Acquisition
- **Applications**: Internet of Things
- **Edge Processing**: Near Research
- **Deep Learning and AI**: Future Research
Data Security Pyramid

Data Security

Design Security

Secure Hardware

DATA SECURITY

PLATFORM SECURITY

HARDWARE SECURITY

IoT

IoT

IoT

Information Assurance:
Key storage using Physically Unclonable Function (PUF)
Advanced Crypto Accelerators
Licensed Patent Protected DPA Resistance Pass through License

Anti-Tamper: Secure Bitstream, Tamper Detection,
Active Mesh, No Copying, Cloning, or Reverse Engineering

Trust: Licensed Patent Protected DPA
Resistance, NIST Certified Crypto
Accelerators, Secure Supply Chain
Big Data & IoT Clusters Security

- Privacy
  - Encryption

- Access Control
  - Reputation

- Security
  - Authentication
Cluster Security in Details

- **PRIVACY**
  - Secrecy
  - Anonymity

- **ACCESS CONTROL**
  - Encryption
  - Intrusion

- **TRUST**
  - Detection
  - Authentication

- **SECURITY**
  - Integrity
  - Authorization

- **Non-Reputation**
  - Reputation

**Clusters and Security Aspects**

- **Cluster Security in Details**
  - Intrusion
  - Encryption
  - Authentication
  - Integrity
  - Secrecy
  - Anonymity
Security, Privacy and Trust Models

• Secure, Trustworthy and Privacy Friendly Interactions
  – Security and trustworthiness protocols
  – Implement security mechanisms at EU research level (Opinion, VITAL, FIESTA)
  – Secure service requests and interactions

• Main Results
  - Investigated existing security frameworks and protocols
  - Centralized authentication and authorization framework
  - Implementation of CAS for OpenIoT
  - Integration of CAS with LSM storage
  - Security client API and tag libraries for integration
Trust Cross Domain Terminology

TRUST

- Distributed
- Hybrid
- Centralized

- Totally
- Partial

- Behaviour-Based
- Certificate-Based
Security Cross Domain Terminology

- Security
  - Resource
    - Availability
  - Platform
    - Authority
    - Accessibility
    - Control
  - Data
    - Authenticity
    - Secrecy
Privacy Cross Domain Terminology

Privacy

Anonymity

- Identity
- Route
- Location

Secrecy

- Data
- Key
- Hard
- Soft

Solitude

- Inspect Centre for Data Analytics
OpenIoT Security Architecture with CAS – OAuth 2.0 Protocol

- Centralized Authentication Server (CAS) for all OpenIoT applications
- Based on open standard framework for authorization OAuth
- Applications not burdened with credentials handling
- Role and permission authorization managed by CAS

OAuth 2.0 abstract authentication and authorization flow
Internet of Things Stack

**Business Level**
e.g. Smart City, Intelligent Manufacturing, M2M, CPS, Smart Appliances.

**Application Level (+ domain specific)**
e.g. Dublin Core, FOAF, SSN and OpenIoT.

**Semantic Level**
Existing vocabularies (e.g., NCI, SSN-XG)

**Sensor Middleware Level**
owl:sameAs, rdf:seeAlso

**Relationships:**
- closeMatch
- exactMatch
- broadMatch
- narrowMatch
- relatedMatch

**Other knowledge base and ontologies**
e.g. DBPedia, Geonames

**Virtual Sensor Level**
e.g. X-GSN

**Physical Level (Device Standards)**
e.g. IPv6, 6Lowpan, IETF CoAP

**Technology and Infrastructure**
e.g. Arduino, Quark, ARM, etc.
Internet of Things Stack

**IoT Security**

- Information Assurance
- Anti-Tamper: Secure Bootstrapping, Tamper Detection
- Hardware Security
- Network Security
- Platform Security
- Data Security

**IoT CAS Implementations**

- Business Level: e.g., Smart City, Intelligent Manufacturing, M2M, CPS, Smart Appliances
- Application Level (+ domain specific): e.g., Dublin Core, FOAF, SSN, and OpenIoT
- Semantic Level: Existing vocabularies (e.g., NCI, SSN-IG)
- Sensor Middleware Level: Specific to applications, e.g., mxtsensor, Nodej
- Physical Level (Device Standards): e.g., IP, ETSI, IETF, CoAP

**IoT Stack**

Insight Centre for Data Analytics
Big Data Industries Landscape

Data Acquisition
- IBM
- SAP
- SAS
- TIBCO
- HP
- Ab Initio
- Microsoft
- Oracle
- Teradata
- Informatica
- Splunk
- Numenta
- Syncsort

VLDW and BI Appliances
- IBM
- SAP
- Oracle
- Tableau
- Alphaworks
- Splunk
- Infoxchange

Analytics
- EMC
- IBM
- SAP
- TIBCO
- Oracle
- GoodData
- Microsoft
- KXen
- Teradata

BPM & Action
- TIBCO
- Oracle
- EMC
- SAP
- Pega
- Software AG
- OpenText

Including Complex Event Processing (CEP) tools

Data Providers
- LexisNexis
- comScore
- INRIX
- Reuters
- Nielsen
- FactSet
- FactSet
- Kaggle
- COO
- Symphony
- Flickr
- Photos
- Instagram

No SQL
- Hadoop
- Oracle
- EM
- IBM
- Google
- Microsoft
- CouchDB
- MongoDB
- Apache
- Amazon

Data Virtualization
- IBM
- Oracle
- SAP
- Tandem
- ELOQUA
- IgraFX

Content Management
- Adobe
- Alfresco
- IBM

BI Tools
- Oracle
- SAP
- Microsoft
- Tableau
- QlikView
- Pentaho

Data Governance
- IBM
- SAP
- TIBCO
- Oracle

And all your own data
And your partners data

Capgemini - Capping IT off
Manuel Sevilla - 2012
CONCLUSIONS

• Big Data is an explicit element in the IoT.
• Bid Data Management strategies are yet in need to be in place before launching Applications for Big Data.
• Internet of Things enable Big Data Generation.
• Big Data analytics remains as open challenge for knowledge generation.
• IoT and Big Data focus area remains in platforms interoperability.
• Technology for Big Data Semantics requires more research work.
IoT Week 2017 – Day 2 Session: 16:15-17:15 Room 2
Panel Emerging IoT Researches and Technologies Big Data
IoT and Big Data | International Conference Centre of Geneva (CICG)

Dr. Martin Serrano
Principal Investigator and Data Scientist
Insight Centre for Data Analytics, Ireland
<martin.serrano@nuigalway.ie>

Thank You!