IoT Meets Big Data: Standardization Considerations

Ali Drissa Badiel
ITU Area Representative for West Africa
http://www.itu.int/ITU-D/afr
Recommendation ITU-T Y.3600

Requirements and capabilities for cloud computing based big data
Rationale for the study

Because of the v’s characteristics, big data technologies and services allow to resolve many new challenges, and also create new opportunities, than ever before:

- **Heterogeneity and incompleteness**
  - Data processed using big data technologies can miss some attributes or introduce noise in data transmission. Even after data cleaning and error correction, some incompleteness and some errors in data are likely to remain.

- **Scale**
  - Managing large and rapidly increasing volumes of data is a challenging issue for data processing. In the past, the data scale challenge was mitigated by evolution of processing and storage resources. But now data volumes are scaling faster than resources evolve.

- **Timeliness**
  - The acquisition rate and timeliness, to effectively find elements in limited time that meet a specified criterion in a large dataset, are new challenges faced by data processing.

- **Privacy**
  - Data about human individuals, such as demographic information, internet activities, commutation patterns, social interactions, energy or water consumption, are being collected and analysed for different purposes. Big data technologies and services are challenged to protect personal identities and sensitive attributes of data throughout the whole data processing.
Scope of Recommendation ITU-T Y.3600

- Overview of big data;
  - Introduction to big data;
  - Big data ecosystem and roles;
  - Relationship between cloud computing and big data;
- Cloud computing based big data system context and benefits;
- Cloud computing based big data requirements;
- Cloud computing based big data capabilities;
- Cloud computing based big data use cases and scenarios.
Definition of “big data"

- big data: a paradigm for enabling the collection, storage, management, analysis and visualization, potentially under real-time constraints, of extensive datasets with heterogeneous characteristics.

Note 1 - Paradigm is used here to express the necessity to find new approaches to address identified issues and phenomena.

Note 2 - Examples of datasets characteristics include high-volume, high-velocity, high-variety etc.
Big data ecosystem

- Roles of Big Data ecosystem
  - data provider,
  - big data service provider,
  - big data service customer.
Cloud computing based big data system context

Cloud Service Partner (CSN)

CSN: Data provider
- Generate data
- Publish data
- Brokerage data

Cloud Service Provider (CSP)

CSP: Big data application provider
- Visualize data
- Analyze data

CSP: Big data infrastructure provider
- Perform data collection
- Provide data pre-processing
- Perform data storage/persistence
- Manage data privacy
- Provide data integration
- Manage data provenance

Cloud Service Customer (CSC)

CSC: Big data service user
- Use big data service

Role

Sub Role

Activity
# Mapping between big data ecosystem and cloud computing based big data system context

<table>
<thead>
<tr>
<th>Big data ecosystem roles</th>
<th>Cloud computing based big data system context sub-roles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data provider</td>
<td>CSN:data provider</td>
</tr>
<tr>
<td>Big data service provider</td>
<td>CSP:big data infrastructure provider, CSP:big data application provider</td>
</tr>
<tr>
<td>Big data service customer</td>
<td>CSC:big data service user</td>
</tr>
</tbody>
</table>
Requirements and capabilities for Cloud computing based big data

- Data collection
- Data pre-processing
- Data storage
- Data analysis
- Data visualization
- Data management
- Data security and protection
Example of Use cases

Web service users

CSP: BDIP
- Perform data collection
- Perform data storage/persistence
- Provide data integration
- Provide data pre-processing
- Manage data privacy
- Manage data provenance

CSP: BDAP
- Visualize data
- Analyze data

CSC: BDSU

Web services
- Shopping
- Medicine information
- Multimedia
- Education
- Travel/business trip

Personalization customized service

CSN:DP
- Generate data
- Publish data
- Brokerage data

BDIP: Big data infrastructure provider
BDAP: Big data application provider
DP: Data provider
BDSU: Big data service user
## Mapping of big data ecosystem roles and sub-roles of CCRA user view

<table>
<thead>
<tr>
<th>Big Data ecosystem</th>
<th>User view of Y.3502</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Provider</td>
<td>CSN</td>
<td>For the cloud computing based big data environment, the new sub-role of CSN for data provider is required.</td>
</tr>
<tr>
<td>Data Broker</td>
<td>CSN</td>
<td>(option 1) the extension of the activities of CSN: Cloud service broker (option 2) the new sub-role of CSN for data brokerage. (option 3) adding activity “brokerage” on the sub-role which is correspond with simple data provider</td>
</tr>
<tr>
<td>Big Data Service Customer</td>
<td>CSC</td>
<td>(option 1) using CSC (option 2) the new role or sub-role of CSC for big data service customer</td>
</tr>
<tr>
<td>Big Data Service Provider</td>
<td>CSP</td>
<td>Big data service provider could be treated with an implemented feature. Nevertheless, to clarify the cloud-based Big data service, an independent sub-role of CSP for big data service provider is required.</td>
</tr>
</tbody>
</table>
IoT Meets Big Data
ITU-T Study Group 20: IoT and its applications including smart cities and communities (SC&C)

Q1/20
Research and emerging technologies including terminology and definitions

WP1/20
Internet of Things (IoT)

WP2/20
Smart cities and Communities (SC&C)

Next meeting:
From 25 July to 5 August 2016
ITU Headquarters

Responsible for international standards to enable the coordinated development of IoT technologies, including M2M communications and ubiquitous sensor networks
SG20 areas of work

Q1/20
Research and emerging technologies including terminology and definitions

WP1/20 Internet of Things (IoT)

Q2/20
Requirements and use cases for IoT

Q3/20
IoT functional architecture including signalling requirements and protocols

Q4/20
IoT applications and services including end user networks and interworking

WP2/20 Smart cities and communities (SC&C)

Q5/20
SC&C requirements, applications and services

Q6/20
SC&C infrastructure and framework
The first 2 Recommendations approved

**ITU-T Y.4702 Common requirements and capabilities**

Approved on 15 March 2016

This Recommendation provides the common requirements and capabilities of device management in the Internet of Things (IoT).

The provided common requirements and capabilities are intended to be generally applicable in device management application scenarios.

**ITU-T Y.4553 Requirements of smartphone as sink node for IoT applications and services**

Approved on 15 March 2016

This Recommendation provides common Requirements of a smartphone working as a sink node (SPSN) for IoT applications and services.

This Recommendation clarifies the concept of a sink node in IoT domain, and identifies the characteristics, work modes and the high level functional requirements of the SPSN. The use cases are also provided in appendix.
Big Data and IoT

- **Big data is about data**, plain and simple. Yes, you can add all sorts of adjectives when talking about “big” data, but at the end of the day, it’s all data.

- **IoT is about data, devices, and connectivity.** Data – big and small – is front and center in the IoT world of connected devices.
Landscape is Crowded - Very Crowded
Everything & Everyone is Generating Data

- Consumer equipment providers
- Hospitals & Doctors
- ICT infrastructure providers
- Insurance companies
- Regulators
- Appliances providers
- Logistics companies
- Facility management
- Public transport companies
- Retail stores
- City authorities
- Application developers
- Manufacturing industries
- Utilities
- Automation equipment providers
- Common IoT Standards & Framework
Big Data – Big Opportunity

The intersection of IoT and Big Data is creating a tremendous business opportunity, but the reality of today’s IoT projects is that this potential has yet to be fully tapped.
Survey of the Use of IoT

- 200 technology and business professionals responsible for IoT projects.
- The goal of the survey was to understand experiences and impacts of using the data captured by the devices that make up the Internet of Things and focused on the untapped potential of IoT data.

Use of IoT for Business Optimization
- 53 per cent are using IoT projects to optimise existing businesses and 47 percent as a strategic business investment
- Target audiences for IoT solutions include consumers (42 percent), business (54 percent) and internal use by employees (51 percent)

Challenges with IoT Projects
- 96 per cent have faced challenges with their IoT projects
- IoT Is Not Delivering Full Potential Because Of Data Challenges
- Only 8 per cent are fully capturing and analysing IoT data in a timely fashion
- 86 per cent of stakeholders in business roles say data is important to their IoT project
- 94 per cent face challenges collecting and analysing IoT data

Better IoT Data Collection And Analysis Would Deliver More Value
- 70 per cent say they would make better, more meaningful decisions with improved data
- 86 per cent report that faster and more flexible analytics would increase the ROI of their IoT investments

Source: PARSTREAM
IoT & Data Challenges

- 44% said that there was too much data to analyze effectively
- 36% said that it was difficult to capture data in the first place,
- 25% saying data was not captured reliably
- 19% saying that data was captured too slowly to be useful.
- Once data is captured, 27% said they weren’t sure what to use it for and were unsure what questions to ask.
- Much like data capture, 26% said that the analysis process was too slow to be actionable,
- 24% said that business processes were too rigid to allow them to act on their findings – even if they were captured and crunched in time to be useful.
- While cost is often a limiting factor in many technology decisions, for IoT stakeholders, ease of use appears to be a more pressing issue than cost.
- More participants (76%) say they would collect and store more data if it were easier than those who said they would collect and store additional data if it were free.”

Source : PARSTREAM
Big Data or IoT?

- Every minute, we send 204 million emails, generate 1.8 million Facebook likes, send 278 thousand tweets, and upload 200 thousand photos to Facebook. *(BIG DATA or IoT)*
- 12 million RFID tags (used to capture data and track movement of objects in the physical world) were sold in 2011. By 2021, it’s estimated this number will increase to 209 billion as *(BIG DATA or IoT)* takes off.
- The boom of *(BIG DATA or IoT)* will mean that the amount of devices that connect to the internet will rise from about 13 billion today to 50 billion by 2020.
- The *(BIG DATA or IoT)* industry is expected to grow from US$10.2 billion in 2013 to about US$54.3 billion by 2017.
Big Data or IoT?

- Every minute, we send 204 million emails, generate 1.8 million Facebook likes, send 278 thousand tweets, and upload 200 thousand photos to Facebook. *(BIG DATA or IoT)*
- 12 million RFID tags (used to capture data and track movement of objects in the physical world) were sold in 2011. By 2021, it’s estimated this number will increase to 209 billion as *(BIG DATA or IoT)* takes off.
- The boom of *(BIG DATA or IoT)* will mean that the amount of devices that connect to the internet will rise from about 13 billion today to 50 billion by 2020.
- The *(BIG DATA or IoT)* industry is expected to grow from US$10.2 billion in 2013 to about US$54.3 billion by 2017.
Some trends in the IoT

- Fragmented market & vertical applications
- One device type per application
- Single domain data
- Specific and proprietary technologies, In-house IT
- Limited service models, device and connectivity focus
- Dominance of cellular networks

Horizontal solutions
- Multi-purpose devices

Data and information driven,
- Integrated data analytics,
- Cloud, data marketplace

Application innovation
- Hyper-growth of Service Providers,
- variety of business models, integrated ecosystems

Multi-access networks
- Cloud Computing, distributed computing

Standardization has to address this evolution
Opportunities of data in the IoT: using and monetizing data via analytics

(Big) Data analytics are a key revenue opportunity in IoT market

Data analytics can allow organizations to drive revenue by sharing, analyzing and interpreting data, for multiple purposes
- Extraction of tangible business and technology value
- Response and action in real time, improving productivity and business processes, lowering operational and other costs
- Long-range forecasts enabling strategic actions
- Better/New service offer to customers, faster and more efficiently

Surveys indicate increasing Big Data analytics deployments and exponential growth of Big Data analytics revenues
IoT and (Big) Data

- Some analysts indicate that by 2020 40% of data will come from sensors
- Multiple data sources: (real-time) data from things and context, historical and social data (cross-domain data exchange and correlation etc.)
- Data are mainly semi-structured and unstructured
- Data may have different precision and confidence levels
- Various operations to be made on data for the extraction of actionable intelligence (collection, de-noising, aggregation, adaptation, analysis)
- Raw data -> Information -> Knowledge (-> Wisdom)
- Target: the right data, at the right time, at the right location (e.g. cloud versus edge computing)

The Industrial Internet Data loop

Data interoperability throughout the cycle is critical (syntax and meaning)
Things generate raw data

Additional information enables creation of structured metadata (first step of IoT data enrichment)

Abstractions and perceptions give detailed insights by reasoning using knowledge (ontologies, rules) of different domains (second step of IoT data enrichment)

Actionable intelligence allows decision making

The knowledge hierarchy applied in IoT Data processing

Source: Barnaghi and al., “Semantics for the IoT: early progress and back to the future” (IJSWIS, 2012)
Semantics is the study of meaning

- A definition: “The rules and conventions governing the interpretation and assignment of meaning to constructions in a language” [ITU-T Z.341]
- Semantics for the IoT: Shared vocabularies for IoT entities and their relationships

Requirements for interoperability, scalability, consistency, discovery, reusability, composability, automatic operations, analysis and processing of data and services are becoming more and more essential in the IoT

- Given the growing number of interconnected things and connections, variety of devices and connectivity, volume and types of generated data, number and type of services

Semantics based approaches reveal outstanding features to support these dynamic and cross-domain requirements of the IoT
Value proposition of semantics based approaches

Outstanding features for data and services, including an answer to some key technical challenges of data in the IoT

- **Interoperability**: the interoperability level of data and services of the IoT within one application domain and/or among different application domains can be improved.
- **Scalability**: data and services can be managed locally to IoT components. This increases the IoT technical component independence (loosely coupled components) and decentralizes the management, leading to enhanced scalability. Service reachability can be more easily expanded to reach more users, functional evolution of services can be rationalized.
- **Consistency**: data and services can refer to same meaning across time, location, IoT components.
- **Re-usability**: data and services can be reused and composed to construct new data, services.
- **Analytics and actionable knowledge**: merging, correlation and analysis of diverse data generated by the IoT, together with data from external sources such as social media, events and news, can be facilitated in order to produce actionable knowledge.
- **Human-machine interaction**: on one hand, since semantic technologies are based on natural human concepts, data and services become easier for humans to understand. On the other hand, since semantic technologies are formal ways to express concepts, data and services can also be understood by machines. This can improve the interaction between humans and the IoT.

Source: Y.2076
But still issues concerning the full applicability of semantic technologies in the IoT

The benefits of semantic technologies for the IoT will be realized in incremental way: current issues include

- Lack of elaborated use cases as drivers to validate the value proposition (value for the different stakeholders of the value chain)
- Insufficient link with IoT architectures
- Immaturity of tools, essential to establish semantics based bridging (domain specific ontologies formalizing the meaning of domain data and information models; semantic merging, matching and alignment strategies; mediators to enable integration of disparate data resources; bridges etc.)
- Semantic discovery of services, devices, things and their capabilities
- Semantic metadata framework and IoT base/core ontology (horizontal cross-domain integration)
- Participation of domain specific communities and creation of social and business incentives for sharing
- Education (entrepreneurs/domain experts – developers interaction)

The initial results are promising, but further research, development, validation and standardization is required.

Big Data technologies: value proposition and current critical aspects

Big Data technologies address several data challenges in the IoT
- Scalability, data integration, massive data mining, data accessibility and create value (Big Data Analytics)

But some aspects are still critical from a deployment perspective
- Integration and interoperability with legacy environments and applications
- System performances and reliability
- Wide spectrum of technologies and products, organizational impact, skilled personnel
- And analytics is becoming a multi-dimensional challenge (data at rest versus data in motion, and the related data cycle operations)
- Data privacy, access, security
Big Data technical standardization

Potential areas of standardization related to Big Data interoperability [including from Y.Suppl.BigData-RoadMap and AFNOR’s 2015 Big Data white paper]

- Requirements and use cases
- Architecture, data model and APIs (APIs with IoT infrastructure, users, auditors)
- Open data frameworks
- Data analytics (network-driven, others)
- Security and data protection, anonymization and de-identification of personal data (and reversibility)
- Framework for data quality and veracity
- Standards and guidelines to address issues related to legal implications of big data in telecommunications (e.g. data ownership)
- Framework and standards for (telecom) big data exchange
- Benchmarks for system performance evaluation (e.g. Hadoop)
- Standardized visualization methods
- NoSQL query languages
Some key technical challenges of data in IoT

- Dealing with the data “V”s: Volume, Variety, Velocity, Variability, Veracity
- Discovery of appropriate device and data sources
- Integration of heterogeneous devices, networks and data
- Scalability to cope with large numbers of devices, diverse and huge data, computational complexity of data interpretation
- Massive data mining, adaptable learning and efficient computing and processing
- Data querying
- Availability and (open) access to data resources and data, security and privacy of data (incl. doing mining, analytics)
- Interpretation: extraction of actionable intelligence from data
- Non-technical challenges are essential though (e.g. data ownership and governance)

Source: partial adaptation from Dr. Barnaghi, iot360 2015
Main ITU-T developments related to Big Data

Completed studies
- **Y.2066: Common requirements of the IoT**
  - It includes requirements on data aspects of the IoT
- **Y.2068: Functional framework and capabilities of the IoT**
  - It includes IoT capabilities for data management and for integration of Big Data technologies and Cloud Computing technologies with the IoT
- **Y.3600: Big data – cloud computing based reqts and capabilities**

Ongoing studies
- **Y.IoT-BigData-reqts**: Specific reqts and capabilities of IoT for Big Data
- **Y.Suppl.BigData-RoadMap**: Big Data Standardization Roadmap
  - standards landscape in ICT sector and al., feasibility studies on Big Data standardization, identification of technical areas (e.g. IoT), roadmap
- **Y.BigDataEX-reqts**: Big Data Exchange Framework and Reqts
- **Y.BDaaS-arch**: Functional architecture of Big Data as a Service
This presentation has been prepared by Cristina Bueti, Tatiana Kurakova and Reyna Ubeda (ITU) and Marco Carugi (Q2/20 Rapporteur, NEC).

- Contact: tsbsg20@itu.int
Contacts:

Maria Cristina BUETI: Cristina.Bueti@itu.int

Ali Drissa BADIEL: AliDrissa.Badiel@itu.int