



Internet of Things and Big Data: the disruption of the value chain and the rise of new software ecosystems

Norbert Jesse^{1,2}

Received: 24 April 2017 / Accepted: 22 January 2018 / Published online: 9 February 2018
© Springer-Verlag London Ltd., part of Springer Nature 2018

Abstract

IoT connects devices, humans, places, and even abstract items like events. Driven by smart sensors, powerful embedded microelectronics, high-speed connectivity and the standards of the internet, IoT is on the brink of disrupting today's value chains. Big Data, characterized by high volume, high velocity and a high variety of formats, is a result of and also a driving force for IoT. The datafication of business presents completely new opportunities and risks. To hedge the technical risks posed by the interaction between “everything”, IoT requires comprehensive modelling tools. Furthermore, new IT platforms and architectures are necessary to process and store the unprecedented flow of structured and unstructured, repetitive and non-repetitive data in real-time. In the end, only powerful analytic tools are able to extract “sense” from the exponentially growing amount of data and, as a consequence, data science becomes a strategic asset. The era of IoT relies heavily on standards for technologies which guarantee the interoperability of everything. This paper outlines some fundamental standardization activities. Big Data approaches for real-time processing are outlined and tools for analytics are addressed. As consequence, IoT is a (fast) evolutionary process whose success in penetrating all dimensions of life heavily depends on close cooperation between standardization organizations, open source communities and IT experts.

Keywords Internet of Things · Smart factories · Big Data · Software platforms · Data science

1 Introduction

Sensor technology, microelectronics, communication technologies and internet-related software frameworks have made huge progress in recent years. Sensor technology has ignited machine-to-machine communication. Embedded microelectronics has led to remarkable levels of automation in production. Wireless and non-wireless communication technology has accelerated the data transfer significantly.

On the computational side, we are witnessing the accelerating dominance of the internet and innovative software paradigms enable new technologies like in-memory computing, non-SQL database technology, cloud computing and Big Data processing. With significant progress in all

technical dimensions, the window for disruptive interconnectivity of machines and humans is now wide open.

In industry, computer-integrated manufacturing (CIM) and machine-to-machine (M2M) communication with their numerous standards and communication protocols have already changed the shop-floor. CIM and M2M connect machines, mainly as proprietary, closed systems.

The Internet of Things (IoT) is much more inclusive than CIM and M2M: everything is connected to everything else (hence also labeled “Internet of Everything”) using Internet fabric and protocols. It is the network of uniquely addressable physical assets equipped with sensors which nudge information systems to capture, process, analyze, and exchange data—while including humans as sources, data sinks or something in-between.

This convergence of the internet and physical objects is a challenge which creates game-changing opportunities and risks for business development. IoT connects devices, people, places, and even ideas or events. As a final consequence, this disruption is a result of the ability to “make sense” of data, i.e., to leverage the value of data.

✉ Norbert Jesse
norbert.jesse@quinscape.de; norbert.jesse@udo.edu

¹ QuinScape GmbH, Wittekindstr. 30, 44139 Dortmund, Germany

² Department of Computer Science, TU Dortmund University, Otto-Hahn-Str. 14, Dortmund, Germany

To hedge the technical risks, IoT requires standards and platforms to

- deal with the new complexity resulting from the interaction between “things”, humans, systems and systems of systems,
- handle the exponentially growing flood of data and the requirements for real-time processing,
- integrate different tools and platforms to guarantee interoperability,
- extract sense from the data.

In this paper, we address this item list and outline critical dimensions for successful IoT projects.

2 The era of the Internet of Things

According to experts, the world has had 27 billion connected devices in 2014, will spend 3 trillion US-\$ on IoT in 2020 to produce 2.5 quintillion (10 to the power of 15) bytes of data per day or 20–40 MBs per machine per day. Several studies support this or a similar forecast (Kubach 2016; Hauptfleisch 2015; Gartner 2014). It is expected that the connectivity of physical assets will change the entire manufacturing business as it becomes possible to

- track when and how they are being used and to price and charge for them respectively,
- use the data from connected assets to remotely operate a customer’s equipment much more efficiently,
- enable condition-based predictive maintenance to minimize unplanned downtime.

Examples for maintenance applications are path breaking:

- Siemens demonstrates predictive maintenance for high-speed trains. For 26 trains, 400 sensors per train deliver 200 bytes of data/s and GPS coordinates are additionally stored every 30 s to anticipate technical problems (VDI nachrichten 2016).
- Amsterdam Airport has equipped critical machines—elevators, baggage transporters, etc.—with sensors while blending their data with available data from the airport (number of people going through the screening process, etc.) to estimate the relative health of machines and identify the best time to check machines for maintenance with the least possible impact (Cloudera 2016).
- Air France/KLM has equipped their A 380 airplane with 24,000 sensors. These sensors generate 1.6 GB of data per flight (Heck and Franco 2016) and enable the detection of problems 10–20 days ahead.

When it comes to specific sectors of industry, utilities are expected to be in the no. 1 spot, followed by manufacturing and government. While the first two fields of application are obvious, it seems to be less evident regarding governmental scenarios. However and in particular, the concept of smart cities, with smart streets and lighting, smart waste disposal or smart traffic management relies heavily on IoT technology. A smart cities scenario might include:

- sensors embedded in water pipes, sanitation services, traffic controller, parking meters and more, to monitor and flag capacity issues and automatically make adjustments to traffic flow, pickup schedules, etc.,
- improved public safety through more effective and strategic usage of policing resources for crime prevention and emergency responsiveness,
- improved safety for fire fighters through wearable sensors that can track their movement, ensuring that everyone gets out of dangerous situations.

In car technology, consumer orders will be connected directly to warehouses, humans to cars, cars to repair services, insurance companies and traffic control. Already on the market is the Remote Online Service by Mercedes for monitoring cars from home or the office. One of the new features is GPS-based car monitoring which raises an alarm if the car leaves a defined radius.

In B2C, wearables are expected to penetrate the market quickly and data might be connected to doctors, hospitals or insurance companies. In retailing, one of the largest sources of value could be the sales lift in real-time with in-store personalized offers. This scenario will require the sophisticated integration of data from many sources:

- real-time location data (the shopper’s whereabouts in a store),
- data from sensors in a building,
- customer-relationship-management data, including the shopper’s online-browsing history,
- data from tags in the items on display, telling the customer to enter a specific aisle,
- data from instant coupon companies using their data to motivate buying by sending a personalized offer to a mobile phone (Bughin et al. 2015).

As a result, companies have to evaluate their business models, tailor and probably enlarge their service business. There is no doubt that IoT technology will change business as we know it today.

3 Smart sensors as a driving force

IoT is sensor-enhanced Internet. Sensors attached to the Internet have the ability to source data and control actuation. In a broader sense, sensors do not only include hardware sensors and sensor networks but also software sensors, able to capture real-world conditions of interest, such as user presence which can be detected via key clicks or mouse movements. Even when referring only to hardware sensors, their data become the standard source of data. These data are often imprecise, sometimes incomplete and flow continuously. They are “flat” in a sense that they line up like “pearls on a chain”.

Sensor technology has made significant progress in recent years and sensors will be the main data source in the future (Siprell 2016). High-speed, low-power, high-resolution, less noisy, compact sensors which consume less energy are the key factors for IoT. Sensor control air quality, traffic lights, pressure or temperature (Simmons 2014). The coupling of physical, chemical and biological components enables sensor labs on-a-chip to be created for medical diagnosis, biotech or chemical applications. Systems-on-a-chip integrates Bluetooth connectivity and creates a fully wearable sensor hub. Increasingly, sensor systems are becoming adaptive with feedback components, self-monitoring or self-calibration abilities. While multi-sensor systems speed up the availability and quality of information, algorithms for data fusion are gaining in importance and paving the way for intelligent applications (Trankler and Kanoun 2001). Some basic dimensions of sensor fusion are outlined in Table 1.

Sensor data pose qualitative and quantitative challenges. With respect to quantity, it is often not possible or not reasonable to handle the huge amount of sensor data with software based on conventional IT paradigms (SQL data bases, standard multi-layer architectures, etc.). As a consequence, most data generated by IoT sensors remain unused. In the oil-drilling industry, one of the first users of sensors, only one percent of the data from 30,000 sensors on a standard

oil rig is used, and even this small fraction of data is not analyzed for optimization, prediction, and data-driven decision-making (Bughin et al. 2015).

The qualitative challenge is to develop and agree upon standards and practices that enable the exchange and integration of data from sensors across devices, users, and domains. The objective is to achieve semantic or conceptual interoperability, i.e., to represent information in a standard whose meaning is independent of the device or format generating or using it. Semantic interoperability enables service-level integration of IoT end-to-end systems with components from different vendors and guarantees the aggregation of data from different domains (Milenkovic 2015).

Meta-data are essential for IoT because they annotate sensor data to provide context. Meta-data of interest might comprise sensor type, serial number, and frequency of reporting, mobile or static location, manufacturer, domain, associations, access rights, privacy policy and restrictions, accuracy, calibration, and others. The primary function is to provide contextual semantics to create “rich data” for post-processing services and applications. Furthermore, meta-data enable valuable searches to be carried out. Sensor data are numbers that depend on meta-data to provide context and semantics, while, on the other hand, internet content search operates on documents “encoded” in natural languages with inherent dictionary-driven semantics.

The challenge for IoT is to devise a coordinated naming, taxonomy/ontology, and meta-data system. However, it is probable that no single data and metadata format will emerge and be adopted on a large-scale to, for instance, interconnect domains like building automation, transportation or energy management (Milenkovic 2015).

4 Objects of interest and model building

IoT significantly changes the equation for modelling information systems. With IoT, physical objects (refrigerators, machines, or cars), systems (shop-floors) and systems of systems (smart cities) must be modeled adequately. In these models, virtual “things” are proxies for physical and abstract entities that are described in terms of metadata, events, and properties. Meta-information becomes a central part of the model of sensors with information like history, place, state, context, event or contact with other objects (Raggett 2015) (Fig. 1).

The physical object model must adequately identify and represent

- the workload, i.e., an object is more or less active,
- time constraints, i.e., data from the object must be processed in precisely defined cycles probably as short as possible,

Table 1 Dimensions of sensor data

| Dimension | Description |
|-------------------|--|
| Sensitivity | Ratio change of output to change of input |
| Linearity | Measure for the constancy of ratio Input to Output |
| Measurement range | Difference between min and max |
| Response time | Time requirements for a change in input to be observable in the output |
| Accuracy | Diff. between actual and measured values |
| Repeatability | Diff. between successive meas. of the same entity |
| Resolution | Small observable increment in input |

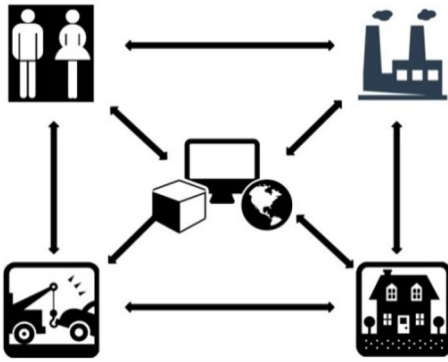


Fig. 1 Modelling interconnected objects in the real-world (according to Volker Gruhn 2017)

- robustness, i.e., data have to be processed with respect to the requirements of the business case (for instance, vital data of a person with respect to an insurance bonus),
- reliability, i.e., any blackout of an object might cause a severe problem for the whole system,
- flexibility, i.e., due to business, regulatory or security reasons the content, format or frequency of data supply might change (Lauby 2015).

Tools for modelling the IoT world have to chart a bridge between machines, machines and humans (operator, car driver, shop manager, etc.) and the internet.

5 Germany's approach for Industry 4.0 and the industrial data space

5.1 Industry 4.0

Industry 4.0 is being promoted by leading German industrial associations. The label stands for the complete digitization and integration of the industrial value chain by closely linking information and communication technology with automation technology. The concept is based on a service-based architecture and includes accepted standards and protocols (ISO, IEC etc.). Ultimately, Industry 4.0 shall identify the relevant standards and protocols needed to enable the implementation of a smart factory and a digital value network (<http://www.plattform-i40.de/I40/Online-Bibliothek>, dated August 29, 2016).

The objectives of Industry 4.0 are to make all relevant information of an industrial domain available *in real time* by *connecting all relevant entities with each other* and to have the capability to *use the data* that is generated to *determine current process statuses* at all times so as to derive the *best possible value-adding decisions*. Obviously, Industry 4.0 will deliver an unending source of potentially valuable data.

Consequently, Industry 4.0 focuses on all three of the following dimensions on integration:

- *Horizontal* integration by value-adding networks: within the context of horizontal integration, interconnected companies—manufacturer, supplier, and development and logistics services—regularly exchange relevant information. This notion is to take account of customer-specific requirements throughout all the different phases of a product's lifecycle—including design, production, delivery and use.
- *Vertical* integration within automation hierarchies: vertical standards link the different hierarchies within the automation technology, i.e., at actuator and sensor, control, and planning units.
- *Self-optimization* of resources: integrating the manufacturing process is essential to self-optimization. The availability of interrelated data and the competence to harness intelligent tools and concepts paves the way for value adding optimizations.

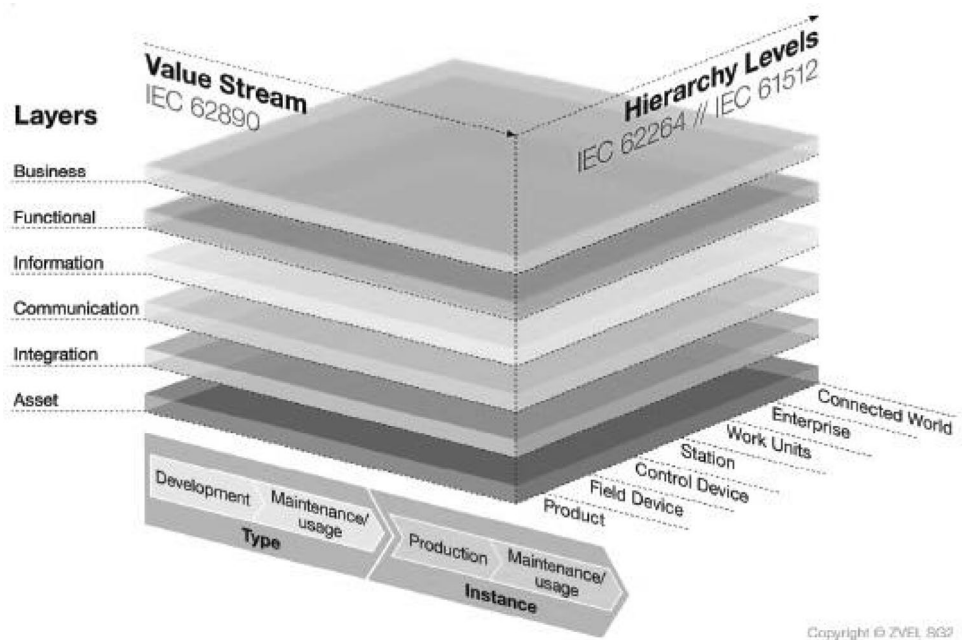
As a comprehensive technical framework for Industry 4.0, a Reference Architecture Modell Industry 4.0 (RAMI 4.0) has been specified. Arranged on three axes—functionalities within factories or facilities, lifecycle and value stream and layer-bases decomposition of a machine—technical standards define a common structure and “universal set of languages” for specific domains (Fig. 2).

To substantiate this reference architecture, the elements of the real manufacturing world must have an adequate virtual representation (both, the real and virtual dimension, coined Cyber-Physical System, CPS). In the context of Industry 4.0, this virtual image is not just a snapshot of *the current status and current connections*. Much more than that, it should also include all the information covering the *complete lifecycle* of the CPS—comprising relevant information from geometric data, mechanical properties or technical and security features. All further lifecycle dimensions—engineering, commissioning and operations, maintenance and service—add additional data (<http://www.zwei.org>).

All in all, Germany's Industry 4.0 approach wants to set an international reference framework for a comprehensive, well-structured, fast and flexible interaction between the CPSs on shop-floor level and the LOB IT systems (ERP, Manufacturing Execution System, CRM, logistics etc.) on the other end of the line.

However, Industry 4.0 is a comprehensive framework which has yet to prove its superiority over more pragmatic or restricted approaches. From the perspective of data management, the challenges are by all means enormous. The sensor-driven flow of data must be controlled and Industry 4.0 demands powerful tools for (big) data integration, routing, validation or security management. There is no doubt,

Fig. 2 The 3-dimensional Industry 4.0 cube



that Talend’s Data Fabric will play a significant role in an Industry 4.0 environment.

Based on the reference architecture and asset models RAMI 4.0 will control and optimize the complete value chain of a production process. Following this concept, the machine producer and integrator can define smart services, and humans on the shop-floor will be able to react flexibly in case of a breakdown of one of the assets (objects). Even the ability to produce goods in batch-size one at reasonable costs seems possible.

5.2 The industrial data space (IDS)

The industrial data space initiative was launched in Germany at the end of 2014 by representatives from business, politics, and research. The overall goal of IDS is to

provide an IT reference architecture model for the safe, secure and transparent exchange of data between the many diverse producer and (possible) consumer of industrial data. Today, it is an explicit goal of the initiative to take both the development and use of the platform to an international level.

The most important requirements to be met by the reference architecture model are summarized in Table 2.

The reference model consists of four architectural elements:

- the business architecture addresses questions regarding the economic value of data, the quality of data, applicable rights and duties (data governance), and data-management processes,

Table 2 Technical challenges and solutions (Kubach 2016)

| Problems | Direction of solutions |
|---|--|
| Complexity Complex technology stack Systems of systems | Ecosystem Single point of contact End-to-end solution Openness and standards |
| Heterogeneity Hardware Scale Data volume | Technology portfolio No one-fits-all size Multiple data base, messaging technologies |
| Security and privacy Huge diversity of IoT device types with different capabilities and deployment scenarios Global distribution of assets Physical access to assets Limited penetration of security mechanisms | Security-by-design Security from the very beginning Data sovereignty |

- the security architecture addresses questions concerning secure execution of application software, secure transfer of data, and prevention of data misuse,
- the data and service architecture specifies in an application and technology independent form the functionality of the IDS, especially the functionality of the data services, on the basis of existing standards (vocabularies, semantic standards, etc.),
- the software architecture specifies the software components required for pilot testing of the IDS.

Central elements of this architecture (Fig. 3) are the connector for the exchange of data, the broker for the mediation of data offers and requests and an app store. The app store shall offer software code which can be injected into the connector to enrich the data with additional value (from meta-data to analytics).

6 Industrial internet consortium (IIC) and the W3C web of things interest group

IIC is an international, open membership, not-for-profit consortium trying to define an architectural framework for an industrial internet. The mission is to coordinate ecosystem initiatives to connect and integrate objects with people, processes and data using common architectures and open standards (<http://www.iiconsortium.org/IIRA.htm>, dated August 29, 2016).

While the focus of RAMI 4.0 is on manufacturing in depth, with representations of the level of hierarchy, life cycle and value-stream level of an object, the Industrial Internet Consortium (IIC) is propagating a different approach: The IIC reference architecture (IIRA) focuses on cross-multiple application domains and aims to provide guidance for the development of systems, solutions and application architectures. Although IIRA is at a high level of abstraction, it will deliver the foundations for the vocabulary and the design patterns for completely different IoT use cases.

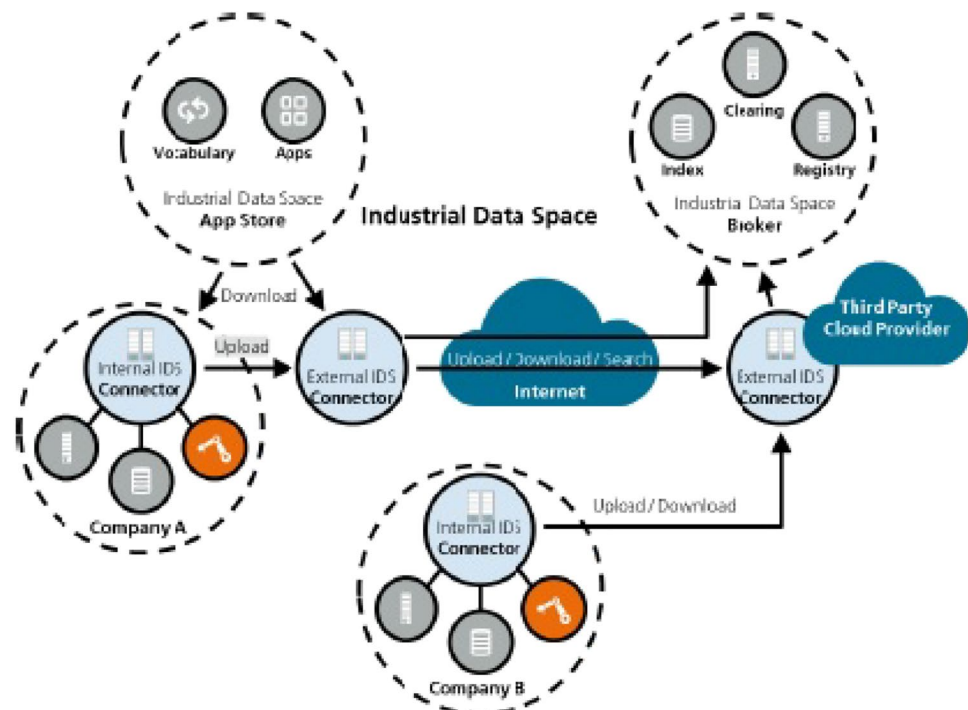
This difference between RAMI 4.0 and IIC can be outlined with an example from the automotive business:

- RAMI 4.0 addresses the shopfloor, i.e., the manufactured car with its many components,
- in an IIC scenario, a car is an element in a broader context, i.e., parked at home, charging its batteries by a connection to a smart grid, talking to other cars on the road, traffic light systems, etc.

To a certain extent, both approaches can be regarded as complementary.

The Web of Things Interest Group propagates a WebOfThings interaction model. This model is based on properties, events, and actions. Properties define the state of the thing and its configuration and settings; events are state changes in properties that the thing is able to report or send; actions are state changes which the application

Fig. 3 Software dimensions in the IDS (according to IDS 2017)



invokes on a thing (Koster 2016, <http://www.w3.org/WoT/IG>, dated August 29, 2016).

7 IoT platforms in the cloud

The ambitious goal to develop accepted, overarching IoT standards has to compete with a growing number of cloud solutions. These solutions allow for straight-forward implementation of IoT scenarios. The promise made by these IoT application enablement platforms (AEP) is to support affordable solutions by easily transferring sensor data to the cloud. Characteristics are ease and flexibility of deployment, scalability, developer-friendly user interfaces and cogent systems architectures. Essential features to enable cloud connectivity are, among others, network security, networking protocols, responsive performance, reliability and resilience, scalability (Ayla 2015). Analytics tools and dashboards enable users to track, monitor or match data to gain immediate insights. Furthermore, data can be forwarded to external systems like ERP or CRM (Fig. 4).

The claim is to be able to connect nearly any device to the cloud while offering all of the interfaces, tools and premium services needed for a manufacturer to manage, provision, and analyze their IoT deployment. With these Software-as-a-Service-platforms it is no longer necessary to invest in stack development, end-to-end security, infrastructure, and other IoT “must haves”.

8 Big Data technology

8.1 From Master Data Management to Big Data

Today, companies have to handle huge amounts of data, stored in a variety of data bases, driven by ERP software, CRM or shop systems. Software for Master Data Management (MDM) is useful to normalize data, detect and correct errors (data quality module) and enrich data—and can be used to create a single source of truth for enterprise critical data. Often, MDM software cooperates with a product management system (PIM) which adds pictures and further marketing relevant data. An Enterprise Service Bus (ESB) can be responsible for synchronizing any change of data via the MDM (Fig. 5).

However, MDM is not Big Data. A huge amount of IoT data, often generated in real-time, poses very specific challenges.

8.2 Characteristics of Big Data

From the IoT perspective, Big Data is a subset of the IoT technology where Big Data software addresses data handling and IoT takes responsibility for sensors, devices, and data delivery (Dull 2015).

Big Data scenarios are usually characterized by the volume, velocity, and variety of data. Additional criteria might be Vs like Validity, Veracity, Value, and Visibility. The basic intention is to collect as many data as possible to detect semantic patterns and correlations

Fig. 4 Platforms in the cloud (e.g., Ayla or Cumulocity)

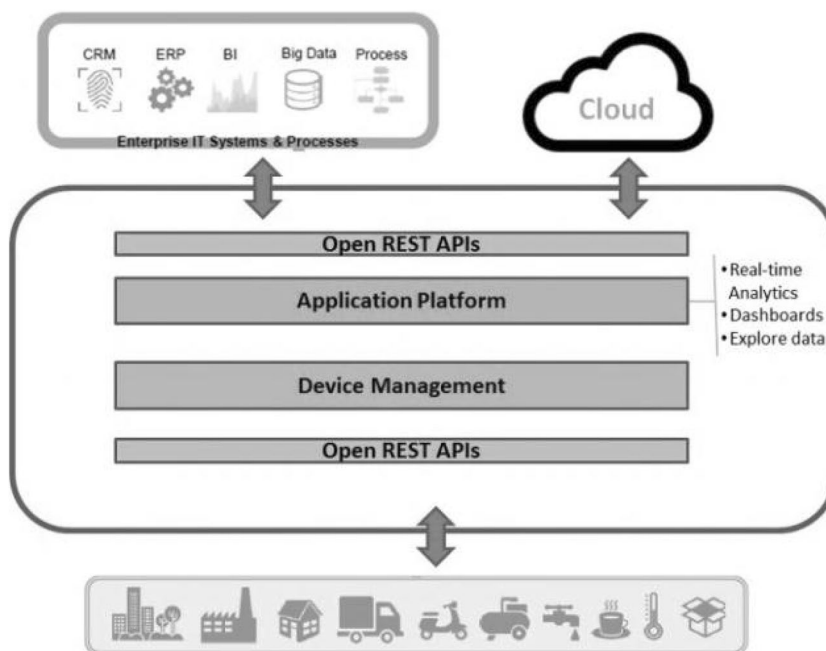
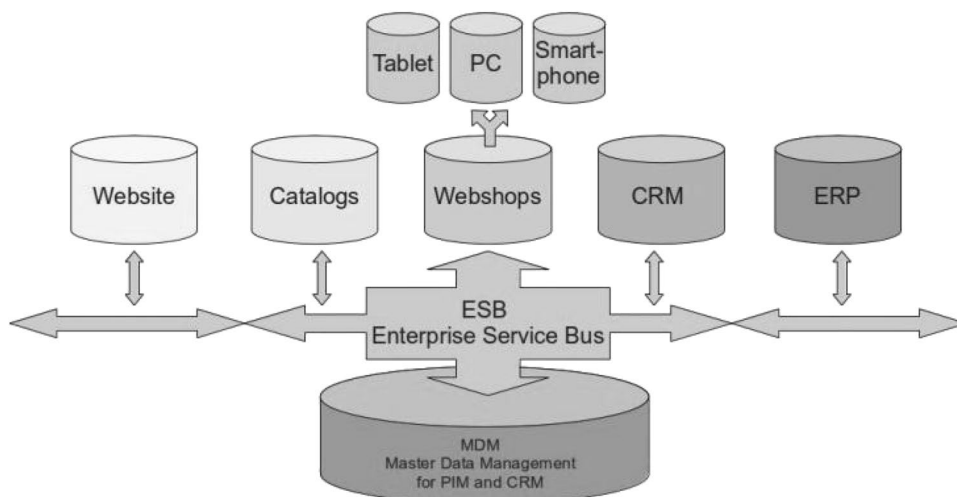


Fig. 5 Master data management (MDM) depending architecture based on an enterprise service bus (Sepia Altera, no year)



- in huge data oceans (volume) which,
- fill up continuously or event-driven (velocity),
- in a structured or unstructured format (variety).

IoT and Big Data are clearly connected intimately as billions of internet-connected “things” generate massive amounts of data (McLellan 2015).

Especially in scenarios with a constant in-flow of data (real-time) it is necessary to efficiently store and analyze data on-line and off-line. This challenge requires new storage concepts, high performance processing approaches, etc. The Hadoop ecosystem is at the forefront of the technology.

8.3 The Hadoop ecosystem

With relational database technology alone Big Data is not possible, cumbersome or expensive. The open source Apache Hadoop software library is the most prominent open source framework that allows for the distributed processing of large, structured and unstructured datasets across clusters of computers (Inmon and Linstedt 2015). Hadoop is designed to scale up from single servers to thousands of machines, each offering local computation and storage.

A complex Hadoop ecosystem has continuously evolved with tools and frameworks for different types of storage, processing, data integration, resource management, security, analytics, search, and data discovery. This evolving ecosystem is based on numerous software “modules” (Hadoop 2016):

- Hadoop Distributed File System (HDFS) provides high-throughput access to application data in a singular virtual place; HDFS facilitates rapid data transfer among the nodes in a cluster, ensuring resilience in case of node failures. The data can be of any format, i.e., structured, semi-structured or unstructured.

- YARN is a framework for job scheduling and cluster resource management (SQL interactive engines, batch engines or real-time streaming engines).
- MapReduce enables the parallel processing of large sets of data in distributed clusters.
- Sqoop provides UNIX-based commands to import and export data from RDBMS to HDFS and vice versa.
- Cassandra, Hbase and the like are NoSQL databases designed to handle large amounts of data across several nodes in a cluster setup.
- Apache Kafka is a tool for reliable ingestion of high volume streaming data.
- Hive and Impala are SQL interfaces for data summarization and ad hoc querying.
- Mahout is a scalable machine learning and data-mining library.
- Pig, a high-level data-flow language and execution framework for parallel computation.

There are manifold technical details and limitations which must be carefully analyzed with respect to the specific application scenario. High-level requirements for a Big Data architecture must address (Fig. 6):

- Reliable data ingestion
- Flexible storage and query options, and
- Sophisticated analytics tools.

To resolve the issue of latency with a Hadoop system, the Lambda architecture was developed (lambda-architecture.net). The basic notion of Lambda is to serve a batch and a speed layer where the batch layer collects the data for off-line analysis and the speed based on Spark (in-memory technology) handles the most recent data by creating real-time views.

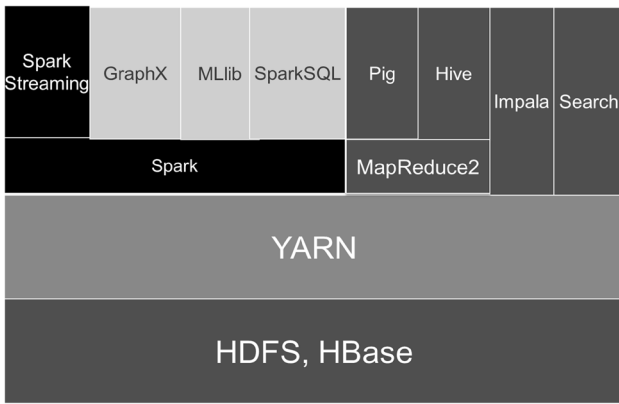


Fig. 6 Hadoop ecosystems and its major components dark and middle grey: Hadoop core, black: spark components, light grey: unsupported add-ons

The Hadoop ecosystem is gradually turning into a general-purpose data-operating system and is the leading framework for dealing with an extraordinary influx of datasets, often with a need to compute in real-time.

9 Platforms for data analytics

The value of IoT lies in the data generated by the connected objects and the ability to identify value in data is a prerequisite for successful B2B. The question is how to become a data-driven enterprise. The answer includes a re-evaluation of the given data integration solution and an analysis of the current software development life-cycle procedures (Wilmer 2016) (Fig. 7).

While methods for descriptive and diagnostic statistical analytics have been well-established for decades it is now possible to address predictive and prescriptive analytics, i.e., to get an insight into “What will happen?” and “How can we make it happen?” Consequently, innovative companies transform themselves into “mathematical

corporations” with data stewards, data scientists, data architects, data artists or information brokers (Charan 2015; Provost and Fowcett 2013).

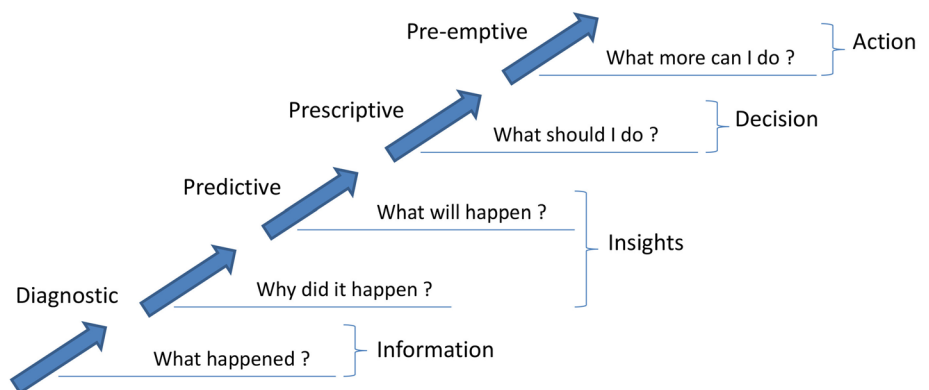
IoT solutions in the cloud and Big Data frameworks have embedded analytics tools, but these tools have to compete with advanced products equipped with sophisticated statistical and machine learning algorithms. Furthermore, self-service requirements pose very specific challenges concerning the usability of the tools.

A close link between Big Data architectures and analytics is essential for the implementation of predictive analytics. The preferred programming language R and tools (libraries and models) for statistics and machine learning have paved the way for fast embedded analytics and self-service for technicians and managers. These tools combine methods like regression analysis, similarity matching, clustering attempts, profiling and many more (<http://www.tibco.com>, dated August 29, 2016).

Nevertheless the significant technical progress: Any statistical or machine-learning toolbox requires a clear understanding of four business dimensions in advance (Provost and Fawcett 2013):

- timeliness, i.e., companies need to understand what timely information is for a specific business case (expiration speed),
- data organization, i.e., data needs to be pre-processed and organized in a way that it can be further analyzed (data preparation),
- accuracy, i.e., companies need indicators to define the quality of the data,
- relevance, resulting only from the specific business requirements.

Fig. 7 From hindsight to insight to foresight (based on HP 2014)



10 Challenges

10.1 Technical challenges

IoT requires high system stability, complete coverage and guaranteed low latency. Any system failure or connection loss can cause damage running into millions (Table 3).

The complexity resulting from the heterogeneity of the different hardware and software components poses a new challenge. Security and privacy issues need special attention for obvious reasons. Mature software platforms have to address these dimensions. There is no “one-fits-all-needs” solution but CIOs have to carefully evaluate what fits best to their line-of-business.

The preferred answer to this technological complexity of Big Data is open source technology based on Hadoop. A discussion of the advantages of open source software can be found in (Jesse 2014). However and in general, the challenges of security and privacy for IoT have not been solved yet and are expected to pose a growing threat (McLellan 2015).

10.2 Transformational challenges for business

From the perspective of business development, IoT, Big Data and analytics disrupt businesses and address numerous action items:

- companies have to align and balance their interests with other companies along the value chain and end users who create, own, or service a product (the data source),
- they become hardware and software companies at the same time; embedded software needs to be updateable and supported, probably with numerous versions (additional complexity),

- as data is constantly streamed, the classic information pull (gather, analyze, decide) has to be complemented by a real-time business process,
- the vulnerability of companies is increasing significantly and IoT has to invest in security to guarantee physical control and monitor any manipulation of data (Ransbotham 2015; Bughin et al. 2015).

Moreover, management has to align their IT and operational strategy tightly and create new organizational responsibilities:

- chief financial, marketing, and operating officers as well as leaders of business units will have to disrupt their silos,
- companies need to endow employees with new skills, so the organization becomes more analytically rigorous and data driven,
- analytics experts and data scientists must be connected with executive decision-makers and with frontline managers to give impact. In some cases, the decision-makers will be algorithms (Bughin et al. 2015; Provost and Fawcett 2013).

11 Conclusions

Sensor technology—from simple proximity measuring to complex bio-sensing—is developing fast. Numerous connectivity standards are available. In this document, the focus is on the need for software standardization and the contribution of software platforms to handle the unprecedented complexity of applications. Data, created by sensors, must be enriched by meta-data to provide meaning. Programming languages, data-encoding formats and protocols need to be regarded with respect to their relevance for IoT. Identity management for devices, users, application and services has to be addressed. To verify data and metadata,

Table 3 User requirements for an industrial data exchange

| User requirements | |
|---------------------------|---|
| Data sovereignty | The data owner specifies the terms and conditions of use for data |
| Decentral data management | Data management remains with the data owner, if desired |
| Data economy | Data is viewed as an economic asset. It can be distinguished into three categories: private data, i.e., data belonging to a specific value creation chain, which is available to selected companies only, and public data (weather or traffic information, geo data etc.) |
| Value creation | The IDS facilitates the creation and use of smart services and digital business models |
| Easy linkage of data | Linked-data concepts and common vocabularies facilitate the integration of data between participants |
| Trust | All participants, data sources, and data services of the IDS are certified against commonly defined rules |
| Secure data supply chain | Data exchange is secure across the entire data supply chain, i.e. from data creation to data capture to data usage |
| Data governance | Participants jointly decide on data management processes as well as on applicable rights and duties |

their provenance and the location of the sensors are relevant. The modelling of these tightly interconnected items leads to further questions.

IoT is on a fast accelerating path with evolving standards, technologies and platforms. As of Jan 2016, with over 275 vendors and products in the data platform and analytics landscape (451 Research 2016) it is no surprise that IoT and Big Data suffers from a lack of interoperability with data silos, high costs and limited market potential.

From the business perspective, integration has a vertical and horizontal dimension. On the vertical level, technical processes will be integrated with business processes. From a horizontal perspective, IoT builds a bridge between the boundaries of companies and integrates the complete value chain.

IoT and Big Data technology come from different backgrounds. IoT is driven by sensor technology and, more generally, from a hardware perspective. Big Data, however, has deep roots in new software paradigms developed by the Internet and social media enterprises like Google, Facebook or Yahoo. Hardware and software are two sides of the same coin. In this paper, we have briefly addressed the software side, i.e., concepts and platforms as enablers for huge data-driven IoT solutions. Furthermore, we indicated the need for powerful tools to analyze the deeper value of data.

At the end of the day, IoT opens up considerable opportunities. With the open-source Hadoop ecosystem, accepted exchange formats and a growing set of standards, a convergence or “blending” of platforms is on its way.

Acknowledgements The author is grateful for the information provided by Talend Inc., TIBCO Inc., Ayla Networks Inc. and Cumulocity GmbH as well as discussions with Dr. Gero Presser.

References

- 451 Research (2016) Open source big data projects—emergence of the converged data platform, Executive Brief 2016
- Ayla Networks (2015) Build vs. buy. Manufacturers’ Biggest IoT Decision: Build or buy an IoT platform? White Paper
- Bughin J, Chui N, Manyika J (2015) An executive’s guide to the Internet of Things The rate of adoption is accelerating. Here are six things you need to know. McKinsey Quarterly, August 2015
- Charan R (2015) The Attackers’ advantage. Turning Uncertainty into breakthrough opportunities, New York
- Cloudera (2016) Webinar
- Dull T (2015) Big data and the Internet of Things: Two sides of the same coin? SAS Best Practices. http://www.sas.com/en_ph/insights/articles/big-data/big-data-and-iot-two-sides-of-the-same-coin.html
- Gartner (2014) Gartner says 4.9 billion connected “Things” will be in use in 2015. In 2020, 25 Billion Connected “Things” Will Be in Use, Nov. 2014. <http://Gartner.com/newsroom/id/2905717adoop.apache.org/>
- Gruhn, Volker 2017. Big Picture: Engineering CPS. Presentation for the CPS.HUB NRW, April 28, 2017
- Hauptfleisch K (2015) Über machine-to-machine und Internet der Dinge zur Industrie 4.0, Computerwoche, 11.07.2015
- Heck D, Franco J-M (2016) Big Data at Air France, talend internal paper
- Hewlett Packard Enterprise (2014) From hindsight to insight to foresight. Extend analytical capabilities with the HPE Vertica Analytics Platform. Technical white paper <http://www.libelium.com/casestudies>. Dated August 29, 2016 http://www.zvei.org/Downloads/Automation/ZVEI-Faktenblatt-Industrie4_0-RAMI-4_0.pdf. Dated August 29, 2016
- IDS (2017) <http://www.industrialdataspace.org/en/>
- Inmon WH, Linstedt D (2015) Data architecture: a primer for the data scientist. Big Daga, Data Warehouse and Data Vault. Amsterdam etc
- Jesse N (2014) Boosting Government Performance with Open Source Software?—A Roadmap for Germany/¿Impulsar el desempeño del estado con software de código abierto? - Plan de trabajo para Alemania, Conferencia Científica Internacional (UCIENCIA 2014), La Habana
- Koster M (2016) Data models for the Internet of Things. <http://iot-datamodels.blogspot.de>
- Kubach U (2016) Das Internet der Dinge als Enabler für Industrie 4.0 Anwendungen. Talk at Hannover Messe, April 28, 2016. <http://lambda-architecture.net/>
- Lauby Sh. (2015) Your data is not magic—friday Distraction. <http://www.hrbartender.com/tag/big-data/>. Accessed 18 Sep 2015
- McLellan (2015) The internet of things and big data: Unlocking the power. ZDNet Special Feature: The Power of IoT and Big Data, March 2, 2015
- Milenkovic M (2015) A case for interoperable IoT sensor data and meta-data formats. Ubiquity symposium: the internet of things. Ubiquity, Volume 2015, Issue November 2015
- Provost F, Fawcett T (2013) Data science for businesses. What you Need to know about data mining and data-analytic thinking. Sebastopol
- Raggett D (2015) Building the web of things, posted May 29, 2015. <http://www3.org/blog/wotig/2015/05/29/building-the-web-of-things/>
- Ransbotham S (2015) Ready or not, here IoT comes. MIT Sloan. Blog December 22, 2015. <http://sloanreview.mit.edu/article/ready-or-not-here-iot-comes/>
- Sepia Alterra MDM-integration in die Unternehmens IT. Düsseldorf
- Simmons Th (2014) Sensoren und Messtechnik: Schlüsseltechnologien des technischen Fortschritts. all-electronics.de. 27.05.2014
- Siprell St (2016) Datenverarbeitung mit Sensoren. Javaspacerturm 3/2016, p 50 ff
- Trankler R, Kanoun O (2001) Recent advances in sensor technology. In: Instrumentation and measurement technology conference (IMTC), 2001 May 21–23. Proceedings of the 18th IEEE (vol 1)
- VDI nachrichten (2016) Big Data weiter im Höhenflug. April 29, 2016, Nr. 17/18
- Wilmer D (2016) The evolution of ETL and continuous integration. Talend blog. 15. June 2016. <http://de.talend.com/blog/>