

# A new architecture for Cognitive Internet of Things and Big Data

Mohammed Saifaeddine HADJ SASS, Faiza GHOUZZI JEDIDI

January 21, 2023

## 1 Summary

### 1.1 Introduction

Big data and Internet of Things(IoT) are considered as the main models for defining new information architecture projects. The new architecture proposed in this paper between Cognitive Internet of Things(CIoT) and Big Data Analysis benefits computing mechanisms and defining the tool. The increasing volume, variety and velocity of the data generated through Iot sensors will continue to fuel the explosion of data, it also gives another responsibility to analyze data and make decisions based of quantitative analysis. The proposed big data tools in this paper are capable of handling masses of data received through the Iot devices. Using this kinds of innovative architecture.

### 1.2 Related Works

There are many various systems that exhibit similar properties but not using Iot. But the tool is capable to be a solution for the heterogeneous data sources. The time of the data and the real-time consideration for the analysis process is slow because of increasing the amount of data and using complex algorithms. Moving the data ingestion tool helps for processing the complex data transformations. High-speed links connect to enterprise data through a programmable interface to ease seamless resource management and control physical resources. Knowing which techniques and tools can fit well in data flow process is important. Authors discuss IoT architectures and classify the architectures by their domains.

#### 1.2.1 Classification of Iot Architectures

- Domains: Domains for smart tourism, logistics
- Services: Mobile ticket booking, Netbanking
- Layers: 3-layered architecture, SoA-based Architecture
- IoT Projects: BeTaas, OpenIot

- Wireless network: Zigbee, 6LoWPAN
- Framework: cascades, I-core
- Security: Trust Management, Security and Privacy

Process of data flow, data control and technical constraints have not been discussed. We study and analyze existing technologies, tools, and techniques from the related works and several survey papers.

### 1.3 Open Research issues

Cognitive Internet of Things (IoT) concept has come a long way but there are still some open research challenges in the field of data management and analysis. These include data collection, data Ingestion, data storage and data analysis as dynamic heterogeneous resources, data-centric issues, scalability, reliability, fault tolerance, context awareness, cost, security, privacy, and data quality. In the following, we discuss some of the technical challenges due to 3V (volume, variety, velocity) problems.

- Volume: The IoT devices generate high volume of data every millisecond and transmitted rapidly, hence the volume of incoming data is high. Some data store can have an insufficient store space such as the Hadoop cluster that needs to have space for MapReduce jobs, other workloads and for the data storage requirements. Data volumes coming from sensors or other data sources can bypass the storage capacity of a Hadoop cluster.
- Variety: Entropies find it so difficult to deal with the raw, semi-structured, and structured data by traditional tools. Interoperability between various sources of data from different systems is very challenging in IoT because of heterogeneous nature of data.
- Velocity: Data delivery in big data has to be able to handle high velocity i.e. the ingestion from various devices to the database would be continuously moving at a certain speed. The frequency of data is the rate at which massive amounts of data are shared and processed.

### 1.4 Basic Concepts of IoT

The Internet of Things (IoT) allows people and things to be connected at any time, anyplace, with anything and anyone, ideally using any path/ network and any service. The IoT bundles many emerging technologies such as sensors, actuators, semantics, context-aware computing, big data analytics, communication technologies, data lake, service management. In this section, the authors presented the following subsections as the most IoT sub-layers for IoT system. The intention will be oriented on the generic IoT architecture, well on the concept of cognitive IoT paradigm and architecture.

#### 1.4.1 Generic Iot Architecture

- Perception layer: It has the ability to perceive, detect, gather and exchange information with various IoT devices. Obtains data from sensors such as cameras, sensors and RFID
- Network layer: This layer forwards the data received from perception layer to application layer using internet. This layer can be divided into 2 layers:
  - Data Exchange layer: Handles transparent transmission of data
  - Information Integration layer: Aggregates, cleans, extracts the data.
- Application layer: It creates a smart environment. It receives information and processes it and deliver to end user. Seamless integration between IoT environment with intelligence leads to cognitive IoT.

#### 1.4.2 Cognitive IoT

Cognitive Internet of Things (CIoT) is the merge of cognitive computing technologies with collected data from connected devices. The evolution of ubiquitous computing leads to heterogeneous infrastructure challenges. Cognitive IoT handles those challenges represented in context awareness by producing a smart system. Data produced by IoT devices and web data are sensing through context management middle-ware.

#### 1.4.3 Cognitive IoT architecture

- Service: Provides frameworks for intelligent services
- Knowledge and decision making: Controls the services, manages information and performs Big Data Analytics
- Information extraction and context management: Performs Data consumption and integration.
- Perception and data exchange: Senses the data through IoT devices or Web data.

#### 1.4.4 Data flow based CIoT

The Internet of Things (IoT) is exploding, and entropies are constantly trying to keep up with the increasing volume, variety, and velocity of data produced by these devices. Entropies finds it so difficult to deal with the raw, semi-structured, and structured data, various system architectures are used to handle data flow from an input source to the desired output that is desired.

#### 1.4.5 Architecture of data flow

- Data sources: The data is obtained from the source.
- Data collection: Data collection is dealing with the collection of data (actively or passively) from a device, system, or as a result of its interactions.
- Extract, Transform and Load ETL: Has the most important components which are extract, transform and load process and allows data manipulation from various data sources.
- Data Ingestion: Transformation of input data source to output data source.
- Data Storage: Huge volume and velocity of collected data is stored in a data storage.
- Data Wrangling: Process of mapping raw data to another format in which the required data is to be identified, extracted, cleaned and integrated.
- Business Analytics: Business analytics will be able to explore data in preparation for data modeling such as service.
- Data Analysis: Creates insight from the output of data wrangling.
- Service: Creates visual representation of the data.
- Data Distribution: Analyzes results and makes information and displays as a conceptual model or application interface to end-user.
- Metadata: The provenance of each data item such as the different data processing, transformation phases, and analysis techniques.
- Data quality: Handles data quality problems in the phases.
- Privacy and Security: Uses all the methods and techniques to ensure the privacy of the stored data and prevent data losses.

#### 1.4.6 New architecture for CIIoT and Big Data

A tool is required to collect data from various sources using smart device features such as human sensors, user input, documents, environmental sensor, and movement. This tool can extract and recognize data from unstructured data which is a big challenge for smart devices. In the knowledge and decision making layer, the output data from this tool and other data can be extracted and loaded into a central storage such as "DL".

## 2 My Views

The architecture proposed in this paper is suitable for all projects that deal with large quantity of data. Users can configure the tool without any technical assistance. The authors proposed automated functions such as capturing data through sensors to locate and track the user movements. The architecture solves the real time consideration by sending the collected data to be ingested in the problem. The speed of data flow in the architecture can be controlled using data analysis technologies such as ELT. Data Lake and Data Warehouse are combined to improve complex data transformation.

The architecture proposed in this paper has been created by analyzing the existing technologies, tools and techniques. There are some limitations due to 3V (Variety, Velocity, Volume), which has been solved by merging Data Lake and Data Warehouse. The architecture proposed in this paper can be very useful since the entire process from data collection to analysis and presentation of data analysis is performed using CIIoT is very useful. This tool solves the speed of data flow. In future the tool can be enhanced using new methods such as deep learning for extracting and recognizing the data from data sources that could be used to improve the data collection, in addition to real cases of the implementation of the approach.

## 3 Agreement, Pitfalls and Fallacies

- Big data and Internet of Things are considered as the main paradigms when defining new information architecture projects
  - Agreed to an extent
  - Both the technologies are useful and can work in hand in hand since entire data control flow is covered.
- The architectures which have described in earlier make cognitive IIoT concept feasible, there are still some open research challenges.
  - Agreed
  - There are some limitations due to huge data generated every millisecond. The data generated can't be stored without proper technologies.
- Cognitive IIoT infuses intelligence into the physical world through physical objects.
  - Not Agreed
  - Every logical thinking a human can perform can't be replicated or can't be solved by CIIoT.
- The increasing volume, variety, and velocity of data produced by the Internet of Things will keep going to fuel the information exploding.

- Agreed
  - Huge data will keep the information and data received exploding only since the data generated is huge.
- Our architecture is suitable for all projects that deal with the large quantity of data. In addition, users can configure the tool without any technical assistance.
  - Agreed
  - Any user can understand this architecture since it is a non-technical and easy concept.

#### **4 Submitted by:**

- Bhavesh Kumar B
- 21011101031
- AI DS-A