Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 3159-0040
Vol. 2 Issue 9, September - 2015

Introduction to Big Data Management Based On Agent Oriented CyberParks

Jamal Raiyn
Computer Science Department
Al Qasemi Academic College
Baqa Al Garbiyya, Israel
raiyn@qsm.ac.il

***Abstract*—Big Data is rising rapidly, with new multiple dynamic parameters. Big Data consists of data that change their attributes in accordance with their environment over time and take on new characteristics. To process this kind of data, a new adaptive functionality is needed. We call this kind of evolution data. Evolution data are produced in a dynamic environment under new conditions. In this paper we will consider evolution data that are produced by different ICT resources in public places. To manage this kind of Big Data, we propose the use of an evolution agent. The main task of an evolution agent is to learn the environment and to update its rules in accordance with environmental changes.**

Keywords—Big Data management, evolution agent, ICT.

I. BACKGROUND AND MOTIVATION

The amount of data is on the rise in various fields. The great number of internet technology users produces huge amounts of data, and they are on the increase. This kind of data is known as Big Data [1]. Big Data is characterized by three aspects according to Madden (2012)[2]:

- The data are numerous,
- The data cannot be categorized into regular relational databases, and
- The data are generated, captured, and processed very quickly.

Big Data has generated significant interest in various fields, including the manufacture of healthcare machines, banking transactions, social media, and satellite imaging. Big Data challenges have been described by Michael and Miller (2013)[3], such as rapid data growth, transfer speeds, the diversity of data, and security issues. Big Data is still in its infancy stage and has not been reviewed in general. Hence, this study comprehensively surveys and classifies the various attributes of Big Data, including its volume, management, analysis, security, nature, definitions, and rapid growth rate. The development of new IT technologies has rapidly increased the volume of information, which cannot be processed using existing technologies and methods [4-6]. In computational sciences, Big Data presents critical problems that require serious attention [7]. In the IT industry as a whole, the rapid rise of Big Data has generated new challenges with respect to data management and analysis. According to Khan et al. (2014)[9] , five common issues involve: volume, variety, velocity, value, and complexity. Madden (2012) [2] note additional issues such as the fast growth of volume, variety, value, management, security, and efficiency. In some fields data have grown rapidly. However, the type of data that increases most rapidly is unstructured data. This type is characterized by “human information” such as high-definition videos, movies, photos, scientific simulations, financial transactions, phone records, genomic datasets, seismic images, geospatial maps, e-mail, tweets, Facebook data, call-center conversations, mobile phone calls, website clicks, documents, sensor data, telemetry, medical records and images, climatology and weather records, log files, and text. According to Khan et al. (2014)[9], written in *Computer World*, unstructured information may account for more than 70% to 80% of all data in organizations. Currently, 84% of IT managers process unstructured data, and this percentage is expected to drop by 44% in the near future [9]. Most unstructured data are not modeled, are random, and are difficult to analyze. According to the Industrial Development Corporation (IDC) and the EMC Corporation, the amount of data generated in 2020 will be 44 times greater by 40 zettabytes (ZB)] than in 2009.

Big Data technology aims to minimize hardware and processing costs and to verify the value of Big Data before committing significant company resources. Properly managed Big Data are accessible, reliable, secure, and manageable. Hence, Big Data applications can be applied in various complex scientific disciplines (either single or interdisciplinary), including atmospheric science, astronomy, medicine, biology, genomics, and biogeochemistry. Khan et al.

(2014)[9] have proposed a new data life cycle that uses the technologies and terminologies of Big Data. This new approach to data management and handling required in e-Science is reflected in the scientific data life cycle management (SDLM) model. With this model, existing practices are analyzed in different scientific communities. The generic life cycle of scientific data is composed of sequential stages, including experiment planning (for research projects), data collection and processing, discussion, feedback, and archiving. The proposed data life cycle consists of the following stages: collection; filtering and classification; data analysis; storing; Sharing and publishing; and data retrieval and discovery. In processing Big Data, users face several challenges [10].

Big Data requires a huge storage capacity; rapidly search, sharing, and analysis capabilities; and in some areas, data visualization. These and others challenges need to overcome to maximize Big Data.

Currently, various techniques and technologies are used, such as SAS, R, machine learning platforms and Matlab to handle extensive data analysis. However, the proposed schemes are limited in managing Big Data effectively and are still lacking. According to Khan et. al. (2014)[9], others challenges to Big Data analysis include data inconsistency and incompleteness, scalability, timeliness, and security.

This paper is organized as follows: Section 1 introduces agent technologies. Section 2 describes ICT resources management in Cyber Park. Section 3 introduces a new scheme for Big Data management based on evolution agent technology. Section 4 discusses open issues of big data and concludes the paper.

II. AGENT TECHNOLOGY

This section introduces agent technology [11]. There is no general definition for the term *software agent*. Basically, when people speak of a software agent, they are talking about a software component that operates fairly independently [12]. In a sense, the agent acts in order to accomplish a task on behalf of its user. The word *agent* is an umbrella term that covers a wide range of specific agent types. "An agent is a software thing that knows how to do things that you could probably do yourself if you had the time." [13]. A specific definition of *intelligent software agent* [14] that many agent researchers might find acceptable is: a software entity that functions continuously and autonomously in a particular environment, often inhabited by other agents and processes [13]. The requirement for continuity and autonomy derives from our desire to be able to carry out activities in a flexible and intelligent manner that is responsive to changes in the environment without requiring constant human guidance or intervention [12][15].

Ideally, an agent that functions continuously in an environment over a long period of time would be able to learn from its experience. In addition, we expect an agent that inhabits an environment with other agents and processes to be able to communicate and cooperate with them, and perhaps move from place to place in doing so Labrou et.al. 1999)[16]. Here we propose the adoption a proposal by Franklin and Graesser (1996)[17]. After analyzing several agent proposals, they offered a definition similar to that of Beer. They defined an autonomous agent as a system situated in, and part of an environment which senses that environment and acts on it. Over time, it pursues its own agenda so as to effect what it senses in the future. Wooldridge and Jennings [18] define an autonomous agent as possessing the following minimal characteristics:

- Autonomy: agents operate without the direct intervention of humans or others, and have some kind of control over their actions.
- Social ability: agents interact with other agents (possibly humans) via some kind of agent-communication language.
- Reactivity: agents perceive their environment, and respond in a timely fashion to changes that occur in it.
- Pro-activeness: agents do not simply act in response to their environment; they are able to exhibit goal-directed behavior by taking the initiative.
- Adaptivity: agents learn and improve with experience.
- Mobility: agents can migrate in self-directed ways from one host platform to another to fulfill assigned tasks, and most important, they exhibit intelligent behavior, including self governing and self-dispatching.
- "Knowledge-level"- communication ability: agents possess the ability to communicate with persons and other agents with language employing human like "speech acts" rather than with typical symbol-level, program-to-program protocols.
- Intelligence: the agent needs to be able to interpret monitored events to make appropriate actuation decisions for autonomous operation.
- Cloning: an agent must be able to carry over its capabilities, facilities, and state information into a new environment. This can be achieved by making a copy, or clone, of the original agent, and placing the new agent in the new environment. The clone must have all of the capabilities and facilities of the original agent, along with any state information.

A. *Classification of Agents*

Agents are classified according to some clear properties. Mostly, we distinguish between stationary (or static) and mobile agents [19].

- *Collaborative Agents*

Collaborative agents cooperate autonomously with other agents in order to complete tasks given by their users. These agents can be capable of learning, but this is not necessary. Collaborative agents possess, autonomy, social ability, responsiveness, and proactiviness. Hence they are able to act rationally and autonomously in open and time constrained multi-agent environments.

- *Interface agents*

Interface agents are autonomous and engage in learning with other agents in order to complete tasks given by their users. Interactions between a human and an interface agent do not necessarily require an agent communication language, as communication between agents does.

- *Mobile agents*

Mobile agents [15] use computational software processes and are capable of migrating from one computer to another via the internet or intranet. They interact with foreign hosts gathering information on behalf of their owners and then return home with the gathered information [20]. This sort of agent can be used in various different applications ranging from flight reservation to the managing of telecommunication networks [21]. Mobile agents are defined as agents because they are autonomous, and they co-operate with other agents [22]. When they co-operate, it is important that they exchange information in the same area of interest without necessarily giving it all away, and on the other hand, that they do not gather the same information twice, or transport to the location where it originated [19][23][25].

- *Information agents*

Information agents are tools for gathering a constantly growing body of information from the internet. Information agents perform the role of managing, manipulating and collating the retrieved and distributed information [26].

- *Reactive agents*

Reactive agents act/respond in a stimulus-response manner to the current state of the environment in which they are embedded. Reactive agents are relatively simple and interact with other agents in basic ways. Therefore, they tend to operate on raw sensor data that need to be processed quickly [27].

- *Hybrid agents*

Hybrid agents are a combination of various kinds of agents. Since each type has its own strengths and deficiencies, the aim is to find the best suitable solution for the problem at hand, such that deficiencies are minimized and the strengths maximized in a straight forward manner.

III. ICT MANAGEMENT IN CYBER PARK

This section introduces the use of ICT in cyber Park [29-30]. ICT resources produce various kinds of data, and various schemes have been proposed to manage data sets. Traditional management schemes can be categorized according to fixed and dynamic management strategies.

- *Fixed Management Strategy*

In a fixed management strategy, every ICT resource manages the data it produces in centralized way. Data sets can be managed based on historical and on real-time information.

- *Dynamic Management Strategy*

In dynamic management, new data produced are placed in a central pool. The data can be managed in a centralized or in distributed way [28].

- *Cognitive Management*

Cognitive management uses modern tools to manage new produced Big Data that is produced. A cognitive manager uses adaptive agents to manage the data. Adaptive agents can add new functionality to dealing with Big Data [4].

A. *The Structure of Cognitive Management*

The cognitive radio scheme consists of two main parts:

i. A resource agent:

ii. An evolution Agent: The Evolution agent follows changing environment conditions and updates its rules according to characteristics of new environments

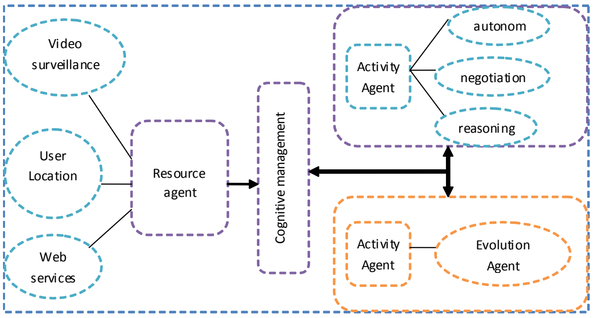

Figure 1: Cognitive radio infrastructure

B. *Methodology*

This section gives an overview of the cognitive management methodology as shown in Figure 2.

- All of data sets of various ICT resources are placed in a central pool (see Figure 3).
- Each ICT resource is assigned a resource agent (RA).
- The resource agent works in a dynamic environment: each resource agent is assigned several tasks.
- The resource agent has access to data shared from other ICT resource.
- The resource agent selects relevant data sets according to planned goals, as shown in Figure 4.
- The resource agent selects data according to the task to be performed.
- The advantage of a dynamic assignment strategy is to give the resource agent opportunities to access various kinds of data from others ICT resources.

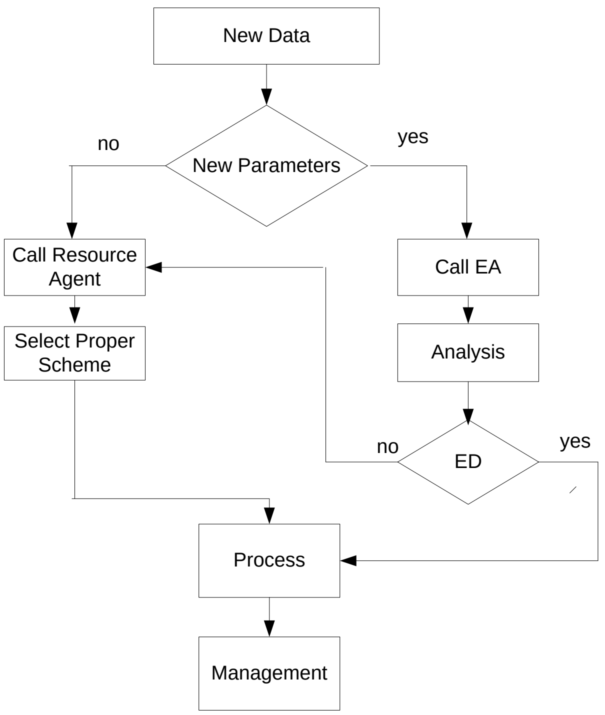

Figure 2: Methodology

C. *Description of the Method*

Data from various ICT resources, such as Video surveillance and mobile and location- based services, are placed in a pool. For instance video surveillance data are indicated as

$$R^{videoS} = \{r_1,\ldots r_n\}.$$

The newly produced data are called *evolution data* and are indicated as

$$\tilde{R}^{videoS} = \{\tilde{r}_1,\ldots,\tilde{r}_n\}.$$

In contrast to regular data, evolution data are produced under changing urban conditions. The evolution agent goes through fours main steps to characterize Big Data.

To determine the evolution data, the evolution agent performs the following steps:

Step 1

Each entity in the training databases is described by a set of binary descriptors (properties). The choice of descriptors is application dependent and requires knowledge of the specific domain in which it is applied.

Step 2

In this step, we determine the weight of a descriptor at each position within the frame . If a certain descriptor has a high rate of occurrence (a value of 1) at a certain position, then its weight increases.

Step 3

The score for each frame is calculated as a summary:

$$A(i,j) = R^{videoS} \oplus \tilde{R}^{video}$$

where *A (i,j)* is a factor for each weight at position *j*, *D(i,j)* is the weight of frame *i* at position *j*, B is the factor for *X* weight and *X* is the result of the vector XOR operation. At the first iteration each of the factors is 1.

Step 4:

In this step, we choose the frames with highest separation score and arrange them according to their natural order, excluding any frame (that corresponds to the original frames R^{videoS}. The separation score is then evaluated using the correlation method.

Step 5

The extracted data are an evolution data; an advanced scheme is needed to process this data.

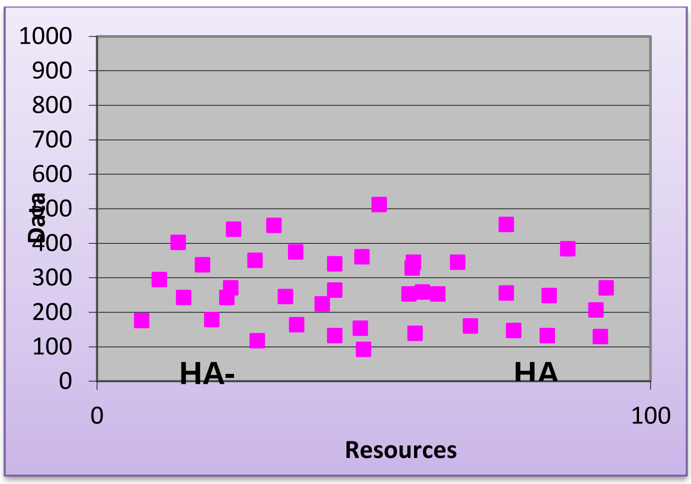

Figure 3: Data from ICT Resources

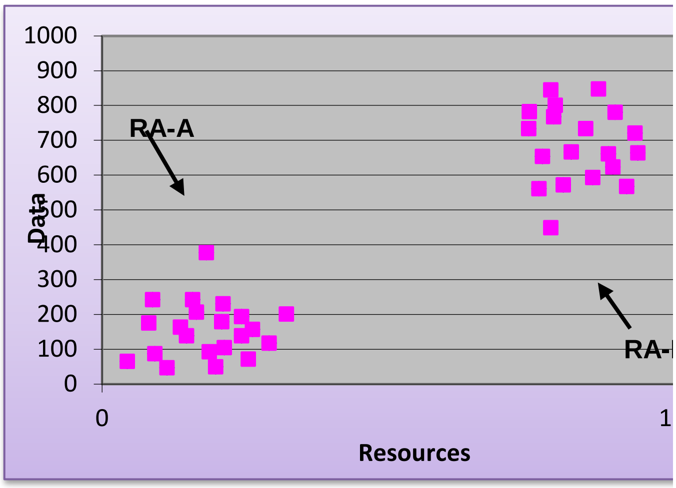

Figure 4: Extracted Big Data

IV. CONCLUSION

This paper has introduced a review of the term Big Data and has discussed its characteristics. The internet and mobile technology are growing rapidly, and the data accumulated over twenty years have becomes Big Data. We introduced a form Big Data management that is based on evolution agent technology. An evolution agent has the task of monitoring t newly produced data. It makes decision if the new data differ from historical data. In this, a case new scheme was proposed to process the Big Data.

REFERENCES

[1] D. Che, M. Safran, and Z. Peng, "From Big Data to Big Data Mining: challenges, issues, and opportunities in Database Systems for Advanced Applications", Springer, Berlin, Germany, 2013. pp. 1–15.

[2] S. Madden, "From databases to big data", IEEE Internet Computing, 16, 3, 2012. pp. 4–6.

[3] K. Michael. and K.W. Miller, "Big data: new opportunities and new challenges", Editorial: IEEE Computer, 46, 6, 2013, pp. 22–24.

[4] J. Raiyn, "Using Cognitive Radio Scheme for Big Data Traffic Management in Cellular Systems", International Journal of Information Technology and Management, 14, 2/3, 2015.

[5] J. Raiyn, "Toward Developing Real- Time Online Course Based Interactive Technology Tools", Advances in Internet of Things, 4, 2014. 13-19.

[6] J. Raiyn, "A Survey of Cyber Attack Detection Strategies", International Journal of Security and Its Applications, 8, 1, 2014, 247-256.

[7] C.A. Steed, D.M. Ricciuto and G. Shipman, "Big data visual analytics for exploratory earth system simulation analysis", Computers and Geosciences, 61, 2013. pp. 71–82.

[8] D.E. O'Leary, "Artificial intelligence and big data", IEEE Intelligent Systems, 28, 2, 2013, pp. 96–99.

[9] N. Khan, I. Yaqoob, I.A.T. Hashem, Z. Inayat, W.K.M. Ali, M. Shiraz and A. Gani, "Big Data: Survey, Technologies, Opportunities, and Challenges", Scientific World Journal, 2014. pp.1-18.AvailableOnline:http://dx.doi.org/10.1155/2014/712826.

[10] M. Hilbert and P. L´opez. "The world's technological capacity to store, communicate, and compute information", *Science*, 332, 6025, 2011. pp. 60–65.

[11] A.L.G. Hayzelden, J. Bigham, „Agent technology in communications systems: an overview", The Knowledge Engineering Review, 14, 4, United Kingdom, 1999. pp. 341-375.

[12] P. Antsaklis, "Defining Intelligent Control", IEEE Control Systems, 14, 3, 1999. pp. 58-66.

[13] M.N. Huhns and M.P. Singh, "Readings in Agents; Morgan Kaufmann Publishers, USA. 1998.

[14] H.S. Nwana. "Software Agents: An Overview", Knowledge Engineering Review, 11:3, 1996. pp. 205-244.

[15] J. Hartroth, aand D. Kottmann, Mobile Softwareagenten: Entwicklungsstand und Einsatzperspektiven; Theorie und Praxis der Wirtschaftsinformatik, heft 190, Juli, Huethig. 1996.

[16] Y. Labrou, T. Fini, Y. Peng, "Agent Communication Languages: The Current Landscape"; IEEE Intelligent Agent Systems, 14, 2, March/April, 1999. pp. 45-52.

[17] S. Franklin and A. Graesser, "Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents",, Proceedings of the 3rd International Workshop on Agent Theorie, Architectures, and Languages, Lecture, Notes in Artificial Intelligence, Springer Verlag, Heidelberg, August, 1996.

[18] M. Wooldridge and N. R. Jennings, "Intelligent Agents: Theory and Practice", Knowledge Engineering Review, 10:2, 1995, pp.115-152.

[19] S. Russel, and P. Norvig, "Artificial Intelligence: A Modern Approach", Englewood Cliffs, NJ, Prentice Hall, 1995.

[20] M. Luck, V. Marik, O. Stepankova, and R. Trappl, "Multi-Agent Systems and Applications", Springer, 2001.

[21] H. Helin, H. Laamanen, and K. Raatikainen, "Mobile Agent Communication in Wireless Networks", ITG-Fachberichte, European Wireless, VDE Verlag Berlin, 1999.

[22] T. Springer, T. Ziegert, and A. Schill, „Mobile Agents as an Enabling Technology for Mobile Computing Applications", Kuenstliche Intelligenz, Oktober, Heft 4. Oktober, 2000. pp.284-289.

[23] M. Dalmeijer, D.J. Hammer, A.T.M. Aerts, "Mobile software agents; Computers in Industry", 41, 3, 2000. pp. 251-260.

[24] J. Castro, "Software Engineering for Large-Scale Multi-Agent Systems", Springer- Verlag Berlin, 2003.

[25] M. Pechovcek, "Multi-agent systems and applications", 3rd International Central and Eastern European Conference on Multi-Agent Systems, Springer, Berlin, 2003.

[26] M. Klusch,"Intelligent Information Agents". Springer, 1999.

[27] Z. Lin, and K. Carley, "Proactive or Reactive: An Analysis of the Effect of Agent Style on Organization Decision-making Performance", Intelligent System in Accounting Finance and Management, 2, 1993. pp. 271-287.

[28] J. Raiyn, "A Self-Optimization of the Dynamic Resource Management Based on the Cognitive Radio", Wireless Engineering and Technology, 2, 87-92. 2011.

[29] J. Raiyn, "Information Security and Safety in Cyberpark",Global Journal of Advanced Engineering, 2,8, 2015.pp.73-78.

[30] J. Raiyn, "Modern Information And Communication Technology and Their Application in Cyperpark", Journal of Multidisciplinary Science and Technology, 2, 8, 2015.