The JUNIPER Project

Kevin Vu

University of California at Los Angeles

VuKevin@icloud.com

Abstract— **This document details a new platform under the *JUNIPER Project* in which to aid businesses in tapping into the full potential of big data. Funded by the European Union for nearly five years, the motivation behind the project includes discovering prototype technologies that can optimize and expand big data analytical software applications.**

I. Introduction

Since its launch in the winter of 2012, the JUNIPER project has backed the development, testing, and evaluation of "prototype technologies" [1]. This collective effort sets sights on integrating smarter and more efficient approaches to handle big data. The Internet today is responsible for carrying out millions of tasks—tasks that range from online business meetings to humoured cat videos [1]. But inasmuch as the capabilities of the current software applications have been successful in managing these data streams and stored data, unveiled projections show that data traffic volumes are "expected to grow twelve-fold by 2018." Within mobile networks, data traffic has seen a rise of nearly 5,000% in the last 12 quarters [2]. The global analyst firm estimated that 85% of the Fortune 500 has remained unable to exploit big data in 2015 [4]. As such, the JUNIPER project aims to provide a platform to assist in the performance of big data processing.

II. Big Data

In short, the term 'big data' entails complex and large data sets that traditional databases and software applications aren't entirely adept to handle. Big data is often subject to three primary aspects upon discussion: volume, variety, and velocity that respectfully correspond to the amount of data, the kind of data, and the speed of information flow [3]. While some attribute web behaviour and social network interactions to big data, big data also encompasses product information, financial transfers, and sale transactions [3]. In effect, big data comprises of both unstructured and multi-structured data.

A. Unstructured Data [3]

This type of data is derived from unorganized and mainly textual information. Meanwhile, it is difficult to utilize traditional data models and databases to represent/analyse them. Examples include:

- Instagram posts
- Snapchats

B. Multi-Structured Data [3]

This type of data is derived from interactions between individuals and machines. It has a plethora of data formats and types which, in turn, similarly make it difficult to represent via traditional means of data representation and processing. Examples include:

- Web log data (textual, visual, and transactional elements)
- Social platforms and customer interaction streams [4].

C. Big Data Processing

Processing of big data involves a data generator and client requests by application. A data generator produces streams of data that is then filtered before it can be stored. Following is the user request by application to "produce a reply back to the end-user" [1]. Figure 1 illustrates this.

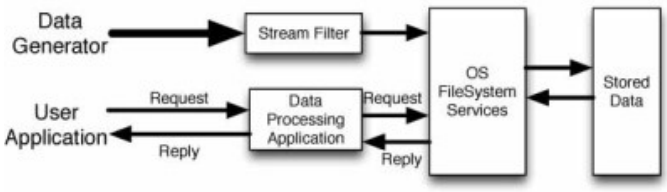

Fig. 1 Typical Big Data processing

III. The JUNIPER Approach & Results

The JUNIPER Project seeks to enable "efficient" and "real-time exploitation" of big data. It addresses potential performance, precision of data streams/data accesses, and scalability in terms of its platforms. Subsequently, the project has worked to create a Java platform that involves the exploits of streaming I/O, parallelism, and optimization [1]. In spite of limited resources, the JUNIPER project focuses its efforts to available, sufficient resources that include the processor and I/O. With this in mind, the project allocates these resources so "application processes such that real-time constraints can be met" [1]. More specifically, traditional systems are not fully aware of the increase in performance from added processing

power until run-time. The JUNIPER Project addresses this by constructing a system that comprises of real-time technologies that will allow for constraint information and added performance—all before run-time.

This four year long effort has amounted to a series of software developments published on GitHub under 'Juniper-Project'. Below is a preview of some of the releases and their accompanying features.

A. Platform

The platform created by the JUNIPER Project is a run-time environment that allow for the execution of applications developed by JUNIPER. It is written in Java and requires Open MPI & the Apache Maven build tool.

B. MONGO DB Modeller

The MONGO DB Modeller is a module that allows for representations via the MongoDB data models [1]. Essentially, the noSQL databases provided by MongoDB enable the integration of data more easily and efficiently.

C. Real-Time Java VM Extensions

The JUNIPER Project utilizes JamaicaVM to optimize systems and applications that handle big data while minimizing footprint (space). This development exactly reflects the previously stated vision of the JUNIPER Project, which included virtual machine extensions and class libraries to apply to scalable platforms [1].

These developments are just some of the releases of the JUNIPER Project. They, along with the rest in Table 1, have collectively outputted a means to exploit big data in terms of efficiency and real-time.

TABLE I
OTHER PROJECT RESULTS [1]

Software	**Description**
Monitoring Library	Writes and reads metrics values from an ElasticSearch database.
Offline MPO	The MDE Environment can run offline.
Caicos Acceleration	Acceleration of Java code
Real-Time OS Extension & Scheduling Analyser	Command line tool or integrated into environment, real-time
Modelling Environment	Modelling of big data real-time applications

IV. POTENTIAL APPLICATIONS

The JUNIPER Project has launched precedents on the development of big data while also expanding the big data horizon overall. Big data that can be processed and analysed cheaply and efficiently is highly desirable in today's world. As such, the developments and results of the JUNIPER Project mean a lot for the future of big data. This isn't merely to one single sector; rather, it is across industries and sectors [1]. For example, big data can be "of significant benefit" to retailers, auto industries, and even health industries.

A. *Auto Industries*

Machines and devices, aided by smart use of big data, become smarter and more automated [1]. For example, Telsa Motor's CEO Elon Musk has been working to build Tesla cars that will self-drive in as small as two years. However, Musk acknowledges that current data is not sufficient to fully support an automated vehicle. Heavy amounts of data are required to create an automated vehicle that is statistically sound in terms of safety compared to manual control [7]

"The data is not yet there to support a fully autonomous vehicle. " – Elon Musk

B. *Retail Applications*

Customer behaviour and preferences can be more easily tracked and analysed. Retailers want to know what shelves are stocking out, what online products see most success, which aisles are heavily visited, etc. Big data allows for the integration of such big ideas.

C. *Health*

Big Data analytics have a lot in store for revolutionizing the health provision [6]. More specifically, medical personnel will be more equipped with information to treat their patients. The health sector as it is now has 80% of the data not only unstructured, but also in varying forms of "lab results, images, and medical transcripts." For this, you would need a powerful analytics platform to process it and intelligence to provide relevant insights. This sheer importance has led to Apple and IBM's collaboration in using big data analytics for the health sector. Where IBM is expanding its global analytics Watson Health, Apple is creating a platform that can store health data via iOS applications. From this, it is clear of the place big data holds for healthcare.

Conclusions

Big Data in today's world is ever expansive. It exists across industries. Companies are receptive of the potential of Big Data; many even have started taking research and development into it. For example, work on big data exists in things such as Amazon's customer purchase tracking and Chevron's research for product optimization. More than that, news outlets have even reported how Target was able to figure

out that a teen girl was pregnant before her own father did from big data analytics.

With respect to the JUNIPER Project, big data has much potential to be unlocked further on the technological horizon. The platform that the team under the project has developed will be substantial in aiding businesses in the advancement of big data analytics.

REFERENCES

[1] "Community Research and Development Information Service - CORDIS." European Commission : CORDIS : News and Events : *New Platform to Help Business Tap Full Potential of Big Data*. N.p., n.d. Web. 16 Feb. 2016

[2] Chua, Godfrey. *Mobile Networks at the Tipping Point: The Data Explosion and Next Generation Network Challenge* (n.d.): n. pag. Print.

[3] *"What Is Big Data?"* Forbes. Forbes Magazine, n.d. Web. 16 Feb. 2016.

[4] "Multi-Structured Data: The New Innovation Opportunity - InformationWeek." InformationWeek. N.p., n.d. Web. 16 Feb. 2016.

[5] "A Java Platform for High-Performance and Real-Time Large Scale Data Management." GitHub. N.p., n.d. Web. 16 Feb. 2016.

[6] "Community Research and Development Information Service - CORDIS." European Commission : CORDIS : News and Events : Trending Science: Big Data Analytics to Revolutionise Health Provision. N.p., n.d. Web. 17 Feb. 2016.

[7] "Elon Musk Says Tesla Vehicles Will Drive Themselves in Two Years." Fortune Elon Musk Says Tesla Vehicles Will Drive Themselves in Two Years Comments. N.p., 21 Dec. 2015. Web. 17 Feb.2016.