OPEN SOURCE BIG DATA ANALYTICS SOFTWARE FOR BUSINESS AND ENTERPRISE

Balaji Janamanchi, Texas A & M International University

Abstract

Stated simply, big data analytics is the process of analyzing enormously large volumes of transactional and other business data to discover hidden patterns, develop useful insights, and identify actionable pointers that help managements take efficient and effective decisions. It is now possible to obtain such analytical metrics and decision support almost in real-time at little or no cost.

This study explores the availability and accessibility of big data analytics application software on the Sourceforge website, undeniably, a large repository of open source software resources. Limiting the search to big data analytics software under the "business & enterprise" category, data of 36 applications is downloaded in October 2016 and analyzed to identify the preferred features and to discern any meaningful trends. It is hoped that the results will help shed some light on the current state and profile of open source big data analytics software.

Keywords: Open Source Software, Big Data Analytics, Real Time Decision Support

INTRODUCTION

What Is Open Source?

The term "open source" refers to something people can modify and share because its design (or source code so to state) is publicly accessible. The term originated in the context of software development to designate a specific approach to creating computer programs (Opensource, 2017a). Open sources software (OSS) is quite distinct from normal usual software in that, open source software is developed by software programmers to be made available free-of-charge to intended users under rather unrestrictive set of terms of usage license (Opensource, 2017a). Primarily most open source software is available free of cost to interested end users who can in most instances download the software applications as they need for their use. However, given the nature of free-of cost supply, the support for software is somewhat lack luster and is nonexistent in some cases.

As the name suggests, "open source" refers to the source code of an application left open and accessible for the users to modify, extend, alter, amend and improve as they need or desire, to suit their usage requirements. Typically, when computer software is sold/provided to end-users, the proprietary software providers do not allow access to the source course of the software to the end-users. A very limited or no opportunity is given to end-users to customize the functionality and finer features of the application depending upon the software providers' marketing policies. This obviously leads to an avoidable dependence on software sellers for almost all issues with the use and implementation of applications at the end-user organizations.

What Is Source Forge?

> *"Sourceforge is a Web-based service that offers software developers a centralized online location to control and manage free and open-source software projects. It provides a source code repository, bug tracking, mirroring of downloads for load balancing, a wiki for documentation, developer and user mailing lists, user-support forums, user-written reviews and ratings, a news bulletin micro-blog for publishing project updates, and other features"* (Wikipedia, 2017).

SourceForge is a forum where software developers can host their software for others to use and improve upon, to be further shared and further improved by others. The process is somewhat similar to crowd funded projects. SourceForge provides developers the tools and facility to help create and promote open source software projects. As of July 2018, SourceForge is home to over 500,000 projects, is serving several millions of registered users and is supporting more than four million downloads a day. SourceForge is owned and operated by Slashdot Media (SourceForge, 2018).

What Is Big Data Analytics?

In simple words, "big data analytics is the process of examining large data sets to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information." It goes without saying that in a competitive business environment an organization that's able to harness the Big Data sources with analytics and is able to make right decisions in a timely manner steals a march over the competition. Cost reduction, better decision support and innovation opportunities are the big drivers of big data analytics adoption in business and enterprise based on General Electrics' wind farm optimization process (HBR, 2015).

According to Statistical Analysis System (SAS) software internal research team's white paper, "Why is big data analytics important?" the three main reasons why big data is important for businesses are, "Cost reduction, Faster and better decision making, and new product and services (creation of)" (SAS, 2017). This white paper was produced by interviewing over 50 businesses to understand how they are using Big Data in their business operations.

Who Uses Big Data Analytics?

Corporations, Business Organizations, Marketing Consultants, Government Agencies, Policy Makers, Predictive Modelers, Scientists and Researchers and whoever needs to make effective use of large volumes of data in making better decisions for better/effective results uses big data analytics. As stated in the above cited SAS research report, big data analytics support faster and better decision making coupled with cost reduction to users. Potentially, the results from the big data analytics open doors for and enable creation of new products and services.

CURRENT STUDY- OBJECTIVE

Prior Studies

Open source software projects have attracted considerable attention from researchers and other stakeholders from multiple perspectives and continue to be researched as OSS projects are evolving into dependable and viable alternatives to proprietary software.

Leaner and Tirole (2005) studied open source project licensing preferences of licensors and the community of developers for different types of applications with the objective of understanding what type of projects tend to have restrictive licenses in contrast to projects that tend to have less restrictive licenses.

Marsan and Pare (2013) studied, antecedents of open source software adoption in health care organizations in Canada and concluded that it's desirable to consider the factors influencing the OSS adoption together as a whole rather than each in isolation of others.

Rafiq (2009) surveyed the Library Information System (LIS) community around the world to assess their perceptions of adoption of OSS in libraries. Rafiq discovered that the regardless other categorical characteristics, respondents from developed countries held significantly different perceptions about adoption of OSS in libraries as opposed to the respondents from developing countries.

Sarrab and Hussain Rehman (2014) presented a case study of OSS selection for adoption based on software quality characteristics. Specifically, they categorized the quality characteristics into system quality, information quality and service

quality. Within each category they adopted the characteristics established in prior researches to identify 13 significant items in all, viz., availability, reliability, performance, usability and functionality under system quality; maintainability, reusability, testability and security under information quality; commercial support, community support, documentation and developer skills under service quality.

Gallego, Bueno, Racero, and Noyes (2015) researched the effect of user-training and user-fit on adoption of OSS and concluded that the user characteristics, technological complexity and trainers support played an important role in the success of adoption of OSS solutions.

Roumani, Nwankpa and Roumani (2017) studied the reasons that lead to the potential adopters' trust in enterprise open source vendors and identified that vendor's security, vendor's embracement of OS standards and vendor's support services are the factors that promote such trust. They concluded that system trust is effective in increasing the adopters' attitudes and intentions in adoption of OSS.

Its common knowledge that several IT giants like, Google, IBM, Microsoft and Yahoo have been sponsoring and supporting the OSS projects. These corporations sponsor OSS projects for a variety of reasons and have slowly perfected the process of benefitting from the open source community's enormous talent. They indulge in acquiring & taking over the software developer teams that created the successful OSS projects or otherwise.

Objective

Current study is aimed at exploring the availability and accessibility of open source big data analytics software. The study is in particular, focused on the SourceForge website which is undeniably, a large repository of open source software resources. SourceForge is the second most popular source code hosting facility in the world as per Wikipedia (2018). SourceForge is next only to GitHub and followed by Bitbucket which is in third place. Of these top three hosting facilities, SourceForge is the oldest having been established in 1999 while the other are established in the year 2008.

Main idea of this study is to understand the user activity in terms of updates & downloads on the SourceForge in respect of the big data analytics so as to learn about success of OSS big data analytics applications in terms of the features of the various projects. The objective is to inform prospective users and other stakeholders of the successful open source big data analytics projects that may be of interest to them in their work. Implicitly, the results from the study will also provide pointers for developers to include the features in their projects that the end-users seem to prefer in order to enjoy better patronage. And for the well-informed, these results may provide insights into possible opportunities to develop newer

alternatives for the existing projects to satisfy unmet needs of certain segments of end-users.

DATA COLLECTION

The data was gathered using SourceForge open source software website www.sourceforge.net. While there are many sources for open source software at present, SourceForge is the first and oldest such repository of OSS. As such, it was decided to use SourceForge for data collection to get a feel for the pulse of the market in this context. As stated before, SourceForge is an open source community resource that aims to support the success of open source projects. Initial data set was collected on October 6th and 7th 2016. By using the browse option on the website, it was possible to only concentrate on big data analytics software in the business & enterprise field (access the list by choosing "business & enterprise" menu item from browse page, and then typing in "big data analytics" in search bar). A total of 36 applications were found and recorded.

As part of the revision effort, data was gathered a second time in 2018 on July 10th,11th and 12th to facilitate comparison and to understand the changes over twenty-one-month interval. Once again on the SourceForge website, a search was started with "big data analytics" key words and then the results were filtered by category of "Office/Business." Although a total of 55 projects appeared in the narrowed results page, only the top 25 projects were recorded that had experienced user downloads, as the rest of the projects were still under development and some have not even been launched and others had no downloads at all.

Metrics Recorded

For each of the data sets (initial 36 and the follow up 25) of big data analytics software in the business & enterprise field, following data was collected: software name, date registered, date last updated, intended audience, programming language, languages, OS versions available, review & rating (on scale of 5.0), total downloads, top download country, top downloaded OS, license type. Any missing data was not imputed but highlighted in the spreadsheet. Analyses are performed based on the available data under the respective fields. As it happens, download details are now available in 2018, in greater detail than they were in 2016, in terms of percentage of downloads for top downloaded country and top downloaded OS.

ANALYSIS AND FINDINGS

As stated under the data collection section, for each of the thirty-six big data analytics software found under "business & enterprise" classification, the following data was collected, software name, date registered, date last updated, intended audience, programming language, languages, OS versions available, review & rating (on scale of 5.0), total downloads, top download country, top

downloaded OS, license type. The idea in gathering these items of data about each of the application is to get good understanding of the profile of typical successful OSS applications for big data analytics.

Date of Registration and Age of Projects

A histogram of these applications based on the date of registration reveals the typical age of the projects since first deployed on SourceForge. Figure 1 below depicts such histogram of these applications on the basis of their registration date.

FIGURE 1
Histogram of date registered
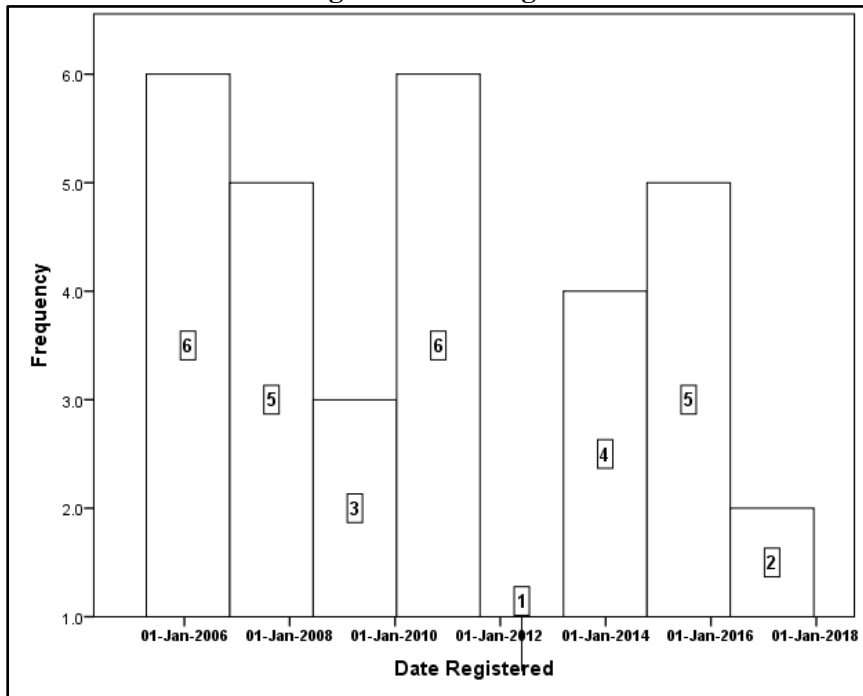
As may be observed from Figure 1, the earliest of these projects was registered before the year 2006 and the most recent one in 2016. Given that there are eight bars, each represents a time interval of about 18 months covering the total time of about 12 years. Between 2012 and 2013, there was a period of complete standstill with no new applications registered for an entire year or so. Data collected in July 2018 revealed that the new project development continued in this area into early 2018 as well. As a matter of fact, there are several projects that are at their early stages of development and are yet to be launched for end-user review and adoption, thereby bearing ample evidence to a robust growth in newer OSS options for big data analytics.

Date Updated and Development Activity

As may be noted from Figure 2, about 16 projects have not been updated after the year 2013. Some of these may have been abandoned or perfected to the point no further revisions were required. Over 30% (11 out of 36, to be precise) are actively updated as recently as middle of 2016 (that's quite impressive given that the above data was gathered in October of 2016). From the more recent July 2018 data, it was noticed that eight of top 25 projects were updated in first half of 2018 showing a more active development and improvement process. Typically, revisions are made to address user requests or to include additional functionality.

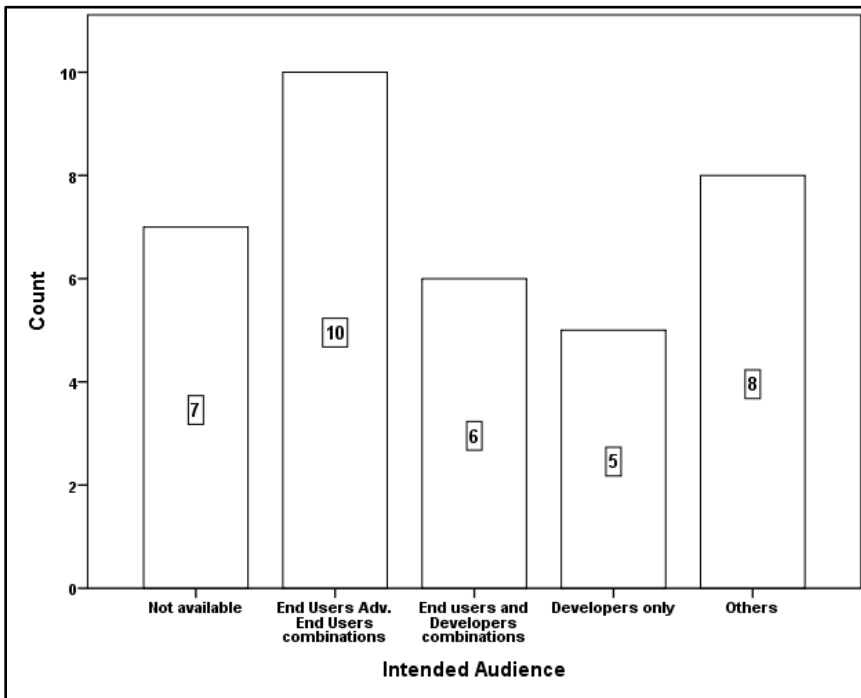**FIGURE 2**
Histogram of Last Updated
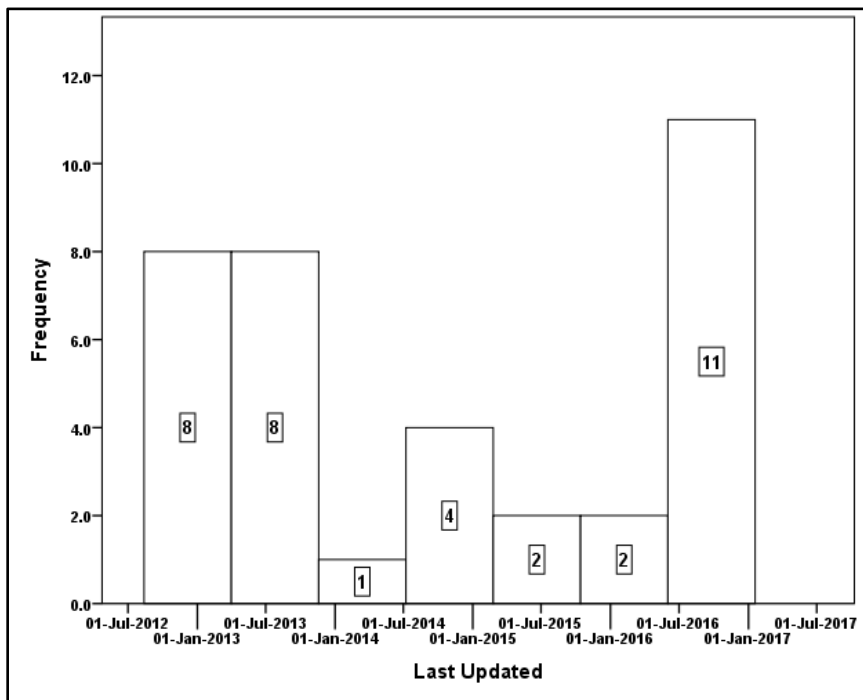
Intended Audience

Data on intended audience has been recoded to group together the audience categories as follows. All applications that didn't specify intended audience as category "0" to mean 'Not available." The applications that specified "End-users" and "Advanced End-users" together with any other except for "developers" as category 1. Applications that targeted both "Developers" and "End-users" with or without others as category 2. "Developers" only without end-user but with or without others as category 3. Then finally all categories that neither targeted "end-

users" nor "developers" but any other combinations as category 4. Figure 3 below depicts the histogram of intended audience thus generated. As may be noted, groups that contain end-users account for nearly 44% (16 out of 36) while groups that contain developers come in next with 30% (11 out of 36). "exclusively developers only," is limited to 14% (5 out of 36). The "others" category also tends to contain industry specific or special purpose applications. These trends indicate that most of these applications are targeting end-users as such, they are mostly "ready to use" applications needing little or no refinements (read as calling for little or no programming skills).

FIGURE 3
Intended Audience

However, the intended Audience chart for the top 25 projects of July 2018 revealed a slightly different trend as seen in Figure 4 below. With greater focus on attracting end-users at 72% (total of 18 out of 25), OSS creators appear to put out ready to use/deploy applications needing no further tweaks or adjustments.

Licensing

Broadly speaking, OSS license types fall on a spectrum ranging from highly restrictive (GNU AGPLv3) through moderately restrictive (GNU LGPLv3), moderately permissive (Apache License 2.0), highly permissive (MIT License)

and ending with Unlicense (public domain) (Choosealicense, 2017). Conventional wisdom and logic suggest that restrictive licenses are likely to attract developers more than end-users because of the restrictions and associated obligations. While highly permissive and unlicense types will be of great interest to end-users more than developers. All licenses are explained in greater detail at the Opensource Initiative website (Opensource, 2017b).

FIGURE 4
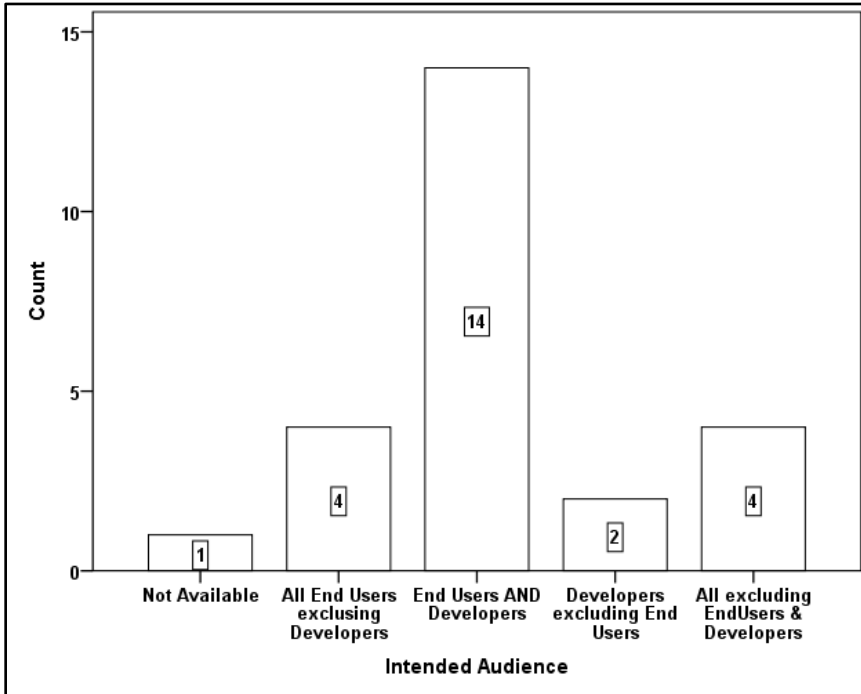**Intended Audience (Top 25 of July 2018)**

Figure 5 below shows the distribution of license types among the applications under this study. It may be noted that GNU GPLv2 is the most popular license. Overall, there is a fairly widespread selection of license types ranging on the spectrum of license from restrictive to permissive and totally unlicensed public domain as well. Nevertheless, the tilt is in the direction of strong copyleft a.k.a. protective, reciprocal and restrictive license types as opposed to permissive and totally free public domain licenses. Data from July 2018 didn't show any noteworthy departure from the pattern seen in 2016.

Downloads

Notwithstanding the fact that downloads do not always guarantee the usage of downloaded software, in an OSS community context, downloads do denote the

popularity of an application in that it generated enough interest in the users to follow through necessary steps to download a specific application either for further study and development, for use in one's work, or simply out of their idle curiosity. Accordingly, "downloads" metric is a good proxy to popularity of an application or that of a license type etc. We analyzed the data in respect of downloads in multiple perspectives and results are reported hereunder.

FIGURE 5
License Type
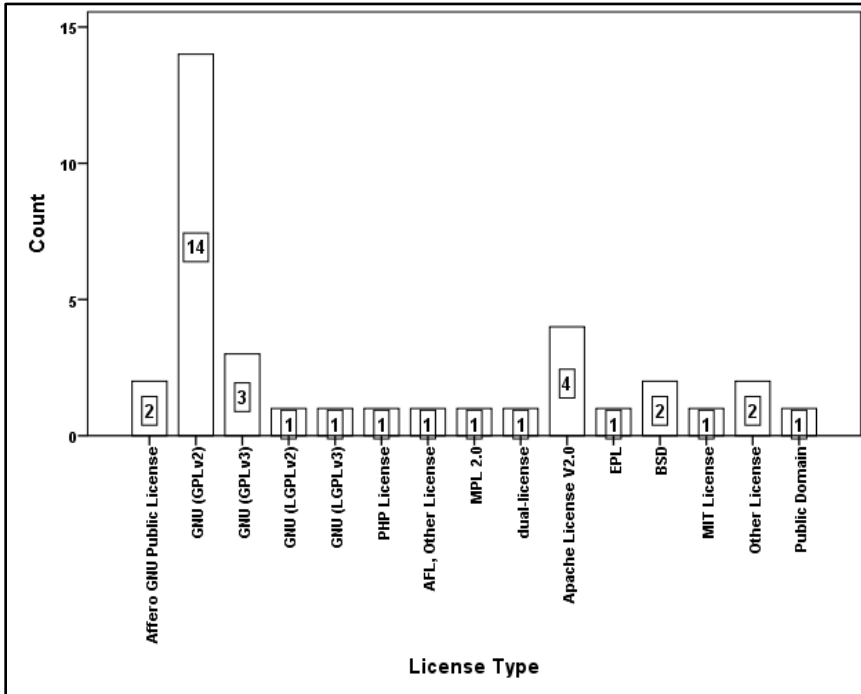
As may be noted from Table 1, of the 36 applications under study, 24 are most downloaded in just 5 countries. The twelve that do not have a most downloaded country entered are more than likely, are new applications that are yet to be downloaded as on the date of data collection. As for the others, United States is the leader with 14 most downloaded projects, with an aggregate of 6, 241,460 downloads, followed by India with 5, Spain with 3 and Brazil and Germany accounting for one project each. While not much can be read into the spread of countries in the Table 1; one can definitely conclude that the OSS initiative is spreading far and wide.

Figure 6 below depicts the downloads captured by the license type and it is interesting to note that GNU GPLv2, despite falling under the restrictive license type category enjoyed the most following in terms of mean downloads of 419,257

per application of the license type. Among the permissive license types Berkeley Software Distribution (BSD) appears to be most popular with 112,373 mean downloads per application.

TABLE 1
Count of projects most downloaded by country

| Country	Count of projects most downloaded	Sum of Downloads
Brazil	1	2,498
Germany	1	1,780
Spain	3	23,661
India	5	14,968
N/A	12	0
United States	14	6,241,460
Grand Total	36	6,284,367

FIGURE 6
Downloads by License Type
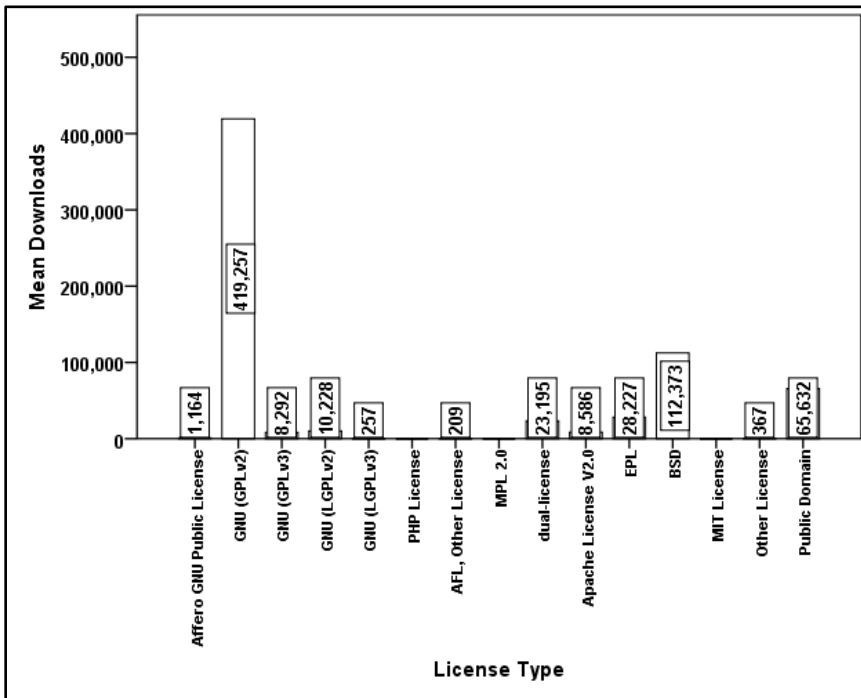
When seen from the operating system (OS) point of view, Windows OS emerges the leader with 16 out of 36 applications under study having been most downloaded

on the Windows OS. Amongst the 2016 data, there is no clear second to Windows in that, none of the other OS score significant downloads to merit attention. Accordingly, they are combined into 'other' for reporting purposes. In about 12 cases, the metric is not reported as they have not had any downloads as yet as on the date of data collection in October 2016. From the 2018 data set too, Windows is seen to be most dominant OS downloaded as may be seen from figure 7 below. Of the top 25 projects, 4 projects had not specified their OS.

FIGURE 7
**Top Downloaded OS as of July 2018**
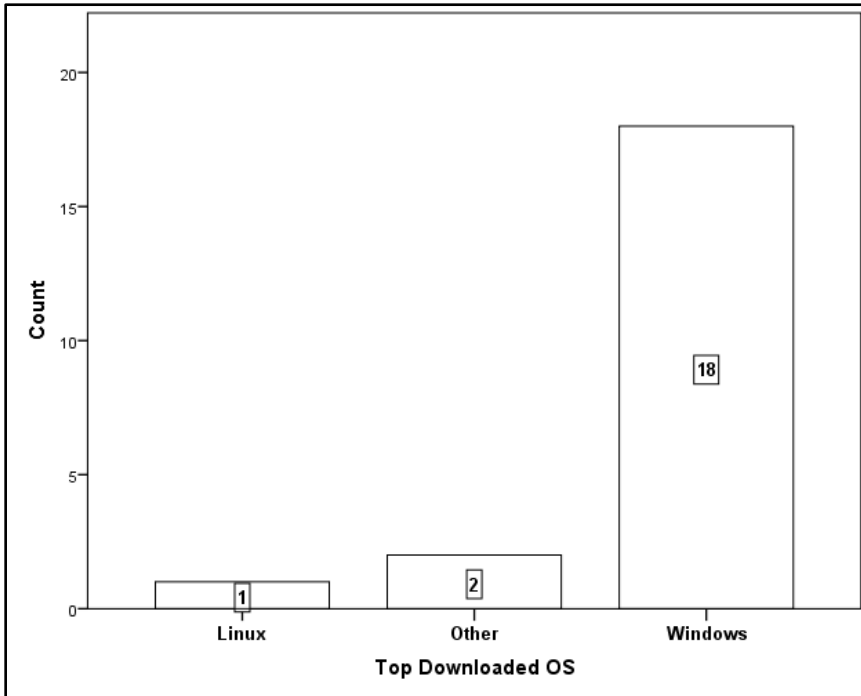
In so far as the preferred programming language is concerned, Java, Javascript, JSP and other Java family languages top the list with nearly 50% of the projects offering the applications in these languages (or in these language versions). Next most popular language family appears to be C, C++ and C# and some combinations thereof. They account for nearly 15% of the applications under study followed by Python, PHP, Ruby and PL/SQL in that order.

Figure 8 above depicts life time downloads of top seven applications until the date of data collection in 2016. As may be observed, one specific application "Pentaho" has had phenomenal success. However, Figure 9 below depicts the current top seven as of July 2018 which demonstrates a significant change in the ranking by

total downloads. Two new projects made it into top three, viz., Jaspersoft Studio and Zenoss Enterprise Hybrid IT monitoring.

FIGURE 8
**Top Seven downloaded applications**
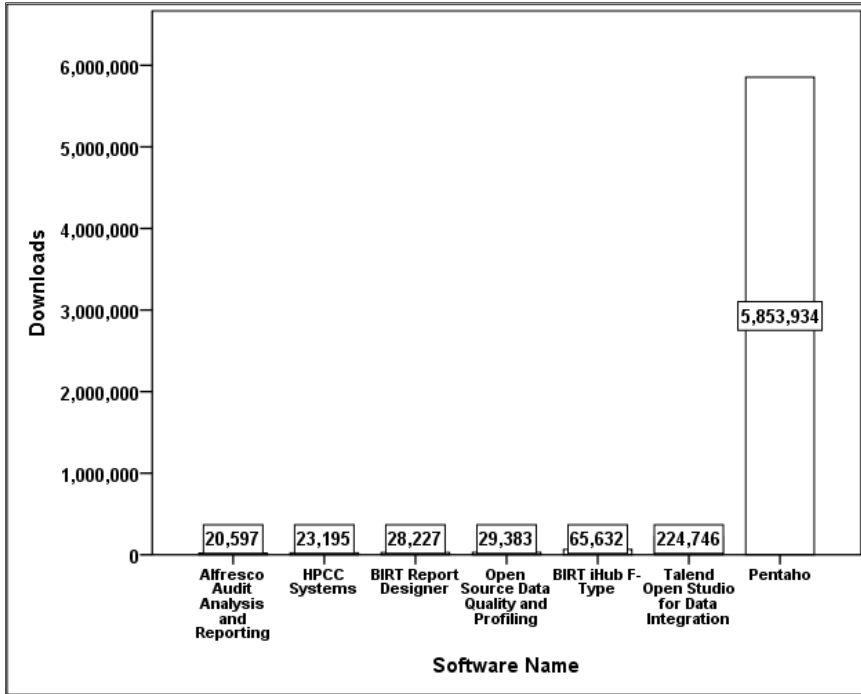
While many factors contribute to the success of an OSS project, Figure 10 above does indicate the popularity of certain family of programming languages over others as being the preferred languages for OSS projects by developers as well as end-users. Java and JavaScript being the most preferred ones. Python and Zope server environment combination appear to be the new entrant making a very strong entry into this OSS arena.

The Top Seven Applications from October 2016

Pentaho (GNU GPLv2 License) is a product of Hitachi Group Company. Claims to provide, easy to use Business Intelligence (BI) for all by combining data integration and business analytics to facilitate access, visualize and explore all data that potentially impacts business results. Pentaho may be accessed in entirety or module by module either in the cloud or on-the-go (mobile) as highlighted on the company's account page in SourceForge website (Pentaho, 2017).

Talend Open Studio for Data Integration and Talend Open Studio for Big Data
(BSD License) are products of Talend Corporation. These products facilitate
Extraction, Transofrmation and Loading (ETL) from any source of data to any
Analytics or other such tools. Makes it easier to connect to databases, flat files,
cloud-based applications and data to simplify ETL processes (Talend, 2017).

FIGURE 9
**Top seven downloaded software as of July 2018**
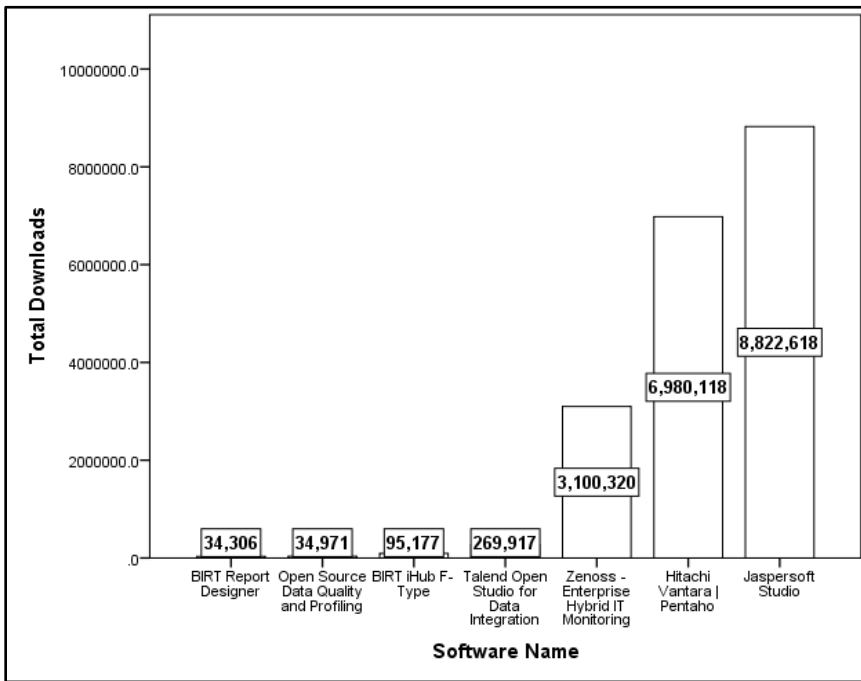
BIRT iHub F-Type (Public Domain License) is a Business Intelligence and
Reporting Tool (BIRT) that claims to provide data support and connectivity to all
types of Relational databases and thereby support data visualization and reporting-
as highlighted on the company's account page in SourceForge website (BIRT
iHUB, 2017).

Open Source Data Quality and Profiling (Apache License V2.0) a.k.a. Aggregate
Profiler: As described on the SourceForge page of the project, "*This project is
dedicated to open source data quality and data preparation solutions. Data
Quality includes profiling, filtering, governance, similarity check, data enrichment
alteration, real time alerting, basket analysis, bubble chart Warehouse validation,
single customer view etc. defined by Strategy.*" (OSDQP, 2017).

BIRT Report Designer (EPL License)- BIRT is a top-level software project within
The Eclipse Foundation which is an independent not-for-profit consortium of

software industry vendors and an open source community. The project is supported by Actuate, IBM, and Innovent Solutions. Main focus of the application is to create, "data visualization and reports that can be easily embedded into rich clients and web applications." (BIRT Report Designer, 2017).

FIGURE 10
Total downloads of projects by programming language as of July 2018
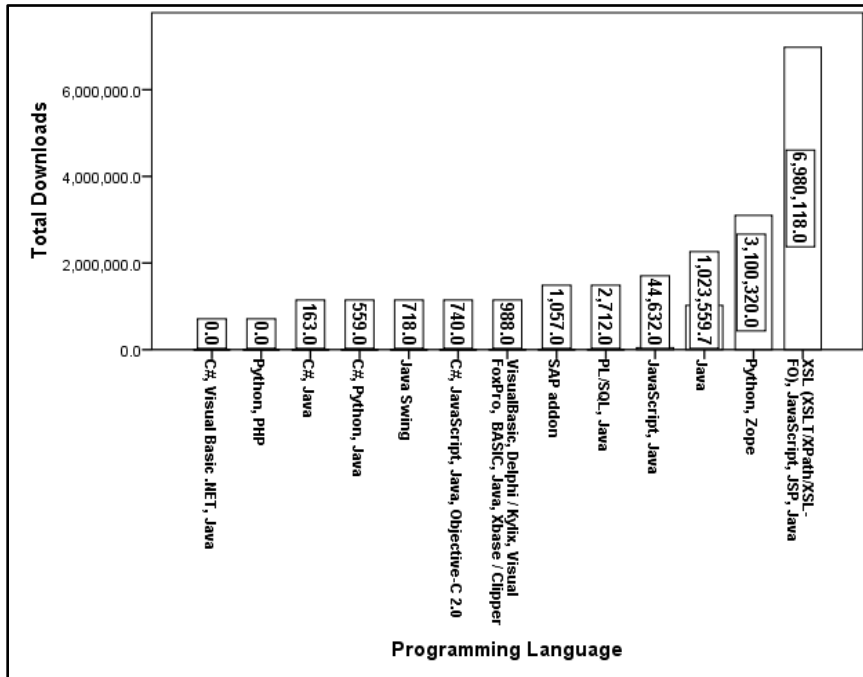
HPCC Systems (Dual License) from LexisNexis Risk Solutions offers an open source, proven, data-intensive supercomputing platform designed for the enterprise to process & solve Big Data analytical problems. The HPCC Systems Data Refinery engine (Thor) helps clean, link, transform and analyze Big Data. The HPCC Systems Data Delivery engine (Roxie) provides highly concurrent and low latency real time query capability (HPCC Systems, 2017).

Alfresco Audit Analysis and Reporting (AAAR) (GNU GPLv3 License): Provides a solution to extract, store and query audit data. AAAR Analytics is a set of powerful tools to analyze data in an interactive and customizable way with a user console of dashboards, reports and free analysis (AAAR, 2017).

Surprisingly, no two of the top seven applications are available under the same version of OS license except for the first and seventh being rather close under GNU GPLv2 and GNU GPLv3 respectively. Each of the other five represents a distinct license type.

Two New Projects in Top Three - July 2018

Jaspersoft Studio (Eclipse Public License): Essentially, Jaspersoft® Studio is an editing software for JasperReports® that facilitates multiple format report generation by accessing a variety of data sources. As described on the SourceForge page, with Jaspersoft, "*You can build documents of any complexity from your data. Print-ready PDFs to interactive dynamic HTML with navigation inside or outside the report. High quality PowerPoint, RTF, Word, spreadsheet documents or raw CSV, JSON, or XML. Different types of data sources are accessible, big data, CSV, Hibernate, Jaspersoft Domain, JavaBeans, JDBC, JSON, NoSQL, XML, or custom data source. Available as an Eclipse plug-in or a standalone application, it comes in two editions: Community and Professional. The Professional edition includes additional features, maps, advanced HTML5 charts and professional support.*" (Jasper Studio, 2018).

It's interesting to note that Jasper studio is a stand-alone platform-independent editing tool that allows end-users to continue using their existing IT setup while simultaneously allowing them to create and use powerful data visualization reports to help keep their organization ahead with timely decision support. In other words, the jasper studio requires little to no interruption in the existing IT setup and serves more as a free add-on.

Zenoss –Enterprise Hybrid IT Monitoring (GPLv2): It's a hybrid IT monitoring and analytics software. As described on its SourceForge page, "Zenoss *is the global leader in hybrid IT monitoring and analytics software, providing complete visibility for cloud, virtual and physical IT environments for more than 40,000 global organizations. Zenoss customers gain IT performance and risk insights into their unique IT ecosystems through real-time analytics that adapt to the ever-evolving data center and cloud, enabling them to eliminate disruptions and accelerate business.*" Further, "*Once Zenoss Core discovers the IT infrastructure, it automatically begins monitoring the performance of each device and provides event and fault management capabilities.*" (Zenoss-Enterprise, 2018).

Given the utility of hybrid IT monitoring, the phenomenal success of Zenoss-Enterprise over the past two years, 2016 to 2018 is understandable. As may be noted, this application is providing real-time analytics support for the IT infrastructure of user organizations, attesting to its immediate utility to end-users.

OBSERVATIONS AND CONCLUSIONS

Based on the charts of various features of the big data analytics OSS software and analyses of the relative data the following important observations may be made along with corresponding conclusions drawn.

- Six of the top seven big data OSS on SourceForge in 2016 are from the USA.
- Five of the top seven big data OSS on SourceForge are on Windows OS.
- Only two of top seven big data OSS on SourceForge offer GNU (GPLv2 or GPLv3) license all others offer different license each. Which means users of different OS will find some projects that are compatible to their operational requirements.
- Five of the top seven big data OSS on SourceForge projects have corporate sponsors. This is quite significant in that corporate sponsorship appears to promote the OSS creation despite the fact that the sponsors only derive some indirect benefit from the overall final product.
- Sponsorship seems positively impact popularity of software which goes to reiterate the end-users' preference for brand name recognition leading to higher reliability for continuity, higher trust and support for future.
- Most applications appear to target "advanced end users," because after all, SourceForge is a OSS platform that is made up of developers as active members and they would like for their creations to be adopted and further developed by the other developers either by combining it into their own creations or by making it that much more robust and wholesome by additions, alternations to provide a viable and low or no-cost alternative to other proprietary options on offer in the market.
- Java and Java script appear to be most preferred programming language which is not surprising given that Sun microsystems enjoys a lot of market share and recognition for its products and enhancements.
- Python programming and associated Zope server environment appears to gain ground in the last couple of years, since 2016 in this area of OSS and it's not surprising given the host of user-friendly features of Python's high-level general purpose programming environment.
- Going by the sample from SourceForge, albeit a convenient sample, there is a very active OSS community in developing big data analytics not limited to any one particular region of the world or a programming platform.
- The activeness of the community is further evidenced by the fact that, from Oct 2016 to July 2018 – the OSS big data projects went up from 36 to 55 while one of the initial 36 was abandoned by the creator. The term used for such abandoned OSS in the forum's ling being "abadonware." This means, nearly twenty more projects made it into the list of big data analytics related projects over the past two years.
- In July 2018, of the 55 projects only the top 25 had user download data indicating several others were either in development stages or have not yet been reviewed and downloaded by users.

Overall, for cash strapped business organizations wanting to implement big data analytics initiatives with least possible cost aid their managerial decision making, there appear to be ample options on offer in OSS forums. Despite some commonalities, the top projects under this category of OSS are quite distinct from

each other in their features and functionality, which denotes that they are catering to different and distinct segments of customers. Given that these offerings are very diverse in their features in terms of functionality, licensing, OS availability, programming language, user manuals in multiple languages etc., all it takes for a business is some serious investment of time and effort to harness the power of big data analytics to help the organization find a project that meets its requirements to help it stay ahead in the market place.

REFERENCES

AAAR, 2017. Alfresco Audit Analysis and Reporting, overview (https://sourceforge.net/projects/aaar/?source=directory; accessed on March 1, 2017.)

BIRT iHub, 2017. Overview (https://sourceforge.net/projects/birtihubftype/; accessed on December 1, 2017).

BIRT Report Designer, 2017. Overview (https://sourceforge.net/projects/opensourcebirtreportdesigner/?source=directory-featured; accessed on December 1, 2017).

Choosealicense, 2017. Licenses (https://choosealicense.com/licenses/; accesses on December 1, 2017).

Gallego, M. D., Bueno, S., Racero, F. J., and Noyes, J., 2015. Open source software: The effects of training on acceptance. Computers in Human Behavior. Elsevier publishers, vol. 49, pp 390-99.

Harvard Business Review, 2015. How the Digital Windfarm Will Make Wind Power 20% More Efficient? Sponsor Content Insight from GE, July 06, 2015.

HPCC Systems, 2017. Overview (https://sourceforge.net/projects/hpccsystems/?source=directory-featured; accessed on March 1, 2017).

Lerner, J., and Tirole, J., 2005. The scope of open source licensing. Journal of Law Economics & Organization vol. 21 (1) 20.

Jaspersoft Studio, 2018. Overview (https://sourceforge.net/projects/jasperstudio/?source=directory accessed on July 28, 2018).

Marsan, J., and Pare, G., 2013. Antecedents of open source software adoption in health care organizations: A qualitative survey of experts in Canada. International Journal of Medical Informatics, Elsevier publishers, vol. 82. pp 731-741.

Opensource, 2017a. What is open source? (https://opensource.com/resources/what-open-source; accessed on March 1, 2017).

Opensource, 2017b. Open Source Licenses by Category (https://opensource.org/licenses/category; accessed on March 1, 2017).

OSDQP, 2017. Open Source Data Quality and Profiling, overview (https://sourceforge.net/projects/dataquality/?source=directory ; accessed on March 1, 2017).

Pentaho, 2017. Hitachi Vantara | Pentaho, overview (https://sourceforge.net/projects/pentaho/?source=directory; accessed on March 1, 2017).

Rafiq, M. 2009. LIS community's perceptions towards open source software adoption in libraries. The International Information & Library Review, Elsevier publishers, vol. 41, pp137-145.

Roumani, Y., Nwankpa, J. K., and Roumani, Y.F. 2017. Adopters' trust in enterprise open source vendors: An empirical examination. The Journal of Systems and Software, Elsevier publishers, vol. 125, pp 256-270.

Sarrab, M., and Hussain Rehman, O.M. 2014. Empirical study of open source software selection for adoption, based on software quality characteristics. Advances in Engineering Software, Elsevier publishers, vol. 69 pp 1–11.

SAS, 2017. Why is big data analytics important? By SAS research team. (https://www.sas.com/en_us/insights/analytics/big-data-analytics.html.; accessed on December 1, 2017).

SourceForge, 2018. About SourceForge. (https://sourceforge.net/about; accessed on July 12, 2018).

Talend, 2017. Overview (http://www.talend.com/download/talend-open-studio/?qt-product_tos_download_new=1&utm_medium=communityext&utm_source=sourceforge&utm_campaign=tosdi; accessed on March 1, 2017).

Wikipedia, 2017. SourceForge (https://en.wikipedia.org/wiki/SourceForge; accessed on March 1, 2017).

Wikipedia, 2018. Comparison of source code hosting facilities. (https://en.wikipedia.org/wiki/Comparison_of_source_code_hosting_facilities; accessed on July 10, 2018).

Zenoss-Enterprise, 2018. Overview (https://sourceforge.net/projects/zenoss/?source=directory; accessed on July 29, 2018).

About the Author

Balaji Janamanchi is, currently, Associate Professor of Management, Division of International Business and Technology Studies of the A.R. Sanchez Jr. School of Business Administration at Texas A&M International University, Laredo, Texas where he is employed since fall 2006. He received Ph.D. and MS degrees from the Rawls College of Business Administration, Texas Tech University majoring in Production and Operations Management (POM) and Management Information Systems (MIS), respectively. He worked as an MIS instructor in the Rawls COBA, TTU (2001–2004). His research interests include supply chain management, project management, online teaching, and e-commerce related issues.