

A REVIEW ON APPLICATION OF OPEN-SOURCE TOOLS

¹Weirong Han, ²Jiarui Chang, ³Chenshun Hong

¹12th Grade, ²11th Grade, ³9th Grade

¹Dalian Yuming Senior High School, ²Dalian No. 24 High School, ³Jiahui Middle School
China

Abstract: Taking into consideration a number of benefits of open-source software packages such as reduction in dependency on closed-source vendors, greater access to tools, access to source code, ability to customize, and increased security over closed-source vendors. The present work highlights various open-source software packages available in the market. Some open-source software packages discussed in the present work include open-source operating systems, distributed databases for handling massive amounts of data, cloud deployment software, virtualization tools, web servers for support of websites, high-quality typesetting systems for technical and scientific documentation, BigData analysis tools, document management systems, and tools for creating quick demos, screencast software demonstrations, and active presentations, among others.

Index Terms -Open Source, RethinkDB, Cassandra, CloudStack, LaTeX, Lighttpd

1. Introduction: Open Source

The term "open source" refers not only to the lack of availability of some source but also to other distribution terms related to open-source software packages that must be taken into consideration [1]:

- Free redistribution
- Source code: The source code must be included along with the program, and there must be some permission for the distribution of the source code as well as the program in the compiled form.
- Derived works: Modifications must be allowed in the license, and redistribution must be allowed under the same terms and conditions as the license for the original software.
- Integrity of the author's source code: In some conditions, there can be some restriction on the distribution of the source code, such as the license allowing the distribution of "patch files" along with the source code for the purpose of modifying the program at the build time.
- No discrimination against persons or groups
- No discrimination against fields of endeavor
- Distribution of license: The right attached to the main program must be applied to all to whom the program is redistributed, and there should be no additional requirement for the license.
- License must not be specific to a particular product.
- License must not restrict other software packages.
- License must be technology-neutral.

2. Open-Source Distributed Database

Definition:

A database management system (DBMS) is a software application that intermingles with applications, users, and the database itself to capture and analyze given data. The DBMS allows users to define, create, querying, update, and administrate databases [2]. There are many open-source DBMSs, including

- RethinkDB
- Cassandra
- MySQL
- PostgreSQL
- SQLite
- Berkeley DB

2.1 Product Name: RethinkDB

Product Description:

RethinkDB is an open-source/NoSQL/scalable/JavaScript Object Notation/distributed document-oriented database designed specifically for real-time web applications. It can reduce the amount of effort and time required to build scalable real-time applications [3]. It has a special database access model that works on web-based applications such that data can be sent directly to the client on a real-time basis.

Product License

The RethinkDB client driver is licensed under Apache License v2.0. The RethinkDB server is licensed under GNU Affero General Public License v3.0.

Some cases that can benefit from the real-time push architecture include the following:

Product Applications/Uses

- Various connected devices
- Marketplaces having requirements for real-time data
- Collaborative mobile and web apps
- Apps with analytic privileges
- Multiplayer games

2.2 Product Name: Cassandra

Product License: Apache License 2.0

Product Description:

Cassandra is an open-source distributed database management system used when there is a strong need for scalability and availability. It supports a number of properties such as atomicity, consistency, isolation, and durability (ACID). It provides for the replication of data across multiple data centers with no single point of failure [4]. It is most suitable for applications that cannot afford to lose data even when the whole data center fails. It provides a complete elastic system on which the throughput increases linearly as new machines are added without any interruption or downtime.

Product Applications/Uses

- It is fault tolerant such that data can be automatically replicated to multiple nodes.
- With no downtime, the failed node can be replicated easily.
- It can handle peta bytes (PBs) of data with thousands of nodes under its control.
- It provides complete control for synchronous or asynchronous replication.

3. Definition of Cloud Computing:

Cloud computing is a model for enabling ubiquitous, convenient, and on-demand network access to a shared pool of configurable computing resources such as networks, servers, storage systems, applications, and services that can be rapidly provisioned and released with minimal management effort or service provider interaction [5]. This cloud model is composed of five essential characteristics, three service models, and four deployment models.

Essential Characteristics:

- On-demand self-service
- Broad network access
- Resource pooling
- Rapid elasticity
- Measured service

Service Models

- Software as a service (SaaS)
- Platform as a service (PaaS)
- Infrastructure as a service (IaaS)

Deployment Models:

- Private cloud
- Community cloud

- Public cloud
- Hybrid cloud

3.1 Open-Source Tools for Cloud Computing

Product Name: Eucalyptus

Description: Eucalyptus is a Linux-based open-source software architecture that implements efficiency-enhancing private and hybrid clouds within the enterprise's existing IT infrastructure [6]. Eucalyptus refers to "Elastic Utility Computing Architecture for Linking Your Programs to Useful Systems. Written in Java, C language eucalyptus helps the pooling of computation power, network resources, and storage.

Product License: It is Licensed under GPLv3 with proprietary relicensing.

3.2 Product Name: CloudStack

Description: An open-source Apache system is a software package designed for deploying and managing large networks of virtual machines and the highly scalable infrastructure as a service (IaaS) as a highly available cloud computing platform [7]. CloudStack is used by a number of service providers to offer public cloud services and by many companies to provide on-premise (private) cloud offerings or as part of a hybrid cloud solution.

Product License: It is licensed under Apache License Version 2.0

CloudStack supports a number of hypervisors such as

- VMware
- KVM
- Citrix XenServer
- Xen Cloud Platform (XCP)
- Oracle VM server
- Microsoft Hyper-V

3.3 Product Name: OpenNebula

Product Description: It is an open-source system for providing cloud solutions [8]. It is written in Ruby, C++, and Shell and used for the purpose of authentication. OpenNebula uses the lightweight directory access protocol (LDAP).

Product License: Apache 2 license

Some design principles OpenNebula include

- Openness related to codes, architectures, interfaces
- Flexibility
- Interoperability and portability
- Stability
- Scalability
- SysAdmin-centrism
- Lightness
- Simplicity

Because of these design principles, there are a number of benefits:

- Faster responses to infrastructure needs for services
- Centralized management
- Higher utilization of existing resources
- Operational saving
- Lower infrastructure expenses

3.4 Product Name: OpenStack

Description: It is a cloud-based operating system that helps to control pools of networking resources, storage, and computation at various data centers [9]. All such management is done using the OpenStack dashboard, which provides the administrator with complete control and helps in providing resources to various users via the web interface. It is written in Python.

Product License: The license agreement is under Apache License 2.0.

Benefits/Uses

- It provides the agility and flexibility necessary for basic business needs.
- There are remarkable differences in productivity and staffing.
- It provides portability and interoperability in bringing a number of cloud service providers to a single platform.
- Consumers can compare, search, and select services based on the specific need.

4. Definition of Virtualization

This refers to an environment that allows for the running of a new operating system on a preinstalled or existing operating system. Such an environment is called a virtual machine, which is created using virtualization software [10]. There are a number of such open-source virtualization software packages available that can help obtain access to hardware components and other features of the host machine.

Benefits of Virtualization

- Testing and data recovery
- Infrastructure consolidation
- Running multiple operating systems simultaneously
- Easier software installation

4.1 Product Name: VirtualBox

Description: This is a general-purpose cross-platform virtualizer for x86 hardware and targets servers, desktops, and embedded systems [11]. It can run anywhere from small embedded systems or desktop machines to data centers and cloud environments. It provides a number of features:

Portability means it can run on 32-bit and 64-bit operating systems.

- No hardware support is required.
- Multigenerational branched snapshots of the state of the virtual machine are provided.
- It enables remote machine displays.
- It allows VM groups for collective control of virtual machines.
- It has a modular design with properly defined internal programming interfaces.

Other open source virtualization tools are shown in TABLE 1.

TABLE 1. Various open source virtualization tools

Name	Virtualization Type	Hypervisor Type
Xen hypervisor	Hardware Assisted Virtualization, Paravirtualization	Bare Metal (Type 1)
Kernel based virtual machines(KVM)	Full Virtualization, Hardware Assisted Virtualization, Paravirtualization	Hosted (Type 2)
Nuxis	Full Virtualization, Hardware Assisted Virtualization, Paravirtualization	Bare Metal (Type 1)
Proxmox VE	Full Virtualization, Operating system virtualization	Bare Metal (Type 1)
Linux-Vserver	Operating system Virtualization	Hosted(Type2)

5. Definition of Web Servers

A web server is a computer system that helps in processing requests via HTTP (hypertext transfer protocol) [12]. Web servers provide a number of benefits:

- A local website can behave like a live one.
- The use of server-side scripting languages such as ColdFusion and PHP can be done.
- It helps in standardizing the coding, creating virtual directories, and mapping to physical directories, among others.
- It facilitates the creation of FTP sites.
- It has customized error pages.
- Before committing to a production environment, the developer perform various tests.

There are a number of open-source web servers available, including the following:

5.1 Product Name: Apache HTTP server

Description: This is a powerful and flexible open-source HTTP server that implements the latest protocols such as HTTP/1.1(RFC2616). It provides the full source code and has an unrestricted license.

Product License: It comes under Apache License 2.0.

5.2 Product Name: Nginx

Description: Nginx is a high-performance open-source HTTP server, a reverse proxy, and an IMAP/POP3 proxy server. It is best known for its stability, simple configuration, high performance, and lower resource consumption.

Product License: It is licensed under 2-clause BSD.

Nginx can provide a number of solutions for various purposes:

- API gateway
- Load balancing
- Streaming media
- Web server
- Accelerates web and mobile performance
- Application defense

5.3 Product Name: Lighttpd

Description: This is an open-source package designed for a high-performance environment with a small memory footprint [13].

Product License: It is licensed under 3-clause BSD.

It provides a number of features such as

- Effective CPU load management
- Fast CGI (common gateway interface)
- SCGI (simple common gateway interface)
- URL rewriting and output compression

Its performance can be measured using the various parameters such as

- **Memory Usage**

Lighttpd and Nginx are best in terms of memory usage for concurrent connections. This is because the Apache HTTP server creates a new thread for every connection. Therefore, as the number of connections increases, the corresponding memory usage also increases. Nginx stays more static in terms of memory consumption in comparison to Lighttpd.

- **Requests Per Second**

Nginx is a web server that can handle the maximum number of requests as compared to Lighttpd and Apache. It can handle about double the amount of requests than Lighttpd and 4 times as many as Apache.

6. Open-Source Tools for Document Management System

There are a number of document management systems available for providing services to end users.

6.1 Product Name: LaTeX

Description: LaTeX is designed specifically for high-quality typesetting. Medium to large scientific or technical documents can be written in a proper manner with the help of markup tagging conventions [14]. It was first developed in 1985 by Leslie Lamport.

Product License: It is available under the LaTeX Project Public License

LaTeX provides a number of features:

- Large documents with tables, figures, cross-references, and sections can be effectively controlled for.
- It allows for multi-lingual typesetting.
- It enables special typesetting for formulas of complex mathematics.
- It allows for typesetting for mathematics with AMS-LaTeX.
- It uses metafont fonts or postscript.
- It enables the insertion of spot colors, processes, and artworks.
- Indexes and bibliographies can be generated automatically.

6.2 Product Name: OpenKM

Description: OpenKM is a free document management system that gives a web interface for the management of arbitrary files [15]. It helps in the storage, production, distribution and management of electronic documents.

Product License: It is available under the GNU GPL license.

OpenKM can provide various solutions for the smooth working of the organization:

- Document management: It simplifies work, manages content easily, and thus increases efficiency.
- Record management: It corresponds to the creation, receipt, maintenance, use, and disposition of records
- It allows for the construction of own applications.
- Modules: It is a flexible module system that provides features of multi-tenancy and helps in enabling multiple independent tenants to be hosted on a single instance.
- OpenKM has barcode privileges for reading and identifying modules.

Other open-source document management systems include the following:

- SeedDMS
- Alfresco
- Kimios

7. Definition of Operating System

An operating system is an interface between the hardware and the user. It is the core of the computer. Although the market is dominated by the Windows operating system, there are certain systems such as open-source-based operating systems, including Ubuntu, the representative system.

There are many open-source operating systems available for providing services to end users:

7.1 Product Name: Ubuntu

Description: It is based upon the Linux operating system, and its source code is available for free [16]. It is the most widely used open-source operating system.

Product License: It is distributed under the GNU and GPL licenses. Other open-source operating systems available include

- FreeBSD
- OpenSolaris
- ReactOS
- Haiku OS

8. Open-Source Tools for Screen Casting

Product Name: CamStudio

Description: CamStudio is an open-source system that acts as a screen-casting system that allows for all screen and audio activities to be recorded [17]. This software is written in the C++ language. It produces outputs in AVI format.

Product License: It is available under the GNU GPL license.

Other open-source screen casting software packages include the following:

- Dahu
- Captura
- Capture Box
- Freeseer
- Kazam
- RecordMyDesktop

9. Conclusions

From the above discussion, it can be concluded that there are a vast number of open-source products available to serve various user requirements, including

- Operating systems
- Distributed databases for handling large amounts of data
- Cloud deployment software
- Virtualization tools
- Web servers for website support
- High-quality typesetting systems for technical and scientific documentation
- Big data analysis tools
- Document management systems
- Tools for creating quick demos, screencasts software demonstrations, and active presentations

References

- [1] "The Open Source Definition," 22 June 2016. [Online]. Available: <https://opensource.org/osd-annotated>.
- [2] "Wikipedia," 22 June 2016. [Online]. Available: <https://en.wikipedia.org/wiki/Database>.
- [3] "The open-source database for the realtime web," Rethinkdb, [Online]. Available: <https://www.rethinkdb.com/>. [Accessed 22 June 2016].
- [4] "Cassandra," Apache, [Online]. Available: <http://cassandra.apache.org/>. [Accessed 22 June 2016].
- [5] "The NIST Definition of Cloud Computing," [Online]. Available: <http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf>. [Accessed 22 June 2016].
- [6] "Wikipedia," [Online]. Available: [https://en.wikipedia.org/wiki/Eucalyptus_\(software\)](https://en.wikipedia.org/wiki/Eucalyptus_(software)). [Accessed 22 June 2016].
- [7] "Cloudstack," [Online]. Available: <https://cloudstack.apache.org/>. [Accessed 22 June 2016].
- [8] "Opennebula," [Online]. Available: <http://opennebula.org/>. [Accessed 22 June 2016].
- [9] "Considering OpenStack? Learn about the Path to Cloud," [Online]. Available: <https://www.openstack.org/>. [Accessed 22 June 2016].
- [10] "Basic Concepts and Challenges against Virtualization," *International Journal of Computer Applications*, vol. 102, pp. 26-30, 2014.
- [11] "Virtualbox," Oracle, [Online]. Available: <https://www.virtualbox.org/>. [Accessed 22 June 2016].
- [12] "Web server," [Online]. Available: <http://whatis.techtarget.com/definition/Web-server>. [Accessed 22 June 2016].

- [13] "Web server performance comparison," DreamHost, [Online]. Available: <https://help.dreamhost.com/hc/en-us/articles/215945987-Web-server-performance-comparison>. [Accessed 22 June 2016].
- [14] "LaTeX – A document preparation system," Apache, [Online]. Available: <https://www.latex-project.org/>. [Accessed 22 June 2016].
- [15] "Document Management and Record Management System," [Online]. Available: <http://www.openkm.com/>. [Accessed 22 June 2016].
- [16] "Ubuntu," [Online]. Available: <http://www.ubuntu.com/>. [Accessed 22 June 2016].
- [17] "CamStudio," [Online]. Available: <http://camstudio.org/>. [Accessed 22 June 2016].

