فصلنامه سیستم‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)

Winter, 2021, 2 (4), 1-11
DOR: 20.1001.1.27832570.1400.2.4.2.9

Received: 22July 2021
Accepted: 06 Sep 2021

A Comparative Study of Open-Source Software for Deployment and Management of Cloud Computing Utilizing a Big Data Processing Quality Model

Mahdi Jafari[1], Amir Kalbasi[2*]

1. MSc. Student, Department of Computer Engineering, Amir Kabir University of Technology, Tehran, Iran.
mahdijafary@aut.ac.ir
2. Assistant Professor, Department of Computer Engineering, Amir Kabir University of Technology, Tehran, Iran.
(Corresponding Author) kalbasi@aut.ac.ir

Abstract

Introduction: The volume of data produced by human society is growing rapidly. Data is being produced in many different industries such as manufacturing, transportation, healthcare, and social networks. Due to the volume of data being produced, data storage and processing are among the most important issues when dealing with big data. The main challenges when dealing with big data are data storage and management, data processing and analytics, and resource management to provide the infrastructure needed to support the first two mentioned challenges. Cloud computing, due to its features and architecture, is a promising infrastructure to store and process big data. Different cloud computing deployment models exist, namely, public cloud, private cloud, community cloud, and hybrid cloud. To store and process big data in a cloud environment, individuals and organizations may be more inclined to deploy and manage private clouds to gain greater control and access to resources and their data. Numerous open-source software has been developed for the deployment and management of private clouds. Evaluating and choosing among them is a challenging task, especially for those who are new to these large-scale software systems. Furthermore, due to the continuous delivery of new releases with major changes or new features and modules for each of the cloud infrastructure management software, choosing among them could be a challenge even for an experienced user.

Method: In this paper, first of all, we provide the Quality Model for Cloud Infrastructure (QMCI) for evaluation of cloud infrastructure management software. QMCI focuses on quality factors that are important when processing big data. The top-level factors of this model are 1- Functionality 2- Usability 3- Reliability 4- Supportability 5- Performance. The top-level factors are then divided into sub-factors to further refine the quality model. Metrics can be considered for the sub-factors to evaluate a cloud infrastructure management software.

Discussion: Based on QMCI, multiple-criteria decision-making can be utilized to choose between cloud infrastructure management software that best suits a given set of criteria. In the remaining of this paper, three of the most popular open-source cloud infrastructure management software, namely, Eucalyptus, OpenStack, and Apache Cloud Stack are evaluated based on QMCI to compare their capabilities, weaknesses, and strengths from big data processing perspective. Previous literatures that considered the selected three cloud infrastructure management software were studies and utilized to perform the comparative study.

Keywords: Big Data Processing, Big Data Storage, Cloud Computing, Cloud Management, Quality Model

فصلنامه سیستم‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)
انجمن علمی تجارت الکترونیکی ایران
واحد زنجان

مطالعه تطبیقی نرم‌افزارهای منبع‌باز استقرار و مدیریت رایانش ابری با استفاده از مدل کیفیت از دیدگاه پردازش کلان‌داده

تاریخ دریافت: ۱۴۰۰/۰۴/۳۱
تاریخ پذیرش: ۱۴۰۰/۰۶/۱۵
سال دوم، زمستان ۱۴۰۰
شمارهٔ چهارم، صص: ۱ - ۱۱

مهدی جعفری[۱]، امیر کلباسی[۲*]

۱. دانشجوی کارشناسی‌ارشد، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران، ایران. mahdijafary@aut.ac.ir
۲. استادیار، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر، تهران، ایران. (نویسندهٔ مسئول) kalbasi@aut.ac.ir

چکیده: روزبه‌روز حجم داده‌های تولیدشده توسط انسان‌ها در حال افزایش است. ذخیره‌سازی و پردازش داده‌ها، یکی از مباحث بسیار مهم در مواجهه با کلان داده‌ها است که نیازمند منابع ذخیره‌سازی و پردازشی گسترده می‌باشد. رایانش ابری به‌دلیل ویژگی‌ها و معماری خود، به‌عنوان یک زیرساخت امیدبخش برای این مسئله مطرح است. برای ذخیره‌سازی و پردازش داده‌ها در بستر ابری، افراد و سازمان‌ها ممکن است برای کنترل و دسترسی بیشتر روی منابع و داده‌ها، تمایل بیشتری به استقرار و مدیریت ابرهای خصوصی داشته‌باشند. برای استقرار و مدیریت ابرهای خصوصی نرم‌افزارهای متعدد منبع‌باز ارائه‌شده‌است که گزینش بین آن‌ها به‌عنوان یک چالش به‌ویژه برای کسانی که تازه در این راه قدم‌برمی‌دارند، مطرح است. هدف ما در این مقاله ارائه یک مدل کیفیت برای مدیریت زیرساخت ابری و مقایسه نرم‌افزارهای محبوب *Eucalyptus*، *OpenStack* و *CloudStack* بر اساس مدل کیفیت ارائه‌شده با تمرکز بر کلان‌داده به‌منظور نمایش قابلیت‌ها، ضعف‌ها و برتری‌های هرکدام در جهت کمک به گزینش بین آن‌ها است.

واژه‌های کلیدی: پردازش کلان‌داده، ذخیره‌سازی کلان‌داده، رایانش ابری، مدیریت ابر، مدل کیفیت.

۱. مقدمه

امروزه، اطلاعات دیجیتال یک موضوع همگانی واجتناب‌ناپذیر است. اخیراً پیشرفت در فناوری‌های اطلاعاتی (نظیر ارتباطات از راه دور، نرم‌افزارهای موبایل‌های هوشمند، اینترنت اشیاء و شبکه‌های اجتماعی) سیل عظیمی از داده‌های دیجیتالی را به‌وجودآورده‌است. این حجم وسیع از داده‌ها اگر به درستی مدیریت شوند، می‌توانند به‌منظور انجام تجزیه و تحلیل و استخراج اطلاعات مفید جهت کاربردهای گوناگون از جمله بهبود کسب‌وکار، افزایش بهره‌وری و کاهش هزینه‌ها استفاده شوند. برای این منظور داده‌ها باید از منابع مختلف جمع‌آوری، ذخیره، پردازش و تجزیه و تحلیل شوند تا اطلاعات مفید از آن‌ها استخراج شود. این داده‌ها که مشخصه‌های متفاوت و پیچیدگی بالایی نسبت به داده‌های سنتی دارند، با عنوان کلان‌داده شناخته‌می‌شوند.

کلان‌داده، به تعریف مؤسسه تحقیقاتی گارتنر(Gartner)، اطلاعات با مشخصه‌های حجم بالا (High Volume)، سرعت بالا (High Velocity) و تنوع بالا (High Variety) هستند که برای امکان‌پذیر کردن تصمیم‌گیری، کشف دانش و بهینه سازی فرایند، به فرم‌های جدید پردازش نیاز دارند [۱]. مطابق این تعریف پایه‌ای، کلان‌داده با سه مشخصه اصلی حجم، سرعت و تنوع بازنمایی شده‌است. این در حالی است که اخیراً تعاریف متعددی که این داده‌ها را با مشخصه‌های بیشتری توصیف‌می‌کنند، ارائه‌شده‌است [۲-۳].

با توجه به مشخصه‌های کلان‌داده‌ها، نمی‌توان از تکنولوژی‌های سنتی برای مدیریت آن‌ها بهره‌برد. مقیاس‌ناپذیری، داشتن ساختاری متمرکز، ثبات و ساختار داده یک‌ریخت به‌عنوان دلایل مهم عدم‌کارایی تکنولوژی‌های داده‌های سنتی در مواجهه با کلان‌داده‌ها شناسایی‌شده‌است [۴]. بنابراین برای جمع‌آوری، ذخیره‌سازی، پردازش و تجزیه و تحلیل کلان‌داده‌ها، قدرت محاسباتی بالایی نیاز است تا بتوان فرم‌های پردازشی جدیدی را برای آن‌ها ارائه‌کرد. در همین راستا، ابرها بستر مناسبی را برای ذخیره‌سازی، پردازش و تجزیه و تحلیل کلان‌داده‌ها فراهم‌می‌آورند طوری‌که به دو مورد از نیازمندی‌های اصلی تجزیه و تحلیل کلان‌داده‌ها یعنی ذخیره‌سازی مقیاس بالا و محاسبات با کارایی بالا پاسخ می‌دهند.

به تعریف مؤسسه ملی فناوری و استانداردهای آمریکا (NIST) به‌عنوان یکی از پراستنادترین تعاریف [۵]، رایانش ابری مدلی است که امکان دسترسی همه‌جانبه، راحت و برحسب تقاضا از طریق شبکه به مجموعه عظیمی از منابع محاسباتی قابل‌تنظیم (از جمله سرورها، ذخیره‌سازها، برنامه‌ها، شبکه و خدمات) را فراهم‌می‌کند. این مدل می‌تواند به‌سرعت تهیه و با حداقل مداخله مدیریتی و یا تعامل با ارائه‌دهنده سرویس آماده ارائه خدمت شود. مطابق این تعریف رایانش ابری مدلی است که از پنج ویژگی اساسی ۱)ارائه سرویس خودکار برحسب تقاضا ۲) دسترسی گسترده از طریق شبکه ۳) مجموعه عظیمی از منابع ۴) انعطاف‌پذیری سریع ۵) خدمات اندازه‌گیری‌شده، سه مدل خدمت ۱) نرم‌افزار به‌عنوان خدمت (Software as a Service - SaaS) ۲) بستر به‌عنوان خدمت (Platform as a Service - PaaS) ۳) زیرساخت به‌عنوان خدمت (Infrastructure as a Service - IaaS)، و چهار مدل استقرار ۱) ابر خصوصی (Private Cloud) ۲) ابر عمومی (Public Cloud) ۳) ابر انجمنی (Community Cloud) ۴) ابر ترکیبی (Hybrid Cloud) تشکیل‌شده‌است. شکل ۱ نمایشی از تعریف فوق است.

شکل ۱: مدل رایانش ابری طبق تعریف مؤسسه ملی فناوری و استانداردها

رایانش ابری با استفاده از روش‌های مختلف مجازی‌سازی و کنترل منابع سعی بر کاهش هزینه‌ها برای کاربران نهایی از طریق بهینه‌سازی منابع و مدیریت بهتر آن‌ها دارد. علاوه‌بر این، رایانش ابری زیرساخت ذخیره‌سازی و پردازش حجم بسیار زیادی از داده‌ها را فراهم‌می‌آورد. با توجه به قابلیت‌های متعددی که رایانش ابری ارائه‌می‌کند، بسیاری از سازمان‌های کوچک و بزرگ در جهت انتقال سرویس‌های خود بر روی آن هستند تا پردازش‌های خود را بر بستر ابر انجام‌دهند. به‌منظور استقرار و مدیریت ابر، نرم‌افزارهای منبع‌باز متعددی از جمله Eucalyptus، OpenStack، Cloud Stack ارائه‌شده که با بهره‌گیری از آن‌ها می‌توان ابرها را در مدل‌های مختلف ابر خصوصی، ابر عمومی و یا ابر ترکیبی مستقرکرد. در این پژوهش، پس از بررسی برخی از نرم‌افزارهای مختلف استقرار و مدیریت ابر، یک مقایسه تطبیقی از آن‌ها ارائه‌شده‌است تا نقاط قوت، نقاط ضعف، و کاربرد هریک از این نرم‌افزارها به‌منظور گزینش بهتر در میان هریک از آن‌ها با تمرکز بر کلان‌داده مشخص‌شود.

ادامه این مقاله به شرح زیر است: در بخش ۲ ضمن معرفی چالش‌های مدیریت و پردازش کلان‌داده‌ها، راه‌حل‌ها و تکنولوژی‌های ارائه‌شده برای حل آن‌ها ذکرمی‌شوند. سپس رایانش ابری به‌عنوان بستری مناسب برای مدیریت و پردازش کلان‌داده‌ها معرفی‌می‌شود. در بخش ۳ ابتدا نرم‌افزارهای منبع‌باز استقرار و مدیریت ابر بررسی و معرفی‌می‌شوند، سپس یک مدل کیفیت برای مدیریت زیرساخت ابری با تمرکز بر کلان‌داده ارائه‌می‌شود و در ادامه مقایسه نرم‌افزارها با دو نگرش مقایسه عمومی و مقایسه بر اساس مدل کیفیت ارائه‌شده و با توجه به مطالعات انجام‌گرفته صورت‌می‌گیرد. در بخش ۴ نتیجه‌گیری از پژوهش فعلی شرح‌داده‌خواهدشد.

۲. مدیریت و پردازش کلان‌داده در بستر ابری

مشخصه‌های کلان‌داده‌ها چالش‌های جدیدی را در مدیریت و پردازش آن‌ها به‌وجودمی‌آورند. این چالش‌ها سبب تحمیل نیازمندی‌های جدید برای پردازش کلان‌داده‌ها می‌شوند. در جهت پاسخ به این نیازمندی‌ها، مدیریت و پردازش کارآمد کلان‌داده‌ها اغلب به تکنولوژی‌های خاصی مانند پردازش موازی گسترده، پایگاه داده‌های توزیع‌شده، سیستم‌های ذخیره‌سازی مقیاسپذیر و زیرساخت مناسب نیاز دارد.

ما در این بخش با تمرکز بر زیرساخت، ضمن معرفی چالش‌ها و نیازمندی‌های نشأت‌گرفته از مشخصه‌های کلان‌داده‌ها، درصدد نمایش رایانش ابری به‌عنوان بستری مناسب برای پردازش کلان‌داده‌ها هستیم.

۱،۲. چالش‌های مدیریت و پردازش کلان داده

همان‌طور که در بخش قبل گفته‌شد، کلان‌داده‌ها در یک تعریف پایه‌ای دارای سه مشخصه اصلی حجم بالا، سرعت بالا و تنوع بالا می‌باشند. مشخصه‌های فوق چالش‌های متعددی را در فازهای مختلف استخراج ارزش از داده‌ها به‌وجودمی‌آورند. با تمرکز بر زیرساخت، سه چالش اساسی که در مدیریت و پردازش کارآمد کلان‌داده‌ها باید مورد توجه باشد عبارتند از:

۱. ذخیره‌سازی و مدیریت داده‌ها: کلان‌داده‌ها با سرعت و تنوع بالا در حال تولید هستند. این امر منجر به تولید حجم بسیار زیادی از داده‌ها می‌شود که ذخیره‌سازی آن‌ها در حافظه فیزیکی سنتی مشکل ساز است [۶]. از طرفی سرعت تولید کلان‌داده‌ها مستلزم آن است که سیستم‌های ذخیره‌سازی مقیاس‌پذیر باشند.

۲. پردازش و تجزیه و تحلیل داده‌ها: پردازش حجم بسیار زیاد داده‌ها که با سرعت بالایی تولیدمی‌شوند، نیازمند منابع محاسباتی گسترده است. با افزایش سرعت پردازنده، شبکه و قدرت ذخیره سازی می‌توان تاحّدی منابع محاسباتی گسترده‌ای را فراهم‌کرد، بااین‌حال، منابع محاسباتی که برای پردازش کلان‌داده نیاز است بسیار فراتر از قدرت پردازش ارائه‌شده توسط این منابع محاسباتی است [۷].

۳. مدیریت منابع: همان‌طور که در چالش‌های ۱ و ۲ اشاره‌شد، ذخیره‌سازی و پردازش کلان‌داده‌ها نیازمند منابع ذخیره‌سازی و پردازشی گسترده است که مدیریت این منابع نیز خود یک چالش جدید را ایجادمی‌کند.

۲،۲. تکنولوژی‌های مدیریت و پردازش کلان‌داده

بر اساس توضیحات بخش ۱-۲، فضای ذخیره‌سازی مقیاس بزرگ و منابع محاسباتی قدرتمند، دو مورد از چالش‌های اصلی مدیریت و پردازش کلان‌داده‌ها هستند. در همین راستا روش‌ها و تکنولوژی‌های مختلفی ارائه‌شده‌است.

برای مقابله با چالش ذخیره‌سازی و مدیریت داده‌ها، باید از فایل سیستم‌هایی که فضای ذخیره‌سازی بالایی را ضمن قابلیت‌های مقیاس‌پذیری، افزونگی و تحمل‌پذیری خطا ارائه‌می‌دهند، بهره‌برد. فایل سیستم گوگل (GFS) [۸] و فایل سیستم توزیع‌شده Hadoop (HDFS) [۹] از جمله نمونه های محبوب هستند که برای ذخیره‌سازی کلان‌داده‌ها بهینه‌سازی‌شده اند [۶]. از طرفی کلان‌داده‌ها علاوه‌بر داشتن حجم بالا، اغلب داده‌هایی بدون ساختار هستند. این ویژگی سبب عدم‌کارایی پایگاه داده‌های رابطه‌ای (RDBMS) برای ذخیره‌سازی و مدیریت آن‌ها می‌شود. پایگاه داده‌های غیررابطه‌ای (NoSQL) [۱۰] مانند MongoDB و Cassandra راه‌حل‌های مناسب برای کلان‌داده‌ها می‌باشند.

برای مقابله با چالش پردازش و تجزیه و تحلیل داده‌ها، تکنولوژی‌های مختلفی بر پایه پردازش موازی و توزیع‌شده، ارائه‌شده‌اند. MapReduce [۱۱] یکی از نمونه‌های مشهور این تکنولوژی‌ها است. MapReduce یک مدل برنامه‌نویسی است که برای پردازش حجم زیادی از داده‌ها آن‌ها را به گروه‌هایی از کارهای مستقل از هم تقسیم می‌کند بدین‌صورت که گروه‌ها می‌توانند به‌صورت موازی توسط ماشین‌های مجزا پردازش‌شوند. MapReduce شامل دو تابع اصلی نگاشت و کاهش‌می‌باشد که با توجه به قابلیت‌های پردازش موازی، مقیاس‌پذیری بالا و تحمل‌پذیری خطا به یک تکنولوژی کارآمد برای پردازش کلان‌داده‌ها تبدیل‌شده‌است. علاوه‌بر این تکنولوژی‌ها و روش‌های متعدد دیگری از جمله Apache Spark برای این کار پیشنهادشده‌است [۱۲].

برای مقابله با چالش مدیریت منابع، می‌توان به دو روش عمل‌کرد. روش اول بدین‌صورت است که منابع ذخیره‌سازی و پردازشی مورد نیاز سازمان از طریق ابرهای عمومی نظیر آمازون (AWS) و مایکروسافت (Microsoft Azure) به‌عنوان خدمت دریافت‌شوند. این ابرها خدمات متعددی در مدل‌های مختلف خدمت برای ذخیره‌سازی و پردازش کلان‌داده ارائه‌می‌دهند. Amazon S3 به‌عنوان یک خدمت ذخیره‌سازی که قابلیت‌های مقیاس پذیری، در دسترس بودن داده‌ها، امنیت و کارایی را فراهم‌می‌کند، و Amazon EMR به‌عنوان یک پلتفرم کلان‌داده ابری که قابلیت‌های اجرای عملیات پردازش داده های توزیع‌شده در مقیاس بزرگ، پرس‌وجو های تعاملی SQL و برنامه‌های کاربردی یادگیری ماشین با استفاده از چارچوب‌های تحلیلی منبع باز مانند Apache Hive، Apache Spark و Presto را ارائه‌می‌دهد، از جمله از این خدمات هستند که توسط ابر عمومی AWS ارائه‌می شود [۱۳]. مطابق این روش افراد و سازمان‌ها داده‌های خود را باید در بستر ابر عمومی، ذخیره سازی و پردازش‌کنند. امنیت، شخصی سازی، هزینه و میزان کنترل بر روی منابع و داده ها محدودیت هایی هستند که سازمان‌ها در استفاده از ابرهای عمومی با آن مواجهند. روش دوم بدین‌صورت است که به‌جای دریافت مستقیم خدمات مورد نیاز از ابرهای عمومی، افراد و سازمان‌های علاقه‌مند به پردازش کلان‌داده مدیریت منابع ذخیره‌سازی و پردازشی را برعهده‌می‌گیرند. برای این کار نرم‌افزارها متعددی از جمله Eucalyptus، OpenStack و CloudStack مطرح‌می‌شوند که در بخش ۳ بررسی‌خواهندشد.

رایانش ابری، با فراهم‌آوردن تکنولوژی‌های ذخیره سازی و پردازش توزیع شده و پایگاه داده‌های غیررابطه‌ای برای مقابله با چالش‌های مورد

ا شاره برای کلان‌داده‌ها، به یک بستر منا سب برای مدیریت و پردازش کلان‌داده‌ها تبدیل شده‌ا ست. به‌طورکلی، ذخیره و پردازش کلان‌داده‌ها در بستر ابری را می‌توان به صورت یک معماری ۳-لایه‌ای مطابق شکل ۲ بازنمایی‌کرد.

مطابق این معماری، در بستر رایانش ابری برای هریک از چالش‌های مورد اشاره در بخش ۱-۲ تکنولوژی‌های مناسب در لایه‌های مختلف درنظرگرفته‌شده‌است. ارائه تکنولوژی‌های مدیریت و پردازش کلان‌داده از جمله فایل سیستم‌های HDFS و GFS، پایگاه داده‌های غیررابطه‌ای و تکنولوژی‌های پردازش موازی نظیر MapReduce در لایه‌های مختلف، رایانش ابری را به یک زیرساخت امیدبخش برای مقابله با چالش‌های کلان‌داده تبدیل‌کرده‌است.

شکل ۲: معماری ذخیره و پردازش کلان داده در بستر ابری

۳. مدیریت ابر

اصطلاح مدیریت ابر یک اصطلاح کلی است که دارای تعریف گسترده و زیربخش‌های مختلفی می‌باشد. ما در این بخش بر روی مدیریت زیرساخت ابری به‌عنوان زیرمجموعه‌ای از مدیریت ابر تمرکز خواهیم‌کرد.

مدیریت زیرســـاخت ابری به خدماتی اشـــاره‌می‌کند که امکان دسترسی و کنترل مجموعه عظیمی از منابع محاسباتی و ذخیره‌سازی قابل‌تنظیم (از جمله سرورها، ذخیره‌سازها و شبکه) را امکان‌پذیرمی‌کند. برای مدیریت زیرســـاخت ابری، نرم‌افزارهای مختلفی مطرح اســـت که قابلیت های زیادی را برای افراد و ســـازمان ها فراهم‌می‌کنند. این نرم‌افزارها به افراد و سازمان‌ها کمک‌می‌کند تا بر اهداف سازمانی خود تمرکز بیشـــتری داشـــته‌باشـــند. همچنین امکان مدیریت ابر را از طریق یک خط فرمان و یا رابط کاربری مناسب فراهم‌می‌کند.

در این بخش با تمرکز بر کلان‌داده، ابتدا سه نرم‌افزار منبع‌باز مدیریت زیرساخت ابری معرفی‌می‌شوند. سپس یک مدل کیفیت برای مدیریت زیرساخت ابری با تمرکز بر کلان‌داده ارائه‌می‌شود و در ادامه مقایسه نرم‌افزارها با دو نگرش مقایسه عمومی و مقایسه بر اساس مدل کیفیت ارائه‌شده و با توجه به مطالعات انجام‌گرفته، صورت‌می‌گیرد.

۱،۳. بررسی نرم‌افزارهای منبع باز مدیریت زیرساخت ابری

نرم‌افزارهای متعددی برای مدیریت زیرساخت ابری پیشنهادشده‌است. این نرم‌افزارها در یک دسته بندی کلی در دو دسته منبع‌باز(Open Source) و منبع‌بسته(Closed Source) قرارمی‌گیرند. ما نرم‌افزارهای منبع‌باز مدیریت زیرساخت ابری را به‌دلیل در دسترس بودن آن‌ها مورد توجه قرارخواهیم‌داد.

همان‌طور که اشـــاره‌شـــد نرم‌افزارهای متعددی برای مدیریت زیرساخت ابری ارائه‌شده‌است. ما نرم‌افزارهایی که با توجه به بازخوردها مورد علاقه بیشـــتر کاربران و ســـازمان‌ها در زمینه کلان‌داده بودند را انتخاب‌کرده‌ایم که عبارتند از: Eucalyptus، OpenStack و Cloud Stack.

Eucalyptus یک نرم‌افزار مدیریت ابر برای ایجاد ابرهای خصوصی و ترکیبی اســـت. اولین نســـخه آن در ســـال ۲۰۰۸ ارائه‌شـــد. پس از آن نســـخه‌های متعددی از این نرم‌افزار ارائه‌شـــد. نســـخه ۵،۱ به‌عنوان جدیدترین نسخه آن در May سال ۲۰۲۱ منتشر شد. ارائه مجموعه‌ای از APIهای ســـازگار با خدمات S3، EC2، ELB، CloudWatch، Auto Scaling، CloudFormation و IAM آمازون یکی از نکات مثبت این نرم‌افزار به‌شمارمی‌رود. از طرفی قابلیت نصب آسان آن از طریق اجرای چند دستور ساده نیز از دیگر ویژگی‌های مهم آن است. Eucalyptus را می‌توان روی بستر لینوکس نصب و اجرا کرد. یک نرم‌افزار ماژولار که از معماری لایه ای پیروی‌می‌کند، قابلیت مقیاس‌پذیری این ســـیســـتم را افزایش‌می‌دهد. از نرم‌افزار euca2ools برای تســـهیل تعامل کاربران با نرم‌افزار به‌عنوان رابط کاربری استفاده می‌کند. فضای کاربر و مدیریت کامل جدا از هم هســـتند و تا حد مناســـبی از درگیرشـــدن کاربر با پیچیدگی‌های زیرســـاختی و اجرایی ســـیســـتم جلوگیری‌می‌کند. برای تسهیل کار با آن، مستندات مناسبی ارائه‌شده‌است.

OpenStack دومین نرم‌افزار مدیریت ابر است که به آن می‌پردازیم. نسخه اول آن در سال ۲۰۱۰ توسط ناسا و رک اسپیس (Rackspace) توسعه‌داده شد. نسخه‌های مختلف آن با اسامی مختلفی ارائه‌می شوند. اولین نســـخه آن تحت‌عنوان Austin منتشـــرشـــد. در حال حاضـــر جدیدترین نســـخه آن تحت‌عنوان Yoga در March ســـال ۲۰۲۲ منتشرشده‌است. کد منبع OpenStack تحت مجوز Apache License 2 قراردارد. OpenStack تعامل کاربر با سیستم را از دو طریق رابط کاربری وب و API فراهم‌می‌کند. ماموریت OpenStack این اســـت که "یک پلتفرم رایانش ابری متن باز ایجادکند که اســـتفاده از آن آســـان اســـت، پیاده سازی آن ساده است، بین استقرارها قابلیت تعامل دارد، در همه مقیاس ها خوب کارمی‌کند و نیازهای کاربران و اپراتورهای ابر های خصوصی و عمومی را برآورده‌می‌کند" [۱۴].

OpenStack در اولین نسخه خود از ۲ مؤلفه اصلی ۱) محاسبات (Nova) که نقش اصلی آن مدیریت منابع پردازشی و شبکه‌های بزرگ اســـت. و ۲) ذخیره‌ســـازی(Swift) به‌عنوان یک ســـیســـتم ذخیره‌ســـازی توزیع‌شـــده که قابلیت مقیاس‌پذیری را با افزایش منابع ذخیره‌ســـازی جدید فراهم‌می‌آورد تشـــکیل‌شـــده‌بود. اما با ارائه نســـخه‌های جدیدتر همواره تعداد مؤلفه ها و ویژگی های آن نیز افزایش یافت به‌طوری که Yoga به‌عنوان جدیدترین نســـخه آن بیش از ۳۰ مؤلفه از جمله NOVA، SAHARA، CINDER و SWIFT دارد.

Apache Cloud Stack سومین نرم‌افزار مدیریت ابر است که به آن می‌پردازیم. Apache Cloud Stack به‌عنوان یک پروژه در سال ۲۰۰۸ در

شرکت نوپای VMOps آغازشد که امکان ایجاد ابرهای عمومی و خصوصی را فراهم‌می‌کند. همچنین برخی از سازمان‌ها از آن به‌عنوان بخشی از ابر ترکیبی نیز استفاده‌می‌کنند. برای تعامل کاربر با زیرساخت، سه روش رابط کاربری تحت وب، API و خط فرمان (CLI) درنظرگرفته-شده‌است. از قابلیت‌های مثبت دیگر آن می توان به پشتیبانی از ناظران ماشین متعدد و مشهور از جمله KVM، Citrix XenServer و XenCloud اشاره‌کرد. علاوه‌براین ارائه APIهای سازگار با S3 و EC2 آمازون از دیگر قابلیت‌های آن به‌شمارمی‌رود.

۴. ارائه یک مدل کیفیت برای مدیریت زیرساخت ابری از دیدگاه پردازش کلان‌داده

مدل کیفیت، به‌تعریف استاندارد ISO 25010 سنگ‌بنای ارزیابی کیفیت یک محصول نرم‌افزاری است. مدل کیفیت، ویژگی‌های کیفیتی را که هنگام ارزیابی ویژگی‌های یک محصول نرم‌افزاری باید درنظرگرفته‌شوند را تعیین‌می‌کند [۱۵]. به‌عبارتی‌دیگر، هر مدل کیفیت یک دیدگاه کیفیتی را به‌صورت مجموعه‌ای از مشخصه‌های کیفیتی و روابط بین آن‌ها اتخاذمی‌کند.

مدل‌های کیفیت مختلفی از جمله مدل McCall، مدل Bohem، مدل FURPS، مدل ISO 9126 و مدل ISO 25010 ارائه شده‌ا ست. در [۱۶] یک مطالعه مقایسه‌ای از برخی مدل‌های کیفیت انجام‌شده‌است.

ما به‌منظور ارائه یک مدل کیفیت به‌عنوان مبنایی برای انجام مقایسه نرم‌افزارهای مورد اشاره، پس از بررسی مدل‌های مختلف کیفیت با تکیه بر ویژگی های کلان‌داده ها و چالش های مطرح در پردازش کلان‌داده‌ها در بستر ابر که در بخش‌های قبل مورد بحث قرارگرفت، سه مدل کیفیت FURPS، ISO 9126 و ISO 25010 را گزینش‌کرده و بر اساس آن‌ها مدل کیفیت مناسب برای ارزیابی کیفیت زیرساخت ابری از دیدگاه پردازش کلان‌داده را ارائه‌می‌کنیم. این مدل کیفیت امکان ارزیابی و مقایسه بهتر نرم‌افزارهای زیرساخت ابری از دیدگاه پردازش کلان‌داده را فراهم‌می‌نماید. مدل کیفیت ارائه‌شده با عنوان مدل کیفیت زیرساخت ابر QMCI که مخفف عبارت Quality Model for Cloud Infrastructure می‌باشد در شکل ۳ نشان‌داده‌شده-است.

مطابق مدل کیفیت ارائه‌شده، مشخصه های کیفیتی Functionality، Usability، Reliability، Supportability و Performance به‌عنوان فاکتورهای اصلی در سطح یک قراردارند و هرکدام از این فاکتورها شامل زیر فاکتورهایی در سطح دوم ا ست که در ادامه به توضیح هرکدام از آن‌ها می‌پردازیم.

Suitability، توانایی محصول نرم‌افزاری برای ارائه مجموعه‌ای از عملکردهای مناسب برای وظایف مشخص‌شده و اهداف کاربر را نشان‌می‌دهد.

Interoperability، توانایی محصول نرم‌افزاری برای تعامل با یک یا چند سیستم مشخص را نشان‌می‌دهد.

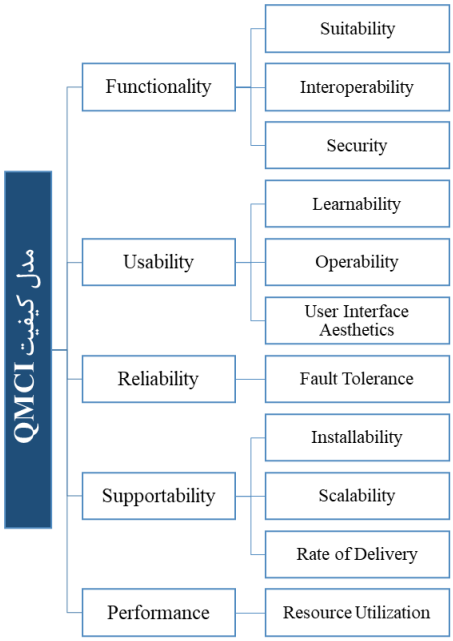
Security، توانایی محصول نرم‌افزاری برای محافظت از اطلاعات و داده‌ها در مقابل دسترسی و اعمال تغییر افراد و سیستم‌های غیرمجاز است‌Tطوری‌که افراد و سیستم‌های مجاز فقط بتوانند به آن‌ها دسترسی و کنترل داشته‌باشند را نشان‌می‌دهد.

Learnability، توانایی محصول نرم‌افزاری که کاربر را قادر می‌سازد تا آن محصول و کاربردهای آن را بیاموزد.

Operability، توانایی محصول نرم‌افزاری که کاربر را قادر می-سازد تا آن را به فعالیت وادارد و کنترل‌کند.

User Interface Aesthetics، توانایی محصول نرم‌افزاری را در ارائه یک رابط کاربری رضایت‌بخش برای تعامل کاربر با محصول نشان‌می‌دهد.

Fault Tolerance، توانایی محصول نرم‌افزاری برای حفظ سطح مشخصی از عملکرد در هنگامی که محصول و یا رابط مشخص‌شده آن با نقص روبروشده‌است را نشان‌می‌دهد.



شکل ۳: مدل کیفیت مدیریت زیرساخت ابری از دیدگاه پردازش کلان‌داده

Install ability، توانایی محصول نرم‌افزاری برای نصب در یک محیط مشخص را نشان‌می‌دهد.

Scalability، توانایی محصول نرم‌افزاری برای رسیدگی به حجم فزاینده‌ای از کارها با افزودن منابع به محصول را نشان‌می‌دهد.

Rate of Delivery، نرخ نسخه های جدیدی از نرم‌افزار که در اختیار کاربران قرار می‌گیرد را نشان‌می‌دهد.

Resource Utilization، توانایی محصول نرم‌افزاری برای استفاده از میزان و انواع مناسب منابع در حین انجام عملیات خود در شرایط ذکر شده را نشان‌می‌دهد.

۱،۴. مقایسه نرم‌افزارهای مدیریت زیرساخت ابری

برای انجام مقایسه بین نرم‌افزارهای مورد اشاره، ما دو نوع مقایسه را در این بخش بررسی‌خواهیم‌کرد. در مقایسه اول ویژگی‌های عمومی مورد بحث ما خواهدبود و در مقایسه دوم بر مبنای مدل کیفیت ارائه‌شده در بخش قبل، مشخصه‌های کیفیتی را در مورد هریک از نرم‌افزارها بررسی-خواهیم‌کرد.

مقایسهٔ عمومی

در مقایسه عمومی، فاکتورهای عمومی نرم‌افزارها موردتوجه قرارمی-گیرند. برخی از فاکتورهای مورد بررسی عبارتند از:

۱. مجوز اجرا: در حالی که نرم‌افزار های منبع باز در این م قا له موردتوجه قرارگرفته‌اند، بااین‌حال، هرکدام مجوزهای اجرای متفاوتی دارند. هریک از نرم‌افزارها با توجه به مجوز اجرایی که دریافت‌کرده‌اند می‌توانند در اهداف مختلفی استفاده‌شوند. به‌عنوان نمونه یکی از این مجوزها Apache License 2.0 است که مجوز نسبتاً سهل‌گیری می‌باشد. به‌طوری که پروژه‌هایی که تحت این مجوز هستند برای هر هدفی قابل‌استفاده می‌باشند.

نتایج حاصل از انجام مقایسه عمومی بر اساس فاکتورهای فوق در جدول ۱ ارائه‌شده‌است.

مقایسه بر اساس مدل کیفیت

برای انجام مقایسه بین نرم‌افزارهای مورد اشاره بر مبنای مدل کیفیت، مدل QMCI را مورد توجه قرارداده و هریک از فاکتورهای آن را در دامنه هر یک از نرم‌افزارها بررسی و مطالعه‌کرده‌ایم. در این مسیر مطالعات پیشین پاسخگوی برخی از نیازهای مدل مورد نظر بوده است که به آن‌ها ارجاع‌شده‌است. بااین‌حال، در جهت تکمیل جدول و پاسخ به نیاز همه فاکتورهای مدل ارائه‌شده، آخرین نسخه هر یک از نرم‌افزارها را مطالعه-کرده‌ایم. بررسی‌ها و دستاوردهای ما عبارتند از:

۱. در [۱۷] فاکتورهای Suitability و Security بررسی‌شده‌اند. بر اساس این مطالعه، برخلاف OpenStack و Cloud Stack که در جهت اهداف شرکت‌ها، ارائه‌دهندگان خدمات و محققان مناسب است، Eucalyptus شرکت‌های تجاری بزرگ و مؤسسات تحقیقاتی را هدف قرارمی‌دهد. از طرفی هریک از نرم‌افزارها مکانیزم‌های امنیتی متفاوتی استفاده‌می‌کنند که می‌تواند بر اساس نیازهای سازمان مهم را باشد.

۲. Interoperability: در دامنه مورد مطالعه ما که کلان‌داده ا ست، سازگاری و تعامل با ابر آمازون به‌عنوان محبوب‌ترین ابر عمومی می‌تواند مزیت بسیار مهم برای هریک از نرم‌افزارها باشد. ابر عمومی آمازون خدمات مختلفی از جمله Amazon S3، Amazon EMR و Amazon EBS برای ذخیره‌سازی و پردازش کلان‌داده ها ارا ئه‌می‌د هد. اگر نرم‌افزارهای مورد مقایسه، APIهایی سازگار برای تعامل با هر یک از این خدمات ارائه‌دهند، ضمن آنکه توسعه‌دهندگان به جهت آشنایی خود با ابر آمازون و وب سرویس آن می‌توانند با سهولت بیشتر از نرم‌افزارهای مورد نظر را استفاده‌کنند، افراد و سازمان‌ها نیز می‌توانند در کنار ابر خصوصی خود از خدمات ارائه‌شده توسط ابر عمومی آمازون نیز بهره‌مندشوند. این کار مزیت‌های بسیاری از جمله مقیاس‌پذیری بیشتر را فراهم‌می‌کند. بنابراین بررسی اینکه هر یک از نرم‌افزارهای مورد مقایسه با کدام یک از خدمات ارائه شده توسط ابر عمومی آمازون سازگار هستند یک موضوع مهم است.

از طرفی هر یک از نرم‌افزارهای مورد مقایسه از ناظران مختلفی پشتیبانی‌می‌کنند. ناظران ماشین علاوه‌بر ارائه قابلیت‌های مختلف، می‌توانند منبع باز یا منبع‌بسته باشند. بنابراین اینکه هر یک از نرم‌افزارهای مورد مقایسه از کدام ناظران ماشین پشتیبانی‌می‌کنند مسئله مهم دیگری است که در Interoperability باید بدان توجه‌شود.

۳. Learnability: به‌طور کلی در مواج هه با ه مه محصولات نرم‌افزاری، قابلیت یادگیری آن فاکتور مهمی تلقی‌می‌شود. از طرفی در دامنه پژوهش فعلی این فاکتور زمانی بیش از پیش اهمیت پیدامی‌کند که مشاهده می‌شود بیشتر افراد و سازمان‌ها که تمایل به استقرار خصوصی ابرها و مدیریت آن‌ها دارند، لزوماً افراد دارای تخصص‌های فنی بالا در حوزه فناوری و اطلاعات نیستند. بنابراین اینکه هریک از مح صولات با ارائه م ستندات و آموزش‌های منا سب، بتوانند ا ستفاده از محصول نرم‌افزاری را برای هر نوع کاربری تسهیل بخشند، فاکتور مهم دیگری است که باید مورد توجه قرارگیرد.

در همین راستا هر سه نرم‌افزار مورد بحث مستندات جامع و کاملی را در اختیار کاربران خود قرارداده اند. مستندات نسخه ۵.۰* Eucalyptus به‌عنوان آخرین نسخه آن در [۱۸] دردسترس است. همچنین اطلاعات نسخه‌های مختلف از جمله تغییرات اعما شده و ویژگی‌های جدید نسبت به نسخه‌های قبل، به‌طور کامل در بخش Release Notes از [۱۹] ارائه شده‌ا ست. م ستندات آخرین ن سخه مورد برر سی Apache Cloud Stack نیز در [۲۰] قابل‌دسترس است افزون بر این، اطلاعات نسخه های مختلف و ویژگی های نسخه جد ید نیز به‌صورت د سته‌بندی شده در بخش مورد ا شاره در د سترس ا ست. Open stack م ستندات خود را در قالب ۹ زبان مختلف از جمله زبان‌های انگلیسی و فرانسوی ارائه‌کرده و به مراتب طیف و سیع تری را ن سبت به دیگر نرم افزارها تحت پوشش قرارداده است. این مستندات در [۲۱] در دسترس ا ست، با این‌حال، همه مستندات آن در ۹ زبان مختلف ارائه‌نشده‌است، بلکه در برخی زبان‌ها تنها بخشی از مستندات در دسترس است.

برای انجام مقای سه، ما نرم‌افزارها را بر ا ساس م ستندات ارائه شده آن ها برای کاربران، مدیران ابر و توسعه‌دهندگان بررسی‌کرده‌ایم. مستندات کاربران، به‌عنوان راهنمای جامعی از نحوه استفاده خدماتی است که نرم‌افزارها ارائه‌می‌دهند. مدیریت ابر و چگونگی انجام آن در مستندات مخصوص مدیران ابر ارائه‌می‌شود. آشنایی با معماری نرم افزار و نحوه م شارکت در تو سعه آن موارد مهم دیگری ا ست که اگر در اختیار تو سعه‌دهندگان قرارگیرد زمینه ساز ایجاد جامعه گ سترده ای از توسعه‌دهندگان و در نتیجه بلوغ نرم‌افزار می‌شود، این اطلاعات در قالب مستندات توسعه‌دهندگان ارائه‌می‌شود. بنابراین بود یا نبود هریک از

جدول ۱: مقایسه عمومی نرم‌افزارهای مدیریت زیرساخت ابری

مشخصه / ابزار	Eucalyptus	OpenStack	CloudStack
نسخه فعلی	v 5.1	Yoga	v 4.16.1.0
نوع ابر پیاده سازی	خصوصی و ترکیبی	خصوصی و عمومی	خصوصی و عمومی
مجوز اجرا	GPLv3	Apache License 2.0	Apache License 2.0

جدول ۲: مقایسه نرم‌افزارهای مدیریت زیر جساخت ابری بر اساس مدل کیفیت

فاکتور کیفیت	زیر فاکتورکیفیت		Eucalyptus	OpenStack	CloudStack
Functionality	Suitability [۱۷]		شرکت های تجاری بزرگ و موسسات تحقیقاتی	شرکت ها، ارائه دهندگان خدمات و محققان	شرکت ها، ارائه دهندگان خدمات و محققان
	Interoperability		سازگاری با خدمات ابر آمازون: ۷ خدمت (S3، EC2، ELB، IAM، CloudWatch، Autoscaling، CloudFormation)	سازگاری با خدمات ابر آمازون: ۲ خدمت (EC2 و S3)	سازگاری با خدمات ابر آمازون: ۲ خدمت (EC2 و S3)
			ناظران ماشین: ۳ ناظر (KVM، Xen و VMware)	ناظران ماشین: ۱۲ ناظر (Ironic، VMware ESX/ESXi، PowerVm، zVM، KVM، LXC، QEMU، UML، Virtuozzo، XenServer، Xen، Hyper-V)	ناظران ماشین: ۶ ناظر (KVM، XenServer، Xen، Hyper-V، LXC و vSphere)
	Security		Cloud Controller یک جفت کد کلید عمومی/خصوصی برای احراز هویت کاربر ایجاد می کند	مؤلفه NOVA که به عنوان هسته محاسبات پروژه استفاده شده است به مدیران امکان مدیریت حساب های کاربری و منابع پروژه را می دهد. از Keystone برای مدیریت احراز هویت استفاده شده است. و همچنین API شامل محافظت در مقابل حملات DoS و مشتریان معیوب است	از گروهی از ماشین های مجازی تحت عنوان Security Group برای امنیت استفاده می کند که ترافیک های ورودی و خروجی خود را بر اساس مجموعه ای از قوانین فیلتر می کنند
Usability	Learnability	Administration Docs	✓	✓	✓
		Developers Docs	×	✓	✓
		User Docs	✓	✓	✓
	Operability		آماده سازی سکوی پردازش کلان داده باید به صورت دستی انجام گیرد	ارائه مؤلفه SAHARA برای آماده سازی سکوی پردازش کلان داده	آماده سازی سکوی پردازش کلان داده باید به صورت دستی انجام گیرد
	User Interface Aesthetics		۳ رابط (AWS CLI، Euca2ools و Management Console)	۲ رابط (GUI و API)	۳ رابط (GUI و API و CLI)
Reliability	Fault Tolerance		پشتیبانی بر اساس رویکرد Replication	پشتیبانی بر اساس رویکرد Replication	پشتیبانی بر اساس رویکرد Replication

این م ستندات عامل مهم دیگری ا ست که می‌تواند بر روی قابلیت یادگیری و استفاده از نرم‌افزار خاص اثرگذار باشد.

ارا ئه رابط های کاربری متنوع، از دیگر موارد مهم تأثیر گذار در افزایش قابلیت یادگیری و اســتفاده در ســیســتم‌های نرم‌افزاری مورد اشاره خواهدبود. دربارهٔ این موضوع، هر سه نرم‌افزار مورد هدف، عملکرد مناســبی را نمایش‌می‌دهند، طوری که رابط های گو ناگونی را برای تعامل کاربران با سی ستم درنظرگرفته‌اند که این مو ضوع به‌تف صیل در فاکتور کیفیتی دیگری بحث‌خواهدشــد.. بنابراین به‌طور کلی می‌توان گفت هر سه نرم‌افزار قابلیت یادگیری منا سبی در اختیار کاربران خود قرارمی‌دهند.

۳. Operability: برای ذ خیره‌ســازی و پردازش کلان‌داده‌هــا ویژگی‌های مقیاس‌پذیری، افزونگی و تحمل‌پذیری خطا بهعنوان ویژگی‌های ضروری مطرح شد. Apache Hadoop به‌عنوان یکی از تکنولوژی‌هایی اســت که این ویژگی‌ها را در خود دارد. Hadoop از دو بخش اصــلی فایل ســیســتم (HDFS) و پردازش (MapReduce) تشــکیل‌شــده اســت. بنابراین اگر بتوانیم به‌عنوان مثال Hadoop را در نرم‌افزارهای مورد مقایسه مستقرکنیم، توانسته‌ایم به یکی از چالش‌های پردازش و ذخیره‌ســازی کلان‌داده پاســخ دهیم. برای این کار نرم‌افزار OpenStack مؤلفه SAHARA را ارائه‌داده‌است، که با کمک آن به‌راحتی می‌توان از Hadoop در بستر ابری استفاده‌کرد. اما در سایر نرم‌افزارهای مورد بررســی، باید آن را به‌صــورت دســتی پیاده‌ســازی‌کنیم. بنابراین چگونگی آماده‌ســازی ســکوی پردازش کلان‌داده در هریک از نرم‌افزارها موضوع مهم دیگری در دامنه مورد مطالعه است.

۴. User Interface Aesthetics: ارا ئه رابط های کاربری متنوع و مختلف یکی از ویژگی‌های مهم در جهت ت سهیل ا ستفاده از نرم‌افزارها

برای افراد و سازمان‌ها است. هریک از نرم‌افزارهای مورد مقایسه روش‌های مختلفی را برای تعامل با کاربر درنظرگرفته‌اند.

۵. Fault Tolerance: همان‌طور که در بخش‌های قبل تأکیدشد، تحمل‌پذیری خطا به‌عنوان یکی از ویژگی‌های ضروری در ذخیره‌سازی و پردازش کلان‌داده‌ها مطرح‌شد. بنابراین پشتیبانی تحمل‌پذیری خطا در هریک از نرم‌افزارها یمورد اشاره، مسئله مهم دیگری است که باید مورد بررسی قرارگیرد.

۶. Install ability: بیشتر نرم‌افزارهای مدیریت زیرساخت ابری روی همه سیستم عامل‌ها قابل پیاده‌سازی و اجرا نیستند. لذا این موضوع که روی چه سیستم عامل‌هایی می‌توان هریک از نرم‌افزارها را استفاده‌کرد، نکته مهم دیگری است که باید مورد توجه باشد.

۷. Scalability: با توجه به ویژگی‌های کلان‌داده‌ها، برای ذخیره‌سازی و پردازش آن‌ها ویژگی مقیاس پذیری مشخصه بسیار مهمی است که باید به آن پرداخته‌شود. بر اساس مطالعه صورت‌گرفته در [۲۲]، OpenStack با توانایی مدیریت داده‌های حجیم تا میزان پتابایت به‌صورت توزیع‌شده، با قابلیت مقیاس‌پذیری بالا تا ۱ میلیون ماشین فیزیکی و ۶۰ میلیون ماشین مجازی و میلیاردها شیء ذخیره‌شده، به‌عنوان یک نرم‌افزار مقیاس‌پذیر در سطح انبوه معرفی‌شده‌است. از طرفی مطالعه [۲۳] نیز مقیاس‌پذیربودن نرم‌افزارهای Cloud Stack و Eucalyptus را نشان‌می‌دهد. بنابراین بر اساس مطالعات فوق می‌توان گفت از نظر سطح مقیاس‌پذیری، OpenStack عملکرد بهتری نسبت به دو نرم‌افزار دیگر دارد.

۸. Rate of Delivery: نیازمندی‌ها همواره در حال تغییر و تحول هستند و اگر نرم‌افزارها در کوتاه‌ترین زمان ممکن خدمات خود را با نیازمندی‌های جدید تطبیق‌ندهند، از رقابت خارج‌خواهندشد. تطبیق نرم‌افزارها با نیازمندی‌های جدید، با ارائه نسخه‌های جدید آن‌ها همراه است. بنابراین در هر نسخه جدید خدمات بهتر و نسخه تکامل‌یافته‌تری از نرم‌افزار در اختیار کاربران قرارمی‌گیرد. در همین راستا فاکتور کیفیت Rate of Delivery نرخ نسخه‌های جدیدی از نرم‌افزار را که در اختیار کاربران قرارمی‌گیرد، نشان‌می‌دهد.

برای اندازه گیری این فاکتور، تعداد نسخه‌های ارائه‌شده برای یک نرم‌افزار در یک بازه زمانی خاص درنظرگرفته‌می‌شود. ازآنجاکه نسخه اولیه هر سه نرم‌افزار در یک اختلاف زمانی کوتاه با یکدیگر ارائه‌شده‌است، ما برای سنجش این فاکتور، تعداد همه نسخه‌های ارائه‌شده برای هر نرم‌افزار را از ابتدا تاکنون مورد سنجش قرارداده‌ایم.

۹. Resource Utilization: در رابطه با کارایی مطالعات مختلفی انجام‌شده‌است. در [۲۴]، نرم‌افزارهای Eucalyptus، Open stack و Open nebula از نظر کارایی بررسی‌شده‌اند. این مطالعه که کارایی نرم‌افزارها در استفاده از منابع را از ۳ نگرش کارایی پردازنده (CPU Performance)، کارایی شبکه (Network Performance) و کارایی ورودی/خروجی (I/O Performance) مطالعه‌کرده‌است در هر ۳ نگرش Eucalyptus عملکرد بهتری نسبت به OpenStack نشان‌داده‌است. در مطالعه [۲۵] نرم‌افزارهای OpenStack و Cloud Stack بررسی و مقایسه‌شده‌اند.

در این مطالعه که مقایسه بر مبنای Unix Bench و Bonnie++ به انجام‌رسیده‌است، در بررسی Bonnie++ که آزمون‌هایی از جمله کارایی ورودی/خروجی، استفاده از پردازنده و سرعت اجرا را مورد هدف قرارمی‌دهد، هر دو نرم‌افزار در استفاده از پردازنده برابر عمل‌کرده‌اند، در عمل نوشتن، Stack Cloud با اختلاف ناچیز، بهتر عمل‌کرده‌است، در سایر موارد OpenStack با اختلاف، عملکرد بهتری از خود نشان‌داده‌است. در طرف دیگر OpenStack در بررسی Unix Bench امتیاز شاخص ۲۰۰۲,۲ را دریافت‌کرده‌است. در مقابل Cloud Stack امتیاز ۱۴۴۳,۷ را دریافت‌کرده‌است. مطالعه [۲۸] نیز به بررسی و ارزیابی کارایی دو نرم‌افزار Cloud Stack و Eucalyptus پرداخته‌است. در این مطالعه کارایی نرم‌افزارها از دیدگاه‌های مختلفی از جمله کارایی پردازنده، کارایی شبکه و کارایی ورودی/خروجی در شرایط مختلف مورد ارزیابی قرارگرفته‌است، که نتایج حاصل به‌عنوان بخشی از اطلاعات برای مقایسه کارایی نرم‌افزارها در این پژوهش استفاده‌شده است.

نتایج حاصل از انجام مقایسه بر اساس مدل کیفیت در جدول ۲ ارائه‌شده‌است.

با تأمل بیشتر در جداول ۱ و ۲ مشاهده‌می‌شود که Eucalyptus رابط‌های کاربری متنوعی را نسبت به سایر نرم‌افزارها ارائه‌می‌دهد. در مقابل OpenStack روی همه سیستم عامل‌های رایج قابل‌اجرا است ولی اجرای نرم‌افزار های Cloud Stack و Eucalyptus محدود به برخی از سیستم‌عامل های خاص است. OpenStack و Cloud Stack هر دو تحت مجوز یکسان Apache License 2.0 هستند در حالی که Eucalyptus تحت مجوز GPLv3 ا ست. Eucalyptus با بی شتر خدمات ارائه‌شده توسط ابر عمومی آمازون برای کلان‌داده سازگار است، درحالی که نرم‌افزارهای دیگر تنها با خدمات S3 و EC2 آمازون سازگار هستند. نکته حائز اهمیت دیگر این است که نرم‌افزارهای Eucalyptus و Cloud Stack مؤلفه‌های مشخصی برای آماده سازی سکوی پردازش کلان‌داده ارائه‌نمی‌دهند. در این نرم‌افزارها آماده‌سازی سکوی پردازشی باید دستی انجام‌گیرد. بدین معنا که پیاده‌سازی و اجرای Apache Hadoop، Apache Spark در این نرم‌افزارها باید توسط خود افراد و سازمان‌ها صورت‌گیرد [۲۶]. اما، OpenStack با ارائه مؤلفه SAHARA آماده‌سازی سکوی پردازش کلان‌داده را به‌راحتی برای کاربران فراهم‌می‌کند. هدف از پروژه SAHARA، ارائه ابزاری ساده به کاربران برای فراهم‌کردن چارچوب‌های پردازش داده از جمله Apache Hadoop، Apache Spark و Apache Storm است [۲۷]. تحمل‌پذیری خطا در هر سه نرم‌افزار پشتیبانی‌می‌شود و از نظر کارایی Eucalyptus عملکرد بهتری از نرم‌افزارهای دیگر دارد. هر سه نرم‌افزار مقیاس‌پذیر می باشد. بااین‌حال، از این نظر OpenStack عملکرد بهتری نسبت به سایرین دارد.

بر اساس توضیحات فوق و مقایسه ارائه‌شده، افراد و سازمان‌ها می‌توانند بر اساس درجه اهمیت فاکتورهای کیفیتی مورد نظر خود به گزینش نرم‌افزار مناسب بپردازند.

۵. نتیجه‌گیری

در این مقاله پس از بررسی مفاهیم کلان‌داده و مشخصه‌های آن، ذخیره‌سازی و پردازش این حجم عظیم از داده‌ها و مدیریت منابع گسترده مورد نیاز به‌عنوان سه چالش اساسی مطرح‌شدند. از طرفی برای پاسخ‌دهی به چالش‌های ذخیره‌سازی و پردازش کلان‌داده‌ها، منابع ذخیره‌سازی و محاسباتی گسترده‌ای مورد نیاز است. با توجه به ویژگی‌ها و معماری ابرها، رایانش ابری به‌عنوان یک بستر امیدبخش برای ذخیره‌سازی و پردازش کلان‌داده‌ها معرفی‌شده‌است. به‌منظور ذخیره‌سازی و پردازش کلان‌داده‌ها در بستر رایانش ابری، منابع گسترده مورد نیاز را می‌توان به‌صورت یک خدمت از ابرهای عمومی مانند ابر آمازون دریافت کرد. با این حال ابرهای عمومی محدودیت‌های از قبیل امنیت، شخصی‌سازی و میزان کنترل بر روی منابع و داده‌ها را همراه‌دارند. بنابراین برای کاهش محدودیت‌های فوق می‌توان مدیریت و استقرار ابرها را بر عهده افراد و سازمان‌های علاقه‌مند به پردازش کلان‌داده‌ها قرارداد که بحث مدیریت زیرساخت ابری مطرح‌می‌شود که خود چالش دیگری است.

مدیریت زیرساخت ابری به‌عنوان یک زیر مجموعه از مفهوم کلی مدیریت ابر عنوان‌شد که نرم‌افزارهای متعدد منبع‌باز و منبع‌بسته را شامل‌می‌شود. نرم‌افزارهای منبع‌باز با توجه به قابلیت شخصی‌سازی آنها بر حسب نیاز سازمان و افراد می‌تواند گزینه جذاب‌تری برای کاربران آن‌ها باشد. هریک از نرم‌افزارهای منبع‌باز مدیریت زیرساخت ابری دارای ویژگی‌های متعدد و گاه مختلفی هستند. در این مقاله ما ضمن ارائه یک مدل کیفیت برای مدیریت زیر ساخت ابری، نسخه‌های فعلی هریک از این نرم‌افزارها را با تمرکز بر کلان‌داده‌ها مطالعه‌کرده و با دو نگرش مقایسه عمومی و مقایسه بر اساس مدل کیفیت ارائه‌شده، به مقایسه پرداختیم. مطابق این بررسی نشان‌داده‌شد که هریک از نرم‌افزارها در برخی از فاکتورها نسبت به سایر نرم‌افزارها برتری و یا نقص‌هایی دارند که مورد بحث قرارگرفت.

برتری‌ها و نقص‌های هریک از نرم‌افزارها بر اساس نیاز سازمان‌های مختلف بسته به اهداف آن‌ها می‌تواند برای هریک از آن‌ها درجه اهمیت متفاوتی داشته‌باشد. بنابراین سازمان‌ها می‌توانند نسبت به نیازهای خود به هریک از فاکتورهای کیفیتی ارائه‌شده در مدل QMCI وزن‌هایی به منزله درجه اهمیت آن‌ها اختصاص‌دهند. بر اساس این وزن‌دهی می‌توان در میان نرم‌افزارهای فوق تصمیم‌گیری نهایی را انجام‌داد. در همین راستا یکی از مواردی که می‌توان به‌عنوان کار آینده در این پژوهش مطرح‌کرد استفاده از روش‌های تصمیم‌گیری چندمعیاره از جمله تکنیک AHP برای گزینش نهایی نرم‌افزار مناسب سازمان بر اساس دو عامل درجه اهمیت فاکتورهای کیفیتی و جدول مقایسه‌ای ارائه‌شده، است.

مراجع

[1] Mark Beyer , Douglas Laney , "The Importance of 'Big Data': A Definition," 21 6 2012. [Online]. Available: https://www.gartner.com/en/documents/2057415/the-importance-of-big-data-a-definition. [Accessed 16 12 2021].

[2] Cheikh Kacfah Emani, Nadine Cullot, Christophe Nicolle, "Understandable Big Data: A survey," *Computer Science Review*, vol. 17, pp. 70-81, 2015.

[3] Amir Gandomi, Murtaza Haider, "Beyond the hype: Big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, pp. 137-144, 2015.

[4] Chun-Wei Tsai, Chin-Feng Lai, Han-Chieh Chao, Athanasios V. Vasilakos, "Big Data Analytics: A Survey," *Journal of Big Data*, vol. 2, no. 21, 14 05 2015.

[5] Peter Mell, Tim Grance, "The NIST Definition of Cloud," *Recommendations of the National Institute of Standards and Technology*, vol. 800, no. 145, 2011/9/28.

[6] Yaoxue Zhang, Ju Ren, Jiagang Liu, Chugui Xu, Hui Guo, Yaping Liu, "A Survey on Emerging Computing Paradigms for Big Data," *Chinese Journal of Electronics*, vol. 26, no. 1, pp. 1-12, 2017.

[7] Nrusimham Ammu, Mohd Irfanuddin, "Big Data Challenges," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 2, pp. 613-615, 2013.

[8] Sanjay Ghemawat, Howard Gobioff, Shun-Tak Leung, "The Google File System," in *Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles*, New York, NY, USA, Association for Computing Machinery, 2003, p. 29–43.

[9] T. White, Hadoop: The Definitive Guide, O'Reilly, 2012.

[10] Ali Davoudian, Liu Chen, Mengchi Liu, "A Survey on NoSQL Stores," *ACM Computing Surveys (CSUR)*, vol. 51, no. 2, pp. 1-43, 2018.

[11] Jeffrey Dean, Sanjay Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107-113, 2008.

[12] Chaowei Yang, Qunying Huang, Zhenlong Li, Kai Liu, Fei Hu, "Big Data and Cloud Computing: Innovation Opportunities and Challenges," *International Journal of Digital Earth*, vol. 10, no. 1, pp. 13-53, 2017.

[13] "Analytics on AWS," [Online]. Available: https://aws.amazon.com/big-data/datalakes-and-analytics/. [Accessed 16 12 2021].

[14] "Introduction: A Bit of OpenStack History," [Online]. Available: https://docs.openstack.org/project-team-guide/introduction.html. [Accessed 16 12 2021].

[15] "ISO/IEC 25010," [Online]. Available: https://iso25000.com/index.php/en/iso-25000-standards/iso-25010. [Accessed 25 3 2022].

[16] R. E. Al-Qutaish, "Quality Models in Software Engineering Literature: An Analytical and Comparative Study," *Journal of American Science*, vol. 6, no. 3, 2010.

[17] Theo Lynn, Graham Hunt, David Corcoran, John P Morrison, Philip D Healy, "A Comparative Study of Current Open-source Infrastructure as a Service Frameworks," in *CLOSER*, 2015.

[18] "Eucalyptus Documentation," 06 08 2021. [Online]. Available: https:/ /docs. Eucalyptus. cloud/eucalyptus/5/.

[19] "Official Documentation for Eucalyptus Cloud," 06 08 2021. [Online]. Available: https://docs.eucalyptus.cloud/eucalyptus/4.4.5/index.html.

[20] "Welcome to Apache Cloud Stack's Documentation," 06 08 2021. [Online]. Available: http://docs.cloudstack.apache.org/en/latest/index.html.

[21] " OpenStack Documentation," 08 06 2021. [Online]. Available: https://docs.openstack.org/yoga/

[22] Isaac Odun-Ayo, Olasupo Ajayi, Sanjay Misra, "Cloud Computing and Open Source Software:Issues and Developments," in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 2018.

[23] Siddharth Jain, Rakesh Kumar, Sunil Kumar Jangir Anamika, "A Comparative Study for Cloud Computing Platform on Open Source Software," *ABHIYANTRIKI: An International Journal of Engineering & Technology (AIJET)*, vol. 1, no. 2, 2014.

[24] Brummett, Travis and Sheinidashtegol, Pezhman and Sarkar, Debadrita and Galloway, Michael, "Performance Metrics of Local Cloud Computing Architectures," in *2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing*, 2015.

[25] Mullerikkal, Jaison Paul and Sastri, Yedhu, "A Comparative Study of OpenStack and CloudStack," in *2015 Fifth International Conference on Advances in Computing and Communications (ICACC)*, 2015.

[26] Vahid Amiry, Shayan Zamani Rad, Mohammad Kazem Akbari, Morteza Sargolzai Javan, "Implementing Hadoop Platform on Eucalyptus Cloud Infrastructure," in *2012 Seventh International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, 2012.

[27] "Welcome to Sahara!," 19 10 2018. [Online]. Available: https://docs.openstack.org/sahara/latest/. [Accessed 16 12 2021].

[28] Mumtaz Al-Mukhtar, Asraa Abdulrazak Ali Mardan, "Performance Evaluation of Private Clouds Eucalyptus versus CloudStack," International Journal of Advanced Computer Science and Applications, vol. 5, 2014.