علوم و مهندسی کامپیوتر
نشریه علمی پژوهشی انجمن کامپیوتر ایران
مجلد ۱۲، شماره ۳ (الف)، ۱۳۹۳
صفحات ۱۶-۲۹
مقاله عادی

ارائه ابزار شبیه‌ساز ابر داده‌ای بر پایه کلادسیم با قابلیت مدیریت داده و کپی

سعیده مهری رضا انتظاری ملکی علی موقر

دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف، تهران، ایران

چکیده

مضاعف‌سازی داده یکی از مهم‌ترین مکانیزم‌های مدیریت داده در سیستم‌های توزیع شده است. از آنجا که ارزیابی کارایی سیاست‌های مربوط به آن و نیز تخصیص منابع در محیط واقعی محاسبات ابر، به علت محدودیت‌های موجود عملاً غیرممکن است، برای این منظور از ابزارهای شبیه‌ساز استفاده می‌گردد. کلادسیم (Cloud Sim) چارچوبی کلی و توسعه‌پذیر برای مدل‌سازی، شبیه‌سازی و تست خدمات کاربردی و زیرساخت محاسبات ابری می‌باشد. اما در کلادسیم برای شبیه‌سازی ابر داده‌ای چندین مشکل اساسی وجود دارد: ۱) نبود قابلیت اجرای درخواست‌های مبتنی بر داده و نحوه تخصیص منبع به آن‌ها ۲) نبود امکان مدیریت داده و کپی ۳) نبود قابلیت پرس‌وجو در مورد تعداد و مکان کپی‌ها، هزینه‌ها، میزان دسترس‌پذیری و موارد مشابه برای ارائه الگوریتم‌های نوین مضاعف‌سازی داده. از طرف دیگر، ابزار شبیه‌ساز دیگری برای شبیه‌سازی ابر داده‌ای وجود ندارد. با هدف حل سه مشکل مطرح شده، بستر کلادسیم برای شبیه‌سازی ابرداده‌ای بدون تغییر درلایه‌های معماری کلادسیم و صرفاً با توسعه موجودیت‌های موجود در معماری آن، گسترش داده شده و ابزار جدیدی برای شبیه‌سازی ابرداده‌ای تحت عنوان Data Cloud Sim ارائه شده است که امکان توسعه سیاست‌های مدیریت کپی و مضاعف‌سازی داده به صورت آگاهانه‌تر با هدف بهبود کیفیت سرویس و کاهش هزینه‌ها فراهم می‌شود.

کلمات کلیدی: ابر داده‌ای، مرکز داده، مضاعف‌سازی داده، مدیریت داده و کپی، شبیه‌ساز کلادسیم.

۱- مقدمه

محاسبات ابر[۱]، سیستم توزیع شده‌ای است که خدمات زیرساخت، بستر و نرم‌افزار را به مشتریان ارائه می‌دهد. نحوه محاسبه هزینه پرداختی مشتریان برای خدمات ارائه شده از مدل پرداخت مبتنی بر استفاده[۲] پیروی می‌کند. خدمات ابر در سه لایه زیرساخت به عنوان سرویس[۳]، بستر به عنوان سرویس[۴] و نرم‌افزار به عنوان سرویس[۵] ارائه می‌گردد [۱]. به عبارت دیگر محاسبات ابر، مدلی برای دسترسی راحت و بر حسب تقاضا به منابع محاسباتی از طریق شبکه است. در محاسبات ابر تخصیص یا آزاد کردن منابع با کمترین تلاش مدیر یا فراهم‌کننده سرویس امکان‌پذیر است [۲].

بیشتر خدمات ابری مثل شبکه‌های اجتماعی، میزبانی وب، تحویل محتوا و پردازش داده بلادرنگ، دارای ترکیب، پیکربندی و نیازمندی‌های توسعه متفاوتی هستند. ارزیابی کارایی سیاست‌های زمان‌بندی، تخصیص منابع و مدیریت داده در محیط واقعی محاسبات ابر برای مدل‌های کاربردی مختلف، بسیار چالش برانگیز است. استفاده از زیرساخت‌های واقعی ابر برای ارزیابی کارایی و تکرار آزمون به علت محدودیت‌های موجود عملاً غیرممکن است. بنابراین از ابزارهای شبیه‌ساز برای این منظور استفاده می‌گردد. ابزارهای شبیه‌ساز متفاوتی در محیط‌های توزیع شده مانند گرید ارائه شده‌اند که از آن جمله می‌توان به گریدسیم [۳]، سیم‌گرید [۴] و گنگ‌سیم [۵] اشاره کرد.

با وجود تشابه‌های زیاد بین دو محیط گرید و ابر، هیچ کدام از این شبیه‌سازها را به صورت مستقیم برای مدل‌سازی و شبیه‌سازی محیط محاسبات ابری نمی‌توان بکار برد. برای شبیه‌سازی محیط ابر، ابزاری به نام کلادسیم (Cloud Sim) [۱] ارائه شده که در واقع چارچوبی کلی و توسعه‌پذیر برای مدل‌سازی، شبیه‌سازی و آزمون خدمات کاربردی و زیرساخت محاسبات ابری است. با استفاده از کلادسیم، محققان و توسعه‌گران می‌توانند کارایی برنامه‌های کاربردی جدید خود را در یک محیط کنترل شده با صرفه‌جویی در زمان، تست و ارزیابی کنند.

با وجود استفاده روزافزون از مراکز داده ابری، چالش‌های زیادی در رابطه با مدیریت داده و کپی برای بهبود کارایی دسترسی مطرح می‌شود. مضاعف‌سازی

داده[۶] (یا کپی‌سازی داده) یکی از مهم‌ترین مکانیزم‌های مدیریت داده در سیستم‌های توزیع شده است که با ایجاد کپی‌های متعدد از داده‌ها و توزیع آن‌ها در شبکه، کارایی دسترسی به داده را بهبود می‌بخشد. مکانیزم مضاعف‌سازی داده در سیستم‌های تجاری ابر مثل سیستم ذخیره‌سازی ساده آمازون[۷] [۶]، سیستم فایل گوگل[۸] [۷] و سیستم فایل توزیع شده هدوپ[۹] [۸] استفاده می‌شود. در این سیستم‌های توزیع شده، به صورت پیش‌فرض الگوریتم ایستای سه کپی به ازای هر داده، برای مضاعف‌سازی داده بکار برده می‌شود. ارائه الگوریتم‌های نوین برای مضاعف‌سازی داده در مراکز داده ابری با هدف بهبود کارایی و کاهش هزینه‌های ذخیره‌سازی و ارزیابی کارایی آن‌ها، موضوع چالش برانگیزی است. مسائل مهمی که در مضاعف‌سازی داده مطرح می‌شود عبارتند از زمان ایجاد کپی جدید از هر داده، تعداد کپی، مکان ذخیره کپی، انتخاب داده مناسب برای کپی کردن از آن و جایگزینی کپی به هنگام نبود فضای کافی برای ذخیره‌سازی. اما کلادسیم امکان شبیه‌سازی ابر داده‌ای[۱۰] و ارزیابی کارایی الگوریتم‌های مضاعف‌سازی داده را فراهم نمی‌آورد.

با وجود چالش‌های مطرح شده، سه مسئله اساسی برای شبیه‌سازی الگوریتم‌های مضاعف‌سازی داده در کلادسیم وجود دارد. مسئله اول این است که کلادسیم قابلیت اجرای درخواست‌های مبتنی بر داده- درخواست‌هایی که برای اجرا نیاز به داده دارند -را ندارد. از طرف دیگر نحوه تخصیص منبع به درخواست های مبتنی بر داده موضوعی چالش برانگیز برای محققان است که در کلادسیم امکان شبیه‌سازی الگوریتم‌های تخصیص منبع برای درخواست‌های مبتنی بر داده فراهم نشده است. مسئله دوم این است که برای مدیریت داده و کپی نیاز به دو موجودیت مدیر کپی[۱۱] برای انجام عملیات مضاعف سازی و کاتالوگ کپی[۱۲] برای نگهداری اطلاعات کپی‌ها وجود دارد که در کلادسیم چنین قابلیتی نیز پیاده‌سازی نشده است. مسئله سوم نیز این می‌باشد که برای پاسخگویی به مسائل مطرح شده در مضاعف‌سازی داده و ارائه الگوریتم‌های نوین مضاعف‌سازی داده نیاز به اطلاعاتی مانند تعداد و مکان کپی‌ها، میزان دسترس پذیری هر داده، هزینه ذخیره‌سازی و هزینه مضاعف‌سازی مراکز داده و موارد مشابه است که چنین امکانی در کلادسیم وجود ندارد.

بنابراین در این مقاله بستر کلادسیم با هدف حل هر سه مسئله اساسیِ مطرح شده، گسترش داده شده تا بستر لازم برای مدیریت آگاهانه داده و کپی در مراکز داده‌ی ابری فراهم گردد. ابزار ارائه شده قابلیت اجرای درخواست‌های مبتنی بر داده را داشته و همچنین بستری برای پیاده‌سازی سیاست‌های تخصیص منبع برای درخواست‌های مبتنی بر داده را فراهم می‌آورد. از طریق ایجاد دو نقش جدید مدیر کپی و کاتالوگ کپی در موجودیت‌های معماری کلادسیم و توسعه آن علاوه بر فراهم نمودن امکان مدیریت داده و کپی، اطلاعات مفیدی در مورد تعداد و مکان کپی‌ها، میزان دسترس‌پذیری هر داده، هزینه ذخیره‌سازی و هزینه مضاعف‌سازی داده‌های ذخیره شده در هر مرکز داده ارائه می‌شود. با استفاده از این اطلاعات، امکان ارزیابی برنامه‌های کاربردی ابر که از مدل پرداخت مبتنی بر استفاده پیروی می‌کنند، در ابزار ارائه شده فراهم گردیده است. هم‌چنین ابزار ارائه شده این امکان را فراهم می‌کند تا محققان بتوانند الگوریتم‌های مضاعف‌سازی نوین را پیاده‌سازی، تست و ارزیابی کنند.

ساختار این مقاله به این صورت می‌باشد که در بخش ۲ و در قسمت کارهای مرتبط، ابتدا شبیه‌سازهای ارائه شده برای سیستم توزیع شده گرید معرفی و سپس شبیه‌سازهای ارائه شده در محیط ابر و ابزارهای توسعه‌یافته بر پایه کلادسیم معرفی و موارد استفاده هر یک بیان شده است. در قسمت دوم این بخش، دانش لازم برای آشنایی با کلادسیم، معماری و نحوه شبیه‌سازی در آن، بیان شده است. بخش ۳، ابزار جدید ارائه شده بر پایه کلادسیم برای شبیه‌سازی ابر داده‌ای را معرفی و معماری آن را توضیح داده است. هم‌چنین قابلیت‌های جدید ارائه شده در این ابزار و نحوه عملکرد آن در این بخش آورده شده است. در بخش ۴، ابزار جدید ارائه شده بر پایه کلادسیم، در سناریوهای مختلف با استفاده از مجموعه داده مصنوعی و واقعی مورد ارزیابی و تحلیل قرار گرفته و در بخش ۵، نتیجه‌گیری و کارهای آتی برای توسعه این ابزار بیان شده است.

۲- پیش‌زمینه

۲-۱- کارهای مرتبط

سیر تکاملی سیستم‌های محاسباتی از محاسبات خوشه‌ای[۱۳]، محاسبات گرید[۱۴] به سمت محاسبات ابر پیش رفته است. برای توسعه و تست اجزای جدید و ارزیابی سیاست‌های زمان‌بندی پیشنهادی در گرید، چندین شبیه‌ساز مثل سیم‌گرید [۴]، گریدسیم [۳] و گنگ‌سیم [۵] ارائه شده است. سیم‌گرید ابزاری مبتنی بر زبان c برای شبیه‌سازی کاربردهای توزیع شده در بستر گرید است که امکان مدل‌سازی منابع زمان اشتراکی[۱۵]، با اعمال بارکاری ثابت یا بارکاری واقعی را فراهم می‌کند.

با استفاده از سیم گرید امکان شبیه‌سازی تخصیص منابع به وظایف براساس سیاست زمان‌بندی تعریف‌شده وجود دارد. از طرف دیگر، گرید سیم ابزار شبیه‌ساز مبتنی بر رخداد برای منابع ناهمگن گرید است که با زبان جاوا پیاده‌سازی شده است و قابلیت مدل‌سازی موجودیت‌های گرید، کاربران، کاربردها، واسط‌های منابع و زمان‌بندها را دارد. هم‌چنین امکان مدل‌سازی منابع در هر دو حالت زمان اشتراکی و فضا اشتراکی[۱۶] را فراهم می‌کند. گنگ‌سیم ابزار شبیه‌ساز دیگری برای مدل‌سازی سازمان‌ها و منابع مجازی مبتنی بر گرید است که تحلیل سیاست‌های مختلف زمان‌بندی را پشتیبانی می‌کند. اما هیچ یک از این شبیه‌سازها قابلیت شبیه‌سازی زیرساخت و نیازمندی‌های سطح کاربردی ابر را ندارند.

کلادسیم [۱] ابزاری برای مدل‌سازی رفتاری اجزای ابری مثل مراکز داده، ماشین‌های مجازی و سیاست‌های تأمین منابع است که به زبان جاوا پیاده‌سازی شده است. سیاست‌های تأمین منابع در کلادسیم عمومی، ساده و قابل توسعه است. هم‌چنین امکان توسعه موجودیت‌های کلادسیم برای افزودن قابلیت‌های بیشتر درشبیه‌سازی محیط‌های ابری توسط محققان و توسعه‌گران وجود دارد. در بخش ۲، کلادسیم، معماری آن و نحوه شبیه‌سازی در آن به تفصیل آمده است. کلادآنالیست (Cloud Analyst) [۹] ابزار شبیه‌سازی و مدل‌سازی دیداری بر پایه کلادسیم برای کاربردهای ابر در مقیاس بزرگ است. کلادآنالیست بر بالای کلادسیم ساخته شده و یک واسط کاربر گرافیکی به آن اضافه شده است. یکی از اهداف اصلی کلادآنالیست، جدا کردن بخش آزمون‌های شبیه‌سازی از بخش برنامه‌نویسی است؛ بنابراین یک مدل‌ساز می‌تواند بدون صرف وقت زیاد برای برنامه‌نویسی، بر روی پیچیدگی‌های شبیه‌سازی تمرکز کند.

از خصوصیات اصلی کلادآنالیست داشتن واسط کاربر گرافیکی است که به آسانی و به سرعت می‌توان آزمون‌های شبیه‌سازی را تنظیم نمود و با ذخیره سناریوهای شبیه‌سازی در قالب xml، آزمون‌های شبیه‌سازی را به تعداد زیاد تکرار نمود. خصوصیت دیگر کلادآنالیست امکان ایجاد خروجی گرافیکی در قالب نمودار و جدول است که خلاصه‌ای از مقادیر زیاد نتایج آماری در حین شبیه‌سازی را نمایش می‌دهد. خصوصیت اصلی دیگر کلادآنالیست، قابلیت تعریف شبیه‌سازی با درجه بالایی از انعطاف‌پذیری و قابل پیکربندی بودن آن است. پیکربندی مراکز داده، ماشین‌های مجازی، سیاست‌های تخصیص منابع (زمان اشتراکی و فضا اشتراکی)، کاربران و توزیع جغرافیایی آن‌ها، پارامترهای متغیر اینترنت شبیه تأخیر شبکه و پهنای باند، زمان شبیه‌سازی و سیاست‌های واسط مراکز داده از جمله تنظیمات موجود در کلادآنالیست است.

گرین‌کلاد (Green Cloud) [۱۰] یک شبیه‌ساز شبکه در سطح بسته[۱۷] برای مراکز داده آگاه از انرژی[۱۸]، با تمرکز بر ارتباطات ابری است. گرین‌کلاد برای گرفتن اطلاعات جزئی در مورد انرژی مصرفی اجزای مراکز داده (سرورها، سوئیچ‌ها و پیوند[۱۹]ها) طراحی گردیده است. گرین‌کلاد بر پایه شبیه‌ساز شبکه NS2، پروتکل TCP/IP را به صورت کامل پیاده‌سازی کرده است و امکان شبیه‌سازی پروتکل‌های دیگر را نیز فراهم می‌آورد. تنها نقطه‌ضعف گرین‌کلاد این است که مناسب مقیاس کوچکی از مراکز داده است زیرا که زمان و حافظه‌ی زیادی برای شبیه‌سازی لازم دارد. نتورک‌کلادسیم (Network Cloud Sim) [۱۱] گسترشی از کلادسیم، برای مدل‌سازی مراکز داده ابر واقعی و کاربردهای تعمیم‌یافته مثل محاسبات با کارایی بالا[۲۰]، تجارت الکترونیکی و گردش‌کاری است.

شبیه‌سازی کاربردهای موازی و توزیع شده در شبیه‌سازهای دیگر به صورت کامل در نظر گرفته نشده است. مدل‌های کاربرد در محاسبات ابری از کاربردهای وب چند لایه‌ای مثل تجارت الکترونیکی تا کاربردهای علمی می‌تواند متغیر باشد. برخی از این کاربردها از چندین وظیفه تشکیل یافته است که نیاز به ارتباط با یکدیگر دارند. برای شبیه‌سازی رفتار چنین کاربردهای پیچیده موازی و توزیع شده در نتورک‌کلادسیم، ساختار و توابع جدیدی به کلادسیم اضافه‌شده است. خصوصیت مهم دیگر نتورک‌کلادسیم، امکان مدل‌سازی کامل همبندی[۲۱] واقعی در داخل مراکز داده بین ماشین مجازی، ماشین میزبان و سوئیچ است. در کلادسیم، صرفاً همبندی بین مراکز داده در نظر گرفته شده است. برای ارزیابی دقیق کاربردهایی با ارتباطات بین وظایف و مهاجرت ماشین‌های مجازی، در نظر گرفتن نحوه همبندی شبکه، تأخیر و پهنای باند بسیار مهم و تأثیرگذار است.

اموسیم (EMUSIM) [۱۲] محیط شبیه‌سازی برای مدل‌سازی، ارزیابی و اعتبارسنجی کارایی کاربردهای محاسبات ابری است که ترکیبی از نمونه‌سازی و شبیه‌سازی است. اموسیم بر روی دو بستر ساخته‌شده است: چارچوب نمونه‌سازی خودکار[۲۲] برای نمونه‌سازی و کلادسیم برای شبیه‌سازی. اموسیم، از طریق نمونه‌سازی، اطلاعاتی خودکار، از رفتار کاربردها به دست می‌آورد و از این اطلاعات برای تولید مدل شبیه‌سازی استفاده می‌نماید تا تأثیر تعداد منابع مختلف و الگوی درخواستی کاربردهای ابری را ارزیابی کند. ام‌دی‌سی‌سیم (MDC Sim) [۱۳] شبیه‌ساز رخداد گسسته تجاری است که در دانشگاه پنسیلوانیا توسعه داده شده است. ام‌دی‌سی‌سیم، امکان مدل‌سازی خصوصیات سخت‌افزاری اجزای مختلف مرکز داده مانند سرورها، پیوندهای ارتباطی و سوئیچ‌ها از تأمین‌کنندگان مختلف و هم‌چنین تخمین انرژی مصرفی آن‌ها را فراهم می‌کند. این ابزار دارای سربار شبیه‌سازی کمی بوده و علاوه بر این، همبندی مراکز داده را به شکل گراف جهت‌دار نگهداری می‌کند [۱۴].

کلادریپورت (Cloud Reports) [۱۵] ابزار گرافیکی برای شبیه‌سازی محیط‌های توزیع شده محاسبات ابری است که از کلادسیم برای شبیه‌سازی استفاده می‌کند. کلادریپورت دارای واسط کاربر گرافیکی است که استفاده از آن را آسان می‌سازد. همچنین این ابزار امکان ایجاد گزارش به قالب Html و فایل‌های داده که با استفاده از نرم‌افزارهای دیگر مثل متلب قابل استفاده است را فراهم می‌کند. کلادریپورت فراهم‌کننده‌ی لایه زیرساخت به عنوان سرویس را با تعداد دلخواه مراکز داده و پیکربندی سفارشی، شبیه‌سازی می‌کند. علاوه بر این، مشتریان لایه زیرساخت به عنوان سرویس نیز به صورت سفارشی در این ابزار قابل شبیه‌سازی هستند. ورک‌فلوسیم (Work Flow Sims) گسترش‌یافته کلادسیم است [۱۶] که آماده‌سازی و اجرای گردش‌کار را از طریق پیاده‌سازی پشته‌ای تجزیه کننده گردش‌کار، موتور گردش‌کار و زمان‌بند وظیفه، فراهم می‌کند. داینامیک کلادسیم (Dynamic Cloud Sim) [۱۷] نیز گسترش یافته کلادسیم است که امکان مدل ۱) ناهمگنی در کارایی منابع محاسباتی، ۲) عدم قطعیت در تغییرات متغیر کارایی ماشین‌های مجازی و ۳) ماشین‌های سرگردان و شکست‌ها در طول اجرا، را فراهم می‌کند.

ریل‌کلادسیم (Real Cloud Sim) [۱۸]، شبیه‌ساز تخصیص ماشین‌های مجازی بر پایه موتور اصلی در کلادسیم است که از موتور خود نیز برای شبیه‌سازی تخصیص استفاده می‌کند که بر پایه الگوریتم ژنتیک، برنامه‌ریزی مختلط صحیح با نرم‌افزار Lingo و شبیه‌ساز شبکه NS2 است. هم‌چنین واسط گرافیکی آن امکان نمایش همبندی درقالب brute را فراهم می‌کند و یک گزارش کامل به عنوان خروجی تولید می‌کند. کلاداکشن (Cloud Auction) [۱۹] نیز گسترش یافته کلادسیم است که با توسعه کتابخانه‌ای، کلادسیم را برای کار کردن با خدمات مبتنی بر حراج قادر می‌سازد. کلادمیگ اکسپرس (Cloud MIG Xpress) [۲۰] امکان مقایسه و برنامه‌ریزی فازهای مرتبط با مهاجرت سیستم‌های نرم‌افزاری به محیط ابری بستر به عنوان سرویس یا زیرساخت به عنوان سرویس را فراهم می‌کند. این ابزار همچنین امکان تبدیل خودکار مدل سیستم به مدل کلادسیم برای شبیه‌سازی مختلف در مورد توسعه ابر را دارد.

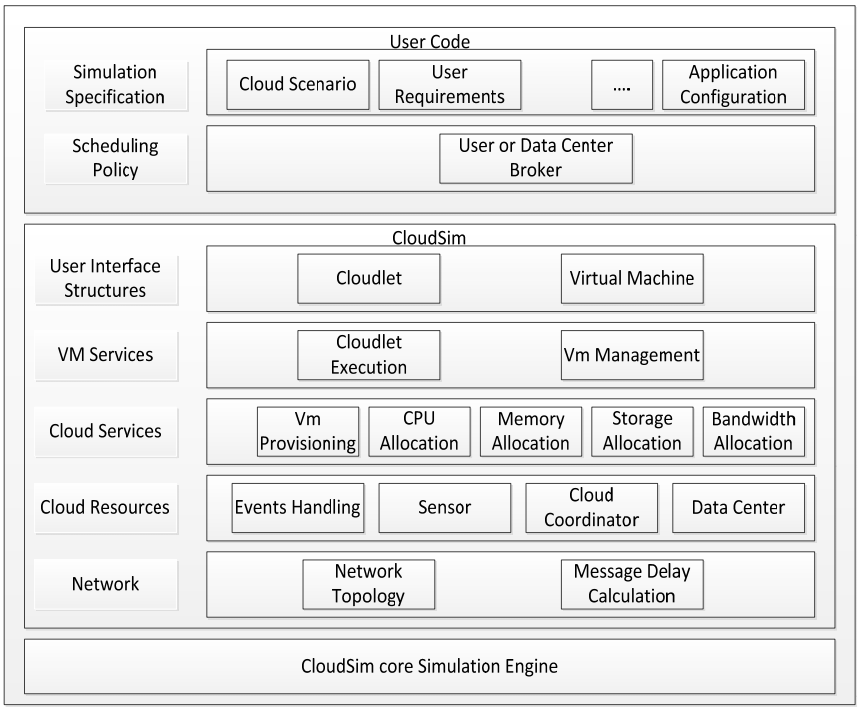
ابزارهای ارائه شده هر یک برای منظور خاصی توسعه داده شده‌اند اما برای ابر ذخیره‌ساز[۲۳] یا ابر داده‌ای که امکان مدیریت داده و کپی در آن وجود داشته باشد، ابزاری با کد باز ارائه نشده است. تنها برای شبیه‌سازی ابر ذخیره‌ساز، لانگ و همکارش [۲۱] ابزاری ارائه کرده‌اند که با اضافه کردن لایه جدید به معماری کلادسیم، قابلیت‌های تقسیم فایل به بلوک قبل از ذخیره‌سازی، ذخیره چند کپی از داده در مراکز داده متفاوت و امکان افزایش دسترس‌پذیری با مدیریت کپی را ارائه نموده است، اما به نحوه‌ی اجرای درخواست‌های مبتنی بر داده و تخصیص منابع به آن، اشاره‌ای نشده است. بر خلاف مقاله مذکور، در این مقاله، بستر کلادسیم برای شبیه‌سازی ابر داده‌ای بدون تغییر در لایه‌های معماری کلادسیم و صرفاً با توسعه موجودیت‌های موجود در معماری کلادسیم، گسترش داده شده و ابزار جدیدی برای شبیه‌سازی ابر داده‌ای ارائه شده است که قابلیت اجرای درخواست‌های مبتنی بر داده را داشته و امکان توسعه و ارزیابی سیاست‌های جدید تخصیص منبع برای درخواست‌های مبتنی بر داده را برای محققان و توسعه‌گران فراهم می‌کند.

ابزار ارائه شده علاوه بر فراهم کردن امکان ذخیره چند کپی از داده در مراکز داده مختلف و امکان پرس‌وجو از مکان کپی، قابلیت پرس‌وجوی تعداد کپی، دسترس‌پذیری و هزینه ذخیره‌سازی هر داده را نیز فراهم کرده است. با استفاده از این اطلاعات، امکان توسعه سیاست‌های مدیریت کپی و مضاعف‌سازی داده به صورت آگاهانه‌تر با هدف بهبود کیفیت سرویس و کاهش هزینه ذخیره‌سازی فراهم شده است. همچنین ابزار ارائه شده دارای قابلیت پرس‌وجو از هزینه ذخیره‌سازی و مضاعف‌سازی هر مرکز داده است که با استفاده از آن، امکان ارزیابی برنامه‌های کاربردی ابر که از مدل پرداخت مبتنی بر استفاده پیروی می‌کنند، فراهم شده است که در بخش ۳ قابلیت‌های جدید ابزار ارائه شده با جزئیات بیشتر توضیح داده شده است.

۲-۲- معرفی کلادسیم

کلادسیم [۱] چارچوبی برای مدل‌سازی و پیاده‌سازی زیرساخت و سرویس‌های ابر است که در آزمایشگاه ابر دانشگاه ملبورن ارائه شده است. کلادسیم مجموعه‌ای از کتابخانه‌ها و کلاس‌های پیاده‌سازی شده به زبان جاوا است که امکان مدل‌سازی رفتاری و سیستمی اجزای ابری مثل مراکز داده، ماشین‌های مجازی و سیاست‌های تأمین منابع را فراهم می‌کند.

شکل ۱ معماری چند لایه‌ای کلادسیم را نشان می‌دهد. بالاترین لایه پشته‌ی کلادسیم به نام کد کاربر[۲۴]، در بر گیرنده دو بخش اصلی مشخصات شبیه‌سازی و سیاست‌های زمان‌بندی است.



شکل ۱- معماری چند لایه‌ای کلادسیم [۱]

که موجودیت‌های اصلی برای میزبان‌ها (مانند تعداد ماشین‌ها و خصوصیات آن‌ها)، برنامه‌های کاربردی (مانند تعداد وظایف و نیازمندی آن‌ها)، ماشین‌های مجازی[۲۵]، تعداد کاربران، نوع برنامه کاربردی و سیاست‌های زمان‌بندی واسط را تعیین می‌کند. امکان توسعه این موجودیت‌ها برای افزودن قابلیت‌های جدید وجود دارد. لایه میانی با نام شبیه‌ساز کلادسیم قابلیت مدل‌سازی و شبیه‌سازی محیط‌های مبتنی بر مراکز داده مجازی ابر را فراهم می‌کند که شامل واسط‌های مدیریتی مشخص برای مدیریت ماشین‌های مجازی، حافظه، فضای ذخیره‌سازی و پهنای باند است.

مسائل اساسی مثل تخصیص میزبان به ماشین‌های مجازی، مدیریت اجرای برنامه کاربردی و نظارت سیستم‌های پویا توسط این لایه صورت می‌گیرد. در پایین‌ترین لایه، کلاسی به نام کلادسیم است که مسئول مدیریت صف رخدادها و کنترل گام به گام اجرای رخدادهای شبیه‌سازی است.

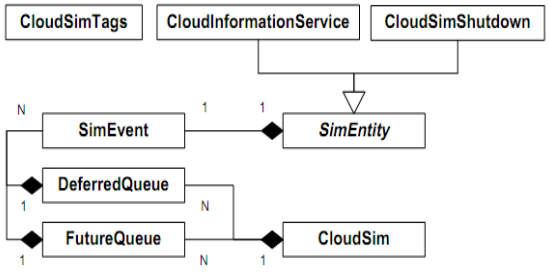
شکل ۲ نمودار کلاسی عنصرهای اصلی و هسته در کلادسیم را نشان می‌دهد. هر عنصری در کلادسیم به نام موجودیت شناخته می‌شود که از کلاس Sim Entity ارث می‌برد و دارای سه تابع Start Entity، Process Entity و Shutdown Entity است که عملیات مقداردهی، پردازش رخدادها و از بین رفتن موجودیت را انجام می‌دهند. هر موجودیتی که ایجاد می‌شود به صورت خودکار در سرویس اطلاعاتی ابر[۲۶] ثبت می‌شود. Cloud Sim Shutdown موجودیتی است که منتظر اتمام همه موجودیت‌ها است تا در پایان شبیه‌سازی به سرویس اطلاعاتی ابر اطلاع بدهد.

هر موجودیت قابلیت ارسال پیغام به موجودیت‌های دیگر و پردازش رخدادها را دارد. هر رخدادی که توسط موجودیتی در کلادسیم ایجاد می‌شود، در صفی به نام صف آینده[۲۷] ذخیره می‌گردد. رخدادها با توجه به پارامتر زمانی مرتب و در صف درج می‌شوند. در هر مرحله از شبیه‌سازی، رخدادها زمان‌بندی شده و از صف آینده حذف و به صف معوق[۲۸] افزوده می‌شوند. مشابه همین عملکرد در هر موجودیت در تابع پردازش رخداد، صورت می‌گیرد؛ رخدادها از صف معوق انتخاب و اقدام متناسب با رخداد انجام می‌شود. اقدام متناسب با رخداد از طریق برچسب موجود در آن رخداد قابل‌شناسایی است. کلاس Cloud Sim Tags در برگیرنده تعداد زیادی برچسب رخداد یا دستور است که نوع اقدام لازم برای انجام، موقع ارسال یا دریافت رخداد را تعیین می‌کند [۲۲]. مدل‌سازی در ابر با استفاده از کلادسیم بسیار گسترده و دارای جنبه‌های مختلفی است. در ادامه موجودیت‌های اصلی ابر در کلادسیم توضیح داده می‌شود:

- **مرکز داده:**

یک مرکز داده شامل تعدادی موجودیت میزبان است که هر میزبان نشان‌دهنده‌ی یک سرور فیزیکی محاسباتی در ابر است. برای هر میزبان می‌توان تنظیمات سخت‌افزاری همگن یا ناهمگن در مورد میزان حافظه، فضای ذخیره‌سازی، قدرت محاسباتی و موارد دیگر را تعیین کرد. هر میزبان می‌تواند دارای یک یا تعداد بیشتری ماشین مجازی باشد که توسط سیاست‌های تخصیص ماشین مجازی[۲۹] تعریف شده در فراهم کننده سرویس ابر، ماشین‌های مجازی به میزبان‌ها تخصیص داده می‌شود.

هر مرکز داده می‌تواند شامل شبکه‌های ذخیره‌سازی به نام SAN[۳۰] باشد که امکان شبیه‌سازی بازیابی داده را فراهم می‌کند. دسترسی به فایل‌ها در SAN دارای تأخیری اضافه بر زمان اجرای کار است که به علت انتقال داده در شبکه داخلی مرکز داده رخ می‌دهد.



شکل ۲- نمودار کلاسی از کلاس‌های اصلی در هسته کلادسیم [۱]

- **واسط ابر یا واسط مراکز داده:**

واسط ابر یا واسط مراکز داده مسئولیت مذاکرات در مورد نیازمندی‌های کیفیت سرویس مطلوب بین خدمات نرم‌افزار به عنوان سرویس و فراهم کننده ابر را بر عهده دارد. واسط با سرویس اطلاعاتی ابر در ارتباط بوده و تصمیمات بر خط[۳۱] برای تخصیص منابع و خدمات را با توجه به نیازمندی کاربران انجام می‌دهد. امکان توسعه واسط ابر برای پیاده‌سازی و ارزیابی سیاست‌های جدید برای توسعه‌دهندگان سرویس‌های ابری وجود دارد. در این مقاله واسط ابر توسعه داده شده و قابلیت مدیریت داده و کپی به آن افزوده شده است.

- **کلادلت[۳۲]:**

کلادلت مدلی برای سرویس‌های کاربردی مبتنی بر ابر مانند تحویل محتوا، شبکه‌های اجتماعی و گردش کار تجاری است. کلادلت درخواستی است که از سوی کاربر برای اجرا به واسط ابر ارسال می‌شود. محققان می‌توانند با توسعه مدل کلادلت قابلیت مدل‌سازی برنامه‌های کاربردی مبتنی بر داده را نیز فراهم کنند [۲۲] که در این مقاله کلادلت به عنوان درخواست مبتنی بر داده توسعه داده شده است.

۳- ابزار شبیه‌سازِ ابرِ داده‌ای ارائه شده (Data Cloud Sim)

با توجه به اینکه در کلادسیم قابلیت تخصیص منابع به درخواست‌های مبتنی بر داده و امکان مدیریت کپی در ابر داده‌ای وجود ندارد، ابزاری برای شبیه‌سازی ابر داده‌ای بر پایه کلادسیم ارائه شده است که قابلیت اجرای درخواست‌های مبتنی بر داده و توسعه سیاست‌های تخصیص منبع برای این درخواست‌ها را دارد. هم‌چنین مدیریت آگاهانه داده و کپی در ابر داده‌ای نیازمند اطلاعاتی در مورد تعداد و مکان کپی‌ها، میزان دسترس‌پذیری و هزینه ذخیره‌سازی هر داده است که در کلادسیم چنین امکانی فراهم نشده است. ابزار ارائه شده با فراهم نمودن امکان پرس‌وجو در مورد این اطلاعات، مدیریت آگاهانه داده و کپی را ممکن می‌سازد.

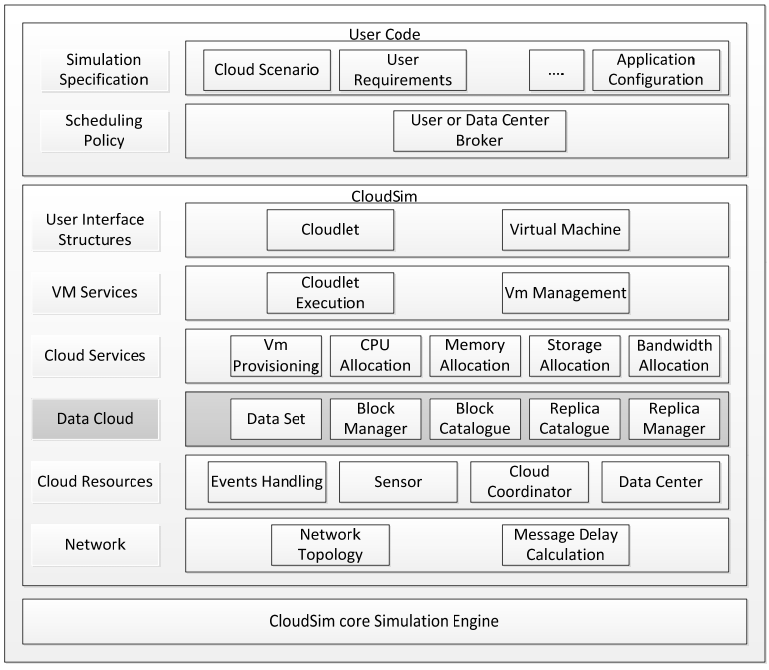
در ادامه معماری ابزار جدید ارائه شده مورد بررسی قرار گرفته و قابلیت‌های جدید ارائه شده معرفی شده است.

۳-۱- معماری ابزار شبیه‌ساز ابر داده‌ای ارائه شده

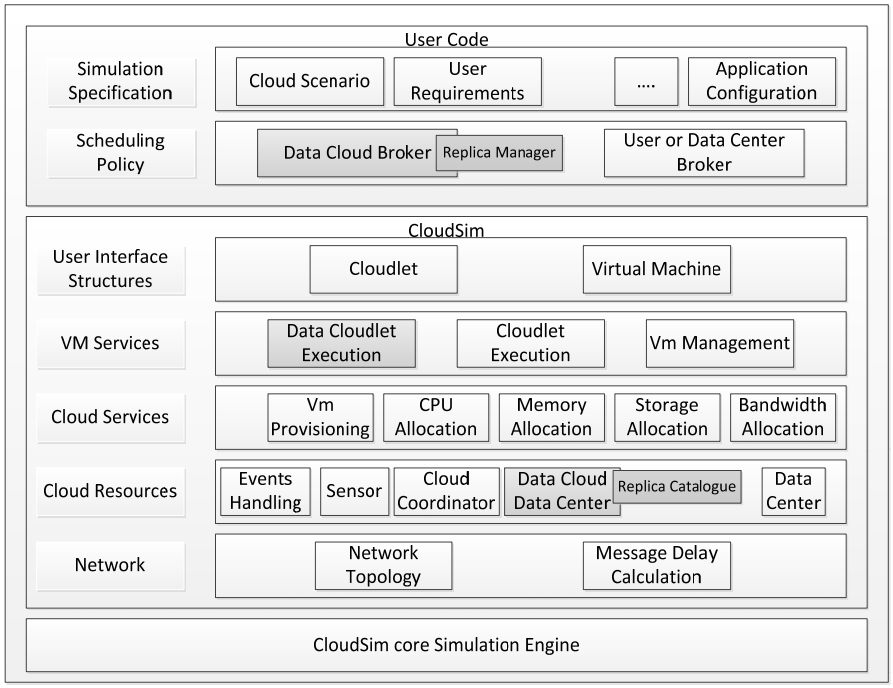
همان‌طور که در بخش ۲ اشاره شد، تنها ابزار ارائه شده برای شبیه‌سازی ابر ذخیره‌ساز [۲۱] یک لایه جدید با نام ابر داده‌ای به معماری کلادسیم اضافه کرده است. شکل ۱ معماری ابزار ارائه شده در [۲۱] بر پایه کلادسیم را نشان می‌دهد. این لایه شامل موجودیت‌های جدید مدیر کپی، کاتالوگ کپی، مدیر بلوک، کاتالوگ بلوک و مجموعه داده است. این ابزار قابلیت‌های تقسیم فایل به بلوک قبل از ذخیره‌سازی، ذخیره چند کپی از داده در مراکز داده متفاوت، امکان پرس‌وجو از مکان کپی و افزایش دسترس‌پذیری با مدیریت کپی را فراهم نموده است.

اما در این مقاله بستر کلادسیم برای شبیه‌سازی ابر داده‌ای بدون تغییر در لایه‌های معماری کلادسیم و صرفاً با توسعه موجودیت‌های موجود در معماری پایه کلادسیم، گسترش داده‌شده و ابزار جدیدی برای شبیه‌سازی ابر داده‌ای با قابلیت‌های جدید ارائه شده است. در معماری ابزار ارائه شده، دو موجودیت جدید واسط ابر داده‌ای و مرکز داده ابر داده‌ای که به ترتیب گسترش یافته موجودیت‌های واسط ابر و مرکز داده هستند، اضافه شده است. با ارسال پیغام‌های متعدد بین این دو موجودیت جدید و اضافه نمودن رخدادهای جدید، قابلیت‌های زیر به کلادسیم اضافه شده است:

- قابلیت اجرای درخواست‌های مبتنی بر داده
  - قابلیت پیاده‌سازی سیاست‌های تخصیص منبع برای درخواست‌های مبتنی بر داده
- قابلیت ذخیره چند کپی از داده در مراکز داده
- امکان پرس‌وجو در مورد تعداد کپی از هر داده و مکان کپی‌ها
- امکان پرس‌وجو در مورد میزان دسترس‌پذیری هر داده در زمان اجرا
- امکان پرس‌وجو در مورد هزینه ذخیره‌سازی هر داده و هزینه مضاعف‌سازی هر مرکز داده درزمان اجرا
- امکان جایگزینی کپی به هنگام نبود فضای کافی برای ذخیره‌سازی



شکل ۳- معماری ابزار ارائه شده برای ابر ذخیره‌ساز در [۲۱]



شکل ۴- معماری ابزار جدید شبیه‌ساز ابر داده‌ای ارائه شده بر پایه کلادسیم (Data Cloud Sim) در این مقاله با قابلیت جدید مدیریت داده و کپی

- بستری برای پیاده‌سازی سیاست‌های جدید انتخاب داده مناسب برای جایگزینی کپی
- قابلیت مدیریت کپی
- قابلیت پیاده‌سازی سیاست‌های جدید برای مدیریت کپی و مضاعف‌سازی داده

شکل ۴ معماری ابزار جدید ارائه شده بر پایه کلادسیم در این مقاله را نشان می‌دهد. موجودیت‌های جدید اضافه‌شده به معماری با رنگ متفاوت نشان داده شده است که در ادامه توضیحات بیشتری در مورد موجودیت‌های جدید و نقش آن‌ها داده شده است.

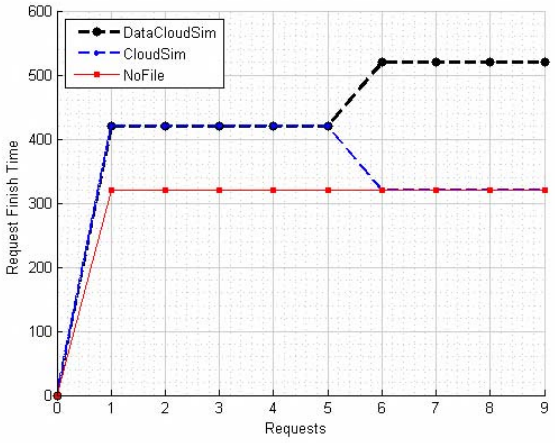
۳-۲- افزودن قابلیت جدید اجرای درخواست‌های مبتنی بر داده به کلادسیم

درخواست مبتنی بر داده، درخواستی است که برای اجرای کار نیازمند داده است. به عبارت دیگر، کلادلت برای اجرای کامل و موفق نیاز به فایلی دارد؛ اما در کلادسیم در تخصیص منبع برای کلادلت، فایل مورد نیاز در نظر گرفته نشده است و به صورت ترتیبی کلادلت را به روی ماشین‌های مجازی موجود ارسال می‌کند. برای نمایش ضعف کلادسیم در اجرای درخواست‌های مبتنی بر داده، تعدادی کلادلت که نیاز به فایل دارند برای اجرا به واسط ابر در کلادسیم فرستاده شده است. شکل ۵ زمان اتمام کلادلت‌ها در کلادسیم و ابزار ارائه شده را نشان می‌دهد. کلادلت ۱ تا ۵ نیاز به فایل ۱ با اندازه (MB) ۱۰۰۰۰ و کلادلت ۶ تا ۹ نیاز به فایل ۲ با اندازه (MB) ۲۰۰۰۰ دارند. فایل ۱ در مرکز داده ۱ و فایل ۲ در مرکز داده ۲ ذخیره شده است.

منحنی No File مربوط به حالتی است که همان کلادلت‌ها نیاز به فایل ندارند. همان طوری که در شکل ۵ دیده می‌شود زمان اتمام کلادلت‌های ۶ تا ۹ در Cloud Sim برابر با حالتی است که نیاز به فایل ندارند. علت این است که کلادلت ۶ تا ۹ نیاز به فایل ۲ داشتند که چون کلادسیم بدون توجه به فایل مورد نیاز، این کلادلت‌ها را برای اجرا به مرکز داده ۱ ارسال کرده است و مرکز داده ۱ ، فایل ۲ را ندارد؛ بنابراین بدون هیچ خطایی فرض می‌کند که کلادلت با موفقیت اجرایش به اتمام رسیده است. اما در Data Cloud Sim ابزار ارائه شده، همان طوری که انتظار می‌رفت کلادلت ۶ تا ۹ زمان اتمام بیشتری دارند زیرا که فایل ۲ دارای اندازه بیشتری نسبت به فایل ۱ است.

بنابراین در ابزار جدید، با توسعه کلاس واسط مراکز داده و ایجاد موجودیت جدید واسط ابر داده‌ایدر قسمت ارسال کلادلت‌ها به روی ماشین‌های مجازی، عملیات زیر انجام می‌شود:

- پیدا کردن فایل مورد نیاز درخواست کاربر یا کلادلت
- پیدا کردن لیست مراکز داده که کپی از فایل مورد درخواست را دارند
- انتخاب مرکز داده مناسب با استفاده از سیاست‌های تعریف شده
- انتخاب ماشین مجازی مناسب مستقر در میزبان‌های آن مرکز داده
- ارسال کلادلت به روی آن ماشین مجازی برای اجرا و برگرداندن نتایج



شکل ۵- ضعف کلادسیم در اجرای درخواست‌هایی که نیاز به فایل دارند

در صورتی که کلادلتی برای اجرا، به مرکز داده دارای آن فایل فرستاده شود، اما به هر علتی موقع رسیدن کلادلت به مرکز داده، فایل پاک شده باشد، در این

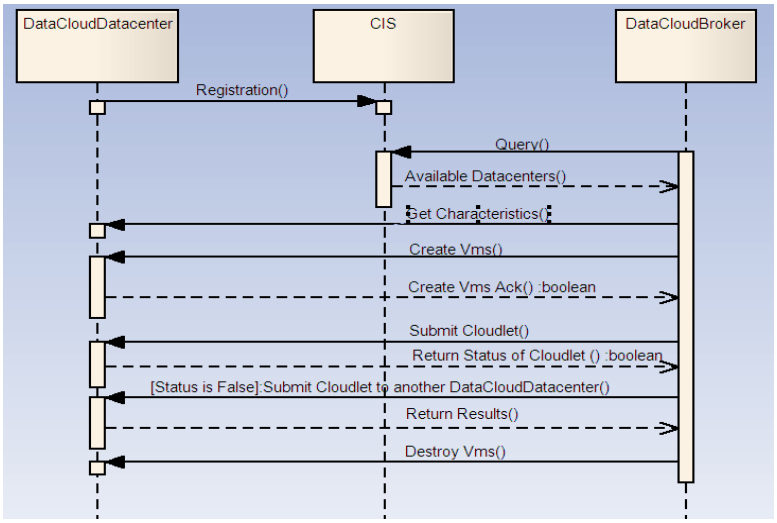
صورت مرکز داده ابر داده‌ای پیغامی به واسط ابر داده ارسال می‌کند تا واسط دوباره منبع مناسب برای آن تخصیص داده و برای اجرا ارسال نماید. اگر درخواستی فایل مورد نظر برای اجرا را نداشته باشد، وضعیت آن درخواست ناموفق خواهد بود. شکل ۶ نمودار توالی مربوط به این سناریو را نشان می‌دهد. با بررسی وضعیت تمامی درخواست‌های ارسال شده برای اجرا در نتایج، می‌توان نتیجه گرفت که عملکرد ابزار صحیح می‌باشد.

۳-۳- افزودن قابلیت جدید مدیریت داده و کپی در مراکز داده به کلادسیم

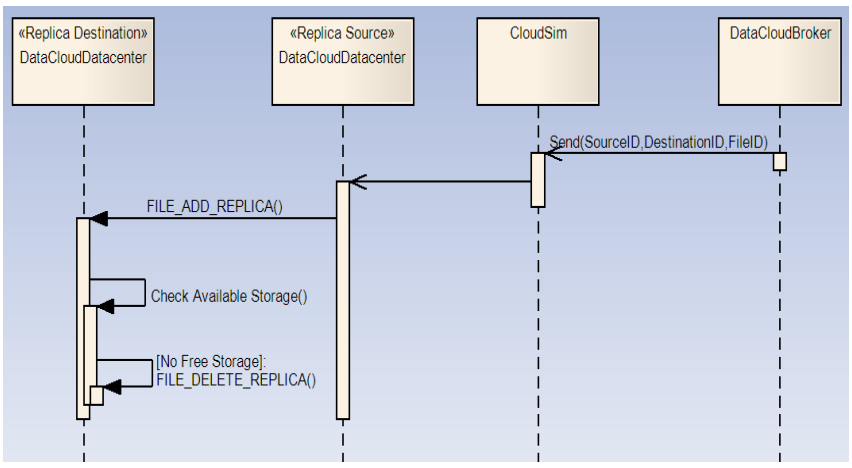
نقش جدیدی که به واسط ابر داده اضافه شده است، مدیرکپی است. در هر مرکز داده ابر، یک کاتالوگ کپی قرار داده شده است که اطلاعات مربوط به کپی‌های موجود در آن مرکز داده را نگهداری می‌کند. مدیر کپی به صورت متناوب اطلاعات موجود در کاتالوگ کپی مراکز داده ابر را گرفته و اطلاعات خود را بروز رسانی می‌نماید. واسط ابر با استفاده از اطلاعات مدیر کپی از جمله تعداد کپی از هر داده، تعداد کل کپی‌ها و میزان دسترس‌پذیری داده‌ها، می‌تواند تصمیمات مضاعف‌سازی داده را با آگاهی بیشتری انجام دهد.

برای پیاده‌سازی امکان ذخیره کپی داده در مراکز داده مختلف، نیاز به ارتباط جدید بین واسط داده و مراکز داده و همچنین ارتباط مراکز داده با یکدیگر وجود دارد. برای ارتباطات جدید از برچسب‌های تعریف‌شده در کلاس Data Cloud Tags موجود در کلادسیم استفاده شده است. در این کلاس تعداد زیادی برچسب تعریف شده است اما در کلادسیم فعلی از این برچسب‌ها برای پردازش رخدادها استفاده نشده است. برای ایجاد یک کپی از داده، واسط ابر داده، مرکز داده مبدأ و مقصد کپی را تعیین می‌کند و از موجودیت Cloud Sim می‌خواهد تا پیغامی با برچسب DataCloudTags.FILE_ADD_REPLICA همراه با کپی فایل از مرکز داده مبدأ به مرکز داده مقصد ارسال نماید. سپس در مرکز داده مقصد با دریافت این رخداد با برچسب جدید DataCloudTags.FILE_ADD_REPLICA باید اقدامی متناسب انجام شود. برای این منظور در قسمت پردازش رخدادهای مرکز داده، این برچسب و اقدام متناسب با آن تعریف شده است تا هرگاه رخدادی با این برچسب دریافت نمود، کپی فایل را در فضای ذخیره‌سازی خود قرار دهد و اطلاعات کاتالوگ کپی خود را بروزرسانی نماید.

تعداد کپی از هر داده یا انتخاب مرکز داده مناسب برای ذخیره در آن از جمله تصمیمات مضاعف‌سازی داده است که با استفاده از اطلاعات مدیر کپی و کاتالوگ کپی اتخاذ می‌شود. زمانی که کپی داده از مرکز داده مبدأ به مقصد فرستاده می‌شود، اگر در مقصد فضای ذخیره‌سازی کافی در دسترس نباشد، به صورت تصادفی داده‌ای انتخاب و با آن جایگزین می‌گردد. امکان پیاده‌سازی سیاست‌های جدید برای انتخاب جایگزین مناسب نیز وجود دارد. در شکل ۵ نمودار توالی مربوط به این سناریو نمایش داده شده است.



شکل ۶- نمودار توالی مربوط به سناریوی ارتباط واسط ابر داده و مرکز داده- زمانی که فایل موردنیاز در مرکز داده، بعد از ارسال پاک شده باشد



شکل ۷- نمودار توالی مربوط به سناریو ایجاد کپی و ارسال آن از مرکز داده مبدأ به مقصد

۳-۴- افزودن قابلیت پرس‌وجو در مورد هزینه مضاعف‌سازی مراکز داده، هزینه ذخیره‌سازی و دسترس‌پذیری داده‌ها در حین اجرا

ابزار ارائه شده این امکان را فراهم می‌کند که هزینه مضاعف‌سازی هر مرکز داده و هزینه ذخیره‌سازی و دسترس‌پذیری هر فایل ذخیره شده در مراکز داده در حین اجرا محاسبه شده تا با استفاده از این اطلاعات، واسط ابر بتواند تصمیمات مضاعف‌سازی آگاهانه تری را با هدف کاهش هزینه یا زمان اجرای درخواست‌ها اتخاذ نماید. همچنین این اطلاعات به عنوان خروجی نمایش داده می‌شود. هزینه مضاعف‌سازی، هزینه صرف شده برای پهنای باند جهت کپی داده از مرکز داده مبدأ به مقصد مورد نظر است. رابطه ۱ نحوه محاسبه هزینه مضاعف‌سازی برای یک مرکز داده را نشان می‌دهد که برابر است با حاصل‌ضرب هزینه به ازای واحد پهنای باند در مرکز داده در اندازه فایل (FS) و هزینه انتقال فایل (TC).

$$replicationCost_{Datacenter} = CostPerBW_{Datacenter} * FS * TC \quad (۱)$$

هزینه ذخیره‌سازی هر فایل برابر است با حاصل ضرب هزینه واحد ذخیره‌سازی در مرکز داده‌ای که در آن ذخیره شده است در اندازه فایل (FS) و تعداد کل کپی‌های آن فایل (NR) که از رابطه ۲ محاسبه می‌شود.

$$storageCost_{File} = CostPerStorage_{Datacenter} * FS * NR \quad (۲)$$

دسترس‌پذیری داده معیار مهم دیگری است که در تصمیم‌گیری مدیریت داده و کپی درنظر گرفته می‌شود. هر چه تعداد کپی بیشتری از داده موجود باشد، دسترس‌پذیری داده افزایش می‌یابد؛ بنابراین مصالحه‌ای بین هزینه ذخیره‌سازی و دسترس‌پذیری داده وجود دارد. رابطه ۳ نحوه محاسبه دسترس‌پذیری یک فایل را نشان می‌دهد. ρ نشان‌دهنده نرخ شکست مرکز داده‌ای است که فایل در آن ذخیره شده است و NR_f نشان‌دهنده‌ی تعداد کپی‌های موجود از فایل مورد نظر هست.

$$Avalability_{f_n} = 1 - \rho^{NR_{f_n}} \quad (۳)$$

۳-۵- اعتبارسنجی ابزار ارائه شده

دقت یک ابزار شبیه‌ساز و اعتبار نتایج آن تا حد زیادی به دقت جزئیات اجزای شبیه‌سازی شده در آن ابزار وابسته است. برای شبیه‌سازیِ رفتار زیرساخت ابر با صحت و دقت بالا، استفاده از اجزای اساسی مانند مدل کاربردی عمومی و مدل شبکه مراکز داده، در ابزارهای شبیه ساز ابر امری ضروری است [۱۱] و با توجه به اینکه ابزار ارائه شده در این مقاله بر پایه کلادسیم بوده که چارچوبی کلی و توسعه‌پذیر برای مدل‌سازی و شبیه سازی ابر است و معماری کلادسیم تغییر داده نشده است، بنابر این نتایج حاصل قابل اتکا خواهد بود.

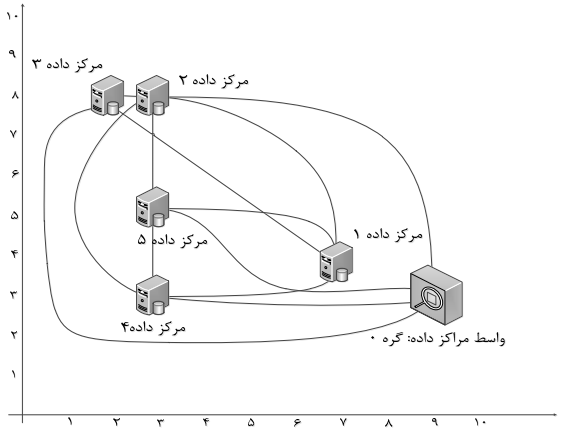
مهم‌ترین بخش اعتبارسنجی ابزار ارائه شده مربوط به اجرای درخواست‌های مبتنی بر داده است که پایه‌ای برای بقیه قابلیت‌های افزوده شده است. همان طور در شکل ۵ نیز نشان داده شده است، درخواست‌هایی که برای اجرا نیاز به فایل ندارند دارای زمان اتمام ۳۲۰ ثانیه هستند. با وجود شرایط یکسان در مراکز داده انتظار می‌رود زمان اجرای درخواست‌هایی که به فایل ۲ نیاز دارند، دو برابر زمان اجرای درخواست‌هایی باشد که به فایل ۱ نیاز دارند. منحنی Data Cloud Sim نشان می‌دهد که زمان اجرای درخواست‌های فایل ۱ به میزان ۱۰۰ ثانیه و درخواست‌های فایل ۲ به میزان ۲۰۰ ثانیه (دقیقا دو برابر حالت اول)، بیشتر از درخواست‌های بدون نیاز به فایل است. هم‌چنین برای تست عملکرد درست ابزار، الگوریتم مضاعف‌سازی ایستای سه کپی به عنوان الگوریتم محک شبیه‌سازی شده است و در هر بخش نشان داده شده است که نتایج حاصل شده صحیح می‌باشد.

۴- شبیه‌سازی و نتایج

برای بررسی نحوه عملکرد ابزار ارائه شده برای شبیه‌سازی ابر داده‌ای، محیطی با ۵ مرکز داده با همبندی نشان داده شده در شکل ۸ را در نظر می‌گیریم. همبندی نشان داده شده توسط ابزاری به نام BRITE [۲۳]، ابزار ایجادکننده همبندی شبیه همبندی اینترنت، تولید شده است که از مدل Waxman با پارامترهای پیش‌فرض آلفا برابر ۰.۱۵ و بتا برابر ۰.۲، با قراردهی تصادفی گره‌ها و توزیع ثابت پهنای باند بین پیوندها استفاده کرده است. همبندی شامل ۶ گره و ۱۱ یال است. نحوه نمایش همبندی قالب brite در دو قسمت گره و یال است که در جدول ۱ نشان داده شده است. اعداد در قسمت گره به ترتیب نشان‌دهنده‌ی شناسه گره، موقعیت x، موقعیت y، درجه ورودی، درجه خروجی، شناسه AS، نوع گره و در قسمت یال به ترتیب نشان‌دهنده‌ی شناسه یال، گره مبدأ، گره مقصد، فاصله هندسی، تأخیر پیوند، پهنای باند پیوند، AS مبدأ، AS مقصد است. این قالب قابل خواندن با کلادسیم می‌باشد.

برای اینکه انواع متفاوتی از مراکز داده در شبیه‌سازی وجود داشته باشد، دو نوع مرکز داده با نام‌های قوی و ضعیف تعریف و استفاده شده است که مشخصات هر یک از مراکز داده در جدول ۲ قابل مشاهده است. مشخصاتی مثل سیستم‌عامل[۳۳]، نام فوق ناظر[۳۴]، حافظه میزبان[۳۵] و پهنای باند بر اساس مقادیر پیش‌فرض موجود در کلادسیم تعریف شده است. مقادیر مشخصات دیگر مثل پهنای باند و تأخیر شبکه SAN با استفاده از اطلاعات موجود در [۲۴] مقدار تأخیر شبکه SAN، ۰.۱ تا ۰.۰۱ ثانیه و پهنای باند آن ۱۰۰ MB/s تعریف شده است. ۲ مرکز داده از نوع قوی و ۳ مرکز داده از نوع ضعیف در نظر گرفته شده است.

جدول ۳ تعداد ماشین‌های مجازی، مراکز داده و فایل‌های ذخیره شده در مراکز داده را نشان می‌دهد. مشخصات ماشین‌های مجازی و جدول ۵ مشخصات کلادلت‌ها را نشان می‌دهد. در ماشین مجازی برای زمان‌بندی کلادلت‌ها از الگوریتم زمان اشتراکی تعریف شده در کلادسیم استفاده شده است.



شکل ۸- همبندی مراکز داده و واسط مراکز داده

جدول ۱- همبندی استفاده شده در شبیه‌سازی

Topology: (۶ Nodes, ۱۱ Edges)
Model (۱ - RTWaxman): ۶۱۰۱۰۱۲۰.۱۵۰.۲۱۱۱۰.۰۱۰۲۴.۰
Nodes: (۶)

۰	۹	۳	۴	۴	-۱	RT_NONE
۱	۷	۴	۴	۴	-۱	RT_NONE
۲	۳	۸	۴	۴	-۱	RT_NONE
۳	۲	۸	۳	۳	-۱	RT_NONE
۴	۳	۳	۴	۴	-۱	RT_NONE
۵	۰	۵	۳	۳	-۱	RT_NONE

Edges: (۱۱)

۰	۲	۱	۵.۶۵۶۸۵۴	۰.۱	۱۰.۰	-۱	-۱	E_RT_BACKBONE U
۱	۲	۰	۷.۸۱۰۲۵	۰.۲	۱۰.۰	-۱	-۱	E_RT_BACKBONE U
۲	۳	۲	۱.۰	۰.۱	۱۰.۰	-۱	-۱	E_RT_BACKBONE U
۳	۳	۱	۶.۴۰۳۱۲۴	۰.۵	۱۰.۰	-۱	-۱	E_RT_BACKBONE U
۴	۴	۱	۴.۱۲۳۱۰۵	۰.۲	۱۰.۰	-۱	-۱	E_RT_BACKBONE U
۵	۴	۲	۵.۰	۰.۳	۱۰.۰	-۱	-۱	E_RT_BACKBONE U
۶	۵	۴	۳.۶۰۵۵۵۱	۰.۲	۱۰.۰	-۱	-۱	E_RT_BACKBONE U
۷	۵	۰	۹.۲۱۹۵۴۴	۰.۴	۱۰.۰	-۱	-۱	E_RT_BACKBONE U
۸	۰	۴	۶.۰	۰.۹	۱۰.۰	-۱	-۱	E_RT_BACKBONE U
۹	۰	۳	۸.۶۰۲۳۲۵	۰.۸	۱۰.۰	-۱	-۱	E_RT_BACKBONE U
۱۰	۱	۵	۷.۰۷۱۰۶۸	۰.۵	۱۰.۰	-۱	-۱	E_RT_BACKBONE U

جدول ۲- مشخصات مراکز داده استفاده شده در شبیه‌سازی

مرکزداده	RAM (MB)	Hosts	MIPS	Host CPU	BW (MB/s)	Host Storage (MB)	OS	VMM	SAN Delay (ms)	SAN BW (MB/s)	SAN Capacity (MB)
قوی	۲۰۴۸	۱۰	۳۰۰۰۰	quad-core	۱۰۰۰۰	۱۰۰۰۰۰	Linux	Xen	$۱۰^{-۴}$	۱۰۰	۲۰۰۰۰۰۰
ضعیف	۲۰۴۸	۱۰	۱۰۰۰	dual-core	۱۰۰۰۰	۱۰۰۰۰۰	Linux	Xen	$۱۰^{-۳}$	۱۰۰	۲۰۰۰۰۰۰

هر کلادلت فرض شده است که به یک فایل برای اجراشدن نیاز دارد که نام فایل درخواستی باید مشخص باشد. از مشخصات کلادلت، اندازه فایل[۳۶] و اندازه خروجی[۳۷] بر حسب بایت، به ترتیب نشان‌دهنده اندازه کلادلت بعد و قبل از اجراشدن است که متفاوت با فایل مورد نیاز کلادلت است و از اندازه پیش‌فرض موجود در کلادسیم ۳۰۰ بایت استفاده شده است. طول[۳۸] کلادلت بر حسب میلیون دستورالعمل در ثانیه[۳۹] نشان‌دهنده تعداد دستورالعمل لازم برای اجرا شدن است.

در تمام سناریوهای شبیه‌سازی از هزینه‌های موجود در جدول ۱ برای محاسبه هزینه ذخیره‌سازی و مضاعف‌سازی استفاده شده است. هزینه ذخیره‌سازی ذکر شده در جدول ۶ با استفاده از مقادیر پیش‌فرض موجود در کلاس‌های پایه کلادسیم تعیین شده است. مجموعه داده مورد نیاز برای آزمون عملکرد ابزار ارائه شده، شامل نحوه توزیع دسترسی به فایل‌ها و نرخ دسترسی به هر فایل است.

توزیع دسترسی به فایل‌ها در [۲۵] و [۲۶] توزیع پواسون[۴۰] فرض شده است و با توجه به این که نرخ دسترسی به فایل‌ها در صفحات اینترنتی از توزیع زیپف[۴۱] پیروی می‌کند [۲۷]، در سناریوی اول شبیه‌سازی توزیع دسترسی به فایل‌ها، پوآسون و توزیع نرخ دسترسی به فایل‌ها، زیپف فرض شده است. در سناریوی دوم شبیه‌سازی از مجموعه داده واقعی رسانه ذخیره‌سازی ویکی مدیا[۴۲] استفاده شده است.

برای تست عملکرد درست شبیه‌ساز، سه الگوریتم ایستای یک کپی، دو کپی و سه کپی به عنوان الگوریتم محک پیاده‌سازی شده است. الگوریتم ایستای سه کپی به صورت پیش‌فرض در سیستم‌های تجاری ابر مثل آمازون S3، سیستم فایل گوگل و سیستم فایل توزیع شده هدوپ استفاده می‌شود [۲۸]. در هر سه الگوریتم، ابتدا داده‌های اصل با توزیع یکنواخت بین مراکز داده پخش و ذخیره می‌شوند که کپی اصل نامیده می‌شوند. سپس از هر داده، یک یا دو کپی ایجاد و با توزیع یکنواخت بین مراکز داده توزیع می‌گردد به طوری که هیچ مرکز داده‌ای دو کپی یکسان از داده را نداشته باشد. همچنین فضای کافی برای ذخیره‌سازی کپی جدید را داشته باشد تا کپی‌های قبلی از مراکز داده پاک نگردد و حالت ایستا بودن و تعداد ثابت کپی پایدار بماند.

به عبارت بهتر در حالت یک کپی، تنها یک کپی اصل از هر داده وجود دارد و در حالت سه کپی، از هر داده یک کپی اصل و دو کپی دیگر نیز ایجاد می‌شود. هم‌چنین برای سایز فایل‌های ذخیره‌شده در مراکز داده مشابه [۲۱]، از توزیع پرتو[۴۳] استفاده شده است که در شکل ۹ نمایش داده شده است.

جدول ۳- مشخصات شبیه‌سازی

مشخصات شبیه‌سازی	
تعداد ماشین‌های مجازی	۱۴۰
تعداد فایل‌ها	۱۰
تعداد مراکز داده	۵

جدول ۴- مشخصات ماشین‌های مجازی استفاده شده در شبیه‌سازی

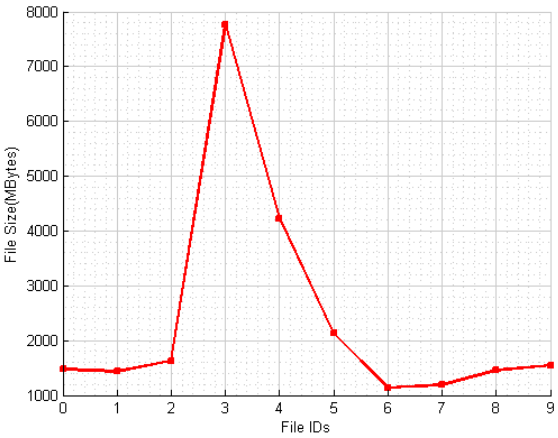
مشخصات ماشین‌های مجازی مورد استفاده						
RAM (MB)	BW (MB)	MIPS	VMM name	Number of CPUs	Cloudlet Scheduler	Size (MB)
۵۱۲	۱۰۰۰	۱۰۰۰	Xen	۱	Timeshared	۱۰۰۰۰

جدول ۵- مشخصات هر کلادلت استفاده شده در شبیه‌سازی

مشخصات کلادلت‌ها				
Output Size (Byte)	File Size (Byte)	Number of CPUs	Number of Required File	Length (MI)
۳۰۰	۳۰۰	۱	۱	۱۰۰۰

جدول ۶- هزینه‌های استفاده شده در شبیه‌سازی

هزینه‌های استفاده شده در شبیه‌سازی	
هزینه ذخیره‌سازی	۰.۱ $ PerMB
هزینه انتقال داده	۱ $ PerMB
هزینه پهنای باند	۰.۱ $ PerMB/s



شکل ۹- نمودار سایز فایل با توزیع پرتو

۴-۱- مدیریت داده و کپی

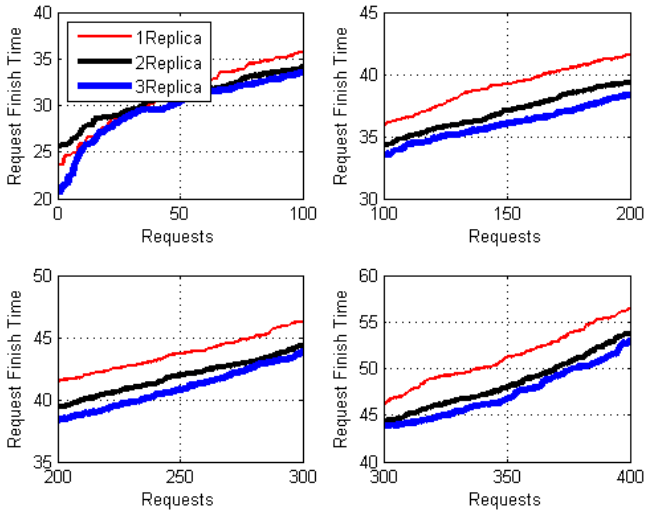
ابزار ارائه شده، اطلاعاتی در مورد تعداد دقیق کپی از هر فایل، مکان دقیق کپی‌های موجود و میزان دسترس‌پذیری هر فایل را به عنوان خروجی تولید می‌کند. به طور مثال یک خروجی از اجرای الگوریتم سه کپی در جدول ۷ نمایش داده شده است. این اطلاعات در زمان اجرا نیز قابل دسترسی است.

جدول ۷- اطلاعات مدیریت داده و کپی در الگوریتم سه کپی

شناسه فایل	سایز فایل (MB)	شناسه مراکز داده	دسترس‌پذیری	تعداد کپی‌ها
۰	۱۴۷۲	[۴و۳و۲]	۹۹.۹۹۹۹	۳
۱	۱۴۳۷	[۵و۴و۳]	۹۹.۹۹۹۹	۳
۲	۱۶۲۹	[۴و۳و۵]	۹۹.۹۹۹۹	۳
۳	۷۷۵۵	[۳و۵و۴]	۹۹.۹۹۹۹	۳
۴	۴۲۳۳	[۶و۴و۵]	۹۹.۹۹۹۹	۳
۵	۲۱۳۳	[۴و۵و۶]	۹۹.۹۹۹۹	۳
۶	۱۱۳۵	[۴و۶و۵]	۹۹.۹۹۹۹	۳
۷	۱۱۹۱	[۲و۵و۴]	۹۹.۹۹۹۹	۳
۸	۱۴۵۷	[۵و۴و۲]	۹۹.۹۹۹۹	۳
۹	۱۵۴۸	[۶و۲و۴]	۹۹.۹۹۹۹	۳

۴-۲- زمان اتمام کلادلت

انتظار می‌رود هر چه تعداد کپی بیشتری از فایل‌ها موجود باشد، زمان دسترسی به داده‌ها کاهش یافته و در نتیجه زمان اتمام کلادلت کاهش یابد. در ابتدای شبیه‌سازی، در هر سه الگوریتم ابتدا فقط یک کپی اصل از داده‌ها وجود دارد و بعد از دوره تناوب اول کپی دوم و یا سوم از داده‌ها در مراکز داده ذخیره می‌شود. همان طوری که در شکل ۱۰ دیده می‌شود، بعد از ۲۰ درخواست اول، الگوریتم ایستای سه کپی، نسبت به الگوریتم دو کپی و الگوریتم دو کپی نسبت به الگوریتم یک کپی، زمان اتمام کلادلت کمتری داشته است.برای نمایش بهتر، نمودار به ازای هر ۱۰۰ درخواست، در نمای بزرگ‌تر رسم شده است.

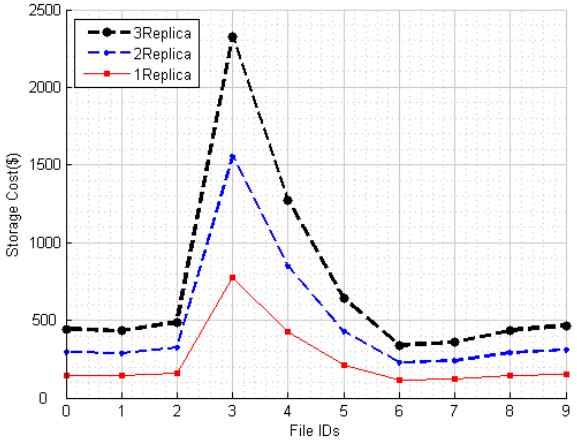


شکل ۱۰- نمودار زمان اتمام کلادلت سه الگوریتم ۱ کپی، ۲ کپی و ۳ کپی

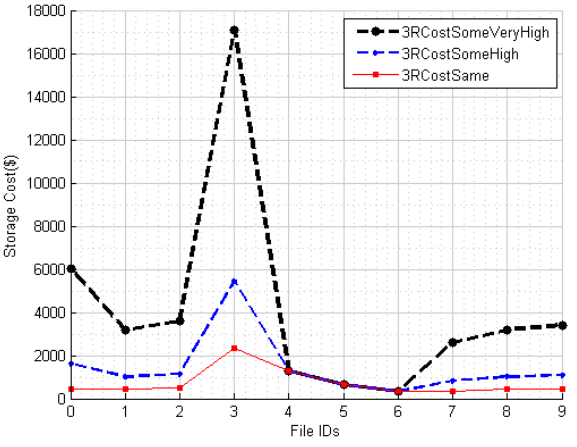
۴-۳- هزینه ذخیره‌سازی

در شکل ۱۱ نمودار هزینه ذخیره‌سازی هر فایل به ازای هر سه الگوریتم ایستای سه کپی ، دو کپی و یک کپی نشان داده‌شده است. براساس رابطه ۲ هزینه ذخیره‌سازی با اندازه فایل رابطه مستقیم دارد، هر چه فایل دارای اندازه بیشتری باشد بنابراین هزینه ذخیره‌سازی بیشتری خواهد داشت. در سناریوی بکار برده شده در شکل ۱۱، هزینه ذخیره سازی در تمام مراکز داده یکسان فرض شده است، به همین دلیل هزینه ذخیره‌سازی تنها تابعی از اندازه فایل است. همین طور که قابل پیش‌بینی بود، هر چه تعداد کپی‌ها بیشتر باشد، هزینه ذخیره‌سازی نیز بیشتر می‌شود که شکل ۱۱ نیز همین مطلب را تأیید می‌کند. دلیل دیگر بر اعتبار نتایج حاصل از ابزار ارائه شده این است که هزینه ذخیره سازی هر فایل در الگوریتم سه کپی دقیقا سه برابر هزینه آن در الگوریتم یک کپی است.

در سناریوی دیگر، هزینه ذخیره‌سازی در مراکز داده متفاوت فرض شده و هزینه ذخیره‌سازی هر فایل با الگوریتم سه کپی در سه حالت مختلف در شکل ۱۲ نشان داده شده است. 3R Cost same حالتی است که مراکز داده دارای هزینه یکسانی برای ذخیره‌سازی هستند. 3R Cost Some High حالتی است که تعدادی از مراکز داده دارای هزینه بیشتر و 3R Cost Some Very High حالتی که همان مراکز داده دارای هزینه خیلی بیشتری هستند.



شکل ۱۱- نمودار هزینه ذخیره سازی هر فایل در سه الگوریتم ۱ کپی، ۲ کپی و ۳ کپی

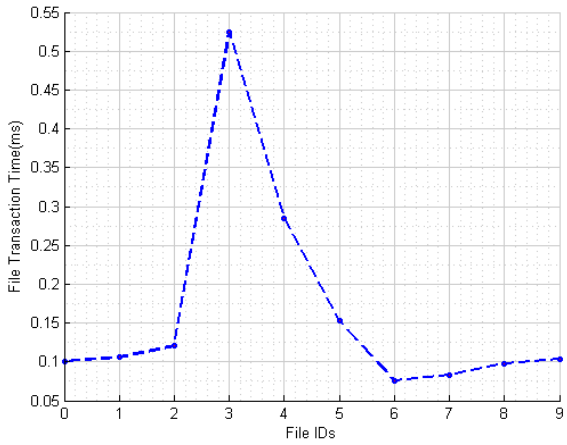


شکل ۱۲- نمودار هزینه ذخیره‌سازی هر فایل در سه الگوریتم ۱ کپی، ۲ کپی و ۳ کپی در سناریویی با هزینه ذخیره‌سازی متفاوت در مراکز داده

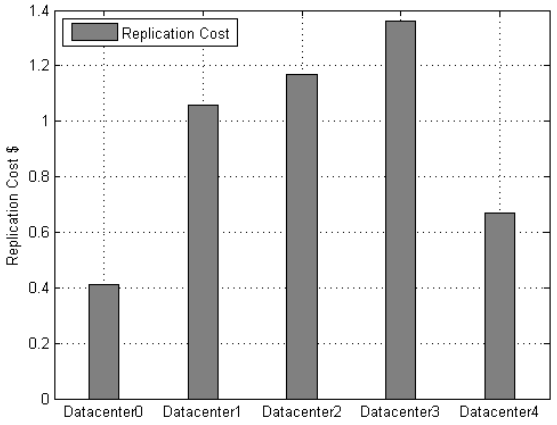
همان طوری که دیده می‌شود هزینه ذخیره‌سازی هر فایل با توجه به مکان ذخیره سازی آن و هزینه آن مرکز داده دارای هزینه ذخیره سازی متفاوتی می‌باشد. به عنوان مثال، فایل شماره ۵ در هر سه حالت در مرکز داده‌ای ذخیره شده است که در هر سه دارای هزینه یکسانی بوده است، بنابراین هزینه ذخیره‌سازی این فایل ثابت مانده است.

۴-۴- زمان تراکنش فایل

زمان تراکنش فایل با اندازه فایل ارتباط مستقیم دارد هر چه فایل دارای اندازه بیشتری باشد، دارای زمان تراکنش بیشتری خواهد بود. همان طور که در شکل ۱۳ دیده می‌شود، زمان تراکنش نیز مشابه نمودار اندازه فایل است که خود این مطلب نیز دلیل دیگری بر اعتبار عملکرد صحیح ابزار ارائه شده است.



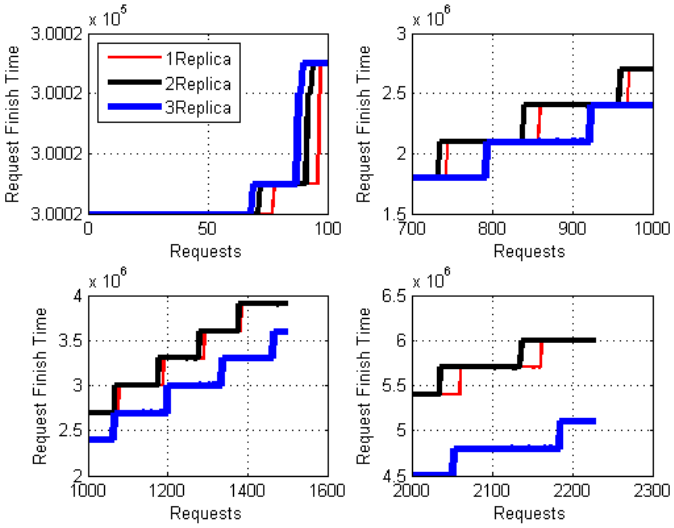
شکل ۱۳- زمان بازیابی فایل



شکل ۱۴- هزینه مضاعف‌سازی هر مرکز داده

۴-۵- هزینه مضاعف‌سازی

در شکل ۱۴ هزینه مضاعف‌سازی هر مرکز داده نشان داده‌شده است که بر حسب تعداد فایل‌ها و اندازه فایل‌های ذخیره شده در آن، هزینه هر مرکز داده متفاوت بوده است. با توجه به نمودار می‌توان فهمید که در مرکز داده سوم، فایل بیشتری یا فایل با اندازه بزرگ‌تری ذخیره شده است.



شکل ۱۵- نمودار زمان اتمام درخواست‌های رسانه ذخیره‌سازی ویکی‌مدیا، در بازه‌های تصادفی مختلف

می‌توان با استفاده از این اطلاعات، ماشین‌های مجازی بیشتری در این مرکز داده فراهم نمود تا درخواست‌ها با سرعت بیشتری پاسخ داده شوند. از موارد دیگر کاربرد اطلاعاتی مثل هزینه ذخیره‌سازی و مضاعف‌سازی در محاسبه هزینه درخواست‌های صورت گرفته برای یک برنامه کاربردی خاص است تا از مدل پرداخت مبتنی بر استفاده پیروی کند.

برای اطمینان از عملکرد درست ابزار ارائه شده، در سناریوی دیگری از مجموعه داده ترکیبی از داده‌های واقعی موجود در رسانه ذخیره‌سازی ویکی مدیا استفاده شده است. رسانه ذخیره‌سازی ویکی مدیا، امکان ذخیره و بارگیری محتواهای مختلف مثل عکس، فایل پی دی اف، فایل صوتی و فایل تصویری را از سایت upload.wikimedia.org فراهم می‌کند. Ganglia نام ابزار نظارتی سیستم توزیع شده مقیاس‌پذیر برای سیستم‌های محاسباتی مثل گرید و خوشه است. این ابزار به کاربران اجازه می‌دهد تا به اطلاعات آماری همه ماشین‌های تحت نظارت به صورت زنده یا اطلاعات قدیمی دسترسی داشته باشند. مجموعه داده مورد استفاده مربوط به آمار درخواست‌های صورت گرفته از سایت ذخیره‌سازی ویکی مدیا از ساعت ۴:۴۰ تا ۶:۴۰ تاریخ ۱۹-۱۲-۲۰۱۳ است [۲۹] که شامل اطلاعات درخواستی ۲۷ ماشین نظارت شده است. از آنجایی که درخواست‌ها در محیط شبیه‌سازی کلادسیم به واسط مراکز داده فرستاده می‌شود، متوسط درخواست‌های ۵ ماشین اول، به عنوان درخواست‌های ورودی شبیه‌سازی در نظر گرفته شده است. آماری در مورد نرخ دسترسی به فایل به تفکیک نام در این سایت در دسترس نبود، به همین علت از توزیع تصادفی برای دسترسی به فایل‌ها استفاده شده است. شکل ۱۵ نمودار زمان اتمام بیش از ۲۰۰۰ درخواست صورت گرفته از رسانه ذخیره‌سازی ویکی مدیا در بازه زمانی دو ساعت را نشان می‌دهد. از آنجایی که تعداد درخواست‌ها زیاد بود و نمایش زمان اتمام همه درخواست‌ها امکان پذیر نبود، به صورت تصادفی چهار بازه مختلف از درخواست‌ها در نمای بزرگتر در شکل ۱۵ نمایش داده شده است. همان طوری که قابل پیش بینی بود، الگوریتم سه کپی نسبت به الگوریتم دو کپی و الگوریتم دو کپی نسبت به الگوریتم یک کپی بعد از ۱۰۰ درخواست اول، زمان اتمام کمتری داشته است.

۵- نتیجه‌گیری

در این مقاله ابتدا چالش‌های موجود در ارزیابی کارایی سیاست‌های زمان‌بندی و تخصیص منابع ابری بر روی بسترهای واقعی ابر مورد بررسی قرار گرفت و سپس شبیه‌سازهای موجود برای سیستم‌های توزیع شده گرید و ابر معرفی شد. شبیه‌ساز ابر به نام کلادسیم که چارچوب کلی برای شبیه‌سازی محیط‌های ابری است با جزئیات بیشتر معرفی و نحوه کار آن نیز بیان شد. با وجود چالش‌های مطرح شده، چند مشکل اساسی برای شبیه‌سازی الگوریتم‌های مضاعف‌سازی داده در کلادسیم مطرح شد که عبارتند از نبود قابلیت اجرای درخواست‌های مبتنی بر داده و نحوه تخصیص منبع به این نوع درخواست‌ها، نبود امکان مدیریت داده و کپی و عدم ارائه اطلاعات لازم برای پاسخگویی به مسائل مطرح شده در مضاعف‌سازی داده، ارائه الگوریتم‌های نوین مضاعف‌سازی داده و ارزیابی برنامه‌های کاربردی ابر با استفاده از مدل پرداخت مبتنی بر استفاده. بنابراین در این مقاله، بستر کلادسیم با هدف حل مشکلات مطرح شده‌ی فوق گسترش داده شده تا ابزاری برای شبیه‌سازی ابر داده‌ای بر پایه کلادسیم ارائه شود تا بدین‌وسیله بستر لازم برای مدیریت آگاهانه داده و کپی در مراکز داده ابری فراهم گردد.

ابزار جدید ارائه شده دارای قابلیت اجرای درخواست‌های مبتنی بر داده، ذخیره چند کپی از داده در مراکز داده متفاوت، امکان پرس‌وجو در مورد تعداد کپی از هر داده و مکان کپی‌ها، امکان پرس‌وجو در مورد میزان دسترس‌پذیری و هزینه ذخیره‌سازی هر داده و هزینه مضاعف‌سازی هر مرکز داده است. همچنین امکان مدیریت کپی و پیاده‌سازی سیاست‌های جدید برای مدیریت کپی‌ها در ابزار جدید ارائه شده است. مضاعف‌سازی داده یکی از مهم‌ترین مکانیزم‌های مدیریت داده در سیستم‌های توزیع شده است که تصمیم‌گیری در مورد انتخاب داده مناسب برای کپی، زمان ایجاد کپی، تعداد کپی از هر داده و مکان کپی از جمله مباحثی است که به عنوان کارهای آینده می‌تواند در نظر گرفته شود که با استفاده از آن می‌توان سیاست‌های جدید مضاعف‌سازی داده در مراکز داده ابری و سیاست‌های جایگزینی کپی ارائه نموده و با استفاده از این ابزار، کارایی الگوریتم‌های پیشنهادی جدید را مورد ارزیابی قرار داده و بهبود بخشید. همچنین گسترش سیاست‌های تخصیص منابع به درخواست‌های مبتنی بر داده از مباحثی است که محققان می‌توانند سیاست‌های نوین خود را با این ابزار ارزیابی کنند.

با گسترش ابزار ارائه شده بر پایه کلادسیم می‌توان قابلیت‌های جدید دیگری مثل تقسیم فایل به بلوک، مدیریت بلوک و کپی‌ها را اضافه نمود تا نتایج واقعی‌تری از ارزیابی کارایی سیاست‌های مضاعف‌سازی داده بدست آورد. با استفاده از اطلاعاتی که این ابزار در حین اجرا به واسط مراکز داده ابری به عنوان مدیر کپی‌ها ارسال می‌نماید، می‌توان سیاست‌های مضاعف‌سازی آگاهانه‌تری پیاده‌سازی

نمود و مورد ارزیابی قرار داد. همچنین می‌توان با طراحی واسط گرافیکی کاربر استفاده از این ابزار را راحت‌تر نمود. ابزارهای دیگر ارائه شده بر پایه کلادسیم مثل کلادآنالیست و کلادریپورت را نیز می‌توان گسترش داد تا بتوان از این ابزارها هم برای تحلیل و گزارش‌گیری از مراکز داده ابری با درخواست‌های مبتنی بر داده نیز استفاده نمود.

مراجع

[1] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. D. Rose, and R. Buyya, "CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms," *Journal of Software: Practice and Experience*, vol. 41, no. 1, pp. 23-50, 2011.

[2] D. R. Malhotra, and P. Jain, "An EMUSIM Techniques and its Components in a Cloud Computing Environment," *Journal of Computer Trends and Technology*, vol. 4, no. 8, pp. 2435-2440, 2013.

[3] R. Buyya, and M. Murshed, "Gridsim: A Toolkit for the Modeling and Simulation of Distributed Resource Management and Scheduling for Grid Computing," *Journal of Concurrency and Computation Practice and Experience*, vol. 14, no. 13-15, pp. 1175-1220, 2002.

[4] A. Legrand, L. Marchal, and H. Casanova, "Scheduling Distributed Applications: The SimGrid Simulation Framework," *Proc. IEEE/ACM Intl Symp. Cluster Computing and the Grid*, pp. 138-145, 2003.

[5] CL. Dumitrescu, and I. Foster, "GangSim: A Simulator for Grid Scheduling Studies," *Proc. IEEE Intl Symp. Cluster Computing and the Grid*, pp. 1151-1158, 2005.

[6] Amazon-S3, Amazon simple storage service, http://www.amazon.com/s, November 2013.

[7] S. Ghemawat, H. Gobioff, and S. Leung, "The Google File System," *Proc. ACM Intl Symp. Operating*, pp. 29-43, 2003.

[8] The Apache Software Foundation, http://hadoop.apache.org/core, November 2013.

[9] B. Wickremasinghe, R. N. Calheiros, and R. Buyya, "CloudAnalyst: A CloudSim-based Visual Modeller for Analysing Cloud Computing Environments and Applications," *Proc. IEEE Intl Conf. Advanced Information Networking and Applications*, pp. 446-452, 2010.

[10] D. Kliazovich, P. Bouvry, and S. U. Khan, "GreenCloud: A Packet-level Simulator of Energy-aware Cloud Computing Data Centers," *Journal of Supercomputing*, vol. 62, no. 3, pp. 1263-1283, 2012.

[11] S. K. Garg, and R. Buyya, "NetworkCloudSim: Modelling Parallel Applications in Cloud Simulations," *Proc. IEEE Intl Conf. Utility and Cloud Computing*, pp. 105-113, 2011.

[12] R. N. Calheiros, M. A. Netto, C. A. De Rose, and R. Buyya, "EMUSIM: An Integrated Emulation and Simulation Environment for Modeling, Evaluation, and Validation of Performance of Cloud Computing Applications," *Journal of Software: Practice and Experience*, vol. 40, no. 4, pp. 1-18, 2012.

[13] S. H. Lim, B. Sharma, G. Nam, E. K. Kim, and C. R. Das, "MDCSim: A Multi-tier Data Center Simulation, Platform," *Proc. IEEE Intl Conf. Cluster Computing and Workshops*, pp. 1-9, 2009.

[14] R. Malhotra, and P. Jain, "Study and Comparison of CloudSim Simulators in the Cloud Computing," *IEEE Trans. Computer Science Engineering and its Applications*, vol. 1, no. 4, pp. 407-419, 2013.

[15] Cloud Reports, https://github.com/thiagotts/Cloud Reports, January 2014.

[16] Work flow Sim, https://github.com/WorkflowSim, January 2014.

[17] M. Bux, and U. Leser, "DynamicCloudSim: Simulating Heterogeneity in Computational Clouds," *Proc. ACM Intl Workshop on Scalable Workflow Execution Engines and Technologies*, pp. 1-12, 2013.

[18] Real Cloud Sim, http://sourceforge.net/projects/realcloudsim, January 2014.

[19] Cloud Simulation Tools, http://www.cloudbus.org/cloudsim, January 2014.

[20] Cloud MIG Xpress, http://sourceforge.net/projects/cloudmigxpress, January 2014.

[21] S. Long, and Y. Zhao, "A Toolkit for Modeling and Simulating Cloud Data Storage: An Extension to CloudSim," *Proc. IEEE Intl Conf. Control Engineering and Communication Technology*, pp. 597-600, 2012.

[22] R. Buyya, R. Ranjan, and R. N. Calheiros, "Modeling and Simulation of Scalable Cloud Computing Environments and the CloudSim Toolkit: Challenges and Opportunities," *Proc. IEEE Intl Conf. High Performance Computing and Simulation*, pp. 1-11, 2009.

[23] A. Medina, A. Lakhina, I. Matta, and J. Byers, "BRITE: An Approach to Universal Topology Generation," *Proc. IEEE Intl Symp. Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, pp. 346-353, 2001.

[24] SAN Bandwidth, and SAN Delay, http://www.allsan.com, January 2013.

[25] Q. Wei, B. Veeravalli, B. Gong, L. Zeng, and D. Feng, "CDRM: A Cost-effective Dynamic Replication Management Scheme for Cloud Storage Cluster," *Proc. IEEE Intl Conf. Cluster Computing*, pp. 188-196, 2010.

[26] X. Xu, S. Wang, K. Yao, and X. Zhou, "Research on the Strategy of FLDC Replication Dynamically Created in Cloud Storage," *Proc. IEEE Intl Conf. Consumer Electronics, Communications and Networks*, pp. 2815-2818, 2012.

[27] A. L. Adami, and A. B. Huberman, "Zipf's Law and the Internet," *Glottometrics Journal*, vol. 3, no. 1, pp. 143-150, 2002.

[28] K. Sashi, and A.S. Thanamani, "A New Dynamic Replication Algorithm for European Data Grid," *Proc. ACM Intl Bangalore Conf.*, pp. 1-4, 2010.

[29] Data Set, http://ganglia.wikimedia.org, January 2013.

سعیده مهری مدرک کارشناسی خود را در رشته مهندسی فناوری اطلاعات در سال ۱۳۸۹ از دانشگاه تبریز و مدرک کارشناسی ارشد مهندسی فناوری اطلاعات با گرایش شبکه‌های کامپیوتری را در سال ۱۳۹۲ از دانشگاه صنعتی شریف دریافت نموده است. پایان نامه وی در مورد الگوریتم‌های مضاعف‌سازی داده در مراکز داده ابری با هدف بهبود کارایی است. زمینه‌های پژوهشی وی در زمینه محاسبات ابری، محاسبات گرید، الگوریتم‌های مضاعف‌سازی داده، تعادل بار و تخصیص منابع با هدف بهبود کیفیت سرویس می‌باشد.

آدرس پست‌الکترونیکی ایشان عبارت است از:

mehri@ce.sharif.edu

رضا انتظاری ملکی در حال حاضر دانشجوی دکتری مهندسی کامپیوتر (گرایش نرم‌افزار) در دانشگاه صنعتی شریف می‌باشد. وی مدرک کارشناسی و کارشناسی ارشد خود را در همین رشته-گرایش درسال‌های ۱۳۸۶ و ۱۳۸۸ از دانشگاه علم و صنعت ایران دریافت نموده است. رساله دکتری وی در رابطه با مدل‌سازی و ارزیابی کارایی‌پذیری در محیط گرید محاسباتی بوده و عمده فعالیت پژوهشی وی نیز در ارتباط با مفاهیمی همچون ارزیابی کارایی و اتکاپذیری، مدل‌سازی با روش‌های ریاضی، الگوریتم‌های زمانبندی و محاسبات گرید و ابر می‌باشد.

آدرس پست‌الکترونیکی ایشان عبارت است از:

entezari@ce.sharif.edu

علی موقر استاد دانشکده مهندسی کامپیوتر دانشگاه صنعتی شریف می‌باشد. ایشان مدرک کارشناسی خود را در سال ۱۳۵۶ در رشته مهندسی برق از دانشگاه تهران، مدرک کارشناسی ارشد و دکتری خود را در سال‌های ۱۳۵۸ و ۱۳۶۴ در رشته مهندسی کامپیوتر، اطلاعات و کنترل از دانشگاه میشیگان دریافت کرده‌اند. زمینه تحقیقاتی وی شامل مدل‌سازی کارایی و اتکاپذیری، درستی‌یابی صوری شبکه‌های بی‌سیم و سیستم‌های بی‌درنگ توزیع شده می‌باشد. وی همچنین عضو ارشد IEEE و ACM می‌باشد.

آدرس پست‌الکترونیکی ایشان عبارت است از:

movaghar@sharif.edu

اطلاعات بررسی مقاله:

تاریخ ارسال: ۹۲/۱۱/۲۶

تاریخ اصلاح: ۹۳/۴/۲۲

تاریخ قبول شدن: ۹۳/۵/۳۰

نویسنده مرتبط: رضا انتظاری ملکی، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف، تهران، ایران.

[1] Cloud Computing
[2] Pay-As-You-Go Model
[3] Infrastructure As a Service (IaaS)
[4] Platform As a Service (PaaS)
[5] Software As a Service (SaaS)
[6] Data replication
[7] Amazon Simple Storage System (Amazon S3)
[8] Google File System (GFS)
[9] Hadoop Distributed File System (HDFS)
[10] Data Cloud
[11] Replica Manager
[12] Replica Catalogue
[13] Cluster Computing
[14] Grid Computing
[15] Time-Shared
[16] Space-Shared
[17] Packet-Level
[18] Energy-Aware
[19] Link
[20] High Performance Computing (HPC)
[21] Topology
[22] Automated Emulation Framework (AEF)
[23] Cloud Data Storage
[24] User Code
[25] VMs (Virtual Machines)
[26] Cloud Information Service (CIS)
[27] Future Queue
[28] Deferred Queue
[29] VM Allocation
[30] Storage Area Network (SAN)
[31] On-Line
[32] Cloudlet
[33] OS
[34] VMM
[35] RAM
[36] Size File
[37] Output Size
[38] Length
[39] MI
[40] Poisson Distribution
[41] Zapf Distribution
[42] Wikimedia's Media Storage
[43] Pareto Distribution