

مرور و مقایسه‌ای بر تکنیک‌های تعادل بار در رایانش ابری

مهدی قاسمی^۱، مائده رحمانی^۲، حمید زنگی آبادی زاده^۱، فرشید وظیفه دوست^۳ و سمیه کدخدا ده‌خانی^۳
^۱ دانشکده مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیک، دانشگاه پیام نور مرکز بین الملل کیش
^۲ دانشکده مهندسی کامپیوتر گرایش نرم افزار، دانشگاه پیام نور مرکز بین الملل کیش
^۳ دانشکده مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیک از دانشگاه پیام نور مرکز بین الملل قشم

چکیده

رایانش ابری یکی از فناوری‌های پرکاربرد در حوزه خدمات مبتنی بر فناوری اطلاعات است. این یک فناوری نوظهور در نظر گرفته می‌شود که رانش جدیدی را برای محاسبات مبتنی بر مجازی سازی منابع فراهم می‌کند. رایانش ابری از نیاز به برون‌سپاری محاسبات و امکانات ذخیره‌سازی به مشتریان برای تبدیل شدن به ضروری در فعالیت‌های روزمره پدید آمده است. مزایای رایانش ابری ارائه حداکثر خدمات با حداقل هزینه در هر زمان و غیره است. یکی از موانع اصلی استفاده از رایانش ابری، ثبات عملکرد ابری است. تعادل بار یکی از عناصری است که تأثیر آن بر پایداری عملکرد را نشان می‌دهد. تعادل بار یکی از پیچیده‌ترین موارد است. در واقع فناوری محاسبات ابری از تعداد زیادی منابع تشکیل شده است که به کاربران اجازه می‌دهد تا در صورت نیاز به منابع صحیح و ضروری در یک شبکه دسترسی داشته باشند. برنامه‌های کاربردی با نیازهای منابع مختلف به قابلیت‌های محاسباتی با کارایی بالا نیاز دارند که به حداکثر رساندن استفاده از منابع، مصرف زمان و جلوگیری از سربار پردازنده کمک می‌کند. در این مقاله، تکنیک‌های مختلف متعادل‌سازی بار که در محیط‌های ابری اعمال می‌شود، مورد بحث قرار گرفته و یک مطالعه مقایسه‌ای ارائه شده است.

کلمات کلیدی: رایانش ابری، ذخیره‌سازی و تعادل بار.

تاریخچه مقاله:

تاریخ ارسال: ۱۴۰۳/۰۶/۱۶

تاریخ اصلاحات: ۱۴۰۳/۱۱/۰۷

تاریخ پذیرش: ۱۴۰۳/۱۲/۱۴

تاریخ انتشار: ۱۴۰۳/۱۲/۲۵

ایمیل نویسنده مسئول: Mahdikmg1@gmail.com

۱ - مقدمه

استفاده از محاسبات ابری چندین مزیت دارد که امکان تهیه منابع فناوری اطلاعات بر حسب تقاضا را فراهم می‌کند. مراکز داده اغلب به دلیل پذیرش گسترده رایانش ابری توسط سازمان‌ها، بارهای غیرمعمول زیاد یا کم را تجربه می‌کنند. مدیریت حجم بالای ترافیک شبکه در فضای ابری مستلزم متعادل کردن بار دقیق سرور است. این زمانی است که تعادل بار مفید است. هدف اصلی تعادل بار این است که با توزیع یکنواخت تقاضا در بین تمام سرورهای موجود، از بارگذاری یا کم بار شدن هر یک از سرورها جلوگیری شود. هنگامی که منابع بیش از حد مورد استفاده قرار می‌گیرند، تعادل بار می‌تواند برای توزیع فشار بر روی چندین گره استفاده شود. اگر یک گره قادر به تحمل بار فعلی نباشد، متعادل‌سازی بار باید انجام شود. اگر یک گره خیلی شلوغ شود، گره‌های بهینه دیگر آن را از آنجا می‌گیرند.

در واقع تعادل بار در زمینه محاسبات ابری، منابع را بین صدها ماشین به هم پیوسته و کم هزینه توزیع می‌کند و به ارائه دهندگان خدمات ابری اجازه می‌دهد تا برنامه‌ها و ترافیک خود را به طور موثر مدیریت کنند. این بدان معنی است که تعادل بار برای میزبانی از توزیع ترافیک وظیفه پاسخگو است تا از انعطاف‌پذیری و قابلیت اطمینان با افزایش تقاضا اطمینان حاصل شود. محیط‌های ابری ممکن است در برابر عدم تعادل منابع به دلیل نوسان تقاضای کاربر آسیب پذیر باشند. به دلیل ناکارآمدی ابر در زمانی که ماشین‌های مجازی به طور یکسان در بین گره‌ها مستقر نمی‌شوند، محققان انگیزه پیدا کرده‌اند تا یک روش متعادل‌سازی بار طراحی کنند. برای حفظ یک توان عملیاتی شبکه ثابت، زیرساخت ابری به شدت تنظیم می‌کند که چه کسی می‌تواند به کدام ماشین‌های مجازی دسترسی داشته باشد.

وظیفه متعادل کننده بار این است که با استفاده بهتر از منابع موجود ابر، پاسخ به درخواست‌های کاربر را

شبکه بدون استفاده از دستگاه یا مکان در اختیار مصرف‌کنندگان قرار می‌گیرد. مؤسسه ملی استانداردها و فناوری، رایانش ابری را به عنوان مدلی برای اجازه دادن همه جا، راحت، یک مجموعه مشترک از منابع محاسباتی سفارشی شده و خدماتی که می‌تواند به سرعت با حداقل کار اداری یا تماس خدماتی عرضه و به کار گرفته شود، تعریف می‌کند.

رایانش ابری را به دلیل گسترده بودن در کاربرد، نمی‌توان تعریف کلی ارائه داد، بنابراین تعاریف آن به این بستگی دارد که برای چه کاربردی استفاده می‌شود. رایانش ابری را می‌توان با نگاه به دو دیدگاه مانند دیدگاه کاربر و سازمان تعریف کرد. بنابراین از دیدگاه کاربر، رایانش ابری برای دستیابی به خدمات مبتنی بر محاسبات بدون نیاز به شناخت عمیق فناوری اساسی مورد استفاده، بسیار مهم است و برای سازمان خدماتی را برای مصرف‌کنندگان و نیازهای تجاری با مقیاس نامحدود و بدون محدودیت، کیفیت خدمات متمایز برای تقویت نوآوری و تصمیم‌گیری سریع ارائه می‌دهد. مفهوم رایانش ابری به سیستمی اشاره دارد که در آن منابع یک مرکز داده با استفاده از فناوری مجازی‌سازی شده به اشتراک گذاشته می‌شود که همچنین می‌تواند خدمات منعطف، درخواستی و فوری را به مشتریان ارائه دهد و به مشتری اجازه دهد با استفاده از روش پرداخت به ازای استفاده پرداخت کند. در واقع رایانش ابری به عنوان دسته‌ای از خدمات محاسباتی بر اساس تقاضا که توسط ارائه دهندگان تجاری مانند مایکروسافت و آمازون ارائه می‌شود، توصیف می‌شود. هدف این اصطلاح ارائه محاسبات، ذخیره‌سازی و نرم‌افزار به عنوان یک سرویس است. هدف رایانش ابری دستیابی به دسترسی مقیاس‌پذیر به منابع محاسباتی و خدمات فناوری اطلاعات است. علاوه بر این، رایانش ابری می‌تواند برای ذخیره مقادیر زیادی داده در موضوعات مختلف مانند موسیقی، کتاب‌های الکترونیکی، پادکست‌ها، برنامه‌ها، ویدئوها و فایل‌ها استفاده شود [۱].

۳- ویژگی‌های کلیدی رایانش ابری

مانند ابرهای واقعی که مجموعه‌ای از مولکول‌های آب هستند، اصطلاح "ابر" در رایانش ابری مجموعه‌ای از شبکه‌ها است. کاربر می‌تواند هر زمان که نیاز باشد از روش‌های محاسبات ابری بی‌حد و حصر استفاده کند. کاربران معمولاً به‌جای راه‌اندازی زیرساخت فیزیکی خود، یک ارائه

سرعت بخشد. توسعه یک روش ارزیابی بار قابل اعتماد برای موفقیت هر طرح متعادل کننده بار بسیار مهم است. ارائه یک روش جامع‌تر برای ارزیابی نیازهای بار ممکن است مهم‌ترین جنبه تعادل بار باشد. این به این دلیل است که بر کارایی، اثربخشی و مصرف انرژی وظایف کاربر تأثیر می‌گذارد. هر دو روش تعادل بار استاتیک و پویا وجود دارد و هر دو ممکن است در فضای ابری با استفاده از نرم‌افزار یا سخت‌افزار انجام شوند.

در این مقاله در بخش اول به مفاهیم عمومی رایانش ابری و توازن بار پرداخته می‌شود و سپس در بخش اصلی، مقایسه مطالعات مرتبط تکنیک‌های توازن بار در رایانش ابری بررسی خواهد شد.

۲- رایانش ابری

توسعه رایانش ابری به طور قابل توجهی نحوه عملکرد بخش فناوری اطلاعات امروز را تغییر داده است. رایانش ابری امکان بررسی خدمات فناوری اطلاعات بهتر با هزینه کمتر و سرمایه‌گذاری کمتر را فراهم می‌کند. محبوبیت نرم‌افزار به عنوان یک سرویس به دلیل تأثیر محاسبات ابری بر نحوه توسعه و تهیه سخت‌افزار فناوری اطلاعات افزایش یافته است. این یک فناوری مبتنی بر اینترنت است که به کاربران امکان می‌دهد هر زمان که بخواهند به داده‌های ذخیره شده در سرور به عنوان یک سرویس دسترسی داشته باشند. مشتریان فقط برای سرویسی که استفاده می‌کنند هزینه می‌پردازند زیرا این یک سرویس پرداختی است. رایانش ابری به عنوان یک مدل محاسباتی است که در آن قابلیت‌های انبوه مقیاس‌پذیر مبتنی بر فناوری اطلاعات به عنوان خدماتی به مشتریان متعدد ارائه می‌شود. استفاده از فناوری رایانه مبتنی بر اینترنت برای انواع خدمات (به عنوان ظرفیت ذخیره‌سازی، قدرت پردازش، برنامه‌های کاربردی تجاری یا اجزاء) است. این مجموعه‌ای از خدمات فعال شبکه است که خدمات مقیاس‌پذیر، تضمین شده، معمولاً سفارشی شده و نسبتاً مقرون به صرفه را به روشی آسان برای استفاده ارائه می‌دهد. رایانش ابری به عنوان یک رویکرد محاسباتی تعریف می‌شود که در آن قابلیت‌های بسیار مقیاس‌پذیر مرتبط با فناوری اطلاعات به عنوان یک سرویس از طریق اینترنت به مصرف‌کنندگان خارجی مختلف ارائه می‌شود. این یک الگوی خدمات فناوری اطلاعات است که در آن سخت‌افزار و نرم‌افزار بر حسب تقاضا در سراسر

میلیون‌ها پول نقد را برای حفظ زیرساخت‌های ابری خرج می‌کنند.

دسترسی به شبکه بزرگ: با کمک اینترنت کاربر نهایی می‌تواند به داده‌های خود در هر کجا دسترسی داشته باشد و آن را به روز کند. برای دسترسی به داده‌ها، کاربر نهایی تنها به دستگاه محاسباتی شخصی با مرورگر نیاز دارد.

مقیاس‌پذیری و در دسترس بودن: مصرف کننده می‌تواند زیرساخت‌های کامپیوتری مورد نیاز خود را در هر زمان تغییر دهد. ارائه دهندگان خدمات ابری همیشه فضای ذخیره‌سازی اضافی و سایر خدمات محاسباتی را با هزینه بسیار کمتر ارائه می‌دهند.

مقرون به صرفه: کاربر نهایی می‌تواند هر زمان که نیاز داشته باشد زیرساخت‌های فناوری اطلاعات را اجاره کند، به این معنی که نیازی به نصب سرورهای پرهزینه و سایر زیرساخت‌های فناوری اطلاعات نیست. ارائه دهندگان خدمات ابری تنها چیزی را که کاربر نهایی استفاده کرده است، هزینه می‌کنند. کاربر نهایی می‌تواند میلیون‌ها پول صرفه‌جویی کند زیرا ارائه دهنده خدمات زیرساخت فناوری اطلاعات را حفظ و به روز می‌کند.

امنیت: فضای ابری از زیرساخت‌های سنتی فناوری اطلاعات امن‌تر است. مشتری نیازی به نگرانی در مورد داده‌های خود ندارد زیرا ابر داده‌های مشابهی را در سرورهای مختلف ذخیره می‌کند، اگر یک سرور خراب شود و سرور دیگری داده‌های کاربر را ارائه می‌دهد. ارائه دهندگان خدمات ابری میلیون‌ها دلار را برای زیرساخت‌های خود خرج می‌کنند.

۴- انواع رایانش ابری

بسته به در دسترس بودن ابر برای مردم، رایانش ابری به سه نوع تقسیم می‌شود. آنها ابر عمومی، ابر خصوصی، ابر ترکیبی هستند [۳]

ابر عمومی: زیرساخت ابری برای عموم یا سازمان-های تجاری بزرگ باز است و متعلق به یک شرکت خدمات ابری است. این سرویس توسط ارائه دهنده خدمات ابری بزرگ برای سیاست پرداخت به ازای استفاده ارائه می‌شود. میکروسافت و گوگل شرکت‌های بزرگی هستند که این سرویس را ارائه می‌دهند. اطلاعات کاربر در سرورهای شخص ثالث ذخیره می‌شود. ابر عمومی در دسترس عموم برای استفاده از ابر قرار دارد. هر مصرف کننده خدمات با کیفیت

دهنده واسطه را برای خدمات اینترنت در رایانش ابری ترجیح می‌دهند. کاربران فقط باید برای خدماتی که استفاده کرده‌اند هزینه پرداخت کنند. برای کاهش حجم کار در رایانش ابری می‌توان حجم کار را تغییر داد. باری از خدمات توسط شبکه‌ها اداره می‌شود که ابر را تشکیل می‌دهند، به همین دلیل بار روی رایانه‌های محلی در حین اجرای برنامه سنگین نیست. بنابراین نیاز به سخت‌افزار و نرم‌افزار در سمت کاربر کاهش می‌یابد. برای استفاده از محاسبات ابری به مرورگر وب مانند کروم نیاز است. ویژگی‌های کلیدی رایانش ابری تجمع منابع و کشش، سلف سرویس و خدمات درخواستی، قیمت گذاری و کیفیت خدمات است [۲].

در مفهوم رایانش ابری، کاربر نهایی معمولاً به جای ایجاد زیرساخت فیزیکی، یک ارائه‌دهنده خدمات ابری را برای نیاز زیرساخت‌های محاسباتی خود ترجیح می‌دهد. کاربر نهایی فقط آنچه را که استفاده کرده است پرداخت می‌کند، این بهترین راه برای کاربر نهایی برای صرفه جویی در پول است. امکانات سخت افزاری و نرم افزاری در فضای ابری سریع و عظیم هستند، بنابراین کاربر نهایی تنها به یک کامپیوتر کوچکتر با مرورگر نصب شده در آن نیاز دارد. ویژگی‌های اصلی رایانش ابری به شرح زیر است [۳]:

ادغام منابع: با کمک مدل چند مستاجر، شرکت خدمات ابری منابع مشابهی را برای مشتریان مختلف فراهم می‌کند. بسته به تقاضای مشتری، منابع فیزیکی و مجازی تخصیص مجدد می‌شوند.

خود سرویس بر حسب تقاضا: کاربر نهایی می‌تواند هر زمان که نیاز داشته باشد به ابر دسترسی پیدا کند. این یک جنبه مفید از رایانش ابری است، زیرا کاربر می‌تواند به طور مداوم زمان کارکرد سرور، قابلیت‌ها و ذخیره‌سازی شبکه را ردیابی کند. مصرف کننده همچنین ممکن است قابلیت‌های دستگاه را با این ویژگی ردیابی کند.

تعمیر و نگهداری آسان: فقط کاربر نهایی باید نرم-افزار و سخت‌افزار رایانه شخصی خود را حفظ کند. سرورها و سایر زیرساخت‌های کامپیوتری در فضای ابری قرار دارند. بنابراین، ارائه‌دهنده خدمات ابری خود را به طور منظم حفظ و به روز می‌کند تا خدمات بهتری به مشتریان خود ارائه دهد. معمولاً ارائه‌دهندگان خدمات ابری شرکت‌های بزرگی هستند، بنابراین برای جذب مشتریان بسیاری،

مهاجرت آسان: شرکت از ابرهای عمومی و خصوصی استفاده می‌کند، بنابراین شرکت می‌تواند هر زمان که نیاز باشد بین آنها مهاجرت کند.

بهبود تخصیص منابع: شرکت می‌تواند منابع بیشتری را برای پروژه‌های موقت تخصیص دهد.

۴- انواع رایانش ابری

رایانش ابری ارائه تسهیلات مرتبط با فناوری اطلاعات بسیار مقیاس‌پذیر به عنوان یک سرویس از طریق اینترنت به چندین مشتری است. مشتریان می‌توانند بر اساس نیازهای خود مدل‌های مختلف خدمات ابری را انتخاب کنند. این شامل خدمات بسیاری به مشتریان است که عموماً به نام XaaS نامیده می‌شوند، که در آن X مخفف هر نوع سرویسی مانند سخت‌افزار، نرم‌افزار، پلت‌فرم، زیرساخت، داده و تجارت و غیره است که ابر به مشتریان ارائه می‌دهد. اساساً X شامل سه نوع سرویس است که به طور گسترده استفاده می‌شود مانند SaaS، PaaS و IaaS. این بخش اجزای ابری مانند سه مدل سرویس و چهار مدل استقرار را توضیح می‌دهد. در زیر سه مدل خدمات به اختصار توضیح داده شده است:

نرم‌افزار به عنوان سرویس (SaaS¹): مجموعه‌ای از نرم افزارها یا سرویس‌ها یا برنامه‌های کاربردی موجود در فضای ابری است که کاربران نهایی بر اساس اشتراک می‌توانند به آنها دسترسی داشته باشند. کاربران نهایی خدمات نرم‌افزاری را از طریق این مدل ارائه خدمات به طور مستقیم از طریق شبکه بر اساس درخواست مصرف می‌کنند [۴]. در این مدل، ارائه دهندگان خدمات ابری وظیفه اجرا و نگهداری نرم افزارهای کاربردی، سیستم عامل و سایر منابع را بر عهده دارند. مدل SaaS به عنوان یک رابط برنامه کاربردی مبتنی بر وب برای مشتری ظاهر می‌شود که در آن از اینترنت برای ارائه خدماتی استفاده می‌شود که با استفاده از یک مرورگر وب به آنها دسترسی پیدا می‌کنند. برنامه‌های میزبانی شده مانند Gmail و Google Docs را می‌توان از طریق دستگاه‌های مختلف مانند تلفن‌های هوشمند و لپ‌تاپ‌ها و غیره در دسترس قرار داد. برخلاف نرم‌افزارهای سنتی، SaaS این مزیت را دارد که مشتری نیازی به خرید مجوز، نصب، ارتقاء، نگهداری یا اجرای نرم افزار بر روی رایانه شخصی خود ندارد. همچنین مزایای دیگری مانند کارایی چند مستاجر، پیکربندی و مقیاس پذیری

یکسانی دریافت می‌کند. مشکل اصلی ابر عمومی این است که مشتری نمی‌داند داده‌های خود در کجا ذخیره شده و تا چه اندازه ایمن هستند. حریم خصوصی داده‌ها و امنیت داده‌ها از معایب اصلی ابر عمومی هستند. نمونه‌هایی از ابر عمومی: Office 365، Google applications، Sales force.com و Amazon AWS [۳].

ابر خصوصی: ابرهای خصوصی توسط یک شرکت خاص ساخته و اداره می‌شوند. این ابر فقط برای یک شرکت کار می‌کند. شخص ثالث یا همان شرکت ابر را مدیریت می‌کند. این ابر برای عموم مردم قابل دسترسی نیست و این ابر فقط برای یک شرکت خاص و در داخل سازمان در دسترس است. ابر دانشگاه بورنموث بهترین مثال برای خصوصی است. ابر دانشگاه بورنموث فقط برای دانشجویان و کارمندان است. در مقایسه با ابر عمومی، ابرهای خصوصی امن‌تر و حریم خصوصی کامل هستند. به دلیل حفظ حریم خصوصی و امنیت، بسیاری از شرکت‌ها مایل به استفاده از ابر خصوصی هستند. عیب اصلی این ابر هزینه راه اندازی و عملیاتی است. برای این نوع از ابر سازمان باید منبع IT خریداری کند و نیاز به استخدام شخصی برای نگهداری دارد. نمونه‌هایی از ابر خصوصی: ابر دانشگاه بورنموث، ابر NHS و ابر ملی اکسپرس [۳].

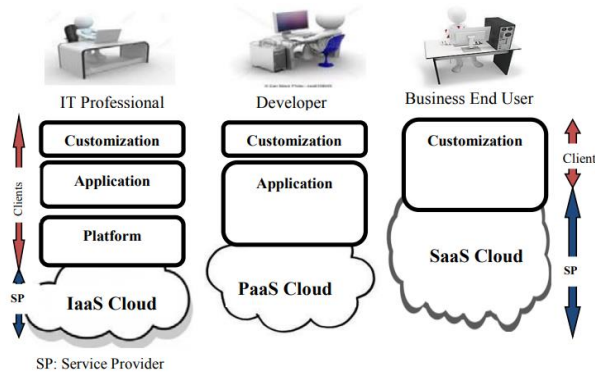
ابر ترکیبی: این نوع ابرها از ترکیب ابرهای عمومی و خصوصی تشکیل می‌شوند. ابر ترکیبی باید حداقل یک ابر عمومی و یک ابر خصوصی داشته باشد. این شرکت‌ها برای دریافت ابر خصوصی با ارائه دهندگان ابر عمومی معامله می‌کنند. ابرهای ترکیبی در مقایسه با سایر ابرها پیچیده‌تر هستند. در این مدل، شرکت‌ها از برخی زیرساخت‌های فناوری اطلاعات خانگی و برخی از زیرساخت‌های فناوری اطلاعات خانگی استفاده خواهند کرد. شرکت‌ها این رویکرد را به دلیل مقیاس پذیری و ارزان تر بودن برای عملیات انتخاب می‌کنند. مزایای ابر هیبریدی عبارتند از [۳]:

هزینه ایجاد زیرساخت فناوری اطلاعات را کاهش می‌دهد زیرا در صورت نیاز از ابر خصوصی استفاده می‌کند.

امن تر و بالاترین مقیاس پذیری: سازمان‌ها از ابر خصوصی برای اطلاعات حساس و از ابر عمومی برای استفاده عادی استفاده می‌کنند.

¹ Software as a service

باند شبکه مصرف شده، زیرساخت شبکه استفاده شده در ساعت، هزینه کاربران محاسبه می‌شود [۴]. زیرساخت به عنوان یک سرویس برای تبدیل منابع فیزیکی به منابع منطقی استفاده می‌شود که می‌تواند به صورت پویا توسط مشتریان در صورت نیاز تهیه و منتشر شوند. برخی از شرکت‌های بزرگی که زیرساخت‌ها را به عنوان خدمات ارائه می‌کنند عبارتند از IBM, Amazon EC2, Google, Rackspace Cloud Servers و Verizon [۵]. شکل ۱ نحوه انتخاب مشتریان از مدل‌های مختلف خدمات ابری را نشان می‌دهد.



شکل ۱: نمایشی از مدل‌های سرویس ابری

۵- تعادل بار

تعادل بار یکی از چالش‌ها و نگرانی‌های اصلی در محیط‌های محاسباتی است. این فرآیند تخصیص، تخصیص مجدد بار در بین منابع موجود به منظور به حداکثر رساندن توان عملیاتی، در عین به حداقل رساندن هزینه و زمان پاسخ، بهبود عملکرد و استفاده از منابع و همچنین صرفه‌جویی در انرژی است. توافقنامه سطح خدمات و رضایت کاربر را می‌توان با تکنیک‌های متعادل کننده بار عالی ارائه کرد. بنابراین ارائه الگوریتم‌ها و مکانیسم‌های متعادل کننده بار کارآمد، کلید موفقیت محیط‌های محاسباتی است. تحقیقات متعددی در زمینه تعادل بار و زمان‌بندی وظایف در محیط‌های محاسباتی انجام شده است [۶].

مدل متعادل‌سازی بار در شکل ۲ نشان داده شده است، جایی که می‌توان دید که متعادل کننده بار درخواست‌های کاربران را دریافت می‌کند و الگوریتم‌های متعادل کننده بار را برای توزیع درخواست‌ها بین ماشین‌های مجازی (VMs) اجرا می‌کند. متعادل کننده بار تصمیم می‌گیرد که کدام ماشین مجازی باید به درخواست بعدی اختصاص یابد. کنترل کننده مرکز داده مسئول مدیریت وظایف است. وظایف به

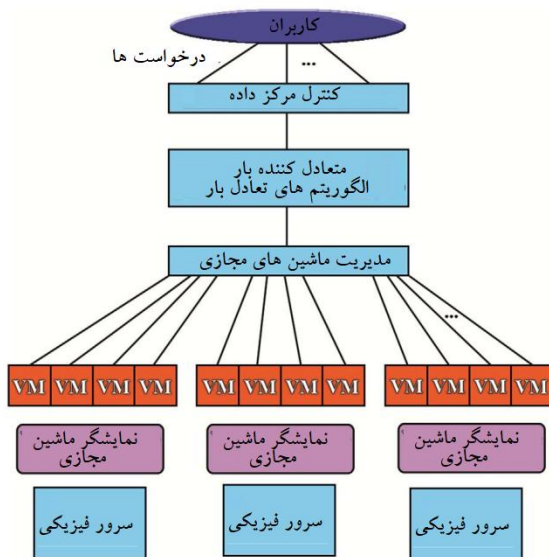
دارد [۵]. مثال‌ها: برنامه‌های Whats، فیس‌بوک، توییتر، Google Docs و صفحات گسترده، salesforce.com، NETSUITE و IBM LotusLive [۴].

پلتفرم به عنوان سرویس (PaaS²): مجموعه‌ای از محیط‌های زمان اجرا مانند نرم‌افزار و ابزارهای توسعه است که بر روی سرورهای ارائه دهنده میزبانی می‌شود. به عنوان پس‌زمینه‌ای عمل می‌کند که محیط زمان اجرا، چارچوب استقرار نرم‌افزار و مؤلفه در پرداخت را برای تسهیل استقرار مستقیم دارایی‌های سطح برنامه یا برنامه‌های کاربردی وب فراهم می‌کند. توسعه‌دهندگان برنامه، پیاده‌کننده‌ها، آزمایش‌کنندگان و مدیران می‌توانند به دنبال توسعه، آزمایش و استقرار نرم‌افزار خود در این پلتفرم باشند و در اینجا کل چرخه عمر نرم‌افزار اجرا می‌شود [۴]. این نوع پیشرفته‌تری از خدمات رایانش ابری است. در PaaS، یک ارائه دهنده خدمات ابری هم نرم‌افزار سیستم (یعنی سیستم عامل) و هم سایر منابع محاسباتی را ارائه، اجرا و نگهداری می‌کند. خدمات PaaS شامل طراحی، توسعه و میزبانی برنامه‌ها می‌باشد. سایر خدمات شامل همکاری، ادغام DB، امنیت، یکپارچه‌سازی وب سرویس، مقیاس‌بندی و غیره است. کاربران نیازی به نگرانی در مورد داشتن منابع سخت‌افزاری و نرم‌افزاری خود یا استخدام کارشناسان برای مدیریت این منابع ندارند. این طرح انعطاف‌پذیری را در نصب نرم‌افزار بر روی سیستم فراهم می‌کند. مقیاس‌پذیری یکی دیگر از مزایای PaaS است. نقص PaaS عدم قابلیت همکاری و قابلیت حمل در میان ارائه دهندگان است [۵]. مثال‌ها: Amazon AWS، Rollbase، jelastic.com، Google App Engine (GAE)، force.com و Microsoft Azure و LONG JUMP [۴].

زیرساخت به عنوان یک سرویس (IaaS): مجموعه‌ای از سرورها، ذخیره‌سازی و شبکه‌ها است. مجازی‌سازی ستون فقرات پشت این مدل است که در آن منابعی مانند شبکه، ذخیره‌سازی، سرورهای مجازی شده، روترها و غیره توسط کاربر از طریق دستکاپ مجازی، ارائه شده توسط CSP³، مصرف می‌شود. بر اساس استفاده از هر ساعت CPU، گیگابایت داده ذخیره شده در ساعت، خدمات ارزش افزوده استفاده شده (مانند نظارت، مقیاس خودکار و غیره)، پهنای

² Platform as a service

³ Cloud Vice Provider



شکل ۲: نمایی از مدل متعدد کننده بار [۶].

متعدد کننده بار ارسال می‌شود، که الگوریتم متعدد کننده بار را انجام می‌دهد تا وظایف را به یک ماشین مجازی مناسب اختصاص دهد. مدیر ماشین مجازی مسئول ماشین‌های مجازی است. مجازی‌سازی یک فناوری غالب در رایانش ابری است. هدف اصلی مجازی‌سازی به اشتراک‌گذاری سخت‌افزار گران قیمت بین ماشین‌های مجازی است. VM یک نرم‌افزار پیاده‌سازی کامپیوتر است که سیستم عامل‌ها و برنامه‌ها می‌توانند روی آن اجرا شوند. VMها درخواست‌های کاربران را پردازش می‌کنند. کاربران در سراسر جهان واقع شده‌اند و درخواست‌های آنها به صورت تصادفی ارسال می‌شود.

درخواست‌ها باید برای پردازش به ماشین‌های مجازی اختصاص داده شوند. بنابراین، تعیین تکلیف موضوع مهمی در سیستم‌های توزیع شده است. اگر برخی از VMها بیش از حد بارگذاری شوند در حالی که برخی دیگر بیکار هستند یا کار کمی برای انجام دادن داشته باشند، QoS کاهش می‌یابد. با کاهش QoS، کاربران ناراضی می‌شوند و ممکن است سیستم را ترک کنند و دیگر برنگردند. یک هایپروایزر یا مانیتور (نمایشگر) ماشین مجازی برای ایجاد و مدیریت ماشین‌های مجازی استفاده می‌شود. مانیتور (نمایشگر) ماشین مجازی چهار عملیات را فراهم می‌کند: چندگانه‌سازی، تعلیق (ذخیره‌سازی)، تامین (رزومه) و مهاجرت زندگی. این عملیات برای متعدد کردن بار ضروری است. متعدد سازی بار باید دو وظیفه تخصیص منابع و زمان بندی کار را در نظر بگیرد. نتیجه این دو وظیفه در دسترس بودن زیاد منابع، صرفه جویی در مصرف انرژی، افزایش بهره‌برداری از منابع، کاهش هزینه استفاده از منابع، حفظ کثرت رایانش ابری و کاهش انتشار کربن است [۶].

۵- معیارهای تعادل بار

در این بخش به بررسی معیارهای تعادل بار در رایانش ابری پرداخته شده است. همانطور که قبلاً ذکر شد، محققان چندین الگوریتم متعدد کننده بار را پیشنهاد کرده‌اند. ادبیات تعادل بار معیارهای پیشنهادی را برای به کارگیری الگوریتم‌های متعدد کننده بار پیشنهاد می‌کند و به صورت زیر خلاصه می‌شود [۶]:

- **توان عملیاتی:** این متریک برای محاسبه تعداد فرآیندهای تکمیل شده در واحد زمان استفاده می‌شود [۶]. در واقع این پارامتر نشان می‌دهد که چند کار در واحد زمان به طور کامل اجرا می‌شوند. حداکثر توان به معنای بهترین عملکرد است [۷].
- **زمان پاسخگویی:** کل زمانی را که سیستم برای انجام یک کار ارسالی صرف می‌کند، اندازه‌گیری می‌کند [۶]. در واقع زمان بین ارسال یک درخواست و دریافت اولین پاسخ از آن کار در یک سیستم مجازی به عنوان زمان پاسخ گفته می‌شود. در نتیجه می‌توان این زمان را کاهش داد تا به سطح بالاتری از کارایی رسید [۷].
- **Makespan:** این متریک برای محاسبه حداکثر زمان تکمیل یا زمانی که منابع به یک کاربر تخصیص داده می‌شود استفاده می‌شود [۶]. در واقع این زمان تجمعی است که یک ماشین مجازی طول می‌کشد تا تمام وظایفی که به آن داده شده را به پایان برساند. این پارامتر باید کوچکترین مقدار ممکن باشد زیرا هرچه تکمیل الگوریتم کمتر باشد، بهتر است [۷].
- **مقیاس‌پذیری:** توانایی یک الگوریتم برای انجام متعدد سازی بار یکنواخت در سیستم با توجه به الزامات با افزایش تعداد گره‌ها است. الگوریتم ترجیحی بسیار مقیاس‌پذیر است [۶]. از جایگاه نشان دهنده

- توانایی سیستم برای کار در شرایط غیرعادی است، این به توانایی یک سیستم سالم برای زنده ماندن زمانی که میزان یا مقیاس وظایف یا حجم کار افزایش می‌یابد اشاره دارد. تغییر مقیاس زیرساخت را می‌توان روزانه در یک محیط ابری انعطاف پذیر انجام داد [۷].
- **تحمل خطا:** توانایی الگوریتم را برای انجام متعادل-سازی بار در صورت بروز برخی خرابی‌ها در برخی گره‌ها یا لینک‌ها مشخص می‌کند [۶]. در واقع این پارامتر به توانایی سیستم برای ارائه تسهیلات پیوسته و ثابت اشاره دارد، حتی اگر یک یا چند گره تصادفی فلاپ شوند [۷].
- **صحت^۴:** صحت به ظرفیت یک بعد برای ارتباط با ارزش واقعی تکمیل کار در حال محاسبه اشاره دارد. کیفیت نتایج اجرای ماموریت را تعیین می‌کند. اخیراً صنعت فناوری اطلاعات تاکید بیشتری بر دقت دستگاه در پاسخ به تقاضای مصرف کننده داشته است. مقدار سازگاری دامنه ساخت سیستم را به مقدار کمی محدود می‌کند [۷].
- استفاده از منابع: موقعیتی که در آن منابع سیستم مانند پردازنده، حافظه، ذخیره سازی و شبکه همه به یک شکل استفاده می‌شوند [۷].
- **زمان مهاجرت:** مدت زمان مورد نیاز برای انتقال یک کار از یک گره اضافه بار به یک گره کم بار [۶].
- **درجه عدم تعادل:** این متریک عدم تعادل بین ماشین‌های مجازی را اندازه‌گیری می‌کند [۶].
- **عملکرد:** کارایی سیستم را پس از اجرای یک الگوریتم متعادل کننده بار اندازه‌گیری می‌کند.
- مصرف انرژی: میزان انرژی مصرف شده توسط تمامی گره‌ها را محاسبه می‌کند. تعادل بار به جلوگیری از گرمای بیش از حد کمک می‌کند و در نتیجه با متعادل کردن بار در تمام گره‌ها، مصرف انرژی را کاهش می‌دهد [۶] و مقدار برق مصرف شده توسط دستگاه‌های مورد استفاده در رایانش ابری یا مراکز داده خاص است [۷].
- **قابلیت اطمینان:** در صورت خرابی دستگاه، کار به کامپیوتر مجازی دیگری منتقل می‌شود تا پایداری سیستم افزایش یابد [۷].
- **انتشار کربن:** مقدار کربن تولید شده توسط همه منابع را محاسبه می‌کند. تعادل بار با جابجایی بارها از گره‌های کم بار و خاموش کردن آنها، نقش کلیدی در به حداقل رساندن این معیار دارد [۶]. این مقدار کربن تولید شده توسط تامین کنندگان برق یک سرویس است، در حالی که تأثیر مخربی بر آب و هوای ما دارد، آلودگی کربنی یک سرویس باید کاهش یا تنظیم شود [۷].
- **پهنای باند:** ترافیک خروجی از شبکه محلی و همچنین ترافیک ورودی از یک عامل اینترنتی را مدیریت می‌کند. این شکاف در ترافیک وب باید آدرس باشد [۷].
- **سربار مرتبط:** این سربار تولید شده توسط اجرای الگوریتم‌ها است. هزینه سربار حمل بار بر روی دستگاه از استراتژی متعادل سازی ناشی می‌شود. اگر بار سیستم به طور مناسب کالیبره شود، حداقل سربار وجود دارد [۷].
- **هزینه مرتبط:** این هزینه بر اساس تعداد منابع استفاده شده تعیین می‌شود. به عنوان مثال، در حالی که منبع به طور کامل استفاده می‌شود، خدمات ارائه شده توسط EC2 هزینه کل را تا ۴۹٪ به حداقل می‌رساند. کاربر ابری می‌خواهد هزینه تامین منابع را با کاهش هزینه منابع درخواستی و همچنین هزینه منابع بیش از حد مشترک که شامل تامین بیش از حد و تامین کمتر است کاهش دهد [۷].
- **زمان مهاجرت:** مدت زمانی که طول می‌کشد تا یک کار یا ماشین مجازی از یک کامپیوتر فیزیکی به کامپیوتر دیگر تحویل داده شود. ممکن است از یک مرکز داده به مرکز دیگر یا از یک میزبان به میزبان دیگر باشد. برای دستیابی به عملکرد متعادل کننده بار مناسب، این زمان باید به نصف کاهش یابد [۷].
- **درجه عدم تعادل:** درجه عدم تعادل معیاری است که تفاوت بین ماشین‌های مجازی را محاسبه می‌کند.

⁴ Accuracy

نظر موقعیت مکانی به عنوان یک سیستم مکانی برای اجرای کارآمد درخواست‌های کاربر در نظر گرفته می‌شوند. برخی از تکنیک‌های توازن بار نیز برای منطقه کوچکتر طراحی شده‌اند که در آن عواملی مانند تاخیر شبکه، تاخیر ارتباطی، فاصله بین گره‌های محاسباتی توزیع شده، فاصله بین کاربر و منابع و غیره در نظر گرفته نمی‌شود. گره‌هایی که در مکان‌های بسیار دور قرار گرفته‌اند نیز یک چالش محسوب می‌شوند، زیرا این الگوریتم‌ها برای این محیط مناسب نیستند. بنابراین، طراحی الگوریتم‌های توازن بار برای گره‌های واقع در دوردست نیز باید در نظر گرفته شود.

تک نقطه خرابی؟ الگوریتم‌های توازن بار پویا مختلفی طراحی شده‌اند که برخی از تکنیک‌های آن توزیع شده نیستند و تصمیم‌گیری برای توازن بار توسط گره مرکزی گرفته می‌شود. اگر ماشین مرکزی از کار بیافتد، بر محیط محاسباتی تأثیر خواهد گذاشت. بنابراین، نیاز به توسعه برخی از الگوریتم‌های توزیع شده وجود دارد که در آن یک گره، کل سیستم محاسباتی را کنترل نمی‌کند.

مهاجرت ماشین مجازی: در مجازی‌سازی امکان ایجاد چندین VM در یک ماشین فیزیکی واحد فراهم می‌شود. این ماشین‌های مجازی ماهیت مستقل و پیکربندی‌های مختلفی دارند. اگر یک ماشین فیزیکی بیش از حد بارگذاری شود، برخی از ماشین‌های مجازی باید با استفاده از روش توازن بار مهاجرت VM به مکان‌های دورتری منتقل شوند.

گره‌های ناهمگن: در تحقیقات اولیه توازن بار ابری، محققان در مورد گره‌های همگن نظریه‌پردازی کردند. در رایانش ابری، نیازهای کاربر به صورت پویا تغییر می‌کند، در نتیجه، برای استفاده موثر از منابع و به حداقل رساندن زمان پاسخ، آنها باید بر روی گره‌های ناهمگن اجرا شوند. بنابراین، ابداع تکنیک‌های توازن بار کارآمد برای محیط ناهمگن یک چالش برای محققان خواهد بود.

مدیریت ذخیره‌سازی: ذخیره‌سازی ابری توانسته است که مسئله سیستم‌های ذخیره‌سازی سنتی قدیمی که به مدیریت پرسنل و هزینه سخت‌افزار بالایی نیاز داشتند، حل کند. ابر به کاربران اجازه می‌دهد تا داده‌ها را به صورت ناهمگن و بدون مشکل دسترسی ذخیره کنند. ذخیره‌سازی

قابل پیش‌بینی بودن: درجه ای است که برای پیش‌بینی توزیع کار، اجرا و دستیابی بر اساس ابزارهای ابری موجود استفاده می‌شود. مقدار قابل پیش‌بینی با رفتار قبلی ورود کار در سیستم و همچنین تقسیم و تکمیل وظایف خاص در سیستم ابری تعیین می‌شود. تخمین تخصیص وظایف بهتر تعادل بار را ارتقا می‌دهد و بهره‌وری را بهبود می‌بخشد [۷].

پیچیدگی: یک الگوریتم باید حتی زمانی که با شرایط پیش‌بینی نشده مواجه می‌شود، عملکرد بهینه‌ای از خود نشان دهد. این نشان می‌دهد که صرف نظر از افزایش کارها و بار، الگوریتم باید مقیاس‌پذیری را حفظ کند و از عملکرد بالا برای الگوریتم متعادل‌کننده بار اطمینان حاصل کند [۸].

سرعت: کل زمان مورد نیاز برای تمام وظایف و توزیع منابع بین کاربران درون سیستم به عنوان زمان تکمیل کل شناخته می‌شود. این معیار در فرآیند زمان‌بندی در محیط ابری بسیار مهم است و معیاری از زمان مورد نیاز برای پردازش مجموعه خاصی از وظایف را ارائه می‌دهد. هدف یک الگوریتم متعادل بار خوب، به حداقل رساندن این معیار است و از تکمیل کارآمد اطمینان می‌دهد [۸].

۵- چالش‌های توازن بار در فضای ابری

فناوری رایانش ابری به هدف تحقیقات پیشرفته در زمینه داده‌ها و محاسبات از نظر جنبه‌های نظری و عملی تبدیل شده است. با این حال، این تحقیقات با مشکلات زیادی مواجه هستند. توازن بار یکی از چالش‌های مهم است که باید به طور ویژه به آن پرداخت. علاوه بر این، چندین موضوع دیگر مانند مهاجرت ماشین مجازی^۵ (VM)، امنیت VM، رضایت کاربر از QoS و استفاده از منابع نیز باید مورد توجه قرار گیرند تا بهترین راه‌حل ممکن برای بهبود استفاده از منابع ابری پیدا شود. در زیر لیستی از چند مسئله توازن بار مورد بحث قرار گرفته است [۹]:

توزیع گره‌ها از نظر جغرافیایی: به طور کلی در فضای ابری، مراکز داده از نظر جغرافیایی برای مقاصد محاسباتی توزیع شده‌اند. در این مراکز گره‌های توزیع شده از

⁶ Single Point of Failure

⁵ virtual machine

گرفتن عواملی مانند مقیاس‌پذیری، مقرون‌به‌صرفه بودن و سازگاری با بارهای کاری پویا است [۱۰].

تکنیک‌های توازن بار اساساً بر اساس وضعیت سیستم و راه‌اندازی فرآیند طبقه‌بندی می‌شوند. رویکردهای توازن بار مبتنی بر وضعیت سیستم به دو دسته ایستا و پویا طبقه بندی می‌شوند، که در شکل ۲-۵ نشان داده شده است. بر اساس شروع فرآیند، آنها به صورت "آغازشونده توسط فرستنده"^۹، آغازشونده توسط گیرنده"^{۱۰}، و متقارن"^{۱۱} طبقه بندی می‌شوند [۹]:

آغازشونده توسط فرستنده: در این تکنیک، اگر یک گره دارای بیش‌بار باشد به دنبال گره‌هایی می‌گردد که دارای بار اندک هستند تا بتواند حجم کار را به اشتراک بگذارد. زمان ازدحام گره‌ها، فرستنده فرآیندی برای یافتن گره‌های بدون بار آغاز می‌کند.

آغازشونده توسط گیرنده: در این تکنیک، گیرنده یا گره‌های کم‌بار به دنبال گره‌های بیش‌بار برای اشتراک‌گذاری حجم کار می‌گردند.

متقارن: در این تکنیک، هر دو تکنیک "آغازشونده توسط فرستنده" و "آغازشونده توسط گیرنده" با هم ترکیب می‌شوند.

بر اساس وضعیت سیستم، تکنیک‌های توازن بار را می‌توان به دسته‌های زیر تقسیم کرد:

ایستا: تکنیک‌های توازن بار ایستا از مجموعه‌ای ثابت از قوانین پیروی می‌کنند که به وضعیت فعلی سیستم وابسته نیستند. الگوریتم‌های ایستا انعطاف‌پذیر نیستند و نیاز به دانش قبلی از منابعی مانند زمان ارتباط، حافظه و ظرفیت ذخیره سازی گره‌ها، قدرت پردازش گره‌ها و غیره دارند. این تکنیک ساده و آسان است، اما قادر به شناسایی سرورهای متصل نیست، در نتیجه، به توزیع نابرابر منابع منجر می‌شود. مسئله اصلی این تکنیک این است که وضعیت فعلی سیستم در هنگام تصمیم‌گیری در نظر گرفته نمی‌شود. بنابراین، برای سیستم‌های توزیع‌شده که به صورت پویا تغییر حالت می‌دهند، نامناسب است. تکنیک‌های ایستا فقط در صورتی بسیار خوب عمل می‌کنند که نوسان بار کمتری در گره‌ها رخ

ابری روزبه‌روز در حال افزایش است، در نتیجه برای دسترسی کارآمد و سازگاری داده‌ها نیازمند ذخیره‌سازی تکثیر^۷ (تهیه نسخه پشتیبان) از داده‌ها است. طرح‌های تکثیر کامل داده‌ها، به دلیل خطمشی ذخیره‌سازی داده‌های تکراری در نقاط تکثیر، بسیار کارآمد نیستند. تکثیر جزئی می‌تواند خوب باشد، اما مسئله در دسترس بودن مجموعه داده‌ها وجود دارد که پیچیدگی تکنیک‌های توازن بار را افزایش می‌دهد. بنابراین، یک تکنیک توازن بار کارآمدی باید توسعه یابد که توزیع اپلیکشن و داده‌های مرتبط را بر اساس سیستم تکثیر جزئی در نظر بگیرد.

مقیاس‌پذیری متوازن‌کننده بار (لود بالانس)^۸:

در دسترس بودن و مقیاس‌پذیری سرویس‌های ابری برحسب تقاضا به کاربران این امکان را می‌دهد که در هر زمان به سرویس‌ها دسترسی داشته باشند و به سرعت آنها را کاهش یا افزایش مقیاس دهند. یک متوازن‌کننده بار خوب باید تغییرات سریع تقاضاها را از نظر توان محاسباتی، ذخیره سازی، توپولوژی سیستم و غیره در نظر بگیرد تا بتواند این تغییرات را به طور کارآمد تسهیل کند.

پیچیدگی الگوریتم: در رایانش ابری، الگوریتم‌ها

باید ساده و راحت اجرا شوند. یک الگوریتم پیچیده باعث کاهش عملکرد و کارایی سیستم ابری می‌شود.

۵- طبقه‌بندی استراتژی‌های توازن بار بر اساس وضعیت سیستم

توازن بار فرآیند تقسیم موثر ترافیک شبکه ورودی بین چندین سرور باطن است. اساساً سه نوع نرم افزار، سخت افزار و ابر بومی در تعادل بار وجود دارد و به منظور پراکنده کردن ترافیک شبکه و برنامه، یک متعادل کننده بار سخت افزاری از سخت افزار واقعی که در محل قرار دارد استفاده می‌کند. دو نوع نرم افزار متعادل کننده بار وجود دارد: منبع باز و تجاری. هر دو قبل از استفاده نیاز به نصب دارند. بر خلاف متعادل کننده بار نرم افزار، یک متعادل کننده بار مجازی (همچنین به عنوان متعادل کننده بار بومی ابری شناخته می‌شود) نرم افزار دستگاه متعادل کننده بار سخت افزاری را بر روی یک ماشین مجازی نصب می‌کند. چالش اصلی نیاز به مقایسه عملکرد آنها در محیط‌های مختلف شبکه با در نظر

⁹ sender initiated

¹⁰ reciver initiated

¹¹ symmetric

⁷ replication

⁸ Load-Balancer

گره‌ها دارای یک پایگاه اطلاعاتی برای توزیع کارآمد وظایف هستند. الگوریتم‌های توزیع شده می‌توانند به صورت مشارکتی یا غیرمشارکتی باشند تا بتوانند با یکدیگر در تعامل باشند. اگر همه گره‌های سیستم با هم کار کنند تا به یک هدف یا تصمیم‌گیری مشترک دست یابند، پس آن را مشارکتی^{۲۰} می‌نامند؛ در غیر این صورت، غیرمشارکتی^{۲۱} است.

توزیع نشده: در تکنیک‌های غیرتوزیعی، یک گره منفرد یا چند گره برای توزیع بار تصمیم‌گیری می‌کنند. تکنیک‌های غیر توزیعی می‌توانند به صورت متمرکز یا نیمه توزیع شده باشند. در تکنیک‌های متمرکز، یک گره منفرد همه فعالیت‌های توزیع بار را انجام می‌دهد و مسئول توازن بار است. تحمل خطا^{۲۲} یکی از چالش‌های تکنیک‌های متمرکز است و در صورت خرابی تک گره، اطلاعات گره می‌تواند از بین برود و بازیابی نشود. در تکنیک نیمه توزیع شده، خوشه‌ها در گره‌ها تشکیل می‌شوند و هر خوشه به عنوان یک تکنیک متمرکز عمل می‌کند.

۹- طبقه‌بندی و مقایسه تکنیک های متعادل کننده بار

برخی از الگوریتم‌های متوازن بار استاتیکی محبوب در این بخش ارائه شده‌اند:

الگوریتم راندرابین^{۲۳}: یک الگوریتم متعادل کننده بار استاتیک، بدون در نظر گرفتن حالت های قبلی عمل می‌کند. سادگی آن در استفاده از روش راندرابین برای توزیع وظایف است. الگوریتم به طور تصادفی اولین گره را انتخاب می‌کند و سپس با استفاده از روش راندرابین، کارها را به طور مساوی به تمام گره‌های دیگر تخصیص می‌دهد. یک مزیت اساسی راندرابین استقلال آن از ارتباطات بین فرآیندی است. این الگوریتم در صورت توزیع بار کار برابر بین فرآیندها، پاسخ سریع‌تری را ارائه می‌دهد، اما به دلیل کمبود اطلاعات در مورد زمان اجرای پردازنده‌ها، برخی وظایف ممکن است بارهای سنگینی را تجربه کنند.

دهد. تکنیک‌های توازن بار ایستا به شرح زیر دسته‌بندی می‌شوند:

- **بهینه:** در تکنیک‌های بهینه، DCN اطلاعات مربوط به منابع را جمع آوری می‌کند و وظایف را به متوازن‌کننده بار ارسال می‌کند، تا تخصیص بهینه را در حداقل زمان انجام دهند.
- **نسبتا بهینه (غیربهینه):** اگر متوازن‌کننده بار قادر به تعیین تصمیم بهینه نباشد، یک راه حل نسبتا بهینه را محاسبه خواهد کرد. چند تکنیک ایستا عبارتند از: مین-مین^{۱۲}، ماکس-مین^{۱۳}، راند رابین^{۱۴}، نخست کوتاهترین کار^{۱۵}، توازن بار فرصت طلبانه دومرحله‌ای^{۱۶} (OLB)، و توازن بار مرکزی برای ماشین‌های مجازی.

پویا: این تکنیک‌ها براساس وضعیت فعلی سیستم تصمیم‌گیری می‌کنند. مزیت اصلی این تکنیک‌ها این است که آنها وظایف را از یک ماشین دارای بیش‌بار به یک ماشین کم‌بار منتقل می‌کنند. تکنیک‌های توازن بار پویا انعطاف‌پذیر هستند که منجر به بهبود عملکرد سیستم می‌شود. در طول پردازش، مراحل زیر را انجام می‌دهند: به طور مداوم بار گره‌ها را نظارت می‌کند. در یک بازه زمانی مشخص، برای محاسبه بار کاری گره‌ها، اطلاعات بار و وضعیت را بین گره‌ها مبادله می‌کند و بار کاری را بین گره‌ها مجدداً توزیع می‌کند. اگر یک گره دارای بیش‌بار شود، بار را به یک گره کم‌بار منتقل می‌کند. برخی از تکنیک‌های توازن بار پویا عبارتند از: توازن بار مبتنی بر عامل^{۱۷}، توازن بار الهام‌گرفته از رفتار زنبور عسل^{۱۸}، بهینه سازی کلونی مورچگان، الگوریتم فشرده‌ساز^{۱۹}، تکنیک‌های متوازن بار پویا را می‌توان به صورت زیر دسته‌بندی کرد:

- **توزیع شده:** در تکنیک‌های توزیع شده، همه گره‌ها در فرآیند توزیع بار، مانند زمان‌بندی وظیفه یا تخصیص منابع، شرکت می‌کنند. همه

¹² Min-Min

¹³ Max-Min

¹⁴ Round Robin

¹⁵ Shortest Job First

¹⁶ Two-phase Opportunistic Load Balancing

¹⁷ Agent-based Load Balancing

¹⁸ Honey Bee

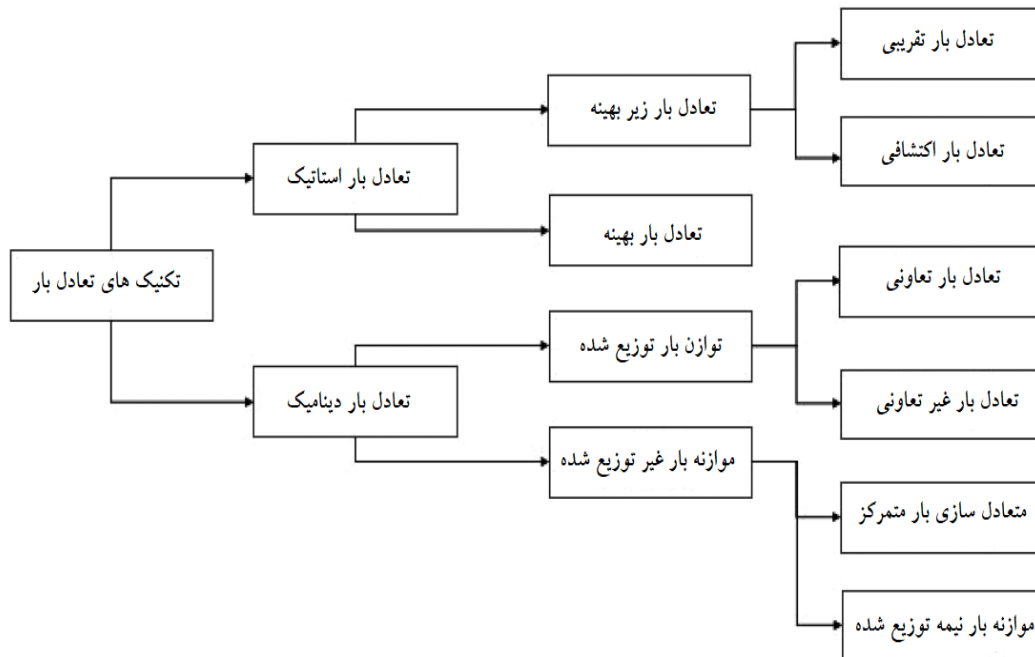
¹⁹ Throttled

²⁰ cooperative

²¹ non-cooperative

²² Fault tolerance

²³ Round Robin



شکل ۳: نمایی از طبقه بندی استراتژی‌های تعادل بار بر اساس وضعیت سیستم

الگوریتم تعادل بار فرصت طلب: اصول اولیه این

الگوریتم مانع از درک سطح فعالیت ماشین مجازی می‌شود. بنابراین، تضمین می‌کند که هر گره در سیستم با تخصیص تصادفی آن وظایف به طور فعال کار می‌کند. با این حال، همه چیز به کندی پیش می‌رود، زیرا مدت زمان تکمیل کارها را محاسبه نمی‌کند. در نتیجه، هنگام تلاش برای توزیع یکنواخت کار، نتایج قابل اعتمادی را ارائه نمی‌دهد. الگوریتم‌های پویا را می‌توان به ترتیب زیر معرفی کرد:

تراز کننده بار دریچه‌ای²⁴: این الگوریتم یک

جدول حاوی ماشین‌های مجازی و وضعیت فعلی آنها (در دسترس یا مشغول) تولید می‌کند. هنگامی که یک کار خاص به یک ماشین مجازی اختصاص داده می‌شود، یک درخواست به واحد کنترل مرکز داده ارسال می‌شود. این واحد به دنبال مناسب‌ترین ماشین مجازی بر اساس قابلیت‌های خود برای کار داده شده است. تخصیص دهنده جدول فهرستی به نام جدول فهرست یا جدول تخصیص، حاوی تمام ماشین‌های مجازی و حالت‌های مربوط به آنها (در دسترس، مشغول یا بیکار) نگهداری می‌کند. اگر ماشین مجازی موجود با فضای کافی پیدا شود، وظیفه به آن ماشین مجازی اختصاص داده می‌شود.

الگوریتم تعادل بار Min-Min: این الگوریتم

کاربرپسند بوده و به سرعت عمل می‌کند. این کار را با تکمیل وظایف در یک سری افزایش می‌دهد. زمان هر کار را محاسبه می‌کند و بر اساس کوتاه‌ترین زمان تکمیل وظایف موجود، آن را به ماشین‌های مجازی اختصاص می‌دهد. این کار تا زمانی ادامه می‌یابد که تمام وظایف به ماشین مجازی‌ها اختصاص داده شود. این الگوریتم به دلیل وظایف کوچکتر متعدد عملکرد خوبی دارد، اما خطر نادیده گرفتن وظایف بزرگتر وجود دارد که منجر به گرسنگی می‌شود زیرا کارهای کوچکتر نسبت به کارهای بزرگتر اولویت دارند.

الگوریتم تعادل بار Max-Min: این رویکرد مشابه

متعادل سازی بار Min-Min با تاکید خاص بر زمان محاسبه کارها است. تحت این روش، تمام وظایف به سیستم داده می‌شود و الگوریتم مدت زمان لازم برای تکمیل هر کار را محاسبه می‌کند. این وظیفه با حداقل زمان مورد نیاز به دستگاه اختصاص داده می‌شود. روش Max-Min به دلیل توانایی آن در مدیریت یک کار بزرگتر علاوه بر وظایف کوچکتر در داخل مجموعه، از الگوریتم Min-Min بهتر عمل می‌کند. الگوریتم Max-Min به طور موثر کارهایی با مدت زمان‌های مختلف را بدون توجه به طول آنها مدیریت می‌کند.

²⁴ Throttled Load Balancer

پردازش آن در حد متوسط زمان پردازش تمام ماشین‌های مجازی است.

تکنیک تعادل بار بر اساس جستجوی فاخته و

گرم شب تاب: الگوریتم جستجوی فاخته از رفتار لانه‌سازی پرندگان فاخته الهام می‌گیرد. مشابه نحوه تخم‌گذاری این پرندگان در لانه پرندگان دیگر، الگوریتم بهترین تخم مرغ‌ها را به نسل بعدی منتقل می‌کند و بدترین لانه‌ها را رها می‌کند. در زمینه رایانش ابری، از الگوریتم جستجوی فاخته با فایرفلای، یک الگوریتم ترکیبی، برای متعادل کردن بار استفاده می‌شود. این الگوریتم به طور موثر وظایف را زمان بندی می‌کند و وظایف بارگذاری شده را به ماشین‌های مجازی با بارگذاری کمتر اختصاص می‌دهد. به سرعت بهترین ماشین مجازی را شناسایی می‌کند، کارایی تعادل بار را افزایش می‌دهد و از عدم تعادل وظایف در سراسر سیستم جلوگیری می‌کند.

الگوریتم اجرا به ترتیب ورود: این الگوریتم با

اختصاص وظایف جدید به منابع با کمترین زمان انتظار عمل می‌کند، به این معنی که این منابع کمترین وظایف را انجام می‌دهند. از یک روش اجرای متوالی پیروی می‌کند، از اولین کار شروع می‌شود و تا تکمیل در صف پیشرفت می‌کند. این الگوریتم به این دلیل انتخاب می‌شود که وظایف را بدون در نظر گرفتن مشخصات ماشین‌های مجازی موجود یا زمان مورد نیاز برای وظایف در صف هنگام تخصیص وظایف جدید به ماشین‌های مجازی اختصاص می‌دهد. برای تعیین موثرترین الگوریتم متعادل کننده بار برای عملکرد بهینه در برنامه‌های ابری، معیارهای رایج مختلفی برای اندازه‌گیری اثربخشی آنها به کار گرفته شده است. در این قسمت در جدول ۲ مقایسه عملکرد الگوریتم‌های متعادل کننده بار بر اساس معیارهای مناسب نامبرده در بخش‌های قبلی ارائه شده است.

اگر ماشین مجازی مناسبی در دسترس نباشد، تخصیص دهنده جدول مقدار ۱- را برمی‌گرداند و درخواست برای پردازش سریع در صف قرار می‌گیرد. نقش متعادل کننده بار این است که پس از دریافت درخواست مشتری، یک ماشین مجازی مناسب را بیابد. در حالی که تخصیص دهنده جدول از الگوریتم راندرابین بهتر عمل می‌کند، نیازهای پیشرفته‌تر تعادل بار، مانند زمان پردازش را در نظر نمی‌گیرد.

بهینه‌سازی کلونی مورچه‌ها: این الگوریتم مبتنی

بر **کلونی مورچه‌ها** به طور موثر بارهای کاری را بین گره‌های ابری توزیع می‌کند. با الهام گرفتن از رفتار مورچه، به دنبال مسیرهای جایگزین در هنگام برخورد با موانع، ایجاد مسیرهای جدید بین گره‌ها است. الگوریتم با مقداردهی اولیه جدول، تسهیل جریان داده و رسیدن به سطح آستانه گره‌های مورد نیاز آغاز می‌شود. همانطور که جریان از گره‌ها عبور می‌کند، الگوریتم هر یک را ارزیابی می‌کند. اگر یک گره کم بار باشد، از بالاترین فرمون دنباله دار برای یافتن مسیر گره کم بار استفاده می‌کند. جدول تا زمانی که گره به حد آستانه خود برسد به روز می‌شود. با این حال، پس از رسیدن به حد مجاز، الگوریتم از یک فرومون خوراک یابی برای کاوش منابع جدید استفاده می‌کند و جدول را تا زمانی که یک گره کم بار پیدا کند، به روز می‌کند در این مرحله منابع را دوباره تخصیص می‌دهد. این روند تا زمانی که کل فرآیند تکمیل شود تکرار می‌شود.

الگوریتم زنبور عسل: این الگوریتم بر اساس رفتار

زنبورهایی که در جستجوی عسل هستند عمل می‌کند. هدف اصلی آن توزیع کارآمد بار کار بر روی ماشین‌های مجازی است و از استفاده بیش از حد از منابع و استفاده ناکافی اجتناب می‌کند. الگوریتم ماشین مجازی را انتخاب می‌کند که دو معیار کلیدی را برآورده می‌کند: وظایف کمتری را در مقایسه با سایر ماشین‌های مجازی دریافت می‌کند و زمان

جدول ۱: مقایسه عملکرد الگوریتم‌های متعادل کننده بار

معیارهای عملکرد	راندرابین	Min-Min	Max-Min	فرصت طلب	جستجوی فاخته و گرم شبتاب	تراز کننده بار دریاچه‌ای	الگوریتم مورچه	الگوریتم زنبور	اجرا به ترتیب ورود
توان عملیاتی	پایین	متوسط	بالا	پایین	بالا	بالا	پایین	پایین	متوسط
پهچیدگی	پایین	پایین	متوسط	پایین	متوسط	متوسط	متوسط	بالا	متوسط
زمان پاسخگویی	پایین	پایین	پایین	متوسط	متوسط	بالا	بالا	بالا	بالا
سرعت	پایین	پایین	بالا	پایین	بالا	بالا	پایین	متوسط	متوسط

سربار شبکه	پایین	بالا	متوسط	پایین	متوسط	پایین	بالا	بالا	پایین
تحمل خطا	پایین	پایین	بالا	پایین	بالا	بالا	بالا	متوسط	پایین
زمان مهاجرت	پایین	متوسط	متوسط	پایین	بالا	بالا	پایین	پایین	پایین
استفاده از منابع	بالا	متوسط	بالا	پایین	بالا	بالا	متوسط	بالا	بالا

استاندارد پرداخته شد. از این بررسی می‌توان این نتیجه را گرفت که محل خدمت باید در تعادل بار در نظر گرفته شود. در آینده، الگوریتم‌های کارآمدتری برای کاهش زمان پاسخ و زمان پردازش داده‌ها گنجانده خواهد شد. می‌توان روش‌های تقسیم مجدد منابع را ادغام کرد که عیب‌یابی را به حداقل برساند.

۱۱- منابع

1. Zangiabadi Zadeh, H., et al., *Cloud Computing Security: A review of basic concepts, security challenges, issues, requirements, security standards, and types of attacks in cloud computing. Computing and distributed systems*, 2023. 6(1): p. 137-153.
2. Srivastava, P. and R. Khan, *A review paper on cloud computing. International Journal of Advanced Research in Computer Science and Software Engineering*, 2018. 8(6): p. 17-20.
3. Khan, R.K.N., *A Review Paper on Cloud Computing. All content following this page was uploaded by Riyas KHAN Navas Khan on 18 December 2020. The user has requested enhancement of the downloaded file.*, 2020.
4. Birje, M.N., et al., *Cloud computing review: concepts, technology, challenges and security. International Journal of Cloud Computing*, 2017. 6(1): p. 32-57.
5. Rashid, A. and A. Chaturvedi, *Cloud computing characteristics and services: a brief review. International Journal of Computer Sciences and Engineering*, 2019. 7(2): p. 421-426.
6. Ghomi, E.J., A.M. Rahmani, and N.N. Qader, *Load-balancing algorithms in cloud computing: A survey. Journal of Network and Computer Applications*, 2017. 88: p. 50-71.
7. Jamal, F. and T. Siddiqui, *Comparative Analysis of Load Balancing Techniques in Cloud Computing, Based on LB Metrics. in 2021 5th International Conference on Information Systems and Computer Networks (ISCON). 2021. IEEE.*
8. Oladunjoye, J.A. and E.O. Jessica, *A Comparative Assessment of Load Balancing Techniques in Cloud Computing: A Review.*

۱۰- نتیجه‌گیری

رایانش ابری ارائه خدمات فناوری اطلاعات را متحول می‌کند و به مصرف‌کنندگان امکان دسترسی به منابع مجازی از طریق اینترنت را می‌دهد. از آنجایی که به سرعت جایگزین زیرساخت‌های داخلی می‌شود، دسته‌های مختلفی از خدمات ابری پدیدار می‌شوند و چالش‌هایی را برای مصرف‌کنندگان در انتخاب ارائه‌دهندگان مناسب ایجاد می‌کنند. تحقیقات تفاوت‌های بین معیارهای ارائه‌دهنده و نیازهای مصرف‌کننده را برجسته می‌کند و دشواری انتخاب‌های آگاهانه را تشدید می‌کند. دسته بندی بر اساس دسترسی (خصوصی در مقابل عمومی) و خدمات تصمیم‌گیری را پیچیده‌تر می‌کند. مدل‌های قیمت‌گذاری متنوع چالش را تشدید می‌کنند و مانع مقایسه کیفیت و هزینه می‌شوند. انعطاف‌پذیری رایانش ابری که با مهاجرت ماشین مجازی و صرفه‌جویی در مقیاس امکان‌پذیر شده است، کسب‌وکارهایی را که به دنبال خدمات برتر فناوری اطلاعات با هزینه‌های کمتر هستند، جذب می‌کند. ظهور تعادل بار در رایانش ابری برای بهینه‌سازی استفاده از منابع، افزایش عملکرد و مدیریت افزایش ترافیک بسیار مهم است.

متعادل سازی بار روشی برای توزیع منابع، پردازنده، درایوهای دیسک، سرورها و غیره بر روی شبکه‌ای از گره‌ها است. یکی از مهمترین و چالش برانگیزترین وظایف در محیط ابری، تعادل بار است. برخی از سرورهای مجازی ابری کمتر مورد استفاده قرار می‌گیرند در حالی که برخی دیگر به دلیل عدم تعادل بار تحت فشار قرار می‌گیرند. بنابراین، تعادل بار مناسب در زیرساخت ابر ضروری است. در واقع تعادل بار به عنوان یک مؤلفه مهم ظاهر می‌شود که استفاده بهینه از منابع، مدیریت یکپارچه تغییرات ترافیک و برنامه‌های کاربردی با کارایی بالا را تضمین می‌کند. تکنیک‌های مختلف متعادل‌سازی بار، از روش‌های سنتی مانند Round Robin گرفته تا الگوریتم‌های پیشرفته‌تر، انعطاف‌پذیری و مقیاس‌پذیری را فراهم می‌کنند. این مقاله به بررسی تکنیک‌های متعادل‌سازی بار و مقایسه آنها بر اساس معیارهای



سمیه کدخدا ده خانی، فارغ التحصیل مقطع کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیکز دانشگاه پیام‌نور قشم می باشد، او به عنوان کارشناس فناوری در دانشگاه پیام‌نور استان کرمان مشغول به کار می باشد و نشانه رایانامه ایشان عبارتند از :
Emailsk65@gmail.com

روش ارجاع: م. قاسمی، م. رحمانی، ح. زنگی آبادی زاده، ف. وظیفه دوست و س کدخدا ده خانی. مرور و مقایسه‌ای بر تکنیک‌های تعادل بار در رایانش ابری، دوفصلنامه محاسبات و سامانه‌های توزیع شده، سال هفتم، شماره ۲، شماره پیاپی ۱۴، صفحه ۱۰ تا ۲۴، سال ۱۴۰۳.

How to cite: M.Ghasemi , M.Rahmani. H.Zangiabadi Zadeh .F.Vazifehdoost .S.kadkhodadehkhan. A review and comparison of load balancing techniques in cloud computing, Journal of Distributed Computing and Systems (JDCS), Vol 7, Issue 2, Pages 10 – 24, 2025.

A review and comparison of load balancing techniques in cloud computing

M.Ghasemi¹ .M.Rahmani² . H.Zangiabadi Zadeh³.

F.Vazifehdoost⁴ . S.kadkhodadehkhan⁵

^{1,3} Payame Noor University, Kish International Center.

² Payame Noor University, Kish.

^{4,5} Payame Noor University, Qeshm International Center.

Abstract

Cloud computing is one of the most widely used technologies in the field of IT-based services. It is considered an emerging technology that provides a new impetus for computing based on resource virtualization. Cloud computing has emerged from the need to outsource computing and storage facilities to customers to become essential in daily activities. The advantages of cloud computing are to provide maximum services at minimum cost at all times, etc. One of the main obstacles to the use of cloud computing is the stability of cloud performance. Load balancing is one of the elements that shows its impact on performance stability. Load balancing is one of the most complex. In fact, cloud

International Journal of Distributed & Cloud Computing, 2024. 12(1).

9.Kumar, P. and R. Kumar, *Issues and challenges of load balancing techniques in cloud computing: A survey. ACM Computing Surveys (CSUR), 2019. 51(6): p. 1-35.*

10. Narsipuram, M., A. Rai, and A. Tiwari, *A Comprehensive Study of Load Balancing Architectures in Cloud Computing. Smart Internet of Things, 2024. 1(2): p. 115-127.*



مهدی قاسمی، دانشجوی کارشناسی ارشد مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیکز ، دانشگاه پیام نور مرکز بین الملل کیش می باشد و نشانه رایانامه ایشان عبارتند از :

Mahdikmg1@gmail.com



مائه رحمانی، دانشجوی کارشناسی ارشد مهندسی کامپیوتر گرایش نرم افزار ، دانشگاه پیام نور مرکز بین الملل کیش می باشد و نشانه رایانامه ایشان عبارتند از :

Maede9708@gmail.com



حمید زنگی آبادی زاده، دانشجوی کارشناسی ارشد مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیکز، دانشگاه پیام نور مرکز بین الملل کیش می باشد و نشانه رایانامه ایشان عبارتند از:

Hamid.zangiabadi@gmail.com



فرشید وظیفه دوست، فارغ التحصیل در مقطع کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی و رباتیکز از دانشگاه پیام نور مرکز بین الملل قشم می باشد و نشانه رایانامه ایشان عبارتند از :

Vazifehdoostfarshid@gmail.com

computing technology consists of a large number of resources that allow users to access the correct and necessary resources in a network when needed. Applications with different resource requirements require high-performance computing capabilities that help maximize resource utilization, time consumption, and prevent processor overhead. In this paper, various load balancing techniques applied in cloud environments are discussed and a comparative study is presented.