بررسی روش‌های مدیریت منابع در رایانش ابری و چالش‌های آن‌ها

ابوالفضل اسفندی[۱]، بابک فولادی‌نیا[۲]، مهدی فرتاش[۳]

[۱] مربی، گروه کامپیوتر، دانشگاه آزاد اسلامی واحد بروجرد
Esfandi@iaub.ac.ir

[۲] دانشجوی دکتری، گروه کامپیوتر، دانشگاه آزاد اسلامی واحد اراک
babakfouladiniya@gmail.com

[۳] استادیار، گروه کامپیوتر، دانشگاه آزاد اسلامی واحد اراک
Fartash.iau@gmail.com

Archive of SID

چکیده

محاسبات ابری، به عنوان یک نمونه محاسبات رایج برای سرویس ها و سیستم های محاسباتی بزرگ، شناخته می شود. اخیرا تحقیقات مهمی روی تکنیک های مدیریت منابع، انجام شده است که بر روی به اشتراک گذاری مؤثر منابع میان چندین کاربر، تمرکز دارد. تکنیکهای مدیریت منابع در ابرها برای محاسبات و ظرفیت کار مؤثر برنامه‌های کاربردی که پارامترهای بهینه دارند، طراحی شده اند.

این مقاله یک مرور جامع بر تکنیک‌های مدیریت منابع و شرح دقیق طبقه بندی گسترده آن براساس خصوصیات مختلف را ارائه می‌دهد و به مقایسه این روشها، مزایا و معایب هر یک می‌پردازد. به علاوه این ره آورد، طراحی اهداف و بررسی چالش هایی را که باید در حین پیشنهادات جدید تکنیک‌های مدیریت منابع در نظر گرفته شود، فراهم می‌کند و با ارائه یک طبقه بندی از تکنیک‌های مدیریت منابع مبتنی بر استانداردهای مدیریت منابع اصلی، راهکاری عالی برای پژوهشگران در این زمینه خواهد بود.

کلمات کلیدی

محاسبات ابری، مدیریت منابع، ظرفیت کار، تعادل بار، تخصیص منابع.

۱- مقدمه

در طی چند سال اخیر، دنیای محاسبات به دلیل افزایش تقاضا برای ابزارهای محاسباتی پیشرفته، تکامل زیادی یافته است. این تحول منجر به ظهور نمونه های محاسباتی جدید مثل محاسبات کلاستر، محاسبات گرید و محاسبات ابری شده است. از میان این نمونه ها، محاسبات ابری از اهمیت بیشتری برخوردار است. محاسبات ابری در سه فرم ارائه می شود: ابر عمومی، ابر خصوصی و ابر ترکیبی. یکی از جنبه های رایانش ابری و مجازی سازی، مدیریت منابع است. مدیریت منابع، یک فرایند است که با تهیه و انتشار منابع سروکار دارد. تکنیک‌های مجازی سازی برای انعطاف پذیری و قانون گذاری روی تقاضای منابع استفاده می‌گردند. روش این است که برای هر تقاضا یک ماشین مجازی ایجاد می‌شود و یا آن را بر روی ماشین مجازی موجود جای می‌دهند. زمانی که کار کامل می شود، همه منابع قابل حصول منتشر شده و قسمت هایی از منابع اشتراکی آزاد خواهند شد. تخصیص منابع روی توافق نامه سطح خدمات (SLA) انجام می‌شود که این توافق نامه بین تولیدکنندگان سرویس و مشتری خواهد بود. SLA شامل جزئیات سطح خدمات است که نیازمند مستاجر می باشد. و همچنین، شامل اطلاعاتی در مورد فرایند پرداخت و جریمه تخلف است[۱].

ساختار بازار محور، یک نمونه از ساختارهای رایانش ابری است که معماری مدیریت منابع واسط سنتی برای چنین سیستم‌هایی مناسب نیستند. معماری واسط سیستم، هیچ انگیزه‌ای را برای ارائه کنندگان خدمات ایجاد نمی‌کند و با همه درخواست‌ها، با اهمیت یکسان برخورد می‌کند. در نتیجه روشن است که تکنیک‌های بازارمحور تدبیری برای محیط های ابری می‌باشد

www.SID.ir

که توانایی قانون گذاری روی منابع و اشتراک گذاری آن بین کاربران را دارد. اینگونه تکنیک‌های مدیریت منابع ، انگیزه اقتصادی را برای ارائه دهندگان خدمات و مشتری فراهم می‌کنند. ارائه دهنده‌های سرویس، منابع شان را در میان چندین متقاضی و مستاجر به شرط پرداخت هزینه به اشتراک می گذارند.

در این مقاله، می‌خواهیم یک بینش ارائه دهیم و تاکید کنیم که تکنیک‌های مدیریت منابع نیازمند اصلاحاتی هستند. اگرچه درباره موضوع مورد بحث، بررسی هایی انجام شده و اهداف با رجوع به جنبه‌ای خاص از تکنیک‌های مدیریت منابع طراحی می‌شوند، اما نیاز به مطالعه‌ای جامع است که دست آورد آن، طبقه بندی تکنیک‌های مدیریت منابع، جزئیات کار آنها، پارامترهای سنجش و بررسی چالش‌ها باشد.

در این مقاله یک طبقه بندی از تکنیک های مدیریت منابع ارائه می‌دهیم که مبتنی بر استانداردهای مدیریت منابع اصلی می‌باشند . معیارهای تحقیق که برای ارائه راه حل‌های مدیریت منابع مورد استفاده قرار می‌گیرند که دربرگیرنده کارایی انرژی، آگاهی توافق سطح سرویس، کمینه سازی بار شبکه ، تعادل بار ، افزایش بهره‌وری، محاسبات ابری ترکیبی می‌باشند.

در ادامه در بخش ۲ خلاصه و مرور کلی بر بررسی‌های موجود را ارائه داده، در بخش۳ رده بندی تکنیک های مدیریت منابع را نشان داده و بررسی میکنیم و در بخش ۴ نتیجه گیری را بیان می‌کنیم.

۲- بررسی‌های موجود

در طی چندین سال گذشته، تکنیک های مدیریت منابع برای محیط های ابری، در تحقیقات مورد توجه قرار گرفته‌اند. تکنیک‌های مدیریت منابع مختلفی برای رسیدگی به چالش های تحقیقات موجود، ابداع شده اند. این درحالی است که منابع و مقالات محدودی، توانایی کمک به محققان در فهم مفاهیم کلیدی و کارکرد تکنیک‌های موجود را دارند. در این بخش ما بر روی مطالعات مختلفی که به هر طریق قصد ارائه بینش روی محیط های ابری را دارند، بحث می‌کنیم.

جنینگز و استادلر [۱] یک بررسی روی مدیریت منابع در پردازش ابری ارائه می‌کنند که در مورد حوزه مدیریت منابع ، انواع مختلف منابع، توانایی تکنولوژی‌ها، نقش های مدیریت منابع و تکنیکهای مختلف مدیریت منابع بحث می‌کند. مولف ها همچنین روی مدیریت ظرفیت کار و بعضی چالش‌های تحقیق بحث می‌کند.

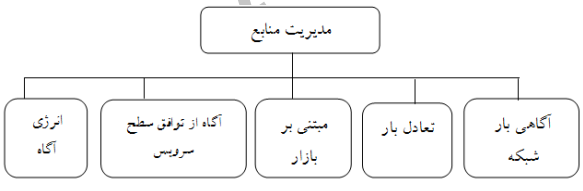
در مرجع [۲]، لین، الگوریتم های زمانبندی مختلف و رده بندی آن ها روی فاکتورهای مختلف، از قبیل زمان، ارزش و انرژی را بررسی می کند. ایده اصلی، یاری کاربران در انتخاب الگوریتم زمان بندی مناسب، با توجه به نوع سرویسی می‌باشد که آنها می‌خواهند استفاده کنند . ریجایلسکی و کوینو [۳] تکنیک‌های مختلف مجازی سازی شبکه و تاثیرشان روی مدیریت منابع آگاه از کیفیت سرویس را مورد بحث قرار می دهند. این مقاله تاثیر تکنیک های مجازی سازی روی کارایی شبکه های مرکز داده را بررسی می‌کند. موضوعات مختلف تحقیق، که در طی مدلسازی کارایی و طرح ریزی تکنیک های مدیریت منابع با آنها مواجه می‌شوند، مورد گفتگو قرار می گیرند.

در مرجع [۱۰]، نویسندگان، معماری، اصول و یک الگوریتم برای انرژی کارآمد و موثر در محیطهای ابری ارائه می‌دهند. بعلاوه، این مقاله در بررسی که ارائه می‌دهد چالش‌های تکنیک‌های تخصیص منبع انرژی کارآمد را مورد پژوهش قرار می دهد. ماروجویک و همکاران[۸] تکنیک های مدیریت منابع متعدد را مورد بحث قرار می دهند که منابع را در ابرهای دارای رادیوی تعریف شده بطریقه نرم افزاری (SDR) مدیریت می کنند. ابر SDR یک ابر بی سیم با مقیاس بزرگ می باشد که سرویس هایی را برای کاربران ارائه می دهد . در این مقاله ، نویسندگان تکنیک های مدیریت منابع ابر SDR را مورد بحث قرار می‌دهند . کانسال و چانا [۱۱] و شاهاپور و جایارخا[۱۲] بررسی‌هایی روی تکنیک های مدیریت منابع تعادل بار ارائه می دهند. آنها استراتژی های گوناگون تعادل بار و پارامترهای ارزیابی را مورد بحث قرار می‌دهند. هیچ یک از مقالات یک رده بندی جامع، کارکرد مفصل و بحث حساس تکنیک های موجود را ارائه نمی‌دهند. در این مقاله، ما روی جنبه‌های برجسته و مهم متمرکز می‌شویم.

۳- تکنیک‌های مدیریت منابع

مدیریت منابع به عنوان یک جنبه از محاسبات ابری، برای ارائه کارایی و استفاده کارامد از سخت افزار اصلی مطرح شده است. در محیط های ابری، اکثر منابع، مجازی شده وهمچنین میان چندین کاربر به اشتراک گذاشته شده‌اند. بیشتر چالش‌های تخصصی در مدیریت منابع، بهینه سازی انرژی، نقض SLA، تعادل بار، سربار شبکه، افزایش بهره وری، ابرهای ترکیبی و یا محاسبات ابری سیار هستند.

بیشتر محققان راه حل‌های بهینه شده چند معیاری را ارائه می‌دهند. بهینه سازی چند معیاری، به بهترین راه حل‌های ممکن که به وسیله چندین استاندارد بهینه سازی مدیریت منابع ارائه شده‌اند، منجر می‌شوند. از این گذشته، بهینه‌سازی چند معیاری می‌تواند ابعاد و چالش های جدیدی را در مدیریت منابع اضافه کند. حتی درصورتی که استانداردهای مطرح شده ناسازگار و ضد و نقیض باشند. این به دلیل ناسازگاری استانداردهای به هم وابسته است و بهینه سازی یک استاندارد، کارایی استاندارد دیگر را پایین می‌آورد. شکل(۱) رده بندی تکنیک‌های مدیریت منابع مبتنی بر بحث‌های بالا را نشان می‌دهد. تکنیک‌ها در طبقه بندی‌های چندگانه براساس مشکل تحقیق و معیار استفاده شده، دسته بندی می‌شوند. در مورد راه حل‌های بهینه‌سازی چندمعیاری، طبقه‌بندی استانداردها، بر اساس استانداردهای مورد توجه، مشخص می‌شود.



شکل (۱): رده بندی تکنیک‌های مدیریت منابع

در زیربخش‌های زیر، ما جزئیات کارهای تکنیک های انتخابی را ارائه می دهیم. به علاوه روی همه تکنیکها و ارائه نظریه برای اصطلاحات و بهبود آن ها تاکید می‌کنیم.

Archive of SID
www.SID.ir

۳-۱- تکنیک های مدیریت منابع آگاه از انرژی

کارایی انرژی یکی از موضوعات اصلی می باشد که نیاز است در سیستم های ابری نشان داده شود. این مورد، بخاطر مصرف انرژی، نه تنها هزینه های نیروی ارائه دهنده های سرویس را افزایش می دهد، بلکه یک نقش در افزایش انتشار مونوکسید کربن بازی می‌کند[۳]. از این گذشته، ماشین ها و سوخت فسیلی استفاده شده برای تولید نیرو، شرکت کنندگان اصلی در انتشار مونو اکسید کربن می باشند.

یک روش بالقوه برای حل این موضوع، تراکم حجم کار می باشد، که در آن مصرف انرژی بوسیله تقویت بیشتر حجم کار روی تعداد کمتر سرویس دهنده ها، کاهش می یابد. ماشین های مجازی میزبان شده، به وضوح روی سرورهای بارگذاری شده، برای حجم کار بالاتر سرورها نسبتا به طور تطبیقی منتقل می شوند، و سرورهای اضافی خاموش می‌شوند. بعضی از تکنیک های مدیریت منابع انرژی آگاه موجود، در بخش بعد مورد بحث قرار می‌گیرند.

مرجع [۱۰] بهترین کاهنده مناسب اصلاح شده(MBFD) مبنی بر الگوریتم بهترین کاهنده مناسب(BFD) می‌باشد که برای بسته بندی مخزن، مورد استفاده قرار می گیرد. الگوریتم MBFD، ماشین‌های مجازی را در ترتیب نزولی براساس نیازهای پردازنده‌شان مرتب می‌کند. بعد از مرتب سازی، همه ماشین های مجازی به میزبان ها بر اساس مدل نیرو تخصیص داده می‌شوند. مدل نیرو، تغییر در مصرف انرژی سرویس دهنده ها را بررسی می‌کند و ماشین مجازی را روی یک سرور جای می دهد که تغییر مصرف انرژی کمینه را نشان دهد. از این گذشته، الگوریتم مقادیر آستانه پویا را برای حفظ کاربرد سرویس دهنده در داخل محدوده و برای اجتناب از تخلفات SLA مورد استفاده قرار می دهد. الگوریتم بهترین کاهنده مناسب سنتی ماشین های مجازی را یا به سرویس دهنده هایی که ظرفیت محاسباتی کمی دارند یا به سرویس دهنده هایی که ظرفیت بلااستفاده بیشتری دارند، تخصیص می‌دهند. مسئله با الگوریتم بهترین کاهنده مناسب این است که مصرف انرژی سرویس دهنده در طی جای دهی ماشین مجازی در نظرگرفته نمی‌شود. از این رو یک ماشین مجازی ممکن است روی سرویس دهنده ای موقعیت داده شود که دارای ظرفیت محاسباتی پایین تری می باشد اما نیروی برق بیشتری را مصرف می‌کند. برای نشان دادن این مشکل، نویسندگان [۱۴] یک الگوریتم بهترین کاهنده مناسب براساس راه حل انرژی کارآمد ارائه می‌دهند. الگوریتم بهترین کاهنده مناسب آگاه به ظرفیت محاسباتی و نیرو (PCA-BFD) موضوع مشخص شده را بوسیله تخصیص دادن ماشین مجازی به یک سرویس دهنده که دارای بیشترین ظرفیت محاسباتی می‌باشد نشان می‌دهند. بوسیله انجام چنین کاری مصرف نیرو الگوریتم بهترین کاهنده مناسب انرژی آگاه محاسباتی و نیرو و تعداد سرویس دهنده‌های استفاده شده به حداقل می‌رسد.

آدیس و همکاران [۱۵] روی معماری سطح سرویس PaaS محاسبات ابری تمرکز می کنند. تمرکز اصلی آنها، حداقل سازی مصرف انرژی و زمان واکنش و حداکثرسازی قابلیت دسترسی می‌باشد. تکنیک مدیریت منابع ارائه شده مبنی بر چهارچوب سلسله مراتبی بحث شده در [۵] می‌باشد، که بهینه سازی غیر خطی منابع در مقابل مقیاس زمانی مختلف انجام می‌شود. مدیر مرکزی، در ریشه سلسله مراتب قرار دارد و وظایف اصلی‌اش را بعد از هر ۲۴ ساعت انجام می‌دهد. از این گذشته آن بخش بندی کلاس و بخش بندی سرویس دهنده را انجام می دهد. بخش بندی یک سرویس دهنده منابع قابل دسترس یک سرویس دهنده را در داخل ماشین های مجازی چندگانه تقسیم می کند. درحالی که هر کار دریافت شده روی یک مجموعه خاص از سرویس دهنده ها مطابق با کلاسش جای داده می‌شود.

در مرجع [۱۶]، نویسندگان حجم کار و تخصیص منابع را به طور پویا به برنامه‌های کاربردی در حال اجرا پیشنهاد می‌دهند. یک کنترل کننده مرکزی برای تخصیص و مدیریت همه منابع مورد استفاده قرار می‌گیرد. تصمیمات مدیریت منابع براساس کارهای ساعتی برای به حداقل رساندن سربار وارد آمده بوسیله تصمیمات، ایجاد می‌شوند. تصمیمات مدیریت منابع، شامل روشن کردن و خاموش کردن سرویس دهنده و انتقال ماشین مجازی از یک سرویس دهنده به سرویس دهنده دیگر می‌باشند و مقدار قابل توجهی از انرژی و منابع شبکه را مصرف می کنند. از این رو این عملیات نمی تواند هر بار انجام شود.

۳-۲- تکنیک های مدیریت منابع آگاه از SLA

کیفیت سرویس یکی از موضوعات اصلی است که باید در حین مدیریت منابع مورد توجه قرار گیرد. زمانی‌که یک ارائه دهنده خدمات، قصد ارائه بیشترین کارایی را برای مشتریان خود دارد، یک قرارداد SLA باید بین هر دو طرف امضا شود. SLA شامل اطلاعاتی درمورد سطوح نیازمندی خدمات و ارزش گذاری بر سرویس است. SLA همچنین شامل یک بند جریمه است که در صورت نقض قرارداد، اجرا می شود. تولید کنندگان باید از تخلفات دوری کنند و در حین ارائه سرویس ها به مشتریان، بررسی هایی را انجام دهند.

آردگنا و همکاران [۱۷] الگوریتم های تخصیص ظرفیت را برای تضمین SLA و مدیریت تغییر پذیری work load ارائه می دهند. الگوریتم‌های ارائه شده با کنترل کننده های منبع پراکنده از لحاظ جغرافیایی، تعامل دارند و همچنین می توانند بار را در صورت مواجهه با تراکم،هدایت کنند. از این گذشته در صورت نیاز، یک برنامه کاربردی روی ماشین های مجازی چندگانه اجرا می شود و حجم کار به طور مساوی روی ماشین های مجازی توزیع می‌شود. به صورتی که حجم های کاری تغییر پذیر می‌باشند، یک پیشگو حجم کار برای پیش بینی الزامات حجم کار آینده مورد استفاده قرار می گیرد و ظرفیت روی نتایج پیش بینی شده تغییر می یابد. از این گذشته، تخلفات SLA بوسیله پایین نگه داشتن زمان پاسخ و واکنش در طی ارتباطات VM میانی لغو می‌شوند. در مورد زمان واکنش بالاتر، VM به ماشین فیزیکی دیگری منتقل می‌شود. هدف اصلی الگوریتم‌های ارائه شده، تهیه حجم‌های کاری تغییرپذیر در حین پایین نگه داشتن تخلفات SLA می‌باشد.

در مرجع [۱۸] تمرکز اصلی محقق بر کاهش منابع بالا/پایین درحال تهیه، بدلیل رخ دادن مدیریت منابع واکنشی است. احتمالات پیش بینی شده مبنی بر چهارچوب کاری در [۴] مورد بحث قرار می گیرند ، که منابع را به وظایف مختلف تخصیص می‌دهند. الگوریتم عوامل مختلف را برای تامین و ذخیره پایانی منابع مورد استفاده قرار می‌دهد. از این گذشته، آن دارای یک ماژول پیشگویی می‌باشد که نیازهای آینده سرویس ها را به شکل اساسی پیش بینی می‌کند. همچنین توانایی دارد منابع را در میان سرویس های متفاوت جریان های کاری برای اجتناب از تخصیص غیر ضروری و آزاد سازی

Archive of SID
www.SID.ir

منابع، مدیریت کند. در مرجع [۱۹]، یک تکنیک تخصیص منابع مبتنی بر کار پیشنهاد می‌شود که روش ماتریس مقایسه زوج سیم و روش سلسله مراتبی تحلیلی را برای رتبه بندی فرآیند تخصیص مورد استفاده قرار می‌دهد. پایه این روش ها، منابع در دسترس و کارایی های کاربر در نظر گرفته میشوند و منابع به کارها بر اساس رتبه بندی مشخص شده تخصیص می یابند . از این گذشته یک روش بر اساس ماتریس انحراف القاء شده برای پیدا کردن عناصر ناسازگار و بهبود نسبت سازگاری برنامه ریزی می‌شود.

در مرجع [۲۰] نویسندگان یک پلت فرم با نام جهت نمای ابر برای محیط به صورت یک سرویس (PaaS) مبتنی بر ابرها معرفی کنند. جهت نمای ابر یک محیط آگاه از SLA می‌باشد که منابع را در تمام چرخه زندگی مدیریت می‌کند. از این گذشته یک انشعاب برای SLA به این پلت فرم به وسیله حفظ در وب معنایی مبنی بر معماری محاسبات ابری اضافه شده است. از این گذشته، حوزه ابر یک مدل SLA ارائه می‌دهد که با استاندارد سطح بالاتر سروکار دارد و می‌تواند به آسانی نیازهای انعطاف پذیر کاربران متعدد را مدیریت کند. بعلاوه، یک چهارچوب ایجاد می‌شود که به طور پویا تخلفات SLA را بوسیله پیکربندی ارتجاعی منابع به حداقل می‌رساند.

تمرکز اصلی تکنیک های ذکرشده، از بین بردن تخلفات SLA می‌باشد. آردگنا و همکاران[۱۷]، وی و همکاران[۱۸]، ارگو و همکاران[۱۹] و گاریکا و همکاران[۲۰] راه حل های انحصاری را برای مسائل مرتبط با SLA ارائه می‌دهند. بنابراین راه حل‌های چند هدفی نه تنها تمرکز روی تخلفات SLA بلکه بررسی چالش‌های دیگر تحقیق را بایستی در نظر گرفته و ارائه کنند، بهینه سازی چند هدفی مانند تخلفات SLA و کاهش انرژی، تخلفات SLA و افزایش بهره وری و تخلفات SLA و بار شبکه می تواند واقعا برای مدیریت جالب باشد. از این گذشته، کارایی تکنیک‌های ذکر شده در بالا برای بهبود بیشتر مورد نیاز می باشند به صورتی که نتایج نشان می دهد تخلفات SLA می‌توانند هنوز ثبت شود .

۳-۳- تکنیک های مدیریت منابع بازار محور

محاسبات ابری نمونه‌ای از ساختار بازار محور می‌باشد و هدف اصلی ارائه دهنده‌های سرویس در این نمونه، افزایش سود می‌باشد. برای این منظور، ارائه دهنده‌های سرویس انواع مختلف راه حل ها را مانند تراکم حجم کاری، اجتناب از تخلف SLA ، مجازی سازی شبکه و حداقل سازی بار شبکه مورد استفاده قرار می‌دهند. همه این راه حل ها مسائل و موضوعات مرتبطی دارند. بنابراین اینجا یک نیاز برای توسعه راه حل ها وجود دارد که بایستی بازار محور و برای ارائه دهنده گان سرویس سودمند باشد. در این بخش ما بعضی از آخرین راه حل های بازار محور را بحث خواهیم کرد که هدف افزایش بازدهی و منفعت ارائه دهندگان سرویس می‌باشد.

یک مزایده مبتنی بر تکنیک تخصیص منابع در [۲۱] برنامه ریزی شده است. این مطلب بحث برانگیز است که معین شود تخصیص منبع بر اساس مزایده می‌تواند برای ارائه دهندگان سرویس در مقایسه با تامین منابع با قیمت ثابت، سودآورتر باشد یا خیر. در تکنیک برنامه ریزی شده کاربر برای منابع اعلام و منتشر شده پیشنهاد قیمت می‌دهد که تنها در صورت برنده شدن در مزایده به او تخصیص داده می‌شود. از این گذشته، تقاضا کاربر در طی تامین منبع در نظر گرفته می شود . قبلا برای تکنیک برنامه ریزی شده ، تقاضای کاربر با تکنیک مبتنی بر مزایده بررسی نمی‌شد.

نویسندگان مرجع [۲۲] یک تکنیک مدیریت منابع چندلایه برای محاسبات ابری برنامه ریزی کرده‌اند. در این تکنیک ، ارائه دهنده نرم افزار به عنوان یک سرویس خدمات را برای کاربرانی که در حال استفاده از سرویس‌های ارائه دهنده‌های زیر بنا به صورت یک سرویس هستند، ارائه می کند. کاربران به ارائه دهنده نرم افزار به عنوان یک سرویس برای سرویس‌هایی که به آنها ارائه می‌شوند، هزینه می پردازند. برای محاسبه مقداری که کاربر باید پرداخت کند، یک تابع بهینه منحصر به فرد مورد استفاده قرار می گیرد که مبنی بر نشان دادن سطح رضایت کاربر می باشد.

هدف اصلی نویسندگان در [۲۳] به حداکثر رساندن بهره وری ارائه دهنده های سرویس بوسیله جلسه توافق سطح سرویس و به حداقل رساندن مصرف انرژی می‌باشد. برای این هدف، یک چارچوب برنامه ریزی می‌شود که هر سرور دارای یک ماژول مقیاس سازی ولتاژ یا فرکانس پویا(DVFS) باشد و فرض می‌شود که یک سرور نمی‌تواند روشن یا خاموش شود و اینجا یک هزینه مشترک انتقال ماشین مجازی وجود دارد. از این گذشته، یک تکنیک بهینه سازی ترکیبی برنامه ریزی می‌شود که قصد دارد مسائل در ارتباط با تعادل بار، مقیاس سازی ولتاژ یا فرکانس پویا، تخصیص منابع و موقعیت دهی سرویس براساس ماشین های مجازی را حل کند.

در مرجع [۲۴]، دو مدل SLA برای سیستم محاسبات ابری برنامه ریزی شده است. مدل SLA طلایی مبنی بر میانگین زمان واکنش، حداکثر میزان سرعت ورودی برای درخواست سرویس گیرنده، میزان پاداش برای هر درخواست سرویس انجام شده می‌باشد و یک جریمه درصورتی که میانگین واکنش درخواست خطا دهد، تعیین می‌کند. مدل توافق سطح سرویس دوم SLA برنز می‌باشد که حداکثر میزان ورودی و تابع مفید را مشخص می‌کند که سود را برای هر درخواست وابسته به زمان واکنش محاسبه می‌کند سپس یک تکنیک با نام تخصیص منبع هدایت نیرو برای مشکل بهینه سازی پیشنهاد داده شده است. در این تکنیک، یک راه حل اولیه به وسیله پردازش سرویس گیرنده ها روی مرتب سازی حریصانه تولید می‌شوند و منابع آنها یک به یک تخصیص داده می‌شوند. سپس میزان ثابت می ماند و اشتراک منابع به وسیله مراحل بهینه سازی بهبود می‌یابد، در نهایت تکنیک تراکم منبع پیشنهاد شده از تحقیق جهت دار نیرو به کار برده می شود.

تکنیک ارائه شده به وسیله نویسندگان در [۲۱،۲۲] تنها بهینه سازی سود و هزینه سرویس می‌باشند. از طرف دیگر، علی و همکاران منفعت ارائه دهنده سرویس را به حداکثر می‌رسانند و هزینه کاربر سرویس را همراه با مصرف انرژی به حداقل می‌رسانند. حداقل سازی مصرف انرژی می‌تواند تاثیر مستقیمی روی منفعت ارائه دهنده سرویس داشته باشد به صورتی که صورت حساب های الکتریکی به حداقل برسد. با این حال ممکن است یک تاثیر منفی داشته باشد اگر حداقل سازی انرژی منجربه تخلفات توافق سطح سرویس شوند و ارائه دهنده سرویس بایستی یک جریمه بپردازد. بنابراین درحین بهینه سازی سود و مصرف انرژی، بایستی همچنین تخلفات SLA در نظر گرفته شود. از این گذشته نویسندگان در [۲۴]، تخلفات SLA را همراه با افزایش بهره وری در نظر می گیرند، ولی آنها چالش های دیگر تحقیق را در نظر نمی‌گیرند. بنابراین، تکنیک های ذکرشده می تواند به وسیله معرفی

Archive of SID

www.SID.ir

معیارها و استانداردهایی مانند تخلفات SLA، بار شبکه و مصرف انرژی بیشتر بهبود یابد.

۳-٤- تکنیک های مدیریت منابع تعادل بار

تعادل بار یک عبارت استفاده شده برای تفسیر مفهوم اشتراک حجم کاری در میان منابع چندگانه می‌باشد[۱۱]. در محیط‌های محاسباتی، ویژگی مهم دیگر سیستم، تعادل بار می‌باشد و وظایف باید بین ماشین‌های فیزیکی مطابق با کاربرد الگوریتم تعادل بار منتقل شوند[۵]. از این گذشته نیاز به ابداع تکنیک های مدیریت منابعی وجود دارد که بایستی به منابع اجازه داده شود که حجم کاری‌شان را به اشتراک بگذارند. محققان برای نشان دادن این موضوع کار می‌کنند و آنها ایده‌های متفاوت دارند. بعضی از تکنیک های خوب شناخته شده و تازه تر در این بخش بحث می‌شوند. یک مدل جدید در [۲۵] برنامه‌ریزی می شود که به طور پویا بار را در میان منابع ابری متعادل می‌کند. به کارگیری منابع سرویس دهنده بررسی می شود، اگر منابع یک سرویس دهنده اضافه بار داشته باشند پس حجم کاری آن سرویس دهنده به سرویس دهنده‌ای منتقل می‌شود که کار کمتری دارد. درحالی که وقتی دو ماشین کم کار باشند حجم کار یک ماشین به ماشین دیگر منتقل می شود و آن ماشین خاموش می شود. برای مدیریت بهتر سرویس دهنده ها و ماشین های مجازی، یک لایه فیزیکی در ساختار درختی جدول درهم سازی توزیع شده(DTH) توسعه می‌یابند.

در [۱۳] یک شبکه عصبی مصنوعی(ANN) بر اساس تکنیک تعادل بار پیشنهاد داده می‌شود. برای توزیع یکسان بار در میان همه سرویس دهنده ها، الگوریتم تکثیر رو به عقب مورد استفاده قرار می‌دهد. تقاضای هر کاربر پیش بینی می شود و منابع مطابق با تقاضای پیش بینی شده تخصیص داده می‌شوند، اما فعالیت سرویس دهنده ها در هر زمان معین، وابسته به تقاضای کاربران در زمان خاص می باشد. در نتیجه فعالیت سرویس دهنده ها به حداقل می‌رسد که منجر به مصرف انرژی کمتر شوند. از این گذشته، ارتباط بین مصرف انرژی و انتشار کربن در این مقاله مشخص می‌شود.

الگوریتم های کمینه تا کمینه و پیشینه تا کمینه برای مدیریت منابع در [۳] برنامه‌ریزی می‌شوند. الگوریتم کمینه تا کمینه با وظایفی شروع می‌شود که هنوز به سرویس دهنده ها تخصیص داده نشده اند . ابتدا زمان تکمیل را برای همه کارها روی سرویس دهنده های متفاوت محاسبه می‌کند. کار با حداقل زمان انجام انتخاب می شود و به سرویس دهنده متناظر تخصیص داده می‌شود. بعد از این تخصیص، کار از لیست کارهای تخصیص داده نشده حذف می شود و زمان انجام محاسبه شده کارهای باقی مانده روی سرویس دهنده‌ای که میزبان یک کار می باشد، به روز می شود. این فرآیند روی تکرار نگه داشته می شود تا لیست خالی شود. از طرف دیگر، الگوریتم پیشینه تا کمینه به همین روش با تنها یک تفاوت کار می‌کند، که آن انتخاب یک کار با حداکثر زمان انجام می‌باشد.

نویسندگان در [۹] یک مدل ارائه می‌دهند که مجموعه‌ای n بعدی منابع هر ماشین مجازی را در نظر می گیرد و عوامل برای کنترل این منابع مورد استفاده قرار می گیرند. همراه با این مدل دو نظریه بر اساس تکنیک‌هایی اختراع می‌شوند که منجر به تعادل بار می‌شوند و منابع به ترتیب تخصیص می‌یابند. تکنیک تعادل بار، بار یک ماشین مجازی را در صورتی که در هر بعد افزایش داشته باشد، کاهش می دهد. قیمت پایداری و حداقل قیمت بی ثباتی به ترتیب ۱ و n می باشند. در حالی که، قیمت حداکثر برای بازی تعادل بار سرویس دهنده $n + 1 + n/m$ و برای بازی موقعیت دهی ماشین مجازی $n + 16/5$ می باشد، که m تعداد سرویس دهنده‌ها می‌باشد.

تکنیک‌های بحث شده در این بخش تعادل بار در میان ماشین های فیزیکی مختلف می‌باشد. تعادل بار همراه با مصرف انرژی یا بار شبکه می‌تواند یک راه حل جالب را ارائه دهد. به صورتی که در سیستم های کارآمد انرژی، تعداد ماشین‌های فیزیکی فعال بوسیله کمک تراکم حجم کار شبکه به حداقل می‌رسند. بنابراین به کاربردن تعادل بار در چنین سناریویی می‌تواند بنابر موقعیت دادن ماشین فیزیکی به سختی کامل شود. از طرف دیگر، تعادل بار انتقال کارها را بین میزبان ها نیاز دارد و این می‌تواند منجر به افزایش بار شبکه شود. بنابراین، درحالی که یک راه حل تعادل بار پیشنهاد می‌شود، پیاده‌سازی سناریو بایستی در نظر گرفته شود و به دقت استانداردهای مدیریت منابع اضافی را انتخاب کند.

۳-۵- مدیریت منابع بار آگاه از شبکه

بار شبکه مقدار ترافیکی می باشد که از طریق یک شبکه در بعضی لحظه‌های مشخص زمان گذرانده می‌شود. در محاسبات ابری، بار شبکه می تواند بخاطر اشتراک اطلاعات بین مدیران منابع ، موقعیت دهی ماشین مجازی، ارتباطات ماشین مجازی و انتقالات ماشین مجازی وارد آید. بارهای بالا شبکه به کاهش کارایی سیستم منجر می شود به صورتی که ارائه دهنده سرویس برای موقعیت دهی ماشین مجازی منتظر خواهد ماند و ارتباطات حساس ماشین مجازی میانی تاخیر خواهد داشت. بنابراین، اینجا یک نیاز برای به حداقل رساندن مقدار ترافیکی وجود دارد که از طریق یک شبکه گذر داده می شود. فعالیت های ذکر شده در بالا که نتایجی در افزایش بار شبکه دارند نمی‌توانند متوقف شوند با این حال، ما می‌توانیم ترافیک تولید شده بوسیله این فعالیت‌ها را به کاهش دهیم. در این بخش، ما درمورد بعضی حالت‌های تکنیک‌های فنی که تلاش می‌کنند بار شبکه را به حداقل برسانند، بحث خواهیم کرد.

در [۷] یک تکنیک برای ارائه دهنده‌های سرویس فراهم شده که یک مرکز داده را انتخاب می کند که می‌تواند در جهت منفعت کاربر، بیشتر خدمت کند. کاربر درخواستی به ارائه دهنده سرویس ارسال کرده و منابع بر اساس الگوریتم تخصیص تطبیقی منابع تخصیص داده می شوند. الگوریتم‌های ذکر شده در بالا یک مرکز داده را بر اساس دو پدیده انتخاب می‌کنند. یکی از این پدیده‌ها فاصله بین کاربر و مرکز داده می‌باشد و پدیده دوم، یک حجم کار روی مرکز داده مشخص می باشد . یک‌بار مرکز داده تصمیم می‌گیرد، کار را روی یکی از سرویس دهنده‌ها با کمک ماشین مجازی جای دهد.

در [۱۴] نویسندگان یک رویکرد کند برنامه ریزی می‌کنند که تنها سودمندترین انتقالات را برای نگه داری موقعیت بهینه ماشین مجازی انجام می‌دهد. الگوریتم تعداد انتقالات را بوسیله انجام دادن انتقال آهسته از موقعیت ماشین مجازی موجود به ماشین مجازی جدید پایین نگه می‌دارد. در اولین قدم، الگوریتم بسته بندی مخزن تداخل پایین(LPBP) لیست سرویس دهنده‌ها بصورت نزولی بر اساس مصرف انرژی مرتب می‌شود . بعد از مرتب سازی سرویس دهنده در بالای لیست بوسیله انتقال یک مجموعه یا حتی

Archive of SID

همه ماشین‌های مجازی‌اش به سرور با کمترین مصرف انرژی تخلیه می‌شود. اگر یک سرویس دهنده نتواند به طور کامل تخلیه شود سپس هیچ انتقالی انجام نمی شود، سرویس دهنده از لیست حذف می شود و الگوریتم سرویس دهنده بعدی را در لیست انتخاب می‌کند.

در [۶] یک سناریو در نظر گرفته می شود که خوشه‌های ماشین‌های مجازی روی یک خوشه از سرویس دهنده موقعیت داده شوند. هر خوشه مجازی تنها یک نوع خاص سرویس را به کاربران ارائه می دهد و آن جوابگوی خوشه مجازی کامل برای ارائه توافق کیفیت سرویس می باشد. از این گذشته نیاز منبع ماشین مجازی ممکن است تغییر کند که بتواند به VMهای پراکنده روی چند سرویس دهنده کمک کند. از این گذشته یک الگوریتم ژنتیک برای تراکم ماشین مجازی ارائه می‌شود که یک حالت سیستم بهینه شده ایجاد می‌کند که طرح ریزی سرویس دهنده ماشین مجازی را ارائه می دهد و منابع را به یک ماشین مجازی تخصیص می‌دهد. در مورد تغییر در تخصیص منابع ماشین مجازی یک حالت سیستم جدید ایجاد می‌شود و سیستم بایستی از حالت قدیمی تر به حالت جدید منتقل شود. با این حال، انتقال از حالت سیستم قدیمی تر به حالت سیستم جدیدتر نتایجی را در سربار سیستم دارد. برای حل این مشکل یک الگوریتم ارائه می‌شود که زمان انتقال را محاسبه می‌کند و هزینه پیکربندی مجدد را به حداقل می‌رساند.

بعضی از تکنیک های بحث شده در این بخش مدیریت بیشتری نسبت به معیارهای تحقیق مدیریت منابع دارند. با این حال، جان و سیم [۷] و مالیک و همکاران [۲] تنها بار شبکه را به حداکثر می‌رسانند و می‌توانند بیشتر برای معرفی بهینه سازی چند معیاری توسعه یابند. الگوریتم بسته بندی مخزن تداخل پایین(LPBP) در [۱۴] و CFMV بارشبکه را همراه با مصرف انرژی به حداقل می‌رسانند. در چنین نوع تکنیک هایی، پیدا شدن راه حل بهینه بخاطر ماهیت ضد و نقیض معیارهای مدیریت منابع می‌تواند سخت باشد. بنابراین، یک مجموعه از راه حل ها شناسایی میشوند و مبتنی بر روی معیار تشخیص بهترین راه حل مناسب انتخاب می شود.

٤- نتیجه

در این مقاله، ما یک رده بندی و بررسی جامع از تکنیک های مدیریت منابع موجود را ارائه دادیم که برای محیطهای ابری ابداع شده‌اند. همچنین یک بحث مفصل روی چالش های تحقیق باز ارائه دادیم و اهدافی را طرح ریزی کردیم که در زمان طرح ریزی یک تکنیک مدیریت منابع بایستی در نظر گرفته شوند و نتیجه گرفتیم که بیشتر تکنیکهای موجود یا راه حل برای استانداردها و معیارهای مدیریت منابع موجود یا حداکثر دو معیار مدیریت منابع ارائه می‌دهند. بنابراین محققین بایستی راه حل‌های چند ضابطه‌ای ارائه دهند که بایستی همچنین قادر باشد استانداردهای ضد و نقیض را مدیریت کند.

اگر ما کارایی یک معیار را بهینه سازی کنیم، کارایی سایر توابع تحت تاثیر قرار می‌گیرد. چنین تابعی می تواند کارایی انرژی، تخلفات توافق سطح سرویس، تخلفات توافق سطح سرویس و حداکثر سازی سود و کارایی انرژی و تعادل بار باشد. این تحقیق و طبقه‌بندی ما می‌تواند یاری دهنده پژوهشگران در زمینه تخصیص منابع در رایانش ابری باشد.

مراجع

[1] Jennings B, Stadler R. Resource management in clouds: survey and research challenges. J Network Syst Manage 2015;23(3):567–619.
[2] Lin CT. Comparative based analysis of scheduling algorithms for RM in cloud computing environment. Int J Comput Sci Eng 2013;1(1):17–23.
[3] Kokilavani T, George Amalarethinam DI. Load balanced min–min algorithm for static meta-task scheduling in grid computing. Int J Comput Appl 2011.
[4] Wei Y, Blake MB. Adaptive service workflow configuration and agent-based virtual RM in the cloud. In: IEEE international conference on cloud engineering (IC2E); 2013.
[5] Nowicki T, Squillante MS, Wu CW. Fundamentals of dynamic decentralized optimization in autonomic computing systems. In: Self-star properties in complex information systems. Springer-Verlag; 2005. p. 204–18.
[6] He L, Zou D, Zhang Z, Chen C, Jin H, Jarvis SA. Developing resource consolidation frameworks for moldable virtual machines in clouds. Future Gener Comput Syst 2014;32:68–81.
[7] Jung G, Sim KM. Agent-based adaptive resource allocation on the cloud computing environment. 40th international conference on parallel processing workshops (ICPPW) 2011:345–51.
[8] Marojevic V, Gomez I, Gilabert PL, Montoro G, Gelonch A. RM implications and strategies for SDR clouds. Analog Integr Circ Sig Process 2012;73(2):473–82.
[9] Ye D, Chen J. Non-cooperative games on multidimensional resource allocation. Future Gener Comput Syst 2013;29:1345–52.
[10] Beloglazov A, Abawajy JH, Buyya R. Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing. Future Gener Comput Syst 2012;28(5):755–68.
[11] Kansal NJ, Chana I. Cloud load balancing techniques: a step towards green computing. IJCSI Int J Comput Sci Issues 2012;9(1):238–46.
[12] Shahapure NH, Jayarekha P. Load balancing in cloud computing: a survey. Int J Adv Eng Technol 2014;6(6):2657–64.
[13] Al Sallami NM, Al Daoud A. Load balancing with neural network. Int J Adv Comput Sci Appl (IJACSA) 2013;4(10):138–45.
[14] Tziritas N, Xu C-Z, Loukopoulos T, Khan SU, Yu Z. Application-aware workload consolidation to minimize both energy consumption and network load in cloud environments. In: 42nd IEEE international conference on parallel processing (ICPP); 2013.
[15] Addis B, Ardagna D, Panicucci B, Squillante MS, Zhang L. A hierarchical approach for the RM of very large cloud platforms. IEEE Trans Dependable Secure Comput 2013;10(5):253–72.
[16] Ardagna D, Panicucci B, Trubian M, Zhang L. Energy-aware autonomic resource allocation in multitier virtualized environments. IEEE Trans Serv Comput 2012;5(1):2–19.
[17] Ardagna D, Casolari S, Colajanni M, Panicucci B. Dual time-scale distributed capacity allocation and load redirect algorithms for cloud systems. J Parallel Distrib Comput 2012;72(6):796–808.
[18]

Archive of SID
www.SID.ir

[19] Wei Y, Blake MB, Saleh I. Adaptive RM for service workflows in cloud environments. In: 2nd international workshop on workflow models, systems, services and applications in the cloud; 2013.
[20] Ergu D, Kou G, Peng Y, Shi Yong, Shi Yu. The analytic hierarchy process: task scheduling and resource allocation in cloud computing environment''. J Supercomput 2013;64(3):835–48.
[21] Garcia AG, Espert IB, Garcی a VH. SLA-driven dynamic cloud resource management. Future Gener Comput Syst 2014;31:1–11.
[22] Zaman S, Grosu D. A combinatorial auction-based mechanism for dynamic VM provisioning and allocation in clouds. IEEE Trans Cloud Comput 2013;1(2):129–41.
[23] Chunlin L, Layuan L. Multi-Layer RM in cloud computing. J Network Syst Manage 2013.
[24] Ali S, Jing Si-Yuan, Kun S. Profit-aware DVFS enabled RM of IaaS cloud. Int J Comput Sci Issues (IJCSI) 2013;10(2):237–47.
[25] Goudarzi H, Pedram M. Profit-maximizing resource allocation for multi-tier cloud computing systems under service level agreements. Large scale network-centric computing systems, Wiley series; 2013.
[26] Ban Y, Chen H, Wang Z. EALARM: an ENHANCE autonomic load-aware RM for P2P key-value stores in cloud. In: 7th IEEE international symposium service-oriented system engineering; 2013.

Archive of SID

www.SID.ir