**به نام خدا**

**توازن بار در ابرهای محاسباتی**

**A comparative study into distributed load balancing algorithms for cloud computing**

**علی - محمد رضایی**

**چکیده:**

اینترنت، از ابتداي آغاز کار خود تاکنون، دست خوش تحولات فراوانی شده است که بعضی از آنها موجب تغيير شيوه زندگی بشر در چند دهه اخير گشته است. یکی از جدیدترین تغييرات در نحوه کارکرد اینترنت، با معرفی رایانش ابري صورت پذیرفته است. این فناوري جدید به دليل ویژگی هایش به سرعت محبوب شده است چرا که در رایانش ابري همه نوع امکانات به کاربران، به عنوان یک سرویس ارائه شده است. طبيعتاً هر تغيير و مفهوم جدیدي در دنياي فناوري، مشکلات و پيچيدگی هاي خاص خود را دارد. بهره گيري از رایانش ابري نيز از این قاعده مستثنی نبوده و چالشهاي فراوانی را پيش روي صاحب نظران این حوزه قرار داده است که از آن جمله میتوان به مواردي نظير: توازن بار، امنيت، قابليت اطمينان، مالکيت، پشتيبان گيري از داده ها و قابليت حمل داده ها اشاره کرد. با توجه به اهميت فرآیند توازن بار در رایانش ابري، هدف این مقاله بررسی این فرآیند و مقایسه روشهاي مطرح در این حوزه میباشد.

گریدهای رایانشی مدل جدیدی برای توسعه رایانشهای موازی سنگین می باشد که هدف آن ایجاد یک ابررایانه انتزاعی با استفاده از منابع آزاددر شبکه های گسترده برای بسترهای نامطمئن چون اینترنت می باشد. از آنجایی که در گریدهای رایانشی بستر ارتباطی و موجودیت منابع بسیار پویا است لذا مدیریت منابع از جمله **توازن بار** از مسائل مهم می باشد.

در این مقاله ابتدا به توضیح و تعریفی از محاسبات ابری و مزایا و معایب آن پرداخته شده است و سپس به شرح توازن بارو بررسی و ارائه ی الگوریتم هایی در زمینه ی فوق و شرح الگوریتم کلونی مورچه پرداخته شده است.

**کلیدواژه ها:**رایانش ابر،چالش درابر،توازن ابر

**1-مقدمه:**

رایانش ابری مدلی بر پایه ی شبکه های بزرگ کامپیوتری مانند اینترنت است که الگویی تازه برای عرضه ،مصرف و تحویل سرویس های فناوری اطلاعات و سایر منابع اشتراکی رایانشی با بکارگیری اینترنت را ارائه می کند.رایانش ابری راهکارهایی برای ارائه خدمات فناوری اطلاعات به شیوه های مشابه با صنایع همگانی پیشنهامی می کند.

دلیل استفاده از واژه ی ابر اینست که پردازش ابری جزئیات فنی اش را از دید کاربران پنهان می سازد ولایه ای از انتزاع را بین این جزییات فنی و کاربران به وجود می آورد.

به عنوان مثال آنچه یک ارائه دهنده ی سرویس نرم افزاری در رایانش ابری ارائه می کند ،برنامه های کاربردی تجاری برخط است که از طریق مرورگر وب یا نرم افزار های دیگر به کاربران ارائه می شود .نرم افزارهای کاربردی و اطلاعات روی سرورها ذخیره می گردند و براساس تقاضا در اختیار کاربران قرار می گیرد.

جزئیات از دید کاربر مخفی می مانند و کاربران نیازی به تخصص یا کنترل در مورد فناوری زیر ساخت ابری که از آن استفاده می کنند ندارند.رایانش ابری را گروهی تغییر الگوواره ای می دانند که دنباله روی تغییری ست که در اوایل دهه 1980 از مدل رایانه بزرگ به مدل کارخواه-کارساز صورت گرفت.[13]

**2-تعریف**

با پیشرفت فناوری اطلاعات نیاز به انجام کارهای محاسباتی درهمه جاوهمه زمان به وجودآمده است.همچنین نیازبه این هست که افراد بتوانند کارهای محاسباتی سنگین خود را بدون داشتن سخت افزارها ونرم افزارهای گران از طریق خدماتی انجام دهند.

رایانش ابری آخرین پاسخ فناوری به این نیازها بوده است.اما از آنجا که تاکنون این فناوری دوران طفولیت خودرامی گذراند هنوزتعریف استاندارد علمی که مورد قبول عام باشد برای آن ارائه نشده است.اما بیشترصاحب نظران برروی قسمتهایی از تعریف این پدیده هم رای هستند.

موسسه ملی فناوری و استانداردها NISTرایانش ابری را اینگونه تعریف می کند:

رایانش ابری مدلی است برای فراهم کردن دسترسی آسان براساس تقاضای کاربر ازطریق شبکه به مجموعه ای از منابع رایانشی قابل تغییروپیکربندی مثل شبکه ها،سرورها،فضای ذخیره سازی ،برنامه های کاربردی و سرویس ها که این دسترسی بتواند باکمترین نیاز به مدیریت منابع ویانیازبه دخالت مستقیم فراهم کننده سرویس به سرعت فراهم شده یا آزاد گردد.

عموما مصرف کننده های رایانش ابری مالک زیرساخت فیزیکی ابرنیستند،بلکه برای اجتناب ازهزینه سرمایه ای آن را از عرضه کنندگان شخص ثالث اجاره می کنند،آنها منابع را در قالب سرویس مصرف می کنند و تنها بهای منابعی که به کارمی برند را می پردازند.[12]

بسیاری از سرویسهای رایانش ابری ارئه شده ،با به کارگیری مدل رایانش همگانی امکان مصرف این سرویسها رابه گونه ای مشابه با صنایع همگانی فراهم می سازند.

این درحالی است که سایر گونه های عرضه کنندگان برمبنای اشتراک سرویسهای خود راعرضه می کنند.به اشتراک گذاردن قدرت رایانشی "مصرف شدنی"میان چند مستاجرمی توانندباعث بهبود نرخ بهره وری شود ،زیرا با این شیوه دیگر سرورها بدون دلیل بیکار نمی مانند .

یک دلیل اینست که رایانه ها به میزان بیشتری مورد استفاده قرار می گیرند زیرا مشتریان رایانش ابری نیازی به محاسبه و تعیین حداکثری برای بار خودندارند.

**2-1-تاریخچه**

پیدایش مفاهیم اساسی رایانش ابری به دهه 1960بازمی گردد زمانی که جان مک کارتی اظهارداشت که رایانش ممکن است روزی به عنوان یکی از صنایع همگانی سازماندهی شود.تقریبا تمام ویژگی های امروزرایانش ابری به همراه مقایسه باصنعت برق و شکل های مصرف عمومی و خصوصی و دولتی و انجمنی راپارک هیل داگلاس درکتابی که باعنوان" مشکل صنعت همگانی رایانه " ارائه نموده است.

متخصصین درسال 1966تنهاخطوط نقطه به نقطه اختصاصی ارائه می کردند اما برخی دیگر شروع به ارائه شبکه های خصوصی و مجازی با کیفیتی مشابه و قیمتهای کمتر نمودند.

نماد ابر برای نمایش نقطه مرزی بین بخشهایی که در حیطه مسئولیت کاربرند و آنهایی که در حیطه مسئولیت عرضه کننده بکار گرفته می شد .رایانش ابری مفهوم ابر رابه گونه ای گسترش می دهدکه سرورها را نیز علاوه برزیرساخت های شبکه در برگیرد.

سایت آمازون با مدرن سازی مرکز داده خودنقش مهمی درگسترش رایانش ابری ایفاکرد.بعدازحباب دات-کام آنها دریافتندکه با تغییر داده های خودکه مانند اغلب شبکه های رایانه ای در بیشتر اوقات تنهااز10%ظرفیت آن استفاده می شد و مابقی ظرفیت برای دوره های کوتاه اوج مصرف در نظر گرفته شده بود-به معماری ابر می توانند بازده داخلی خود را بهبود بخشند.

آمازون ازسال2006امکان دسترسی به سامانه خودازطریق وب سرویسهای آمازون را برپایه رایانش همگانی ارائه کرد.[12]

**3-مقایسه با مدلهای دیگر رایانش**

رایانش ابری اگرچه برخی از ویژگیهایش را از مدلهای رایانشی دیگر به ارث میبرد؛ اما خود متفاوت از آنهاست. برخی از این مدلها عبارتند از:

1. **رایانش شبکه ای**:شکلی از رایانش توزیع شده و رایانش موازی که در آن یک رایانه مجازی بزرگ ازرایانه هایی تشکیل شده است که با جفتگیری ضعیف بهم شبکه شده اند و باهماهنگی با یکدیگر کارمی کنندتاوظایف سنگین رابه انجام برسانند.
2. **رایانش خودمختار** :سامانه های رایانه ای با قابلیت خودمدیریت .
3. **مدل کارخواه کارساز**:بصورت گسترده به هر برنامه ای توزیع یافته ای گفته می شود که بین ارئه دهنده سرویس ودرخواست کننده سرویس تمایزقایل می شود.
4. **رایانه بزرگ**:رایانه های قدرتمند توسط سازمانهای بزرگ برای کاربردهای بحرانی بکاربرده می شود این نوعا شامل پردازشهای حجم زیادداده می باشد.
5. **رایانش همگانی**:عبارتست از بسته بندی منابع رایانشی مانند منابع محاسباتی و ذخیره سازی درقالب سرویس های قابل اندازه گیری به گونه ای مشابه باصنایع همگانی.
6. **نظیربه نظیر**:گونه ای ازمعماری توزیع شده بدون هماهنگی مرکزی است که در آن شرکت کنندگان می توانند در آن واحد عرضه کننده و نیز مصرف کننده منابع باشند.[11]

**4-مزایای اصلی رایانش ابری:**

****

* **چابکی**:کاربر می تواند در زمان نیاز میزان منابع مورد استفاده راکاهش یا افزایش دهد.
* **هزینه** :ادعا می شود که این فناوری هزینه ها را به میزان زیادی کاهش می دهدو هزینه سرمایه ای را به هزینه عملیاتی تبدیل می کند.این به ظاهر موانع ورودبه بازارراکاهش می دهد،زیرارایانش ابر مشتریان رااز مخارج سخت افزارو خدمات و همچنین از درگیری با نصب و نگهداری نرم افزارهای کاربردی به شکل محلی میرهاند**.**

.همچنین هزینه توسعه نرم افزاری راکاهش داده و فرآیندرامقیاس پذیرتر می نماید.

* **نابستگی به دستگاه و مکان**:کاربران می توانند درهرمکانی و باهر دستگاهی به وسیله یک مرورگروب از راه اینترنت به سامانه دسترسی داشته باشند.
* **چندمستاجری**:این ویژگی امکان به اشتراک گذاری منابع و هزینه ها بین گروهی ازکاربران را به وجود می آوردوبدین وسیله موارد زیرراامکان پذیر می سازد:
* متمرکزسازی: زیرساختها در مکانهایی با هزینه کمتر (مثل مکانهایی با هزینه برق یا قیمت کمتر)
* افزایش بکارگیری و کارایی: برای سامانه هایی که اغلب مواقع بیش از 10تا20درصدبکارگیری نمی شوند.
* **قابلیت اطمینان**:درصورتی که از سایتهای چندگانه استفاده شودقابلیت اطمینان افزایش می یابد**.**
* **مقیاس پذیری** :کاربران می توانند درزمان تقاضاوبه صورت داینامیک منابع راتدارک ببینند و نیازی به تدارک پیشین برای زمانهای حداکثربار مصرف منابع نیست.
* **امنیت:**به دلیل تمرکز داده ها و منابع امنیتی بیشتر و پیچیده تر امنیت افزایش می یابد،امانگرانی ها به دلیل ازدست دادن کنترل روی دادهای حساس همچنان پابرجاست.امنیت دررایانش ابری اغلب بیشتریابرابر با سیستم های سنتی می باشد،زیراارئه دهندگان رایانش ابری به منابع اختصاصی امنیتی دسترسی دارند که بیشتر مشتریان ازعهده خرید این منابع برنمی آیند.
* **نگهداری**:به دلیل عدم نیاز به نصب برنامه های کاربردی برای نگهداری آسانتر و با هزینه کمتر انجام می شود.شرکت هایی که سکوهای خودشان را پیاده سازی و اجرا می کنند باید زیرساخت های سخت افزاری و نرم افزاری خودشان راخریداری ونگهداری نمایندو کارمندانی را برای مراقبت ازسیستم استخدام کنند ،همه ی اینها می تواند پرهزینه و زمان بر باشددرحالیکه رایانش ابرنیاز به انجام این کارها را ازمیان می بردهردستگاه ساده که توانایی اتصال وبرقراری ارتباط باسرورراداشته باشد برای استفاده از خدمات رایانش ابرکافی است و می تواند نتایج را بادیگران به اشتراک گذارد.
* **سنجش پذیری**:منابع دررایانش ابری باید قابل اندازه گیری و لازم است که میزان مصرف منابع برای هر کاربر و هر منبع براساس واحدهای ساعتی ،روزانه،هفتگی،ماهانه اندازه گرفت.[12],[11]

**5-چالش ها**

* **آسیب پذیری دربرابر رکوداقتصادی:**

مدل خدمات رایانه ای در مقابل رکورداقتصادی بسیارآسیب پذیراست.

همانگونه که شرکت ها در طی یک رکود محتاطانه عمل می کنند ،هزینه های صرف شده برای خدمات رایانه ای را نیز کاهش می دهند.

* **شکل جدید نرم افزارها**:

متخصصین نرم افزاردرراه ایجاد نرم افزاری که میلیون ها کاربربه جای اجرای آن برروی کامپیوترهای شخصی خود بتوانندازآن مانند یک سرویس استفاده کنند ،باچالش های متعدد جدیدی مواجه شده اند.

* **پذیرش:**

این رویکردنسبتاتازه است ودربسیاری مواردهنوزپذیرفته نشده است.دپارتمان های IT هنوزنسبت به آن بسیارمحتاط عمل می کنند زیراسکوی رایانش ابرتوسط آنها کنترل نخواهد شد .تاکنون سرمایه گذارنی که جرات سرمایه گذاری درپروژه های مخاطره آمیز رادارند پول زیادی در رایانش ابرسرمایه گذاری نکرده اند.توانایی کنترل هزینه ها و تهیه و تدارک زیرساخت ها بهنگام نیاز به ویژه باعث جذب کسب و کارهای جدیدی که منابع کمتری در اختیارداشتند شد.همچنین شرکت های Web2.0که درحالت عادی منابع کمتری دارندوبه دنبال کسب توانایی افزایش یاکاهش آسان تقاضا بهنگام نیاز هستند.

شرکت های بزرگتر که عموما صبر می کنند تاتکنولوژی جدیدپذیرفته شوند،ازبرای پروژه های موقت و گاه و بیگاهی استفاده می کنندکه منابع اضافی زیادی رامی طلبند.مثل همه رویکردهای تازه پدیدار شده ،میزانی ازبیم ،عدم اطمینان و قطعیت،و نگرانی هایی درباره بالندگی این تکنولوژی وجوددارد.

* **کنترل:**

ارائه دهندگان خدمات،معمولا سکوها رابرای پشتیبانی از شیوه های تجاری وIT ی یک شرکت خاص طراحی نمی کنند.همچنین کاربران قادربه تغییر تکنولوژی سکوها به هنگام نیاز نخواهند بود.گرچه ارائه دهندگان می توانند با توجه به اینکه چه تکنولوژی ای به بهترین نحو نیازها راپاسخ می دهد و بهنگام نیازآن راتغییردهندکه این کاربدون موافقت یا رضایت مشتریان انجام می گیرد.

* **هزینه های پهنای باند:**

به لطف پهنای باند بالای شبکه ،کاربرحتی هنگامی که درحال استفاده ازوب به عنوان یک کامپیوترفراگیراست،احساس کاربرروی سیستم محلی راداردبااین حال مشکل زیر پیش می آید:

درحالیکه شرکت ها به کمک رایانش ابر می توانند درهزینه ی تجهیزات و نرم افزار ها صرفه جویی کنند اما باید متحمل هزینه ی شارژ بالاتری برای پهنای باندبشوند.احتمالا هزینه ی پهنای باند باید برای نرم افزار های کاربردی مبتنی بر وب کوچک که داده-متمرکز نیستندکمترخواهد بود،اما هنگامی که مثلا یک شرکت پایگاه داده ای چندترابایتی راازطریق رایانش ابراجرامی کند این هزینه می تواند بسیار بالا باشد.

* **محبوس شدن توسط ارائه دهندگان و استانداردها**

نیاز به استانداردهای باز برای تمام شیوه های استفاده از وب به عنوان یک کامپیوتر فراگیروجوددارد با افزایش تعداد ارائه دهندگان خدمات ابری ،اهمیت جابجایی بیشترخواهدشد.اگرشرکتی از خدمات یکی از ارائه کنندگان ناراضی باشد یا اگرفروشنده از این کسب وکارکناربکشدنمی تواندلزوماآسان و باهزینه ای کم،به ارائه دهنده ی دیگر منتقل شود و یااینکه خدمات مذکوردوباره به درون شرکت برگرداند.درعوض ،شرکت باید داده ها و نرم افزارهای کاربردی اش را قالب بندی مجددنموده و آنهارابه یک ارائه دهنده ی جدید منتقل کند که فرآیندی بالقوه پیچیده است.

واگربخواهدخدمات رابه درون شرکت بیاوردباید کارمندانی راکه واجدمهارت لازم برای کاربا تکنولوژی هستند استخدام کند.

کاربران به طور روزافزون به وب و ارائه دهندگان آن وابسته خواهند شد به این ترتیب هنگامی که ارائه دهندگان خدمات شرایط استفاده ازخدمات ویاروش های عملیاتی خود رابعد ازمدتی تغییربدهند،کاربران آنها احساس به دام افتادن و درماندگی می کنند.برای مثال تحمیل محدودیت های جدیدبراستفاده از یک قابلیت و یا از کارانداختن آن به مدت چندماه بمنظور بهبود بخشیدن به آن.

همچنین ممکن است ارائه دهندگان تصمیم به حذف یک قابلیت که سال ها درسایت رایگان ارائه می شد اما در مقابل بخش بهادار خودراحفظ کندوحتی افزایش قیمت بدهد.

* **شفافیت دسترسی**

اگرشرکت ها نتوانند نشان دهند که چه کسی به داده های مشتریان دسترسی دارد و چگونه مانع دستیابی کارمندان غیرمجازبه اطلاعات می شوند نخواهند توانست ازحسابرسی ظرفیت های خود،به وسیله مشتریان آینده با موفقیت بیرون بیایند.ارائه دهندگان رایانش ابر این نگرانی رابه کمک نظارت قبلی third partyها برسیستم و به وسیله ی مستند سازی رویه های طراحی شده برای پاسخگویی به نیازهای امنیت داده برای مشتریان رفع می کنند.

* **قابلیت اطمینان:**

رایانش ابر همیشه قابلیت اطمینان مسمری راارائه نکرده است.مثلا مشتریان Salesforce.comدرتاریخ 12فوریه 2008به مدت 6 ساعت قادر به دریافت خدمات نبودندو سه روز بعد خدمات sS3 Amazon,EC2به مدت 3 ساعت دچاروقفه شدند.

* **حفظ حریم خصوصی:**

طرفداران حفظ حریم های خصوصی مدل ابر رامورد انتقاد قرار می دهند زیرا ارائه دهندگان سرویسهای ابر می توانند کنترل و نظارت کامل قانونی و یاغیرقانونی برروی داده ها و ارتباطات بین کاربران سرویس و میزبان ابرداشته باشند.رویدادهایی همچون برنامه مخفی آژانس امنیت ملی ایالات متحده آمریکا به همراه شرکتهای T&AT,Verizonکه بیش ازده میلیون مکالمه تلفنی شهروندان امریکایی راضبط نمودند،باعث بوجودآمدن بی اعتمادی میان طرفداران حفظ حریم خصوصی شده است.

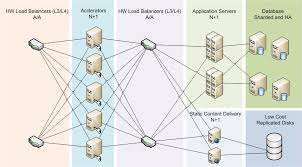
* **امنیت:**

امنیت نسبی رایانش ابری موضوعی بحث برانگیز است که ممکن است پذیرش ابری رابه تاخیربیندازد.گروهی براین باورند که امنیت داده ها وقتی که درداخل سازمان اداره شوند بالاتر است،درحالی که گروهی دیگر عقیده دارند که

ارائه دهندگان سرویس انگیزه ای قوی برای حفظ اعتماد دارندو از این رو سطح امنیت بالاتری را بکار می گیرند.

* **میزان دردسترس بودن و کارایی:**

علاوه برامنیت داده ها ،میزان دردسترس بودن و کارایی برنامه های کاربردی که روی ابر میزبانی می شوند برای کاربران از اهمیت بالایی برخورداراست. [12],[11]



**6-توازن بار**

رایانش ابري یکی از جدیدترین تحولات در فناوري اطلاعات است و با مرور زمان فراگير میشود. از دید حضور در بازار، درک تأثيرات توازن بار در ابر، مهم است. بستر رایانش ابري، یک بستر سرویس دهنده کاملاً اتوماتيک است که به کاربران اجازه ي خرید، ایجاد از راه دور، مقياس پذیري پویا و مدیریت سيستم را میدهد در حال حاضر توازن بار در سيستمهاي رایانش ابري یک چالش میباشد.

هميشه یک راه حل توزیعی مورد نياز میباشد، زیرا همواره نگهداري یک یا چند سرور بی کار و غير فعال تنها براي انجام برخی از خواسته هاي مورد نياز، امکان پذیر نمی باشد و یااینکه به صرفه نيست.

واضح است که به دليل مقياس و پيچيدگی این سيستمها انتساب متمرکز کارها به سرورهاي خاص غيرممکن است . براي مدیریت مناسب منابع فراهم کننده سرویس، به توازن بار نيازمندیم که به ارائه دهنده ي سرویس پيشنهاد میشود.

**6-1-توازن بار چیست؟**

سطح سنجی بار معمولاً براي تداوم یک سرویس، هنگامی که یک یا اجزاي بيشتر مجموعه دچار شکست میشوند، مکانيزه میشود

. اجزا به طور مرتب پایش می شوند و هنگامی که یکی از اجزا پاسخ ندهد، موازنه کننده بار بالا میآید و ترافيک را روي آن نمی فرستد. باسطح سنجی مناسب بار در مکان مصرف منابع اغلب می توان مشکلات را به حداقل رساند که نه تنها هزینه را کاهش و رایانش سبز را ایجاد میکند، بلکه فشار را روي مدارهاي منحصربه فرد پایين نگه میدارد که عمر آنها را به طور بالقوه طولانی تر میکند.

**در واقع می توان گفت هدف از توازن بار،يافتن نگاشتي مناسب از كارها بر روي پردازند ه هاي موجود در سيستم است به طوري كه در هر پردازنده مقدار تقريباً مساوي از كارها اجرا گردد تا زمان اجراي كلي به كمترين مقدار خود برسد**.[4]

**6-2-اهمیت توازن بار**

با توازن بار، میتوان بار را با انتقال پویاي حجم کار محلی از یک ماشين به ماشينی در گره راه دور و یا ماشينی که کمتر مورد استفاده قرار میگيرد،موازنه کرد. این عمل رضایت کاربر را حداکثر، زمان پاسخ را حداقل، بهره برداري از منابع را افزایش، تعداد رد کارها را کاهش و کارایی سيستم را بالامیبرد.[2]

توازن بار هم چنين براي دستیابی به رایانش سبز در ابرها مورد نياز است

.

**عوامل مسئول عبارتند از:**

**مصرف انرژی محدود** : توازن بار میتواند با اجتناب از تعامل بيش از حد گره ها یا ماشينهاي مجازي با توجه به حجم کار، مقدار انرژي مصرفی را کاهش دهد.

 **کاهش انتشار کربن** : مصرف انرژي و انتشار کربن دو روي یک سکه اند. هر دو به طور مستقيم با یکدیگر متناسب هستند. توازن بار به کاهش مصرف انرژي کمک میکند که به طور خودکار منجر به کاهش انتشار کربن میشود و بنابراین به رایانش سبز میرسيم.[7]

**اهداف توازن بار**

بهبود قابل توجه کارایی

داشتن یک برنامه و طرح پشتيبان در مواقعی که سيستم و یا حتی بخشی از آن دچار شکست میشود.

 براي حفظ پایداري سيستم

 براي تطبيق اصلاحات آتی در سيستم

**6-3-انواع الگوريتمهای توازن بار[3]**

بر اساس اینکه چه کسی روند توازن بار را آغاز میکند، الگوریتمهاي توازن بار میتوانند در سه دسته بندي قرار گيرند:

 **فرستند آغاز کننده باشد** (Sender Initiated) : اگر الگوریتمهاي توازن بار توسط فرستنده آغاز شده باشد.

 **گيرنده آغاز کننده باشد** (Receiver Initiated) : اگر الگوریتمهاي توازن بار توسط گيرنده آغاز شده باشد.

 **متقارن** (Symmetric) : ترکيبی از هر دو حالت بالا

بر اساس حالت جاري سيستم، الگوریتمهاي توازن بار میتوانند در دو دسته قرار بگيرند:

** ایستا** (Static) : بستگی به وضعيت جاري سيستم ندارد و نياز به دانش و آگاهی قبلی از سيستم میباشد.

 **پویا** (Dynamic) : نسبت به وضعيت جاري سيستم روي توازن بار تصميم گيري میشود.

**6-3-1-توازن بار ايستا**(Static)

توازن بار ایستا، یک کار داده شده را به یک پردازنده یا گره ثابت تخصيص میدهد. هر زمانی که سيستم دوباره راه اندازي شود، همان پردازنده ي وظيفه ي متصل شده (Binding) ، بدون در نظر گرفتن تغييراتی که ممکن است در طول دوران زندگی سيستم رخ دهد، استفاده میشود. ممکن است تخصيص وظيفه به همان پردازنده، نتيجه ندهد.

اما اختصاص کارهایی که تازه از راه میرسند، در یک ترتيب یا مدل ثابت است . کارایی ماشينهاي مجازي، زمان رسيدن کار تعيين میگردد.

پردازنده رئيس Master حجم کاري را به دیگر پردازندههاي برده (Slave) بر اساس کارایی آنها،تخصيص میدهد. کار اختصاص داده شده به این ترتيب توسط پردازنده برده اجرا شده و نتيجه به پردازنده رئيس برگردانده میشود.[8]

الگوریتمهاي توازن بار ایستا وظایف را به گره تنها بر اساس توانایی گره براي پردازش درخواست جدید تخصيص میدهند. این فرآیند صرفاًبر اساس دانش قبلی از خواص گره است که میتواند شامل : قدرت پردازش گره، حافظه، ظرفيت ذخيره سازي و جدیدترین کارایی ارتباطاتی شناخته شده باشد. اگرچه ممکن است شامل دانش کارایی ارتباطی قبلی نيز باشد. الگوریتمهاي ایستا معمولاً تغييرات پویاي زمان اجراي این صفات را در نظرنمیگيرند. علاوه بر این، این نوع الگوریتمها نمیتوانند با تغييرات بار در زمان اجرا سازگار شوند هدف آنها به حداقل رساندن زمان اجراي کار ومحدود کردن سربار ارتباطات و تأخير است.[2]

**6-3-2-توازن بار پويا**(Dynamic)

در الگوریتمهاي توازن بار پویا حجم کاري، در زمان اجرا ميان پردازندهها، توزیع میشود. پردازنده رئيس، فرآیند جدید را بر اساس اطلاعات جمع آوري شده ي جدید به پردازنده برده تخصيص میدهد. الگوریتمهاي توازن بار پویا ویژگیهاي مختلف گرهها، تواناییها و پهناي باند شبکه، را به حساب میآورند.

بيشتر این الگوریتمها بر ترکيبی از اطلاعات که از قبل درباره ي گره ها در ابر جمع آوري شده و خصوصيات زمان اجراي جمع آوري شده، تکيه میکنند. به طوري که گره هاي انتخابی اجزاي وظایف را پردازش میکنند.

این الگوریتم وظایف را تخصيص داده و ممکن است آنها را به صورت پویا دوباره به گرهها بر اساس ویژگیهاي جمع آوري و محاسبه شده، تخصيص دهد. این الگوریتمها به نظارت مستمر گرهها و جریان وظایف نياز دارند و معمولاً براي پياده سازي مشکلتر هستند. اما این الگوریتمها دقت بالاتري دارند و میتوانند نتایج توازن بار بهينه تري توليد کنند. این الگوریتمها اجازه ي پيشدستی پردازشها را میدهد که در روش توازن بار ایستا، پشتيبانی نمیشود.

یک مزیت مهم توازن بار پویا این است که تصميم توازن بار بر اساس حالت فعلی سيستم است که به بهبود کارایی کلی سيستم با مهاجرت پویاي بار، کمک میکند .یک استراتژي پویا معمولاًچندین بار اجرا میشود و ممکن است یک کار زمانبندي شده را دوباره به یک گره جدید بر اساس حالت پویاي محيط سيستم، تخصيص بدهد.[9]

توازن بار پویا میتواند به دو روش متفاوت انجام شود : توزیع شده (Distributed) و توزیع نشده (Non- distributed) . در نوع توزیع شده، الگوریتمهاي توازن بار پویا توسط تمام گره هاي موجود در سيستم اجرا میشود و وظيفه توازن بار ميان تمامی آنها تقسيم میشود. تعامل ميان گرهها براي رسيدن به توازن بار میتواند به دو شکل باشد : همکاري و بدون همکاري .[5]

در فرم اوّلی گره ها در کنار هم براي رسيدن به یک هدف مشترک کار میکنند. براي مثال، بهبود زمان پاسخ کل، و در فرم دوّمی هر گره به صورت مستقل در جهت یک هدف محلی کار میکند مثل بهبود زمان پاسخ یک کار محلی

الگوریتمهاي توازن بار پویا با ماهيت توزیع شده، پيامهاي بيشتري را نسبت به نوع توزیع نشده توليد میکنند به خاطراینکه هر گره درون سيستم نياز به برقراري ارتباط با هر گره دیگر دارد. یک مزیت این روش در این است که اگر یک یا تعدادي از گرهها درون این سيستم خراب (Fail) شوند، باعث نمیشود که کل فرآیند توازن بار متوقف شود، در عوض تا حدي بر روي کارایی سيستم تأثير می گذارد [12].توازن بار پویاي توزیع شده میتواند روي یک سيستمی که در آن هر گره به تبادل اطلاعات وضعيتی با هر گره ي دیگر درون سيستم نياز دارد، فشار زیادي را ایجاد میکند. این روش زمانی سودمندتر است که بيشتر گره ها به صورت انفرادي با تعامل اندکی با گره هاي دیگر کارمی کنند.[2]

در نوع توزیع نشده، یک گره و یا یک گروهی از گره ها وظيفه توازن بار را انجام می دهند. الگوریتمهاي توازن بار پویاي توزیع نشده میتواند به دو نوع متمرکز ) Centralized و نيمه توزیع شده Semi-Distributed ( باشد.

در نوع متمرکز، الگوریتم توازن بار فقط بر روي یک گره سيستم )گره مرکزي( اجرا میشود. این گره به تنهایی وظيفه توازن بار کل سيستم را به عهده دارد. گرههاي دیگر فقط با گره مرکزي تعامل دارند. درنوع نيمه توزیع شده، گرههاي سيستم به خوشه هایی تقسيم شده اند که در آن توازن بار هر خوشه به صورت متمرکز می باشد. [12]

به وسيله ي تکنيکهاي مناسب براي هر خوشه یک گره مرکزي انتخاب میشود تا توازن بار را در خوشه مراقبت کند. از این رو توازن بار کل سيستم توسط گره هاي مرکزي هر خوشه انجام می شودتوازن بار پویاي متمرکز، پيامهاي کمتري را دریافت میکند.

از این رو از تعداد کل تعاملات درون سيستم نسبت به نوع نيمه توزیع شده،کاسته می شود. با این حال الگوریتمهاي متمرکز میتوانند روي گره مرکزي گلوگاه ایجاد کنند و هم چنين روند توازن بار هنگامی که گره مرکزي سقوط می کند، بی نتيجه می شود. بنابراین این الگوریتم براي شبکه هایی با اندازه ي کوچک مناسبتر است[2].

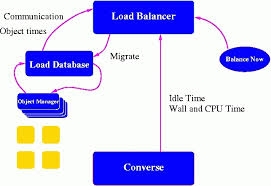
**6-4-مسأله توازن بار را مي توان به صورت يك فرآيند تصميم گيري چهار مرحله اي دانست:**

- تعيين زمان مهاجرت كارها

- تعيين محل و اينكه كدام پردازنده تقاضاي مهاجرت كار را بيان ميكند

- تعداد كارهايي كه بايد مهاجرت كنند

- در نهايت اينكه كدا ميك از كارها مي بايست انتقال يابند.



**6-5-معيارهای الگوريتمهای توازن بار**[1]

**سربار مرتبط**: هنگامی که یک الگوریتم توازن بار اجرا میشود مقدار سربار درگير را تعيين می کند. این شامل سربار مقرر براي بهبودوظایف، ارتباطات ميان پردازندهها و پردازشها میشود. این مقدار باید به حداقل برسد تا تکنيک توازن بار به طور کارآمد عمل نماید.

**توان عملياتی**: براي محاسبه ي وظایفی که اجراي آنها به پایان رسيده، استفاده میشود. این مقدار باید براي بهبود کارایی سيستم بالاباشد.

**کارایی**: براي کنترل بهره وري سيستم استفاده میشود. این معيار باید هزینه هاي معقول را بهبود بخشد. به عنوان مثال، زمان پاسخ وظایف راکاهش دهد در حالی که تأخيرها را قابل قبول نگه دارد.

**بهره برداری از منابع:** براي کنترل بهره برداري از منابع استفاده میشود. این مقدار باید براي توازن بار کارآمد بهينه شود.

**مقياس پذیری**: توانایی یک الگوریتم براي اجراي توازن بار براي یک سيستم با تعداد متناهی گره است. این معيار باید بهبود داده شود.

**زمان پاسخ:** مقدار زمان گرفته شده براي پاسخ دادن، توسط یک الگوریتم توازن بار خاص در یک سيستم توزیع شده است. این پارامترباید به حداقل رسانده شود.

**تحمل خطا**: توانایی یک الگوریتم براي اجراي یکنواخت توازن بار در موارد خرابی اتصال است. توازن بار باید یک تکنيک تحمل خطاي خوب باشد.

**زمان مهاجرت**: زمانی براي مهاجرت کارها یا منابع از یک گره به گره دیگري است. این زمان براي افزایش کارایی سيستم باید به حداقل رسانده شود.

**انتشار کربن**: انتشار کربن همه ي منابع در سيستم را محاسبه میکند، مصرف انرژي بيشتر، انتشار کربن بالاتري دارد. بنابراین براي یک راه حل با توازن بار انرژي کارآمد، این مقدار باید کاهش داده شود.

**6-6-الگوریتمها و روشهای توازن بار**[7]

**الگوریتم خوشه بندی فعال**

بر اساس گروه بندي گرهه اي شبيه به هم کار میکنند. فرایندها شامل: یک گره فرآیند را آغاز می کند واز بين گره هاي همسایه معتبر خود یک گره متفاوت نسبت به قبلی به عنوان گره واسط انتخاب می کند. گره واسط پس از آن یک ارتباط بين یک همسایه هم نوع گره اوّليه آن شکل می دهد. سپس گره واسط ارتباط گره اوّليه با خودش را جدا میکند.

مجموعه فرآیندهاي بالامرتب تکرار می شوند .در این الگوریتم، کارایی سيستم با افزایش منابع افزایش می یابد در نتيجه توان عملياتی با استفاده از این منابع کارآمد بالا می رود. این الگوریتم، با افزایش تنوع (Diversity) سيستم، تنزل پيدا می کند.

**الگوریتم کاوش زنبور عسل**

حالتی از توازن بار در وب سرورها افزایش یا کاهش تقاضا میباشد، سرویسهاي تخصيصی به صورت پویا تقاضاهاي کاربر را تنظيم می کنند. این سرورها تحت عنوان سرورهاي مجازي گروه بندي شده اند و هر سرور مجازي صف هاي سرویس مجازي خود را دارد. هر سرور با پردازش یک درخواست و تقاضا از صف، مقدار منافع و مزایا مورد نياز را محاسبه می کند، مشابه آنچه زنبورهاي عسل در حرکات موزون از خود نشان می دهند یک اندازه گيري از این مزایا مقدار زمان مصرفی پردازنده براي پردازش یک درخواست میباشد. قسمت رقص زنبور عسل مشابه تابلو آگهی در اینجا می باشد. همچنين این صفحه براي تبليغ مزایا )منافع مورد نياز(در کل کلونی استفاده می شود. هرکدام از این سرورها نقش یک کاوشگر و یا یک دیده بان )پيشاهنگی( رامی گيرد.[12]

سرور پس از پردازش یک درخواست میتواند مزایا و منافع را به تابلوهاي آگهی با احتمال rp ارسال کند. یک سرور میتواند یک صف از سرورهاي مجازي رابا احتمال x p انتخاب کند ) xp نشان دهنده رفتار کاوشگر( و یا میتواند آگهی ها را بررسی کند )دیدن رقص( و سرویس دهی کند که بنابراین رفتار پيشاهنگی را نشان میدهد.

یک سرور براي سرویس دهی یک درخواست، پس از محاسبه منافع و مقایسه ي آن با کل منافع کلونی، مقدار xp را براي آن در نظر میگيرد. اگر این منافع بالا بود سرور در همان سرور مجازي جاري باقی میماند و یک آگهی بااحتمال rp براي آن ارسال میشود، اگر منافع کم باشد سرور نقش پيشاهنگی و یا کاوشگري را میگيرد.این روش به توازن بار عمومی از طریق فعاليتهاي سرورهاي محلی دست پيدا میکند.

با افزایش تنوع سيستم، کارایی سيستم افزایش مییابد. اما توان عملياتی با افزایش اندازه سيستم، افزایش نمییابد. براي شرایطی مناسب است که در آن جمعيت متنوعی از انواع سرویس مورد نياز است.[7]

**گرایش نمونه گيری تصادفی**

در اینجا یک گراف مجازي ساخته شده است که در آن هر گره )سرور به عنوان گره( یک اتصال براي نمایش بار سرور دارد. در گراف، هر سرور به صورت یک گره با هر درجه اي به منابع آزاد سرور هدایت میشود. هرگاه یک گره یک کاررا انجام دهد، آن از لبه هاي ورودي حذف می شود که نشان دهنده ي کاهش دسترسی به منبع آزاد میشود.

بعد از تکميل یک کار، گره یک لبه ورودي ایجاد می کند که نشان دهنده ي افزایش دسترسی به منبع آزاد می باشد. افزودن و حذف فرآیند به صورت نمونه گيري تصادفی انجام می شود.

در ابتداي مسير در هر گره به صورت تصادفی یک همسایه انتخاب می شود. آخرین گره، براي تخصيص بار انتخاب شده است. همچنين میتوان روش دیگري براي انتخاب گره جهت تخصيص بار استفاده شود و انتخاب گره بر اساس معيارهاي خاص باشد.

با این حال انتخاب و تخصيص بار میتواند براي گرههاي کم بار باشد.

در نهایت آنچه که ما بدست میآوریم یک گراف جهتدار می باشد این روش، بار را در عرض تمام گرههاي سيستم متعادل می کند. کارایی سيستم با جمعيت مشابه و بالاي منابع افزایش می یابد، درنتيجه توان عملياتی با افزایش منابع سيستم، افزایش مییابد طرح توازن بار در اینجا کاملاً غير متمرکز می باشد؛ لذا بکار گيري آن درسيستمهاي شبکه هاي بزرگ مثل ابر مناسب می باشد.[14]

**روش کارتن**

براي کنترل استفاده از توازن بار و نرخ توزیع شدهي محدود (DRL) ، مکانيسم کارتن را براي ابر، پيشنهاد کرده است.

توازن بار براي توزیع برابر کارها به سرورهاي متفاوت استفاده می شود، بنابراین هزینه هاي مربوطه می تواند کاهش یابد و DRL براي حصول اطمينان از اینکه منابع به روشی توزیع شده اند که تخصيص عادلانه ي منابع حفظ میشود، استفاده می شود. با سربار محاسباتی و ارتباطی خيلی کم، این الگوریتم ساده و براي پياده سازي آسان است. محيط استفاده از این الگویتم، چارچوب متحد براي کنترل ابر است .ازميان معيارهاي توازن بار، معيار سربار و بهره برداري از منابع در نظر گرفته شده است.[6]

**روش رویداد محور**

یک الگوریتم توازن بار رویداد محور را براي بازيهاي آنلاین چند بازي کننده حجيم زمان واقعی (MMOG)پيشنهاد کرده است.

این الگوریتم پس از دریافت رویدادهاي ظرفيتی به عنوان ورودي، اجزاي خودش را در زمينه ي منابع و حالت کلی نشست بازي (Game Session) ، تجزیه و تحليل می کند. در نتيجه فعاليتهاي توازن بار، نشست بازي را توليد می کند. این روش قادراست که مقياس یک نشست بازي را روي چندین منبع بر اساس بار متغير کاربر کاهش یا افزایش دهد اما نقض کيفيت سرویس گاه و بی گاهی دارد.در این الگوریتم از ميان معيارهاي توازن بار، معيار بهره برداري از منابع در نظر گرفته شده است .[7]

**توازن بار مبتنی بر سرور برای سرویسهای توزیع شدهی اینترنتی**

یک سياست توازن بار مبتنی بر سرویس جدید براي سرورهاي وب که در سرتاسر جهان توزیع شده اند پيشنهاد شده است.

این سياست به کاهش زمان سرویس با استفاده از محدود کردن تعدادتغيير مسيرهاي یک درخواست به نزدیکترین سرور راه دور بدون سربار کردن آنها، کمک میکند. یک ميان افزار براي پياده سازي این پروتکل توصيف شده است، همچنين از یک روش اکتشافی براي کمک به سرورهاي وب به منظور تحمل بار استفاده می کند[18].

**منطق فازی**

یک الگوریتم توازن بار مبتنی بر نوبت گردشی را در محيطهاي ماشين مجازي در رایانش ابري، به منظور دستیابی به زمان پاسخ و زمان پردازش بهتر، طراحی کرده است. الگوریتم توازن بار قبل از اینکه سرورهاي پردازشی از راه برسند، انجام شده است.

کار براساس پارامترهاي مختلفی مانند سرعت پردازنده، بار تخصيص داده شده به ماشين مجازي و غيره، زمانبندي شده است. این الگوریتم اطلاعات داخل هر ماشين مجازي و تعداد درخواستهاي جاري تخصيص داده شده به ماشين مجازي سيستم را نگهداري می کند. زمانی که یک کار براي تخصيص میآید این الگوریتم ماشينی با حداقل بار را مشخص میکند و اگر بيشتر از یک ماشين با این ویژگی وجود داشته باشد، این الگوریتم ماشين اول را مشخص میکند. نویسندگان تلاش کردند که تکنيک توازن بار جدید را بر اساس منطق فازي پياده سازي کنند.

از آنجایی که منطق فازي مانند زبان طبيعی است میتوانند مسائل خود را تدوین و فرموله نمایند. در این معماري Fuzzifier فرآیندFuzzification را اجرا میکند که دو نوع از داده ها ي ورودي مانند سرعت پردازنده و بار تخصيص داده شده به ماشين مجازي را تبدیل می کند و یک خروجی مانند بار موازنه شده را که در استنتاج سيستم مورد نياز است.

این طرح سرعت پردازنده و بار ماشين مجازي را به عنوان دو پارامتر ورودي، به منظور توازن بار بهتر در ابر با استفاده از منطق فازي در نظر می گيرد. این پارامترها به عنوان ورودي بهFuzzifier داده می شوند که براي اندازه گيري بار متعادل به عنوان خروجی استفاده می شود. دو پارامتري که به عنوان سرعت پردازنده وبار تخصيصی ماشين مجازي ناميده میشوند با هم براي ارزیابی بار متعادل شده روي مراکز داده ي محيطهاي رایانش ابري از طریق منطق فازي، استفاده می شوند. نتایج به دست آمده با ارزیابی کارایی میتواند با کاهش زمان پردازش و نيز بهبود زمان پاسخ به توازن بار دست یابدکه منجر به استفاده ي حداکثر از منابع می شود.

سرعت پردازنده و بار اختصاص یافته به ماشين مجازي براي توازن بار در رایانش ابري ازطریق منطق فازي اعمال می شوند.[11]

**مدل پيام گرا**

خوشه ها فرصت استفاده از برنامه هاي کاربردي توزیع شده توسط کامپيوترهاي مختلف روي شبکه ها را فراهم می کنند.

این موضوع مرتبط با خوشه هاي مطرح شده در عملکرد شبکه می باشد، اگر کل بار در شبکه توزیعی توسط یک کامپيوتر توزیع شود، باعث کندي شبکه میشود. براي جلوگيري از این وضعيت، مدیریت منابع میتواند به صورت معيارهاي نرم افزاري براي توزیع ترافيک بين ایستگاهها استفاده می شود؛ به طوري که کارایی شبکه در حد بالایی حفظ شود.

وب سرویسها عمدتاً در برنامه هاي پيام فوري آنلاین استفاده می شود، این فناوري براي ارتباطات بلادرنگ بين طرفهاي مختلف میباشد با این حال وجود و یا دسترس بودن برنامه کاربردي مهم میباشد.

مدلی ارائه شده است که از XMPP براي توازن بار استفاده میکند. کلاینت هاي XMPP اطلاعات حاضر را به سرورحاضر XMPP می فرستند و جریانهاي XML جزئيات اطلاعات حاضر مشتریان توليد شده توسط این سرورها را در بردارد. با استفاده ازیک موازنه کننده بار در بالاي یک سرور XMPP اجازه می دهد درخواستهاي ورودي توسط سرویسهاي عمومی، اولویت بندي و به کار گرفته شوند در این الگوریتم از ميان معيارهاي توازن بار، معيار زمان پاسخ و کارایی در نظر گرفته شده است.[6]

**الگوریتم Min-Min**

این الگوریتم با مجموعه اي از وظایف تخصيص داده نشده، شروع میشود. اول از همه، زمان اتمام حداقل برايتمام وظایف یافت می شود. سپس در ميان این زمانهاي حداقل، کمترین مقدار انتخاب می شود که حداقل زمان ميان تمام وظایف روي هرمنبع موجود است. سپس مطابق با زمان حداقل، وظيفه روي ماشين مربوطه زمانبندي می شود.

سپس زمان اجرا براي همه ي وظایف دیگرروي ماشين با اضافه کردن زمان اجراي وظيفه ي تخصيص داده شده به زمان اجراي دیگر وظایف روي ماشين به روز رسانی می شود ووظيفه تخصيص داده شده از ليست وظایفی که به ماشين تخصيص داده می شود، حذف می شود.

دوباره همين روال تا زمانی که همه يوظایف به منابع تخصيص داده شوند، دنبال میشود. اما این روش یک اشکال اصلی دارد که میتواند منجر به گرسنگی شود.[17] در اینالگوریتم از ميان معيارهاي توازن بار، معيار بهره برداري از منابع، سربار، توان عملياتی، زمان پاسخ و کارایی در نظر گرفته شده است .[1]

**الگوریتم Min-Max**

تقریباً همان الگوریتم Min-Min است به جز موارد زیر: بعد از یافتن زمانهاي اجراي حداقل، بيشترین مقدار انتخاب می شود که زمان حداکثر ميان تمام وظایف روي هر منبع موجود است. سپس مطابق با زمان حداکثر، وظيفه روي ماشين مربوطه زمانبندي می شود.

سپس زمان اجرا براي همه ي وظایف دیگر روي ماشين با اضافه کردن زمان اجراي وظيفه ي تخصيص داده شده به زمان اجراي دیگر وظایف روي ماشين به روز رسانی میشود و وظيفه تخصيص داده شده از ليست وظایفی که به ماشين تخصيص داده می شود،حذف می شود .در این الگوریتم از ميان معيارهاي توازن بار، معيار بهره برداري از منابع، سربار، توان عملياتی، زمان پاسخ و کارایی درنظر گرفته شده است.[1]

**الگوریتمهای توازن بار دو مرحلهای OLB + LBMM**

یک الگوریتم زمانبندي دو مرحلهاي پيشنهاد شده است که الگوریتمهاي زمانبندي OLB )فرصت توازن بار( و LBMM )توازن بار Min-Min ( را براي استفاده از اجراي بهتر و حفظ توازن بارسيستم، ترکيب می کند. الگوریتم زمانبندي OLB هر گره را براي رسيدن به هدف توازن بار، در حالت کار نگه میدارد و الگوریتم زمانبندي LBMM براي کاهش زمان اجراي هر وظيفه روي گره استفاده می شود، در نتيجه زمان اجراي کل را کاهش میدهد.

محيط استفاده از این الگوریتم شبکه هاي رایانش ابري سه سطحی است و در آن، معيار کارایی و بهره برداري از منابع در نظر گرفته شده است .این الگوریتم ترکيبی، در استفاده ي موثر از منابع کمک می کند و بهره وري را افزایش می دهد نتایج بهتري را نسبت به کاوش زنبور عسل، نمونه برداري تصادفی و خوشه بندي فعال ارائه می دهد.[1]

**الگوریتم Queue-Idle-Join**

یک الگوریتم توازن بار را براي مقياس پذیري پویاي وب سرویسها پيشنهاد میکند. این الگوریتم توازن بار را در مقياس بزرگ را با توزیع کننده هاي توزیع شده فراهم می کند. ابتدا، توازن بار پردازنده هاي بی کار در عرض توزیع کننده ها براي دسترسی هر پردازنده بی کار به هر توزیع کننده و سپس تخصيص کارها به پردازنده ها به منظور کاهش ميانگين طول صف هرپردازنده صورت می پذیرد[1]

. با از بين بردن کار توازن بار از مسيرهاي حياتی پردازش درخواستها، این الگوریتم به طور موثر بار سيستم راکاهش می دهد، هيچ سربار ارتباطی هنگام ورود کارها رخ نمیدهد و زمان پاسخ واقعی را افزایش نمیدهد.

محيط استفاده از این الگوریتم، مراکز داده ابري است و در آن معيار کارایی، زمان پاسخ و سربار در نظر گرفته شده است .[7]

**سياست توازن بار مرکزی برای ماشينهای مجازی**

یک سياست توازن بار مرکزي براي ماشينهاي مجازي را پيشنهاد کرده اند که بار را به طور مساوي در یک محيط رایانشی ابر یا ماشين مجازي توزیع شده، متعادل میکند. این سياست کارایی کلی سيستم را افزایش می دهد اما سيستمهایی که داراي تحمل خطا هستند را در نظر نمی گيرد.

این روش از اطلاعات حالت عمومی براي تصميمات توازن باراستفاده می کندو بيشتر از 20 % کارایی را افزایش می دهد. محيط استفاده از این الگوریتم رایانش ابري است و در آن معيار کارایی و زمان پاسخ، توان عملياتی و بهره برداري از منابع در نظر گرفته شده است.[7]

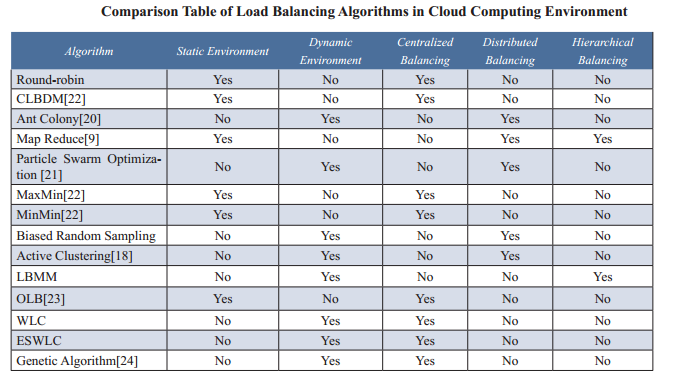
* **روش كولوني مورچه ها**[14]

زيركي گروه مورچه هاي مصنوعي به ويژه بهينه سازي منسوب به محاسبات و (ACO) كولوني مورچه ها نمونه رفتار جديد براي حل مسائل بهينه سازي تركيبي مي باشد و بر اساس اصولي است كه مي تواند رفتار سيستم هاي طبيعي را كنترل كند. هنگامي از موجودات هوشمند براي مسائل شبكه بهره گرفته (Agents) مي شود كه يك گروه از برنامه ها با هم كار مي كنندو مي توانند براي كاربردهايي مانند يافتن كوتاهترين مسير، مسيريابي، توازن بار و مديريت ومانند آن استفاده شوند.

Crausو همكاران، از الگوريتم مورچه براي انتقال كارها بين پردازنده هاي مختلف استفاده نموده است. كارها توسط مورچه ها بين گرههاي يك گراف كه پردازنده هاي سيستم مي باشند جابه جا مي گردند.

قاعده اين جابجايي بر مبناي الگوريتم زير مي باشد:ابتدا هر مورچه متوسط بار خود را محاسبه كرده وسپس خطاي محلي را نيز براساس متوسط بار خود ومورچه هاي قبلي تعيين مي كند. با استفاده از يك فاكتور تحمل پذيري خطا و ميزان ردپاهاي به جاگذاشته شده توسط ساير مورچه ها، احتمال ماندن يا ترك نمودن آن پردازنده را به دست آورده و اقدام به مهاجرت می کند الگوريتم پيشنهاديCraus در بعضی از شرايط، پاسخهاي بهتري را نسبت به الگوريتم DASUD ارائه نموده است.[15]

**مقایسه ی الگوریتم ها ی مطرح شده** [18]

****

**6-6-1-چگونگی یافتن کوتاهترین مسیر توسط مورچه ها**[22]

در طبیعت ، کلونی مورچه ها بدون اینکه اطلاع سراسری از مسیرها داشته باشند ، معمولا کوتا هترین مسیر لانه تا محل غذا را پیدا می کنند. زمانیکه مورچه راه می رود، از خود فرمون ترشح می کند و ردپایی به جای می گذارد . مورچه ها برای انتخاب مسیر مسیری را برمی گزینندکه دارای فرومون بیشتری است. در شکل 1 چگونگی یافتن کوتاهترین مسیر توسط کلونی مورچه ها نشان داده شده است.



در شکل Aمورچه ها به نقطه ای میرسند که بایستی تصمیم ، بگیرند که آیا به راست بروند یا به چپ. از آنجاییکه هیچ سرنخی وجود ندارد که کدام بهتر است، مورچه ها یکی را بصورت تصادفی بر می گزینند.

در شکل BوC با فرض اینکه مورچه ها دارای سرعت یکسانی هستند، مسیری که دارای طول کمتری می باشد پس از مدتی توسط مورچه های بیشتری طی می شود و بنابراین میزان، فرومون بیشتری روی آن مسیر بجای گذاشته میشود. در Dمورچه در نقطه تصمیم گیری مسیری را که دارای فرومون بیشتری است را با احتمال بالاتر انتخاب می کند. بدین گونه کلونی مورچه هاپس از مدتی کوتاهترین مسیر را پیدا میکند. علاوه بر این فرومون بجای گذاشته شده با نرخ معینی بخار می شود و به همین دلیل مسیری که کمتر مورد تردد مورچه ها واقع می شود پس از مدتی فرومونش را از دست می دهد. لذا اگر در محیط تغییری حاصل شود)مثلا مسیری مسدود شود(، مورچه ها می توانند مسیر بهینه دیگری را بیابند. مسئله توازن بار در یک گرید رایانشی را می توان به صورت زیر مدل کرد. هر گره با مواجهه با حجم زیاد کار، اقدام به تولید تعدادی مورچه می کند و آنهارا در شبکه رها می سازد. مورچه ها با اطلاعاتی که از سرویسهای مورد نیاز و وضعیت گره تولید کننده خود دارند،در شبکه جستجو کرده و پس از یافتن گره مادون بار شده آنرا به گره تولید کننده خود گزارش می دهند. براساس اطلاعات دریافت شده توسط مورچه ها، گره تولید کننده مورچه اقدام به فرستادن بارخود به گره های مادون بارشده می نماید.

دو الگوریتم توازن بار مبتنی بر کلونی مورچه در گریدهای رایانشی گزارش شده است : الگوریتم تمام تصادفی و

الگوریتم نشانه گذاری  **.** الگوریتم تمام تصادفی**،** همانند الگوریتم رسننیک عمل می کند، بدین صورت که هر مورچه بصورت تصادفی از گره ای به گره دیگر می رود تا به یک گره مادون بار شده برخورد نماید. در الگوریتم نشانه گذاری**،** هر مورچه مسیر بعدی خود را به صورت تصادفی برطبق مقدار فرومون مسیرهای ممکن انتخاب می کند. پس از انتخاب یک مسیر، امتیاز فرومون آن مسیر افزایش می یابد در این الگوریتم مقدار فرومون مسیرهای منتهی به گرههای مادون بارشده افزایش یافته و از این طریق این گره ها توسط مورچه های بیشتری ملاقات خواهند شد.

**شرح الگوریتم:**

هدف اصلی این الگوریتم کاهش متوسط زمان کاوش مورچه ها برای یافتن گرههای مادون بار شده می باشد.

در این الگوریتم جدول فرومون هر گره فقط توسط دمون آن گره بهنگام می شود و بهنگام سازی بدین ترتیب است که به میزانی که به یک گره کار ارسال می شود به آن گره، در گره فرستنده امتیاز داده می شود که دراینجا امتیاز همان مقدار فرومون است. بنابراین هر گره خصوصا گره هایی که بار زیادی تولید می کنند پس از گذشت مدتی، جدول فرومونشان بگونه ای پیکر بندی می شود که مورچه ها با احتمال بالایی به گره های مادون بارشده هدایت میشوند.

**نتیجه:**

توازن بار یکی از چالشهاي اساسی در رایانش ابري است که نيازمند توزیع حجم کاري محلی پویا به طور مساوي در عرض تمامی گره ها براي رسيدن به رضایت کاربر و نرخ استفاده ي بالاي منابع با حصول اطمينان از تخصيص عادلانه و کارآمد هر منبع محاسباتی است.

در این مقاله الگوریتم هاي مختلف توازن بار در رایانش ابري را مقایسه کردیم و به این نتيجه رسيدیم که ما می توانيم از یک الگوریتم خاص بر طبق نيازهایمان استفاده کنيم.

امارایانش ابري حوزه هاي وسيعی را پوشش می دهد و همانطور که نتيجه گرفتيم هيچ یک از الگوریتمهاي فوق معيارها را ارضا نمیکنند. بنابراین نياز به توسعه ي یک الگوریتم تطبيقی است که براي محيطهاي ناهمگن مناسب باشد و هزینه ها را کاهش دهد.

***REFERENCES***

[1] Sran.N and Kaur.N, Comparative Analysis of Existing Load Balancing Techniques in Cloud Computing, International Journal of Engineering Science Invention, Vol.2, Issue 1,2013.

[2] Khetan.A, Bhushan.V and Chand Gupta.S, A Novel Survey on Load Balancing in Cloud Computing International Journal of Engineering Research & Technology (IJERT) , Vol.2, Issue 2 ,2013.

[3] Ranjith Kumar.Y, Madhu Priya.M and Shahu Chatrapati.K, Effective Distributed Dynamic Load Balancing For The Clouds, International Journal of Engineering Research & Technology (IJERT), Vol.2, Issue 2, 2013.

[4] Begum.S and Prashanth.C.S.R, Review of Load Balancing in Cloud Computing. IJCSI International Journal of Computer Science Issues , Vol.10, Issue 1,2013.

[5] Hamo.A and Saeed.A, Towards a Reference Model for Surveying a Load Balancing, IJCSNS International Journal of Computer Science and Network Security , Vol.13, 2013.

[6] Rastogi.D, Bansal.A and Hasteer.N, A Techniques of Load Balancing in Cloud Computing : A Survey, International Conference on Computer Science and Engineering (CSE) 7th, April 2013.

[7] Kansal.N.J and Chana.I, Cloud Load Balancing Techniques : A Step Towards Green Computing, IJCSI International Journal of Computer Science Issues , Vol. 9, Issue 1, No 1, January 2012.

[8] Rajguru.A and A.S, A Comparative Performance Analysis of Load Balancing Algorithms in Distributed System using Qualitative Parameters, International Journal of Recent Technology and Engineering (IJRTE), Vol.1, Issue 3, 2012.

[9] Al Nuaimi.K, Mohamed.N, Al Nuaimi.M and Al-Jaroodi.J, A Survey of Load Balancing in Cloud Computing: Challenes and Algorithmsg, 2012 IEEE Second Symposium on Network Cloud Computing and Applications, 2012.

[10] Membrey.P, Hows.D and Plugge.E, Load Balancing in the Cloud, pp.211 – 224,2012.

[11] Sethi.S, Sahu.A and Jena.S.K, Efficient load Balancing in Cloud Computing using Fuzzy Logic, IOSR Journal of Engineering (IOSRJEN) ,Vol. 2, Issue 7, pp.65-71,2012.

*[12]* Distinguishing Cloud Computing from Utility Computing *(http:/ / www. ebizq. net/ blogs/ saasweek/ 2008/ 03/ distinguishing\_cloud\_computing/ )*

[13].**Cloud computing: web based applications that change the way you work and**

**collaborate online** By michel Miller

[14] P. Mell and T. Grance. (2009, Oct.). The NIST definition of cloud com-puting, NIST special publication 800-145, National Institute of Standards and Technology. Available: http://www.csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

[15] S. Marston*, et al.*, "Cloud computing — The business perspective," *Decision Support Systems,* vol. 51, pp. 176-189, 2011 .

[16] S.Iqbal, , G.F.Care, “Performance Analysis of Dynamic Load Balancing Algorithms with Variable Number of Processors”, *Journal of* *Parallel and Distributed Computing*, Vol. 65, pp. 934-948, 2005.

[17] S. RAOGURURAJ, S. STONEHAROLD and T.C. HU, “ Assignment of Tasks in A Distributed Processor System With Limited Memory*”, IEEE Trans. On Computers*, Vol. C- 28, No. 4, April 1979.

[18]Comparative Study on Load Balancing Techniques in Cloud Computing N. S. Raghava\* and Deepti Singh

[19] Chiang.M.L, Luo.J.N, Lin.C.B and Wang.S.S, High-Reliable Dispatching Mechanisms for Tasks in Cloud Computing. Department of Information and Communication Engineering, Chaoyang University of Technology,2011.

[20] Hung.C, Wang.H and Hu.Y, Efficient Load Balancing Algorithm for Cloud Computing Network,2011.

[21] Mondal.B and D.K, Load Balancing in Cloud Computing using Stochastic Hill Climbing-A Soft Computing Approach, Published by Elsevier Ltd, 2011

[22] M. Dorigo and L. M. Gambardella, "Ant colony system: a cooperative learning approach to the traveling salesman problem," *IEEE Transactions on Evolutionary* *Computation ,*Volume 1, No.1, 1997, pp. 53-66