



دانشگاه کردستان
دانشکده فنی و مهندسی
گروه کامپیوتر و فناوری اطلاعات

پروژه پایانی درس رایانش فراگیر

استاد درس: سعدون عزیزی

تدریس یار: آرمین چوپانی

نیمسال دوم تحصیلی ۱۴۰۲-۰۱

مقیاس‌پذیری خودکار^۱ یک تکنیک مقیاس‌پذیری است که می‌توانید برای بارهای کاری^۲ میزبانی شده در یک محیط ابری^۳ اعمال کنید. یکی از مزایای عمده میزبانی مبتنی بر ابر این است که می‌توانید به راحتی ظرفیت را به هر میزبانی که برای پشتیبانی از تقاضای خدمات^۴ خود نیاز است، افزایش دهید. مقیاس‌پذیری خودکار این مزیت را یک قدم جلوتر می‌برد. با مقیاس‌پذیری خودکار، از آنجایی که تقاضا برای یک حجم کاری معین در طول زمان تغییر می‌کند، مقدار منابع تخصیص یافته برای پشتیبانی از آن حجم کاری به طور خودکار با نیازهای عملکرد شما سازگار می‌شود.

قبل از اینکه مقیاس‌پذیری خودکار یک گزینه باشد، مقیاس‌پذیری حجم‌های کاری اغلب چالش برانگیز بود. تخصیص منابع به صورت دستی برای پشتیبانی از حجم کاری ذاتاً مستعد خطا است. زیرا پیش‌بینی دقیق تغییرات در تقاضا یا دانستن اینکه چه مقدار منابع برای رسیدگی به این تغییرات نیاز است، دشوار می‌باشد. این ابهام می‌تواند از یک سو منجر به تامین بیش از حد^۵ و پرهزینه منابع شود یا از سوی دیگر به اختلالات احتمالی خدمات به دلیل تامین ناکافی^۶ منابع منجر شود. مقیاس‌پذیری خودکار با افزایش یا کاهش خودکار مقدار منابع اختصاص داده شده به حجم کاری شما به نسبت مستقیم با مقداری که تقاضا نیز افزایش یا کاهش می‌یابد، به حل این مشکلات کمک می‌کند.

برای درک بهتر مقیاس‌پذیری خودکار، ابتدا باید در نظر بگیرید که دو نوع مقیاس‌پذیری ممکن است:

۱. مقیاس‌پذیری افقی^۷

در این نوع مقیاس‌پذیری تعداد گره‌ها را که در یک حجم کاری معین شرکت می‌کنند، افزایش یا کاهش می‌یابد. مزیت مقیاس‌پذیری افقی این است که به شما امکان می‌دهد، تقریباً مقدار نامحدودی از ظرفیت جدید را بدون تأثیر بر گره‌های موجود یا ایجاد خرابی^۸ اضافه کنید. این روش در مقایسه با مقیاس‌پذیری عمودی، برای مقیاس‌پذیری ظرفیت، سریع‌تر است. با این حال، هر برنامه یا حجم کاری را نمی‌توان به صورت افقی مقیاس‌پذیر کرد.

۲. مقیاس‌پذیری عمودی^۹

این نوع مقیاس‌پذیری، حافظه^{۱۰} و یا قدرت پردازش^{۱۱} موجود را برای گره‌های موجود افزایش یا کاهش می‌دهد. به عنوان مثال، می‌توانید دو گره سرور مجهز به ۱۶ گیگابایت رم و ۴ vCPU را به صورت عمودی مقیاس‌دهی کنید تا هر کدام ۶۴ گیگابایت رم و ۱۶ vCPU داشته باشند. در برخی

¹ Autoscaling

² Workloads

³ Cloud

⁴ Services

⁵ Over-provisioning

⁶ Under-provisioning

⁷ Scaling out / Horizontal scaling

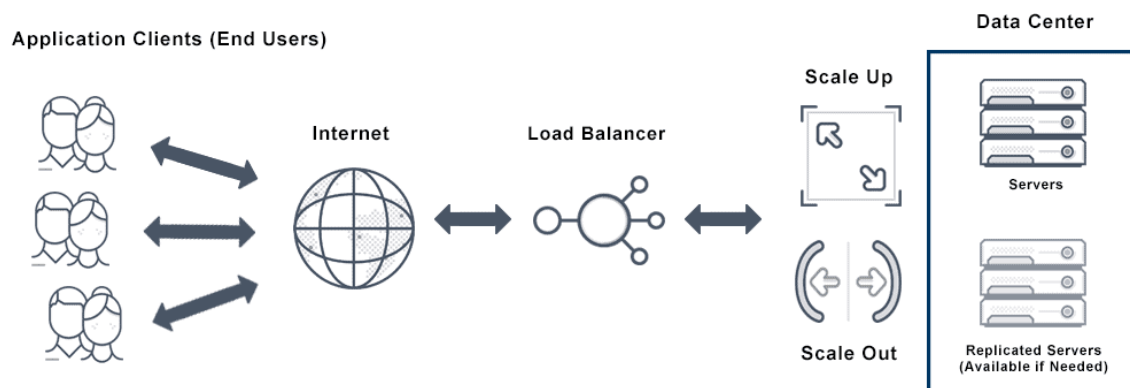
⁸ Downtime

⁹ Scaling up / Vertical scaling

¹⁰ Memory

¹¹ Processing power

موارد، مانند پایگاه‌های داده رابطه‌ای که بدون هیچگونه تقسیم‌بندی پیاده‌سازی شده‌اند، مقیاس-پذیری عمودی این مزیت را دارد که تنها راه مناسب برای مقیاس‌پذیری در پاسخ به افزایش تقاضا باشد.



مقیاس‌پذیری خودکار سازمان‌ها را قادر می‌سازد خدمات ابری مانند ظرفیت‌های سرور یا ماشین‌های مجازی را بر اساس موقعیت‌های تعریف‌شده مانند سطوح استفاده از ترافیک، به‌طور خودکار افزایش یا کاهش دهند. ارائه‌دهندگان محاسبات ابری، مانند خدمات وب آمازون (AWS)، مایکروسافت آژور، و پلتفرم ابری گوگل (GCP)، ابزارهای مقیاس‌پذیری خودکار را ارائه می‌دهند.

از ویژگی‌های اصلی مقیاس‌پذیری خودکار همچنین با افزایش و کاهش بی‌وقفه نمونه‌های^۱ جدید با افزایش و کاهش تقاضا، هزینه کمتر و عملکرد^۲ قابل‌اعتماد را ممکن می‌سازد. به این ترتیب، مقیاس‌پذیری خودکار با وجود تقاضای پویا و در مواقعی غیرقابل پیش‌بینی برای برنامه‌ها، ثبات را فراهم می‌کند. مزیت کلی مقیاس‌پذیری خودکار این است که با تغییر خودکار تعداد سرورهای فعال، نیازی به پاسخ دستی در زمان واقعی به جهش‌های ترافیکی که نیازمند منابع و نمونه‌های جدید هستند را از بین می‌برد.

• تعریف پروژه

در این پروژه شما باید با ارائه و پیاده‌سازی الگوریتم‌هایی، مکانیزم مقیاس‌پذیری خودکار را برای پردازش درخواست‌ها ایجاد کنید. سیستم شما باید قادر به ایجاد و از بین بردن خودکار گره‌های محاسباتی باشد، با افزایش و کاهش درخواست‌ها.

هدف این پروژه ارائه و پیاده‌سازی یک الگوریتم ابتکاری برای مسئله مقیاس‌پذیری خودکار و همچنین پیاده‌سازی مسئله با یکی از الگوریتم‌های یادگیری تقویتی است.

در این پروژه باید یک برنامه‌ای بنویسید که شامل سه بخش باشد:

۱. تولید درخواست‌ها

¹ Computing Node Instances

² Performance

در این سیستم، درخواست‌ها باید براساس توزیع پواسن با نرخ ثابت تولید شوند. هر کدام از درخواست‌ها باید دارای: شناسه یکتا^۱، زمان ایجاد درخواست و زمان پردازشی مورد نیاز باشند.

۲. بخش اصلی (الگوریتم‌های ارائه شده)

پیاده‌سازی دو الگوریتم پیشنهادی

۱. الگوریتم ابتکاری

۲. حل مسئله با یکی از الگوریتم‌های یادگیری تقویتی

۳. پردازش درخواست‌ها (پردازش درخواست‌ها در گره‌های محاسباتی)

در این بخش با شبیه‌سازی گره‌های محاسباتی، درخواست‌های ایجاد شده در بخش اول را پردازش می‌کنید. هر کدام از گره‌ها که برای پردازش درخواست‌ها ایجاد می‌کنید، باید دارای شناسه یکتا، زمان ایجاد، تعداد درخواست‌های که پردازش کرده و طول عمر^۲ آن نمونه باشند.

• معیارهای ارزیابی

معیارهای که باید برای نشان دادن کیفیت الگوریتم‌های ارائه شده در نظر بگیرید شامل، درصد موفقیت انجام درخواست‌ها و تعداد گره‌های ایجاد شده است.

نکته: برای الگوریتم یادگیری تقویتی استفاده شده، علاوه بر دو مورد ذکرشده در بالا، باید حالت^۳، اعمال^۴، تابع پاداش^۵ مورد استفاده و پاداش تجمیعی^۶ حاصل، در طول آموزش مدل نیز گزارش شود.

¹ Unique ID

² Lifespan

³ State

⁴ Actions

⁵ Reward Function

⁶ Accumulated Reward