



دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

سمینار کارشناسی ارشد گرایش مهندسی نرم افزار

بهینه سازی توزیع بار برای سکوهاى خدمات پردازش بدون میزبان با استفاده از روش های مدل سازی

Optimizing Load Balancing for Serverless Computing Platforms Using Modeling Approaches

نگارش: فرید شاهین فر - ۹۹۲۰۱۴۷۶

استاد راهنما: دکتر علی موقر رحیم آبادی

استاد ممتحن داخلی: دکتر جعفر حبیبی

بهمن ۱۴۰۰

- پردازش بدون‌میزبان راهکار جدیدی برای ارائه خدمات ابری است.
- توزیع مناسب بار کاری یکی از بخش‌های اساسی آن است.
- هدف از این پروژه ارزیابی کارایی الگوریتم‌های توزیع بار موجود برای سکوه‌های پردازش بدون‌میزبان است.
- ایجاد یک مدل مناسب از این سیستم امکان تحلیل مناسب را ایجاد می‌کند.

فهرست مطالب

اهمیت پردازش ابری

ارائه خدمات ابری باعث رشد پیشرفت در علم و صنعت شده است.

- CERN به کمک خدمات ابری گوگل چندین پتابایت داده را در سال پردازش می کند[؟].
- تمام زیرساخت Netflix از خدمات آمازون است[؟].
- شرکت Micros از خدمات آمازون برای پویانمایی فیلم ها استفاده می کند[؟].

ویژگی پردازش ابری

در خدمات ابری دو مورد زیر اهمیت بسزایی دارند.

- مقایسه پذیری
- اتکال پذیری

ویژگی پردازش ابری

در خدمات ابری دو مورد زیر اهمیت بسزایی دارند.

- مقایس پذیری
- اتکاپذیری

راهکار

استفاده از معماری های توزیع شده مانند میکروسرویس راهکار مورد استفاده در عمل برای رسیدن به این دو مورد است [؟].

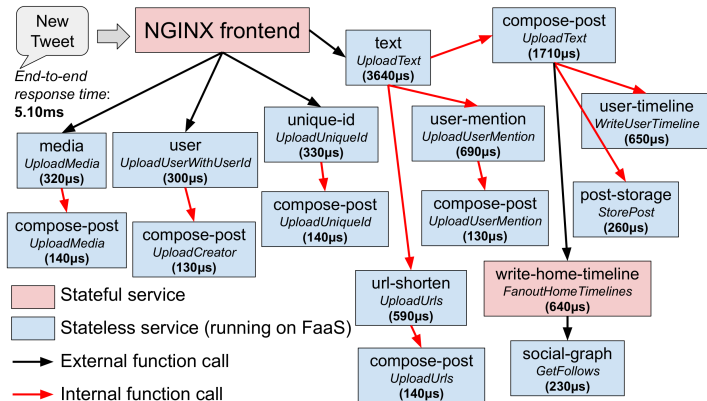
معماری میکروسرویس

- در این معماری، سیستم به مولفه‌هایی ریزدانه و مستقل تقسیم می‌شود.
- موله‌ها با یک دیگر در بستر شبکه‌های کامپیوتری در ارتباط هستند.
- هر مولفه می‌تواند به صورت جداگانه بر روی تعدادی خدمت‌رسان اجرا شود.

معماری میکروسرویس

- در این معماری، سیستم به مولفه‌هایی ریزدانه و مستقل تقسیم می‌شود.
- موله‌ها با یک دیگر در بستر شبکه‌های کامپیوتری در ارتباط هستند.
- هر مولفه می‌تواند به صورت جداگانه بر روی تعدادی خدمت‌رسان اجرا شود.
- در صورت نیاز به توان پردازشی یک مولفه می‌توان تعداد خدمت‌رسان‌های آن را افزایش داد.
- در صورت کاهش بار کاری می‌تواند تعداد خدمت‌رسان‌ها را کاهش داد تا در هزینه صرفه‌جویی شود.
- در صورت رخدادن اشکال برای یک خدمت‌گزار، دیگر خدمت‌گزارها می‌تواند پردازش را ادامه دهند.

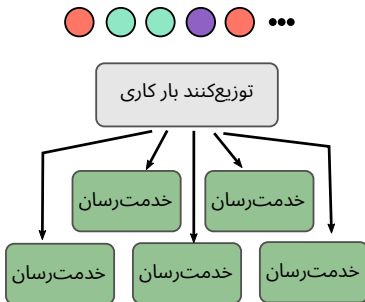
مثالی از یک معماری میکروسرویس



شکل: مثالی از زنجیره مولفه برای ارسال یک تویت (برگرفته از [؟])

نیاز به توزیع کننده بار

- در این معمار چندین خدمت گزار یک مولفه یکسان را اجرا می کنند.
- توزیع کننده بار تصمیم می گیرد که درخواست ها در کدام خدمت رسان پردازش شوند.
- هدف این است که دریافت کننده خدمات چندین خدمت گزار به صورت یک خدمت گزار مشاهده کند.



پیچیدگی زیرساخت معماری میکروسرویس

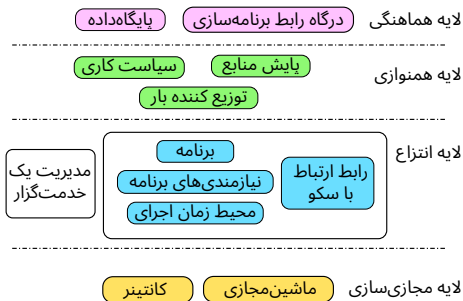
چالش

ایجاد و پیاده‌سازی سیستم‌های نرم‌افزاری با ویژگی‌های شرح داده شده هزینه‌بر و دارای چالش‌های فنی بسیاری است.

خدمات پردازش بدون میزبان

- پردازش بدون میزبان یک چارچوب پردازش ابری است.
- برنامه ساز می تواند برنامه خود را بدون نیاز به فعالیت های اجرایی مانند تهیه و تخصیص منابع پردازشی ایجاد و اجرا کند.
- وظیفه مسائل مربوط به **اتکاپذیری** و **مقیاس پذیری** بر عهده ارائه دهنده خدمات است.
- میزان هزینه این خدمات بر اساس میزان مصرف واقعی برنامه از منابع موجود محاسبه می شود [؟].

معماری سکوی پردازش بدون میزبان



شکل: نگاه سطح بالا به لایه های مختلف در سکوهای پردازش بدون میزبان

ویژگی‌های برنامه‌های پردازش بدون میزبان

- مولفه‌ها در قالب توابع مبتنی بر رخداد تعریف شده‌اند.
- در هنگام دریافت رخداد یک نسخه از این توابع در یک محیط زمان اجرا بارگیری می‌شود و رخداد را پردازش می‌کند.
- در طول اجرا، یک مولفه می‌تواند دیگر مولفه‌ها را فراخوانی کند و یک زنجیره اجرا ایجاد کند.
- در ذات اجرای مولفه‌ها مواردی همچون **همروندی** اجرای چندین مولفه، **وقوع اشکال** در زمان اجرا، **تلاش مجدد یک اجرا** و **استفاده مجدد از محیط زمان اجرای** یک مولفه وجود دارد.
- محیط‌های اجرا در ابتدا سرد هستند و پس از پردازش اطلاعات گرم در نظر گرفته می‌شوند.
- پس از پایان پردازش یک محیط اجرا می‌تواند بسته شود یا برای مدتی باز بماند.

فهرست مطالب

توزیع بار کاری در سکوهای پرداز بدون میزبان

- سرعت ایجاد محیط اجرا یک چالش برای کارایی سکو پردازش بدون میزبان است.
- برای ایجاد محیط اجرا باید نیازمندی‌ها بار گیری و اجرا شوند.
- به همین منظور خدمت گزاران این نیازمندی‌ها را در یک حافظه نهان نگهداری می‌کنند.

کار مرتبط

PASch یک سیستم توزیع بار است که با توجه به نیازمندی‌های بارگیری شده روی خدمت گزارها گذردهی سیستم را ۱.۲۹ برابر افزایش و تاخیر صدک ۸۰ام را تا ۲۳ برابر کاهش داده است [۱].

توزیع بار کاری در سکوه‌های پرداز بدون میزبان

- محیط‌های اجرا در حالت سرد عملکرد ضعیف‌تری نسبت به محیط گرم دارند.
- محیط‌های اجرا پس از اتمام کار برای مدتی باز نگه‌داشته می‌شوند.
- توزیع‌کننده بار می‌تواند با آگاهی از محیط‌های گرم کارایی را بهبود دهد.

کار مرتبط

با افزودن آگاهی از مکان محیط‌های گرم به توزیع‌کننده بار، نشان داده شده است که می‌توان ۶۳ درصد تاخیر در صدک ۹۹ام را کاهش داد [۱].

توزیع بار کاری در سکوه‌های پرداز بدون میزبان

- تشخیص زمان بستن یک محیط اجرا اهمیت فراوانی دارد.
- اگر محیط اجرا فوراً بسته شود تعداد محیط‌های سرد ایجاد شده افزایش پیدا می‌کند.
- اگر محیط اجرا بسته نشود هزینه خدمات بسیار زیاد می‌شود.
- تمام مولفه‌ها فرکانس فراخوانی یکسانی ندارند.

کار مرتبط

با پایش و جمع‌آوری آمار فراخوانی مولفه‌های مختلف روش‌هایی برای تعیین مدت زمان مناسب برای باز نگهداشتن یک محیط اجرا پیشنهاد شده است [۱].

توزیع بار کاری در سکوهای پرداز بدون میزبان

کار مرتبط

برای اجرای کاربردهای بلادرنگ نیاز است تا تضمین پایان پردازش تا موعد مقرر رخداد ارائه شود. به همین منظور پیاده سازی یک الگوریتم آگاه از موعد رخدادها و توان پردازشی موجود نیاز است [؟].

کار مرتبط

تضمین کیفیت سرویس برای ارائه دهندگان خدمات ابری اهمیت دارد. از این منظر پیاده سازی الگوریتم های مختلف که انصاف را برای انواع بارهای کاری رعایت کند اهمیت دارد [؟].

کار مرتبط

ارسال پیام از یک خدمت گزار به یک خدمت گزار دیگر زمان بر است. از این روی طراحی سیستم توزیع بار به نحوی که زنجیره اجرا بر روی یک خدمت گزار قرار بگیرد باعث افزایش کارایی می شود [؟].

جدول: جمع‌بندی کارهای مرتبط

معيار مورد توجه	استفاده از مدل تحليلی	سال انتشار	مرجع
ارزیابی کارایی	بله	۲۰۱۹	[۱]
هزینه سرویس در مقابل کارایی	بله	۲۰۲۰	[۲]
توجه به نیاز سیستم‌های بی درنگ	خیر	۲۰۲۱	[۳]
افزایش بهره‌وری حافظه نهان	خیر	۲۰۲۱	[۴]
تعداد محیط‌های اجرای سرد	خیر	۲۰۲۰	[۵]
توجه به حافظه نهان نیازمندی‌های برنامه	خیر	۲۰۱۸	[۶]
توجه به چرخه‌های لایه مجازی‌سازی	خیر	۲۰۲۲	[۷]
توجه به اجرای زنجیره روی یک خدمت‌گزار	خیر	۲۰۲۱	[۸]

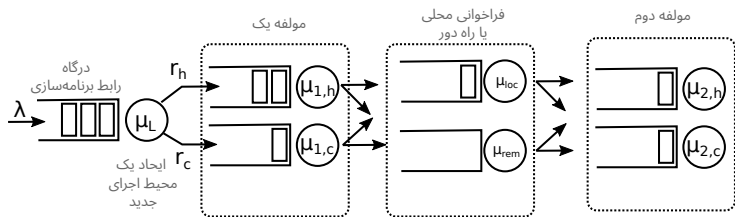
فهرست مطالب

- ایجاد مدلی برای ارزیابی سیستم توزیع بار (لایه همناوی).
- پیاده سازی و ارزیابی الگوریتم‌های ارائه شده در مدل تهیه شده.
- ایجاد یک الگوریتم توزیع بار جدید با توجه به اطلاعات بدست آمده.

جزئیات مهم مدل

- تعداد محیط‌های اجرا پیوسته تغییر می‌کند.
- سرد و یا گرم بودن محیط اجرا در کارایی تاثیر دارد.
- قرار دادن زنجیره بر روی یک ماشین باعث افزایش کارایی می‌شود.
- نسبت تعداد محیط‌های در حال اجرا به توان سخت‌افزاری ماشین در کارایی تاثیر گذار است.

جزئیات مهم مدل



شکل: مدل پیشنهادی برای سکو پردازش بدون میزبان

جزئیات مهم مدل

نماد	توضیح
$m_{i,j}$	تعداد محیط اجرا گرم برای مولفه f_i برای سرور j
$select(j)$	احتمال انتخاب خدمت‌گزار j (وابسته به وضعیت سیستم و الگوریتم انتخابی)
$q_{i,h,j}$	تعداد درخواست‌های درون صف پردازش گرم
$r_{i,h,j}$	احتمال ورود به صف گرم (وابسته به $m_{i,j}$)
$r_{i,c,j}$	احتمال ورود به صف سرد ($1 - m_{i,j}$)
μ_L	نرخ سرویس‌دهی توزیع‌کننده بار
$\mu_{i,c}$	نرخ سرویس‌دهی محیط اجرای مولفه f_i در حالت سرد.
μ_{loc}	نرخ سرویس‌دهی درخواست‌های محلی.
μ_{rem}	نرخ سرویس‌دهی درخواست‌های راه‌دور.

- برای ارزیابی درستی مدل ارائه شده، نتایج آن را با مشاهدات عملی مقایسه می شود.
- برای این منظور می توان از ردپاهای ارائه شده در [؟]، که مربوط به سکوی پردازش بدون میزبان آژور برای شرکت مایکروسافت است، استفاده کرد و یا به جمع آوری ردپا پرداخت.

- برای توزیع بار در سکوه‌های پردازش بدون‌میزبان الگوریتم‌های متعددی پیشنهاد شده است [؟، ؟، ؟، ؟، ؟].
- ارزیابی این الگوریتم‌ها و مقایسه آن‌ها برای فهم برتری و نقاط ضعف هر کدام حائز اهمیت است. همچنین هر کدام از این الگوریتم‌ها از جنبه‌ی متفاوتی مانند هزینه ارائه خدمات، گذردهی سیستم، تعداد محیط‌های اجرای سرد و تاخیر مشاهده‌شده در صدک‌های بالا به مسئله توزیع بار نگاه کرده‌اند.
- در این پروژه با شناسایی این الگوریتم‌ها و مقایسه آن‌ها با استفاده از روش‌های مدل‌سازی ویژگی‌های هر کدام را شناسایی می‌کنیم.
- در نهایت با توجه به اطلاعات بدست آمده الگوریتم جدیدی برای توزیع بار ارائه می‌شود.

پیشرفت مراحل پایان نامه

جدول: زمان بندی فعالیت های پایان نامه

عنوان فعالیت	مدت زمان لازم	میزان پیشرفت	زمان شروع
طالعه پژوهش های پیشین	سه ماه	۱۰۰ %	مهر ۹۹
طراحی مدل ارزیابی سکو پردازش بدون میزبان	دو ماه	۴۰ %	دی ۹۹
پیاده سازی برنامه شبیه ساز بر اساس مدل	یک ماه	۷۰ %	دی ۹۹
انجام آزمایش و جمع آوری مشاهدات	یک ماه	۰ %	
ارائه یک الگوریتم توزیع بار جدید	دو ماه	۰ %	
جمع بندی و نگارش پایان نامه	یک ماه	۰ %	

با تشکر از توجه شما

واژنامه I

- ۱ سکو Platform
- ۲ بدون میزبان Serverless
- ۳ خدمت رسان Server
- ۴ مقیاس پذیری Scalability
- ۵ اتکا پذیری Fault Tolarence
- ۶ معماری میکروسرویس Microservice Architecture
- ۷ مولفه Component
- ۸ کاربرد Application
- ۹ چارچوب Framework
- ۱۰ مشتری Client
- ۱۱ مجازی سازی Virtualization
- ۱۲ همنازی Orchestration
- ۱۳ شیوه نامه Protocol

واژنامه II

- ❶۴ درگاه رابط برنامه‌نویسی API Gateway
- ❶۵ توسعه و عملیات اجرایی DevOps
- ❶۶ مبتنی بر رخداد Event Driven
- ❶۷ اشکال Failure
- ❶۸ مدل معنانشناسی ساخت‌یافته عملیاتی Structural Operational Semantic (SOS)
- ❶۹ میانگیری Buffering
- ❷۰ تئوری صف Queuing Theory
- ❷۱ شبکه‌های فعالیت احتمالی Stochastic Activity Networks (SAN)
- ❷۲ حافظه نهان Cache Memory
- ❷۳ بلادرنگ Realtime
- ❷۴ موعد Deadline
- ❷۵ ردپا Trace