



دانشگاه شهید بهشتی

دانشکده مهندسی و علوم کامپیوتر

## کاهش شروع‌های سرد در پلتفرم‌های بدون سرور

گزارش سمینار کارشناسی ارشد مهندسی کامپیوتر  
گرایش نرم‌افزار

نگارش

امیرمحمد کرمزاده

استاد راهنما

دکتر علیرضا شاملی

خرداد ۱۴۰۰

## چکیده

لورم ایپسوم متن ساختگی با تولید سادگی نامفهوم از صنعت چاپ و با استفاده از طراحان گرافیک است. چاپگرها و متون بلکه روزنامه و مجله در ستون و سطرآنچنان که لازم است و برای شرایط فعلی تکنولوژی مورد نیاز و کاربردهای متنوع با هدف بهبود ابزارهای کاربردی می باشد. کتابهای زیادی در شصت و سه درصد گذشته، حال و آینده شناخت فراوان جامعه و متخصصان را می طلبد تا با نرم افزارها شناخت بیشتری را برای طراحان رایانه ای علی الخصوص طراحان خلاقی و فرهنگ پیشرو در زبان فارسی ایجاد کرد. در این صورت می توان امید داشت که تمام و دشواری موجود در ارائه راهکارها و شرایط سخت تایپ به پایان رسد و زمان مورد نیاز شامل حروفچینی دستاوردهای اصلی و جوابگوی سوالات پیوسته اهل دنیای موجود طراحی اساسا مورد استفاده قرار گیرد.

**واژگان کلیدی:** رایانش ابری، رایانش بدون سرور، شروع های سرد، FaaS، function-as-a-service

# فهرست مطالب

۱	مقدمه	۱
۲	۱.۱ صورت مسئله	۲
۲	۲.۱ انگیزه‌ی تحقیق	۲
۳	۳.۱ مثال مرتبط	۳
۴	۴.۱ اهمیت موضوع	۴
۵	۵.۱ نتیجه‌های مهم تحقیق	۵
۶	۲ مروری بر ادبیات	۶
۸	۳ کارهای مرتبط	۸
۱۰	۴ نتیجه‌گیری	۱۰
۱۲	مراجع	۱۲

## فهرست شکل‌ها

۴	۱.۱ اهمیت شروع سرد . . . . .
---	------------------------------

# فهرست جداول

# فصل ۱

## مقدمه

## ۱.۱ صورت مسئله

چه روش‌هایی برای کاهش تعداد شروع‌های سرد در پلتفرم‌های بدون سرور<sup>۱</sup> با حداقل سربرار<sup>۲</sup> و زمان اجرایی وجود دارند؟

## ۲.۱ انگیزه‌ی تحقیق

رایانش بدون سرور<sup>۳</sup> یکی از مسائل داغ و محبوب این‌روزهای دنیای مهندسی نرم‌افزار و رایانش ابری است. رایانش بدون سرور حوزه‌ی جدیدی را در توسعه‌ی محصول و استقرار<sup>۴</sup> اپلیکیشن‌ها باز کرده است. یکی از دلایل محبوبیت استفاده از پلتفرم‌های بدون سرور و تمایل توسعه‌دهندگان برای مهاجرت به سمت آن، استفاده بیش از پیش از معماری میکروسرویس و نانوسرویس در توسعه‌ی محصولات و حرکت معماران و مهندسين نرم‌افزار در تولید و مهاجرت برنامه‌های کاربردی با این معماری‌ها است.

از دید توسعه دهنده، رایانش ابری با حذف دخالت مستقیم کاربران انتهای در مدیریت زیرساخت از جمله IaaS یا IaaS-like یا IaaS-like موجب بهبود سرعت توسعه محصول و تمرکز کاربران بر روی منطق<sup>۵</sup> برنامه است. همچنین، برای علاوه بر آسانی استفاده و پنهان‌سازی پیچیدگی مدیریت سرور از کاربر، به علت اینکه ارائه‌دهندگان خدمات ابری در نقاط مختلف جهان حضور دارند و همچنین کانفیگ بهینه CDN ها؛ ارتباطات بین سرورها و کاربران با حداقل تاخیر<sup>۶</sup> صورت می‌گیرد.

به طور کلی، یک پلتفرم بدون سرور را هر پلتفرم محاسباتی تعریف کرد که در آن مدیریت مستقیم سرور از کاربران مخفی شده و برنامه‌های کاربردی به صورت اتوماتیک در آن مقیاس‌پذیر می‌شوند و تنها هنگامی که در حال استفاده از پلتفرم هستیم، هزینه آن را پرداخت می‌کنیم. [۱]

یکی از قابلیت‌هایی که در رایانش بدون سرور باعث محبوبیت آن شده است، قابلیت Scale-to-Zero است. این بدان معنی است که هنگامی که از یک کانتینر استفاده‌ای نداریم، منابع آن گرفته می‌شوند و کانتینر اصطلاحاً

<sup>1</sup>serverless

<sup>2</sup>Overhead

<sup>3</sup>Serverless Computing

<sup>4</sup>Deployment

<sup>5</sup>logic

<sup>6</sup>Latency

Zero-Scaled می‌شود. این خود موجب قابلیت پرداخت تنها در حین مصرف ما از تابع می‌شود. اما مشکل اصلی زمانی است که درخواست جدیدی برای کانینر Zero-Scale شده می‌رسد؛ در این حالت باید درخواست منتظر مانده تا سلسله‌ای از آماده‌سازی‌ها انجام شوند تا کانینر مربوطه مجدداً اجرا شود. این خود باعث تاخیری مضاف برای پاسخ‌دهی به درخواست را موجب می‌شود که به این تاخیر مشکل شروع سرد<sup>۱</sup> گفته می‌شود. در واقع می‌توان گفت تاخیر شروع سرد ناشی از تلاش ما در تعادل بین تاخیر در پاسخگویی به درخواست‌ها و هزینه (هزینه‌های استفاده از رم و سی‌پی‌یو و ...) است.

متأسفانه در سالیان اخیر و در عین داغ‌بودن مبحث و نیاز بازار به حل این مشکل، این مشکل چندان در محیط‌های آکادمیک مورد بررسی قرار نگرفته است. البته در نگاه کلی‌تر، مشکلات و مسائل بار مربوط به رایانش بدون سرور، اکثراً در محیط‌های آکادمیکی مثل دانشگاه‌ها با کم محلی روبرو شده‌اند. در این میان، پلتفرم‌های متن‌باز<sup>۲</sup> که دارای جوامع بسیار گسترده‌ای نیز هستند، به خاطر این سری مسائل باز که شرکت‌های تجاری در حال صرف هزینه‌های هنگفتی برای حل و فصل مشکلات مربوط به آن هستند، به شدت از رقابت عقب مانده‌اند. انگیزه ما برای انجام این پژوهش در این است که اولاً بتوانیم به راهکار مناسب‌تری برای حل مشکل مربوط به شروع سرد در پلتفرم‌های بدون سرور برسیم و ثانیاً بتوانیم با مشارکت در بهبود یکی از پلتفرم‌های آزاد در توسعه این پلتفرم‌ها تاثیر کوچکی داشته باشیم.

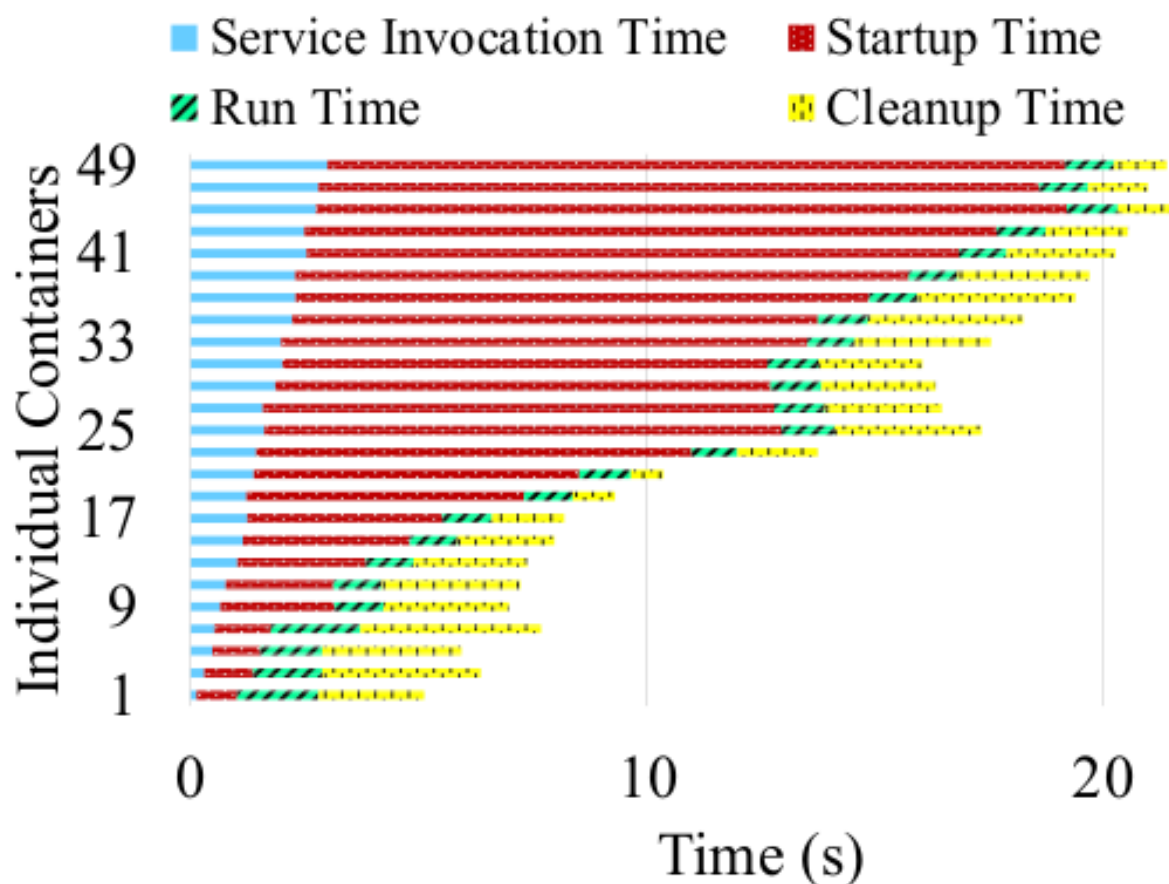
## ۳.۱ مثال مرتبط

در مقاله [۲] آقای لین و همکارش توانستند تا با اسفاده از استخر گرم و نگهداری کانینر توابعی که محبوبیت استفاده دارند، مدت زمان پاسخ را در حدود ۸۵٪ کاهش دهند. این بهبود در پلتفرم knative اجرا شد و ایده‌ی مقالات دیگری نیز بوده است.

<sup>۱</sup> Cold Start

<sup>۲</sup> Open Source





شکل ۱.۱: اهمیت شروع سرد

## ۴.۱ اهمیت موضوع

اگرچه پلتفرم‌های بدون سرور از نظر هزینه، scaling، راحتی استفاده و گستردگی پوشش جغرافیایی برای ما بهینه هستند؛ اما مشکل شروع سرد مشکلی نیست که بتوان به سادگی از آن گذر کرد. تصویر ۱.۱ که از [۳] گرفته شده است، نشان می‌دهد که بیش از ۸۰٪ زمان اجرای کامل یک کانتینر در پلتفرم‌های بدون سرور به آماده‌سازی آن یا شروع سرد اولیه<sup>۱</sup> مربوط می‌شود.

ستون قرمز رنگ زمان آماده‌سازی کانتینر را نمایش می‌دهد. این زمان همان زمان شروع سرد است که دلایل مختلفی از جمله آماده‌سازی کانتینر یا حضور در صف انتظار برای تخصیص منابع باشد. بنابراین، با به‌کارگیری یک استراتژی مناسب می‌توان این زمان را به حداقل رسانید.

متأسفانه تاخیر شروع سرد باعث شده تا توسعه‌دهندگان اقبال کمتری به استفاده از پلتفرم‌های بدون سرور

<sup>1</sup>First Cold Start

داشته باشند. به گونه‌ای که از بین ۱۰۰۰ برنامه‌ی کاربردی بزرگ در پلتفرم Microsoft Azure، تنها یک مورد مربوط به یک برنامه تجاری باشد [۴]. این موضوع نشان می‌دهد که علی‌رغم پتانسیل بالای رایانش بدون سرور، وجود مشکلات جدی از جمله شروع سرد، باعث امتناع توسعه‌دهندگان از مهاجرت به پلتفرم‌های بدون سرور باشد.

## ۵.۱ نتیجه‌های مهم تحقیق

لورم ایپسوم متن ساختگی با تولید سادگی نامفهوم از صنعت چاپ، و با استفاده از طراحان گرافیک است، چاپگرها و متون بلکه روزنامه و مجله در ستون و سطرآنچنان که لازم است، و برای شرایط فعلی تکنولوژی مورد نیاز، و کاربردهای متنوع با هدف بهبود ابزارهای کاربردی می‌باشد، کتابهای زیادی در شصت و سه درصد گذشته حال و آینده، شناخت فراوان جامعه و متخصصان را می‌طلبد، تا با نرم افزارها شناخت بیشتری را برای طراحان رایانه‌ای علی‌الخصوص طراحان خلاق، و فرهنگ پیشرو در زبان فارسی ایجاد کرد، در این صورت می‌توان امید داشت که تمام و دشواری موجود در ارائه راهکارها، و شرایط سخت تایپ به پایان رسد و زمان مورد نیاز شامل حروفچینی دستاوردهای اصلی، و جوابگوی سوالات پیوسته اهل دنیای موجود طراحی اساساً مورد استفاده قرار گیرد.

ساختار این گزارش به این ترتیب خواهد بود:

درادبیات موضوع مروری بر واژگان، مفاهیم تخصصی و هر آنچه که در ادامه به آن نیاز پیدا خواهیم کرد، خواهیم داشت. سپس در فصل کارهای مرتبط به بیان مشروح تحقیقات انجام شده خواهیم پرداخت و در نتیجه‌گیری و کارهای آینده خلاصه‌ای از مطالعات انجام شده و مسائل باز خواهیم پرداخت.

## فصل ۲

### مروری بر ادبیات

در این بخش سعی داریم تا با مروری بر اصطلاحات و ابزارهای مورد استفاده در پژوهش‌های بررسی شده، با پیش‌نیازهای مبحث موردنظر آشنا شویم.

## فصل ۳

### کارهای مرتبط

لورم ایپسوم متن ساختگی با تولید سادگی نامفهوم از صنعت چاپ، و با استفاده از طراحان گرافیک است، چاپگرها و متون بلکه روزنامه و مجله در ستون و سطرآنچنان که لازم است، و برای شرایط فعلی تکنولوژی مورد نیاز، و کاربردهای متنوع با هدف بهبود ابزارهای کاربردی می باشد، کتابهای زیادی در شصت و سه درصد گذشته حال و آینده، شناخت فراوان جامعه و متخصصان را می طلبد، تا با نرم افزارها شناخت بیشتری را برای طراحان رایانه ای علی الخصوص طراحان خلاقی، و فرهنگ پیشرو در زبان فارسی ایجاد کرد، در این صورت می توان امید داشت که تمام و دشواری موجود در ارائه راهکارها، و شرایط سخت تایپ به پایان رسد و زمان مورد نیاز شامل حروفچینی دستاوردهای اصلی، و جوابگوی سوالات پیوسته اهل دنیای موجود طراحی اساسا مورد استفاده قرار گیرد.

## فصل ۴

### نتیجه گیری

در فصل قبل در ارتباط با راهکارهای ارائه شده توسط هر یک از مقاله‌ها به تفصیل، تشریح نمودیم. در این فصل می‌خواهیم به بیان نتایج و مقایسه راهکارهای ارائه شده بپردازیم. در ابتدا به مقایسه‌های روش‌های مبتنی بر مکانیزم DRX/DTX با یکدیگر می‌پردازیم.



## مراجع

- [1] P. Castro, V. Ishakian, V. Muthusamy, and A. Slominski, "The rise of serverless computing," *Commun. ACM*, vol.62, p.44-54, Nov. 2019.
- [2] P.-M. Lin and A. Glikson, "Mitigating cold starts in serverless platforms: A pool-based approach," *arXiv preprint arXiv:1903.12221*, 2019.
- [3] A. Mohan, H. Sane, K. Doshi, S. Edupuganti, N. Nayak, and V. Sukhomlinov, "Agile cold starts for scalable serverless," in *11th {USENIX} Workshop on Hot Topics in Cloud Computing (HotCloud 19)*, 2019.
- [4] M. Shahradd, R. Fonseca, I. Goiri, G. Chaudhry, P. Batur, J. Cooke, E. Laureano, C. Tresness, M. Russinovich, and R. Bianchini, "Serverless in the wild: Characterizing and optimizing the serverless workload at a large cloud provider," in *2020 {USENIX} Annual Technical Conference ({USENIX} {ATC} 20)*, pp.205-218, 2020.