

بسم الله الرحمن الرحيم



## دانشکده مهندسی برق و کامپیوتر

شبکه‌های پیشرفته

*Kua: A Distributed Object Store over Named Data Networking*

سید مهدی رضوی

استاد : آقای دکتر یزدانی

آذر ماه ۱۴۰۲

## فهرست مطالب

۳	..... خلاصه مقاله	۱۰۰
۴	..... معرفی برخی رویکردها	۲۰۰

## ۱.۰ خلاصه مقاله

برنامه‌هایی مانند سیستم‌های آموزش یادگیری ماشین یا لاگ مجموعه، حجم زیادی از داده را تولید و مصرف می‌کند. سیستم‌های ذخیره‌سازی اشیاء یک انتزاع ساده برای ذخیره و دسترسی به چنین مجموعه داده‌های بزرگی ارائه می‌دهند. این مجموعه از داده‌ها معمولاً دارای ظرفیتی بیشتر از سرورهای خاص منظوره هستند و نیاز به تحمل خطا بالایی دارند. در این مقاله، Kua یک ذخیره‌سازی شی توزیع‌شده را ارائه می‌کنیم. ماهیت داده‌محور kua به NDN (Named Data Network) کمک می‌کند که طراحی ساده خود را حفظ کند، همچنین ارائه تحمل خطا، تأخیر کم و تضمین قوام قوی، همراه با امنیت داده محور از دیگر مزیت‌های آن است. ارزیابی اولیه نشان می‌دهد که Kua می‌تواند از قابلیت‌های NDN برای تحویل داده‌های چندپخششی و حافظه پنهان درون شبکه برای دستیابی به کارایی بالاتر نسبت به سیستم‌های ذخیره‌سازی اشیاء موجود استفاده کند.

## ۲.۰ معرفی برخی رویکردها

داده‌ها در سیستم‌های نرم افزاری همه جا حاضر هستند، و اندازه این داده‌ها طی دو دهه گذشته به طور تصاعدی افزایش یافته است، سیستم‌هایی که این داده‌ها را ذخیره و پردازش می‌کنند نیز همینطور است. سیستم‌های ذخیره سازی توزیع شده اولیه شامل سیستم‌های فایل توزیع شده، مانند سیستم فایل Hadoop (HDFS) و سیستم فایل گوگل (GFS) ذخیره سازی مقیاس شده به پتابایت داده، با این حال محدود به یک واحد بودند. با ظهور پلتفرم ابری، سیستم‌های ذخیره سازی جدید مانند سرویس ذخیره سازی ساده آمازون (S3) ظهور کرد، که همان مقیاس را ارائه کرد، تحمل خطا و قابلیت اطمینان در یک معماری چند واحدی سیستم‌های مشابه در حال حاضر در ابرهای مختلف ارائه می‌شوند پلتفرم‌هایی مانند Cloud Google، Storage Azure و غیره.

آمازون S3 و سیستم‌های مشابه خدمات ذخیره سازی داده‌ها را ارائه می‌دهند به کاربران به عنوان یک فروشگاه اشیاء، که در آن داده‌ها به عنوان اشیایی که توسط یک کلید منحصر به فرد ارجاع داده شده اند ذخیره می‌شوند. این اجسام می‌توانند از نظر اندازه بسیار بزرگ باشند، و داده‌ها را می‌توان با استفاده از اولیه‌های ساده ارائه شده توسط فروشگاه روی اشیاء نوشت و خواند. خدمات ذخیره سازی اشیاء در دهه گذشته محبوبیت زیادی پیدا کرده است. و به ذخیره سازی داده‌ها تبدیل شده اند انتخاب برای انواع حجم کار، از جمله مجموعه داده‌ها برای یادگیری ماشین، آیت‌های رسانه، داده بدون ساختار، حافظه پنهان فایل‌های بزرگ و غیره.

سیستم‌های ذخیره سازی اشیاء می‌توانند از ذخیره سازی روی دیسک مانند S3 متغیر باشند باقی ماندن اشیاء روی دیسک، سیستم‌های درون حافظه مانند پلاسما و Redis برای برنامه‌های کاربردی با کارایی بالا. برخی از این سیستم‌ها، به عنوان مثال. پلاسما بخشی از یک سیستم توزیع شده بزرگتر مانند Ray است که اهمیت ذخیره سازی اشیاء را در سیستم‌های نرم افزاری که امروزه استفاده می‌کنیم را نشان می‌دهد.

امروزه اکثر این سیستم‌ها با استفاده از پشته پروتکل TCP/IP. متاسفانه، سرویس تحویل دیتاگرام نقطه به نقطه IP با ماهیت داده محوری ناسازگار است.

خدمات ذخیره سازی اشیاء از آنجایی که IP میزبان‌های پایانی را شناسایی می‌کند و برنامه‌ها تکه‌های داده را شناسایی می‌کنند، سیستم‌های ذخیره سازی مبتنی بر IP نیاز دارد به به طور مداوم شناسه‌های داده‌های برنامه را به مکان‌های نقشه برداری کنید. داده‌ها، یعنی آدرس‌های IP گره‌های ذخیره سازی. علاوه بر این، چندین مورد از این سیستم‌ها به تعداد زیادی اتصال TCP نیاز دارند، در بدترین حالت  $N \times N$  اتصالات با  $N$  تعداد گره‌های ذخیره‌سازی است (مثل در مورد Redis، Cluster یک پیاده‌سازی توزیع‌شده از Redis)

این نشان دهنده یک چالش مقیاس بندی است. چندپخشی بودن IP، یک عامل کلیدی برای تکثیر کارآمد داده‌ها، به طور گسترده در شبکه‌های IP مستقر نیست و فاقد پروتکل‌های حمل و نقل کارآمد و پیاده سازی‌های امنیتی قابل استفاده است.