



چکیده

متن درباره Kua است، یک سامانه ذخیره سازی اشیاء توزیع شده که از شبکه داده های نام دار (NDN) استفاده می کند. هدف Kua ارائه یک راه حل ساده و کارآمد برای ذخیره و دسترسی به مجموعه داده های بزرگ با بهره گیری از ماهیت محور داده ای NDN است. این سامانه از یک پروتکل حراج توزیع شده برای توزیع و تکرار داده در انبارهای ذخیره سازی چندگانه استفاده می کند، که اطمینان از مقاومت در برابر خطا، کاهش تاخیر، تضمین قوی هماهنگی، و امنیت محور داده ای را فراهم می کند. ارزیابی Kua نتایج امیدوارکننده ای را در افزایش کارایی و عملکرد نسبت به سیستم های ذخیره سازی اشیاء موجود نشان می دهد.

خلاصه

مقدمه

متن مورد بحث درباره نیاز به ذخیره و پردازش داده ها در سیستم های نرم افزاری است و نشان می دهد که با گسترش حجم داده ها، سیستم های ذخیره سازی نیز رشد یافته اند. نظیر به سیستم های فایل توزیع شده از قبیل Hadoop File System و Google File System، که امکان ذخیره سازی داده های پتابایتی را فراهم می کردند. با ظهور پلتفرم ابری، سیستم های جدیدی نظیر Amazon S3 ظاهر شدند که امکان ذخیره سازی داده ها با مقیاس، مقاومت در برابر خطا و قابلیت اطمینان در یک معماری چندمشتري (مستاجری) فراهم می کنند. معماری ساده و ویژگی های Kua نسبت به سیستم های موجود بر مبنای TCP/IP مورد بررسی قرار می گیرد، و همچنین نحوه اجرای تضمینات قوی موجود در Kua تشریح می شود.

اشیاء ذخیره سازی: تحقیقات اولیه در زمینه ی ذخیره سازی توزیع شده به ایجاد دیسک های وابسته به شبکه (Network Attached Disks) به عنوان یک جایگزین برای سیستم های فایل توزیع شده منجر شد، و این ایده به سیستم های ذخیره سازی شیء (Object Storage) امروزی تکامل یافت.

شبکه در ذخیره سازی توزیع شده: بهینه سازی های محلی در ذخیره سازی توزیع شده معمولاً بر روی حافظه پنهان (Caching) در حافظه، دسترسی گمان زنی و سرعت بیشتر با استفاده از تکنیک هایی مانند RDMA تمرکز دارند. ساختار محوری پروتکل IP به افزایش پیچیدگی سیستم و هماهنگی بین گره ها منجر می شود. عدم وجود مالتی کست در TCP/IP به هزینه ها و پیچیدگی بیشتری در سیستم ها منجر می شود.

شبکه ی نام دار (NDN): یک لایه ی شبکه ای متمرکز بر داده را ارائه می دهد که از نام های معنایی درخواست ها و امنیت داده با استفاده از امضاءهای رمزنگاری شده ی (Cryptography) بر مبنای نام (Name-Based) استفاده می کند. در NDN، داده با نام خود توسط کاربران درخواست و امنیت یافته و تا زمانی که توانایی شناسایی داده را داشته باشند، نیازی به اطلاع از مکان ذخیره سازی آن نیست. پشتیبانی از تحویل داده به صورت مالتی کست را فراهم می کند و این امکان را ایجاد می کند که دیگر نیازی به ارسال جداگانه ی کپی داده به هر گره نباشد.

اهداف این طراحی به شرح زیر است: طراحی ساده، عملیات اتوماتیک، راحتی استفاده و امنیت.

چه ذخیره کنیم؟ واحدهای ذخیره سازی یا شی را کوچک ترین واحد ذخیره سازی در این طراحی می نامیم و اپلیکیشن به آن یک آی دی معنی دار می دهد. اشیا در سطح پایین با یکدیگر جمع شده و به صورت مجموعه ای از بسته های NDN ذخیره می گردند.

کجا ذخیره کنیم؟ در یک فضای ذخیره‌سازی توزیع شده که نیازمند مکانیزمی است تا مشخص شود هر داده در دست کدام گره است. برا اینکار از تابع hash استفاده شده است تا داده را به زیرمجموعه‌هایی به نام سبد تقسیم و به گره‌ها انتساب دهد. چگونه پیدا کنیم؟ با تشخیص اینکه دیتا برای کدام سبد است جای آن مشخص می‌شود.

مکانیزم حراج در Kua

به منظور تخصیص و تکثیر داده‌ها بین گره‌های ذخیره‌سازی در سیستم به کار می‌رود. این مکانیزم برای ایجاد توازن در توزیع داده‌ها و مدیریت این فرآیند به طور کارآمد و غیرمتمرکز طراحی شده است. در ابتدا، فضای نام کلیه واحدهای قابل آدرس‌دهی در سیستم به بخش‌های ثابتی با استفاده از هش ثابت تقسیم می‌شود که به عنوان "سبدها" شناخته می‌شوند. این سبدها هر کدام شامل یک گروه از داده‌ها هستند که توسط گره‌های ذخیره‌سازی حراج می‌شوند. پروسه حراج به صورت دوره‌ای توسط یک گره مستر اجرا می‌شود. در هر دوره، گره‌های ذخیره‌سازی برای هر سبد پیشنهاد خود را با ملاحظه شرایط محلی ارسال می‌کنند. مستر همه پیشنهادها را جمع‌آوری کرده و برندگان هر سبد را اعلام می‌کند. برندگان سبد با تأیید پیروزی خود اعلام می‌کنند. برندگان سبد به عنوان گره‌های مقصد برای ذخیره و تکثیر داده‌های متعلق به آن سبد مشخص می‌شوند. پروسه حراج به این ترتیب بر روی تمام سبدها اجرا شده و کل فضای نام قابل آدرس‌دهی را پوشش می‌دهد. این مکانیزم حراج از شفافیت و قابلیت مقیاس‌پذیری بالا برخوردار است. با توجه به این که گره‌ها مستقل از یکدیگر عمل می‌کنند و هر گره تنها شرایط محلی خود را در نظر می‌گیرد، نیاز به یک کنترل‌گر مرکزی یا استفاده از مکانیزم‌های پیچیده برای تضمین تطابق و قفل‌زنی نداریم. این ساختار غیرمتمرکز اجازه می‌دهد تا سیستم به طور مستقل و با پویایی درخشان به تغییرات در شرایط محلی و ایجاد توازن در توزیع داده‌ها واکنش نشان دهد.

پروتکل اشیا

با توجه به اینکه Named Data Networking (NDN) برای انتقال داده‌ها به میزبان‌های نهایی نیازی به نام‌گذاری ندارد، Kua همچنین یک طراحی محور داده را برای واسطه ذخیره‌سازی انتخاب می‌کند. این موضوع امکان پیکربندی و تنظیم حداقلی را برای مشتریان قبل از دسترسی به یک خوشه Kua فراهم می‌کند، زیرا مشتری برای ذخیره و دسترسی به داده‌ها نیازی به شناخت آدرس IP یک مستر یا هر گره از خوشه ندارد. این طراحی محور داده باعث سهولت تعامل مشتریان می‌شود، زیرا آنها می‌توانند بدون نگرانی از جزئیات پایین‌تر از آدرس‌های شبکه، بر روی نام‌های داده‌ها تمرکز کنند. مشتریان می‌توانند به طور بی‌دردسری به خوشه Kua دسترسی یابند بدون نیاز به پیکربندی پیچیده، که باعث سادگی استفاده و کاهش هزینه‌های تنظیم می‌شود. این با اهداف کلان طراحی Kua همخوانی دارد و بر سادگی و عملیات خودکار به منظور ذخیره‌سازی توزیعی با کارایی و کاربرپسند تأکید دارد. عملیات درج شی، بازیابی شی و بازپسگیری فضا نیز به همراه API‌های مورد نیاز در این پروتکل تعریف شده است.