پاسخ به سولات چالش همراه اول

پارسا اسدنژاد

۱ .تحلیل مقیاس‌پذیری و مدیریت نرخ بالای داده (۵۰ تا ۵۰,۰۰۰ رکورد بر ثانیه)

ترکیب فعلی پایتون تک‌رشته‌ای و نوشتن مستقیم روی فایل CSV در نرخ‌های بالا پاسخگو نیست؛ چراکه گلوگاه‌های اساسی نظیر محدودیت GIL در پارس هم‌زمانJSON ، سربار الحاق رشته‌ها در بافر، و قفل‌های I/O ناشی از فراخوانی‌های مکرر دیسک مانع پردازش بلادرنگ می‌شوند. همچنین فایل CSV فاقد قابلیت‌های چندنویسنده‌ای، ایندکس‌گذاری و فشرده‌سازی است. برای مقیاس‌پذیری پایدار، باید لایه «دریافت داده» از «پردازش و ذخیره‌سازی» تفکیک شود. با قرار دادن یک صف پیام توزیع‌شده و بادوام) مانند(Apache Kafka ، پیام‌های خام بلافاصله پس از دریافت بافر می‌شوند. سپس چند پردازشگر مستقل (Consumer Groups)به‌صورت موازی اعتبارسنجی را انجام می‌دهند. داده‌ها به‌صورت دسته‌ای (Batch) در فرمت‌های ستونی بهینه‌سازی‌شده نظیر Parquet یا پایگاه‌های داده سری‌زمانی) مانند ClickHouse یا (TimescaleDB ذخیره می‌شوند و محاسبات آماری لحظه‌ای نیز از طریق موتورهای پردازش جریانی (Stream Processing) انجام می‌گیرد.

۲ .معماری مدیریت چند جریان هم‌زمان (سوکت‌های مستقل به ازای هر مدل)

برای پردازش هم‌زمان چند سوکت مجزا، پیاده‌سازی بر پایه رویکرد ناهمگام (Async I/O) با (`asyncio` یا مدل چندرشته‌ای (Multi-threading)همراه با صف مشترک ضروری است. در این ساختار، به ازای هر اتصال یک Task یا Thread مجزا برای دریافت، بافرینگ و پارس اولیه داده‌ها اختصاص می‌یابد، اما جهت جلوگیری از تداخل و مسابقه داده‌ها(Race Condition) ، تمامی داده‌های پردازش‌شده به یک صف thread-safe هدایت می‌شوند تا یک نویسنده واحد (Single-Writer) آن‌ها را روی دیسک بنویسد. علاوه بر این، مکانیزم گزارش‌گیری ۲۰ ثانیه‌ای باید از چرخه دریافت رویدادها کاملاً جدا شود؛ به‌طوری‌که یک تایمر مستقل و دوره‌ای، معیارهای تجمیعی را به‌صورت Thread-safe از ساختارهای مشترک خوانده، در گزارش ثبت و سپس شمارنده‌ها را بازنشانی کند.

۳ .راهکارهای تضمین عدم اتلاف داده(Zero Data Loss)

با توجه به حساسیت تله‌متری شبکه آب و لزوم حفظ تمامی شواهد احتمالی خرابی، باید پایایی داده‌ها در برابر قطع اتصال، کرش و خطاهای ذخیره‌سازی تضمین گردد. در این راستا، ثبت سریع پیام‌های خام در لاگ‌های محلی (Write-Ahead Log) یا صف‌های ماندگار پیش از اعتبارسنجی الزامی است. فرآیند تأیید مصرف پیام (Acknowledgment) تنها پس از درج موفقیت‌آمیز و قطعی در لایه ذخیره‌سازی نهایی انجام می‌شود) الگوی .(At-Least-Once Delivery برای رفع مشکل رکوردهای تکراری ناشی از بازیابی پس از کرش، فرآیند ذخیره‌سازی بر پایه `reading_id` به‌صورت یکتا و پایدار (Idempotent) طراحی می‌شود. در صورت بروز خطای پایگاه داده یا پر شدن حافظه، داده‌های ناموفق به صف خطای موقت (Dead-Letter Queue) منتقل می‌شوند تا پس از رفع نقص پردازش گردند. در نهایت، سیستم کلاینت باید مجهز به مکانیزم اتصال مجدد خودکار با تأخیر تصاعدی (Exponential Backoff)باشد.