بررسی پیکربندی نامناسب سرویس‌های NoSQL در اشل پروژه‌های بزرگ به تفکیک تکنولوژی‌های NoSQL

علیرضا سلطانی نشان

۱۶ آبان ۱۴۰۲

فهرست مطالب

۱ تعریف مسئله

ذخیره‌سازی اطلاعات از مهم‌ترین نیازهای تحلیل کنندگان داده است. امروزه با توجه به پیشرفت صنعت IoT و یادگیری ماشین، تولید داده‌ها بسیار افزایش یافته است به گونه‌ای که بتوان این داده‌ها را به سریع‌ترین روش ممکن در محلی مناسب ذخیره‌سازی و نگهداری کرد. افراد برای ذخیره‌سازی این داده‌ها نیاز به نصب و راه‌اندازی یک سیستم DBM هستند که از طریق یک واسط با زبانی مناسب بتوانند به آن متصل شده و داده‌های دریافتی را بعد از تجزیه و تحلیل آنها در این محل ذخیره‌سازی و مدیریت کنند. امروز محققان ترجیح می‌دهند به دلیل مقیاس پذیری بیشتر، سیستم‌های توزیع شده و قابلیت پایداری بالا از دیتابیس‌های رابطه‌ای به سمت دیتابیس‌های NoSQL مهاجرت کنند. این نوع دیتابیس‌ها امروزه توسط تمام اپلیکیشن‌های جدید پشتیبانی می‌شوند و برای استفاده آسان طراحی شده‌اند. حتی می‌توان متذکر شد که تعداد زیادی از سرویس‌های ذخیره‌سازی ابری امروزه‌از سرویس‌های دیتابیسی NoSQL پشتیبانی گسترده‌ای می‌کنند. این ارائه دهندگان اغلب شرکت‌های معروفی مانند Amazon DynamoDB MS Azure CosmosDB Google Cloud Database می‌باشند. همچنین بیشتر این موتورهای دیتابیسی به صورت متن‌باز هستند و توسعه دهندگان زیادی از سرتاسر جهان روی آنها مشغول توسعه هستند.

در سال‌های اخیر، با پدید آمدن و رشد سریع سرویس‌های دیتابیسی NoSQL بین عموم توسعه‌دهندگان استفاده از این نوع سرویس‌ها افزایش یافته است. دلیل اصلی این محبوبیت نصب و راه‌اندازی و استقرار آسان آنها در هر محلی است. همچنین قابل اعتماد هستند، روش‌ها و مکانیزم‌های زیادی برای تهیه نسخه‌های پشتیبان‌گیر به صورت منظم از داده‌ها را ارائه می‌دهند. دلیل اصلی آسان بود این سیستم آن است که در هنگام راه‌اندازی آنها زمان زیادی را صرف نمی‌کنید، زیرا بعد از نصب اولیه و طی کردن فرایند نصب با زدن روی دکمه "بعدی" دیتابیس شما آمادست و می‌توانید از آن در برنامه خود استفاده کنید. بعد از این فرایند هیچ عملیاتی بر روی تعریف دسترسی‌ها، مدیریت

کاربران در استفاده از دیتابیس مانند اختصاص سطح دسترسی، توسط راه‌انداز سیستم DBM صورت نمی‌گیرد. نتیجه این موارد پیکربندی غیر اصولی و اشتباه [۱] سیستم ذخیره‌سازی داده می‌شود که در نتیجه افشای اطلاعات حساس [۲] را به دنبال خواهد داشت.

سوالی که ممکن است در اینجا مطرح شود آن است که چه زمانی پیکربندی نادرست موجب افشای اطلاعات می‌شود؟

در ابتدا بعد از راه‌اندازی این نوع دیتابیس‌ها اولین هدف استفاده از آنها در محیط لوکال در یک شبکه است. اما افشای اطلاعات و پیکربندی اشتباه زمانی رخ می‌دهد که این دیتابیس‌ها در شبکه اینترنت مورد دسترسی قرار گیرند.

محققان با توجه به موارد گفته شده بالا توانسته‌اند یک ابزار خودکار جهت آنالیز و جست و جوی سیستم‌های دیتابیسی NoSQL را توسعه داده‌اند که به وسیله آن می‌توانند پیکربندی نامناسب این سیستم‌های مستقر شده را متوجه شده، موارد آسیب پذیری را گزارش و سپس به صاحبان این دیتابیس‌ها هشداری در جهت در خطر بودن اطلاعاتشان ارسال کنند.

در این گزارش به طور خلاصه تمام موارد انجام شده را در پنج عنوان توضیح می‌دهیم. در ابتدا در مورد چالش‌ها و نحوه تحقیق روی این آسیب پذیری‌ها و عدم وجود پیکربندی مناسب می‌پردازیم. در بخش مدل پیشنهادی بیشتر ماهیت ابزار توسعه داده شده را مطرح می‌کنیم و سپس نتایج اجرای این ابزار را نمایش می‌دهیم و در نهایت به نوآوری و کارهای آینده می‌پردازیم.

۲ چالش‌ها

ابزاری توسعه داده شده است که در یک رنج گسترده‌ای از آدرس‌های IP می‌تواند اینگونه دیتابیس‌ها را اسکن کند و افشای سرویس آنها را تشخیص دهد. این تشخیص به شکل ایمن بدون هیچ نگهداری داده‌ها و یا افشای اطلاعات حساس آنها صورت می‌گیرد. بررسی ضعف پیکربندی‌های صورت گرفته بر روی ۶۷ میلیون ۷۲۶ هزار و ۶۴۱ آدرس IP بوده است که بین بازه زمانی اکتبر ۲۰۱۹ و مارچ ۲۰۲۰ تکمیل شده است. نکته جالب از آنجایی شروع می‌شود که این سرویس‌ها نه تنها به صورت شخصی راه‌اندازی شده‌اند بلکه تعداد ۱۲ هزار و ۲۷۶ نمونه از آنها در ارائه دهندگان سرویس‌های ابری معروف یافت شده است. با توجه به این موضوع در این تحقیق ۷۴۲ مورد آسیب پذیری پیدا شده است که به صورت مستقیم وب سایت این کاربران به دلیل ضعف در پیکربندی به دیتابیس‌های آنها ارجاع دارد این بدان معناست با وجود تنظیمات و پیکربندی پیش فرض و بدون هیچ گونه استراتژی امنیتی، هر کاربر ناشناس دیگری می‌تواند وارد این دیتابیس‌ها شده و آنها را با نظر و سلیقه خودش تغییر و حتی تخریب به قصد اخاذی کند.

۱.۲ بررسی نمونه‌ها

در مارچ ۲۰۲۰، ۷ ترابایت از داده‌های سایت بزرگسالان به صورت صریح از یک نمونه دیتابیس Elastic Search با اطلاعاتی از قبیل، نام کاربران، جنسیت و گرایش‌ها، لاگ‌های مربوط به پرداخت‌هایشان، ایمیل، با ۱۰۸۸ میلیارد رکورد مورد افشار قرار گرفت.

[۱] Misconfigured

[۲] Data Leakage

در نوامبر سال ۲۰۱۹ یک محقق توانست یک نمونه با پورت باز با بیشتر از ۲.۱ میلیارد رکورد از یک دیتابیس را پیدا کند که شامل اطلاعات حساس از تعداد زیادی از کاربران بود.

۳ مدل پیشنهادی

۴ آزمایش‌ها و تحلیل نتایج

۵ نوآوری‌های تحقیق

۶ بخش‌های باقی مانده