پیش نیازهای محصول SAWMILL

(SAWMILL SYSTEM REQUIREMENTS)

نیازهای سیستمی Sawmill

در این راهنمای نیازهای سیستمی Sawmill در زمان نصب و اجرا برنامه شرح داده شده است.

OS و پلتفرم CPU

Sawmill بر روی هر پلتفرمی اجرا می‌شود اما ما Linux x64 را توصیه می‌کنیم. برای بهترین سرعت، مطابقت و ثبات از Red Hat Enterprise Linux استفاده نمایید. برای دانلودهای RedHat labeled binary می‌توان از سیستم CentOS build استفاده نمود. برای پلتفرم‌های دیگر شاید نیاز به ساختن Sawmill از source code باشد. دیگر OS هایی که مناسب هستند شامل (including x64 Windows, x64 Mac OS, x64 FreeBSD, and x64 Solaris) می‌باشند. معماری‌های CPU غیر از x64 را می‌توان استفاده نمود اما بهترین عملکرد با x64 در مقایسه با معماری‌های دیگر مانند SPARC یا RISC دیده‌شده است. برای حجم داده‌ای بزرگ معماری 32-bit توصیه نمی‌گردد. محدودیت‌های OS های 32-bit در Sawmill در زمان پروسه کردن حجم بالای داده error ایجاد می‌کند.

حجم نصب Sawmill حدود ۱۰ گیگ یا بیشتر در نظر گرفته‌شده است. با این نوع نصب بهتر است از سیستم 64-bit استفاده شود. در نصب‌های حجم بسیار بالا بهتر است از یک سرور اختصاصی استفاده شود.

Sawmill در حالت معمول بر روی یک سرور مجازی بدون مشکل اجرا می‌شود اما سرعت آن ممکن است کندتر باشد. به خاطر محیط مجازی ممکن است در مواردی error نیز دیده شود. ما اجرا در سخت‌افزار را توصیه می‌کنیم بخصوص برای حجم داده‌های بالا.

هارد و فضای لازم:

اصولاً شما نیاز به ۲۰۰ الی ۴۰۰٪ فضا بیش از داده‌های لاگ خام خود برای ذخیره پایگاه داده‌های Sawmill نیاز دارید و فضای بیشتر از این نیز توصیه می‌گردد. پایگاه داده‌ها در سیستم‌های x64 بیش از ۴۰۰٪ فضا نیاز دارند بخصوص با داشتن فیلدهای اعدادی بزرگ (بیش از ۱۰ رقم). برای مثال در زمان گزارش‌گیری از داده‌های لاگ ۱ ترابایتی در یک پروفایل، نیاز به ۴ ترابایت فضا برای پایگاه داده‌ها است. این جمع داده‌ها در پایگاه داده‌ها است و شامل داده‌های افزوده روزانه نمی‌شود. اگر ۱ ترابایت داده‌های لاگ در روز داشته باشید و می‌خواهید ۳۰ روز را پیگیری کنید شما یک سری داده ۳۰ ترابایتی دارید و حدود ۶۰ الی ۱۲۰ ترابایت فضا در هارد نیاز خواهید داشت.

توجه: اگر پروفایل از یک فیلتر پایگاه داده‌ها برای sort جدول‌های اصلی استفاده می‌کند (برای مثال: یک جلسه تحلیل که بیشتر سرورهای لاگ در وب آنرا به صورت پیش‌فرض انجام می‌دهند) یک مرحله build پایگاه داده‌ها وجود دارد که در آن تمامی جدول‌های اصلی پایگاه داده‌ها کپی و sort می‌گردد. در این مرحله پایگاه داده‌ها دو برابر فضای هارد را اشغال خواهد نمود و پس از build به حجم نهایی بازخواهد گشت. پس در مثال بالا اگر تحلیل session یا فیلترهای sort پایگاه داده‌های دیگری وجود دارد موقتاً بین ۴ الی ۸ ترابایت فضا در روز استفاده خواهد شد و سپس بازگشت به ۲ الی ۴ ترابایت با اتمام build امکان‌پذیر است.

اگر از یک سرور پایگاه داده‌های SQL جداگانه استفاده می‌کنید، برای اجرا سرور فضای کافی در نظر بگیرید. پایگاه داده‌ها از هارد سرور استفاده خواهد کرد و بقیه Sawmill در فضایی حدود ۱ گیگ جای خواهد گرفت.

در زمان build و تولید گزارش Sawmill از هارد استفاده سنگین خواهد کرد و توصیه می‌گردد که برای عملکرد بهتر از یک هارد پرسرعت استفاده کنید. استفاده از هارد پرسرعت RAID ۱۰ ردیفی بهتر است. RAID ۵ یا ۶ عملکرد را بسیار

کاهش می‌دهد و حدوداً ۲ برابر کندتر می‌کند. Write buffering در تنظیمات RAID بهتر است فعال باشد چون به شما امکان دو برابر عملکرد در build پایگاه داده‌ها را می‌دهد.

Network mounts معمولاً برای ذخیره پایگاه داده‌های Sawmill عمل می‌کند اما به دلیل عملکرد ضعیف توصیه نمی‌شود. با Network mounts در مواقعی به دلیل موارد locking و synchronization خطاهایی دیده می‌شود.

حافظه:

در سرور Sawmill حداقل حافظه RAM دو گیگ در هر core برای حجم داده‌های بالا توصیه می‌شود. اگر از Sawmill Professional استفاده می‌کنید که فقط Single-thread اجرا می‌شود از حافظه ۲ گیگ استفاده کنید.

پروسسر ها (Processor(s)):

برای پیش‌بینی مقدار توان پروسسر فرض کنید که Sawmill ۲۰۰۰ ردیف لاگ در ثانیه را در هر پروسسر (Intel یا AMD) پروسه می‌کند. برای پروسسر های SPARC و غیره این ۱۰۰۰ ردیف لاگ در ثانیه می‌باشد.

توجه: این‌یک مثال میانه راه است و Sawmill برای برخی از گروه داده‌ای سریع‌تر عمل می‌کند و می‌تواند به‌سرعت های ۱۰۰۰۰ الی ۲۰۰۰۰ ردیف لاگ در ثانیه در هر core برسد. اما برای دانستن اندازه پروسسر بهتر است از یک برآورد میانه راه استفاده کنید تا سیستم متناسبی داشته باشید.

تعداد ردیف‌ها را در گروه داده‌های روزانه خود محاسبه کنید (۲۰۰ بایت در هر ردیف پیشنهاد می‌گردد) و آنرا بر تعداد ردیف‌ها در ثانیه تقسیم کنید تا سرعت پروسه شدن توسط Sawmill به دست آید. حال شما تعداد ثانیه‌ها را برای build پایگاه داده‌ها توسط Sawmill به دست آورده‌اید. ثانیه‌ها را به ساعت تبدیل کنید. اگر بیش از ۶ ساعت محاسبه شد به بیش از یک پروسسر نیاز است. بهتر است تعداد کافی پروسسر داشته باشید که در زمان تعداد ساعات پروسه کردن شما / تعداد ساعات موردنیاز برای پروسه کردن = کمتر از ۶ ساعت

برای مثال:

50 Gigabytes (GB) of uncompressed log data per day

- divide by 200 -> ~250 million lines of log data
- divide by 2000 -> ~125000 seconds
- divide by 3600 -> ~35 hours
- divide by 6 -> 6 processors

زمان ۶ ساعت بر اساس این فرضیه است که نمی‌خواهید بیش از ۶ ساعت در شب را برای به‌روزرسانی و اضافه کردن داده‌های جدید در پایگاه داده‌ها صرف کنید. برای شروع کار زمان build ۶ ساعت در شب مناسب است. شما با این کار انعطاف برای تغییر یا تنظیم دقیق پایگاه داده‌ها و فیلترها رادارید که می‌توان باعث کندی پروسه کردن شود و در محدوده زمان موجود برای پروسه کردن در طول روز خواهید بود. گروه داده‌های مثال بالا را می‌توان با ۴ پروسسر در ۹ ساعت پروسه نمود اگر زمان شبانه اجازه بدهد.

پارس فناوران خوارزم