راهنمای محاسبات با عملکرد بالا (High-Performance Computing - HPC)

درک معماری‌های خوشه‌ای، ابزارها و پیاده‌سازی‌ها

فهرست مطالب

خلاصه اجرایی ۴

دینامیک سیالات محاسباتی (Computational Fluid Dynamics - CFD) ۶

تحلیل داده‌های با عملکرد بالا (High-performance Data Analytics - HPDA) ۷

هوش مصنوعی (AI) و یادگیری ماشین (Machine Learning - ML) ۷

خوشه‌های HPC چیستند؟ ۸

ایستگاه‌های کاری ۸

سرورها ۸

گره‌های محاسباتی ۸

گره‌های هدایتگر ۸

گره‌های ذخیره‌سازی ۸

سیستم‌عامل ۸

خوشه Provisioner ۸

شبکه ۸

راه‌حل‌های ذخیره‌سازی ۹

سرورها ۱۰

گره‌های محاسباتی ۱۰

گره‌های هدایتگر ۱۱

گره‌های ذخیره‌سازی ۱۱

سیستم‌عامل ۱۲

لینوکس در HPC ۱۲

تأمین خوشه ۱۳

راه‌حل‌های تأمین خوشه ۱۳

شبکه‌ها ۱۵

راه‌حل‌های شبکه‌سازی ۱۶

ذخیره‌سازی ۱۹
ذخیره‌سازی عمومی ۱۹
سیستم فایل خوشه‌ای ۱۹
Object Storage ۲۰
راه‌حل‌های ذخیره‌سازی ۲۰
برنامه‌ریزی، بارهای کاری و قابلیت جابجایی بارهای کاری ۲۵
زمان‌بندی‌کننده‌ها ۲۶
راه‌حل‌های مدیریت بار کاری در HPC ۲۶
کتابخانه‌های MPI و کتابخانه‌های محاسبات موازی ۳۱
MPI چیست؟ ۳۱
کتابخانه‌های MPI: ۳۲
راه‌حل‌های MPI ۳۲
بارهای کاری ۳۵
راه حل های بارهای کاری ۳۵
کانتینرها ۳۸
راه‌حل‌های کانتینر ۳۹
خدمات کمکی ۴۱
راه‌حل‌های مدیریت هویت ۴۱
نظارت و مشاهده‌پذیری ۴۲
راه‌حل‌های مشاهده‌پذیری ۴۳
کجا خوشه‌های HPC را اجرا کنیم؟ ۴۵
HPC در فضای عمومی ابری ۴۵
خوشه‌های اختصاصی HPC خصوصی ۴۷
HPC هیبریدی ۴۸
HPC در لبه ۴۸
گام‌های بعدی خود را در HPC با Canonical بردارید ۴۸

خلاصه اجرایی

این کتاب راهنمایی مقدماتی درباره محاسبات با عملکرد بالا یا (HPC) است. این کتاب استفاده‌های مختلف، انواع بارکاری و انواع پردازشی که در HPC وجود دارند را خلاصه می‌کند.

همچنین نمای کلی از خوشه‌های HPC و معماری آن‌ها ارائه داده و بررسی می‌کند که این خوشه‌ها کجا می‌توانند پیاده‌سازی شوند، چه در محل (on-premise) و چه در فضای ابری عمومی.

این کتاب همچنین اجزای مختلفی که در خوشه‌های HPC دخیل هستند را معرفی می‌کند.

به طور کلی، این راهنما به شما کمک می‌کند تا ساختار داخلی خوشه‌های HPC، معماری آن‌ها، موارد استفاده معمول و ابزارهای مرتبط با پیاده‌سازی HPC را درک کنید.

پس از مطالعه این کتاب، باید درک کافی از دنیای HPC داشته باشید و قادر باشید نیازهای خود را برای شروع ارزیابی نمائید.

محاسبات با عملکرد بالا (HPC) چیست؟

محاسبات با عملکرد بالا، منابع محاسباتی را به صورت یک منبع واحد ترکیب می‌کند. این منابع ترکیبی، اغلب به عنوان یک ابررایانه یا خوشه محاسباتی شناخته می‌شوند.

HPC امکان پردازش شدید محاسباتی و اجرای بارهای کاری و برنامه‌های پیچیده را با سرعت بالا و به صورت موازی فراهم می‌کند.

HPC در مقابل ابررایانه‌ها

این روزها، ابررایانه‌ها به عنوان مترادف محاسبات با عملکرد بالا شناخته می‌شوند. با این حال، این دو دقیقاً قابل جایگزینی نیستند، ابررایانه‌ها و ابررایانه‌سازی معمولاً به استقرارهای خوشه‌ای بزرگ‌تر و محاسباتی که در آنجا انجام می‌شود اشاره دارند.

HPC بیشتر به محاسباتی اشاره می‌کند که با استفاده از رایانه‌های بسیار سریع در خوشه‌های کوچک تا بزرگ انجام می‌شوند. اغلب، خوشه‌های HPC و ابررایانه‌ها معماری مشابهی دارند و از سرورهای کالایی ساخته می‌شوند.

تاریخچه ابررایانه‌سازی به استفاده از مجموعه‌ای از سیستم‌های خاص اشاره دارد که مشابه خوشه‌های HPC امروزی، بر روی مسائل بسیار پیچیده یا داده‌محور کار می‌کردند. با این حال، مقایسه قدرت محاسباتی این دو شبیه مقایسه سیب و گلابی است. حتی یک تلفن همراه امروزی قدرتمندتر از اولین ابررایانه‌ها است.

برای مثال، برخی تلفن‌های همراه می‌توانند به چندین گیگافلاپس (GFLOPS)[1] برسند، در حالی که CDC 6600، یک ابررایانه که در دهه ۱۹۶۰ توسط سی‌مور کری طراحی شده بود، حدود سه مگافلاپس را تحویل می‌داد.

در آن زمان، ابررایانه‌ها قدرتمندتر از هر چیز دیگری در بازار بودند و بسیار گران‌قیمت بودند.

برخلاف خوشه‌های HPC فعلی، ابررایانه‌ها از لحاظ معماری بسیار متفاوت بودند. در نهایت، آن‌ها سیستم‌های چندپردازنده‌ای بزرگ با قابلیت‌های بسیار خاص بودند و عمدتاً در حوزه دولت‌ها و دانشگاه‌های پژوهشی استفاده می‌شدند. اما امروز، سیستم‌های HPC در طیف وسیعی از صنایع کاربرد دارند.

موارد استفاده اصلی HPC

HPC برای حل برخی از پیچیده‌ترین و دشوارترین مسائل محاسباتی امروزی استفاده می‌شود. این مشکلات در تمام بخش‌ها، از جمله علم، مهندسی، یا تجارت وجود دارند. برخی از موارد استفاده محبوب HPC شامل موارد زیر است:

- مدل‌سازی اقلیمی و پیش‌بینی آب و هوا
- اکتشاف نفت و گاز
- مهندسی خودروسازی و هوافضا
- تحلیل مالی و ارزیابی ریسک
- کشف دارو و مطالعه پروتئین
- پردازش تصویر و ویدئو
- رمزگشایی معکوس و تشخیص نفوذ در امنیت سایبری
- پژوهش و تحلیل ژنوم‌ها

[1] Giga Floating Point Operations Per Second

این موارد استفاده با معادلات عددی حل می‌شوند، مانند آن‌هایی که در دینامیک سیالات محاسباتی (CFD) وجود دارند. این معادلات داده‌های بزرگ را تحلیل یا پردازش می‌کنند، مانند آن‌هایی که در تحلیل داده‌های با عملکرد بالا (HPDA)، هوش مصنوعی و یادگیری ماشین استفاده می‌شوند.

بارهای کاری برای این موارد استفاده مختلف را می‌توان به یک یا چند نوع مختلف دسته‌بندی کرد، بسته به اینکه چگونه اجرا یا پردازش می‌شوند.

به عنوان مثال، پردازش دسته‌ای (Batch processing) شامل اجرای تعداد زیادی کار مشابه به صورت پشت سر هم است.

پردازش بلادرنگ (Real-time processing) به پردازش داده‌ها در لحظه دریافت آن‌ها اشاره دارد.
پردازش تعاملی (Interactive processing) شامل اجرای برنامه‌های تعاملی مانند شبیه‌سازی‌ها یا بصری‌سازی داده‌ها است.

بیایید برخی از این موارد استفاده را با جزئیات بیشتری بررسی کنیم، چرا که ارتباط نزدیکی با HPC دارند.

دینامیک سیالات محاسباتی (Computational Fluid Dynamics - CFD)

دینامیک سیالات محاسباتی (CFD) شاخه‌ای از علم است که از روش‌های عددی و الگوریتم‌ها برای حل و تحلیل جریان سیالات استفاده می‌کند. این علم برای مطالعه حرکت جامدات، مایعات و گازها و همچنین برای تحلیل و پیش‌بینی تأثیرات جریان‌های سیال بر روی ساختارها و سیستم‌ها به کار می‌رود.

CFD ابزاری مهم برای مهندسان و دانشمندان است، چرا که می‌توان از آن برای مطالعه رفتار سیستم‌های پیچیده در طیف وسیعی از کاربردها، از جمله مهندسی هوافضا، خودروسازی و زیست‌پزشکی استفاده کرد.

تحلیل داده‌های با عملکرد بالا (High-performance Data Analytics - HPDA)

تحلیل داده‌های با عملکرد بالا (HPDA) فرایند تحلیل مقادیر زیادی از داده‌ها با سرعت و کارایی بالا به منظور کسب بینش‌های مفید است.

این فرایند شامل استفاده از تکنیک‌های تخصصی، سخت‌افزار و نرم‌افزارهای خاص برای تحلیل داده‌ها در مقیاس بزرگ، شناسایی الگوها و روندها و اتخاذ تصمیمات بلادرنگ است. تحلیل داده‌های با عملکرد بالا می‌تواند در حوزه‌های مختلفی مانند امور مالی، مراقبت‌های بهداشتی و بازاریابی مورد استفاده قرار گیرد. هدف نهایی آن بهبود بهره‌وری و افزایش سود است.

هوش مصنوعی (AI) و یادگیری ماشین (Machine Learning - ML)

هوش مصنوعی (AI) و یادگیری ماشین (ML) دو حوزه مرتبط از علوم رایانه هستند که بر توسعه سیستم‌های رایانه‌ای که می‌توانند یاد بگیرند، استدلال کنند و تصمیم بگیرند، تمرکز دارند.

AI و یادگیری ماشین شامل استفاده از الگوریتم‌ها برای شناسایی الگوها و روندها در مجموعه داده‌ها و انجام پیش‌بینی‌ها و تصمیم‌گیری بر اساس آن داده‌ها هستند.

AI و ML در کاربردهای متنوعی از جمله داده‌کاوی، پردازش زبان طبیعی، خودروهای خودران و موارد دیگر مورد استفاده قرار می‌گیرند.

حال چگونه HPC کار می‌کند؟ بیایید اجزای پشت معماری خوشه HPC و ابزارهای رایج برای اجرای بارهای کاری HPC را بررسی کنیم. ابتدا تعریف کنیم منظورمان از یک خوشه HPC چیست.

خوشه‌های HPC چیستند؟

خوشه‌های HPC مجموعه‌ای از منابع هستند که به‌طور اصلی برای اجرای بارهای کاری محاسباتی استفاده می‌شوند.

خوشه‌های HPC شامل موارد زیر هستند:

ایستگاه‌های کاری

که با بارهای کاری برای پردازش‌های پیش و پس از پردازش تعامل دارند.

سرورها

که به عنوان گره‌های محاسباتی (Compute Nodes)، گره‌های هدایتگر (Head Nodes) و گره‌های ذخیره‌سازی (Storage Nodes) مستقر می‌شوند.

گره‌های محاسباتی

وظیفه پردازش بارهای کاری را بر عهده دارند.

گره‌های هدایتگر

برای دسترسی کاربران و تعامل با خوشه مورد استفاده قرار می‌گیرند.

گره‌های ذخیره‌سازی

برای ذخیره داده‌ها، چه ذخیره‌سازی عمومی و چه محاسباتی به کار می‌روند.

سیستم‌عامل

که برای اجرای سرورها استفاده می‌شود.

خوشه Provisioner

که اطمینان از همگنی گره‌ها را تضمین کرده و برای پیاده‌سازی سیستم‌عامل سرورها استفاده می‌شود.

شبکه

برای ارتباط بین گره‌ها

راه‌حل‌های ذخیره‌سازی

یک راه‌حل ذخیره‌سازی عمومی برای ذخیره برنامه‌ها و داده‌های کاربر.

یک سیستم فایل خوشه‌ای با سرعت بالا و تأخیر کم که معمولاً برای ذخیره‌سازی محاسباتی استفاده می‌شود.

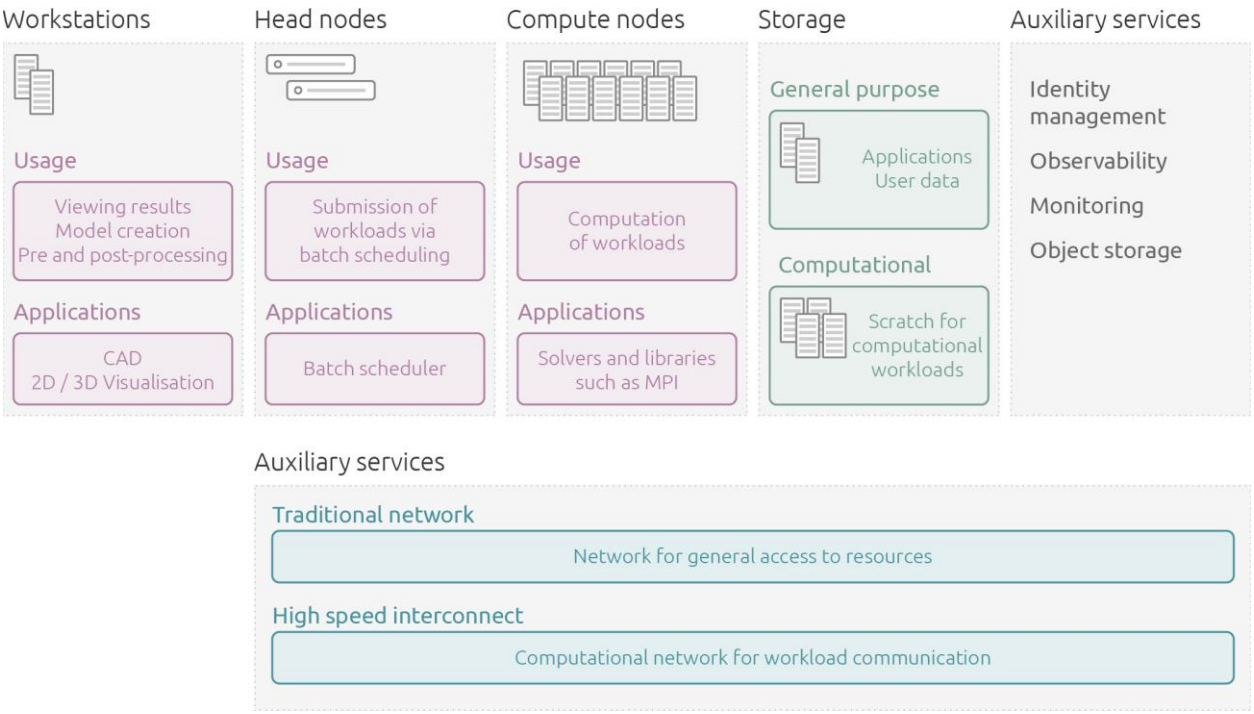
قابلیت‌های زمان‌بندی

بارهای کاری و کتابخانه‌هایی که آن بارهای کاری به آن‌ها وابسته هستند.

خدمات کمکی شامل:

مدیریت هویت برای حفظ یکپارچگی دسترسی کاربران در سراسر خوشه.

یک پشته نظارت و مشاهده‌پذیری که به شما بینشی از استفاده از منابع بارهای کاری می‌دهد.



معماری خوشهHPC

در صفحات بعدی، این اجزا را با جزئیات بیشتری بررسی خواهیم کرد.

سرورها

سرور، کامپیوتر یا سیستمی است که منابع، داده‌ها، خدمات یا برنامه‌ها را به سایر کامپیوترها، که به آن‌ها مشتریان گفته می‌شود، از طریق شبکه ارائه می‌دهد.

سرورها می‌توانند عملکردهای مختلفی را ارائه دهند که معمولاً به آن‌ها خدمات گفته می‌شود، مانند اشتراک‌گذاری داده‌ها یا منابع میان چندین مشتری، یا انجام محاسبات برای یک مشتری.

نمونه‌های رایج از انواع سرورها شامل وب سرورها، سرورهای برنامه، سرورهای پایگاه داده و سرورهای فایل است. در محاسبات با عملکرد بالا، سرورها برای دو هدف اصلی استفاده می‌شوند:

۱. محاسبه مدل‌های ریاضی یا پردازش داده‌ها
۲. ارائه داده‌ها از طریق سرورهای فایل

سرورهایی که برای محاسبه و پردازش داده‌ها استفاده می‌شوند، به‌طور کلی گره‌های محاسباتی (compute nodes) نامیده می‌شوند.

سرورهایی که داده‌ها را ارائه می‌دهند، به‌طور کلی گره‌های ذخیره‌سازی (storage nodes) نامیده می‌شوند.

گره‌های محاسباتی

گره‌های محاسباتی، اجزای پردازش در یک خوشه محاسباتی با عملکرد بالا (HPC) هستند. این گره‌ها بار کاری را با استفاده از منابع محلی مانند CPU، GPU، FPGA و دیگر واحدهای پردازش اجرا می‌کنند.

این بارهای کاری همچنین از دیگر منابع موجود در گره محاسباتی برای پردازش استفاده می‌کنند، مانند حافظه، ذخیره‌سازی و کارت شبکه.

بارهای کاری از پهنای باند موجود در این اجزای زیرین استفاده می‌کنند. بسته به اینکه بار کاری چگونه از این اجزا استفاده می‌کند، ممکن است در هنگام اجرا توسط یکی یا چند مورد از آن‌ها محدود شود. برای مثال، برخی بارهای کاری که مقدار زیادی حافظه استفاده می‌کنند، ممکن است در پهنای باند یا ظرفیت حافظه محدود شوند.

بارهای کاری که مقدار زیادی داده را استفاده می‌کنند یا مقدار زیادی داده در طول محاسبات تولید می‌کنند، ممکن است به دلیل محدودیت‌های پهنای باند شبکه یا عملکرد ذخیره‌سازی، در سرعت پردازش خود محدود شوند.

اگر آن داده‌ها به عنوان بخشی از محاسبات بار کاری در ذخیره‌سازی نوشته شوند. برخی بارهای کاری ممکن است فقط به منابع محاسباتی زیادی نیاز داشته باشند و با توان پردازشی خوشه محدود شوند.

هنگام ایجاد و طراحی این خوشه‌ها، مهم است که از استفاده منابع بار کاری آگاه باشید و خوشه را با در نظر گرفتن آن طراحی کنید. بهترین راه برای درک استفاده از منابع بار کاری، نظارت بر منابع استفاده شده است. این کار به شما کمک می‌کند تا محدودیت‌ها را شناسایی کنید.

گره‌های هدایتگر

گره‌های هدایتگر یا گره‌های دسترسی به عنوان نقطه ورود به یک خوشه HPC عمل می‌کنند. کاربران در اینجا با ورودی و خروجی بارهای کاری خود تعامل دارند و به سیستم‌های ذخیره‌سازی محلی که برای خوشه در نظر گرفته شده دسترسی پیدا می‌کنند.

همچنین در اینجا بارهای کاری خود را زمان‌بندی می‌کنند. برنامه‌ریز (Scheduler) به نوبه خود فرآیندها را روی گره‌های محاسباتی اجرا می‌کند.

گره‌های ذخیره‌سازی

گره‌های ذخیره‌سازی کامپیوترها یا سرورهایی هستند که مسئول ذخیره‌سازی و ارائه دسترسی به داده‌ها از طریق شبکه هستند.

گره‌های ذخیره‌سازی معمولاً به دیگر گره‌های ذخیره‌سازی در یک خوشه ذخیره‌سازی متصل هستند و دسترسی به داده‌های ذخیره‌شده در آن خوشه را فراهم می‌کنند. این گره‌ها اغلب از طریق یک شبکه پرسرعت مانند InfiniBand یا Ethernet به دیگر گره‌های ذخیره‌سازی یا محاسباتی متصل می‌شوند و دسترسی به داده‌ها را به طور مستقیم یا از طریق سیستم فایل فراهم می‌کنند.

پروتکل‌های مختلفی برای ارائه دسترسی به ذخیره‌سازی وجود دارد، از اشتراک‌گذاری سنتی NFS گرفته تا پیاده‌سازی ذخیره‌سازی های مشترک مانند Lustre یا BeeGFS

سیستم‌عامل

برای اجرای گره‌ها، نیاز به یک سیستم‌عامل (OS) است. سیستم‌عامل مسئول مدیریت حافظه، پردازنده، ذخیره‌سازی و دیگر اجزای کامپیوتر است. همچنین، واسطی بین کاربر و سرورها فراهم می‌کند که به کاربران این امکان را می‌دهد تا با کامپیوتر تعامل داشته و برنامه‌ها را اجرا کنند.

سیستم‌عامل‌های رایج در HPC شامل Windows، macOS و Linux هستند.

لینوکس در HPC

سیستم‌عامل لینوکس، که احتمالاً یکی از شناخته‌شده‌ترین پروژه‌های متن‌باز است، به عنوان یک محرک برای نرم‌افزار متن‌باز در HPC و همچنین به وسیله استفاده از موارد کاربرد HPC به توسعه آن کمک کرده است. ناسا یکی از اولین کاربران لینوکس بود و لینوکس به نوبه خود برای اولین خوشه Beowulf نقشی حیاتی داشت.

خوشه‌های Beowulf در اصل خوشه‌هایی بودند که با استفاده از سرورهای تجاری و اتصالات پرسرعت ایجاد شدند و به جای استفاده از سیستم‌های اصلی یا ابررایانه‌های سنتی اولین خوشه Beowulf در ناسا راه‌اندازی شد و به شکل‌گیری HPC به آنچه که امروز می‌شناسیم کمک کرد.

این امر از آن زمان به پذیرش لینوکس در دولت و همچنین به دیگر بخش‌ها کمک کرد. امروزه این نوع خوشه توسط شرکت‌ها نیز استفاده می‌شود.

HPC تلاش‌های زیادی را برای توسعه لینوکس انجام داده است، که تماماً بر کاهش تأخیر و افزایش عملکرد در سطح کل استک (از شبکه‌سازی تا ذخیره‌سازی) متمرکز بوده است.

Ubuntu سیستم‌عاملی است که ۶۶٪ از توسعه‌دهندگان آن را ترجیح می‌دهند و برای HPC ایده‌آل است. این سیستم‌عامل می‌تواند برای ایستگاه‌های کاری، دسترسی به خوشه‌های HPC یا نصب بر روی سرورها استفاده شود و تجربه یکنواختی را در هر دو محیط فراهم می‌کند.

تأمین خوشه

یکنواختی گره‌ها در HPC برای اطمینان از ثبات بار کاری اهمیت دارد. به همین دلیل، معمول است که خوشه‌های HPC با استفاده از راه‌حل‌های Metal-as-a-Service تأمین شوند که به سازمان‌ها کمک می‌کند این زیرساخت را در مقیاس وسیع مدیریت کنند.

راه‌حل‌های تأمین خوشه

MAAS

Metal as a Service یا MAAS، یک پروژه متن‌باز است که توسط Canonical توسعه و نگهداری می‌شود.

MAAS با هدف خاصی ایجاد شده است:

تأمین منابع با تمرکز بر API و بدون استفاده از ماشین‌های مجازی، MAAS تمام جنبه‌های تأمین سخت‌افزار را به‌طور خودکار انجام می‌دهد، از شناسایی یک دستگاه رک گرفته تا استقرار یک سیستم‌عامل سفارشی و در حال اجرا.

این ابزار مدیریت خوشه‌های سرور بزرگ، مانند آن‌هایی که در HPC وجود دارند را از طریق انتزاع و اتوماسیون آسان می‌کند.

MAAS به گونه‌ای طراحی شده است که استفاده از آن آسان باشد، دارای رابط کاربری جامع است و بر خلاف بسیاری از ابزارهای دیگر در این حوزه به دلیل طراحی غیرمتمرکز آن، به شدت مقیاس‌پذیر است.

MAAS به دو بخش تقسیم می‌شود:

۱. کنترل‌کننده منطقه‌ای که وضعیت کلی خوشه را مدیریت می‌کند، از جمله نگهداری اطلاعات مربوط به مشخصات سخت‌افزاری کلی و اطلاعات در مورد سرورهای تأمین‌شده و در دسترس و ارائه تمامی اطلاعات به کاربر
۲. MAAS همچنین با یک کنترل‌کننده رک بدون حالت که وظیفه بوت کردن PXE و کنترل قدرت را بر عهده دارد، ارائه می‌شود. چندین کنترل‌کننده رک می‌توانند مستقر شوند که مقیاس‌پذیری آسان را بدون توجه به اندازه محیط ممکن می‌سازد.

لازم به ذکر است که MAAS می‌تواند در یک پیکربندی با دسترسی بالا مستقر شود، که این ویژگی باعث می‌شود که از نظر تحمل خطا نسبت به پروژه‌های مشابه در صنعت برتری داشته باشد.

xCAT

Extreme Cloud Administration Toolkit یا xCAT، یک پروژه متن‌باز است که توسط IBM توسعه داده شده است.

تمرکز اصلی آن بر روی فضای HPC است و ویژگی‌های آن به‌ویژه برای ایجاد و مدیریت خوشه‌های بدون دیسک، نصب موازی و مدیریت گره‌های خوشه لینوکس طراحی شده است. همچنین، برای راه‌اندازی استک‌های محاسباتی با عملکرد بالا مانند برنامه‌ریزهای دسته‌ای مناسب است.

این ابزار قابلیت کلون کردن و تصویر برداری از ماشین‌های لینوکس و ویندوز را نیز دارد. برخی از ویژگی‌های آن به‌طور ویژه برای سرورهای IBM و Lenovo طراحی شده‌اند. بسیاری از سایت‌های HPC دولتی بزرگ از xCAT برای استقرار خوشه‌های HPC بدون دیسک استفاده می‌کنند.

Warewulf

هدف اعلام‌شده Warewulf این است که "سیستم تأمین سیستم‌عامل بدون حالت و بدون دیسک برای خوشه‌های بزرگ از سیستم‌های فیزیکی و/یا مجازی" باشد. این ابزار برای تأمین خوشه‌های HPC در دو دهه گذشته استفاده شده است. Warewulf به‌تازگی با استفاده از Golang در آخرین نسخه خود، Warewulf v4، دوباره نوشته شده است.

شبکه‌ها

همان‌طور که در بالا ذکر شد، بارهای کاری موازی HPC به شدت به ارتباط بین‌فرایندی وابسته هستند. زمانی که این ارتباط درون یک گره محاسباتی انجام می‌شود، تنها از یک فرآیند به فرآیند دیگر از طریق حافظه آن گره محاسباتی منتقل می‌شود. اما زمانی که یک فرآیند با فرآیندی در گره محاسباتی دیگری ارتباط برقرار می‌کند، این ارتباط باید از طریق شبکه انجام شود.

این ارتباط بین‌فرایندی ممکن است به‌طور مکرر انجام شود. اگر این‌طور باشد، مهم است که شبکه دارای تأخیر کم باشد تا از تاخیر در ارتباط بین فرآیندها جلوگیری شود.

در نهایت، نباید زمان ارزشمند محاسباتی را صرف انتظار برای دریافت پیام‌ها کنید. در مواردی که اندازه داده‌ها بزرگ است، مهم است که این داده‌ها به سرعت منتقل شوند. این امر با شبکه‌های با توان بالا امکان‌پذیر است. هر چه شبکه سریع‌تر بتواند داده‌ها را منتقل کند، زودتر هر فرآیند می‌تواند بر روی بار کاری شروع به کار کند. ارتباط مکرر و اندازه‌های بزرگ پیام و داده ویژگی‌های معمول بارهای کاری HPC هستند.

این امر به ایجاد راه‌حل‌های شبکه‌ای تخصصی منجر شده است که معمولاً تأخیر کم و توان بالا را برای برآورده کردن نیازهای خاص HPC ارائه می‌دهند.

راه‌حل‌های شبکه‌سازی

اترنت

اترنت، فناوری است که به‌طور گسترده برای ارائه اتصال شبکه استفاده می‌شود. برای درک اترنت، اغلب لازم است که مدل OSI را درک کنیم، که اتصال را در هفت لایه توصیف می‌کند:

۱. فیزیکی

۲. پیوند داده

۳. شبکه

۴. حمل و نقل

۵. نشست

۶. ارائه

۷. کاربرد

این مدل جامع است و نیاز به ارتباط قابل اعتماد را برآورده می‌کند. در HPC، جایی که عملکرد و تأخیر از اهمیت بالایی برخوردار است، لایه حمل و نقل که توسط اترنت ارائه می‌شود، گاهی اوقات به‌عنوان لایه ناکارآمد در نظر گرفته می‌شود.

برای مثال، TCP، یک پروتکل حمل و نقل برای اترنت است که نیاز به ارتباط‌های تأیید شده زیادی دارد که بار اضافی را به همراه می‌آورد.

این مشکل با UDP کمتر است، که برای همان سطح از قابلیت اطمینان طراحی نشده است. البته تلاش‌هایی برای بهبود کارایی شبکه‌های مبتنی بر اترنت صورت گرفته است که در ادامه به آن ها اشاره خواهد شد.

RDMA بر روی اترنت همگرا (RoCE) پروتکل شبکه‌ای است که دسترسی به حافظه مستقیم از راه دور (RDMA) را بر روی شبکه اترنت امکان‌پذیر می‌سازد، این کار را با کپسوله کردن یک بسته حمل و نقل InfiniBand (IB) بر روی اترنت انجام می شود.

این امر بسیاری از بارهای اضافی مرتبط با پروتکل‌های حمل و نقل سنتی را حذف می‌کند و در نهایت تأخیر کمتر، بار CPU کمتر و پهنای باند بالاتری را فراهم می‌آورد.

Nvidia InfiniBand

InfiniBand یک فناوری شبکه با سرعت بالا است که در خوشه‌های HPC و سوپرکامپیوترها استفاده می‌شود. این فناوری برای ارتباط داده‌ها بین کامپیوترها و درون کامپیوترها استفاده می‌شود.

همچنین برای اتصال سرورها به سیستم‌های ذخیره‌سازی به‌طور مستقیم یا از طریق یک سوئیچ و همچنین برای اتصال سیستم‌های ذخیره‌سازی به یکدیگر نیز کاربرد دارد.

InfiniBand سرعت بسیار بالا و تأخیر بسیار کم را ارائه می‌دهد، که آن را برای برنامه‌های ذخیره‌سازی و محاسبات با عملکرد بالا، مانند برنامه‌هایی که به MPI برای ارتباط موازی وابسته‌اند، ایده‌آل می‌سازد.

به‌عنوان مثال، نسل آخر InfiniBand اتصال ۴۰۰Gb/s را در هر پورت ارائه می‌دهد. تأخیر یک سوئیچ InfiniBand حدود ۱۰۰ نانوثانیه در مقایسه با حدود ۲۳۰ نانوثانیه برای سوئیچ‌های اترنت است. این ویژگی‌ها باعث شده است که InfiniBand به‌عنوان یک گزینه محبوب برای ارتباطات با سرعت بالا در خوشه‌های HPC شناخته شود.

HPE Cray Slingshot

Slingshot با اترنت سازگار است و در عین حال قابلیت‌هایی مشابه با InfiniBand از نظر توان عملیاتی و تأخیر ارائه می‌دهد. نسل آخر آن اتصال ۲۰۰Gb/s را در هر پورت ارائه می‌دهد. از آنجا که بر پایه اترنت است، ویژگی‌های مناسبی مانند اتصال مستقیم بین سوئیچ‌های HPE Cray Slingshot و سوئیچ‌های اترنت سنتی را ارائه می‌دهد.

Cornellis OmniPath

Cornellis OmniPath که پیش‌تر به‌نام Intel OmniPath شناخته می شد، یک فناوری ارتباط با سرعت بالا است که بر اساس ترکیبی از دو فناوری که اینتل خریداری کرده بود، طراحی شده است:

TrueScale InfiniBand و ارتباط Aries از خط سوپرکامپیوترهای Cray XC.

پس از خرید شرکت Barefoot Networks در سال ۲۰۱۹، اینتل تصمیم گرفت تمرکز خود را بر روی فناوری به‌دست‌آمده از این خرید به جای OmniPath بگذارد.

دلیل این تصمیم این بود که آن‌ها به فرصتی برای ایجاد امکان استفاده‌های ارتباطی با سرعت بالا در سوئیچ‌های دارای ASICS قابل برنامه‌ریزی دست یافته بودند.

اینتل خط تولید OmniPath مبتنی بر این فناوری را به یک شرکت جدید به نام Cornellis منتقل کرد، که به توسعه و نگهداری خط تولید OmniPath خارج از اینتل ادامه می‌دهد.

شبکه بدون سوئیچ Rockport

Rockport یک راه‌حل ارتباط با سرعت بالا مبتنی بر اترنت است که از سوئیچ‌ها اجتناب می‌کند. در این فناوری، NIC ها به‌طور مستقیم در یک شبکه بزرگ متصل هستند و اتصال و مسیریابی را به‌صورت موردنیاز بین یکدیگر فراهم می‌کنند.

NIC هایی که به‌طور مستقیم متصل نیستند، می‌توانند از طریق NIC های متصل، به NIC های دیگر متصل شوند. به‌طور ساده، NIC ها به‌عنوان سوئیچ عمل می‌کنند تا اتصال را فراهم سازند.

ذخیره‌سازی

راه‌حل‌های ذخیره‌سازی در فضای HPC اغلب به‌صورت فایل محور با پشتیبانی از POSIX هستند.

این راه‌حل‌های مبتنی بر فایل، به طور کلی به دو دسته ذخیره‌سازی عمومی و ذخیره‌سازی موازی تقسیم می‌شوند.

راه‌حل‌های دیگر، مانند ذخیره‌سازی شیء یا ذخیره‌سازی Blob (اشیاء باینری بزرگ) که گاهی اوقات در HPC به آن اشاره می‌شود، می‌توانند به‌طور مستقیم توسط برخی بارهای کاری استفاده شوند، اما همه بارهای کاری این قابلیت را ندارند.

ذخیره‌سازی عمومی

دو استفاده اصلی از ذخیره‌سازی عمومی در یک خوشه HPC وجود دارد. یکی برای ذخیره‌سازی باینری‌های برنامه‌های در دسترس و کتابخانه‌های آن‌ها است. این امر به این دلیل مهم است که تمام باینری‌ها و کتابخانه‌ها در سراسر خوشه هنگام اجرای برنامه یکسان باشند، که این امر ذخیره‌سازی مرکزی را مناسب می‌سازد.

دیگری برای دایرکتوری‌های خانگی کاربران و داده‌های دیگر کاربران است، زیرا مهم است که کاربر به‌طور مداوم به داده‌های خود در سراسر خوشه HPC دسترسی داشته باشد. استفاده از یک سرور NFS برای این منظور رایج است، اما پروتکل‌های ذخیره‌سازی دیگری وجود دارد که امکان دسترسی به فایل‌های مبتنی بر POSIX را فراهم می‌آورد.

سیستم فایل خوشه‌ای

سیستم فایل خوشه‌ای یا سیستم فایل موازی، یک سیستم فایل اشتراکی است که منابع ذخیره‌سازی را از چندین سرور ارائه می‌دهد و می‌تواند به‌طور هم‌زمان توسط چندین مشتری استفاده شود.

این به مشتریان دسترسی مستقیم به داده‌های ذخیره‌شده را می‌دهد، که به نوبه خود بارهای اضافی را با اجتناب از انتزاع حذف کرده و منجر به تأخیر کم و عملکرد بالا می‌شود. برخی سیستم‌ها حتی قادر به دستیابی به عملکرد مشابه با عملکرد کلی سخت‌افزار زیرین هستند.

Object Storage

ذخیره‌سازی شیء (Object Storage) معمولاً برای ذخیره‌سازی در خوشه‌های HPC استفاده می‌شود، چه برای بایگانی نتایج محاسباتی گذشته یا داده‌های مرتبط دیگر. به‌علاوه، ممکن است به‌طور مستقیم توسط بارهای کاری که از API های بومی ذخیره‌سازی شیء پشتیبانی می‌کنند، استفاده شود.

راه‌حل‌های ذخیره‌سازی

راه‌حل‌های ذخیره‌سازی مختلفی موجود است که شامل گزینه‌های اختصاصی و متن‌باز می‌شود. در ادامه به برخی از راه‌حل‌های متداول در HPC پرداخته می‌شود

Ceph

Ceph یک راه‌حل ذخیره‌سازی نرم‌افزاری تعریف شده متن‌باز است که بر اساس ذخیره‌سازی شیء پیاده‌سازی شده است.

این فناوری در ابتدا توسط Sage Weil برای یک پایان‌نامه دکترا ایجاد شد و ریشه‌هایی در سوپرکامپیوترها دارد. ایجاد آن توسط برنامه پیشرفته شبیه‌سازی و محاسبات (ASC) که شامل مراکز سوپرکامپیوتر مانند آزمایشگاه ملی لاس آلاموس (LANL)، آزمایشگاه‌های ملی سندیا (SNL)، و آزمایشگاه ملی لارنس لیورمور (LLNL) است، حمایت شد.

آغاز کار بر روی Ceph در یک برنامه تابستانی در LLNL شروع شد. Sage پس از اتمام تحصیلاتش، به‌طور تمام وقت به توسعه Ceph ادامه داد و شرکتی به نام Inktank برای پیشبرد توسعه آن تأسیس کرد.

Inktank در نهایت توسط Red Hat خریداری شد. Ceph همچنان یک پروژه متن‌باز قدرتمند است و توسط شرکت‌های بزرگی از جمله اعضای بنیاد Ceph مانند Canonical، Red Hat، Intel و... نگهداری می‌شود.

Ceph به منظور جایگزینی Lustre در زمینه سوپرکامپیوترها ایجاد شد و از طریق تلاش‌های توسعه قابل توجه ویژگی‌هایی مانند CephFS را اضافه کرد که قابلیت سازگاری با POSIX را فراهم می‌آورد و آن را به یک سیستم ذخیره‌سازی شبکه‌ای فایل‌محور قدرتمند تبدیل می‌کند.

پایه‌های Ceph واقعاً بر اساس تحمل خطا به جای عملکرد بنا شده است و مدل ذخیره‌سازی آن مبتنی بر تکرار، بارهای عملکردی قابل توجهی دارد. بنابراین، هنوز به سطح سایر راه‌حل‌ها از نظر ارائه عملکرد نزدیک به سخت‌افزار زیرین نرسیده است. اما Ceph در مقیاس بزرگ به‌خوبی مقیاس‌پذیر است و می‌تواند مقدار زیادی از عملکرد کلی خوشه Ceph را ارائه دهد.

Lustre

Lustre یک سیستم فایل توزیع‌شده موازی است که برای محاسبات خوشه‌ای در مقیاس بزرگ استفاده می‌شود.

واژه Lustre ترکیبی از کلمات Linux و Cluster است. Lustre به‌طور مداوم در IO500، (یک معیار دوسالانه که عملکرد راه‌حل‌های ذخیره‌سازی را در رابطه با موارد استفاده در محاسبات با عملکرد بالا مقایسه می‌کند)، رتبه بالایی داشته و استفاده قابل توجهی در لیست TOP500، (یک انتشار معیار دوسالانه متمرکز بر عملکرد کلی خوشه)، داشته است.

Lustre در ابتدا به‌عنوان یک پروژه تحقیقاتی توسط Peter J. Braam، که در دانشگاه کارنگی ملون کار می‌کرد، ایجاد شد و سپس شرکت خود را (Cluster File Systems) برای کار بر روی Lustre تأسیس کرد. مشابه Ceph، Lustre تحت برنامه پیشرفته شبیه‌سازی و محاسبات (ASC) و پروژه PathForward که بودجه آن از طریق وزارت انرژی ایالات متحده (DoE)، Hewlett-Packard و Intel تأمین شد، توسعه

یافت. Sun Microsystems در نهایت Cluster File Systems را خریداری کرد که پس از مدت کوتاهی توسط Oracle خریداری شد.

Oracle and Lustre

پس از خرید Sun، Oracle به‌سرعت اعلام کرد که توسعه Lustre را متوقف خواهد کرد. بسیاری از توسعه‌دهندگان اصلی Lustre تا آن زمان از Oracle خارج شده بودند و علاقه‌مند به ادامه نگهداری و توسعه Lustre تحت یک مدل جامعه‌ی باز بودند.

به همین منظور، سازمان‌های مختلفی از جمله Open Scalable File System (OpenSFS) و EUROPEAN Open File Systems (EOFS) تأسیس شدند. برای پیوستن به این تلاش، استارتاپی به نام Whamcloud توسط چندین نفر از توسعه‌دهندگان اصلی تأسیس شد.

OpenSFS بسیاری از فعالیت‌های Whamcloud را تأمین مالی کرد و این امر به‌طور قابل توجهی به توسعه Lustre کمک کرد. در نهایت، Whamcloud توسط Intel خریداری شد و از طریق بازسازی در Intel، بخش توسعه متمرکز بر Lustre به یک شرکت جدید به نام DDN منتقل شد.

BeeGFS

BeeGFS یک سیستم فایل موازی است که برای HPC توسعه یافته است. این فناوری در ابتدا در مرکز تحقیقاتی Fraunhofer برای محاسبات با عملکرد بالا توسط تیمی به رهبری Sven Breuner توسعه یافت. Breuner مدیر عامل ThinkParQ شد، شرکتی که برای نگهداری و تجاری‌سازی پیشنهادات حرفه‌ای مرتبط با BeeGFS تأسیس شد.

BeeGFS توسط چندین نهاد اروپایی که خوشه‌های آنها در لیست TOP500 قرار دارد، استفاده می‌شود.

DAOS

Distributed Asynchronous Object Storage (DAOS) یک راه حل ذخیره‌سازی متن‌باز است که هدف آن بهره برداری از جدیدترین نسل تکنولوژی‌های ذخیره‌سازی، مانند حافظه غیرفرار یا NVM است.

این فناوری از حافظه‌های دائمی Intel Optane و دستگاه‌های ذخیره‌سازی NVMe استفاده می‌کند تا منابع ذخیره‌سازی را به‌عنوان یک راه‌حل توزیع‌شده ارائه دهد. به‌عنوان یک رقیب جدید، DAOS در چالش ۱۰ نود IO500 در ISC HP'22 به‌خوبی عمل کرد و چهار رتبه در میان ۱۰ رتبه برتر را کسب کرد. این پروژه توسط Intel ایجاد شده و به‌طور فعال نگهداری می‌شود.

GPFS

IBM General Parallel File System (که به عنوان IBM Spectrum Scale نیز شناخته می‌شود) یک سیستم فایل خوشه‌ای با عملکرد بالا است که توسط بسیاری از استقرارهای تجاری HPC به‌عنوان یک راه‌حل ذخیره‌سازی تسریع‌شده استفاده می‌شود.

این فناوری همچنین در چندین خوشه سوپرکامپیوتری در لیست TOP 500 یافت می‌شود. GPFS از پروژه تحقیقاتی Tiger Shark در مرکز تحقیقاتی Almaden IBM در سال ۱۹۹۳ آغاز شد و در ابتدا برای برنامه‌های چندرسانه‌ای با تمرکز بر نرخ انتقال طراحی شد. این طراحی متمرکز بر نرخ انتقال به‌طور قابل توجهی برای محاسبات علمی مناسب بود.

VAST Data

VAST Data یک بازیگر نسبتاً جدید در بازار ذخیره‌سازی است که دستگاه‌های ذخیره‌سازی را با استفاده از برخی از جدیدترین تکنولوژی‌ها ارائه می‌دهد.

به‌عنوان مثال، آنها از SSDهای Intel Optane / 3D XPoint NVMe و حافظه غیرفرار مبتنی بر ۳ D XPoint به‌عنوان بخشی از معماری داده‌های خود استفاده می‌کنند.

این حافظه ها به‌عنوان یک لایه تسریع‌شده در مقابل SSD های با چگالی بالاتر و مقرون‌به‌صرفه‌تر مبتنی بر NAND Flash عمل می‌کنند.

VAST Data می‌تواند از طریق NVMe-oF با استفاده از Ethernet یا InfiniBand متصل شود و از RDMA برای NFS نسخه ۳ پشتیبانی می‌کند.

Weka

Weka یک راه‌حل ذخیره‌سازی خوشه‌ای مبتنی بر دستگاه است که به‌عنوان "پلتفرم داده برای عصر ابری و AI" توصیف می‌شود.

این سیستم با رعایت استانداردهای POSIX و فراهم کردن دسترسی از طریق S3، NFS، و SMB، به کاربر این امکان را می‌دهد که به داده ها دسترسی پیدا کند. مانند VAST Data، Weka نیز به عنوان یک تازه وارد

در این حوزه، از تکنولوژی‌های نسبتاً جدیدی مانند NVMe over Fabric و NVIDIA GPUDirect Storage پشتیبانی می‌کند.

PanNFS

PanNFS، که توسط Panasas ایجاد شده است، یک سیستم فایل خوشه‌ای است که از پروتکل‌های DirectFlow (pNFS)، NFS و CIFS برای دسترسی به داده‌ها پشتیبانی می‌کند.

Panasas یکی از کمک‌کنندگان کلیدی به Parallel NFS (pNFS) بوده است، که به مشتریان این امکان را می‌دهد که درخواست‌های فایل را به‌طور همزمان به چندین سرور یا دستگاه ذخیره‌سازی ارسال کنند، به جای اینکه یکی یکی آن‌ها را پردازش کنند. این ویژگی به استاندارد NFS 4.1 افزوده شده است.

برنامه‌ریزی، بارهای کاری و قابلیت جابجایی بارهای کاری

تا کنون، ما به اجزای اساسی خوشه‌های HPC از جمله سخت‌افزار، راه‌حل‌های ذخیره‌سازی، تأمین منابع و سیستم‌عامل پرداخته‌ایم.

اما خوشه‌های HPC تنها به این اصول پایه‌ای وابسته نیستند؛ نرم‌افزار نیز برای عملکرد خوشه‌ها بسیار حیاتی است. زمان‌بندی‌کننده‌ها برای بهینه‌سازی استفاده از خوشه و اطمینان از اجرای بارهای کاری استفاده می‌شوند. کتابخانه‌های MPI توسط بارهای کاری برای فعال کردن ارتباطات موازی استفاده می‌شوند، که باعث می شود که بارهای کاری بتوانند در سطح خوشه‌ها اجرا شوند به جای اینکه روی یک ماشین واحد اجرا شوند.

بارهای کاری خود نرم‌افزارهایی هستند که پایه‌گذار محاسبات هستند. و در نهایت، قابلیت حمل بارهای کاری در حال تبدیل شدن به یک موضوع مهم‌تر از همیشه است، به همین دلیل است که استفاده از کانتینرها در HPC در حال افزایش است.

زمان‌بندی‌کننده‌ها

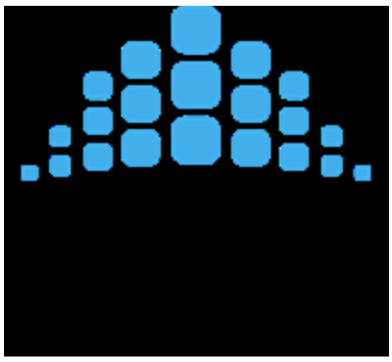
در HPC، زمان‌بندی‌کننده یا (Scheduler) بارهای کاری را در برابر منابع خوشه صف‌بندی می‌کند تا استفاده از آن‌ها را هماهنگ کند.

زمان‌بندی‌کننده‌ها به‌عنوان مغز خوشه‌ها عمل می‌کنند. آن‌ها درخواست‌های بارهای کاری که نیاز به زمان‌بندی دارند را از کاربران خوشه دریافت می‌کنند، آن‌ها را پیگیری می‌کنند و سپس این بارهای کاری را زمانی که منابع در دسترس هستند، اجرا می‌کنند. زمان‌بندی‌کننده ها از هرگونه در دسترس بودن و استفاده از منابع آگاه هستند و تلاش می‌کنند تا هر عاملی که ممکن است بر عملکرد تأثیر بگذارد را در نظر بگیرند.

هدف اصلی آن‌ها زمان‌بندی شغل‌های محاسباتی بر اساس توزیع بهینه بارهای کاری است. زمان‌بندی معمولاً بر اساس نیازهای سازمانی انجام می‌شود. زمان‌بندی‌کننده بارهای کاری را پیگیری کرده و آن‌ها را به یک فرآیند کاربردی دیگر ارسال می‌کند که در گره های محاسباتی برای اجرای آن بار کاری عمل می‌کند.

راه‌حل‌های مدیریت بار کاری در HPC

مدیر بار SLURM



مرور کلی

SLURM که ابتدا به عنوان Simple Linux Utility for Resource Management شناخته می‌شد، یک برنامه‌ریز (job scheduler) متن‌باز و محبوب برای مدیریت و زمان‌بندی بارهای محاسباتی است.

تاریخچه

توسعه یافته به طور مشترک توسط آزمایشگاه ملی لارنس لیورمور، SchedMD، HP و Bull. SchedMD به عنوان نگهدارنده اصلی SLURM و ارائه‌دهنده پشتیبانی تجاری آن عمل می‌کند.

استفاده

SLURM روی حدود ۶۰٪ از خوشه‌های TOP500 نصب شده و به عنوان رایج‌ترین برنامه برای خوشه‌های بزرگ شناخته می‌شود.

نصب

از مخازن Universe در اوبونتو قابل نصب است.

Open OnDemand

مرور کلی

در حالی که Open OnDemand خود برنامه‌ریز نیست، به طور قابل توجهی به SLURM کمک می‌کند. این ابزار یک رابط کاربری تحت وب است که استقرار و مدیریت بارهای کاری را از طریق یک پورتال وب آسان می‌کند.

منشاء

ایجاد شده توسط مرکز ابررایانه اوهایو با بودجه از بنیاد ملی علوم.

Grid Engine

مرور کلی

Grid Engine یک برنامه‌ریز دسته‌ای با تاریخچه پیچیده‌ای است که شامل نسخه‌های متن‌باز و غیرمتن‌باز می‌شود.

تاریخچه

انتشار اولیه

به عنوان یک برنامه غیرمتن‌باز توسط Gridware آغاز شد.

خرید توسط Sun: به Sun Grid Engine (SGE) تبدیل شد و سپس به صورت متن‌باز منتشر شد.

خرید توسط Oracle: بعد از خرید توسط Oracle، به صورت غیرمتن‌باز درآمد و به Oracle Grid Engine تغییر نام یافت.

فُرک‌ها

Son of Grid Engine

که به طور اولیه توسط دانشگاه لیورپول نگهداری می‌شد ولی اکنون به طور فعال نگهداری نمی‌شود.

Grid Community Toolkit

یک فورک متن‌باز ولی به طور فعال نگهداری نمی‌شود.

Univa Grid Engine

یک فورک غیرمتن‌باز از Univa که اکنون متعلق به Altair است و به طور فعال نگهداری می‌شود.

دسترس‌پذیری

نسخه Grid Community Toolkit از مخازن Universe در اوبونتو قابل دریافت است.

OpenPBS

مرور کلی

PBS (Portable Batch System) که ابتدا برای ناسا توسعه یافته بود، در سال ۱۹۹۸ به صورت متن‌باز منتشر شد.

وضعیت فعلی

OpenPBS:

نسخه متن‌باز که به وسیله Altair نگهداری می‌شود.

TORQUE

یک فورک از PBS که ابتدا متن‌باز بود ولی اکنون به صورت غیرمتن‌باز نگهداری می‌شود.

HTCondor

HTCondor یک برنامه زمان‌بندی (Scheduler) مستقل است که به‌طور خاص برای استفاده از منابع غیرقابل استفاده در ایستگاه‌های کاری (workstations) نوشته شده است، به جای استفاده از خوشه‌های HPC.

این سیستم قادر است بارهای کاری را بر روی سیستم‌های بیکار اجرا کند و هنگامی که فعالیتی در سیستم شناسایی می‌شود، آن‌ها را متوقف کند. HTCondor در مخزن بسته‌های Universe در Ubuntu در دسترس است.

Kubernetes

مرور کلی

Kubernetes یک برنامه‌ریز کانتینر است که برای برنامه های بومی ابری محبوبیت زیادی پیدا کرده است. علاقه به گسترش استفاده از Kubernetes در بارهای کاری محاسباتی که به موازی‌سازی وابسته‌اند، افزایش یافته است.

استفاده در یادگیری ماشین:

برخی از بارهای کاری یادگیری ماشین حول Kubernetes اکوسیستم گسترده‌ای ساخته‌اند و این گاهی اوقات نیاز به استقرار Kubernetes به عنوان بار کاری موقت بر روی یک زیرمجموعه از منابع برای مدیریت بارهایی که به شدت به آن وابسته‌اند را ایجاد کرده است.

توسعه

تلاش‌هایی برای گسترش قابلیت‌های برنامه‌ریزی Kubernetes به منظور بهتر خدمت‌رسانی به نیازهای بارهای کاری محاسباتی در حال انجام است.

کتابخانه‌های MPI و کتابخانه‌های محاسبات موازی

استفاده از MPI

در حالی که می‌توان بارهای کاری HPC را بر روی یک سرور یا نود واحد اجرا کرد، پتانسیل واقعی محاسبات با عملکرد بالا از اجرای وظایف محاسباتی فشرده به عنوان فرایندهایی در چندین نود به دست می‌آید. این فرایندهای مختلف به صورت موازی به عنوان یک برنامه واحد با هم کار می‌کنند. برای اطمینان از ارتباط بین فرایندها در نودهای مختلف، به مکانیزم انتقال پیام نیاز دارید که در HPC با عنوان MPI (Messa Passing Interface) شناخته می‌شود.

MPI چیست؟

MPI یک پروتکل ارتباطی و استاندارد است که برای انتقال پیام از حافظه یک سیستم به سیستم دیگر در رایانه‌های موازی طراحی شده است. انتقال پیام به بارهای کاری محاسباتی این امکان را می‌دهد که در نودهای محاسباتی متصل به یک لینک شبکه با سرعت بالا اجرا شوند.

MPI به توسعه HPC کمک کرد تا سازمان‌ها بتوانند مشکلات محاسباتی خود را با هزینه کمتر و مقیاس بیشتری حل کنند و دیگر محدود به توان محاسباتی یک سیستم واحد نباشند.

کتابخانه‌های MPI:

کتابخانه‌های MPI انتزاعاتی را برای ارتباط نقطه به نقطه و جمعی بین فرایندها فراهم می‌آورند. این کتابخانه‌ها برای اکثر زبان‌های برنامه نویسی در دسترس هستند و توسط اکثر بارهای کاری موازی برای رسیدن به مقیاس بی‌نظیر در خوشه‌های بزرگ استفاده می‌شوند.

راه‌حل‌های MPI

OpenMP

مرور کلی:

OpenMP یک رابط برنامه‌نویسی کاربردی (API) و کتابخانه برای برنامه‌نویسی موازی است که از پردازش چند رشته‌ای با حافظه مشترک پشتیبانی می‌کند.

هنگامی که با OpenMP برنامه‌نویسی می‌شود، تمام نخ ها حافظه و داده ها را به اشتراک می‌گذارند.

قابلیت حمل

OpenMP بسیار قابل حمل است و به برنامه‌نویسان یک رابط ساده برای توسعه برنامه‌های موازی که می‌توانند بر روی هر چیزی از دسکتاپ های چند هسته‌ای تا بزرگترین ابررایانه ها اجرا شوند، می‌دهد.

کاربرد

OpenMP به فرآیند ها امکان برقراری ارتباط با یکدیگر در داخل یک نود در یک خوشه HPC را می‌دهد، اما برای پردازش بین نودها به کتابخانه و API اضافی نیاز است که MPI این نقش را ایفا می‌کند.

دسترس‌پذیری

OpenMP از طریق اکثر کامپایلرها، مانند GCC، در اوبونتو در دسترس است.

OpenMPI

مرور کلی

OpenMPI یک پیاده‌سازی متن‌باز از استاندارد MPI است که توسط یک کنسرسیوم از شرکای آکادمیک، تحقیقاتی و صنعتی توسعه و نگهداری می‌شود. این پروژه از ادغام سه پیاده‌سازی شناخته شده MPI که دیگر به طور جداگانه نگهداری نمی‌شوند، ایجاد شده است: FT-MPI از دانشگاه تنسی، LA-MPI از آزمایشگاه ملی لاس آلاموس و LAM/MPI از دانشگاه ایندیانا.

ویژگی‌ها

هدف این پروژه ترکیب بهترین ایده‌ها و تکنولوژی‌ها از هر یک از این پیاده‌سازی‌ها به یک پیاده‌سازی متن‌باز جدید با کد پایه کاملاً جدید و برتر است. OpenMPI در اوبونتو در مخزن بسته‌های Universe در دسترس است.

MPICH

مرور کلی

MPICH که قبلاً با نام MPICH2 شناخته می‌شد، یک پیاده‌سازی متن‌باز و آزاد از MPI است. این پروژه توسط آزمایشگاه ملی آرگون و دانشگاه ایالتی میسیسیپی آغاز شد. نام MPICH از ترکیب MPI و CH گرفته

شده است که CH به معنای Chameleon، یک کتابخانه برنامه‌نویسی موازی که توسط یکی از بنیان‌گذاران MPICH توسعه داده شده است، می‌باشد.

ویژگی‌ها

MPICH یکی از محبوب‌ترین پیاده‌سازی‌های MPI است و به عنوان پایه بسیاری از کتابخانه‌های MPI موجود امروز، از جمله Intel MPI، IBM MPI، Cray MPI، Microsoft MPI و پروژه متن‌باز MVAPICH استفاده می‌شود. MPICH در اوبونتو در مخزن بسته‌های Universe در دسترس است.

MVAPICH

مرور کلی

MVAPICH که بر اساس MPICH توسعه یافته است، آزاد و متن‌باز است و توسط دانشگاه ایالتی اوهایو رهبری می‌شود. اهداف آن ارائه "بهترین عملکرد، مقیاس‌پذیری و تحمل خطا برای سیستم‌های محاسباتی و سرورهای پیشرفته" است که از اتصالات با عملکرد بالا استفاده می‌کنند.

ویژگی‌ها

توسعه آن بسیار فعال است و نسخه‌های مختلفی که بهترین سازگاری با سخت‌افزار و عملکرد ممکن برای زیرساخت‌های موجود را ارائه می‌دهند، در دسترس هستند.

یکی از توسعه‌های قابل توجه آن، پشتیبانی از DPU offloading است، جایی که MVAPICH از SmartNICs (شبکه‌های هوشمند) برای بارگذاری فرآیندهای MPI استفاده می‌کند. SmartNICs و Data Processing Units (DPU) یک فرم پیشرفته از کارت‌های شبکه هستند که دارای اجزای سنتی کامپیوتر، مانند CPU، هستند و به آنها اجازه می‌دهند به عنوان یک کامپیوتر عمل کنند و حتی داده‌ها یا ترافیک شبکه‌ای که از طریق آنها عبور می‌کند را پردازش کنند. این اجازه می‌دهد که برخی از وظایف بارگذاری

میزبان را پردازش کنند، برای مثال، مدیریت ارتباطات MPI که به پردازنده‌های میزبان اجازه می‌دهد به طور کامل بر روی بار کاری تمرکز کنند.

بارهای کاری

منابع

بسیاری از بارهای کاری HPC از توسعه داخلی یا متن‌باز نشأت می‌گیرند که به طور عمده از تلاش‌های جامعه قوی پشتیبانی می‌شود. این بارهای کاری اغلب از پس‌زمینه تحقیقاتی قوی و ابتکار از کارهای دانشگاهی یا منافع ملی ناشی می‌شوند و معمولاً به چندین موسسه یا کشور خدمت می‌کنند.

مثال‌ها

در دنیای متن‌باز، بارهای کاری متعددی برای سناریوهای مختلف وجود دارد، از تحقیقات آب و هوا گرفته تا فیزیک.

راه حل های بارهای کاری

BLAST

مرور کلی

BLAST (Basic Local Alignment Search Tool) یک الگوریتم در بیوانفورماتیک است که برای مقایسه اطلاعات توالی‌های زیستی، مانند توالی‌های پروتئین یا نوکلئوتیدهای DNA یا RNA استفاده می‌شود. این ابزار به پژوهشگران این امکان را می‌دهد که یک توالی را با کتابخانه یا پایگاه داده‌ای از توالی‌های شناخته شده مقایسه کنند، که شناسایی را آسان‌تر می‌کند.

کاربردها

BLAST می‌تواند برای مقایسه توالی‌های یافت شده در حیوانات با آن‌هایی که در ژنوم انسان یافت می‌شود، استفاده شود، که به دانشمندان کمک می‌کند ارتباطات بین آن‌ها و نحوه بیان آن‌ها را شناسایی کنند.

OpenFOAM

مرور کلی

OpenFOAM (Open-source Field Operation And Manipulation) یک جعبه‌ابزار متن‌باز است که برای توسعه حل‌کننده های عددی در دینامیک سیالات محاسباتی استفاده می‌شود.

OpenFOAM در ابتدا به صورت تجاری به نام FOAM فروخته می‌شد، اما تحت مجوز GPL متن‌باز شد و به نام OpenFOAM تغییر نام یافت.

توسعه

در سال ۲۰۱۸، کمیته‌ای برای تعیین مسیر پروژه OpenFOAM تشکیل شد که بسیاری از اعضای آن از بخش خودروسازی هستند. OpenFOAM در مخازن بسته‌های اوبونتو در دسترس است.

ParaView

مرور کلی

ParaView یک پلتفرم متن‌باز برای تحلیل و بصری‌سازی داده‌ها است که با معماری سرور-کلاینت نوشته شده است. این ابزار اغلب برای مشاهده نتایج برنامه‌هایی مانند OpenFOAM و دیگر برنامه‌ها استفاده می‌شود. برای عملکرد بهینه، نیازهای رندرینگ یا پردازش ParaView می‌تواند به عنوان یک کار زمانبندی شده در خوشه اجرا شود تا از منابع محاسباتی خوشه‌ای استفاده شود.

توسعه

ParaView به صورت یک برنامه واحد نیز قابل اجرا است و نیازی به اجرا به صورت انحصاری بر روی خوشه ها از طریق معماری کلاینت، سرور ندارد.

ParaView از طریق همکاری بین KitWare Inc و آزمایشگاه‌های ملی لاس آلاموس با تأمین مالی از وزارت انرژی ایالات متحده آغاز شد. از آن زمان، آزمایشگاه‌های ملی دیگری نیز به تلاش‌های توسعه پیوسته‌اند.
ParaView در مخازن بسته‌های اوبونتو در دسترس است.

WRF

مرور کلی

WRF (Weather Research & Forecasting) مدل پیش‌بینی آب و هوای مقیاس میانه متن‌باز است.
این مدل از محاسبات موازی پشتیبانی می‌کند و توسط جامعه وسیعی برای تحقیقات جوی و پیش‌بینی عملیاتی استفاده می‌شود.

WRF امروزه توسط اکثر نهادهای مربوط به پیش‌بینی آب و هوا استفاده می‌شود.

توسعه

این مدل از طریق همکاری مرکز ملی تحقیقات جوی (NCAR)، سازمان ملی اقیانوسی و جوی (NOAA)، نیروی هوایی ایالات متحده، آزمایشگاه تحقیقاتی نیروی دریایی، دانشگاه اوکلاهما، و اداره هوانوردی فدرال (FAA) توسعه یافته است.

این یک تلاش چندرشته‌ای و چند سازمانی واقعی است و دارای یک جامعه وسیع از حدود ۵۶،۰۰۰ کاربر در بیش از ۱۶۰ کشور است.

شبیه‌ساز دینامیک آتش و مشاهده‌گر دود

شبیه‌ساز دینامیک آتش (FDS) و مشاهده‌گر دود (SMV) ابزارهای متن‌باز هستند که توسط مؤسسه ملی استانداردها و فناوری (NIST) توسعه یافته‌اند.

شبیه‌ساز دینامیک آتش (FDS) مدل دینامیک سیالات محاسباتی (CFD) برای جریان‌های ناشی از آتش است.

این ابزار از محاسبات موازی برای حل عددی فرم خاصی از معادلات ناویر،استوکس استفاده می‌کند. این مدل برای جریان‌های حرارتی با سرعت پایین مناسب است و به ویژه برای شبیه سازی پخش و انتقال دود و حرارت ناشی از آتش استفاده می‌شود.

مشاهده‌گر دود (SMV) بخش بصری‌سازی FDS است و برای تحلیل خروجی‌های FDS مورد استفاده قرار می‌گیرد.

این ابزار به کاربران کمک می‌کند تا بهتر بفهمند و مشاهده کنند که چگونه دود، حرارت و آتش پخش می‌شود. این ابزار به ویژه برای درک تاثیرات بر روی ساختارهای بزرگ و تحلیل سناریوهای فاجعه‌آمیز کاربرد دارد.

کانتینرها

محیط‌های HPC (محاسبات با کارایی بالا) اغلب به وابستگی‌های پیچیده‌ای برای اجرای بارهای کاری وابسته هستند.

تلاش زیادی به توسعه سیستم‌های مبتنی بر ماژول مانند Lmod معطوف شده است که به کاربران اجازه می‌دهد تا برنامه‌ها یا وابستگی‌ها مانند کتابخانه‌ها را خارج از مسیرهای معمول سیستم بارگذاری کنند. این به دلیل نیاز به کامپایل برنامه‌ها در برابر مجموعه‌ای خاص از کتابخانه‌ها است که به نسخه‌های خاص عددی یا فروشنده وابسته‌اند.

برای اجتناب از این وابستگی‌های پیچیده، سازمان‌ها می‌توانند در کانتینرها سرمایه‌گذاری کنند. این روش به طور مؤثر به کاربر اجازه می‌دهد تا یک برنامه را به همراه تمام وابستگی‌های آن در یک کانتینر اجرایی واحد بسته‌بندی کند.

راه‌حل‌های کانتینر

LXD یک سیستم مدیریت کانتینر و ماشین مجازی نسل بعدی است. این ابزار تجربه کاربری یکپارچه‌ای را حول سیستم‌های کامل لینوکس که در داخل کانتینرها یا ماشین‌های مجازی اجرا می‌شوند، ارائه می‌دهد.

برخلاف سایر محیط‌های اجرایی کانتینر، LXD امکان مدیریت ماشین‌های مجازی را نیز فراهم می‌آورد و قابلیت آن برای اجرای محیط‌های چند برنامه‌ای کامل منحصر به فرد است.

با استفاده از LXD، می‌توان به طور مؤثر یک محیط HPC کامل را در داخل یک کانتینر LXD اجرا کرد که این کارایی و جداسازی را بدون هزینه اضافی برای عملکرد فراهم می‌کند.

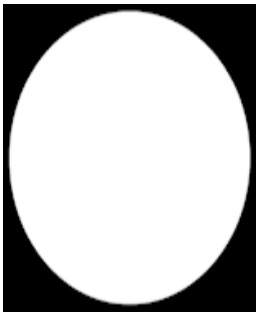
Docker به عنوان محیط اجرایی پیش‌رو برای برنامه‌های بومی ابری شناخته می‌شود و تا حدی در محیط‌های HPC نیز مورد استفاده قرار گرفته است. پذیرش Docker در سیستم‌های چند کاربره واقعی مانند محیط‌های خوشه‌ای بزرگ محدود بوده است، زیرا Docker اساساً به دسترسی‌های ویژه نیاز دارد.

یکی دیگر از معایب ذکر شده معمولاً اندازه کلی تصاویر Docker است که به وابستگی‌های برنامه از جمله کتابخانه‌های MPI نسبت داده می‌شود.

این موضوع معمولاً کانتینرهای برنامه بزرگ را ایجاد می‌کند که ممکن است به راحتی اجزای دیگر کانتینرهای برنامه را تکرار کنند. با این حال، وقتی به درستی انجام شود، Docker می‌تواند برای مدیریت وابستگی‌ها مؤثر باشد.

به‌ویژه در هنگام توسعه و فعال‌سازی یک استک سخت‌افزاری خاص، این امکان را می‌دهد که برنامه‌ها در یک استک یکپارچه بسته‌بندی شوند که برخی نقاط قوت را به همراه دارد.

Singularity



Singularity (یا Apptainer، نام شاخه جدیدتر آن) تلاشی برای ایجاد کانتینرهای اپلیکیشن است که به برخی از معایب احتمالی کانتینرهای Docker پاسخ می‌دهد.

Singularity به دلیل عدم وابستگی به دسترسی‌های ویژه، به خوبی در محیط‌های چند کاربره بزرگ قابل استفاده است. به جای ایجاد کانتینرهای کامل اپلیکیشن با تمام وابستگی‌ها، Singularity می‌تواند توسط مؤلفه‌های سطح سیستم مانند کتابخانه ها و پیاده‌سازی های MPI اجرا شود و کانتینرهای سبک‌تری با اهداف و وابستگی‌های خاص‌تر تولید کند.

Charliecloud

Charliecloud تلاشی برای کانتینر ization است که از نظرهایی شبیه به Singularity است و از Docker برای ساخت تصاویری استفاده می‌کند که سپس می‌توانند توسط محیط اجرایی Charliecloud به صورت غیرمستقیم اجرا شوند. این ابتکار توسط آزمایشگاه ملی لاس آلاموس (LANL) انجام شده است.

خدمات کمکی

بسیاری از مؤلفه‌های نرم‌افزاری می‌توانند برای بهبود استفاده از خوشه‌های HPC به کار روند. این مؤلفه‌ها شامل هر چیزی از مدیریت هویت گرفته تا نرم‌افزارهای نظارت و مشاهده‌پذیری هستند.

مدیریت هویت

مدیران دسترسی هویت در خوشه‌های HPC رایج هستند. آنها به عنوان منبع واحدی برای مدیریت هویت و دسترسی عمل می‌کنند. دسترسی یکپارچه، ورود به هر نود در خوشه را آسان می‌کند. این موضوع اغلب پیش‌نیازی برای زمان‌بندی منابع است.

برای مثال، اگر بخواهید یک شغل موازی را در چندین نود در خوشه از طریق زمان بند دسته‌ای اجرا کنید، نیاز به دسترسی یکسان به نود های محاسباتی و منابع ذخیره‌سازی دارید.

یک راه حل مدیریت هویت می‌تواند به شما کمک کند تا اطمینان حاصل کنید که این سازگاری حفظ می‌شود. بدون آن، به عنوان یک مدیر، باید از ایجاد کاربران، هویت و پیکربندی‌های ذخیره‌سازی اطمینان حاصل کنید که همه آنها در سراسر خوشه به صورت جداگانه پیکربندی شده‌اند.

راه‌حل‌های مدیریت هویت

LDAP (پروتکل دسترسی دایرکتوری سبک) برای دسترسی به خدمات دایرکتوری مانند آنهایی که توسط Active Directory ارائه می‌شود، استفاده می‌شود.

LDAP راهی برای مشتریان فراهم می‌آورد تا اطلاعات دایرکتوری مانند حساب‌های کاربری، کلمات عبور و حقوق دسترسی را در شبکه جستجو و به‌روزرسانی کنند.

LDAP استانداردی برای دسترسی به خدمات دایرکتوری است و توسط پروژه‌های مختلفی از جمله GLAuth، OpenLDAP و FreeIPA استفاده می‌شود.

Active Directory یک راه‌حل خدمات دایرکتوری است که توسط مایکروسافت برای شبکه‌های ویندوز ایجاد شده است.

این سرویس مکان مرکزی برای ذخیره، مدیریت و امنیت اطلاعات کاربران و رایانه ها، از جمله حساب های کاربری، کلمات عبور و حقوق دسترسی را فراهم می‌آورد.

Active Directory همچنین ابزارهایی برای مدیریت منابع شبکه، مانند پوشه‌های به اشتراک گذاشته شده، چاپگرها و برنامه‌ها ارائه می‌دهد.

FreeIPA

FreeIPA یک راه حل متن‌باز برای مدیریت هویت و دسترسی است که توسط Red Hat ایجاد شده است. این سیستم یک راه‌حل یکپارچه برای هویت و احراز هویت در محیط های لینوکس فراهم می‌آورد.

FreeIPA احراز هویت، مجوزدهی و اطلاعات حساب را به‌طور متمرکز ارائه می‌دهد.

FreeIPA بر روی تعدادی از راه‌حل‌های متن‌باز دیگر ساخته شده است، از جمله سرور دایرکتوری ۳۸۹ که یک سرور LDAP را فراهم می‌آورد.

به طور کلی، FreeIPA یک راه‌حل جامع است که بر اساس لیستی گسترده از راه‌حل‌های متن‌باز ساخته شده است.

نظارت و مشاهده‌پذیری

ابزارهای نظارت و مشاهده‌پذیری بینش عمیق‌تری به استفاده از منابع بارهای کاری فراهم می‌آورند و بنابراین کلید حل هر گونه مسائل عملکردی یا شناسایی مشکلات با سلامت کلی خوشه هستند.

معیارهایی که معمولاً در خوشه‌های HPC مورد نظارت قرار می‌گیرند شامل استفاده از CPU و حافظه، پهنای باند شبکه و حافظه، و معیارهای زمان‌بند مانند توان خروجی بارها هستند که میزان کارهای انجام شده در یک

دوره مشخص را اندازه‌گیری می‌کند. زمان‌های انتظار شغل و زمان تکمیل شغل، و همچنین معیارهای استفاده از صف زمان‌بند نیز کلیدی هستند.

نظارت و مشاهده‌پذیری در HPC در گذشته محدود به ابزارهای نظارتی بود که کمترین تأثیر را بر بارهای کاری داشته باشند.

امروزه این موضوع، به دلیل افزایش کارایی سخت‌افزار کمتر اهمیت دارد اما همیشه ایده خوبی است که زمان اجرای بارهای کاری را با و بدون ابزارهای اضافی بررسی کنید تا مبنای زمان‌های اجرای شغل‌های بارهای کاری را تعیین کنید.

این تغییرات باعث شده‌اند که استک نظارتی مدرن برای نظارت بر خوشه‌های HPC بیشتر مرتبط باشد.
راه‌حل‌های مدرن مانند Prometheus و Grafana در این خوشه‌ها به طور فزاینده‌ای مشاهده می‌شوند.

راه‌حل‌های مشاهده‌پذیری

Prometheus یک سیستم نظارت متن‌باز است که برای جمع‌آوری، ذخیره و تحلیل معیارها از برنامه‌ها و خدمات استفاده می‌شود.

Prometheus یک زبان پرس وجو و ذخیره‌سازی معیارها را فراهم می‌آورد و همچنین ویژگی‌هایی برای هشداردهی و سایر ویژگی‌ها دارد.

از Prometheus برای نظارت بر برنامه‌ها و خدمات در مراکز داده، محیط‌های ابری و خوشه‌های Kubernetes استفاده می‌شود.

Grafana یک پلتفرم نظارتی متن‌باز محبوب است که برای تجسم، تحلیل و هشدار بر روی معیارها از سیستم‌ها و خدمات استفاده می‌شود.

Grafana از انواع منابع داده، از جمله Prometheus، پشتیبانی می‌کند و مجموعه‌ای غنی از ویژگی‌ها برای ساخت داشبوردها و هشداردهی بر روی معیارها ارائه می‌دهد.

از Grafana برای نظارت و هشدار در مورد عملکرد سیستم، عملکرد برنامه و سایر معیارها استفاده می‌شود.

Grafana Loki

Loki یک سیستم تجمیع لاگ متن‌باز است که برای استفاده در محیط‌های مبتنی بر ابر طراحی شده است. Loki برای لاگ‌های خدماتی مانند Kubernetes، Prometheus و Grafana طراحی شده و راهی برای ذخیره، پرس‌وجو و تحلیل داده‌های لاگ فراهم می‌آورد.

Loki قابلیت های جستجوی قدرتمندی ارائه می‌دهد و می‌تواند برای رفع اشکال، نظارت بر عملکرد سیستم و موارد دیگر استفاده شود.

Canonical Observability Stack (COS)

Canonical Observability Stack (COS) یک راه‌حل نظارتی یکپارچه است که Prometheus، Grafana و Loki را در یک راه‌حل قابل استقرار ترکیب می‌کند و یک راه حل کامل برای نظارت بر خوشه ها ارائه می‌دهد که یک نمای جامع از معیارها فراهم می‌کند.

کجا خوشه‌های HPC را اجرا کنیم؟

خوشه‌های HPC اکنون تقریباً در هر مکانی قابل استقرار هستند. با توجه به پیشرفت‌های فناورانه، اجرای خوشه‌های HPC در فضای ابری در سال‌های اخیر محبوبیت زیادی پیدا کرده است.

برخی از سازمان‌ها حتی این گزینه‌ها را ترکیب کرده و یک خوشه خصوصی با هزینه بهینه ایجاد کرده و در صورت نیاز از فضای ابری برای گسترش استفاده می‌کنند.

HPC در فضای عمومی ابری

بسیاری از ارائه‌دهندگان فضای ابری عمومی منابع تخصصی در فضای HPC ارائه می‌دهند که برای سازمان‌ها با هر اندازه‌ای قابل دسترسی است.

رایانش ابری امکان ایجاد HPC را برای سازمان‌هایی فراهم کرده که ممکن است نیاز به گسترش یا افزایش مقیاس فراتر از آنچه با خوشه‌های اختصاصی منطقی است، داشته باشند. همچنین، امکان اجرای خوشه‌های آزمایشی کوچک برای کسانی که تازه با HPC آشنا می‌شوند نیز وجود دارد.

خدمات HPC در Amazon Web Services (AWS)

AWS یکی از کلیدی‌ترین بازیگران در ارائه خدمات ابری عمومی برای HPC است. پیاده‌سازی AWS Nitro System برای حذف سربار مجازی‌سازی و دسترسی مستقیم به سخت‌افزار اصلی، از جمله نوآوری‌های مهم AWS بود که باعث کاهش تأخیر و افزایش عملکرد شد.

برای برآورده کردن نیازهای بارهای کاری HPC در زمینه ارتباط بین گره‌ها، AWS Elastic Fabric Adaptor را توسعه داد که به کاهش تأخیر و افزایش عملکرد برای بارهایی که نیاز به ارتباطات با سرعت بالا دارند کمک کرد.

Amazon Web Services (AWS)

برای پوشش نیازهای ذخیره‌سازی کاربران HPC، Amazon یک راه‌حل ذخیره‌سازی تخصصی مبتنی بر Lustre به نام Amazon FSx for Lustre اضافه کرده است. علاوه بر آن، آن‌ها راه‌حل‌های زمان‌بندی مانند AWS ParallelCluster و AWS Batch را نیز ارائه می‌دهند.

Microsoft Azure

Azure یکی دیگر از بازیگران اصلی در ارائه HPC در فضای ابری عمومی است و انواع قوی از ماشین‌های مجازی را ارائه می‌دهد که از فناوری‌های سنتی HPC مانند Infiniband استفاده می‌کنند.

Infiniband عملکرد RDMA را برای بهینه سازی تأخیر و عملکرد فراهم می‌آورد. آنها همچنین انواع ماشین‌های مجازی‌ای را ارائه می‌دهند که برای کاربرانی که به کاهش تعداد هسته‌های نمایان شده به بار کاری نیاز دارند، طراحی شده‌اند و بارهایی را که عمدتاً محدود به پهنای باند حافظه هستند، پشتیبانی می‌کنند.

آن‌ها حتی راه‌حلی را ارائه می‌دهند که ابررایانه ها را به عنوان یک سرویس ارائه می‌دهد، که راه‌حل Cray نامیده می‌شود.

علاوه بر این، آن‌ها ذخیره‌سازی متمرکز بر HPC با Cray ClusterStor را نیز ارائه می‌دهند.

Google Cloud Platform (GCP)

Google Cloud Platform ماشین‌های مجازی HPC پیش‌پیکربندی‌شده‌ای را ارائه می‌دهد. پیشنهادات آن‌ها همچنین شامل اتوماسیون و اسکریپت‌نویسی است که تولید اسکریپت‌های مبتنی بر Terraform را که تأمین محیط HPC مبتنی بر Google Cloud را مدیریت می‌کند، آسان می‌کند.

Google راه‌اندازی محیطی که متناسب با نیازهای کاربر باشد را ساده می‌کند. آن‌ها همچنین مستنداتی دارند که به کاربران کمک می‌کند مراحل مشابه آنچه اتوماسیون زیرساخت مبتنی بر Terraform ارائه می‌دهد را دنبال کنند، با راهنماهای روشن برای هر چیزی که به بارهای کاری MPI و HPC تصاویر مربوط می‌شود.

آن‌ها اطلاعات واضح و عملی در مورد چگونگی بهره برداری بهینه از فضای ابری برای بارهای کاری HPC ارائه می‌دهند.

Oracle Cloud Infrastructure

Oracle یکی از پیشگامان در فعال‌سازی HPC در فضای ابری عمومی بود. آن‌ها رویکرد Bare Metal را در HPC در فضای ابری عمومی اتخاذ کرده‌اند و انواع ماشین‌های مجازی با شبکه‌سازی RDMA با تأخیر فوق‌العاده پایین را ارائه می‌دهند.

راه حل حاصل نزدیک به چیزی است که می‌توان از یک خوشه HPC خصوصی اختصاصی انتظار داشت.

خوشه‌های اختصاصی HPC خصوصی

خوشه‌های خصوصی گزینه‌ای مناسب در HPC برای کسانی هستند که به دنبال بهینه‌سازی هزینه، کنترل و حتی مالکیت یا الزامات امنیت داده خاص هستند.

برخی از راه‌حل‌ها قابلیت‌های مدیریت مشابه فضای ابری را برای منابع محلی در محل ارائه می‌دهند. چالش اصلی با خوشه‌های خصوصی HPC سرمایه‌گذاری اولیه بالا و نیاز به تخصص است.

این مشکل می‌تواند با همکاری با شرکایی مانند Canonical که دسترسی به دانش و راه‌حل‌های تخصصی را فراهم می‌کنند، کاهش یابد. بنیادهای چنین خوشه‌هایی به راه‌حل‌های تأمین خوشه مانند MAAS که در بخش تأمین خوشه‌ها در بالا پوشش داده شده، متکی است.

HPC هیبریدی

استفاده ترکیبی از منابع ابری خصوصی و عمومی در فضای HPC بسیار محبوب بوده است. ابرهای هیبریدی بهترین‌های هر دو دنیا را به کاربران ارائه می‌دهند:

بهینه‌سازی هزینه و کنترل ارائه شده توسط سرورهای محلی، همراه با مقیاس‌پذیری فوق‌العاده خوشه‌های ابری عمومی.

به نوعی، ابرهای هیبریدی راه‌حلی تکمیلی ارائه می‌دهند که در آن نقاط ضعف یکی با نقاط قوت دیگری جبران می‌شود. چالش اضافی اصلی که از چنین تنظیمی ناشی می‌شود، ممکن است پیچیدگی افزایش یافته باشد، اما به طور کلی این امکان را دارد که پایداری کلی بیشتری را به ارمغان بیاورد.

با راه‌حل‌هایی برای هر دو فضای ابری عمومی و خصوصی، Canonical می‌تواند به ساده‌سازی پیچیدگی‌های افزایش یافته کمک کند.

HPC در لبه

بسیاری از بارهای کاری HPC، به ویژه آن‌هایی که نیاز به پردازش واقعی‌زمان یا حساسیت بسیار بالا به تأخیر دارند، اکنون در لبه پیاده‌سازی می‌شوند. این به این معنی است که آن‌ها اغلب در خوشه‌های کوچک یا حتی به عنوان یک رایانه بسیار متمرکز که اغلب به آن کامپیوتر با عملکرد بالا (HPC) گفته می‌شود، مستقر می‌شوند.

گام‌های بعدی خود را در HPC با Canonical بردارید

Canonical می‌تواند به شما کمک کند تا گام‌های بعدی خود را در سفر HPC خود بردارید. راه‌حل‌های ما می‌توانند به شما در برآورده کردن نیازهای HPC از لایه سیستم‌عامل تا اتوماسیون زیرساخت و بیشتر، در سراسر ابرها و در محل کمک کنند.

Ubuntu بهترین توزیع لینوکس برای محاسبات با عملکرد بالا است. برخی از مزایای Ubuntu شامل:

هسته اخیر

مخازن پکیج گسترده

چرخه انتشار ثابت ۲ ساله برای نسخه‌های LTS با پشتیبانی ۵ ساله

نگهداری و رفع اشکال قابل تمدید تا ۱۰ سال با Ubuntu Pro

Ubuntu برای محیط‌های در حال اجرا طولانی‌مدت ایده‌آل است و با Ubuntu Pro، می‌توانید اطمینان حاصل کنید که محیط شما در طول عمر آن پشتیبانی می‌شود.

برای پیاده‌سازی‌های در محل، می‌توانید فرآیند تأمین سرور خود را با استفاده از MAAS بهبود دهید، که توسط سازمان‌های زیادی که به خوشه‌های HPC در محل متکی هستند، مورد اعتماد است.

معماری بسیار در دسترس آن باعث می‌شود MAAS مقاوم به خطا باشد و اطمینان حاصل کند که می‌تواند به مقیاس بزرگ پیاده‌سازی شود.

بدون توجه به اندازه خوشه شما، می‌توانید به MAAS اعتماد کنید تا قابلیت‌های تأمین و تجربه ابری نهایی در مدیریت خوشه‌های برهنه را ارائه دهد، که عملکرد و انعطاف‌پذیری نهایی را فراهم می‌کند.

Juju، راه‌حل ما برای اتوماسیون زیرساخت، می‌تواند به شما کمک کند تا یک خوشه مبتنی بر SLURMرا راه‌اندازی کنید و آماده استفاده برای کاربران باشید و با تشکر از Juju، عملیات روز دوم شما نیز مدیریت می‌شود.

Juju می‌تواند در انتهای ابر عمومی استفاده شود و همچنین می‌تواند برای پیاده‌سازی‌های در محل با MAAS نیز استفاده شود. برای بهره‌برداری از پیاده‌سازی‌های ابری بومی، در نظر داشته باشید که از Charmed Kubernetes و Canonical استفاده کنید.

برای پیاده‌سازی‌های پیچیده‌تر زیرساخت، ما Charmed OpenStack را داریم که به شما امکان می‌دهد تا ابری با بهینه‌سازی هزینه‌های خود داشته باشید، که بهترین ارزش را برای عملکرد با حاکمیت کامل ارائه می‌دهد.

Charmed Ceph، راه‌حل مناسبی برای کسانی که به دنبال راه‌حل ذخیره‌سازی مقاوم به خطا هستند، به شما کمک می‌کند تا بر روی نرم‌افزار متن‌باز بسازید و ذخیره‌سازی مشترک مبتنی بر فایل یا ذخیره‌سازی شیء را ارائه دهید.

می‌توانید از ترکیبی از راه‌حل‌ها برای استراتژی مناسب ابر هیبریدی استفاده کنید و بارهای کاری خود را بسته به نیازهای خود اجرا کنید.