

دانشگاه صنعتی اصفهان

مرکز ابررایانش

گزارش کارآموزی



مرکز ابر رایانش ملی شیخ بهایی

National High Performance Computing Center



نویسنده: مجتبی ملائی

استادکارآموزی: دکتر فاطمه دلدار
سرپرست کارآموزی: آقای محمد صالح محمدی

۲۹ شهریور ۱۴۰۴

چکیده

من و همکارانم در این دوره کارآموزی ضمن آشنایی با مرکز ابررایانش و افراد متخصص شاغل در این مرکز با استفاده از ابزار های kubernetes, kubervirt و ceph بستری ایجاد کردیم تا بتوانیم ماشین های مجازی ویندوز را با هدف در دسترس بودن کامل^۱ و با قابلیت هایی مانند تخصیص منابع به صورت برخط و امکان خودکار سازی^۲ به صورت ایمن در اختیار مصرف کنندگان قرار دهیم.

فهرست مطالب

۱	معرفی	۲
۱.۱	پیش زمینه من	۲
۲.۱	مرکز ابررایانش شیخ بهائی	۲
۱.۲.۱	منظور از یک خوشه HPC چیست؟	۲
۲.۲.۱	معماری	۳
۳.۲.۱	نحوه کار	۳
۳.۱	پروژه	۴
۲	آماده سازی	۵
۱.۲	آماده سازی کامپیوترها	۵
۲.۲	آماده سازی سرورها	۵
۱.۲.۲	Proxmox	۵
۳	ارکستراسیون	۷
۱.۳	Kubernetes	۷
۴	پیاده سازی شبکه	۹
۱.۴	calico	۹
۲.۴	multus	۹
۳.۴	macvtap	۹
۴.۴	bridge	۹
۵.۴	kubemacpool	۹
۵	بستر مجازی سازی در k8s	۱۰
۱.۵	Kubevirt	۱۰
۲.۵	منابع ماشین ها	۱۰
۳.۵	انتقال زنده	۱۰
۴.۵	تخصیص منابع به صورت زنده	۱۰
۶	Windows image	۱۱
۱.۶	گرفتن اطلاعات	۱۱
۲.۶	تنظیمات اضافی	۱۱
۷	kubevirt-manager	۱۲
۸	نظارت	۱۳
۱.۸	Prometheus	۱۳
۲.۸	Graphana	۱۳
آ	اطلاعات اضافی	۱۴

فصل ۱

معرفی

۱.۱ پیش زمینه من

من کارآموزی‌ام را در بعد از ترم ششم آغاز کردم. در این دوران با برخی از گرایش‌ها، زمینه‌ها و مسیرهای مختلف در حوزه کامپیوتر آشنا شدم که از جمله آنها برنامه نویسی سمت سرور^۱، امنیت سایبری، شبکه و هوش مصنوعی بود. در این میان به فناوری‌های سمت سرور مانند Linux علاقه‌مند شدم و این موضوع باعث شد تا تصمیم بگیرم که کارآموزی‌ام را در حوزه‌های مربوط به سرورها بگذرانم.

۲.۱ مرکز ابررایانش شیخ بهائی

شبکه ملی محاسبات ابری در سال ۱۳۸۸ با هدف ارائه خدمات کامل و جامع محاسباتی به صنایع و مراکز علمی و تحقیقاتی کشور و با حمایت معاونت علمی و فناوری ریاست جمهوری افتتاح شد. مرکز ملی محاسبات ابری^۲ (NHPCC) به عنوان یکی از چندین مرکز ابررایانه کشور، در آن زمان بخشی از این طرح ملی را در قالب تأسیس مرکز ملی ابررایانه شیخ بهائی اجرا کرد. تمهیدات در نظر گرفته شده در مرکز ملی محاسبات ابری شیخ بهائی به گونه‌ای است که امکانات سخت‌افزاری و نرم‌افزاری جامع و کامل و همچنین خدمات آموزشی و مشاوره‌ای فنی و تخصصی برای استفاده بهینه (در محل یا از راه دور) در دسترس باشد.

۱.۲.۱ منظور از یک خوشه HPC چیست؟

محاسبات با کارایی بالا یا HPC^۳ هنر داشتن مجموعه‌ای از گره‌های پردازشی قدرتمند است که به عنوان یک خوشه (یا گروه‌هایی از آنها) به هم متصل شده‌اند تا مجموعه داده‌های عظیم یا مسائل ریاضی پیچیده چند بعدی را به صورت موازی و در سریع‌ترین زمان ممکن حل کنند. بنابراین HPC در مورد محاسبات است. همچنین ممکن است اصطلاح ابررایانه را بشنوید که مترادف خوشه HPC (یا گروه‌هایی از آنها) است. این نام بیشتر به دیدگاه سخت‌افزاری و قدرت آن اشاره دارد که معمولاً به عنوان یک واحد ممیز شناور به نام فلاپ (مانند ترافلاپس، پتافلاپس و ...) محاسبه می‌شود.

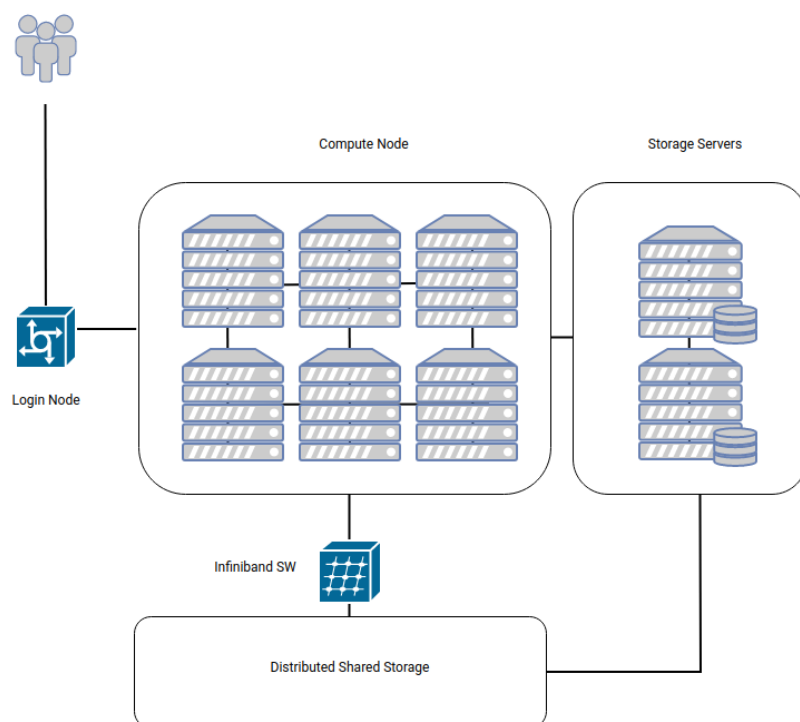
به عنوان مثال، می‌توانید دانش‌آموزی را در نظر بگیرید که یک کد بزرگ (نوشته شده در بیش از هزاران خط) دارد و اگر تمایل داشته باشد کد خود را در لپ‌تاپ شخصی خود اجرا کند، چند سال طول می‌کشد تا پردازش تمام شود و نتیجه نهایی را ببیند؟ بنابراین اگر کد خود را به ابررایانه منتقل کنند، در مدت زمان بسیار کوتاهی به نتایج دست می‌یابد.

^۱back-end

^۲National High Performance Computing Center

^۳High Performance Computing

۲.۲.۱ معماری



شکل ۱.۱: اجزای یک خوشه

معمولاً یک خوشه پایه از این اجزای اصلی تشکیل شده است:

- یک گره ورود (گره سر)
- انواع گره‌های محاسباتی
- یک فضای ذخیره‌سازی مشترک توزیع‌شده
- سرورهای GPU (برای بهره‌مندی از محاسبات گرافیکی عظیم)
- یک سوئیچ Infiniband برای اتصال گره‌ها
- یک زمان‌بند کار و یک سیستم صف‌بندی
- یک گره انتقال
- یک گره پس‌پردازش

۳.۲.۱ نحوه کار

کاربران به گره ورود (گره سر) وارد می‌شوند و فایل‌های ورودی خود را آپلود می‌کنند. سپس کارهای خود را به زمان‌بند ارسال می‌کنند و در صف انتظار قرار می‌گیرند. سپس زمان‌بند شروع به ارسال کارهای آنها به تمام گره‌های محاسباتی دیگری که کاربر مجاز به دسترسی به آنها است، می‌کند.

۳.۱ پروژه

در برخی از موارد مرکز نیاز دارد تا ماشین‌های مجازی ویندوزی را در اختیار مشتریان خود بگذارد. دلیل استفاده از ویندوز راحتی و سادگی بیشتر برای دانشجویان و مشتریانی است که رشته و تخصص آن‌ها در حوزه کامپیوتر نیست و نیاز دارند تا از قدرت محاسباتی مرکز استفاده کنند. برای این افراد کار با سیستم عامل Linux که به طور معمول در خوشه‌ها استفاده می‌شود بسیار دشوار است.

نیاز دیگر این بود که اختصاص ماشین‌ها به مشتریان به صورت خودکار و با کمترین مداخله ممکن از طرف یک نیروی انسانی صورت بگیرد. به همین دلیل استفاده از ابزارهایی مانند Proxmox که در این مرکز استفاده می‌شود ممکن نبود. از طرفی در دسترس بودن و سلامتی این ماشین‌ها مهم‌ترین رکن آن‌ها است. در واقع این پردازش‌ها به دلیل زمان‌بر بودن آن‌ها از اهمیت خاصی برخوردار هستند. به عنوان مثال در صورتی که سرور میزبان دچار خطا شود، ماشین مجازی ویندوز نیز از کار می‌افتد و تمام منابع و زمانی که برای پردازش استفاده شده بود به هدر می‌رود. از دید یک مصرف‌کننده این اتفاق کاملاً غیر قابل تحمل است زیرا که ممکن است مصرف‌کننده زمان لازم برای این اجرای دوباره این پردازش را نداشته باشد. هدف این پروژه برطرف سازی این مشکلات و نیازها بود. در کنار این‌ها قابلیت‌هایی مانند تخصیص منابع به صورت زنده، اضافه کردن و حذف دیسک‌ها به صورت زنده و برقرار سازی سطح مناسبی از امنیت از اهداف مهم دیگر این پروژه بودند.

فصل ۲

آماده سازی

۱.۲ آماده سازی کامپیوترها

در اولین ورود به مرکز، نیاز بود تا کامپیوتر های مورد نیاز خودمان را شخصی سازی کنیم. در واقع میبایست سیستم عامل دیگری بر روی کامپیوترها نصب می کردیم تا اطلاعات قبلی آنها پاک شود و همچنین بتوانیم ابزار های مورد نیاز را روی آنها نصب کنیم. کامپیوترهای مرکز دارای BIOS هستند و همچنین از کارت های گرافیکی قدیمی Nvidia استفاده می کنند. به همین دلیل هنگامی که میخواستیم سیستم عامل های جدید را بر روی آنها نصب کنیم دچار خطا می شدند. بنابراین نهایتا مجبور به استفاده از ISO های قدیمی شدیم. پس از نصب سیستم عامل Manjaro بر روی سیستم ها، باید نرم افزار ها و ابزار های لازم مانند مرورگر، ابزارهای مدیریت بسته، درایورها، کامپایلرها، ابزارهای نصب، ویراشگرهای کد و مدیریت را نصب می کردیم.

۲.۲ آماده سازی سرورها

پس از آماده سازی کامپیوترها، یک سرور قدرتمند به نام "رخش" در اختیار ما قرار گرفت. این سرور شامل ۶۴ هسته پردازشی و ۱ ترابایت حافظه RAM می شد. از ما خواسته شد تا با استفاده از ابزاری به نام Proxmox که یک ابزار برای مدیریت مجازی سازی سطح یک است، چند ماشین مجازی لینوکس ایجاد کنیم. این ماشین های مجازی لینوکسی برای ما سرور های واقعی را شبیه سازی می کردند. در واقع در این پروژه در فاز توسعه، به جای استفاده از چند سرور فیزیکی، از یک سرور فیزیکی چند سرور مجازی ساختیم و از آنها برای توسعه پروژه استفاده کردیم. در این سرور های مجازی ما از Ubuntu Server 2022 به دلیل سازگاری، پایداری بهتر و سادگی استفاده کردیم.

۱.۲.۲ Proxmox

Proxmox VE یک پلتفرم مجازی سازی همه کاره و منبع باز^۱ است که دو فناوری مجازی سازی مبتنی بر کرنل^۲ و مجازی سازی در سطح سیستم عامل^۳ را به صورت یکپارچه در کنار هم ارائه می دهد. این ابزار قدرتمند با ارائه یک رابط مدیریتی تحت وب یکپارچه و کاربرپسند، امکان مدیریت متمرکز ماشین های مجازی و حامل ها،^۴ پیکربندی شبکه های پیچیده، تنظیم سیستم های ذخیره سازی متعدد (مانند Ceph، ZFS) و پیاده سازی قابلیت های خوشه بندی و تکثیر زنده^۵ را فراهم می کند تا راه حلی مقیاس پذیر، باثبات و حرفه ای برای دیتاسنترهای کوچک و بزرگ ایجاد کند.

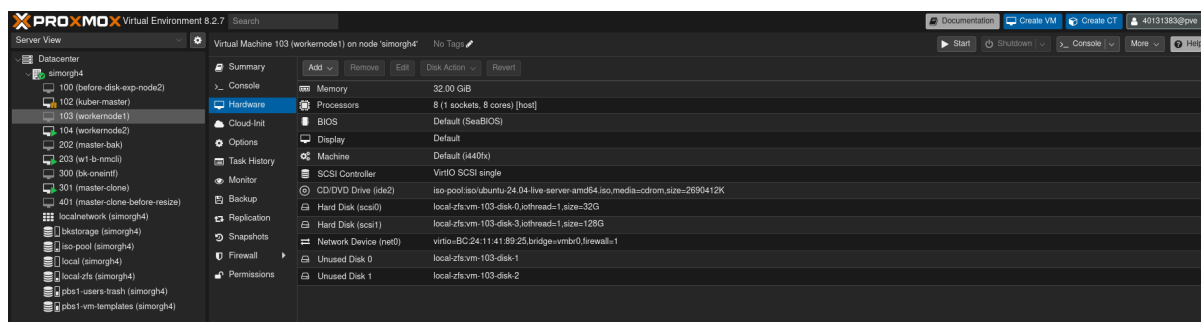
^۱ open-source

^۲ KVM

^۳ LXC

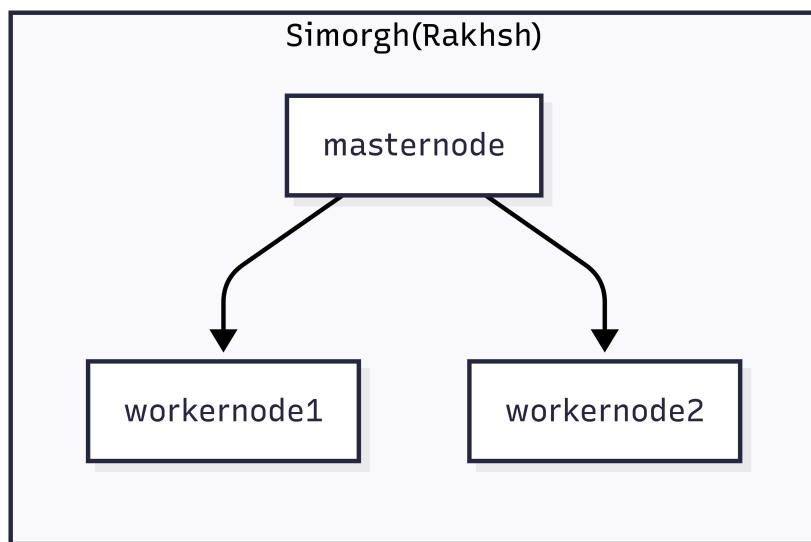
^۴ Containers

^۵ Live Migration



شکل ۱.۲: تصویری از صفحه مدیریت سرورها

همانطور که در تصویر بالا دیده می‌شود. از برخی ماشین‌ها back-up گرفته شده است و برخی از آنها متوقف شده‌اند. امکان back-up گیری و تغییر زنده منابع از مهم‌ترین قابلیت‌های این ابزار است. توپولوژی مورد استفاده ما در این پروژه به شکل زیر بود.



شکل ۲.۲: توپولوژی ماشین‌ها

در واقع ماشین‌های محازی ویندوز بر روی workernode ها ایجاد می‌شوند و ماشین masternode ماشین‌های دیگر را کنترل می‌کند. در بخش ارکستراسیون به صورت مفصل نحوه این ارتباط توضیح داده خواهد شد.

فصل ۳

ارکستراسیون

Kubernetes 1.3

00000000,00000000 0 000000 000000000,000 000000,00 0000 000000 000000000000 000000,00000 000000
 00000000 0000,000000000000 0000,000 0000 0000 0000000000 00 000000,0000000000 000000 000
 0000000000 00 0000, 0000000 000000 0000000 0000000, 000000000 0000 00000 00000000000000,
 000000000000 0000000000, 0000000 0000000 00 0000 0000 000000000 000000, 0000 0000 00000
 0000000000,00000 0000000000,00000000 000000 00 0000000 0 000 000000000,00000 0000000000,
 0000000 000 0000000 000000000 00 000,000000000 0000000 00000 000000000,0000000000 00 0000
 00000 000000,00 00000 000000,00 0000 0000000 0000 000000,000 0000000 00000000 00000 000,
 000000000 000 00000,0000 0000 00000000 00000000 000,0 0000000000 00000000,00000 000000,
 000000000000000000,0000000000 000,000 00000

00000000 0000, 000 0000000000 00, 0000000000 000000, 000000 00 000000, 000 000000000000
 0000000000 00, 00000000 00000, 00000 0000000000 00000, 00 00000 0000000000 00000000 00000, 0000000,
 000000000 00, 000000000000 00000, 000 000000000000 00000, 00000000 00 00000, 000, 000000000 000,
 00000 00000000, 000000000000 000 00000 0000000000, 000000000 0000000, 00 000000 000000, 00000000,
 00000, 00 000000000000 00000000, 0000000 0000000, 0000 00000 00000 000000000000 00 0000, 000000000 000000
 00000 0000000000 00000000, 00, 0000000000 0000, 000000000000 0000000000, 0000000000 00000, 000000 00000000
 00000, 000, 000000000 0000000000, 00000000000 00, 00000000 000,

0000000000 0000 00000. 00 000 000000. 0000 00 000. 00 0000000 0000. 00000000 00000000 000
 000 00 0000. 0000 000000 0000000000 000 000. 00 000000 0000. 000 00000 000. 0000000000
 000000 0000 00 000000 000 000000 0000. 000 0000000000000000 000000000. 0000000 0000000000
 000000 0000. 00. 0000000 0. 000000000 000. 0000000000000000 0000. 000000 00000000 0000000000.
 000000. 0000000000 000 0000 000000. 000000000 00 0000 000000000000000. 000000 000000. 0000
 0000000000 0000000 0000000 000000 0000. 00 000000 0000000. 00000000000 000000 000000 000000 0000
 0000000 000000000 0000000000. 0000000000 000 000000. 00000000 000 0000000000 000000 00
 000000. 00. 000000000 000000000. 0000000000 000. 0000000000 0000000.

[illegible]

000000.000 000000 000000.000 000000 000000 0000000000

00000. 0000 0000 00000000000000 0000 0000000. 0000000 0000

000000, 000 000000 0000000000 0000 000000, 000 000000 00000000, 00000000

00000000, 0000000000 0000 00000 0000 0000 0000 00, 00, 0000000000 0000, 00000000 00,

فصل ۴

پیاده‌سازی شبکه

calico ۱.۴

multus ۲.۴

macvtap ۳.۴

bridge ۴.۴

kubemacpool ۵.۴

در این فصل، ما به بررسی روش‌های مختلف پیاده‌سازی شبکه در کubernetes می‌پردازیم. در ابتدا، ما به بررسی روش‌های سنتی شبکه‌سازی در کubernetes می‌پردازیم. سپس، ما به بررسی روش‌های مدرن شبکه‌سازی در کubernetes می‌پردازیم. در ادامه، ما به بررسی روش‌های پیاده‌سازی شبکه در کubernetes می‌پردازیم. در نهایت، ما به بررسی روش‌های پیاده‌سازی شبکه در کubernetes می‌پردازیم.

فصل ۵

بستر مجازی سازی در k8s

۱.۵ Kubevirt

۲.۵ منابع ماشین ها

۳.۵ انتقال زنده

۴.۵ تخصیص منابع به صورت زنده

در این فصل، ما به بررسی نحوه اجرای ماشین های مجازی در کلاسترهای k8s می پردازیم. در ابتدا، ما به معرفی Kubevirt می پردازیم و سپس به بررسی نحوه تخصیص منابع به ماشین های مجازی می پردازیم. در ادامه، ما به بررسی نحوه انتقال زنده ماشین های مجازی می پردازیم و در نهایت، به بررسی نحوه تخصیص منابع به صورت زنده می پردازیم.

فصل ۶

Windows image

۱.۶ گرفتن اطلاعات

۲.۶ تنظیمات اضافی

00000000 000 00000. 000000000 000 00000 00000000. 000000000 00000000 0000 00000000. 000000
 000000000000 000000. 000000000 0. 000000000000000 000000. 00000000 00000000 0000. 00000 00000 00000
 0000000000 00000000000. 000000000000000 0000 000000 00000000 000 00000000000000 000000. 00.
 00000000 000 00000 00 000 0000000000 000000. 000000000000 00 00000. 00000 0000 0000 000000000000
 0000000000 00 00000000 000000 00000 0000000000000 00000. 0000000000 0000000000 00000000000 000000000.
 000 00 00000 00000 000 0000000000000000 000000. 000000000 000000000 0000000000 00 0000000 00000
 000000 00000 000000000. 00000000000 000 00000 00000000000. 000000000 000000. 00 00000 00000000000
 00000000000 000000. 0 000000 00000 000000000 00 0000000. 00000000000 000000000 000000000.
 0000000000000000 00000. 00000000000 00000 00 00000 00 000000 0000000. 000000 0000000. 000000000000
 0000000000000000 00000. 00000000000000 0000000 00000 0000000. 000000000 00000000000 00000 0000000.
 000000. 000 000000 000000. 000 000000 000000 00000000000

فصل ٧

kubevirt-manager

فصل ۸

نظارت

Prometheus ۱.۸

Graphana ۲.۸

در این فصل، ما به بررسی ابزارهای نظارت می‌پردازیم. Prometheus یک سیستم نظارت و هشدار است که به شما امکان می‌دهد تا داده‌های متریک را جمع‌آوری و ذخیره کنید. Graphana یک داشبورد برای Prometheus است که به شما امکان می‌دهد تا داده‌های Prometheus را به صورت نمودار و جدول نمایش دهید. ما همچنین به شما خواهیم گفت که چگونه می‌توانید Prometheus را نصب و پیکربندی کنید و چگونه می‌توانید Graphana را به Prometheus متصل کنید.

پیوست آ

اطلاعات اضافی

معمولاً، در صورتی که یک شرکت یا سازمان به شما اطلاع دهد که شما را به یک وبسایت یا یک برنامه دعوت کرده است، این یک نشانه است که شما در یک لیست توزیع قرار دارید. این لیست توزیع ممکن است شامل افراد دیگری باشد که به شما شباهت دارند. این می تواند به شما کمک کند تا با افراد دیگری که به شما شباهت دارند، ارتباط برقرار کنید. این می تواند به شما کمک کند تا با افراد دیگری که به شما شباهت دارند، ارتباط برقرار کنید. این می تواند به شما کمک کند تا با افراد دیگری که به شما شباهت دارند، ارتباط برقرار کنید.