

دانشگاه صنعتی اصفهان

مرکز ابررایانش

گزارش کارآموزی



مرکز ابررایانش 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

۳	۱ معرفی
۳	۱.۱ پیش زمینه من
۳	۲.۱ مرکز ابررایانش شیخ بهائی
۳	۱.۲.۱ منظور از یک خوشه HPC چیست؟
۴	۲.۲.۱ معماری
۴	۳.۲.۱ نحوه کار
۵	۳.۱ پروژه
۶	۲ آماده سازی
۶	۱.۲ آماده سازی کامپیوترها
۶	۲.۲ آماده سازی سرورها
۶	۱.۲.۲ Proxmox
۸	۳ ارکستراسیون
۸	۱.۳ Kubernetes معرفی
۸	۲.۳ نحوه کار
۹	۳.۳ کاربرد در پروژه
۹	۴.۳ نصب k8s
۱۰	۴ پیاده سازی شبکه
۱۰	۱.۴ calico
۱۰	۲.۴ multus
۱۰	۳.۴ macvtap
۱۰	۴.۴ bridge
۱۰	۵.۴ kubemacpool
۱۱	۵ بستر مجازی سازی در k8s
۱۱	۱.۵ Kubevirt
۱۱	۲.۵ منابع ماشین ها
۱۱	۳.۵ انتقال زنده
۱۱	۴.۵ تخصیص منابع به صورت زنده
۱۲	۶ Windows image
۱۲	۱.۶ گرفتن اطلاعات
۱۲	۲.۶ تنظیمات اضافی
۱۳	۷ kubevirt-manager
۱۴	۸ نظارت
۱۴	۱.۸ Prometheus

۱۴ Graphana ۲.۸

۱۵ آ اطلاعات اضافی

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.

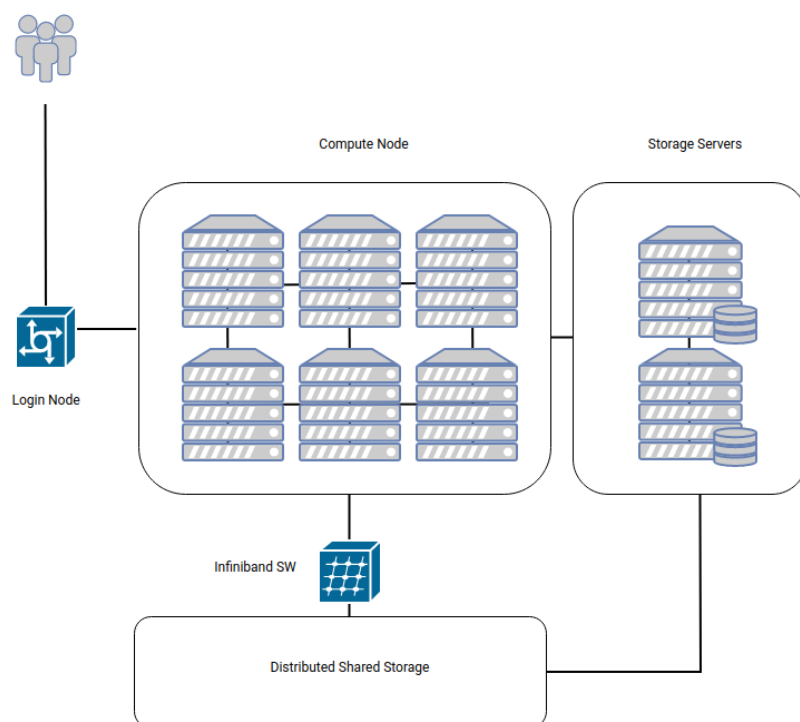
به عنوان مثال، می‌توانید دانش‌آموزی را در نظر بگیرید که یک کد بزرگ (نوشته شده در بیش از هزاران خط) دارد و اگر تمایل داشته باشد کد خود را در لپ‌تاپ شخصی خود اجرا کند، چند سال طول می‌کشد تا پردازش تمام شود و نتیجه نهایی را ببیند؟ بنابراین اگر کد خود را به ابررایانه منتقل کنند، در مدت زمان بسیار کوتاهی 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 کاربر پسند، امکان مدیریت متمرکز ماشین های مجازی و حامل ها،^۴ پیکربندی شبکه های پیچیده، تنظیم سیستم های ذخیره سازی متعدد (مانند ZFS، Ceph) و پیاده سازی قابلیت های خوشه بندی و جابه جایی زنده^۵ را فراهم می کند تا راه حلی مقیاس پذیر، باثبات و حرفه ای برای دیتاسنترهای کوچک و بزرگ ایجاد کند.

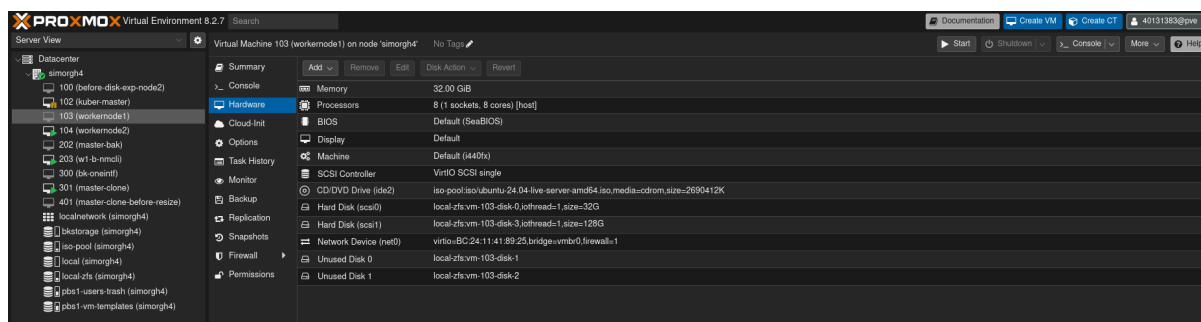
^۱ open-source

^۲ KVM

^۳ LXC

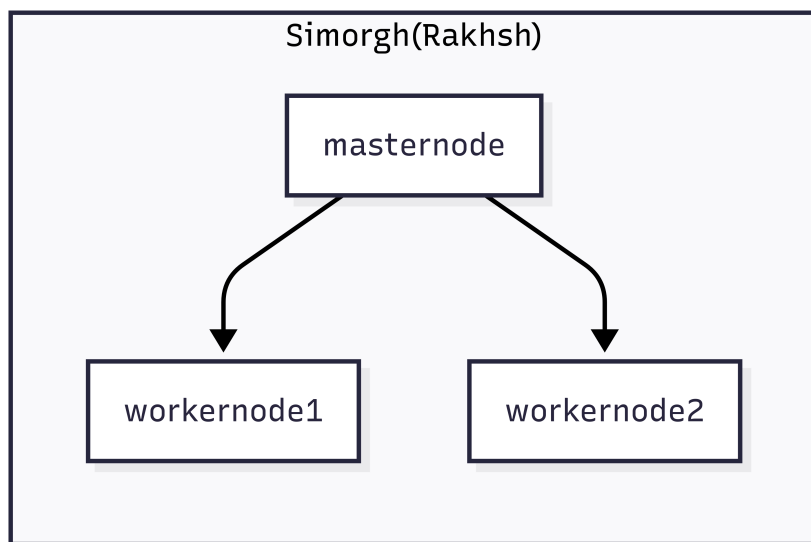
^۴ Containers

^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 مجازی 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

ارکستراسیون در دنیای فناوری اطلاعات به فرآیند خودکارسازی استقرار، مدیریت، مقیاس گذاری و شبکه سازی برای حامل ها اطلاق می شود. در یک معماری مبتنی بر میکروسرویس ها، که یک برنامه به ده ها یا صدها جزء مستقل تقسیم می شود، ارکستراسیون نقشی حیاتی در هماهنگ سازی این اجزا، توزیع بار پردازشی، و تضمین در دسترس بودن و مقاومت در برابر خطا^۱ برنامه ایفا می کند. این مفهوم، مدیریت چرخه عمر حامل ها را از سطح یک میزبان منفرد فراتر برده و آن را در مقیاس یک خوشه از سرورها تعریف می کند.

۱.۳ Kubernetes معرفی

Kubernetes یک پلتفرم ارکستراسیون مبتنی بر حامل^۲ متن باز و گسترده است که به عنوان یکی از اصلی ترین راه حل ها در این حوزه شناخته می شود. این پلتفرم با به کارگیری یک معماری Master-Worker^۳، امکان تعریف حالت مطلوب^۴ برنامه را از طریق فایل های اعلانی^۵ فراهم می سازد و سپس به صورت خودکار وظایفی مانند زمان بندی حاملها بر روی گره های کارگر، نظارت بر سلامت آن ها، و بازیابی خودکار را انجام می دهد.

مزایای کلیدی Kubernetes شامل مقیاس پذیری الاستیک^۶ و خودکار (هم به صورت افقی و هم عمودی)، بهینگی در مصرف منابع، قابلیت حمل بالا^۷ بین محیط های ابری مختلف، و تسهیل فرآیندهای DevOps^۸ و CI/CD^۹ است. در نتیجه، این پلتفرم به عنوان استاندارد اصلی برای اجرا و مدیریت برنامه های کاربردی مدرن و مقیاس پذیر در محیط های هایبرید^{۱۰} و چندابری^{۱۱} تبدیل شده است.

۲.۳ نحوه کار

Kubernetes با دریافت تعریف حالت مطلوب برنامه از کاربر، به صورت خودکار فرآیندهای مدیریت و هماهنگی حامل ها را انجام می دهد. کاربران مشخص می کنند که برنامه شان باید چگونه اجرا شود و Kubernetes مسئولیت رساندن سیستم به آن حالت و حفظ آن را بر عهده می گیرد.

سیستم به طور مداوم وضعیت فعلی را با حالت مطلوب مقایسه می کند. اگر بین این دو تفاوتی وجود داشته باشد، Kubernetes به صورت خودکار اقدامات لازم را انجام می دهد. به عنوان مثال، اگر یک حامل از کار بیفتد، سیستم به طور

^۱resilience

^۲Container Orchestration Platform

^۳Control Plane and Worker Nodes

^۴Desired State

^۵Declarative YAML/JSON

^۶Elastic Scalability

^۷High Portability

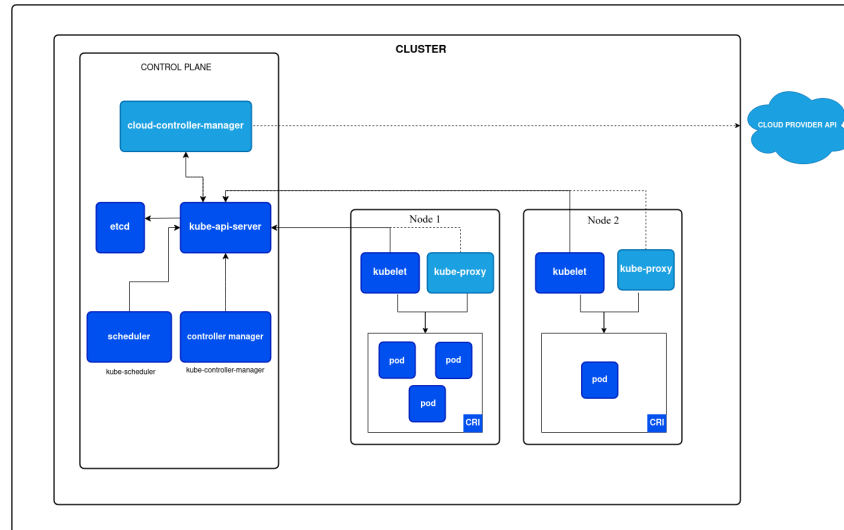
^۸Development and Operations

^۹Continuous Integration and Continuous Delivery/Deployment

^{۱۰}Hybrid Cloud

^{۱۱}Multi-Cloud

خودکار حامل جدیدی را جایگزین آن می‌کند تا برنامه همیشه در حالت مطلوب باقی بماند. Kubernetes حامل‌ها را بین گره‌های مختلف توزیع می‌کند و اطمینان حاصل می‌نماید که از منابع سخت‌افزاری به بهترین شکل ممکن استفاده شود. همچنین شبکه‌های لازم برای ارتباط بین حامل‌ها و تعادل بار بین آنها را به صورت خودکار مدیریت می‌کند تا برنامه به صورت یکپارچه و با کارایی بالا قابل دسترسی باشد.



شکل ۱.۳: معماری کوبرنتیز

ممکن است نام K8s را هم به جای kubernetes شنیده باشید که معنای آن حرف k و ۸ حرف میان آن و حرف s است.

۳.۳ کاربرد در پروژه

k8s به دلیل مقاومت در برابر خطا، بازگشت به حالت مطلوب پس از وقوع خطا، مقیاس پذیری و انعطاف پذیری خوب و ساختار ماژولار آن یک گزینه عالی برای مدیریت ماشین‌های مجازی ویندوزی و دیگر چیزهای مورد نیاز آنها است. در اصل یکی از نیازهای اصلی پروژه که در دسترس بودن دائمی و مقاومت در برابر خطا است که به کمک k8s رفع می‌شود. نکته حائز اهمیت این است که k8s به تنهایی تمام کارها را انجام نمی‌دهد. در واقع این ابزار فقط بستر را برای حامل‌ها فراهم می‌سازد. به دلیل امکان تعریف منابع دلخواه و در نتیجه آن، قابلیت توسعه پذیری و بالای این ابزار، افزونه‌های بسیار زیادی برای انجام کارهای مختلف وجود دارد. برخی از این افزونه‌ها مربوط به سیستم داخلی خود k8s است مانند افزونه‌های شبکه که در ادامه معرفی می‌شوند و یا افزونه‌هایی که یک سرویس کامل را اجرا می‌کنند مانند kubevirt که در ادامه پروژه به آن نیاز داشتیم.

۴.۳ نصب k8s

فصل ۴

پیاده‌سازی شبکه

calico ۱.۴

multus ۲.۴

macvtap ۳.۴

bridge ۴.۴

kubemacpool ৫.২

0000000000 0000 00000. 00 0000 000000. 00000 00 0000. 00 00000000 0000. 000000000 00000000 0000
 0000 00 00000. 00000 000000 0000000000 0000 0000. 00 000000 00000. 0000 00000 0000. 00000000000
 0000000 00000 00 0000000 0000 0000000 00000. 0000 00000000000000000000 0000000000. 00000000 00000000000
 0000000 00000. 00. 00000000 0. 0000000000 0000. 00000000000000000 00000. 0000000 00000000 00000000000.
 000000. 0000000000 0000 00000 00000000. 00000000000 00 00000 000000000000000. 0000000 0000000. 0000
 00000000000 00000000 00000000 0000000 00000. 00 000000 00000000. 000000000000 000000 0000000 0000000 00000
 00000000 0000000000 00000000000. 0000000000 0000 0000000. 000000000 0000 0000000000 0000000 00
 000000. 00. 0000000000 0000000000. 0000000000 0000. 00000000000 0000000.

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 00000000000
 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. 000000000000 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 و نمایش داده‌ها می‌پردازیم. Prometheus یک سیستم نظارت و هشداردهی است که به شما امکان می‌دهد تا داده‌های سیستم‌های مختلف را جمع‌آوری و ذخیره کنید. Graphana یک ابزار نمایش داده‌ها است که به شما امکان می‌دهد تا داده‌های Prometheus را به صورت نمودار و داشبورد نمایش دهید. در این فصل، ما به بررسی نحوه نصب و پیکربندی این ابزارها می‌پردازیم و همچنین نحوه جمع‌آوری و نمایش داده‌ها را یاد می‌گیریم.

پیوست آ

اطلاعات اضافی

معمولاً، در صورتی که یک شرکت یا سازمان به منظور ارائه خدمات به مشتریان، نیاز به جمع‌آوری اطلاعات اضافی دارد، باید این کار را به روشی قانونی و شفاف انجام دهد. این شامل اطلاع‌رسانی به مشتریان در مورد نحوه جمع‌آوری و استفاده از اطلاعات آنها، و همچنین گرفتن رضایت آنها است. در صورتی که یک شرکت یا سازمان به منظور ارائه خدمات به مشتریان، نیاز به جمع‌آوری اطلاعات اضافی دارد، باید این کار را به روشی قانونی و شفاف انجام دهد. این شامل اطلاع‌رسانی به مشتریان در مورد نحوه جمع‌آوری و استفاده از اطلاعات آنها، و همچنین گرفتن رضایت آنها است.