



BIG DATA WEEK
A GLOBAL FESTIVAL OF DATA

Haciendo **Big Data** en los tiempos de **Contenedores y Kubernetes**

Nicolás Lecaros

<https://www.linkedin.com/in/nicolaslecaros/>



BIG DATA WEEK
A GLOBAL FESTIVAL OF DATA



Nicolás Lecaros

<https://www.linkedin.com/in/nicolaslecaros/>

¿Quién soy?

Nicolás Lecaros () {



Corporate IT Project Manager;



Professor, School of Computing and Telecommunications;

/* Eterno estudiante, profesor y profesional TI con +11 años de experiencia en análisis, diseño, desarrollo, implementación, soporte y gestión de proyectos TI. En donde me he especializado en Business Intelligence (BI), Business Analytics (BA), Big Data y Cloud Computing. */

}

Mi trayectoria:



Nicolás Lecaros

<https://www.linkedin.com/in/nicolaslecaros/>

¿Puede una plataforma concebida para permitir la ejecución de aplicaciones efímeras convertirse en el sistema operativo elegido para ejecutar cargas de trabajo de datos?



Antecedentes

- La mala latencia de la red ***no es un problema*** para los proveedores de la nube.
- Hay muchos más de ellos para elegir, por lo que estamos hablando de un entorno ***multicloud***.
- Muchas organizaciones están implementando sus propias nubes privadas, por lo que estamos hablando de una ***nube híbrida***.

Nos enfrentamos a una situación en la que los datos siguen siendo cada vez más grandes, pero también ***deben ser flexibles*** para seguir escalando en múltiples infraestructuras.

Aquí es donde entran los **Contenedores y Kubernetes**.

La Infraestructura On-Premise rápidamente nos queda pequeña



¿Por qué Cloud, por qué ahora?

“Hubo un tiempo en que cada hogar, pueblo o aldea tenía su propio pozo de agua. Hoy en día, los servicios públicos compartidos nos dan acceso a agua limpia simplemente abriendo la llave.

Al igual que en la compañía de agua, ***hay un equipo de profesionales dedicados que se aseguran de que el servicio proporcionado sea seguro y esté disponible 24/7.***

Cuando la llave no está abierta, no solo está ahorrando agua, sino que no está pagando por los recursos que no necesita actualmente.”



Las empresas están llevando su trabajo a la nube.



Existen muchas opciones, y sigue creciendo el mercado...





Nicolás Lecaros

<https://www.linkedin.com/in/nicolaslecaros/>

Contenedores

- ✓ Necesitan ser administrados.
- ✓ La creación de las redes es difícil.
- ✓ El **contenedor** debe ser programado, distribuido y con carga equilibrada.
- ✓ Y los datos deben persistir en alguna parte.

Un **contenedor** proporciona un contexto aislado en el que se puede ejecutar una aplicación, junto con su entorno. Pero los contenedores aislados a menudo necesitan ser gestionados y conectados al mundo externo. Los sistemas de archivos compartidos, las redes, la programación, el balanceo de carga y la distribución son todos desafíos.

THE #1 PROGRAMMER EXCUSE
FOR LEGITIMATELY SLACKING OFF:

"My Docker containers are building!"





kubernetes

Nicolás Lecaros

<https://www.linkedin.com/in/nicolaslecaros/>

¿Qué es Kubernetes?

Según el sitio web kubernetes.io, Kubernetes es:

”...un sistema de código abierto para automatizar la implementación, el escalado y la administración de aplicaciones en contenedores”.

Kubernetes

- Ejecutar un contenedor en una computadora portátil es relativamente simple. Pero conectar contenedores en varios hosts, escalarlos, implementar aplicaciones sin tiempo de inactividad y descubrir servicios entre varios aspectos puede ser difícil.
- Kubernetes aborda esos desafíos desde el principio con un conjunto primitivo y una potente API abierta y extensible. ***La capacidad de agregar nuevos objetos y controladores permite una fácil personalización para diversas necesidades de producción.***

Kubernetes

- Κυβερνήτης significa “timón” o “piloto” en griego.
- Administra grupos de contenedores.
- Fue diseñado por **Google** y donado a **Cloud Native Computing Foundation** (parte de la **Linux Foundation**).
- Soporta múltiples nubes y metales.
- 100% de código abierto.
- Administra las aplicaciones, no las máquinas.

¿Para que sirve Kubernetes?

Kubernetes

- ✓ Ejecutar aplicaciones en contenedores encima de **Kubernetes** está de moda.
- ✓ El nuevo mundo de los contenedores no siempre es amable con los desarrolladores de aplicaciones que están acostumbrados a una experiencia rápida de desarrollo local.
- ✓ Con **Kubernetes**, puede haber muchas diferencias entre la forma en que se desarrolla y ejecuta una aplicación localmente en comparación con un clúster de Kubernetes.



Kubernetes

- ✓ El aprendizaje automático (ML) y otras aplicaciones de computación intensiva se vuelven cada vez más populares.
- ✓ Existe una creciente necesidad de aceleradores de hardware como GPU.
- ✓ Si tiene cargas de trabajo de ML en contenedores, [puede usar GPU](#) en [Google Kubernetes Engine \(GKE\)](#), que proporciona un entorno listo para producción.
- ✓ No existe un entorno de desarrollo local de Kubernetes que admita cargas de trabajo de GPU.
- ✓ Para desarrollar y probar aplicaciones de GPU, debe usar un entorno local que no se parece a su entorno de producción GKE.



Kubernetes



Keubernetes



Release it!



Contenedores y Kubernetes

- **Docker** ha permitido aprender, investigar y testear a un ritmo muy rápido, y de manera limpia y ordenada, un montón de nuevas herramientas.
- **Docker** ha revolucionado la filosofía de **DevOps** y tiene una gran comunidad de desarrolladores detrás, que amplían y mejoran las imágenes sobre las que se están generando nuevas aplicaciones.
- **Docker** añade las características de la portabilidad, la agilidad, el aislamiento y la abstracción, lo que permite crear soluciones mucho más rápidamente, y con mucha más fiabilidad.

Contenedores y Kubernetes

- Además, **Kubernetes**, como herramienta de orquestación de contenedores, ha traído la *fácil escalabilidad, la resiliencia, la facilidad de la comunicación entre los contenedores, la despreocupación por el balanceo de carga y la facilidad para gestionar la actualización de versiones*, entre otras muchas características.
- También es importante resaltar que *ha permitido conectar contenedores de Docker en diferentes servidores, en la nube y en entornos mixtos*.



Where is
OUR DATA?!

Ingenieros
de Datos

Agile
DEVELOPMENT!

Ingenieros
de Software

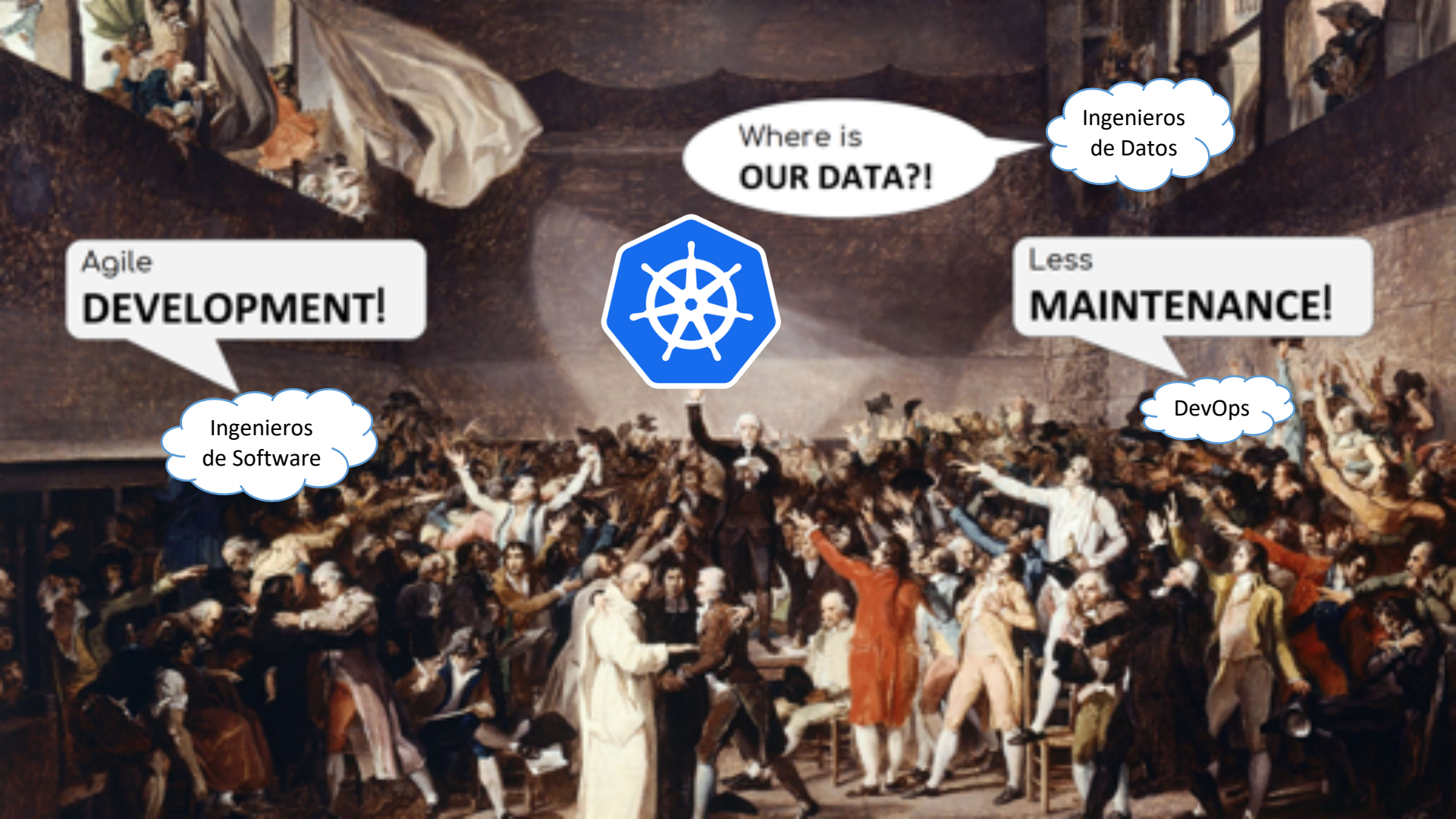
Less
MAINTENANCE!

DevOps

Situación Actual

- Diferentes equipos típicamente involucrados en el ciclo de vida del desarrollo de software y sus objetivos principales.
- Los **ingenieros de software** quieren un desarrollo de software ágil.
- Los **ingenieros de datos** se preocupan por dónde están los datos y quieren que los sistemas de bases de datos sigan funcionando.
- Los equipos de **DevOps** quieren que todos los sistemas funcionen con menos mantenimiento e interrupciones.

Debido a la revolución de las tecnologías de Contenedores y Kubernetes, todos estos objetivos son posibles de lograr en las organizaciones.



Where is
OUR DATA?!

Ingenieros
de Datos

Agile
DEVELOPMENT!

Ingenieros
de Software



Less
MAINTENANCE!

DevOps

Modernos Pipelines bajo Kubernetes



Modernos Pipelines bajo Kubernetes

- Las tecnologías de gestión de contenedores como **Kubernetes** hacen posible implementar tuberías modernas de big data.
- En el pasado, las soluciones de Big Data solían estar basadas principalmente en Hadoop , pero el ecosistema ha evolucionado en los últimos años con nuevas bases de datos, transmisión de datos y soluciones de aprendizaje automático que requieren más que el modelo de implementación de Hadoop (MapReduce, YARN y HDFS).
- Estas soluciones también requieren una capa de programación de clúster para alojar diversas cargas de trabajo.

Modernos Pipelines bajo Kubernetes

- Debemos considerar un marco común para crear aplicaciones de análisis de extremo a extremo nativas en la nube.
- Los desarrolladores deben desacoplar los servicios de datos de las aplicaciones y los marcos para que las soluciones de big data sean flexibles y eficientes.
- También ayuda con los servicios de datos que generalmente se usan para administrar diferentes tipos de datos, como datos no estructurados, estructurados o de transmisión.



Modernos Pipelines bajo Kubernetes

- *Los mundos de Kubernetes y los datos se están fusionando bastante rápido.* De hecho, es bastante común no solo conectar servicios Kubernetes a sistemas de datos con estado como almacenamiento, bases de datos o Spark, sino también para *ejecutar esos sistemas como servicios en Kubernetes.*
- Por cierto, también *hay una suposición común de que Kubernetes matará cualquier emoción persistente sobre Hadoop,* porque Kubernetes proporciona una capa de orquestación más simple y flexible para los datos, servicios, así como soporte para una gama más amplia de plataformas de almacenamiento.

Modernos Pipelines bajo Kubernetes

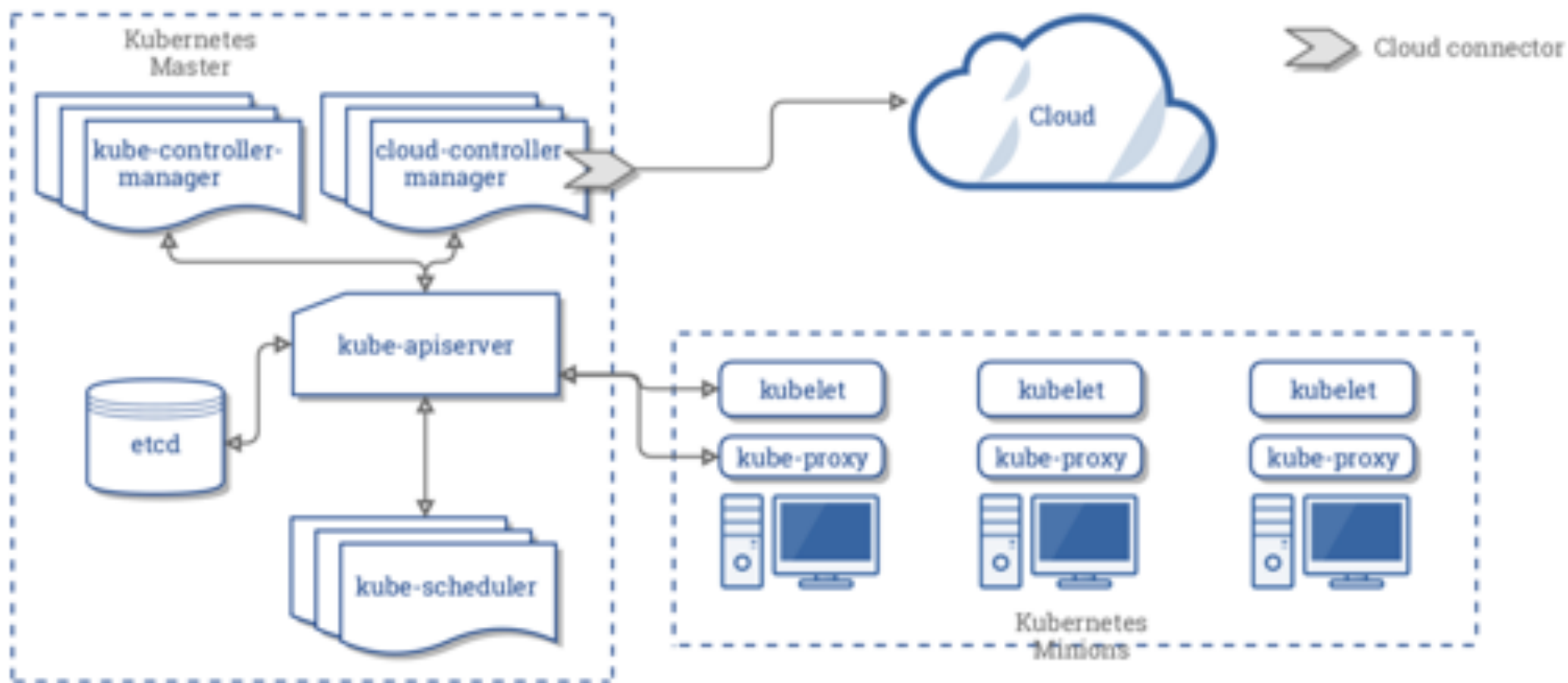
Hay un par de razones para esta unión de mundos.

- Una es que Kubernetes es una parte integral de muchos esfuerzos de modernización, que también incluyen: machine learning / inteligencia artificial, IoT y otros componentes con muchos datos. ***Solo tiene sentido que la integración entre estos sistemas sea una prioridad.***
- La otra razón es que ***cada vez más aplicaciones heredadas se están portando a contenedores que se ejecutan en clústeres de Kubernetes***, y estas aplicaciones requieren almacenamiento con estado de una amplia variedad de fuentes.

KUBERNETES

KUBERNETES EVERYWHERE





Opciones para usar Kubernetes

Kubernetes Auto Administrado

- GKE es un entorno de producción administrado que permite implementar aplicaciones en contenedores.
- Kubernetes Engine se lanzó en 2015, con base en los más de 12 años de experiencia de Google en la ejecución de servicios como Gmail y YouTube en contenedores. Permite la puesta en marcha de [Kubernetes](#) sin demoras, ya que elimina por completo la necesidad de instalar, administrar y operar tus propios clústeres.



Google Kubernetes Engine (GKE)
Kubernetes-as-a-service

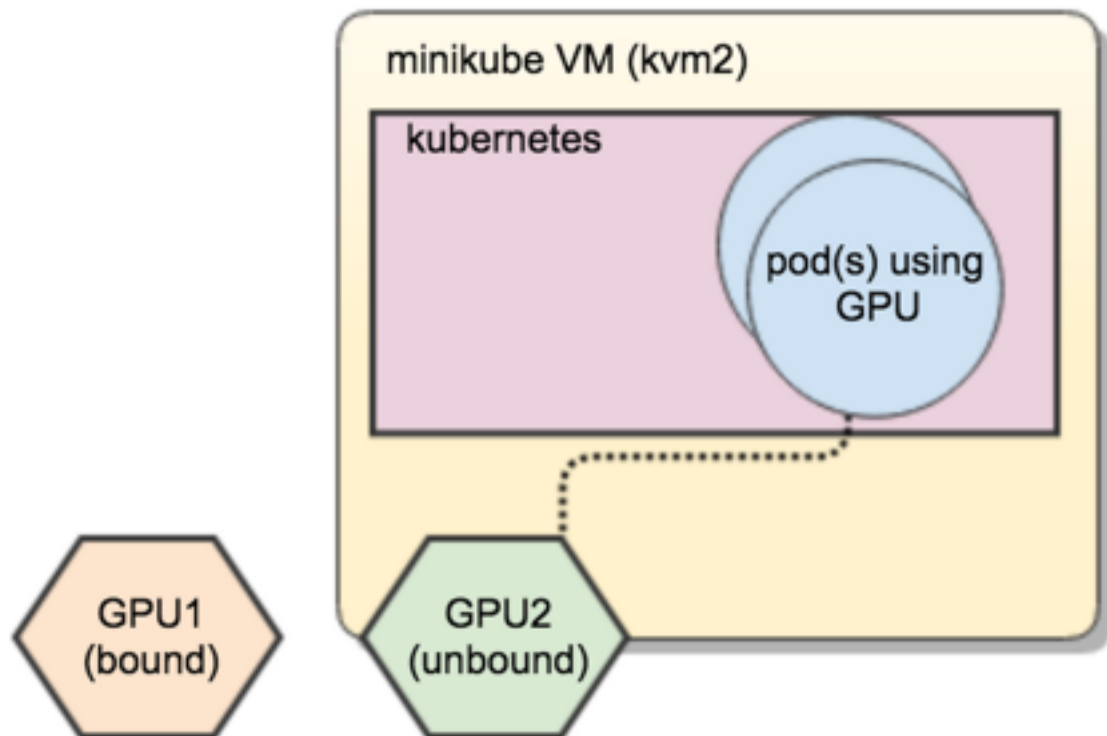
MiniKube

- Minikube es la solución para ejecutar cargas de trabajo de Kubernetes localmente.
- Ejecuta un clúster de Kubernetes de un solo nodo dentro de una VM en su máquina de desarrollador.
- Cuenta con soporte de GPU, mejorando la experiencia del desarrollador al crear cargas de trabajo de GPU en Kubernetes.

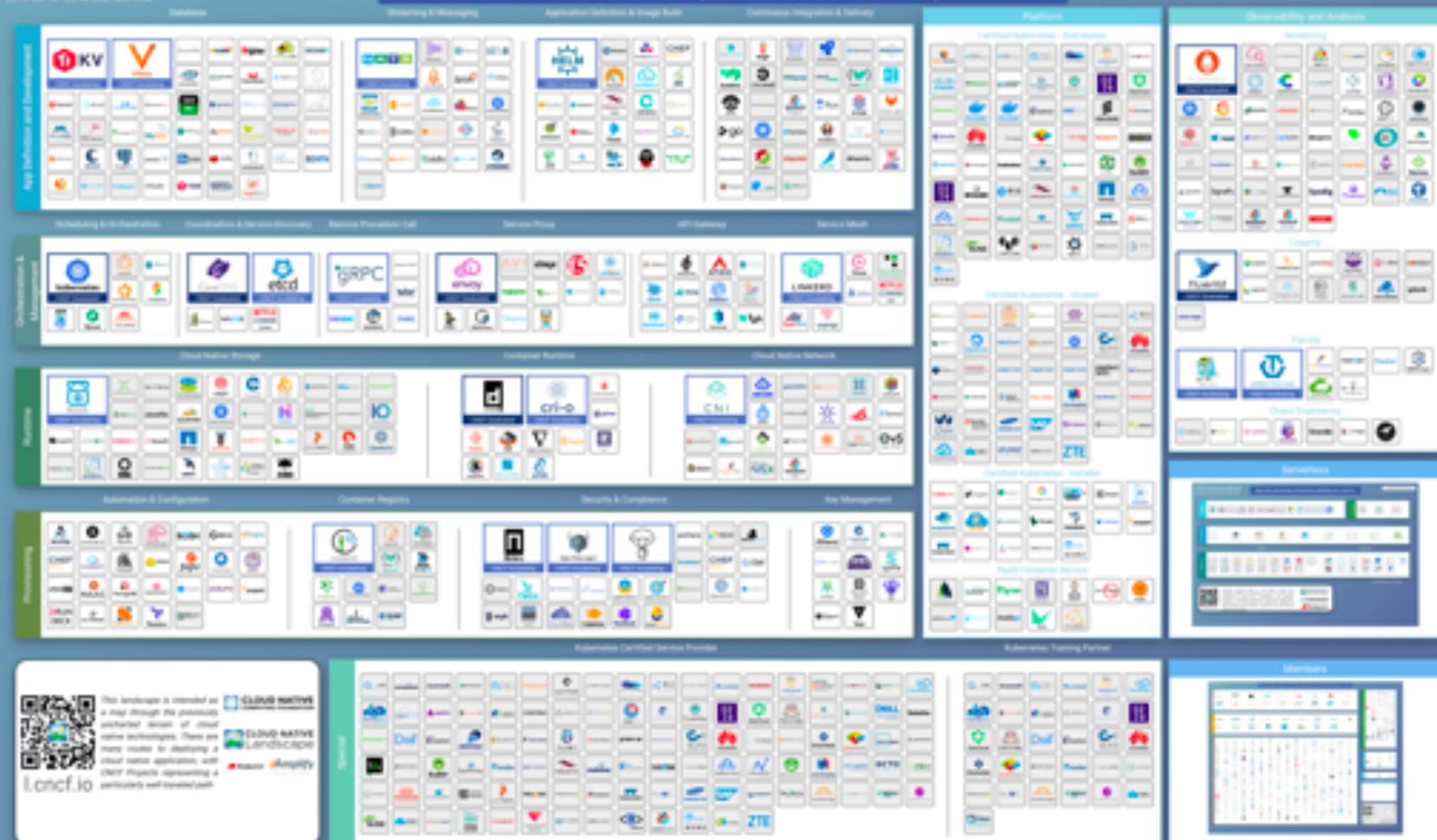


minikube

Linux host



Kubernetes es un ecosistema...



l.cncf.io

This landscape is intended as a map through the previously uncharted terrain of cloud native technologies. There are many routes to deploying a cloud native application, with CNCF Projects representing a particularly well-trodden path.



Partners

Deployed my blog on Kubernetes



Me dijiste que K8s
era lo mejor

Lo mejor para crear
sistemas agnósticos



Muchas gracias.

Nicolás Lecaros

<https://www.linkedin.com/in/nicolaslecaros/>

¡Nunca dejes de aprender!

