

Serverless Research with Ion Stoica

Nombre: Fabián Concha Sifuentes

El podcast **Serverless Research with Ion Stoica** nos da una introducción al laboratorio de investigación AMPLab de Berkeley, donde nacieron proyectos revolucionarios como Apache Spark y Mesos. Esta cercanía con Silicon Valley ha permitido que Berkeley sea un punto de encuentro entre la investigación académica y las necesidades de la industria, especialmente en sistemas distribuidos.

Introducción al paradigma serverless

Ion Stoica explica que la computación en la nube comenzó en 2006 con Amazon Web Services (AWS), que permitió a los usuarios acceder rápidamente a máquinas virtuales (VMs) sin tener que construir centros de datos. Sin embargo, aunque las VMs facilitaron la elasticidad y el escalado, los usuarios aún debían gestionarlas manualmente.

La computación serverless representa la evolución de ese modelo, eliminando la necesidad de administrar servidores. El proveedor de la nube se encarga de toda la infraestructura, y el usuario solo paga por el tiempo de ejecución de su código. Las características clave del serverless incluyen:

- Separación entre almacenamiento y cómputo.
- Pago por ejecución (no por recursos reservados).
- Escalabilidad automática sin intervención manual.

Tipos de plataformas serverless

Stoica distingue dos categorías principales:

1. **Function-as-a-Service (FaaS)**, como AWS Lambda, que ejecuta funciones específicas sin preocuparse por la infraestructura.
2. **Backend-as-a-Service (BaaS)**, como Google BigQuery, que ofrece servicios completos como bases de datos o motores de consulta sin administración explícita.

Como investigador, Stoica se interesa más por FaaS debido a la flexibilidad que ofrece para construir sistemas complejos y realizar investigaciones.

Desafíos técnicos del serverless

Al intentar utilizar serverless para aplicaciones más generales como análisis de datos o aprendizaje automático, surgen varios retos:

- **Latencia de arranque en frío**: cargar entornos de ejecución grandes, como bibliotecas de Python, puede tomar varios segundos.

- **Intercambio de datos entre funciones:** actualmente se hace mediante almacenamiento en S3, lo cual es lento, costoso y con límites de I/O.
- **Falta de control sobre la ubicación de ejecución:** afecta el rendimiento de operaciones como difusión de datos o entrenamiento distribuido.
- **Seguridad y aislamiento:** los contenedores usados en serverless tienen menos aislamiento que las VMs.

Estos retos abren nuevas oportunidades de investigación, como diseñar almacenamiento efímero rápido o algoritmos que se adapten a estas limitaciones.

Pyron: un proyecto del laboratorio RISE

Para abordar algunas de estas limitaciones, Stoica y su equipo han desarrollado **Pyron**, una capa de abstracción sobre AWS Lambda que permite paralelizar tareas en Python de forma sencilla. Pyron ofrece primitivas de alto nivel como **map** y **reduce**, facilitando la ejecución de algoritmos complejos (por ejemplo, álgebra lineal) sin que el usuario deba preocuparse por la gestión de funciones individuales.

Este esfuerzo sigue la filosofía del paper "*Occupy the Cloud: Distributed Computing for the 99%*", que busca democratizar el acceso a la computación paralela, especialmente para científicos que no son expertos en infraestructura.

Dificultades de adopción en la nube

Se reflexiona que, aunque la nube promete simplificar la vida de los científicos de datos, en la práctica muchos aún la encuentran compleja:

- Deben elegir entre cientos de opciones de instancias.
- Les preocupa olvidar apagar recursos y generar altos costos.
- Escalar dinámicamente sigue siendo difícil.

Estas barreras son similares tanto para investigadores individuales como para empresas. Las grandes organizaciones, además, enfrentan desafíos adicionales relacionados con la seguridad y la migración de datos.

DataBricks y su evolución

Stoica cofundó DataBricks en 2013 para facilitar la adopción del análisis de datos y aprendizaje automático en la nube. Inicialmente centrados en Spark, evolucionaron hacia una plataforma integral de analítica unificada, que automatiza la gestión de clústeres y facilita la experimentación.

Casos de uso destacados incluyen:

- **Apple:** detección de intrusiones en tiempo real en dispositivos internos, procesando petabytes de logs.
- **Regeneron:** análisis genómico y clínico para diagnósticos personalizados y ensayos clínicos más eficaces.

En ambos casos, la preparación de datos es una tarea compleja que consume gran parte del esfuerzo, y DataBricks ofrece herramientas optimizadas para estas necesidades.

Cultura de investigación en Berkeley

Finalmente, Stoica describe la cultura única de los laboratorios de Berkeley, como AMPLab y el actual RISELab, inspirados por líderes como Dave Patterson. Estos laboratorios tienen una duración limitada (aprox. 5 años), objetivos claros y una fuerte conexión con la industria. Este modelo ha demostrado ser muy efectivo para producir resultados tangibles, como Spark, Mesos, y Tachyon.

Se comenta brevemente el interés de Stoica en criptomonedas, a pesar de su experiencia previa en redes peer-to-peer, y destaca el enfoque estratégico de Patterson para detectar tendencias tecnológicas antes de que se vuelvan mainstream.