



**EDUCACIÓN**  
**PROFESIONAL**

**Unidad 2:**

# **Introducción a las Herramientas de la Nube**

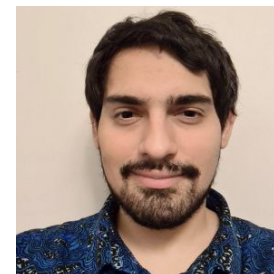
**Clase 1 - Introducción al curso**

**Profesor: Germán Leandro Contreras Sagredo**

## Introducción - Equipo

Antes de partir... ¿quiénes imparten este curso?

- **Profesor:** Germán Leandro Contreras Sagredo, Ingeniero Civil de Computación UC  
**Contacto:** [glcontreras@uc.cl](mailto:glcontreras@uc.cl)
- **Ayudante:** Valentina Rojas Mercier, Ingeniera Civil de Industrias, Diploma en Ingeniería de Computación UC en formación  
**Contacto:** [varojasm@uc.cl](mailto:varojasm@uc.cl)



¡Ambos trabajamos en **Kimche!**

## Introducción - Equipo

¿Qué es Kimche?

- Empresa *EdTech* dedicada a la mejora de la toma de decisiones en gestión escolar **a través de los datos**.
- Hacemos uso de **herramientas de la nube** que nos permiten abarcar eficientemente el problema de **Big Data**.



Conoce más en <https://www.kimche.co/>

## Introducción - Equipo

A lo largo de la clase, relataremos muchos casos de uso de las herramientas de la nube que nos han permitido salir adelante como empresa en Kimche, de forma que puedan apreciar el **potencial** de los contenidos a estudiar dentro de esta unidad.

## Introducción - Aspectos formales

¿Esta unidad no se llamaba “Ecosistema Hadoop”?

- Reestructuración del diplomado
- Esta será la primera versión del curso.
- **Objetivo:** Curso con fuerte énfasis en lo práctico en lo que respecta a herramientas de la nube

Por este motivo, ¡tendremos muchas actividades prácticas!

## Introducción - Aspectos formales

- Dos evaluaciones escritas: Evaluaciones **individuales** a ser realizadas en las clases 3 y 6.  
**Cada una vale un 20% de la calificación final.**
- Tarea de programación: Evaluación **grupal** que involucra la aplicación de las herramientas a ver en el curso.  
**Esta vale un 60% de la nota final.**
- Actividades prácticas todas las clases, **sin nota**. Estas tendrán por objetivo acercarlos a las herramientas y, por otra parte, permitirles avanzar gradualmente en su tarea.

## Introducción - Aspectos formales

Cronograma del curso. Cambios a ser informados oportunamente.

- **Clase 1:** 21 de julio
- **Clase 2:** 28 de julio
- **Clase 3:** 4 de agosto
  - Evaluación escrita N° 1 en horario de cátedra, primer módulo.
  - Subida de enunciado de la tarea grupal.
- **Clase 4:** 11 de agosto
- **Clase 5:** 18 de agosto
- **Clase 6:** 25 de agosto
  - Evaluación escrita N° 2 en horario de cátedra, primer módulo.
- **Entrega tarea grupal:** 1 de septiembre a las 23:59.

## Introducción - Un poco de ustedes

Antes de seguir esta cátedra, me interesaría saber un poco más de ustedes.

¿Conocen el concepto? ¿Han trabajado con la “nube” antes?



## La “nube”

¿Qué es la “nube”? ¿Qué entendemos de este concepto?

- “Modelo para permitir, de manera ubicua, conveniente y según demanda, el acceso a un conjunto de herramientas computacionales configurables (...) que pueden ser rápidamente provisionadas o liberadas con mínimo esfuerzo administrativo (...)” (National Institute of Standards and Technology).
- **“Colección de servicios que ayudan a los desarrolladores a enfocarse en proyecto en vez de la infraestructura que lo alimenta”** (Geewax, 2018).

## La “nube”

En resumen, entenderemos por la *nube* a las herramientas de desarrollo que nos permiten **abstraernos** de la arquitectura que sostendrá nuestros proyectos.

Ahora... ¿por qué nos interesa eso? ¿Qué potencial tiene el *cloud computing*?

## ¿Por qué nos interesa el *Cloud Computing*?

### Escalabilidad

Nuestras herramientas escalan según demanda (para arriba o para abajo), sin necesidad de hacer modificaciones manuales para adaptarnos.

La mayoría de los servicios a los que podemos acceder cobra según **cuotas de uso**. Por lo tanto, solo nos preocupamos de que el cobro por servicio esté dentro de un margen establecido\*.

\* Existen sistemas de alarma para poder ver en qué casos estamos excediendo un límite autoimpuesto.

## ¿Por qué nos interesa el *Cloud Computing*?

### Costos

Como se mencionó antes, las herramientas de la nube cobran según cuotas de uso. Esto es ideal para los costos de la empresa, ya que se “paga por lo que usa”.

Por otra parte, gracias a la **abstracción** de la infraestructura (IaaS - *Infrastructure as a Service*), nos podemos olvidar de la compra y mantención de servidores.

## ¿Por qué nos interesa el *Cloud Computing*?

### Flexibilidad

Si almacenamos nuestros datos en la nube, podemos escoger el sistema de gestión de bases de datos (MySQL, PostgreSQL, etc.), así como también la configuración de almacenamiento que queramos.

Por otra parte, podemos ejecutar en los servidores códigos de diversos lenguajes de programación.

Más aún, podemos hacer que dos servicios de nubes **distintas** interactúen entre sí.

## ¿Por qué nos interesa el *Cloud Computing*?

### **Confiabilidad**

Los sistemas de almacenamiento suelen poseer opción de respaldo en caso de caídas del sistema o errores inesperados, lo que aumenta la confianza en su uso.

Además, proveen una alta seguridad y privacidad a nuestros datos, lo que nos facilita ese trabajo\*.

\* Existen servicios que nos permiten configurar el nivel de privacidad que queramos de ciertos datos, por ejemplo, archivos que necesitamos que puedan ser consultados desde cualquier dispositivo.

## ¿Por qué nos interesa el *Cloud Computing*?

### Otras aplicaciones

Además de lo antes mencionado, existen otras herramientas *cloud* que pueden ser de interés:

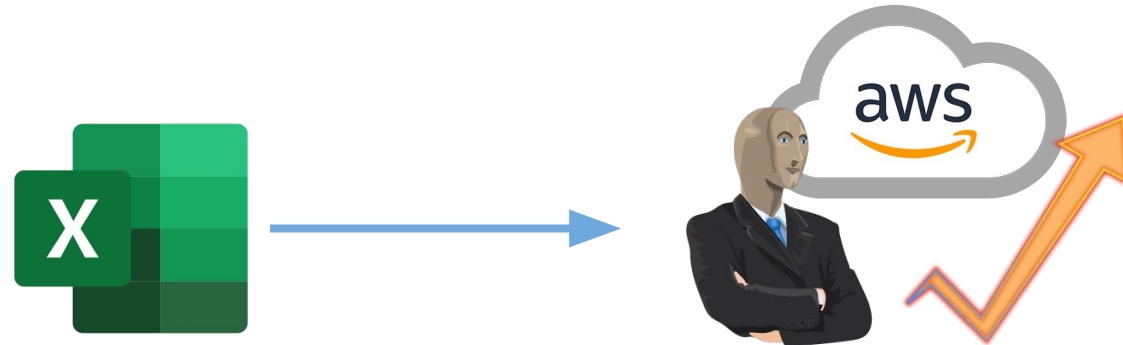
- *Machine Learning*
- Visualización de datos

Estos temas serán ahondados en otras unidades de este diplomado, pero es importante saber que son áreas abarcadas dentro de la nube.

## Panorama actual

Impacto de la nube en empresas pequeñas y medianas (Deloitte, 2014).

Caso particular: **Kimche**

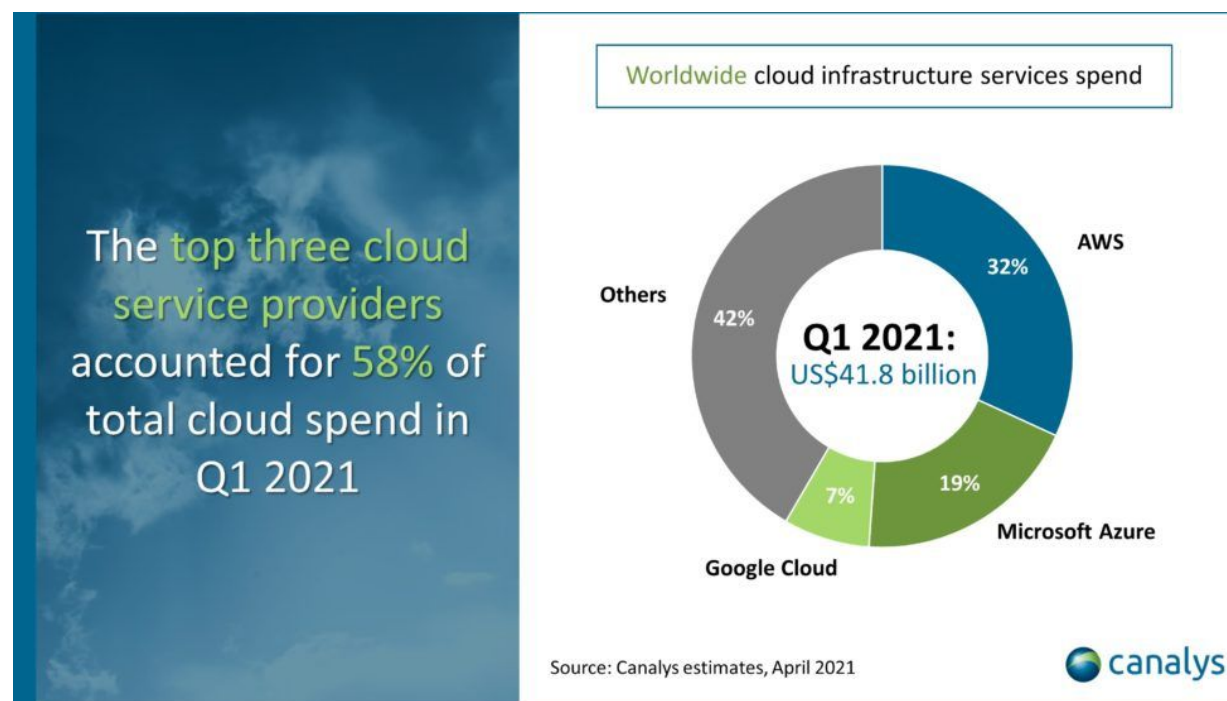




## Panorama actual

Total invertido en servicios de la nube al 2021 (Stalcup, 2021)

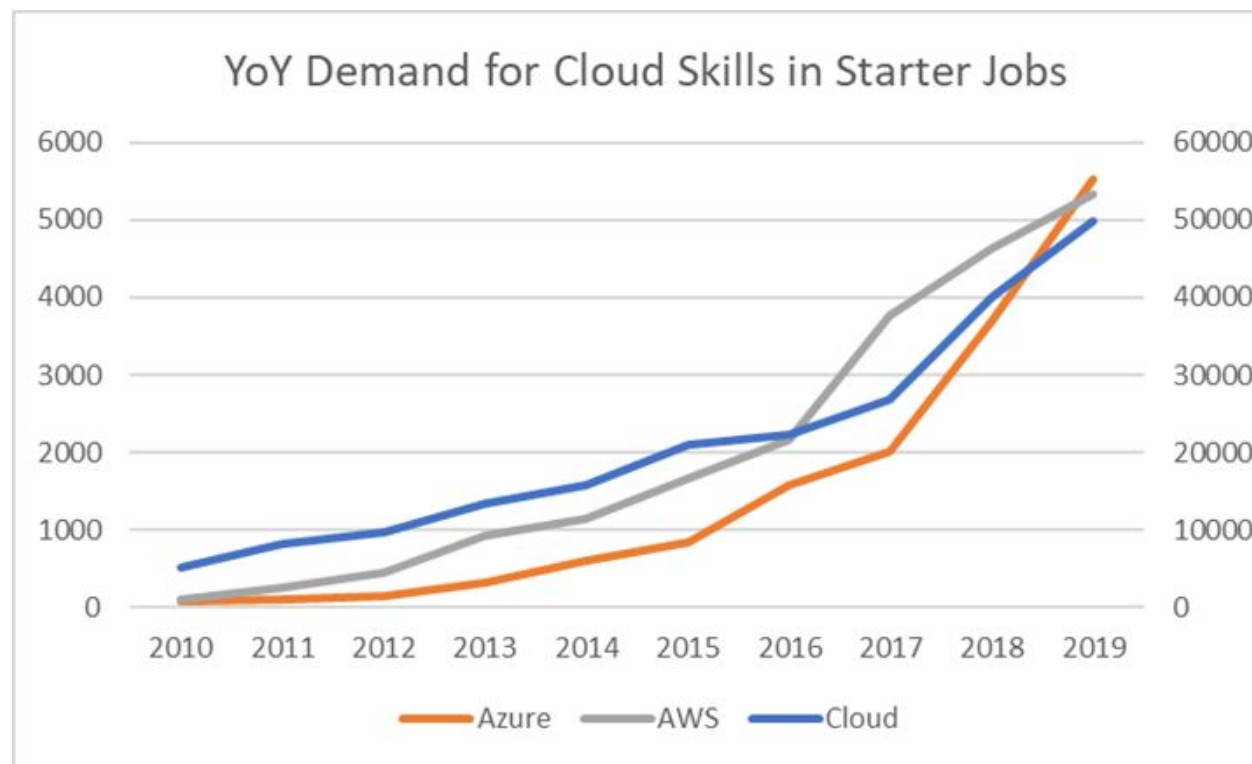
¡En Kimche utilizamos **Amazon Web Services (AWS)** y **Google Cloud Platform (GCP)**!



## Panorama actual

Crecimiento de la solicitud de habilidades en herramientas de la nube (Wortman, 2020).

¡Es un buen momento para aprender a hacer uso de estas herramientas!



## Herramientas

En el contexto de este curso, nos interesan las herramientas útiles con el **Big Data**, foco de este diplomado.

Se hará uso de herramientas de los proveedores Amazon (AWS) y Google Cloud Computing (GCP). El uso de estas les entregará la base para poder usar las de cualquier otro proveedor (Azure, por ejemplo).

¿Cuáles serán estas herramientas?

# Herramientas

- Almacenamiento de datos
  - Objetos: S3 (AWS); Cloud Storage (GCP)
  - Bases de datos relacionales: RDS (AWS); Cloud SQL (GCP)
- Ejecución de código
  - Funciones *serverless*: Lambda (AWS); Cloud Functions (GCP)
  - Paralelización: Simple Queue Service (AWS)
- Almacén de datos (Warehouse)
  - BigQuery (GCP)

## Herramientas - Almacenamiento de datos

### Almacenamiento de objetos

En muchas aplicaciones y empresas, es de interés poder almacenar datos en cualquier formato. Por ejemplo:

- Material audiovisual: Fotos de perfil, logos, grabaciones
- Código: Sitios web estáticos
- Otros: Credenciales

Si bien es una tarea sin mayor complejidad, es importante poder almacenarlos de forma **escalable**.

## Herramientas - Almacenamiento de datos

### Bases de datos relacionales

Los proveedores *cloud* proporcionan servicios de bases de datos para almacenar nuestra información. Podemos tener más de una base de datos y, a través de otras herramientas, podemos hacer que interactúen entre sí. Esto es válido también para bases de datos almacenadas en **nubes distintas**.

Si bien también existen servicios de bases de datos no relacionales, no serán el foco dentro de esta unidad.

## Herramientas - Ejecución de código

### Funciones *serverless*

Los proveedores *cloud* proporcionan herramientas que nos permiten ejecutar código *serverless* (sin servidor) a través de funciones. Esto significa que lo podemos correr **abstrayéndonos** del servidor en el que se ejecuta.

Podemos hacer que estas funciones se encarguen de interactuar con otros recursos de la nube, o bien, podemos hacer que funcionen como un *backend* que atienda distintas solicitudes (ejemplo: microservicios).

## Herramientas - Ejecución de código

### Paralelización

Existen ocasiones en las que nos interesa administrar una carga de solicitudes de forma eficiente, atendiendo varias de estas de forma simultánea.

En el contexto de esta unidad veremos Simple Queue Service de AWS, servicio de colas de mensajes que tienen como propósito resolver la situación antes descrita.



## Herramientas - Almacén de datos

Uno de los conceptos más importantes a tener en cuenta al momento de estudiar la Big Data es el ***data warehousing***.

Un *data warehouse* consiste en un “almacén de datos”, cuyo fin particular es almacenar los datos procesados de una empresa en un formato que permita su análisis y posterior toma de decisiones.

Para conocer un poco más cómo funciona, haremos uso de la herramienta BigQuery de GCP.

## Un caso real - Monitoreo de Google Classroom

En el inicio de la pandemia, Kimche estuvo a punto de perder gran parte de su clientela por la falta de clases presenciales (gran fuente de sus datos hasta ese momento).

Para poder seguir siendo una empresa competitiva, fue necesario adaptarse a los tiempos y ofrecer un servicio relativo a las clases *online*: así nace el monitoreo de Google Classroom.

## Un caso real - Monitoreo de Google Classroom

**Idea:** Entregar a colegios reportes de asistencia y desarrollo de tareas realizadas en Google Classroom para **mejorar la toma de decisiones** en lo que respecta a las clases *online*.

**Ejecución:** Uso de las API de Google para obtener los datos y llevar a cabo reportes de visualización que resuman el estado actual de la escuela.

## Un caso real - Monitoreo de Google Classroom

### Problemas

- Almacenamiento de credenciales de las cuentas de las escuelas para el uso de la API
- Llamados simultáneos a la API; cuotas de uso (QPS - *Queries per Second*)
- Formato de datos para la visualización (nivel colegio, nivel sostenedor)

## Un caso real - Monitoreo de Google Classroom

### Primer acercamiento sin apoyo de herramientas de la nube

- *Scripts* en Python que hacían llamados a la API.
- Subida manual de archivos obtenidos a servidor.
- Límite de archivo a subir (10 MB); segmentación “automatizada” de los archivos (y subidas por parte).

Cuando más de 10 colegios contrataron el servicio, esta dinámica ya no era escalable.

## Un caso real - Monitoreo de Google Classroom

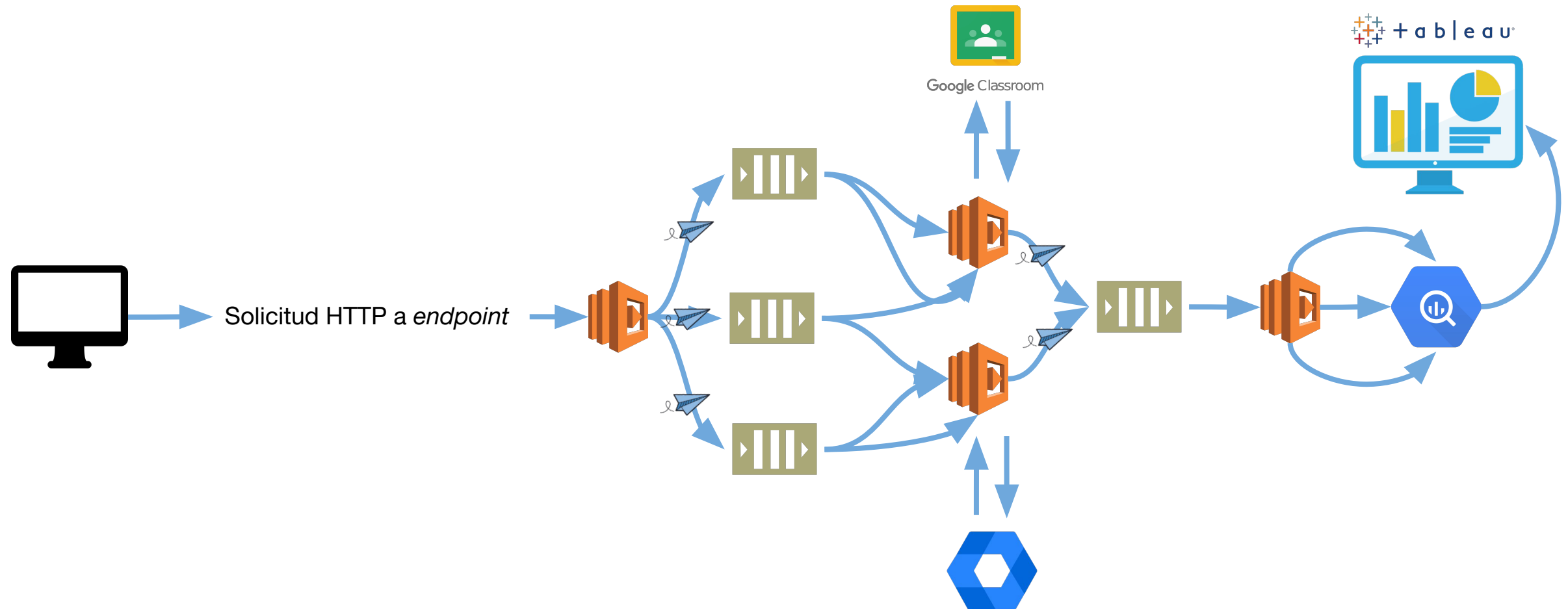
### Segundo acercamiento usando herramientas de la nube

- *Login* y almacenamiento de credenciales automatizado, usando AWS RDS, Lambda y S3.
- Distribución de carga y paralelización usando AWS SQS.
- Inserción de datos automatizada usando AWS Lambda y Google BigQuery

Actualmente más de 150 colegios usan el servicio (¡incluso en !)

# Un caso real - Monitoreo de Google Classroom

## Esquema general

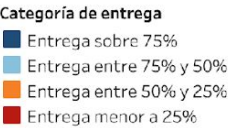


# Un caso real - Monitoreo de Google Classroom

## Ejemplo de visualización de los datos con Tableau

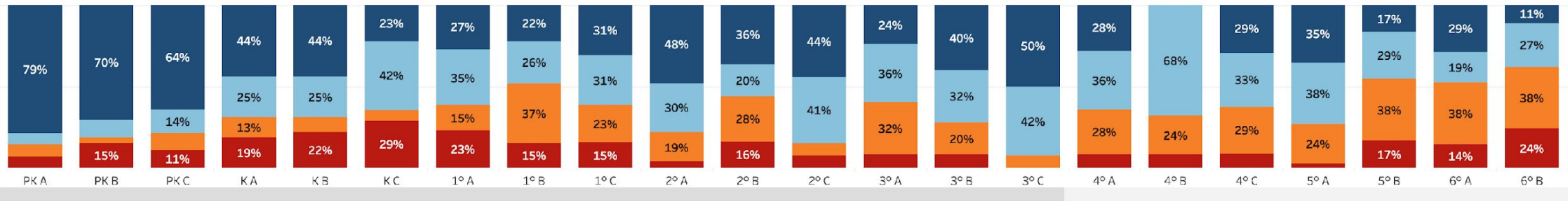
### Casos de riesgo

La siguiente visualización muestra la distribución por curso de la entrega de tareas en Classroom. Permite identificar a aquellos estudiantes que tienen un bajo porcentaje de entrega respecto a sus tareas asignadas.



### Distribución de entregas por curso

Haz click en una distribución para ver a los estudiantes que la componen.





## Un caso real - Monitoreo de Google Classroom

### En resumen...

- Las herramientas de la nube nos permitieron escalar.
- Las herramientas de la nube nos permitieron expandir el negocio.
- Las herramientas de la nube nos mantuvieron a flote en un contexto de incertidumbre.

¡Y dentro del curso veremos la mayoría de las herramientas involucradas en esta historia!

## Actividad

Antes de terminar la clase, haremos una pequeña actividad práctica para empezar a familiarizarnos con las herramientas.

Pueden acceder en el siguiente enlace:  
<https://colab.research.google.com/drive/1bgih7Zuj7H7F9uCrJVqvSLB4GyOvjQYK?usp=sharing>

## Cierre

Antes de cerrar esta primera clase, les pido que respondan esta breve encuesta: <https://forms.gle/ZHMHlQqRvWKQRp3e7>

Los que lo deseen, pueden comentarme a mí y a sus compañeros/as sobre las preguntas de la encuesta.

Si no quieren, está bien. ¡Es opcional!

Servirá para ver cómo vamos armando el resto del curso. 😊

## Referencias

Dar, A. (2018). Cloud Computing-Positive Impacts and Challenges in Business Perspective. *Journal of Computer Science & Systems Biology*, 12. Recuperado de <https://www.hilarispublisher.com/open-access/cloud-computingpositive-impacts-and-challenges-in-business-perspective-jcsb-1000294.pdf>

Deloitte (2014). *Small business, big technology. How the cloud enables rapid growth in SMBs.* Recuperado de <https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Technology-Media-Telecommunications/gx-tmt-small-business-big-technology.pdf>

## Referencias

Stalcup, K. (2021). *AWS vs Azure vs Google Cloud Market Share 2021: What the Latest Data Shows*. ParkMyCloud. Recuperado de <https://www.parkmycloud.com/blog/aws-vs-azure-vs-google-cloud-market-share/>

Wortman, R. (2020). *Azure skills and certifications can boost grads' salary potential*. Microsoft. Recuperado de <https://techcommunity.microsoft.com/t5/microsoft-learn-blog/azure-skills-and-certifications-can-boost-grads-salary-potential/ba-p/1820642>

## Bibliografía

Geewax, J. (2018). *Google Cloud Platform in Action*. Manning.

Wittig, A., & Wittig, M. (2019). *Amazon Web Services in Action, Second Edition*. Manning.



**EDUCACIÓN**  
**PROFESIONAL**

**Unidad 2:**

# **Introducción a las Herramientas de la Nube**

**Clase 1 - Introducción al curso**

**Profesor: Germán Leandro Contreras Sagredo**