

SOFTWARE GURU & PROPIEDADES.COM

Aventuras en el procesamiento de datos

BY ANDRÉS BRAVO



HOLA! YO SOY

ANDRÉS BRAVO

HEAD OF DATA @PROPIEDADES.COM



@BitsHoroscope



abravo-m

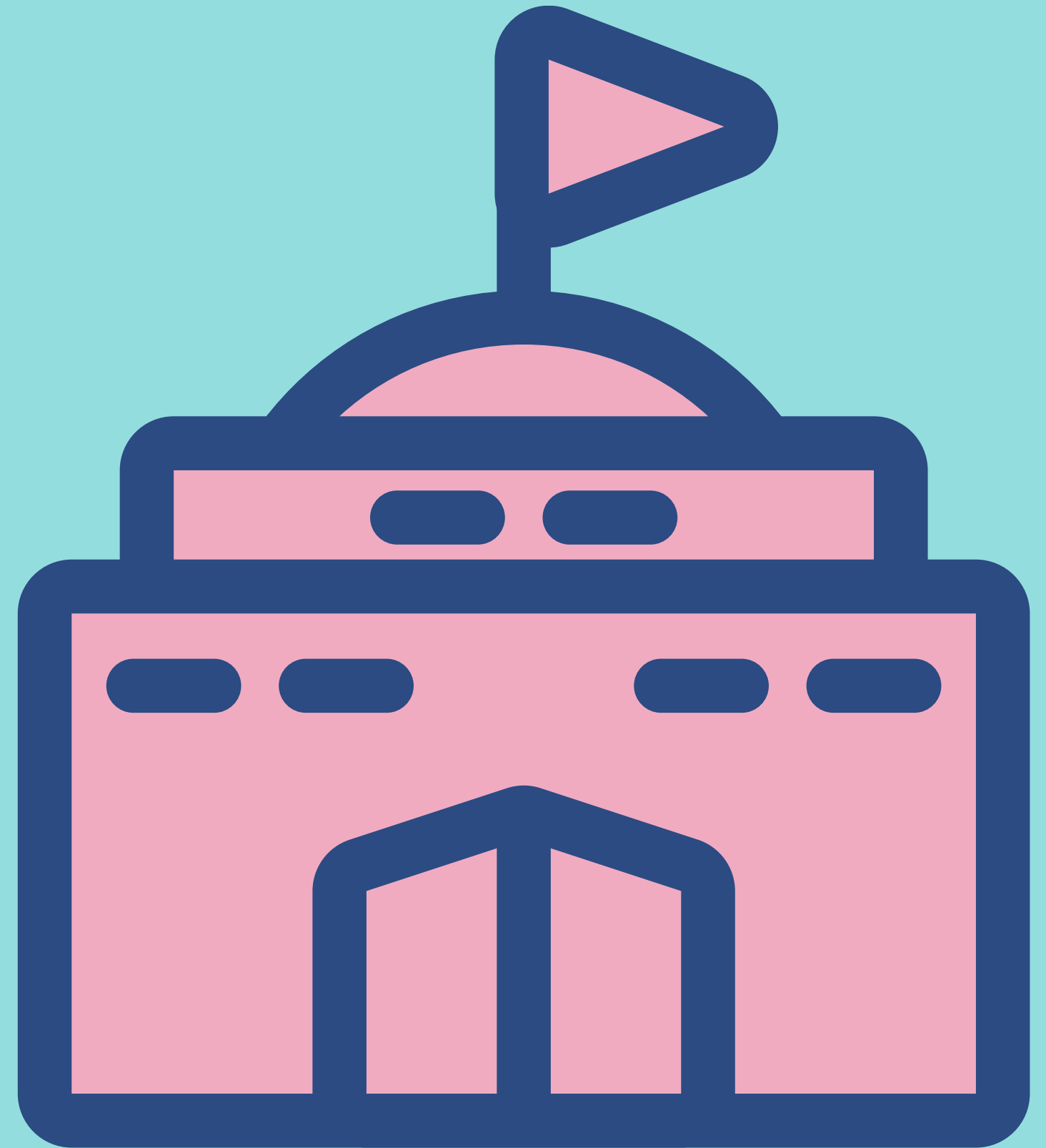


andybravo.dev



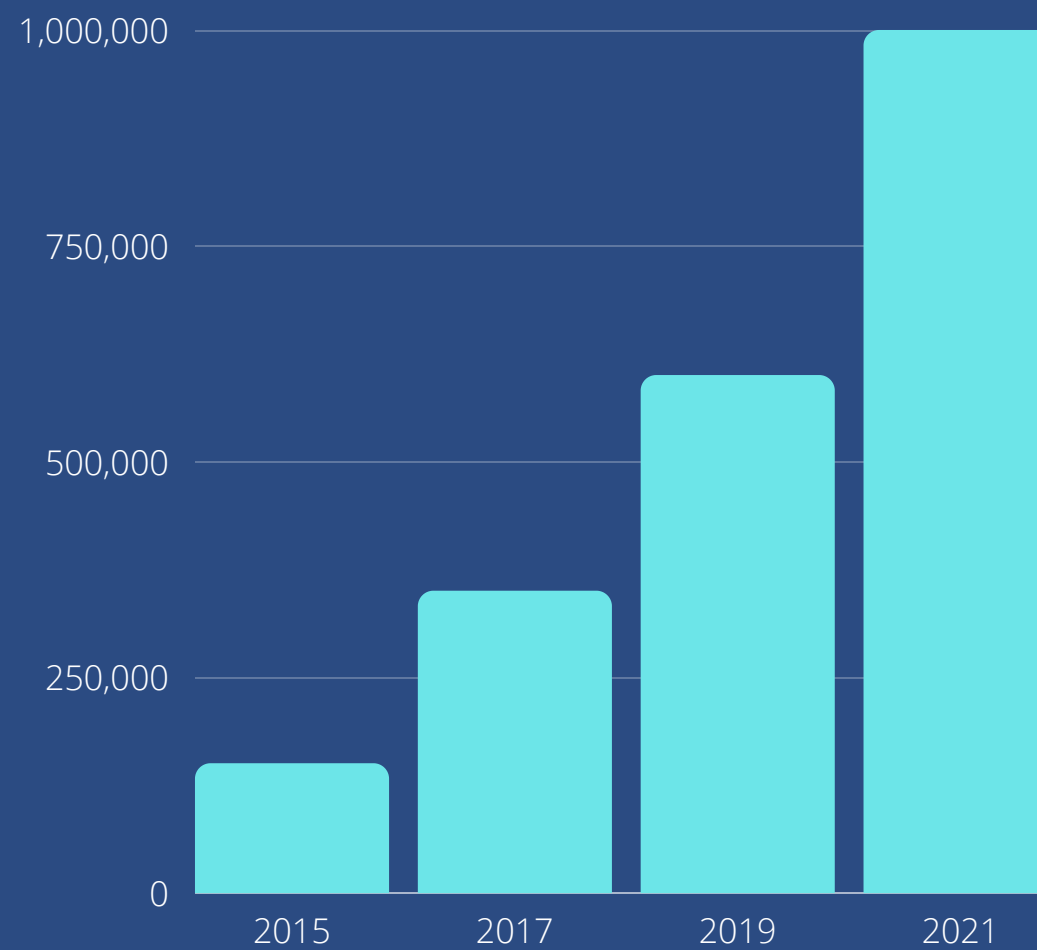
Nuestro propósito

Porque todo empieza con el porqué

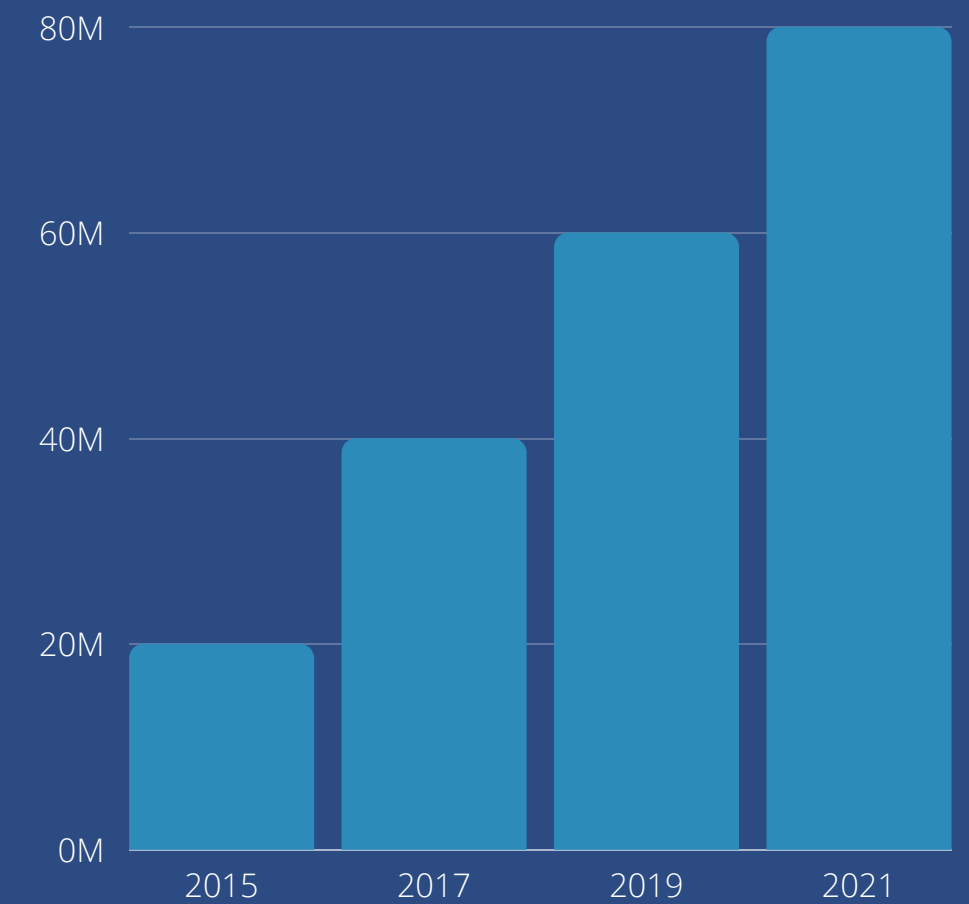


¿Cuál es nuestra oferta?

MAYO 2021



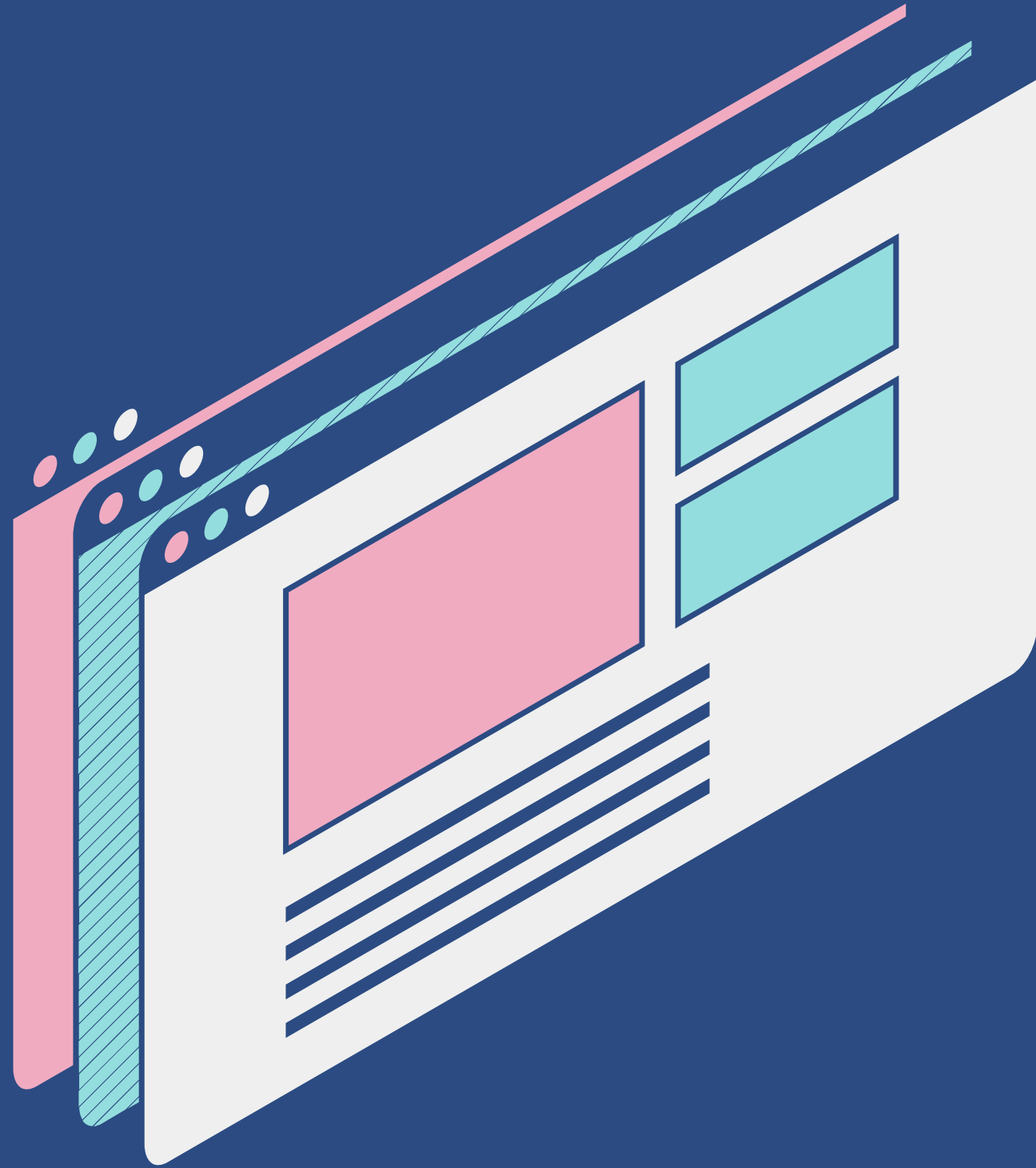
Anuncios Listados



Imágenes procesadas



25,000 usuarios por hora



¿Cómo
plantear una
arquitectura?

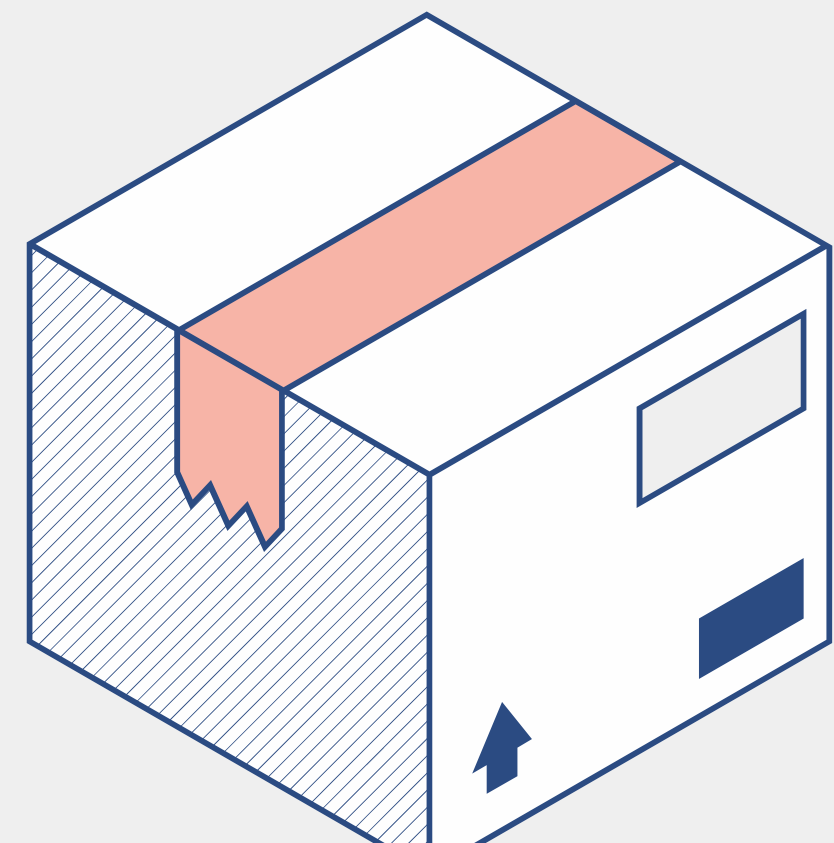
Cargar datos desde el portal

- Carga desde formulario
- Expuesto al usuario



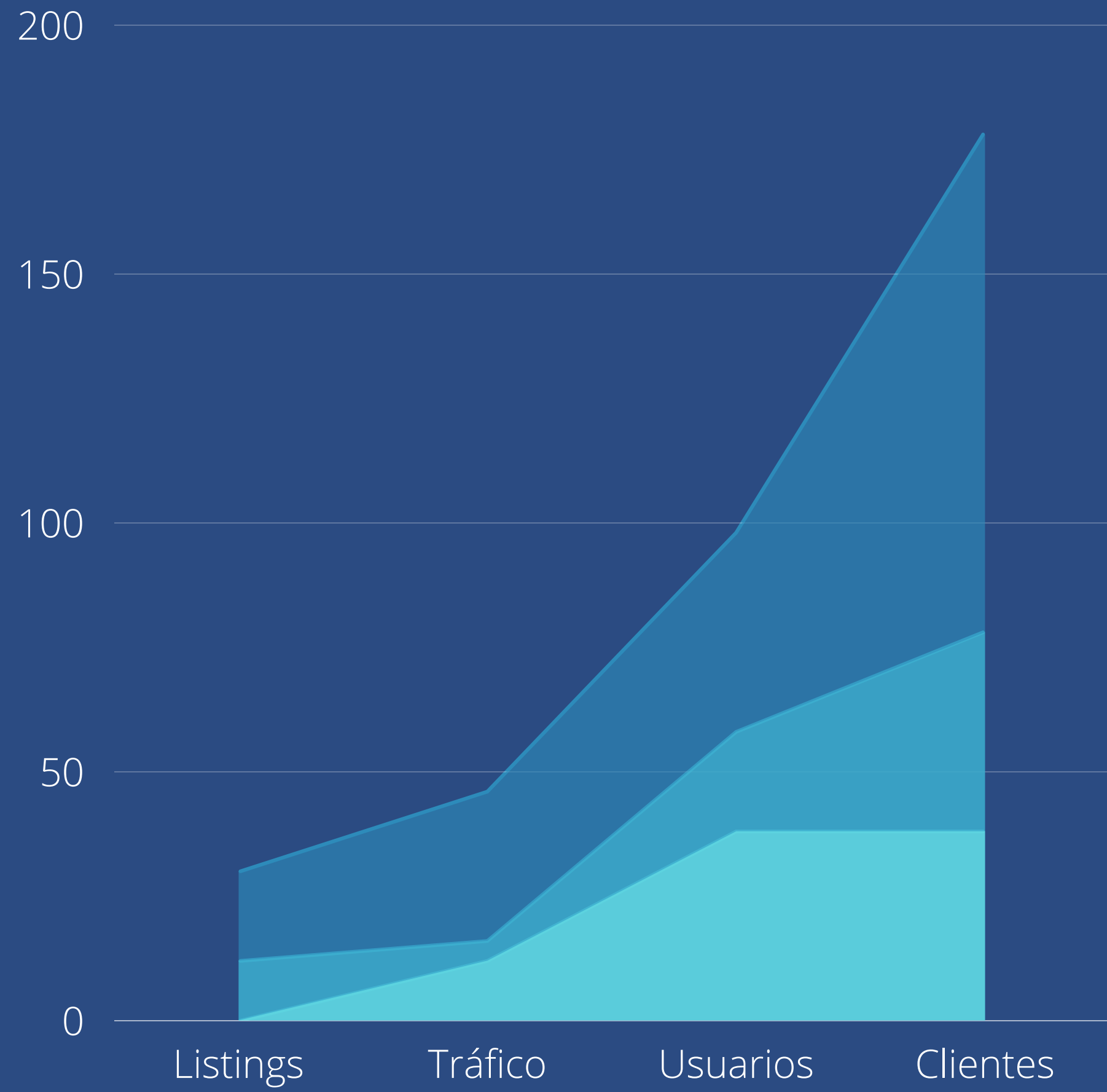
Cargar datos de forma masiva

- Procesamiento para usuarios desde 50 anuncios.
- Procesamiento Batch
- Transferencia con XML



Volvamos al 2013

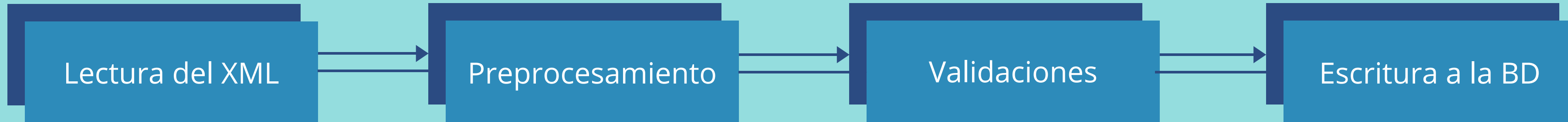




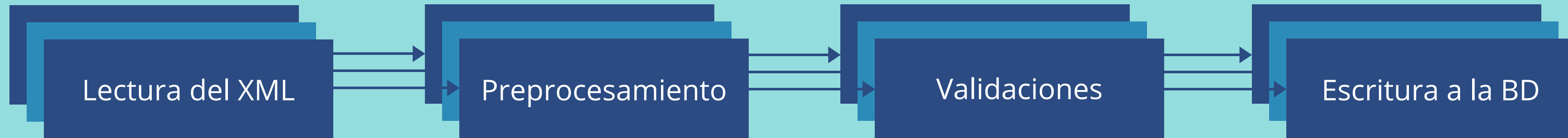
Flujo simplificado del Pipeline



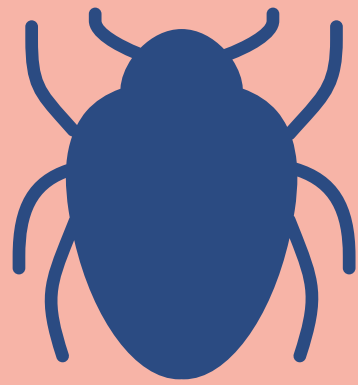
Flujo simplificado del Pipeline



Flujo simplificado del Pipeline



Problemas asociados



**Mucha labor manual
para corregir la
operación**

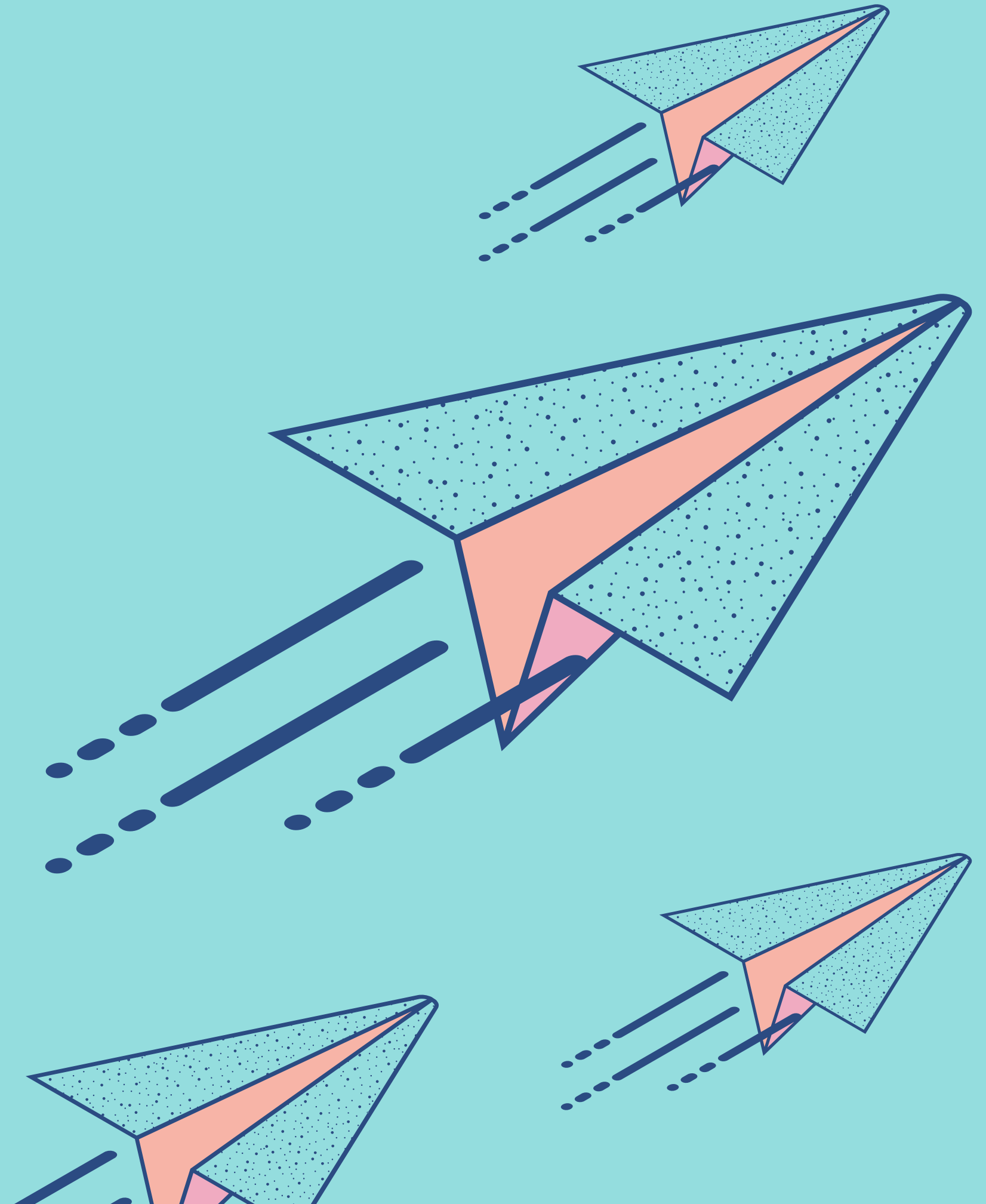


**Problemas de
infraestructura**



**Disminución en el tiempo
de desarrollo**

Ingeniería de datos al rescate



ESTRATEGIA



DATOS

Generar un sistema concurrente de procesamiento masivo de datos



OPERACIÓN

Disminuir el tiempo operativo para acelerar el proceso de desarrollo



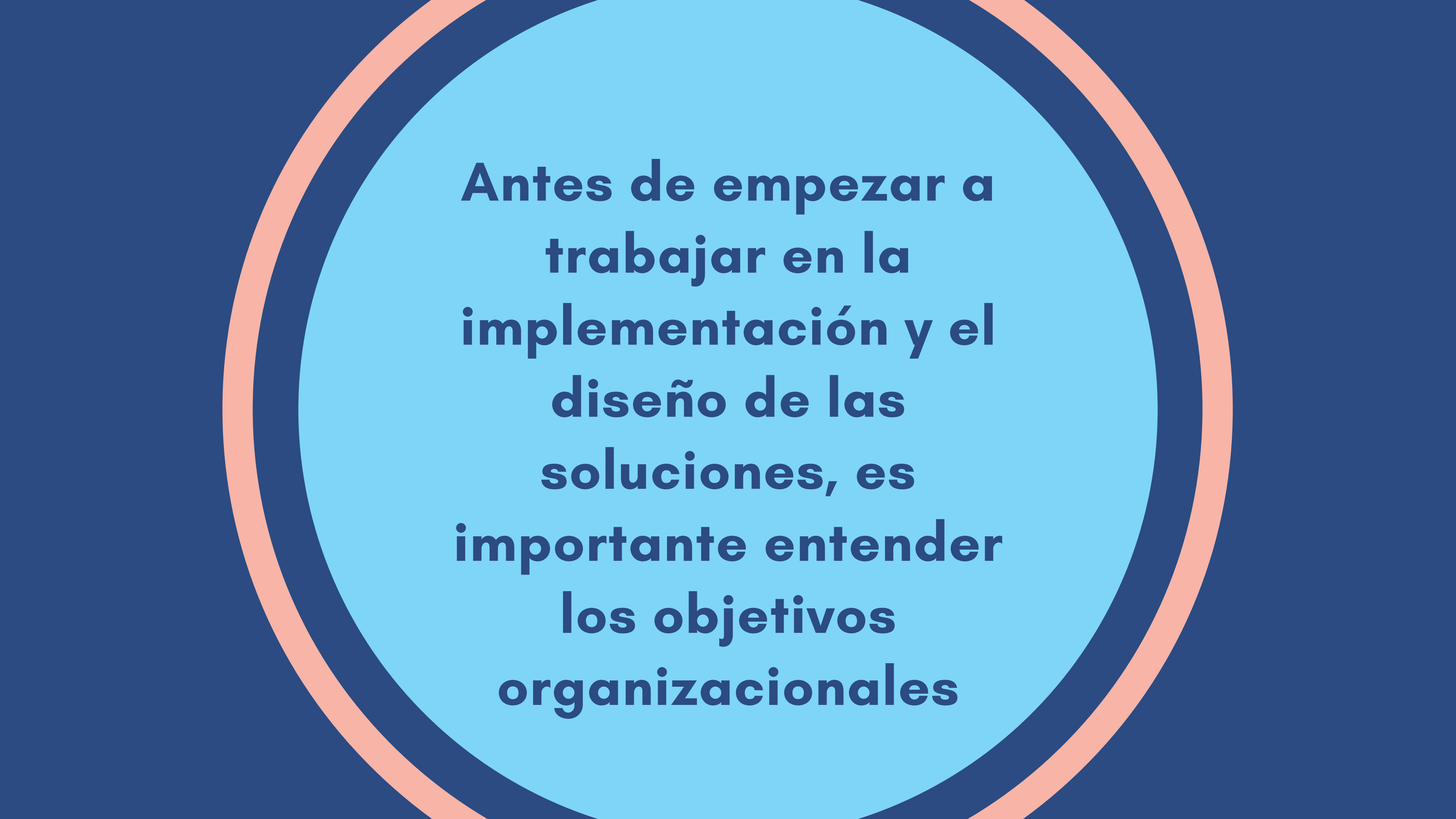
INFRAESTRUCTURA

Montar infraestructura elástica capaz de soportar la operación

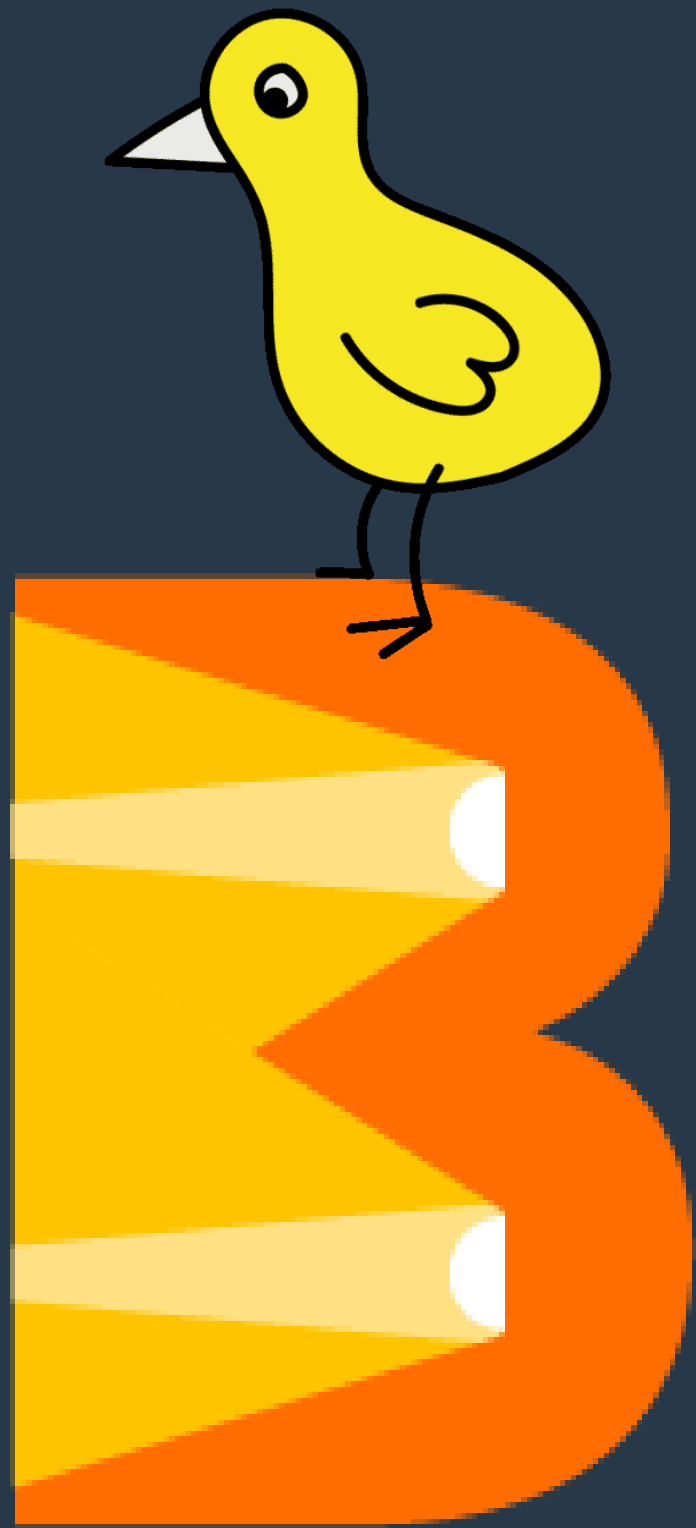


INCIDENCIAS

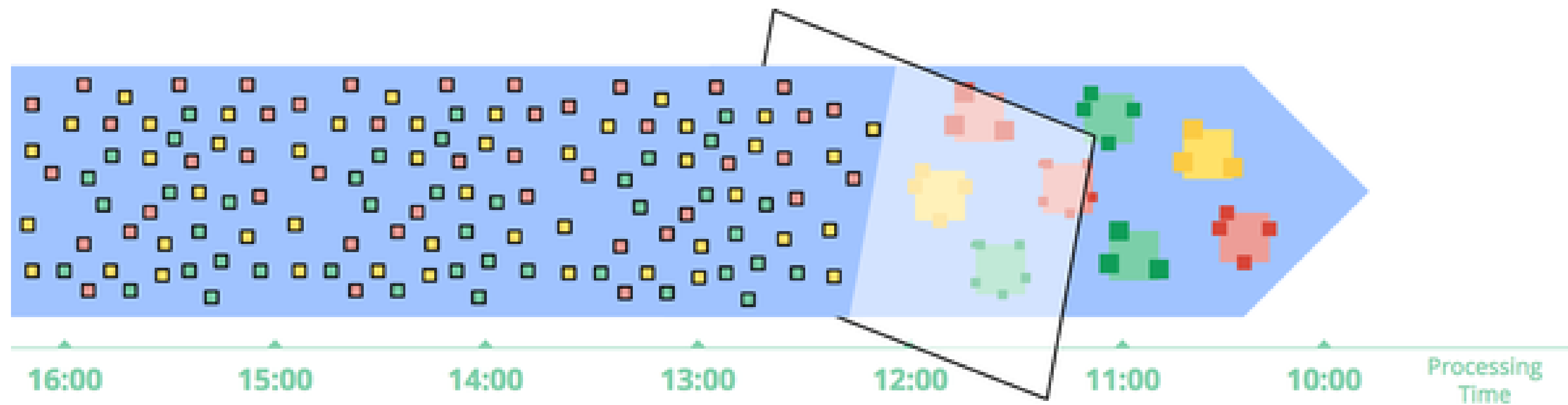
Reducir el tiempo de recuperación de un incidente



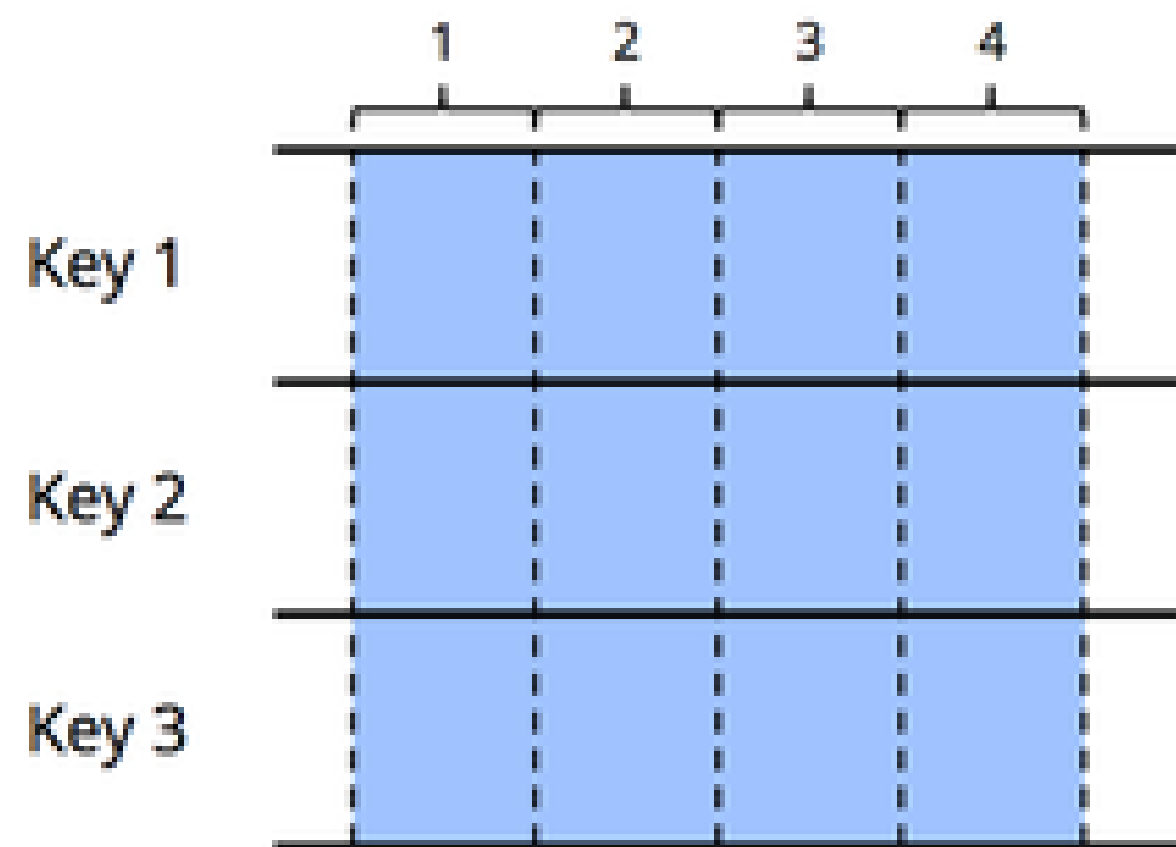
**Antes de empezar a
trabajar en la
implementación y el
diseño de las
soluciones, es
importante entender
los objetivos
organizacionales**



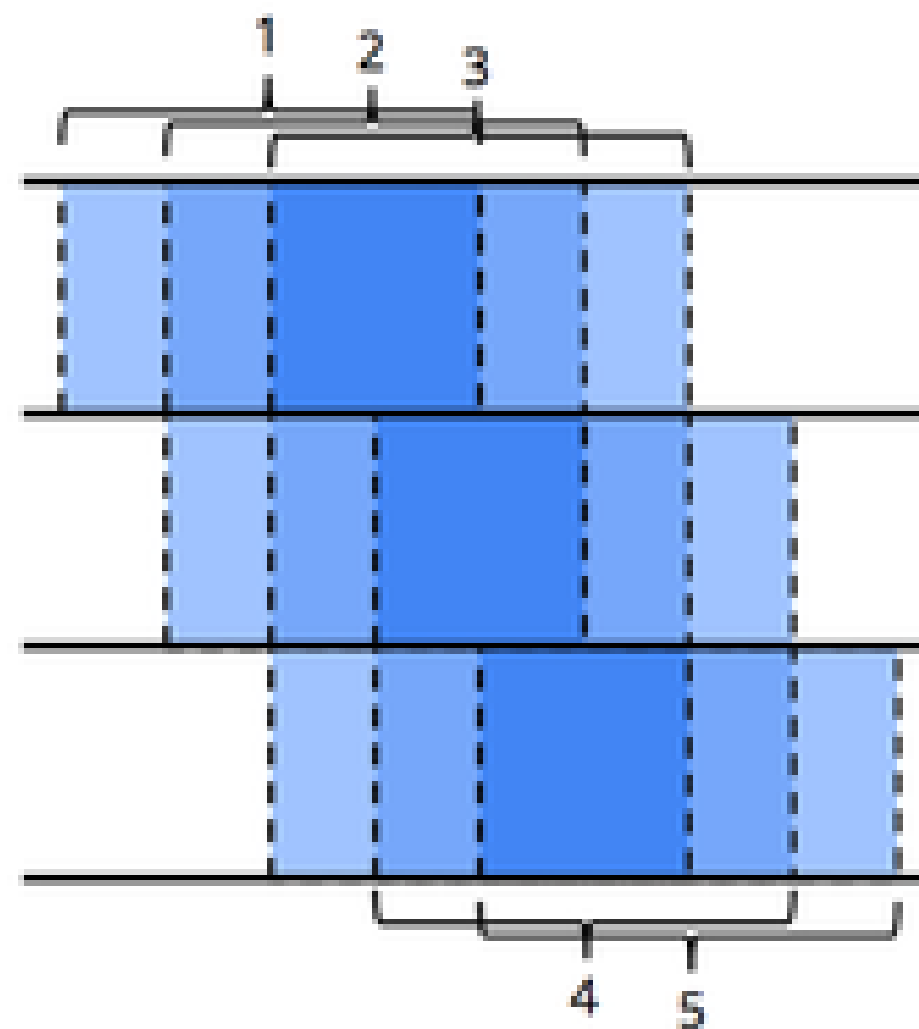
beam



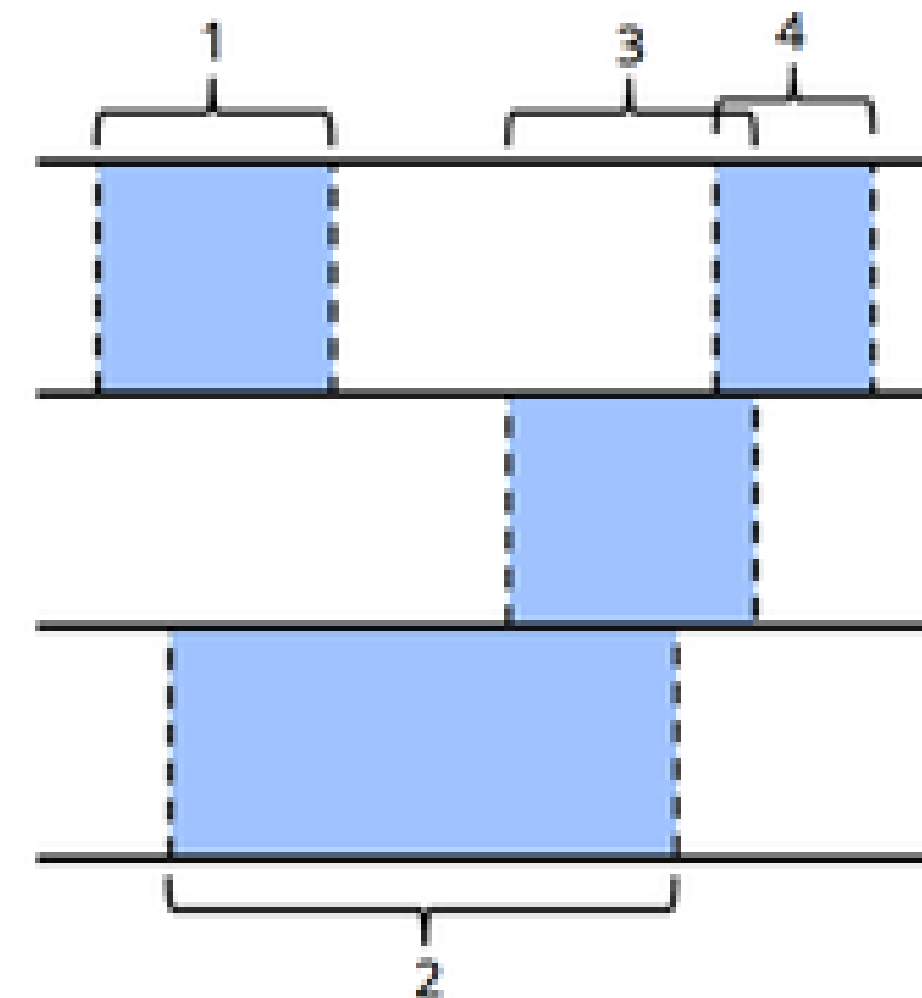
Fixed



Sliding



Sessions



Time

Pipeline 1.0



1

2

3

4

5

STEP

STEP

STEP

STEP

STEP

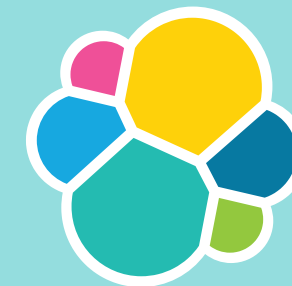
Unificación de
fuentes de
entrada

Verificación y
consistencia de
la información

Conciliación del
inventario

Propiedades nuevas,
actualizaciones y
eliminaciones

Localización

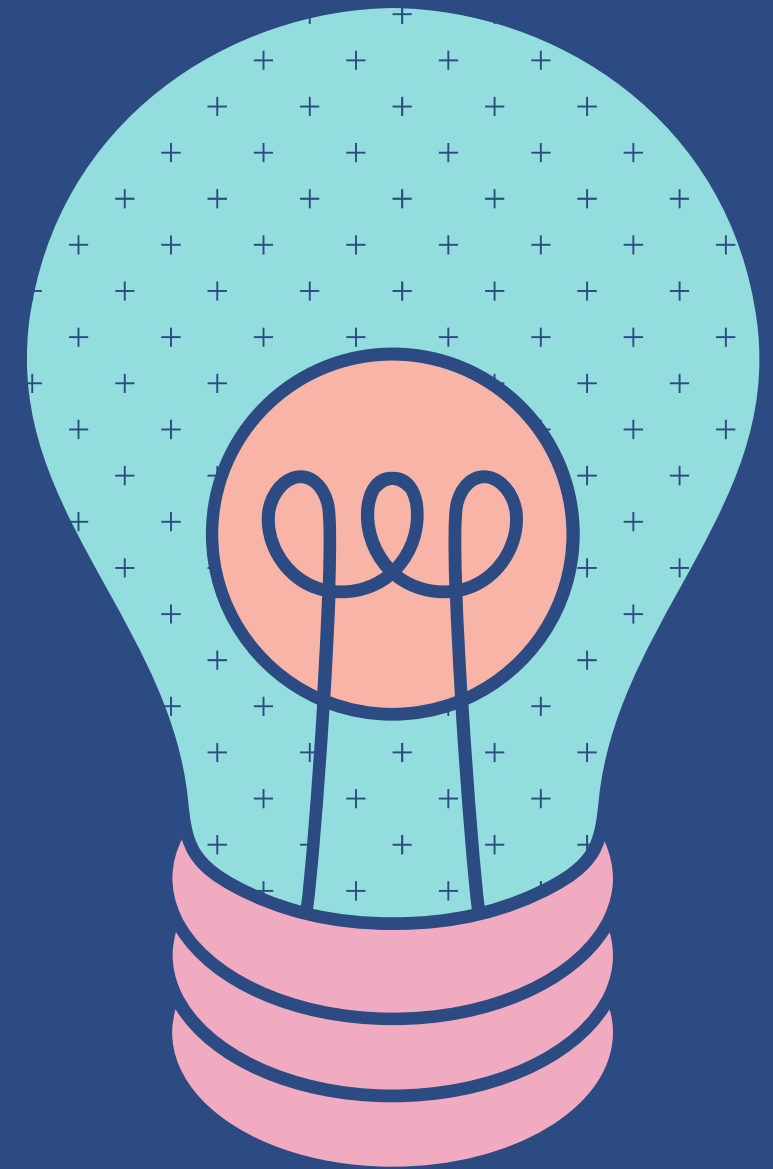


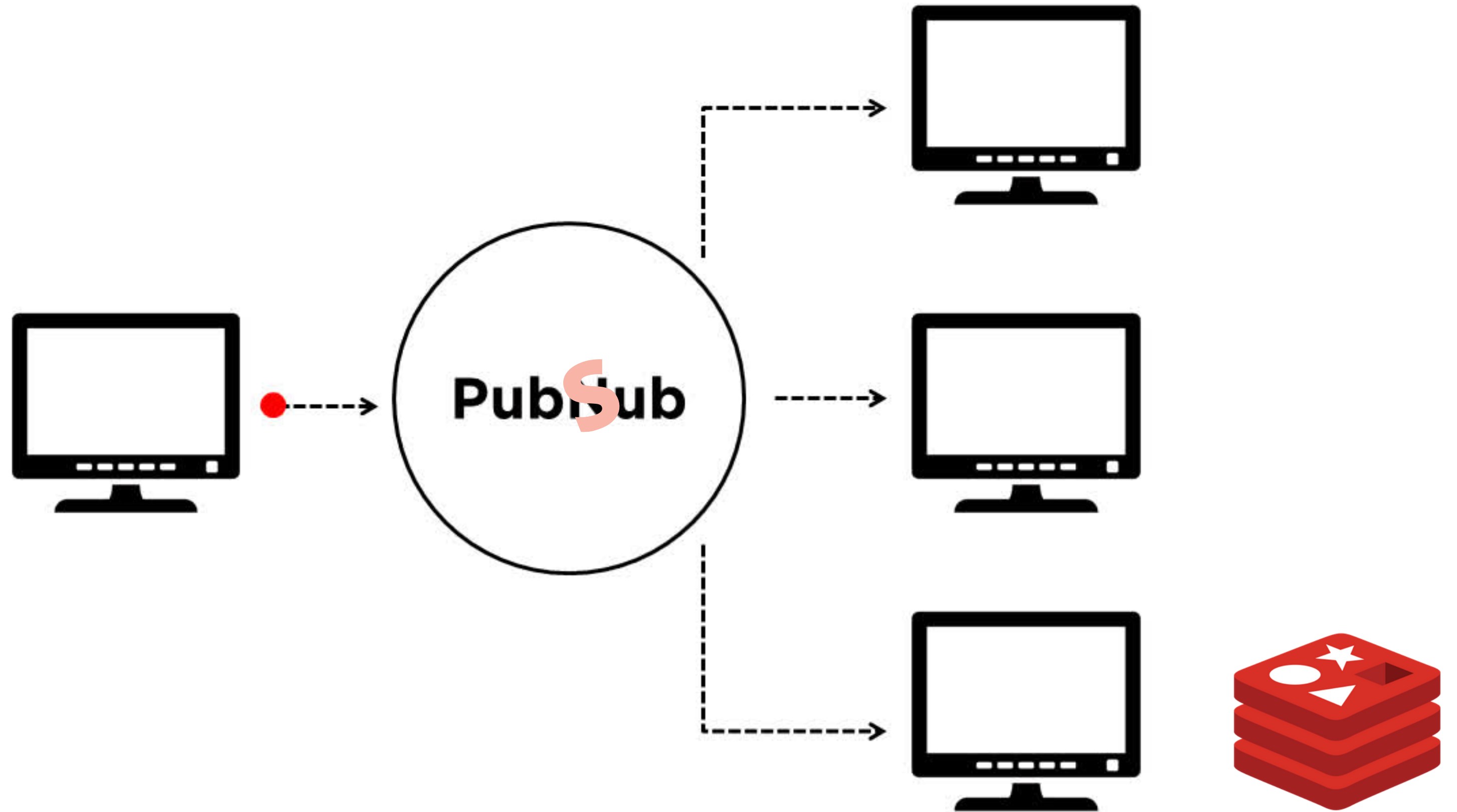
Procesamiento
de imágenes

Catálogo de
rechazos



PATRONES DE DISEÑO





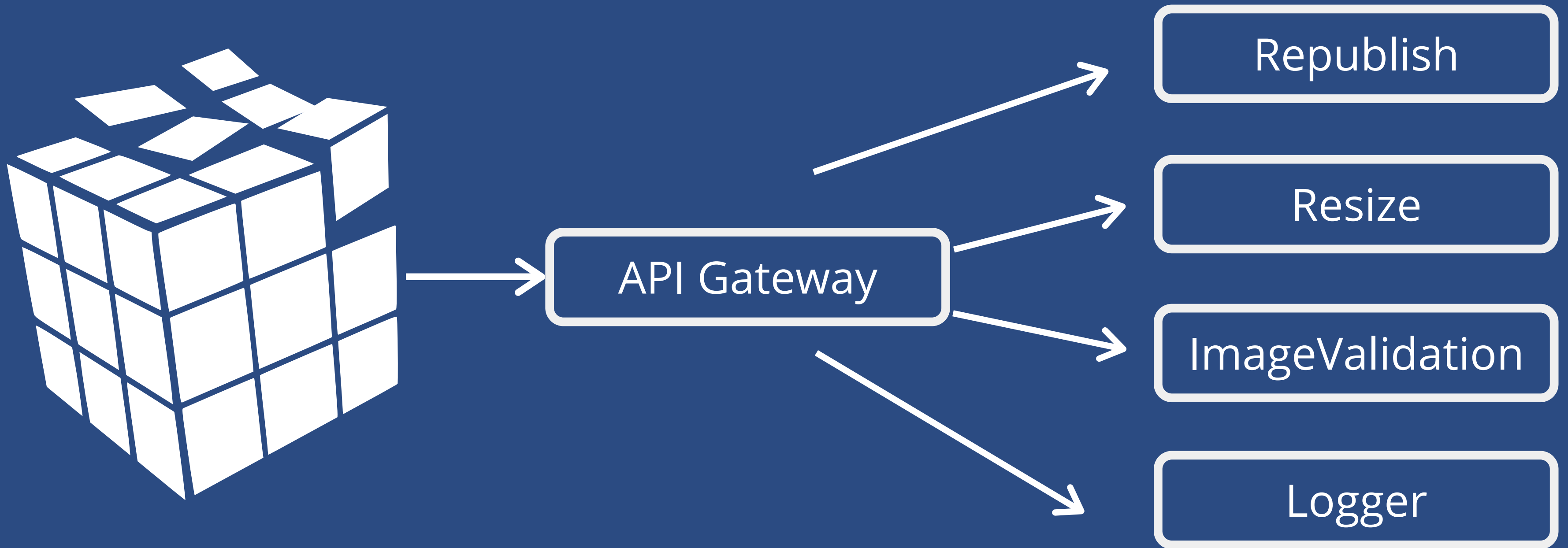


Inventario actualizado cada 24 hrs

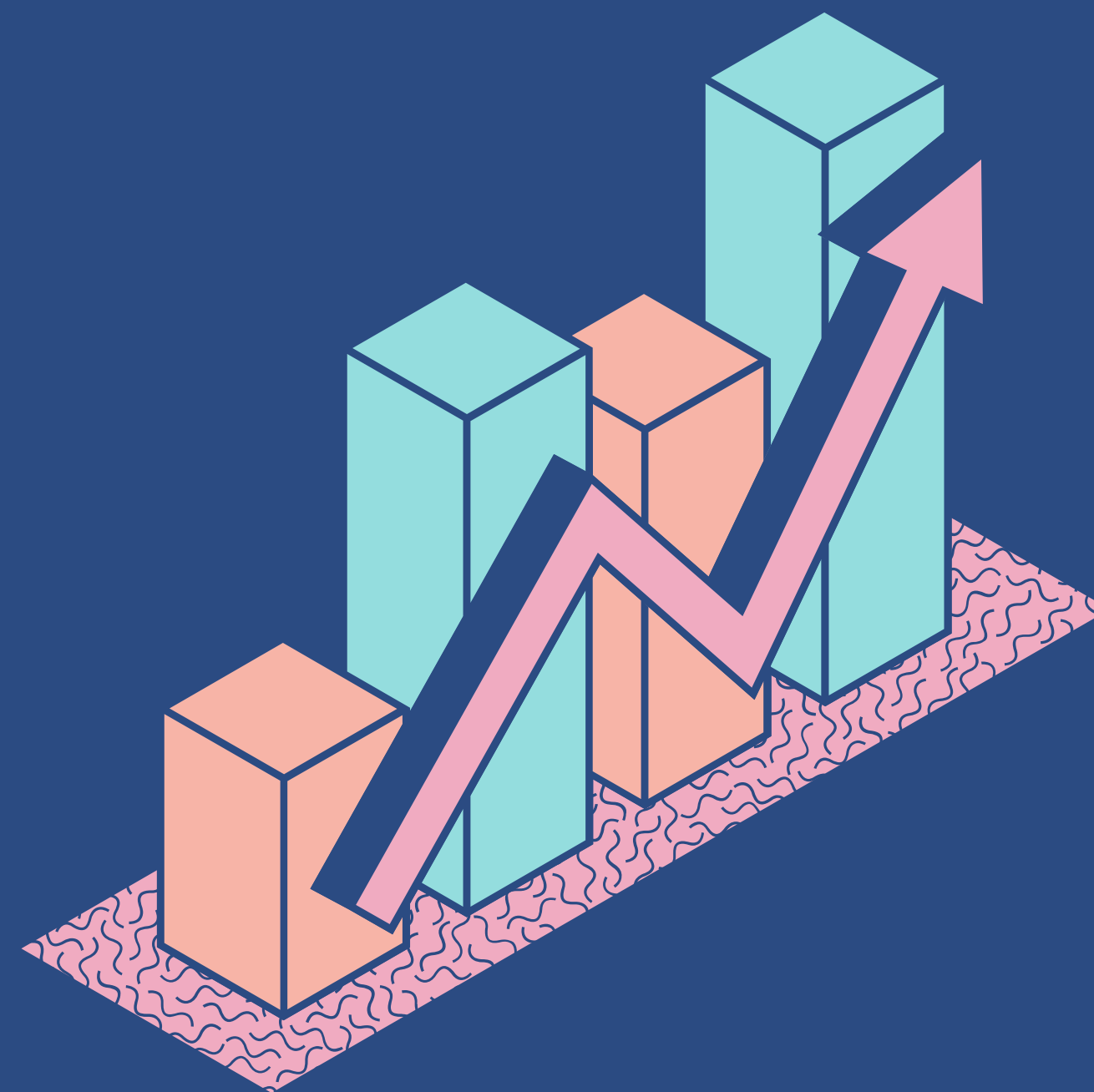
MÁS TIEMPO PARA DESARROLLAR Y
MENOS PARA DEPURAR

bueno, más o menos...

El lado oscuro...
la depuración!

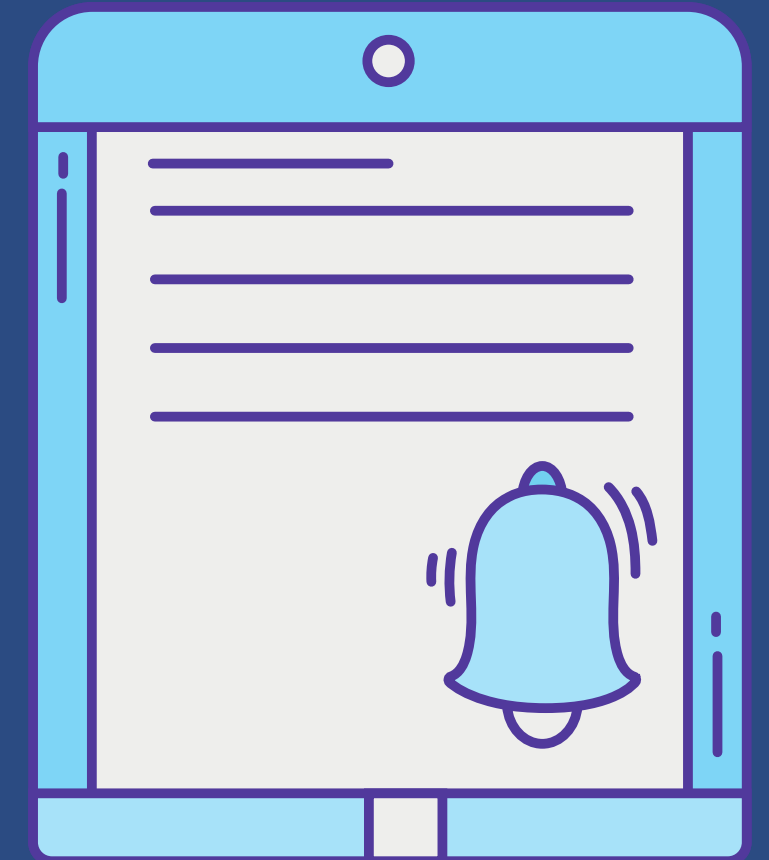


Selección de la infraestructura

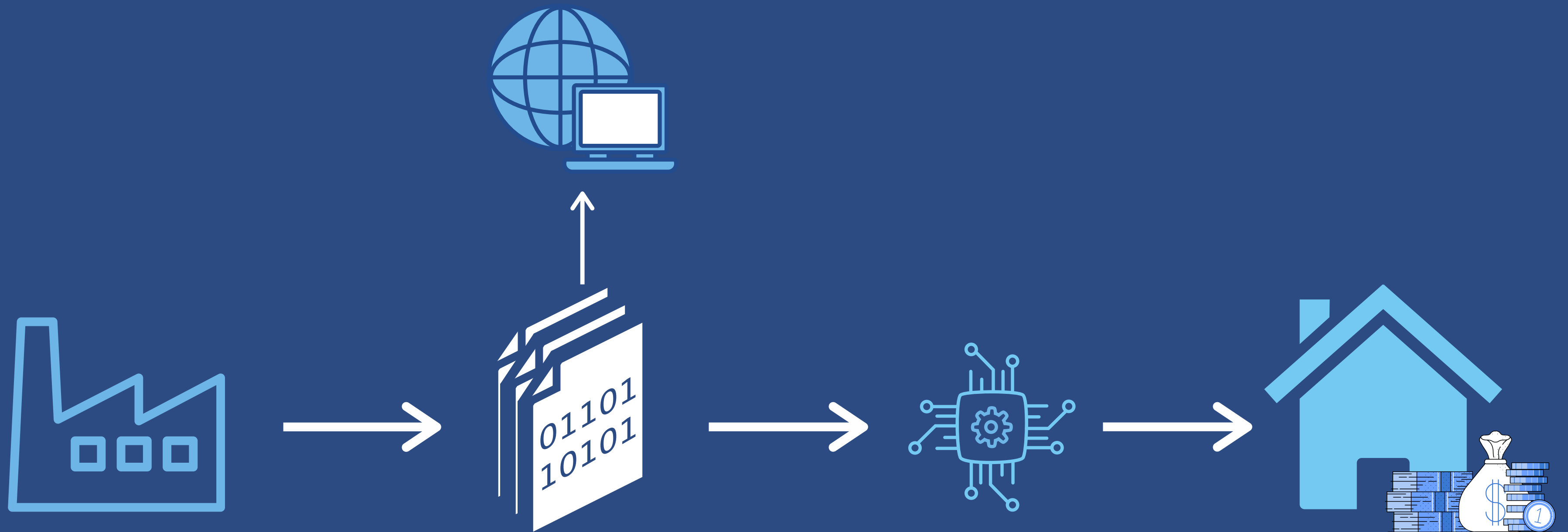




kubernetes



Machine Learning





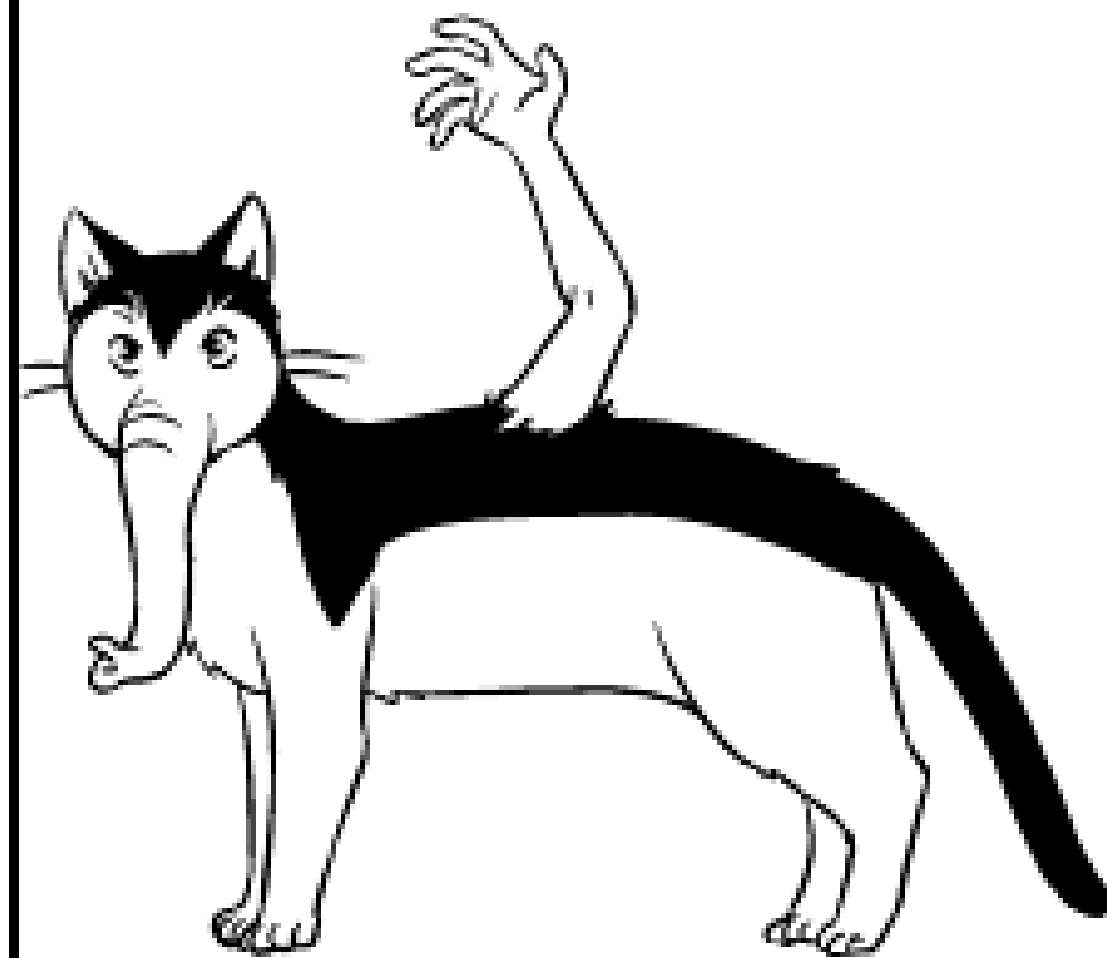
La arquitectura de
software es un
organismo vivo y en
constante evolución



HOW THE SOFTWARE
IS DESIGNED.



HOW THE SOFTWARE
IS DESIGNED.



HOW THE SOFTWARE LOOKS
TWO VERSIONS LATER.



TIEMPO DE LA CONVERSACIÓN



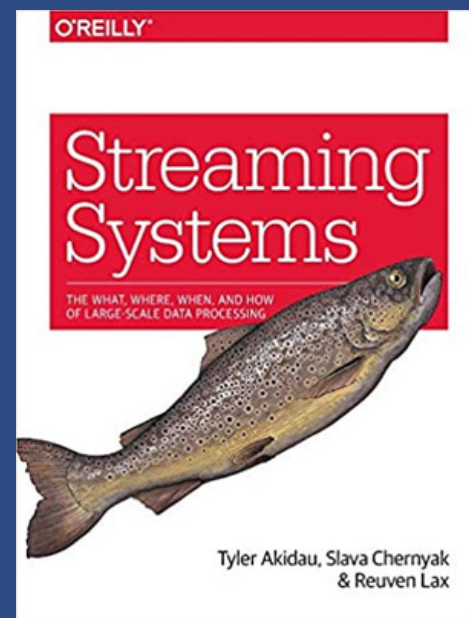
@BitsHoroscope



abravo-m

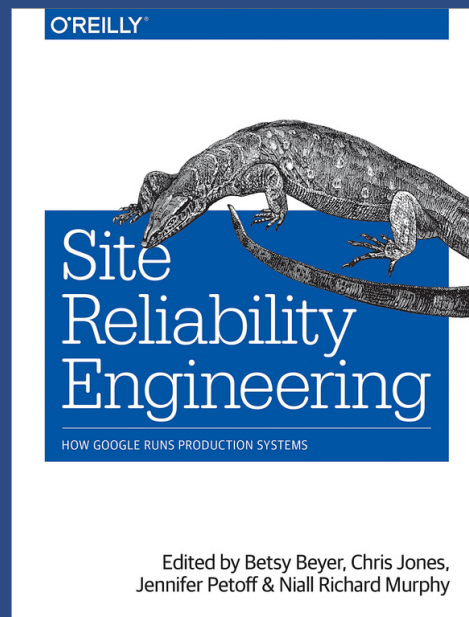
Recursos

PORQUE EL
CONOCIMIENTO
CRECE CUANDO SE
COMPARTE



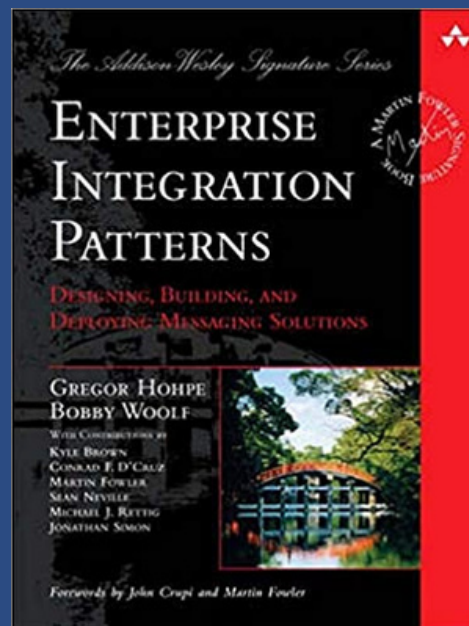
Streaming Systems - Akidau, Chernyak, Lax

<https://www.oreilly.com/library/view/streaming-systems/9781491983867/ch04.html>



Site Reliability Engineering - Bayer, Jones, Petoff

<https://sre.google/workbook/table-of-contents/>



Enterprise Integration Patterns Messaging Solutions - Hohpe, Woolf

La última y nos vamos