



ETL

Extract, Transform and Load



**Por
Francisco Contreras**

Francisco Contreras



Miembro del core del JUG Nicaragua

Master en informática empresarial

Entusiasta de las tecnologías

**Más de 10 años trabajando en Desarrollo de software para
empresas a nivel mundial**

Twitter
@Frank_JCG

LinkedIn
/in/fjcontrerasg/

Temas a abordar

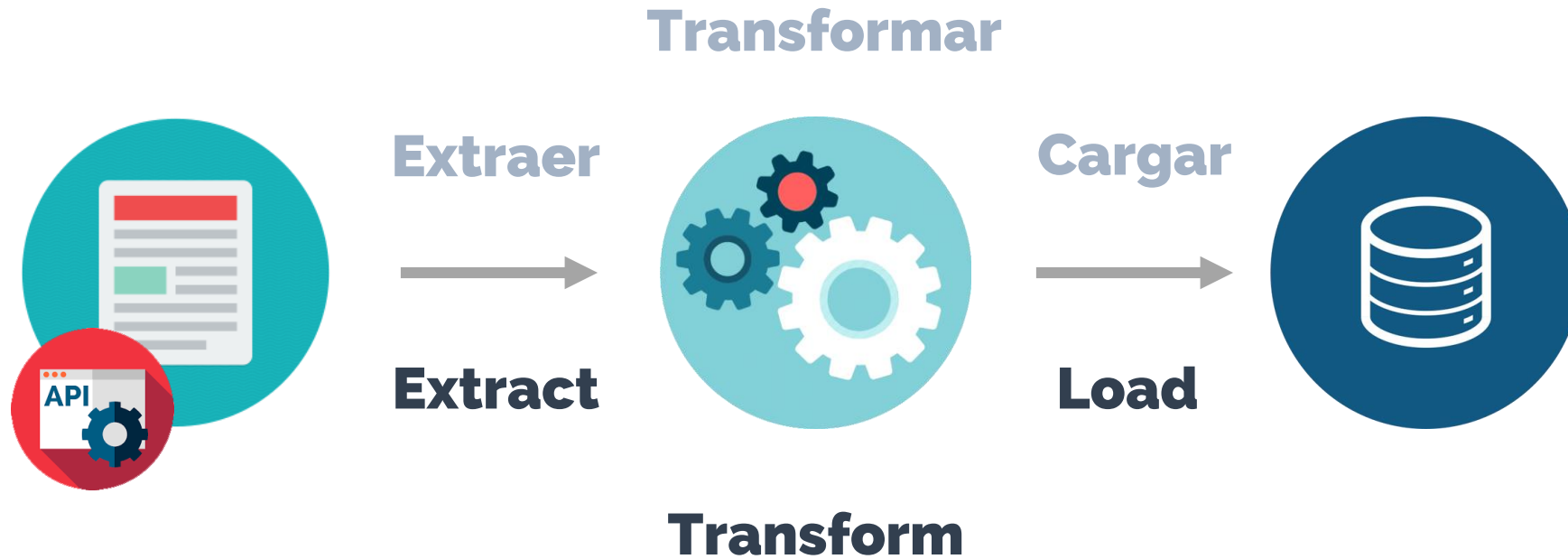


¿Qué es ETL?

Buenas practicas en el diseño de ETL

Problemas comunes

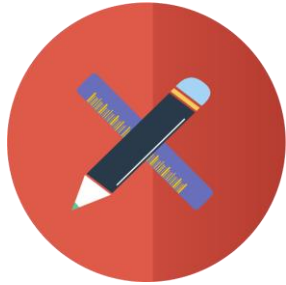
¿Qué es ETL?



¿Cómo diseñar un sistema ETL?



**Recolección de
requerimientos**



**Selección de
herramientas**



**Implementación y
revisión**



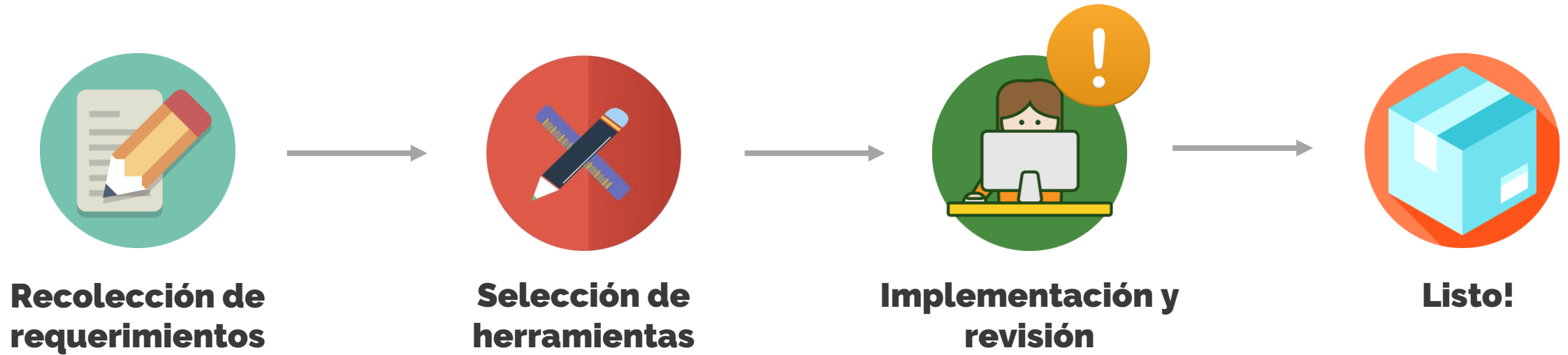
Listo!

¿Cómo diseñar un sistema ETL?



**Cada sistema es diferente, por lo tanto, no existen
formulas genericas para el diseño de sistemas
específicos**

¿Cómo diseñar un sistema ETL?



Recolección de requerimientos



Funcionales

Datos, calculos, destinos



Calidad de los datos

Usabilidad según el origen



Regulaciones de ley

Sobre el uso y almacenamiento de la información



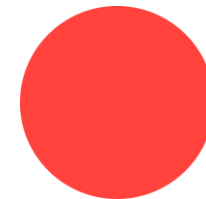
Retención

Para auditorías o revisiones internas



Seguridad

Respaldos, redundancia, etapas



Otros

Costos de servicios, licencias y demás

Selección de herramientas y metodologías



**Sistemas y
herramientas**



**Patrones de
diseño**

Consideraciones para la etapa de Extracción



Detectar cambios

Las extracciones deben ser diferenciales a la medida de lo posible



Pull vs Push

Podes recibir datos en lugar de ir a buscarlos



Metadatos

Que ayuden a determinar estados e brinden información del proceso



Area Staging

Base de datos intermedia que permita pre procesamiento

Cosideraciones para la etapa de transformación



Perfil de registros

Como se van a crear los registros, que informacion deben contener



Manejo de defectos

Como reaccionar ante datos Corruptos.



Manejo de duplicados

Establecer políticas de como se van a filtrar duplicados.



Tipos de transformaciones

Cálculos y demas

Cosideraciones para la etapa carga



Post procesos

Ejecución de pos procesos, re-indexación,
notificaciones, etc.

Consideraciones técnicas



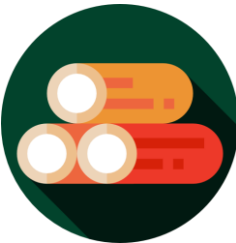
Sistemas de monitoreo

Se deben contar con herramientas para verificar las etapas del proceso



Backups

Información que permita conocer el estado y posibles errores



Logs / Vitácoras

Información que permita conocer el estado y posibles errores

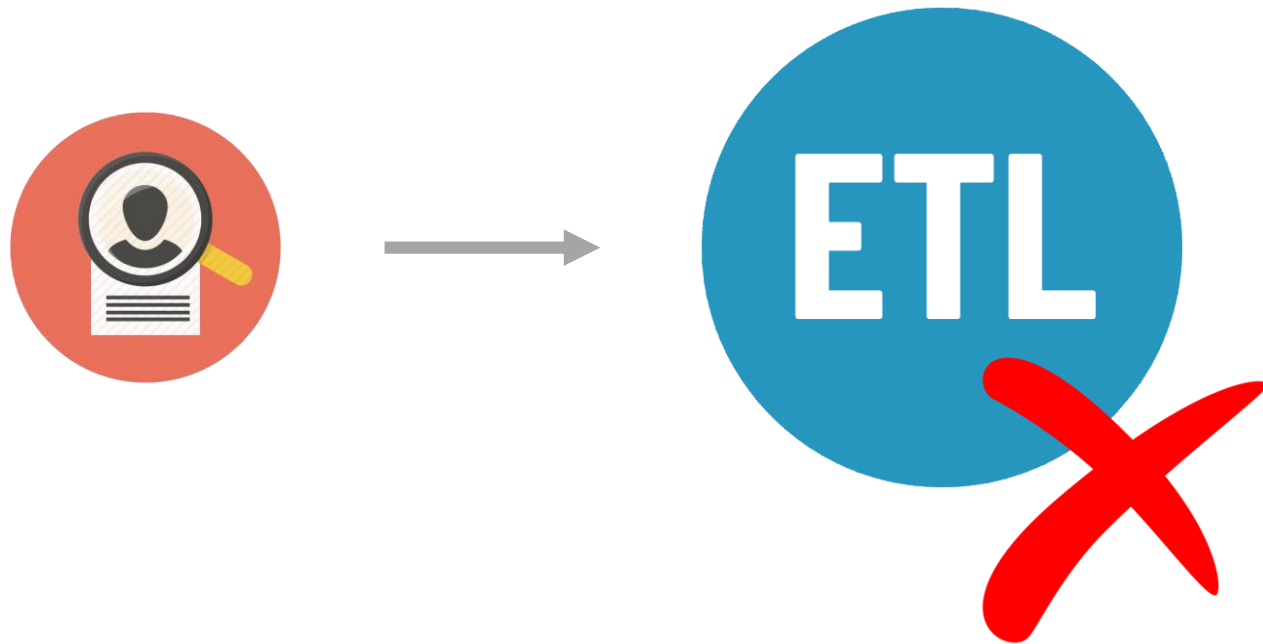


Capacidad de recuperación

Puede el sistema recuperarse ante una falla

Problemas comunes

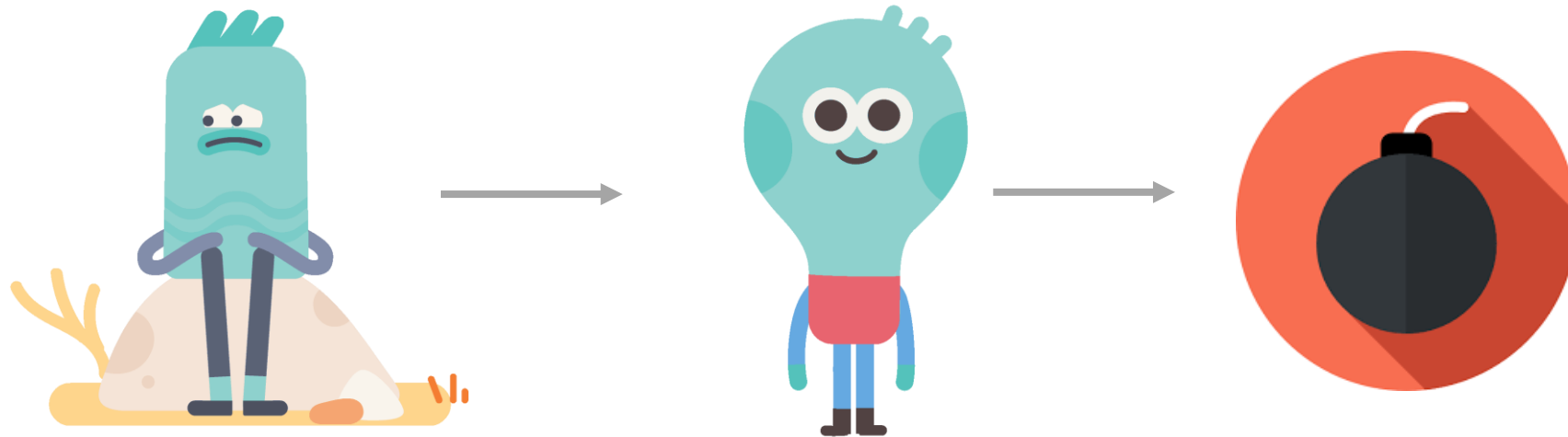
La prisa por empezar a implementar



- **Entender la información que se va a procesar**
- **Darse una idea de las posibles herramientas a utilizar**
- **Hacer una mejor estimación de los tiempos de trabajo**

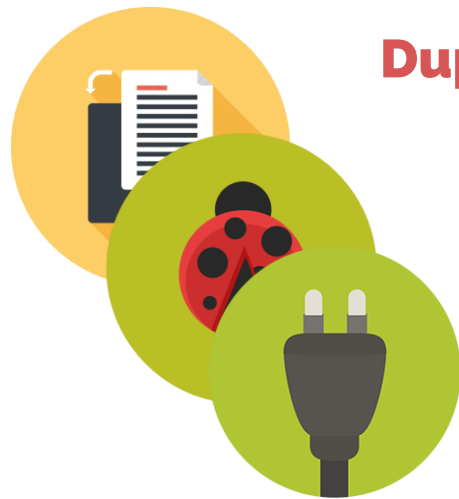
Problemas comunes

Desconocimiento de la lógica y objetivos del negocio



Problemas comunes

Incapacidad de manejar excepciones

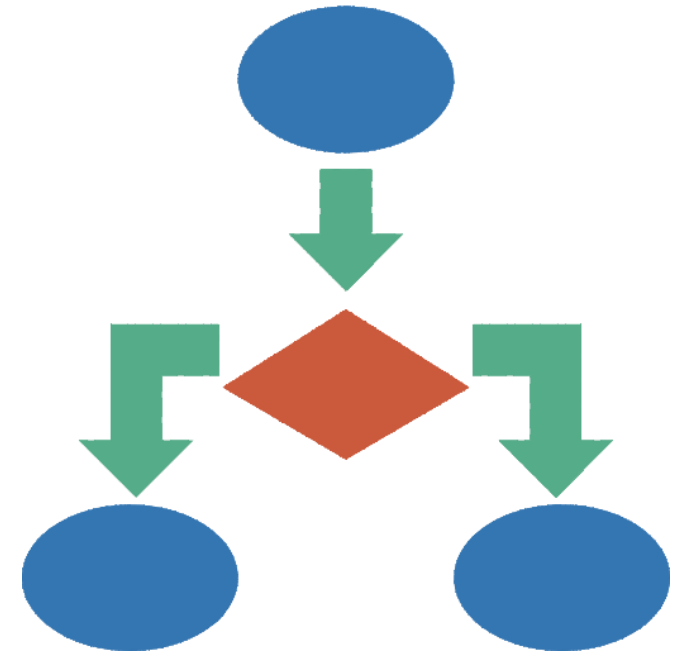


Duplicidad de datos

Datos defectuosos

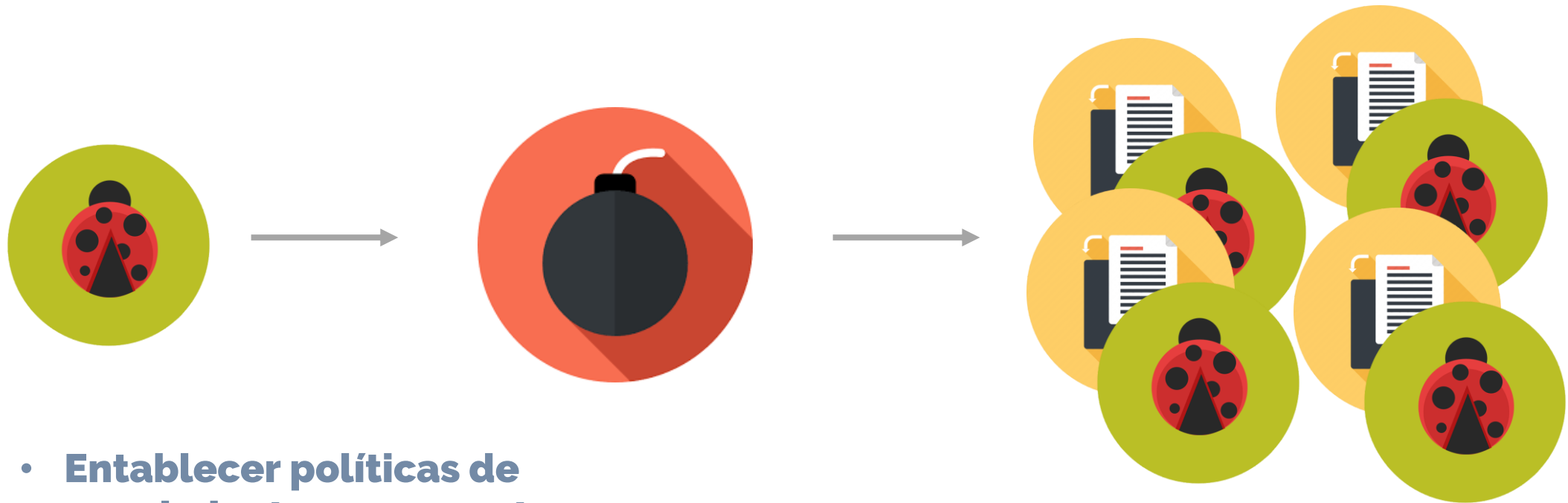
Fallos de conexión

- **Flujos a que manejen excepciones**



Problemas comunes

Insuficiente información para la revisión de procesos



- **Establecer políticas de seguimiento para encontrar errores**
- **Reportes**

Problemas comunes

Incapacidad de reiniciarse automáticamente



- **Reinicio manual del proceso**
 - Trabajo de horas extra
 - Alta probabilidad de fallo
 - Duplicados
 - Indices



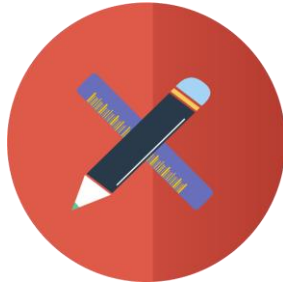
- **Reinicia automáticamente**
- **Notifica del problema encontrado**
- **Maneja el problema apropiadamente**
- **Utiliza índices para reanudar el trabajo**
- **No duplica**

**SUB PROCESOS – PROCESOS MAESTROS
– MANEJO DE ESTADOS**

Conclusiones



**Recolección de
requerimientos**



**Selección de
herramientas**



**Implementación y
revisión**

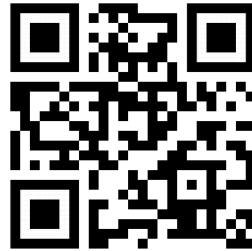


Listo!

¿Cómo integrarse?



<https://jugnicaragua.slack.com>



En la WEB

<http://javanicaragua.org/>



Telegram

<https://t.me/jugnicaragua>



¡Gracias!