

# Trabajo práctico integrador.

## Bases de datos

### 1. Introducción y motivación

Les proponemos para este trabajo final realizar el análisis y diseño de arquitecturas de bases de datos en aplicaciones reales. Vamos a considerar dos etapas, la primera trabajando con arquitecturas híbridas donde veremos como basado en una aplicación específica podemos plantear el uso de múltiples bases. Tengan en cuenta que este es un problema abierto con lo que no hay una única solución válida.

Además, a este, también contarán con un problema único para todos los grupos en el cual deberán implementar una base de datos relacional basados en su DER y con datos para cargar la misma. Durante la clase 3 comenzaremos a trabajar sobre el script de SQL para poder crear las tablas correspondientes en Amazon RDS.

### 2. Diseño de arquitecturas híbridas

Para esta parte, la idea es que con lo visto en clase sepan a partir de un requerimiento específico determinar como puede armarse una arquitectura de bases híbridas que nos permita alcanzar el mejor desempeño posible de nuestras aplicaciones. Como consideración general, vamos a tomar que todos los proyectos propuestos tienen un alto impacto y por ende necesitan de una gran disponibilidad de datos en cada una de sus partes.

Es válido mencionar que no se espera que realicen el deployment de la solución pero si que profundicen cuanto más puedan sobre el problema a analizar.

#### 2.1 Aplicaciones propuestas

##### 1. Plataforma de *Intelligent Documents* conectada:

- Para esta arquitectura debemos considerar un caso en el que tenemos una plataforma que debe poder relacionar documentos por distintos atributos (ejemplo: [connectedPapers](#))
- Asimismo, esta plataforma cuenta con un asistente virtual que permite al usuario realizar búsquedas usando lenguaje natural. Este asistente cuenta con historial de búsqueda y permite hacer sugerencias basadas en interacciones previas o similares.

##### 2. Plataforma de clasificación y procesamiento de documentos:

- Para este caso consideramos una plataforma de ML que nos permite hacer un procesamiento *end-to-end* de procesamiento de documentos, esto significa tomar un documento en pdf o jpeg y clasificarlo (si es un contrato, un balance, una política de trabajo, un instructivo), procesarlo (leer el documento mediante CV y OCR), entenderlo (NLP) y guardarlo.
- Esta plataforma debe poder guardar la información del documento completo, su resumen, sus palabras clave, las imágenes correspondientes para hacer referencia

topológica (en que línea se encuentra el texto). Como así la metadata necesaria para interconectar documentos.

**3. Plataforma de e-commerce tipo *marketplace*.**

- Consideraremos una plataforma de compraventa de cosas donde tenemos además una componente social para poder ver las interacciones entre vendedor y comprador.
- La plataforma tiene que contar con los servicios necesarios para poder hacer análisis de comportamiento entre los usuarios clientes, targeted advertising basado en su historial.
- Además, debe contar con puntas para poder realizar analítica sobre los ítems que se publican con su metada asociada.

**4. Asistente virtual de customer service con conexión a multiples servicios para CX.**

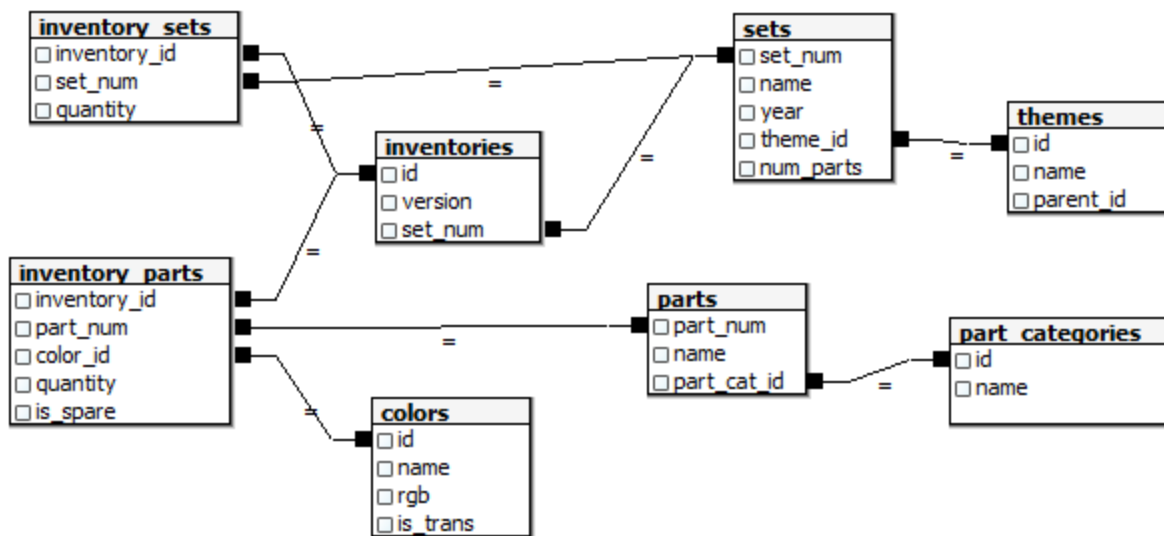
- Consideramos un asistente virtual embebido en la aplicación de market place que permite a los usuarios tanto compradores como vendedores de poder resolver distintas dudas sobre el uso de la plataforma.
- Además, el AV debe contar con información necesaria para poder relacionar distintos usuarios y las actividades entre sí mismas.
- El AV también debe contar con capacidad de poder discernir entre usuarios y administradores del sitio para ofrecer distintas herramientas.

**5. Plataforma de interconexión de sensores IoT para una fábrica con línea de producción.**

- Consideremos una aplicación o plataforma que nos permite interconectar una fábrica inteligente con múltiples sensores tanto en máquinas como en personas.
- La plataforma debe poder interconectar operario y máquina para poder hacer operaciones guiadas y ofrecer al operario documentación en caso de ser necesaria para el procedimiento en cuestión.
- Además, los sensores y máquinas interconectadas deben contar con un sistema de interconexión que les permite censar de manera continua el entorno y el avance de la producción (ejemplo: sistemas Industry X.0, 5G and industry digital transformation)

### 3. Análisis en conjunto Base de datos relacional

Vamos a considerar la [base datos de LEGO](#), la misma cuenta con el siguiente modelo relacional:



La idea de esta sección es que podamos montar esta base de datos en AWS RDS, para ello deberán valerse de lo visto en clases para poder levantar la misma con los datos provistos en el link.

Dentro de esta base, vamos a suponer que el equipo de marketing les pidió que hagan un reporte que contenga la siguiente información:

- ¿Cuáles son los 10 colores más utilizados en los 90?
- ¿Cuántos colores tenemos que son únicos?
- ¿Cómo es la tendencia en cantidad de piezas por sets a través de los años?
- ¿Cuáles son las temáticas más populares de los 2000?

**Pequeña ayudita:** si bien es un notebook en R, de acá pueden sacar algunas ideas interesantes

## 4. Presentación de resultados

Como presentación de este trabajo, deberán realizar una presentación para lo cual tendrán un bloque de tiempo de máximo 15 min por grupo. Durante la cursada también les recomendamos fuertemente presenten un informe de varianza muy cortito para poder guiarlos en el desarrollo.