

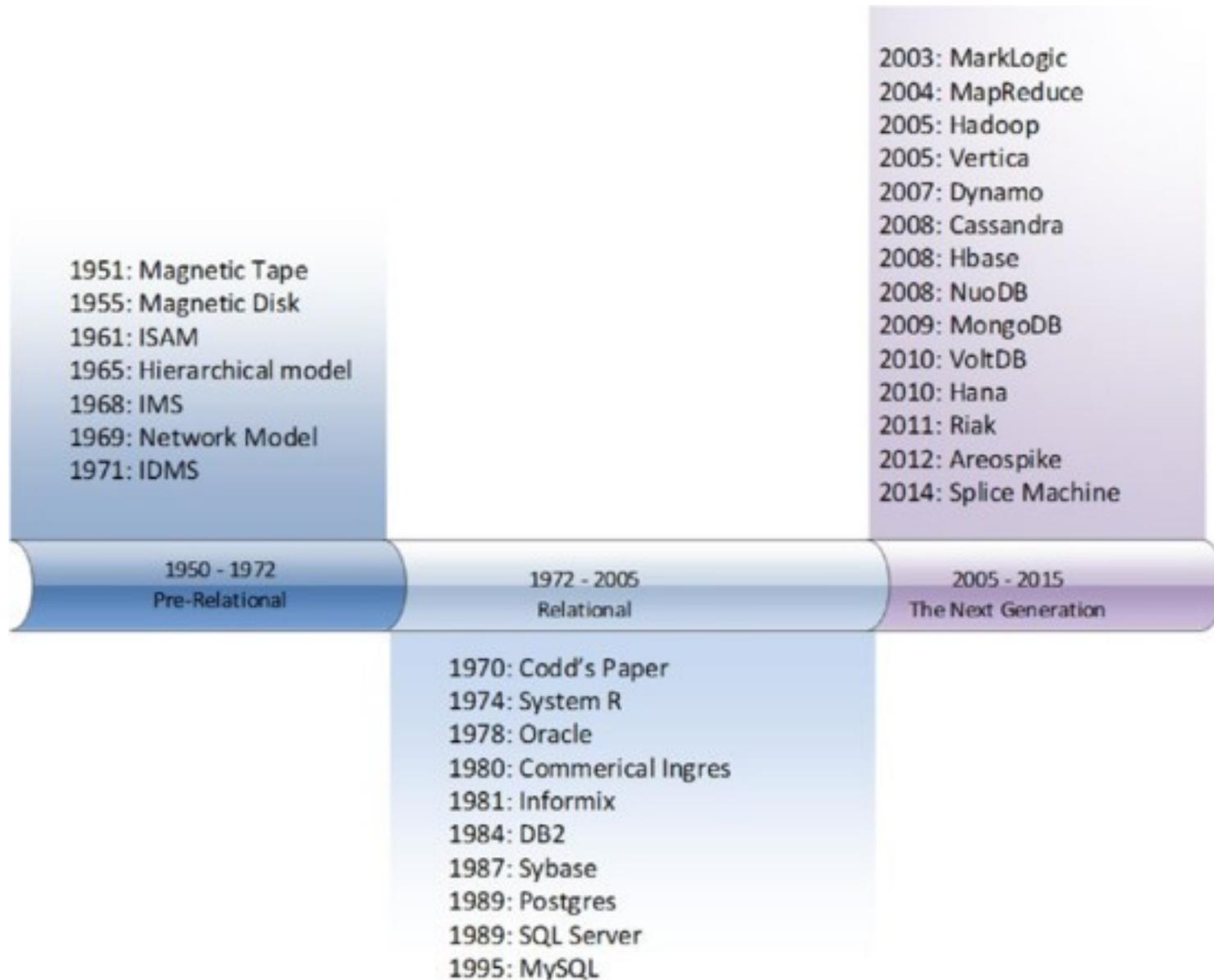
Evolución de los Sistemas de Bases de Datos



Rogelio Ferreira Escutia

Evolución

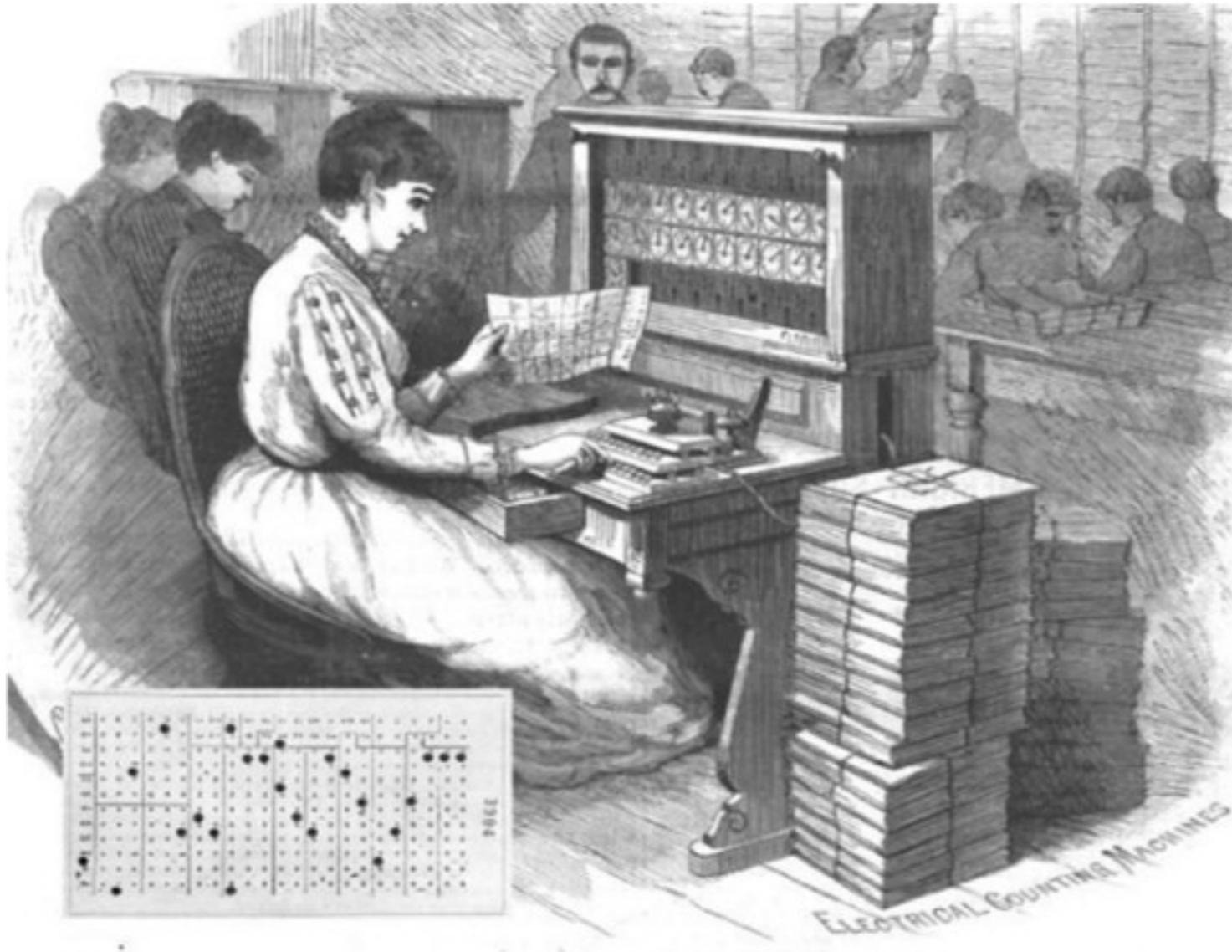
Evolución en los Sistemas de BD



Primera Generación

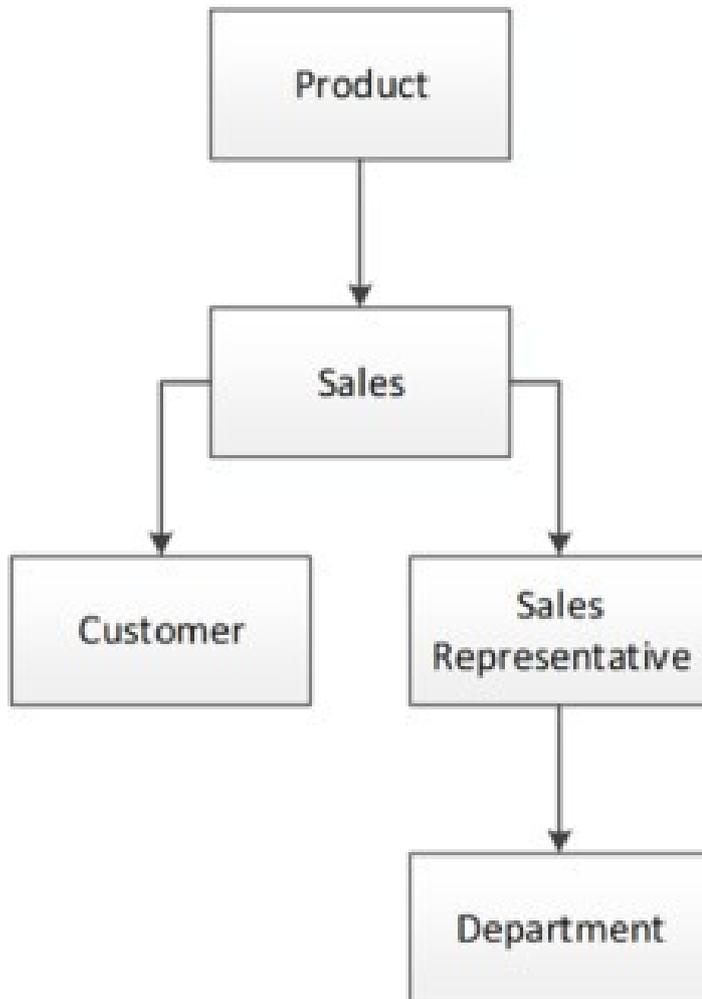
Tarjetas perforadas

- Censo de Estados Unidos en 1890

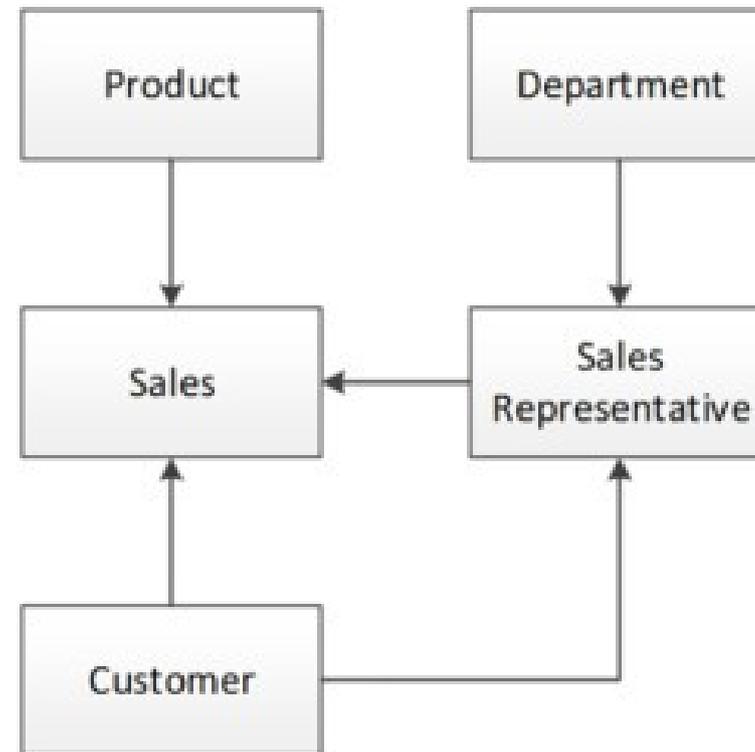


Modelos Usados

Hierarchical Model



Network Model

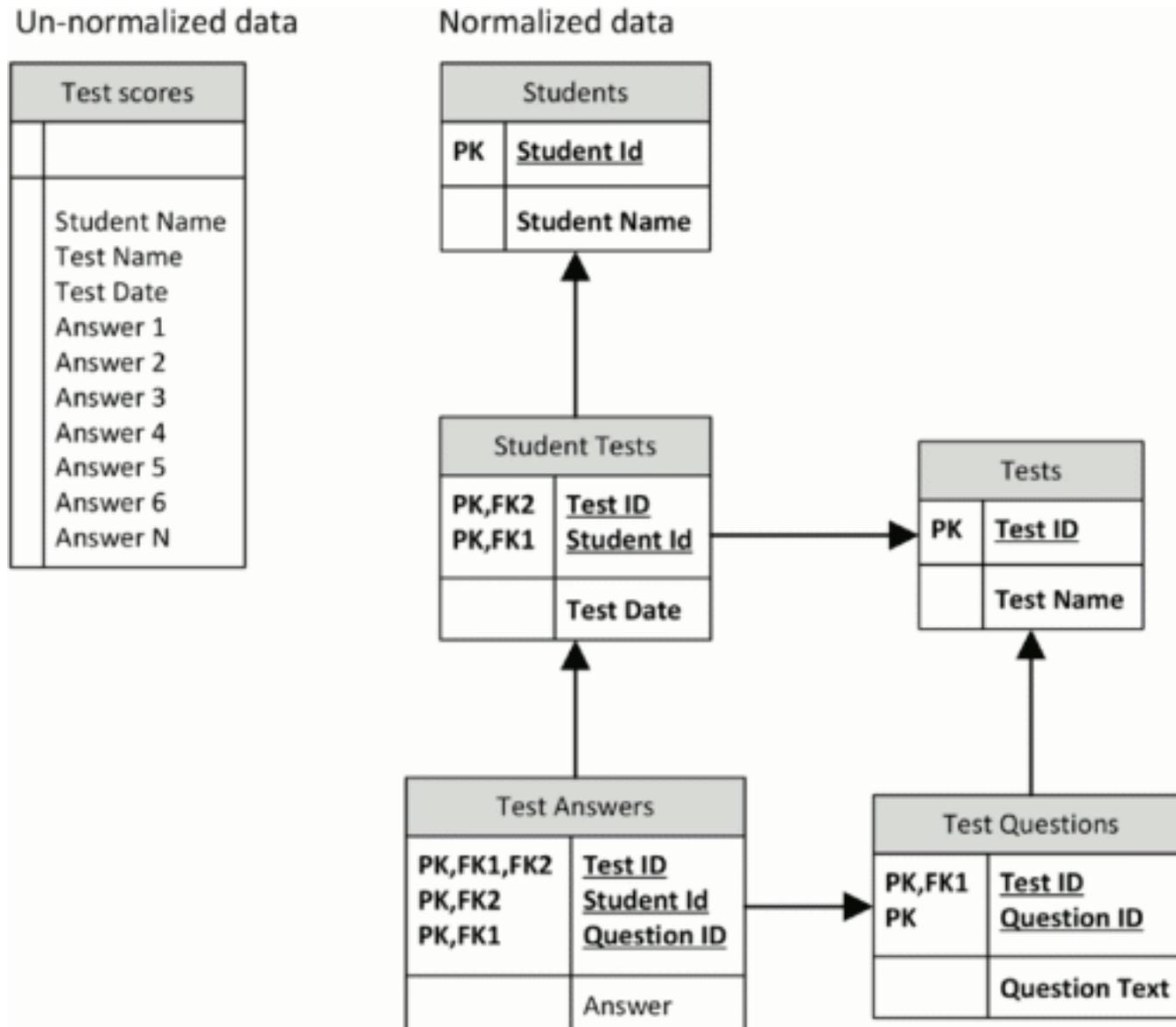


Segunda Generación

Modelo Relacional

- **Tuplas.**
- **Relaciones.**
- **Llaves.**
- **Uniones.**

Normalización



Modelo de Transacciones ACID

- **Propuesto por Jim Gray a finales de los 70s:**
 - **A (Atomic).**
 - **C (Consistency).**
 - **I (Isolated).**
 - **D (Durable).**

System R

- **IBM System R es un sistema de base de datos construido como un proyecto de investigación en el Laboratorio de Investigación San Jose de IBM a partir de 1974.**
- **System R fué la primera implementación de SQL.**

System R

- Computadora IBM 3032 (abril de 1980) donde se implementó System R.



Surgimiento de los DBMS

- **Oracle.**
- **Ingres.**
- **Sybase.**
- **Microsoft SQL Server.**
- **Informix.**
- **MySQL.**
- **DB2.**

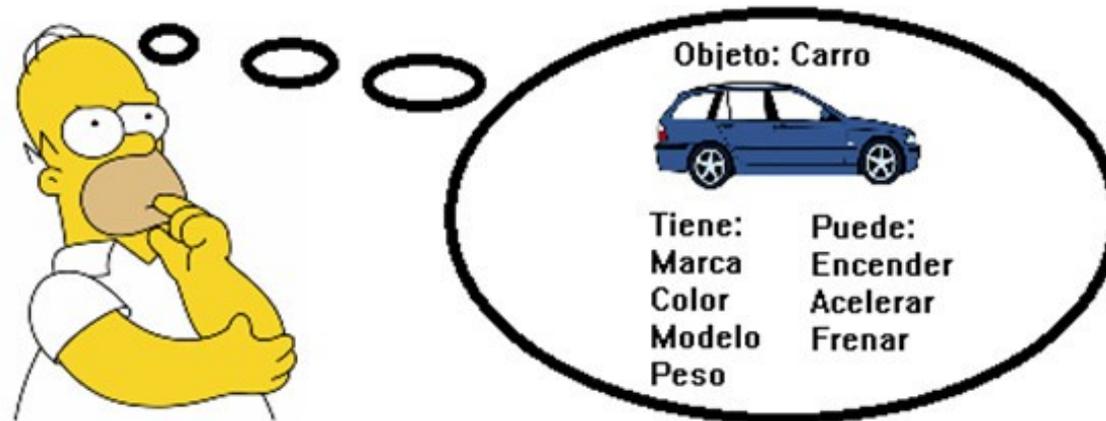
Computación Cliente - Servidor

- Con el surgimiento de las microcomputadoras basadas en las plataformas con el estándar de las IBM PC, los sistemas con interfaces gráficas como Microsoft Windows permitieron el nuevo paradigma Cliente - Servidor.



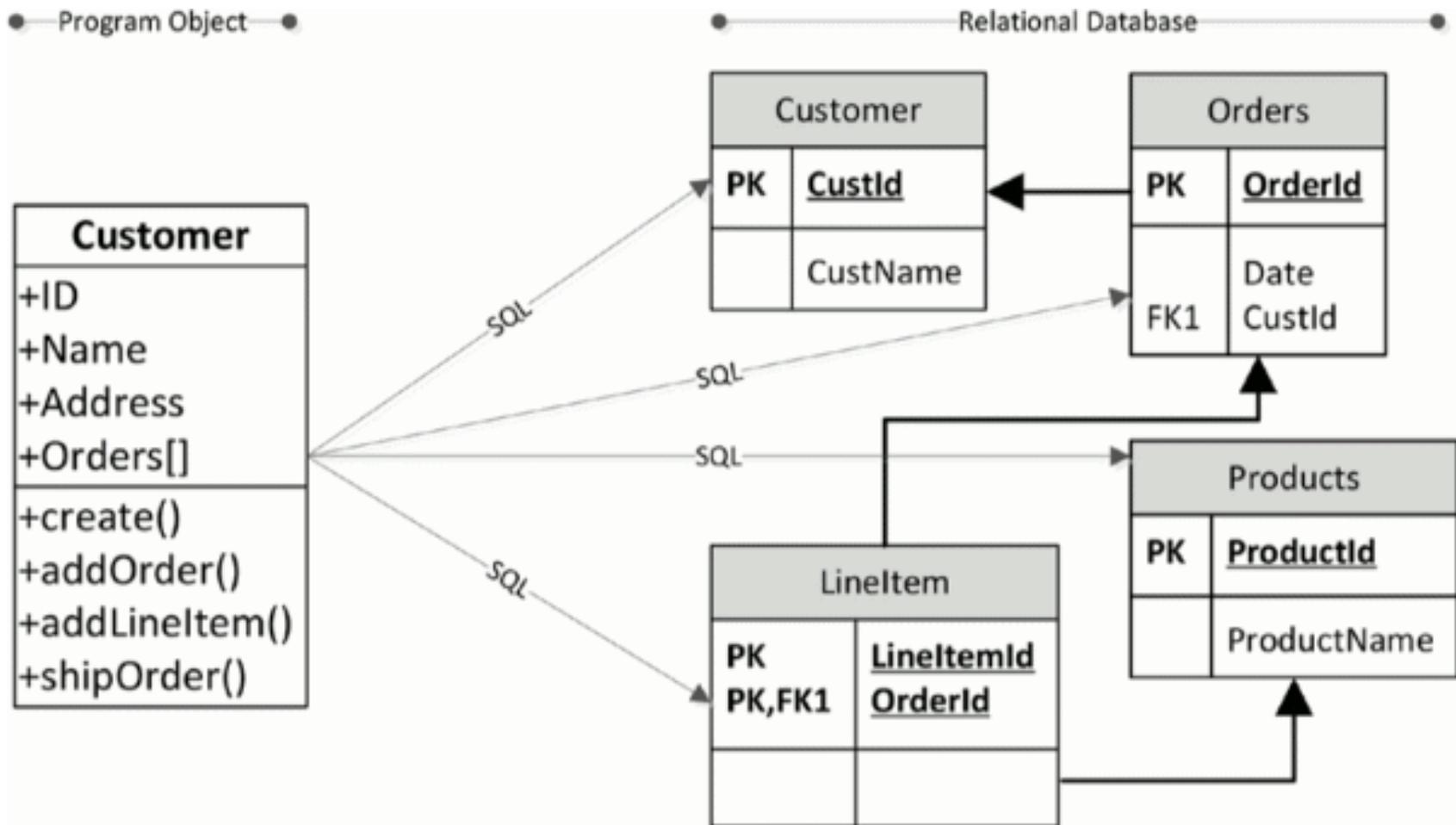
OODBMS

- Los “Sistemas de Bases de Datos” y la “Programación Orientada a Objetos” se unen y se crea la “Object Oriented Data Base Management Systems”.
- En este paradigma el objeto “carro” puede representar todos los registros asociados con el “carro”, así cómo su comportamiento (las operaciones que se pueden hacer con dichos registros).



OODBMS

- Objetos y su codificación en un esquema relacional:



OODBMS

- **Ventajas:**
 - **Se ganó en productividad del programador y desempeño de la aplicación.**
- **Desventajas:**
 - **Se requieren varias operaciones para trasladarlo a un sistema relacional.**
- **Debido a que pocas compañías implementaron este paradigma, muchos fabricantes a finales de los 90s empezaron a discontinuar su implementación.**

Tercera Generación

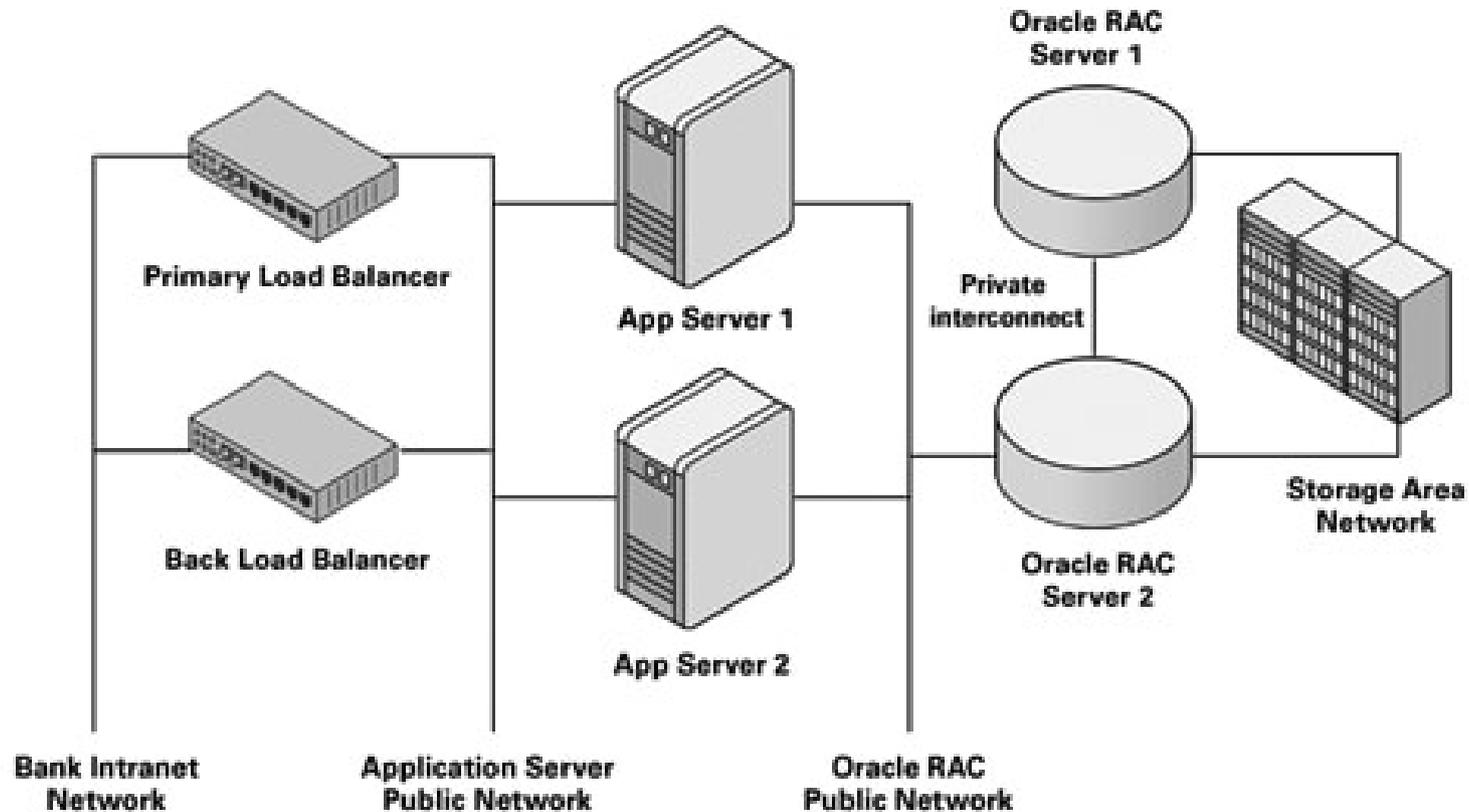
Inicio

- **A mitad de la década del 2000, los modelos relacionales ya no soportaron las nuevas tendencias como son los nuevos modelos cliente/servidor, las Tecnologías Web y el uso masivo de datos, por lo lo cual se propusieron nuevas plataformas para el uso de datos, iniciando la tercera generación de Sistemas de Bases de Datos.**

Desarrollos de Oracle

Oracle RAC (2001)

- Es un sistema orientado a sistemas de Clusters que provee alta disponibilidad en ambientes de bases de datos de Oracle.



Desarrollos de Google

Google GFS (2003)

- **El Sistema de Archivos Google, en inglés Google File System (GFS, GooFS o GoogleFS), es un sistema de archivos distribuido propietario desarrollado por Google Inc, que soporta toda su infraestructura informática de procesamiento de información en nube.**
- **Está especialmente diseñado para proveer eficiencia, fiabilidad de acceso a datos usando sistemas masivos de cluster de procesamiento en paralelo. La actual versión de Google File System tiene el nombre clave Colossus.**

Google MapReduce (2004)

- **MapReduce es un modelo de programación para dar soporte a la computación paralela sobre grandes colecciones de datos en grupos de computadoras.**

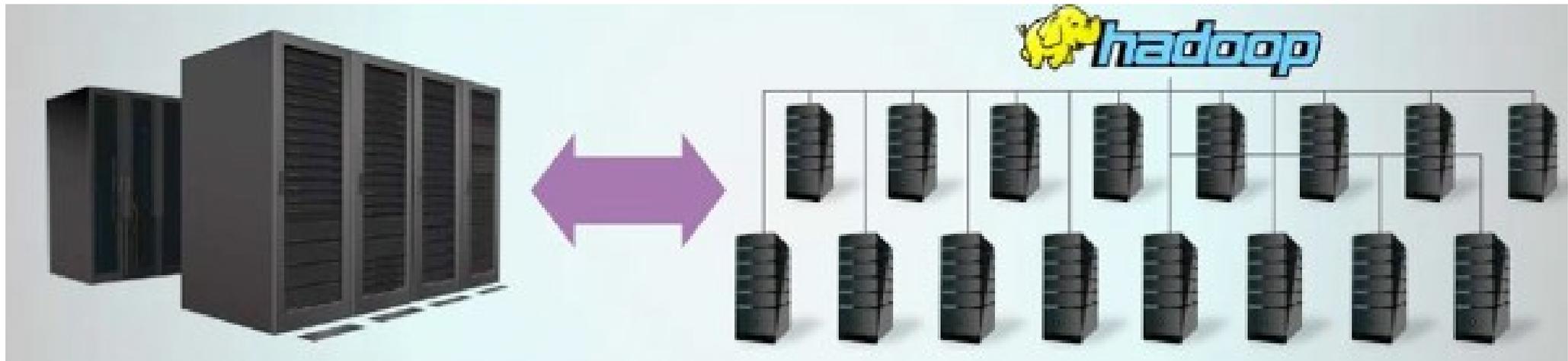


Hadoop (2004)

- Hadoop es una implementación OpenSource inspirada en los documentos de Google para MapReduce y Google File System (GFS).
- Su desarrollo fué liderado inicialmente por Yahoo y actualmente lo realiza el proyecto Apache.
- Apache Hadoop es un framework de software que soporta aplicaciones distribuidas bajo una licencia libre.
- Permite a las aplicaciones trabajar con miles de nodos y petabytes de datos. Hadoop se inspiró en los documentos Google para MapReduce y Google File System (GFS).

Hadoop (2004)

- Almacenamiento de datos distribuido, tolerante a fallos.
- Analogía con un sistema de ficheros, no una BD.
- HDFS ha demostrado que es escalable hasta 200 PB de almacenamiento en un cluster de 4500 nodos



Desarrollos de Amazon

Amazon AWS (2006)

- **Amazon Web Services (AWS abreviado) es una colección de servicios de computación en la nube pública (también llamados servicios web) que en conjunto forman una plataforma de computación en la nube, ofrecidas a través de Internet por Amazon.com.**



Public data storage



Private data storage



Dynamic web hosting



Web Apps hosting



Web Apps hosting



VPN gateway



Multimedia streaming



Load balancing



Auto scaling

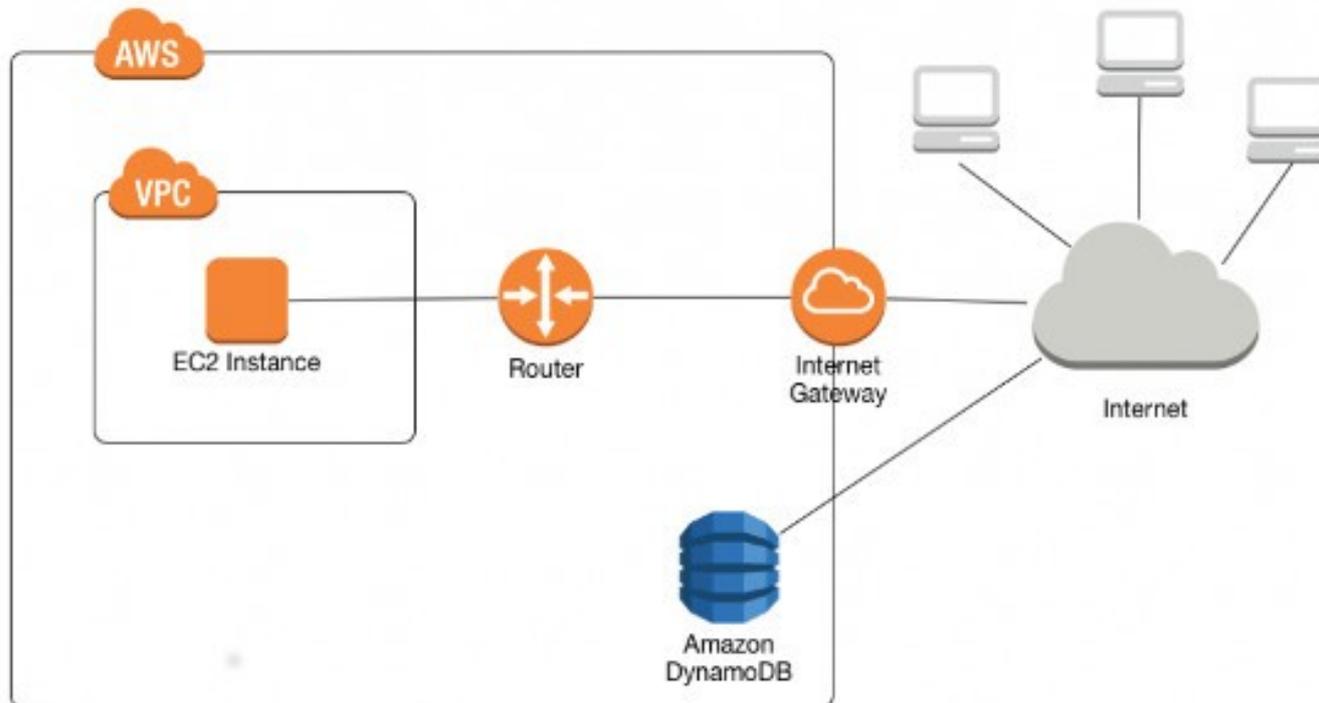
... and much, much more!

Amazon Dynamo (2008)

- **Fu  propuesto por Amazon y es un sistema de almacenamiento interno utilizado inicialmente para su propio sitio web.**
- **Es un modelo alternativo al esquema ACID ya que es muy dif cil mantener un sistema relacional en un ambiente grande de informaci n.**

Amazon DynamoDB (2012)

- Es un servicio de base de datos noSQL ofrecido por Amazon como parte de Amazon Web Services.
- DynamoDB puede gestionar más de 10 billones de solicitudes por día y admite picos de más de 20 millones de solicitudes por segundo.



Sistemas “document”

MongoDB (2017)

- **MongoDB (del inglés humongous, "enorme") es un sistema de base de datos NoSQL orientado a documentos de código abierto.**
- **En lugar de guardar los datos en tablas, tal y como se hace en las bases de datos relacionales, MongoDB guarda estructuras de datos BSON (una especificación similar a JSON) con un esquema dinámico, haciendo que la integración de los datos en ciertas aplicaciones sea más fácil y rápida.**

Sistemas “key-value”

Cassandra (2008)

- Apache Cassandra es una base de datos NoSQL distribuida y basada en un modelo de almacenamiento de «clave-valor», de código abierto que está escrita en Java.
- Permite grandes volúmenes de datos en forma distribuida.
- Lo usa Twitter para su plataforma. Su objetivo principal es la escalabilidad lineal y la disponibilidad.

Sistemas “NoSQL”

NoSQL

- **NoSQL (a veces llamado "no sólo SQL") es una amplia clase de sistemas de gestión de bases de datos que difieren del modelo clásico de SGBDR (Sistema de Gestión de Bases de Datos Relacionales) en aspectos importantes, siendo el más destacado que no usan SQL como lenguaje principal de consultas.**
- **Los datos almacenados no requieren estructuras fijas como tablas, normalmente no soportan operaciones JOIN, ni garantizan completamente ACID (atomicidad, consistencia, aislamiento y durabilidad) y habitualmente escalan bien horizontalmente.**

NoSQL

DESCRIPCIÓN	MICROSOFT AZURE TABLE STORAGE	AMAZON SIMPLEDDB	GOOGLE CLOUD BIGTABLE
MODELO DE BASE DE DATOS	Wide column store	Wide column store	Key-value store
DESARROLLADOR	Microsoft	Amazon	Google
PUBLICACIÓN INICIAL	2012	2007	2015
ESQUEMA DE DATOS	Libre	Libre	Libre
TIPOS PREDEFINIDOS	Si	no	no
MÉTODOS DE ACCESO	RESTfull API	RESTfull API	gRPC (using protocol buffers) API HBase compatible API (Java)
LENGUAJES SOPORTADOS	.Net (todos), C++, Java, JavaScript, PHP, Python, Ruby	.Net (todos), C, C++, Erlang, Java, PHP, Python, Ruby, Scala	Go, Java
MÉTODO DE PARTICIONADO	Sharding	Sharding	ninguno
CONSISTENCIA	Immediate Consistency	Eventual Consistency Immediate Consistency	Immediate Consistency
TRANSACCIONES	Bloqueo optimísitico	no	Operaciones atómicas en un registro
CONTROL DE ACCESO	Derechos de acceso basados en permisos sobre le servicio, autenticación por clave privada y firmas compartidas	Derechos de acceso vía AWS Identity and Access Management (IAM)	no



Rogelio Ferreira Escutia

***Instituto Tecnológico de Morelia
Departamento de Sistemas y Computación***

***Correo: rogelio@itmorelia.edu.mx
 rogeplus@gmail.com***

***Página Web: http://sagitario.itmorelia.edu.mx/~rogelio/
 http://www.xumarhu.net/***

Twitter: http://twitter.com/rogeplus

Facebook: http://www.facebook.com/groups/xumarhu.net/