

# Introducción y Conceptos básicos de bases de datos

---

Alex Di Genova

April 4, 2022

Universidad de O'higgins

Bienvenida curso de Bases de datos

Planificación curso BD

# **Bienvenida curso de Bases de datos**

---

Alex Di Genova

- 2003–2008 Ingeniero en Bioinformática.
- 2013-2017 Doctor en Sistemas Complejos.
- 2017-2021 Postdoctorado en algoritmos y cancer (Francia).
- 2022 - Profesor Asistente UOH.
  - Di Genoma Lab
    - Combinamos el desarrollo de nuevos algoritmos, análisis de genomas y tecnologías ómicas de última generación para estudiar sistemas biológicos complejos.



## Original article

# SalmonDB: a bioinformatics resource for *Salmo salar* and *Oncorhynchus mykiss*

Alex Di Génova<sup>1</sup>, Andrés Aravena<sup>1,2,\*</sup>, Luis Zapata<sup>1,3</sup>, Mauricio González<sup>1,4</sup>, Alejandro Maass<sup>1,2</sup> and Patricia Iturra<sup>3</sup>

<sup>1</sup>Laboratory of Bioinformatics and Mathematics of the Genome, Center for Mathematical Modeling (UMI 2807 CNRS) and Center for Genome Regulation (Fondap 15090007), University of Chile, Santiago, Chile, <sup>2</sup>Department of Mathematical Engineering, Faculty of Physical and Mathematical Sciences, University of Chile, Santiago, Chile, <sup>3</sup>ICBM Human Genetics Program, Faculty of Medicine, University of Chile, Santiago, Chile and <sup>4</sup>Laboratorio de Bioinformática y Expresión Génica, INTA, University of Chile, Santiago, Chile

\*Corresponding author: Tel: +56(2) 978 48 70; Fax: +56(2) 688 97 05; Email: andres.aravena@dim.uchile.cl

Submitted 1 July 2011; Revised 21 September 2011; Accepted 16 October 2011

SalmonDB is a new multiorganism database containing EST sequences from *Salmo salar*, *Oncorhynchus mykiss* and the whole genome sequence of *Danio rerio*, *Gasterosteus aculeatus*, *Tetraodon nigroviridis*, *Oryzias latipes* and *Takifugu rubripes*, built with core components from GMOD project, GOPArc system and the BioMart project. The information provided by this resource includes Gene Ontology terms, metabolic pathways, SNP prediction, CDS prediction, orthologs prediction, several precalculated BLAST searches and domains. It also provides a BLAST server for matching user-provided sequences to any of the databases and an advanced query tool (BioMart) that allows easy browsing of EST databases with user-defined criteria. These tools make SalmonDB database a valuable resource for researchers searching for transcripts and genomic information regarding *S. salar* and other salmonid species. The database is expected to grow in the near future, particularly with the *S. salar* genome sequencing project.

Database URL: <http://genomicasalmones.dim.uchile.cl/>

Alumnas y Alumnos

# Planificación curso BD

---

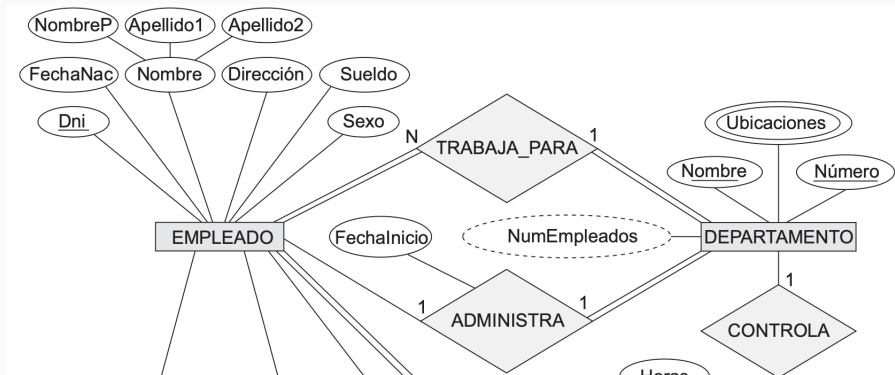
- 4 Unidades (14 semanas)
  - Modelamiento de bases de datos (3 semanas)
  - Modelo Relacional (4 semanas)
  - Lenguaje de Consulta SQL (3 semanas)
  - Transacciones y Bases de datos no relacionales (4 semanas)



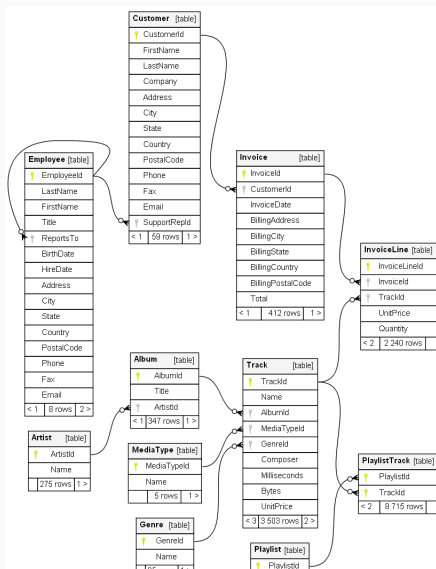
## Modelamiento de bases de datos

- Cada **persona** puede habitar en solo una **vivienda** y estar registrada en solo un **municipio** pero puede ser propietaria de varias viviendas. . . .
- La empresa está organizada en departamentos. Cada uno tiene un nombre único, un número único y un empleado concreto que lo administra. Un departamento controla una cierta cantidad de proyectos, cada uno de los cuales tiene un nombre único, un número único y una sola ubicación.

## Esquema Relacional



## Diagrama Entidad-Relacion (Modelo UML)




## SQL (SQLite3)

```
[10] 1 %%sql
      2 PRAGMA table_info([Artist]);

* sqlite:///content/chinook-database/ChinookDatabase/DataSources/Chinook_Sqlite.sqlite
Done.
cid name      type      notnull dflt_value pk
0  ArtistId INTEGER      1      None      1
1  Name  NVARCHAR(120) 0      None      0
```

```
1 %%sql
2 select * from Artist limit 10;
```

 \* sqlite:///content/chinook-database/ChinookDatabase/DataSources/Chinook\_Sqlite.sqlite  
Done.

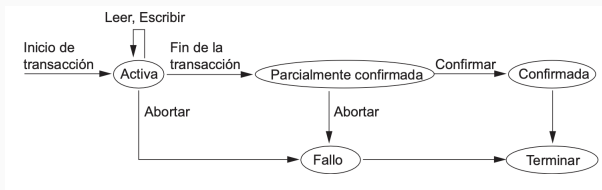
ArtistId	Name
1	AC/DC
2	Accept
3	Aerosmith
4	Alanis Morissette
5	Alice In Chains
6	Antônio Carlos Jobim
7	Apocalyptica
8	Audioslave
9	BackBeat
10	Billy Cobham

GoogleColab – <https://colab.research.google.com/>

Video descriptivo de GoogleColab

## Transacciones

- Atomicidad
- Conservación de la consistencia.
- Aislamiento
- Durabilidad



## No-SQL

- levelDB, rocksDB ...

- Controles :
  - Control 1: Semana del 9 Mayo.
  - Control 2: Semana del 13 Junio.
  - Control 3: Semana del 11 Julio.

- Controles :
  - Control 1: Semana del 9 Mayo.
  - Control 2: Semana del 13 Junio.
  - Control 3: Semana del 11 Julio.
- Tareas:
  - Tarea 1: Semana del 9 Mayo.
  - Tarea 2: Semana del 6 Junio.

## Condiciones y Políticas de Evaluación

- El promedio de actividades complementarias se considerará como un cuarto control (control IV) y tendrá una ponderación de 15%. El promedio de controles I,II,III y IV con sus respectivas ponderaciones corresponderán a la nota final del curso. El curso será aprobado con una nota promedio igual o superior a 4,0.



## Condiciones y Políticas de Evaluación

- El promedio de actividades complementarias se considerará como un cuarto control (control IV) y tendrá una ponderación de 15%. El promedio de controles I,II,III y IV con sus respectivas ponderaciones corresponderán a la nota final del curso. El curso será aprobado con una nota promedio igual o superior a 4,0.
- Estudiantes que se ausenten a un control tendrán la oportunidad de recuperarlo durante el periodo correspondiente al final del semestre. El control recuperativo es de carácter **acumulativo**, por lo tanto, contendrá contenido de las cuatro unidades del curso. Adicionalmente, alumnos que quieran remplazar una calificación en un control o actividades complementarias, también podrán rendir el control recuperativo.

## Condiciones y Políticas de Evaluación

- El promedio de actividades complementarias se considerará como un cuarto control (control IV) y tendrá una ponderación de 15%. El promedio de controles I,II,III y IV con sus respectivas ponderaciones corresponderán a la nota final del curso. El curso será aprobado con una nota promedio igual o superior a 4,0.
- Estudiantes que se ausenten a un control tendrán la oportunidad de recuperarlo durante el periodo correspondiente al final del semestre. El control recuperativo es de carácter **acumulativo**, por lo tanto, contendrá contenido de las cuatro unidades del curso. Adicionalmente, alumnos que quieran remplazar una calificación en un control o actividades complementarias, también podrán rendir el control recuperativo.
- Un/a estudiante que cometa plagio obtendrá un 1,0 en la evaluación y el caso será informado a Escuela de Ingeniería.

# Resultados de Aprendizaje

- Diseñar diagramas de Entidad/Relacional para satisfacer las necesidades de un problema enunciado.
- Realizar a partir de un diagrama Entidad/Relación un diseño relacional.
- Normalizar un diseño relacional de bases de datos.
- Formular consultas de distinto tipo en SQL.
- Reconocer la noción de transacción y operar el sistema de recuperación de un sistema de administración de bases de datos.
- Conocer sistemas de bases de datos no relacionales.