

Evaluación de una nueva línea de crédito empresarial basada en análisis de datos financieros.

Jhoan Sebastian Beltran Pabon¹, Juan Guillermo Ballesteros²

Carlos Arturo Diaz³, Darly Rodríguez Jiménez^{3,4}

¹⁻⁴Dpto. de Ingeniería y ciencias básicas
Universidad Central
Analítica de datos

Bogotá, Colombia

May 4, 2024

Contents

1	Introducción	3
2	Características del proyecto de investigación que hace uso de Bases de Datos	3
2.1	Titulo del proyecto de investigación	3
2.2	Objetivo general	3
2.2.1	Objetivos especificos	3
2.3	Alcance	4
2.4	Pregunta de investigación	4
2.5	Hipotesis	4
3	Reflexiones sobre el origen de datos e información	6
3.1	¿Cual es el origen de los datos e información?	6
3.2	¿Cuales son las consideraciones legales o eticas del uso de la información?	6
3.3	¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación?	6
3.4	¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto?	6

4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)	7
4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto	7
4.2	Diagrama modelo de datos	8
4.3	Imágenes de la Base de Datos	8
4.4	Código SQL - Lenguaje de definición de datos (DDL)	8
4.5	Código SQL - Manipulación de Datos (DML)	11
4.6	Código SQL + Resultados: Vistas	11
4.7	Código SQL + Resultados: Triggers	12
4.8	Código SQL + Resultados: Funciones	13
4.9	Código SQL + Resultados: Procedimientos Almacenados	13
5	Bases de Datos No-SQL (<i>Segunda entrega</i>)	14
5.1	Diagrama Bases de Datos No-SQL (<i>Segunda entrega</i>)	14
5.2	SMBD utilizado para la Base de Datos No-SQL (<i>Segunda entrega</i>)	14
6	Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos (<i>Tercera entrega</i>)	15
6.1	Ejemplo de aplicación de ETL y Bodega de Datos (<i>Tercera entrega</i>)	15
6.2	Automatización de Datos (<i>Tercera entrega</i>)	15
6.3	Integración de Datos (<i>Tercera entrega</i>)	15
7	Proximos pasos (<i>Tercera entrega</i>)	16
8	Lecciones aprendidas (<i>Tercera entrega</i>)	17
9	Bibliografía	18

1 Introducción

En un entorno empresarial cada vez más complejo y competitivo, las empresas en Colombia deben contar con herramientas precisas y eficientes para tomar decisiones informadas. La recopilación y análisis de datos financieros desempeña un papel fundamental en la planificación estratégica y la evaluación del desempeño empresarial. La Financiera del Desarrollo Territorial (Findeter) ha jugado un papel relevante en el impulso del crecimiento económico y el desarrollo sostenible en las diversas regiones de Colombia. Para mantener su enfoque innovador y apoyar la evolución empresarial, Findeter busca comprender mejor el estado financiero de las empresas colombianas. Este estudio se centra en el diseño y la implementación de una base de datos en Oracle para gestionar y analizar la información financiera de las 1000 empresas más grandes de Colombia. La base de datos permitirá aplicar técnicas de modelado, normalización y SQL para organizar datos de manera eficiente y producir informes y visualizaciones que generen insights sobre el desempeño empresarial. El análisis de estos datos será clave para desarrollar nuevas estrategias y oportunidades financieras que impulsen la competitividad y contribuyan al fortalecimiento de la economía nacional. Con esta iniciativa, Findeter aspira a mejorar el entendimiento del entorno financiero empresarial en Colombia, facilitando la toma de decisiones fundamentadas y fomentando el desarrollo sostenible.

2 Características del proyecto de investigación que hace uso de Bases de Datos

2.1 Título del proyecto de investigación

Evaluación de una nueva línea de crédito empresarial basada en análisis de datos financieros.

2.2 Objetivo general

Diseñar e implementar una base de datos en Oracle para gestionar y analizar la información financiera de las 1000 empresas más grandes de Colombia, aplicando técnicas de modelado, normalización y SQL, con el fin de generar informes y visualizaciones que permitan obtener insights sobre el desempeño empresarial.

2.2.1 Objetivos específicos

- Analizar los datos de las 1000 empresas, identificando entidades, atributos y relaciones para diseñar un modelo conceptual y lógico de la base de datos usando Oracle Modeler.
- Implementar la base de datos en Oracle Developer, creando tablas, relaciones, restricciones y objetos necesarios usando SQL, y aplicando principios de normalización para garantizar integridad y eficiencia.

- Desarrollar consultas SQL y visualizaciones en Oracle Developer para obtener información relevante sobre el desempeño financiero de las empresas, aplicando conceptos de álgebra relacional y cálculo relacional.
- Elaborar un dashboard interactivo en Oracle Developer que permita a los usuarios explorar y analizar de manera intuitiva los datos financieros de las empresas, generando gráficos y tablas dinámicas basados en las consultas SQL desarrolladas.

2.3 Alcance

Se contempla realizar el diseño y modelado de una base de datos relacional sobre las 1000 principales empresas de Colombia utilizando el motor Oracle de bases de datos y modelado, Data Modeler y el Oracle Developer. La información se extrae únicamente del informe de las 1000 empresas más grandes del país publicado en la página de la supersociedades el 22 de junio del 2023 con corte al 31 de diciembre del 2022.

Se plantea realizar una base de datos que contenga las siguientes tablas de información:

Empresa: Contiene toda la información de registro legal de la compañía.

Sector: Contiene la información del sector económico de la compañía.

Finanzas: Contiene la información financiera de la compañía.

Ubicación: Contiene la información de la localización de la compañía.

Supervisores: Contiene la información gerencial de la compañía.

2.4 Pregunta de investigación

¿Cómo puede el análisis de datos financieros de las principales empresas colombianas, mediante un sistema de gestión de bases de datos, contribuir a evaluar la viabilidad y potencial impacto de una nueva línea de crédito empresarial en la economía nacional?

2.5 Hipotesis

Mediante la gestión y análisis de la información financiera de las 1000 empresas más grandes de Colombia, utilizando técnicas de modelado, normalización y SQL, se podrá obtener una visión integral del desempeño empresarial que permitirá identificar patrones, tendencias y relaciones significativas. Así mismo, al desarrollar consultas SQL basadas en conceptos de álgebra y cálculo relacionales, se podrán obtener insights relevantes sobre el desempeño financiero de las empresas, lo que facilitará la toma de decisiones informadas. Con esta información se espera que la implementación de una nueva línea de crédito empresarial con tasa compensada tenga un impacto positivo y significativo en la actividad empresarial en Colombia, reflejado en un aumento en la cantidad de proyectos y programas

estratégicos respaldados por la Financiera del Desarrollo Territorial (Findeter) y una mejora en la competitividad empresarial a nivel nacional.

3 Reflexiones sobre el origen de datos e información

3.1 ¿Cual es el origen de los datos e información?

Investigación de la Supersociedades de las 1.000 Empresas y se realiza cada año teniendo en cuenta la información financiera de las empresas del país. El actual informe corresponde a la información de estados financieros individuales de las empresas para el año 2022 con corte a 31 de diciembre. Contiene la información financiera incluyendo activos, pasivo, patrimonio, ingresos operacionales y ganancias/pérdidas. El estudio tiene en cuenta la información financiera presentada bajo normas internacionales de información financiera (NIIF) y toma como criterio para determinar el tamaño de las empresas los ingresos operacionales.

3.2 ¿Cuales son las consideraciones legales o eticas del uso de la información?

Integración de información de la Superintendencia de Sociedades y de otras entidades como Superintendencia Financiera, Superintendencia de Vigilancia y Seguridad Privada, Superintendencia de Salud, Superintendencia de Servicios Públicos Domiciliarios y Superintendencia de Transporte. El 70 de las sociedades que hacen parte de las 1000 empresas más grandes pertenecen a un conglomerado societario y reportan ingresos operacionales equivalentes a 941 billones de pesos.

3.3 ¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación?

Los retos son la información son la generación de matrices o controlantes que permiten conocer información financiera que incluye activos, pasivo, patrimonio, ingresos operacionales y ganancias/pérdidas. Información relevante sobre el sector empresarial del país elaborada por la Superintendencia de Sociedades.

3.4 ¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto?

Facilidad de acceso y búsqueda de información: Un sistema de base de datos que permita acceder fácilmente a los datos recopilados sobre estas empresas, así como realizar búsquedas rápidas y precisas para obtener información específica, organizada de manera estructurada y coherente, lo que facilita su análisis y comprensión. Esto puede incluir categorización por sector, tamaño de la empresa, ingresos, ganancias, entre otros aspectos relevantes. Análisis y generación de informes. Además, de facilitar la generación de informes personalizados y la visualización de datos, lo que ayudaría a comunicar de manera efectiva los hallazgos de la base de datos.

4 Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)

4.1 Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto

Oracle Developer es una herramienta de desarrollo integrada (IDE) gratuita diseñada específicamente para trabajar con la base de datos Oracle.

El Oracle Developer permite escribir, editar y ejecutar código SQL de forma gráfica, facilitando la construcción y depuración de consultas complejas. Adicional también permite hacer procedimientos almacenados, funciones y triggers en el lenguaje PL/SQL.

Esta herramienta fue desarrollada a inicios de los años 90s junto con todo el desarrollo de los sistemas Oracle. Actualmente está disponible para ser usada en sistemas operativos Windows, Linux y macOS.

Se escoge este gestor de bases de datos porque nos parece que cumple con todos los requisitos que se requieren para este proyecto:

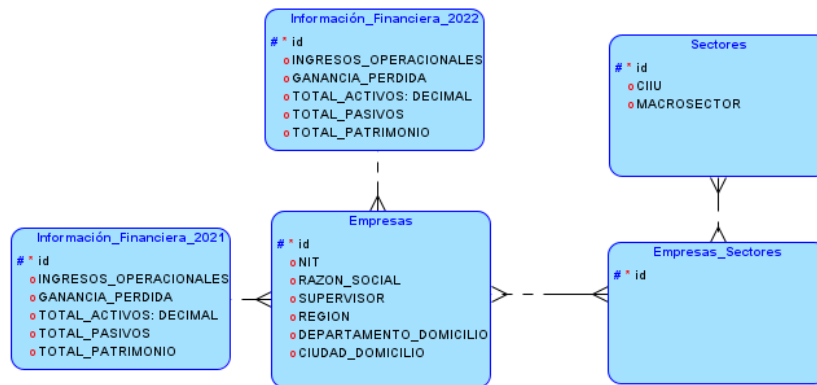
Tiene un diseñador grafico de bases de datos.

Se maneja en lenguaje PL/SQL.

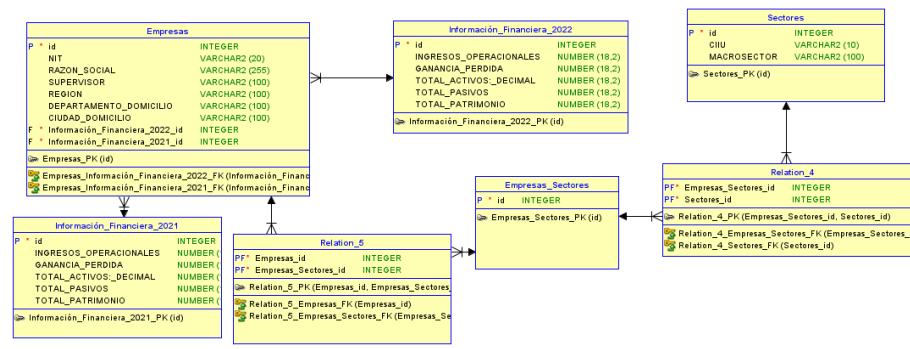
Es de libre acceso.

Cuenta con todo lo requerido por la base de datos relacional: Constraints, triggers, configuración de llaves únicas y llaves foraneas.

4.2 Diagrama modelo de datos



4.3 Imágenes de la Base de Datos



4.4 Código SQL - Lenguaje de definición de datos (DDL)

```

CREATE TABLE empresas (
  id INTEGER NOT NULL,
  nit VARCHAR2(20),
  razon_social VARCHAR2(255),
  supervisor VARCHAR2(100),
  region VARCHAR2(100),
  departamento_domicilio VARCHAR2(100),
  ciudad_domicilio VARCHAR2(100),
  informacion_financiera_2022_id INTEGER NOT NULL,
  informacion_financiera_2021_id INTEGER NOT NULL

```



```

);

ALTER TABLE empresas
    ADD CONSTRAINT empresas_pk PRIMARY KEY (id);

-- Corregir nombres de las claves foráneas
ALTER TABLE empresas
    ADD CONSTRAINT emp_info_fin_2021_fk
    FOREIGN KEY (informacion_financiera_2021_id)
    REFERENCES informacion_financiera_2021 (id);

ALTER TABLE empresas
    ADD CONSTRAINT emp_info_fin_2022_fk
    FOREIGN KEY (informacion_financiera_2022_id)
    REFERENCES informacion_financiera_2022 (id);

CREATE TABLE sectores (
    id            INTEGER NOT NULL,
    ciu           VARCHAR2(10),
    macrosector   VARCHAR2(100)
);

ALTER TABLE sectores
    ADD CONSTRAINT sectores_pk PRIMARY KEY (id);

CREATE TABLE informacion_financiera_2021 (
    id                INTEGER NOT NULL,
    ingresos_operacionales  NUMBER(18, 2),
    ganancia_perdida       NUMBER(18, 2),
    total_activos          NUMBER(18, 2),
    total_pasivos          NUMBER(18, 2),
    total_patrimonio       NUMBER(18, 2)
);

ALTER TABLE informacion_financiera_2021
    ADD CONSTRAINT info_fin_2021_pk PRIMARY KEY (id);

CREATE TABLE informacion_financiera_2022 (
    id                INTEGER NOT NULL,
    ingresos_operacionales  NUMBER(18, 2),
    ganancia_perdida       NUMBER(18, 2),
    total_activos          NUMBER(18, 2),
    total_pasivos          NUMBER(18, 2),
    total_patrimonio       NUMBER(18, 2)
);

```

```

ALTER TABLE informacion_financiera_2022
    ADD CONSTRAINT info_fin_2022_pk PRIMARY KEY (id);

CREATE TABLE empresas_sectores (
    id INTEGER NOT NULL
);

ALTER TABLE empresas_sectores
    ADD CONSTRAINT empresas_sectores_pk PRIMARY KEY (id);

CREATE TABLE relation_4 (
    empresas_sectores_id INTEGER NOT NULL,
    sectores_id          INTEGER NOT NULL
);

ALTER TABLE relation_4
    ADD CONSTRAINT relation_4_pk PRIMARY KEY (empresas_sectores_id, sectores_id);

ALTER TABLE relation_4
    ADD CONSTRAINT rel_4_emp_sect_fk
    FOREIGN KEY (empresas_sectores_id)
    REFERENCES empresas_sectores (id);

ALTER TABLE relation_4
    ADD CONSTRAINT rel_4_sect_fk
    FOREIGN KEY (sectores_id)
    REFERENCES sectores (id);

CREATE TABLE relation_5 (
    empresas_id          INTEGER NOT NULL,
    empresas_sectores_id INTEGER NOT NULL
);

ALTER TABLE relation_5
    ADD CONSTRAINT relation_5_pk PRIMARY KEY (empresas_id, empresas_sectores_id);

ALTER TABLE relation_5
    ADD CONSTRAINT rel_5_emp_fk
    FOREIGN KEY (empresas_id)
    REFERENCES empresas (id);

ALTER TABLE relation_5
    ADD CONSTRAINT rel_5_emp_sect_fk
    FOREIGN KEY (empresas_sectores_id)
    REFERENCES empresas_sectores (id);

```

4.5 Código SQL - Manipulación de Datos (DML)

4.6 Código SQL + Resultados: Vistas

```
CREATE VIEW vista_empresas_fin_2022 AS
SELECT
    e.id AS empresa_id,
    e.nit,
    e.razon_social,
    e.supervisor,
    e.region,
    e.departamento_domicilio,
    e.ciudad_domicilio,
    f2022.ingresos_operacionales,
    f2022.ganancia_perdida,
    f2022.total_activos,
    f2022.total_pasivos,
    f2022.total_patrimonio
FROM
    empresas e
INNER JOIN
    informacion_financiera_2022 f2022
    ON e.informacion_financiera_2022_id = f2022.id;

CREATE VIEW vista_empresas_por_sector AS
SELECT
    e.id AS empresa_id,
    e.nit,
    e.razon_social,
    s.ciiu AS sector_ciiu,
    s.macrosector
FROM
    empresas e
INNER JOIN
    relation_5 r5
    ON e.id = r5.empresas_id
INNER JOIN
    empresas_sectores es
    ON r5.empresas_sectores_id = es.id
INNER JOIN
    relation_4 r4
    ON es.id = r4.empresas_sectores_id
INNER JOIN
    sectores s
    ON r4.sectores_id = s.id;

CREATE VIEW vista_financiera_resumida AS
```

```

SELECT
    e.id AS empresa_id,
    e.nit,
    e.razon_social,
    f2021.ingresos_operacionales AS ingresos_2021,
    f2021.ganancia_perdida AS ganancia_perdida_2021,
    f2021.total_activos AS total_activos_2021,
    f2021.total_pasivos AS total_pasivos_2021,
    f2021.total_patrimonio AS total_patrimonio_2021,
    f2022.ingresos_operacionales AS ingresos_2022,
    f2022.ganancia_perdida AS ganancia_perdida_2022,
    f2022.total_activos AS total_activos_2022,
    f2022.total_pasivos AS total_pasivos_2022,
    f2022.total_patrimonio AS total_patrimonio_2022
FROM
    empresas e
INNER JOIN
    informacion_financiera_2021 f2021
    ON e.informacion_financiera_2021_id = f2021.id
INNER JOIN
    informacion_financiera_2022 f2022
    ON e.informacion_financiera_2022_id = f2022.id;

```

4.7 Código SQL + Resultados: Triggers

```

-- Trigger para registrar inserciones en la tabla empresas
CREATE OR REPLACE TRIGGER trg_empresas_insert
AFTER INSERT ON empresas
FOR EACH ROW
BEGIN
    INSERT INTO log_cambios (
        id,
        tabla,
        accion,
        fecha,
        id_registro
    ) VALUES (
        log_cambios_seq.NEXTVAL,
        'empresas',
        'INSERT',
        SYSDATE,
        :NEW.id
    );
END;

-- Trigger para validar ingresos operacionales en informacion_financiera_2022

```

```
CREATE OR REPLACE TRIGGER trg_validar_ingresos
BEFORE INSERT OR UPDATE ON informacion_financiera_2022
FOR EACH ROW
BEGIN
    IF :NEW.ingresos_operacionales < 0 THEN
        RAISE_APPLICATION_ERROR(-20001, 'Ingresos operacionales no pueden ser negativos');
    END IF;
END;
```

4.8 Código SQL + Resultados: Funciones

4.9 Código SQL + Resultados: Procedimientos Almacenados

5 Bases de Datos No-SQL (*Segunda entrega*)

5.1 Diagrama Bases de Datos No-SQL (*Segunda entrega*)

5.2 SMBD utilizado para la Base de Datos No-SQL (*Segunda entrega*)

- 6 Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos** (*Tercera entrega*)
 - 6.1 Ejemplo de aplicación de ETL y Bodega de Datos** (*Tercera entrega*)
 - 6.2 Automatización de Datos** (*Tercera entrega*)
 - 6.3 Integración de Datos** (*Tercera entrega*)

7 Proximos pasos (*Tercera entrega*)

8 Lecciones aprendidas (*Tercera entrega*)

9 Bibliografía