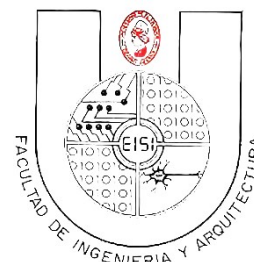




**UNIVERSIDAD DE EL SALVADOR**  
**FACULTAD DE INGENIERÍA Y ARQUITECTURA**  
**ESCUELA DE INGENIERÍA DE SISTEMAS**  
**INFORMÁTICOS**



**DISEÑO INSTRUCCIONAL CURSO DE ESPECIALIZACION**

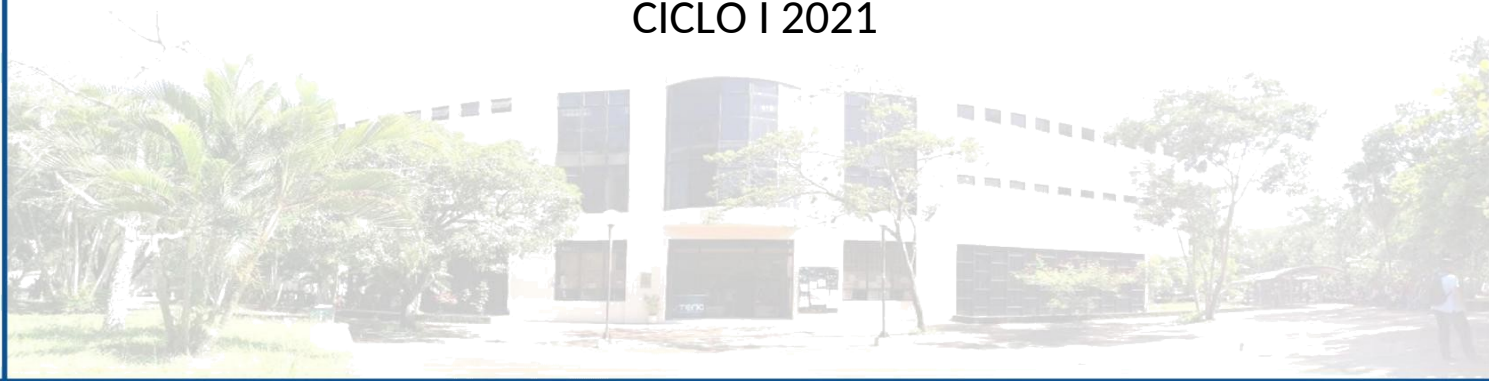
**ESPECIALIZACIÓN EN**  
**INGENIERÍA DE DATOS**

Código:

8 Unidades valorativas

Facilitador: Marlon Menjivar

CICLO I 2021



# UNIVERSIDAD DE EL SALVADOR – FIA – EISI

## DISEÑO INSTRUCCIONAL - CURSO DE ESPECIALIZACIÓN

### Contenido

1.	Descripción del curso:.....	1
2.	Objetivos:.....	1
2.1	Módulo 1: Datawarehousing.....	1
2.2	Módulo 2: Introducción al Big Data y Cloud Computing.....	1
2.3	Módulo 3: Procesamiento Masivo en Paralelo con Spark.....	2
2.4	Módulo 4: Visualización de Datos.....	2
3.	Competencias a desarrollar.....	2
4.	Duración y horarios.....	3
5.	Contenido temático.....	3
6.	Requisitos para optar al curso.....	6
7.	Metodología de la enseñanza y aprendizaje.....	6
7.1	Contenidos teóricos.....	6
7.2	Clases magistrales.....	6
7.3	Talleres.....	6
7.4	Exámenes Parciales.....	6
7.5	Trabajo de aplicación.....	7
7.6	Consultas.....	7
8.	Consideraciones importantes.....	7
8.1	Generales.....	7
8.2	Sobre la elaboración de trabajos o tareas evaluados.....	7
9.	Sistema de evaluación.....	8
10.	Actividades evaluadas.....	8
10.1	Exámenes parciales.....	8
10.2	Tareas.....	9
8.3	Trabajo final.....	9
11.	Ficha resumen de actividades evaluadas.....	10
12.	Comunicación en la asignatura.....	10
12.1	Comunicación con su profesor.....	10
12.2	Comunicación con coordinador de especialización.....	10
13.	Bibliografía recomendada.....	11
14.	Servicios de biblioteca.....	11

# UNIVERSIDAD DE EL SALVADOR – FIA - EISI

## DISEÑO INSTRUCCIONAL - CURSO DE ESPECIALIZACIÓN

### 1. Descripción del curso:

**Nombre el curso:** Especialización en Ingeniería de Datos

**Modalidad:** Semipresencial

**Cupo:** 20

#### **Justificación:**

La ingeniería de datos es la disciplina que se encarga de recolectar, trasladar y validar datos para su explotación. Utiliza principios de ingeniería de software para construir modelos de datos como almacenes de datos (Datawarehouse) y lagos de datos (Data Lake) garantizando la disponibilidad, consistencia, mantenibilidad, seguridad y limpieza de datos para que estos puedan ser utilizados.

Con la explosión de Big Data en el mundo actual, se vuelve necesario contar con métodos sofisticados y eficientes de manejo de grandes volúmenes de datos. Las clásicas bases de datos relacionales cada vez se ven más limitadas especialmente ante datos no estructurados por lo que un sinnúmero de nuevas tecnologías han surgido para llenar estos vacíos.

Un ingeniero de datos se especializa en estas herramientas y se vale del dominio del negocio o industria en la que trabaje, para volver a los datos, uno de los activos más valiosos que una organización puede tener.

### 2. Objetivos:

#### 2.1 Módulo 1: Datawarehousing

- 1) Que el estudiante comprenda las diferencias entre un modelo de base de datos diseñado para alta transaccionalidad y otro diseñado para analítica.
- 2) Explicar las características, componentes y funciones principales para el diseño, modelado y mantenimiento de un Data Warehouse.
- 3) Aplicar técnicas de carga de datos históricos e incrementales para las diferentes estructuras de un data warehouse.
- 4) Que el estudiante aprenda técnicas y desarrolle criterios de diseño de estructuras multidimensionales.

#### 2.2 Módulo 2: Introducción al Big Data y Cloud Computing

- 1) Que el estudiante comprenda el concepto de Big Data, sus características y como este nuevo paradigma de entender los datos ha generado la explosión de nuevas tecnologías.
- 2) Explicar el concepto, utilidad y técnicas de implementación de Data Lakes.

# UNIVERSIDAD DE EL SALVADOR – FIA - EISI

## DISEÑO INSTRUCCIONAL - CURSO DE ESPECIALIZACIÓN

- 3) Que el estudiante adquiera conocimientos teóricos y prácticos acerca de computación en la nube en Amazon Web Services (AWS).
- 4) Utilizar componentes en AWS para crear un data lake.

### 2.3 Módulo 3: Procesamiento Masivo en Paralelo con Spark

- 1) Explicar el concepto de procesamiento en paralelo, escalado horizontal y clústeres spark.
- 2) Que el estudiante adquiera habilidades técnicas en el lenguaje Scala, comprendiendo el paradigma de programación funcional.
- 3) Que el estudiante sea capaz de escribir y mantener series de transformaciones de datos utilizando spark
- 4) Explicar los componentes de Amazon Web Services que permiten utilizar la potencia de spark en computación en la nube.
- 5) Explicar el concepto de base de datos columnar e introducir al estudiante a Amazon Redshift.
- 6) Construir un modelo de datos multidimensional haciendo uso de Data Lakes, Procesamiento Masivo en paralelo en Spark y utilizando Redshift como base de datos analítica.

### 2.4 Módulo 4: Visualización de Datos

- 1) Que el estudiante adquiera aptitudes y habilidades de construcción de reportes y dashboards utilizando Power BI
- 2) Desarrollar conocimientos técnicos acerca de consumir datos de diversas fuentes desde Power BI

## 3. Competencias a desarrollar

- 1) Técnicas de modelado de bases de datos con propósitos analíticos
- 2) Habilidades en computación en la nube en Amazon Web Services
- 3) Construcción de Data Lakes
- 4) Programación funcional con Scala
- 5) Utilización de Spark para transformaciones de datos masivos
- 6) Presentación de datos en Power BI

# UNIVERSIDAD DE EL SALVADOR – FIA - EISI

## DISEÑO INSTRUCCIONAL - CURSO DE ESPECIALIZACIÓN

### 4. Duración y horarios

La especialización será desarrollada a lo largo de 9 meses distribuyendo el tiempo de acuerdo con la siguiente tabla:

N°	Modulo	Duración
1	Datawarehousing	3 meses
2	Introducción al Big Data y Cloud Computing	1 mes
3	Procesamiento Masivo en Paralelo con Spark	4 meses
4	Visualización de Datos	1 mes

Los horarios de clase serán los martes y jueves de 6:30pm a 8:00pm en los cuales se variará dependiendo de la programación entre clases teóricas y clases prácticas.

Los sábados se habilitará un espacio de consultas con horario de 3pm a 4pm con el objetivo de brindar soporte a los estudiantes, especialmente en el desarrollo del trabajo final. El espacio de consulta deberá ser reservado con anticipación de al menos un día y la disponibilidad será sujeta a demanda.

### 5. Contenido temático

Mes	Semana	Contenidos	Metodología	Duración	Evaluaciones	Ponderación
<b>MODULO 1: DATAWAREHOUSING</b>						
1 (24h)	1	Comparación Bases de Datos Transaccionales OLTP vs Bases de Datos Analíticas OLAP.  Objetivos de Datawarehousing y Business Intelligence	Clase magistral/ Recurso didáctico	6 h		
	2	Introducción a diseño dimensional	Clase magistral/ Recurso didáctico	6 h		
	3	Arquitectura Kimball	Clase magistral/ Recurso didáctico	6 h		
	4	Arquitecturas DW alternativas	Clase magistral/ Recurso didáctico	6 h		
2	5	Metodología para	Clase	6 h		

**UNIVERSIDAD DE EL SALVADOR – FIA - EISI**  
**DISEÑO INSTRUCCIONAL - CURSO DE ESPECIALIZACIÓN**

(24h)		diseño de DW	magistral/ Recurso didáctico			
	6	Tipos de Fact Tables	Clase magistral/ Recurso didáctico	6 h		
	7	Tipos de dimensiones	Clase magistral/ Recurso didáctico	6 h		
	8	Técnicas de modelado de Fact Tables	Clase magistral/ Recurso didáctico	6 h		
3 (22h)	9	Técnicas de modelado de Dimensiones	Clase magistral/ Recurso didáctico	6 h		
	10	Caso de estudio, diseño completo de un modelo	Clase magistral/ Recurso didáctico	4 h		
	11	Diseño de ETL	Clase magistral/ Recurso didáctico	6 h		
	12	Cargas Full e Incrementales	Clase magistral/ Recurso didáctico	6 h	Parcial 1	20%

**MODULO 2: INTRODUCCIÓN AL BIG DATA Y CLOUD COMPUTING**

4 (20 h)	13	Introducción a Amazon Web Services	Clase magistral/ Recurso didáctico	6 h		
	14	Data Lakes vs DW	Clase magistral/ Recurso didáctico	6 h		
	15	Consola AWS y servicios relacionados a datos	Práctica	4 h		
	16	Consola AWS y servicios relacionados a datos	Práctica	4 h	Tarea Fase Inicial	5%
					Parcial 2	15%

**MODULO 3: PROCESAMIENTO MASIVO EN PARALELO CON SPARK**

**UNIVERSIDAD DE EL SALVADOR – FIA - EISI**  
**DISEÑO INSTRUCCIONAL - CURSO DE ESPECIALIZACIÓN**

5 (24h)	17	Introducción a Scala	Taller/ Recurso didáctico	6h		
	18	Introducción a Scala	Taller/ Recurso didáctico	6 h		
	19	Introducción a Scala	Taller/ Recurso didáctico	6 h		
	20	Arquitectura Spark	Clase magistral/ Recurso didáctico	6h		
6 (22h)	22	Instalando spark en local	Taller/ Recurso didáctico	4 h		
	23	Spark SQL, RDDs, Dataframes y Datasets	Clase magistral/ Recurso didáctico	6 h		
	24	Spark SQL, RDDs, Dataframes y Datasets	Clase magistral/ Recurso didáctico	6 h		
	25	Joins y transformaciones	Clase magistral/ Recurso didáctico	6h		
7 (16h)	22	Importando Datos a S3	Taller	4 h		
	23	AWS EMR	Taller	4 h		
	24	Jupyter Notebook Scripting	Taller	4 h		
	25	Implementación de ETL en Jupyter	Taller	4 h		
8 (16h)	22	Automatización de ETL en AWS	Taller	4 h		
	23	AWS Pipelines	Taller	4 h		
	24	AWS Glue	Taller	4 h		
	25	Alternativas a automatización	Clase magistral	4h	Parcial 3	30%
<b>MODULO 4: VISUALIZACIÓN DE DATOS</b>						
9 (16h)	22	Importancia de visualización de datos	Clase magistral	4 h		
	23	Tour Power BI y servicios	Clase magistral	4 h		
	24	Importando diversas fuentes de datos	Taller	4 h		

# UNIVERSIDAD DE EL SALVADOR – FIA - EISI

## DISEÑO INSTRUCCIONAL - CURSO DE ESPECIALIZACIÓN

	25	Creación de dashboards	Taller	4h	Parcial 4	15%
					Tarea Final	15%

### 6. Requisitos para optar al curso

- 1) Diseño de base de datos relacionales
- 2) Conocimientos avanzados de SQL
- 3) Conocimientos de Oracle
- 4) Conocimientos de PLSQL
- 5) Proyectos Maven
- 6) Dominio de Java SE
- 7) Dominio de Python
- 8) Versionado de código con GIT
- 9) Aprobar el examen de admisión al curso que se desarrollará en la primera semana del mismo

### 7. Metodología de la enseñanza y aprendizaje

La asignatura se impartirá semipresencial, se presenta un panorama general de los componentes de la metodología de enseñanza y aprendizaje:

#### 7.1 Contenidos teóricos

La especialización está conformada por 4 módulos, cada uno de ellos contará con material específico en donde se expandirán los temas desarrollados en las clases y además se incluirán materiales adicionales de lectura obligatoria, como documentaciones oficiales de tecnologías utilizadas o artículos relacionados a la temática en desarrollo.

#### 7.2 Clases magistrales

Clases expositivas en las que el docente desarrollará contenidos teóricos o metodológicos acerca de los conceptos en desarrollo. Serán complementadas con materiales teóricos y podrán ser virtuales o presenciales de acuerdo con las indicaciones del docente.

#### 7.3 Talleres

Sesiones desarrolladas por el docente en donde se resolverán ejercicios, implementarán configuraciones y se aplicarán los conceptos teóricos estudiados.

#### 7.4 Exámenes Parciales

Se realizará un examen parcial por módulo, 4 en total. Contendrán una parte teórica y otra parte práctica y podrán ser realizados en línea o de manera presencial de acuerdo con las indicaciones del docente.



### 7.5 Trabajo de aplicación

La fase inicial consistirá en la propuesta de un modelo dimensional elaborado tomando como base un modelo relacional que el estudiante elija, pudiendo ser este didáctico o una implementación real. La fase final será la implementación técnica del modelo utilizando tecnologías cubiertas en la especialización o tecnologías que el estudiante considere aplicables.

### 7.6 Consultas

Los sábados se habilitará un espacio de consultas con horario de 3pm a 4pm con el objetivo de brindar soporte a los estudiantes, especialmente en el desarrollo del trabajo final. El espacio de consulta deberá ser reservado con anticipación de al menos un día y la disponibilidad será sujeta a demanda.

## 8. Consideraciones importantes

### 8.1 Generales

La especialización comprenderá 4 módulos enfocados en un área específica de ingeniería de datos. Las evaluaciones parciales se realizarán una por cada módulo finalizado, en el periodo establecido y con una ponderación establecida en la especificación de las evaluaciones.

En adición a las evaluaciones parciales, el estudiante desarrollará un proyecto conformado por dos fases, la primera a ejecutarse inmediatamente después de finalizar el primer módulo en la que se planteará un modelo dimensional. La segunda fase se ejecutará a partir del penúltimo mes de la especialización y comprenderá la implementación y presentación del modelo especificado en la primera fase.

### 8.2 Sobre la elaboración de trabajos o tareas evaluados

Es importante que al momento en que elabore informes de trabajos o tareas como parte de las actividades evaluadas de esta y otras asignaturas no se agregue información extraída de Internet u otra fuente sin que se anote la procedencia, ya que la ausencia de esa información implicaría una falta.

Trabajos o tareas con Plagio: Si se detecta que el trabajo o tarea que se presenta es copia textual ya sea en su totalidad o bien de algunos párrafos o se toman de ideas de otros autores sin darse el crédito correspondiente, al no anotarse la fuente donde fueran extraídas, en forma automática el profesor asignará al trabajo o tarea “un cero” al apartado que evidencie el plagio.

No se aceptarán trabajos o tareas en forma extemporánea, si el/la estudiante previa a la presentación de la misma, no presenta una justificación escrita ante la Coordinación de la Cátedra, en la que explique el motivo de su incumplimiento con la entrega del trabajo o tarea, así como demostrarlo con el dictamen profesional correspondiente que autentique la misma.

# UNIVERSIDAD DE EL SALVADOR – FIA - EISI

## DISEÑO INSTRUCCIONAL - CURSO DE ESPECIALIZACIÓN

Los trabajos o tareas deben ser individuales y originales no se aceptan trabajos o tareas en parejas o en grupos (excepto autorización expresa del encargado de la cátedra). Si se detecta copia del trabajo o tarea entre dos o más estudiantes, o se comprueba que el trabajo o tarea se obtuvo de otra fuente y que no se hacen las citas bibliográficas respectivas, el tutor respectivo asignará “un cero” al apartado que evidencia el plagio.

### 9. Sistema de evaluación

El sistema de evaluación que se empleará en la asignatura es el siguiente:

Etapa	Aspecto a Evaluar	Descripción	Ponderación
1	Propuesta de perfil de trabajo de aplicación de acorde al módulo técnico de especialización elegida	El estudiante planteará la propuesta de un trabajo de aplicación el cual deberá de presentar luego de haber finalizado los cursos de especialización inscritos	5%
2	Sumatoria de las evaluaciones programadas en cada módulo o semanas del curso	Exámenes parciales por módulo	80%
3	Trabajo final	Trabajo de aplicación	15%

### 10. Actividades evaluadas

A continuación, pueden verse las actividades evaluadas que se desarrollarán en la especialización. Se detalla la descripción, así como las fechas de asignación y entrega de estas.

#### 10.1 Exámenes parciales

	<b>Ponderación: 20%</b>
<b><u>Parcial 1</u></b>	Módulo 1: Datawarehousing
<b><u>Fecha</u></b> Lunes 31 de mayo 2021	
	<b>Ponderación: 15%</b>
<b><u>Parcial 2</u></b>	Módulo 2: Introducción al Big Data y Cloud Computing
<b><u>Fecha</u></b> Miércoles 30 de junio 2021	
	<b>Ponderación: 30%</b>
<b><u>Parcial 3</u></b>	Módulo 3: Programación Masiva en Paralelo con Spark
<b><u>Fecha</u></b>	

# UNIVERSIDAD DE EL SALVADOR – FIA - EISI

## DISEÑO INSTRUCCIONAL - CURSO DE ESPECIALIZACIÓN

Viernes 29 de octubre 2021	
	<b>Ponderación: 15%</b>
<b>Parcial 4</b>	Módulo 4: Visualización de Datos
<b>Fecha</b>	
Martes 30 de noviembre 2021	

### 10.2 Tareas

Tarea 1	Ponderación: 5%
<p><b><u>Descripción: Diseño de un modelo multidimensional para soportar un proceso de negocio seleccionado.</u></b></p> <p>Se formarán grupos de 2 o 3 estudiantes, elegirán un modelo de base de datos que podría ser un esquema de ejemplo de Oracle o un esquema real proporcionado por una empresa. Basado en lo anterior se seleccionará un proceso de negocio y se realizará el diseño multidimensional que soporte el análisis de este.</p>	<p><b><u>Indicaciones de entrega:</u></b> Los entregables serán el diagrama del modelo, creado en algún software seleccionado por los estudiantes.</p> <p>Un documento explicando la lógica, reglas de negocio y consideraciones generales aplicados a dimensiones, y fact tables.</p> <p>Pasada la entrega se programarán defensas del modelo en donde el docente sugerirá mejoras al diseño de ser necesario.</p>
<p><b><u>Fecha de entrega:</u></b> Miércoles 30 de Junio 2021</p>	

### 8.3 Trabajo final

Trabajo de aplicación	Ponderación: 15%
<p><b><u>Descripción: Creación de modelo dimensional utilizando Amazon Web Services.</u></b></p> <p>Manteniendo de ser posible, los grupos de la fase inicial de diseño; los alumnos implementarán el modelo dimensional utilizando S3 como data lake, Spark como motor de transformación de datos y Redshift para presentar el modelo implementado.</p> <p>Finalmente deberán crear un tablero en Power BI mostrando datos interesantes encontrados en el modelo creado</p>	<p><b><u>Indicaciones de entrega:</u></b> Los entregables serán un repositorio en GitHub conteniendo todos los scripts de spark utilizados para ETL, scripts de creación de tablas en redshift y demás elementos de código que el grupo considere necesario.</p> <p>El repositorio deberá contener documentación del producto en el README del repositorio indicando el modelo de datos implementado, reglas de negocio, tipos de dimensiones implementadas, entre otros.</p> <p>Además, se deberá especificar los componentes de AWS utilizados.</p>
<p><b><u>Fecha de entrega:</u></b> Viernes 10 de Diciembre 2021</p>	

## **11. Ficha resumen de actividades evaluadas**

Actividad	Ponderación	Fecha de Asignación	Fecha de Entrega
Tarea Fase Inicial	5	Martes 1 de junio de 2021	Miércoles 30 de junio 2021
Tarea Fase Final	15	Lunes 1 de noviembre 2021	Viernes 10 de diciembre 2021
Examen Parcial 1	20	Lunes 31 de mayo 2021	
Examen Parcial 2	15	Miércoles 30 de junio 2021	
Examen Parcial 3	30	Viernes 29 de octubre 2021	
Examen Parcial 4	15	Martes 30 de noviembre 2021	

## **12. Comunicación en la asignatura**

En su avance en la asignatura será de mucha importancia que usted mantenga una constante comunicación con su profesor. A continuación, se presentan los medios de comunicación de los cuales usted podrá disponer, es importante que tome en cuenta las indicaciones y sugerencias que en este apartado se realizan a fin de poder hacer uso adecuado de los mismos.

### **12.1 Comunicación con su profesor**

#### **12.1.1 Plataforma (aula virtual):**

Es el medio de comunicación primordial para exponer dudas y/o comentarios de la asignatura. Único medio oficial para adjuntar tareas, realizar evaluaciones, consulta y visualización de los contenidos académicos, recepción de indicaciones diversas, entre otras; salvo que por motivos de fuerza mayor la plataforma no se encuentre disponible, en cuyo caso se expondrán en los medios oficiales como la página en Facebook de Educación a Distancia y página web de Educación a Distancia, las indicaciones pertinentes a la asignatura se harán por los medios antes escritos.

#### **12.1.2 Correo electrónico:**

Se utilizará como herramienta de apoyo de comunicación en caso que la interacción del tutor y estudiante por medio de la plataforma sea inaccesible.

#### **12.1.3 Otros medios de comunicación:**

Otros medios como redes sociales o teléfono no son válidos en la resolución o mediación de problemas académicos administrativos, tampoco para realizar consultas académicas o recibir asesorías. Sin embargo, no se descarta que sean medios utilizados para dar anuncios e información de carácter urgente.

## 12.2 Comunicación con coordinador de Modalidad de especialización

### 10.2.1 Correo electrónico:

El Coordinador de la modalidad no atenderá consultas académicas, sino aquellas consultas u observaciones que usted crea que es importante hacerle saber sobre el desempeño anómalo del curso. Los contactos del coordinador de la modalidad son:

Coordinador de Cátedra: MSc. Ing. Patricia Estrada

Correo electrónico: [pestrada@fia.ues.edu.sv](mailto:pestrada@fia.ues.edu.sv)

## 13. Bibliografía recomendada

- Kimball, R., & Ross, M. (2013). The data warehouse toolkit (3rd ed.). Wiley.
- Gupta, S., & Giri, V. (2018). Practical enterprise data lake insights. Distributed to the Book trade worldwide by Springer Science+Business Media New York.
- García Márquez, F., & Lev, B. (2017). *Big data management*. Springer International.
- Kienzler, R. (2017). Mastering Apache Spark 2.x - Second Edition. Packt Publishing.