

Programa del Curso

Información General

Profesores	Correo Electrónico	Atención a estudiantes
Cristian Camilo Castellanos	cc.castellanos87@uniandes.edu.co	Solicitar cita por correo electrónico
Gabriel Paredes	gjparedest@gmail.com	Solicitar cita por correo electrónico

Descripción

La adopción de computación en la nube continúa creciendo en una variedad de organizaciones y dominios y han permitido la evolución y escalamiento en diversas áreas, especialmente en soluciones de datos. Esta adopción implica entender los cambios arquitectónicos, funcionales, de seguridad y gobierno junto con el modelo de costos para tomar decisiones bien informadas que permiten aprovechar las ventajas específicas de la nube como agilidad, elasticidad, escalabilidad, y resiliencia. Este curso se enfoca en el diseño y evaluación de soluciones de datos que van desde almacenamiento, transaccionalidad, seguridad y gobierno, como sus usos en analítica básica y avanzada. Una solución de datos en la nube bien diseñada, parte de la clara definición de las necesidades de negocio, casos de uso y su traducción en criterios de diseño, selección de servicios, tecnologías y configuraciones específicas para dar respuesta a dichas necesidades. Este curso aproxima a los estudiantes al marco conceptual y práctico de diseñar soluciones de datos que respondan a las necesidades actuales y futuras de las organizaciones y cómo las tecnologías de nube son un habilitador y potenciador para abordarlas.

El temario está compuesto por cuatro módulos:

Módulo # 1 - Conceptos básicos, tipos de almacenamiento, seguridad y gobierno de datos en la nube.

Módulo # 2 - Soluciones de bases de datos SQL, No-SQL y de propósito específico.

Módulo # 3 - Ingeniería de datos, soluciones de ETL, analítica básica y procesamiento de big data.

Módulo # 4 - Soluciones de analítica avanzada, aprendizaje automático e inteligencia artificial.

Objetivos Pedagógicos

- Identificar los conceptos básicos de computación en la nube y las estrategias de datos modernas.
- Identificar y aplicar los componentes de ciberseguridad de información en nube.
- Analizar las ventajas estratégicas que motivan el diseño y despliegue de soluciones de datos y analítica sobre nube pública.
- Evaluar las necesidades, motivadores y restricciones que conducen el diseño de soluciones de datos en nube.
- Identificar y aplicar los patrones de soluciones de datos y analítica de nube a partir de los requerimientos de negocio y TI.
- Diseñar soluciones de datos de manera práctica sobre un servicio de nube pública.
- Identificar los elementos claves de consumo y estimación de costos asociados a una solución de datos en nube.
- Evaluar las diferentes estrategias de migración y modernización de soluciones de datos y analítica en la nube a partir de sus características claves.

- Entender y evaluar la factibilidad, calidad, y buenas prácticas de soluciones de datos en nube.

Plan de Temas

El curso se desarrolla en una introducción más cuatro módulos. En cada uno de ellos se proponen un conjunto de patrones agnósticos y algunas de sus implementaciones que buscan contextualizar al estudiante y aplicar las temáticas vistas.

Módulo	Temas
M1: Conceptos básicos, almacenamiento, gobierno y seguridad en la nube.	1- Conceptos básicos de nube pública y arquitectura de información. 2- Tipos de almacenamiento y transferencia de datos. 3- Gobierno y ciberseguridad de datos en la nube 4- Práctica: almacenamiento, gobierno y seguridad, ejercicio de costeo.
M2: Servicios de bases de datos SQL y No-SQL	1- Base de datos transaccionales administradas y serverless 2- Bases de datos No-SQL y de propósito específico (IoT, Blockchain) 3- Rompiendo el monolito de datos (Lectura) 4- Práctica: Bases datos SQL y No-SQL (DynamoDB, IoT)
M3: Servicios de analítica y procesamiento de grandes volúmenes de datos	1- Soluciones de ingestión de datos, ETL, Ingeniería de datos, data streaming 2- Soluciones de analítica datos (DWH, Data Lake, Data Lake House, Data Mesh) 3- Grandes volúmenes de datos y escalabilidad (Patrones Lambda, Kappa, in-memory, motores de indexación) 4- Soluciones de explotación de datos (BI, APIs, Consultas Ad-hoc) 5- Práctica: ETL, DWH, Data Lake, BI
M4: Servicios de aprendizaje automático e inteligencia artificial.	1- Conceptos básicos de aprendizaje automático (ML), Infraestructura para ML 2- Soluciones de desarrollo de ML, democratización, análisis de sesgo. 3- ML como servicio: transcripción, NLU, computación visual, pronóstico, GenAI. 4- Práctica: Desarrollo de ML, ML como servicio y GenAI

Evaluación del curso

Exámenes	Descripción	%	Fecha
Parcial 1		20%	
Parcial 2		20%	
Quices y Presentaciones		10%	
Total Exámenes		50%	

Proyecto	Descripción	%	Fecha
Entrega 1		20%	
Entrega 2		20%	
Presentación final		10%	
Total proyecto		50%	

Política de manejo de notas

- No existe aproximación automática en la nota definitiva
- Las notas de las evaluaciones de los diferentes módulos se califican entre 1.50 y 5.00 con dos decimales y no hay aproximaciones
- Para pasar el curso es indispensable lograr en el puntaje ponderado 3.00/5.00

Cronograma

S	Fecha	Tema	Profesor
1	05/08	Presentaciones. M1 . Intro, Conceptos básicos de nube pública y arquitectura de información. arquitecturas de referencia Evaluación de soluciones PoC y Well-Arch Frameworks	GP+CC
2	12/08	M1 . Soluciones de almacenamiento en la nube y transferencia de datos, Costeo de soluciones. Gobierno de datos y ciberseguridad en la nube	CC
3	24/08 (Sab)	M1. Lab Virtual : Almacenamiento objetos, Sistema de archivos, seguridad, y ejercicio de costeo	CC
4	26/08	M2 . ACID, Base de datos transaccionales: administradas y serverless, HA, Réplicas Lectura - Rompiendo el monolito de datos	GP
5	02/09	M2 . Teorema CAP, Bases de datos No-SQL y de propósito específico (IoT, Blockchain, in-memory DBs)	GP
6	09/09	M2. Lab Virtual : Bases datos SQL y No-SQL	GP
7	16/09	Parcial 1 (M1+M2)	
-	20/09	Entrega 1	
8	23/09	Sustentación Entrega 1	GP+CC
9	7/10	M3 . Ingeniería de datos. Calidad de datos. ETL, Orquestación, Data streaming, DWH, BI	CC
10	19/10 (Sab)	M3 . Estrategia de datos y analítica moderna Data Lake, Data Lake House, Data Mesh, Patrones Big data	CC
11	21/10	M3. Lab Virtual Serverless Data Lake	CC
12	28/10	M4 . Conceptos Tipos AI/ML. AI Responsable, Sesgo, Concept Drift HW para ML, Soluciones desarrollo, evaluación, despliegue y monitoreo	GP
13	09/11 (Sab)	M4 . Democratización de AI, AutoML, ML/AI como servicio, Transcripción, OCR, LNU, Pronóstico, Computación Visual, GenAI	GP
14	16/11 (Sab)	M4. Lab Virtual : Desarrollo de ML, ML/AI como servicio y GenAI	GP
15	18/11	Parcial 2 (M3+M4)	
-	22/11	Entrega 2	
16	25/11	Sustentación Entrega 2	GP+CC

Bibliografía

- [1] Indrasiri, K., & Suhothayan, S. (2021). *Design patterns for cloud native applications: Patterns in practice using apis, data, events, and streams*. O'Reilly.
- [2] Erl, T., Puttini, R., & Mahmood, Z. (2013). *Cloud computing: concepts, technology, & architecture*. Pearson Education.
- [3] Wilder, B. (2012). *Cloud architecture patterns: using microsoft azure*. " O'Reilly Media, Inc."
- [4] Erl. T., Khattak, W., and Buhler, P. (2016) *Big Data Fundamentals: Concepts, Drivers & Techniques*. Pearson Service Technology Series
- [5] Godinez, M., Hechler, E., Koenig, K., Lockwood, S., Oberhofer, M., & Schroeck, M.J. (2010). *The Art of Enterprise Information Architecture: A Systems-Based Approach for Unlocking Business Insight*. IBM Press
- [6] W.H. Inmon, Daniel Linstedt, Mary Levins (2019). *Data Architecture: A Primer for the Data Scientist, 2nd Edition*. Academic Press

[7] Marz, N.. and Warren, J. (2015). *Big Data, Principles and best practices of scalable real-time data systems*. Manning Publications.

[8] Marco Tranquillin, Valliappa Lakshmanan, Firat Tekiner (2023). *Architecting Data and Machine Learning Platforms*.

[9] Huyen, C. (2022). *Designing machine learning systems*. O'Reilly Media, Inc

[10] Nauman Sheikh (2013) *Implementing Analytics: A Blueprint for Design, Development, and Adoption*. Morgan Kaufmann

[11] Kunigk, J., Buss, I., Wilkinson, P., & George, L. (2018). *Architecting modern data platforms: a guide to enterprise hadoop at scale*. O'Reilly Media.