

Técnicas Avanzadas de Repositorios para Analítica de Datos
Mgr. Cecilia Ana Ruz (Prof. Adjunta, Depto de Computación, FCEN-UBA) con la colaboración del
Esp. Sergio D'Arrigo (Prof. Adjunto, Depto de Computación, FCEN-UBA)

Programa:

La materia está orientada a que los estudiantes sean capaces de:

- Conocer la utilidad de los repositorios analíticos de datos y el rol de la analítica.
- Conocer los fundamentos de Data Warehousing, su concepción, diseño e implementación.
- Conocer las nuevas tendencias tecnológicas en la era del Big Data
- Resolver casos de uso concretos aplicando los conocimientos brindados de la materia

Temario:

Unidad 1: Datos y repositorios analíticos: Introducción. Importancia de los datos y rol de la analítica. Motivación y utilidad de los repositorios analíticos de datos. Clasificación de los tipos de datos. Evolución histórica de

Unidad 2: Data Warehousing: fundamentos, modelización y diseño: Conceptos Básicos de Data Warehousing. Arquitectura. Modelización y diseño. Linaje y gobierno de datos.

Unidad 3: Data Warehousing: implementación, carga y actualización: Implementación. Ingesta, almacenamiento y actualización. Integridad y calidad.

Unidad 4. Big Data y nuevos tipos de DWH: Variantes de data warehouses. Data warehouses y Big Data. Data Lakes, Data *lakehouses*. Estado del arte.

Unidad 5: Analítica y Data Mining: Introducción a la analítica. Tipos de analítica. Exploración, BI y Data Mining. La utilización de sandbox analíticos. Protección de datos personales.

Bibliografía:**Unidad 1**

- Silberschatz A., Korth H. & Sudarshan S. (2020). Database System Concepts (7ma. Ed.), Mc Graw Hill
- Kitchin R. (2014). The Data Revolution: Big Data, Open Data, Data Infrastructures & their consequences (1ra. Ed.). SAGE Publications Ltd
- DAMA International (2017). DAMA-DMBOK: Data Management Body of Knowledge (2nd Ed.). Technics Publications
- Vaisman, A. & Zimanyi, E. (2022). Data Warehouse Systems: Design and Implementation (2da. Ed.). Springer-Verlag GmbH
- Inmon W. (2005) Building the Data Warehouse (4ta. Ed.). Willey Publishing

Unidad 2

- Vaisman, A. & Zimanyi, E. (2022). Data Warehouse Systems: Design and Implementation (2da. Ed.). Springer-Verlag GmbH

- Golfarelli, M. & Rizzi, S.. (2009) Data Warehouse Design, Modern Principles and Methodologies (1ra. Ed.), McGraw-Hill
- Inmon W. (2005) Building the Data Warehouse (4ta. Ed.). Wiley Publishing
- Kimball R. & Ross M. (2013) The Data Warehouse Toolkit - The Definitive Guide to Dimensional Modeling (3ra. Ed.). Wiley & Sons
- Silberschatz A., Korth H. & Sudarshan S. (2020). Database System Concepts (7ma. Ed.), Mc Graw Hill

Unidad 3

- Golfarelli, M. & Rizzi, S.. (2009) Data Warehouse Design, Modern Principles and Methodologies (1ra. Ed.), McGraw-Hill
- Vaisman, A. & Zimanyi, E. (2022). Data Warehouse Systems: Design and Implementation (2da. Ed.). Springer-Verlag GmbH

Unidad 4

Libros

- Vaisman, A. & Zimanyi, E. (2022). Data Warehouse Systems: Design and Implementation (2da. Ed.). Springer-Verlag GmbH
- Inmon B., Levin M. & Srivstava R. (2021). Building the Data Lakehouse (1ra. Ed.). Technics Publications
- Gorelik A. (2019). The Enterprise Big Data Lake – Delivering the Promise of Big Data and Data Science (1ra. Ed.). O'Reilly Media
- Krishnan K. (2013). Data Warehousing in the Age of Big Data (1ra. Ed.) Morgan Kaufmann.

Artículos

- Lakehouse: A New Generation of Open Platforms that Unify DataWarehousing and Advanced Analytics, Armbrust, Ghodsi, Xin, Zaharia. 11th Annual Conference on Innovative Data Systems Research (CIDR '21), January 11–15, 2021
- Golfarelli, M., & Rizzi, S. (2018). From Star Schemas to Big Data: 20+ Years of Data Warehouse Research. A Comprehensive Guide Through the Italian Database Research.
- Khalid, Fakhardien, Zain & Majid (2015) Big Data Analysis and Storage. Proceedings of the 2015 International Conference on Operations Excellence and Service Engineering, Orlando, Florida, USA, September 10-11, 2015
- Armbrust M., Das T., Sun L., Yavuz B., Zhu S., Murthy M., Torres J., van Hovell H., Ionescu A., Łuszczak A., Switakowski M., Szafranski M., Xiao Li, Ueshin T., Mokhtar M., Boncz P., Ghodsi A., Paranjpye S., Senster P., Xin R. & Zaharia M. (2020). Delta Lake: HighPerformance ACID Table Storage over Cloud Object Stores. Proceedings of the VLDB Endowment, Vol. 13, No. 12
- Costa V. & Santos M.. (2018) Evaluating Several Design Patterns and Trends in Big Data Warehousing Systems. En Advanced Information Systems Engineering . Springer International Publishing AG, part of Springer Nature 2018

Unidad 5

Libros

- Sharda R., Delen D. & Turban E..(2017) Business Intelligence, Analytics and Data Science: A Managerial Perspective (4ta. Ed.) Pearson
- Eckerson W. (2011) Performance Dashboards. Measuring, Monitoring and Managing your Business (2da. Ed.). Wiley
- Silberschatz A., Korth H. & Sudarshan S. (2020). Database System Concepts (7ma. Ed.), Mc Graw Hill

- Tan P., Steinbach M., Kumar V. & Karpatne A. (2019) Introduction to Data Mining. (2da Ed.). Pearson
- Provost F. & Fawcett T. (2013) Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking. O'Reilly

Artículos

- Fayyad U., Piatetsky-Shapiro G. & Smyth P. (1996) From Data Mining to Knowledge Discovery in Databases., 1996. En AI Magazine Volume 17 Number 3
- Komorowsky M., Marshall D., Scalciccioli J. & Crutain Y. (2016) Exploratory Data Analysis., En Secondary Analysis of Electronic Health Records, MIT Critical Data
- Delen & Ram (2018) Research challenges and opportunities in business analytics.. En Journal of Business Analytics, Vol.1 2018