

TALLERES DE GESTIÓN DE DATOS E INFORMACIÓN.

México, noviembre 2019.



Agenda

(al 9 de octubre de 2019)

Miércoles, 6 de noviembre de 2019

8 : 3 0 - 9 : 0 0	REGISTRO
	DWTM1 Data Quality Fundamentals: For Big and Little Data LAURA SEBASTIAN -COLEMAN
	DWTM2 Landing the plane: Data Ethics where it hits the ground KATHERINE O'KEEFE
9 : 0 0 - 1 3 : 0 0	DWTM3 Machine and Deep Learning: Casos de estudio PATROCINADOR: UNAM Dr. Fabián Torres Robles M. en C. Blanca Vázquez
	DWTM4 ¿Cómo abordar la definición de un DATA Hub? PATROCINADOR: STAMBIA Dr. Luis Villanueva Peyre
1 3 : 0 0 - 1 4 : 0 0	DESCANSO PARA COMIDA
	DWTV1 Modern Textual Analytics BILL INMON
	DWTV2 Ciencia de Datos con Herramientas Open Source en Python FAVIO VÁZQUEZ
1 4 : 0 0 - 1 8 : 0 0	DWTV3 ¿Están tu organización (y tú) listos para el Gobierno de Datos? PATROCINADOR: ERWIN/DAYSET Mtro. José Luis León Gómez Ing. Dynorah Karina García Oviedo
	DWTV4 Gobierno y Calidad de Datos: De la Teoría a la práctica. PATROCINADOR: INFORMATICA/INFOMANAGEMENT Jorge Zamora Joaquín García

DESCRIPCIÓN DE LOS TALLERES

DWTM1

Organizations today get value from their data in the face of challenging odds. Optimal management of traditional data requires a wide skillset and strategic perspective. Changes in technology have increased the volume, velocity, and variety of data, but many organizations do not yet have a handle on veracity in traditional data management environments, never mind big data environments. And, while big data is on the rise, more traditional forms of data are not going away. Instead, different kinds of data will co-exist and must be managed in conjunction with one another.

This tutorial will revisit the fundamentals of data quality management in the light of big data and explore how to apply them in traditional and big data environments. Participants will learn how to assess the current state of their data environment and deliver more reliable data to their stakeholders.

The tutorial will include four sections:

- Quality management concepts and principles
- Applying quality management to traditional data
- Big Data challenges
- Data Quality Practices for Big Data and Little Data

DWTM2

This workshop introduces attendees to fundamental information ethics concepts and tools, providing a value-add to regulatory compliance, helping move beyond mere compliance to ethical information management as a competitive differentiator.

Long Description/Abstract: From the Gartner hype cycle to the front page of the news, the importance of a grounding in data ethics is becoming increasingly more visible. But is it just hype? How does "we need data ethics" look in day to day business decision making? Do you have appropriate systems of management in place to support ethical decisions and actions? This course is designed for leaders and information management professionals and provides a detailed framework and practical tools and techniques for implementing an ethical information management strategy.

Key takeaways for this session include:

1. An overview of fundamental Ethical Concepts as related to Information Management
2. Risk management, Information management practices
3. Using Data Privacy and Data Protection Law Compliance to focus Ethical Data use beyond compliance
4. Methods to align ethics with Information Governance
5. Practical tools for implementing an ethical information strategy

DWTM3

En este taller se aprenderá el uso de distintas herramientas involucradas en el ciclo de vida del desarrollo de modelos predictivos a partir de técnicas de Machine y Deep Learning. Estas herramientas se implementarán para la solución de distintos casos de estudio dentro de las principales áreas de investigación y desarrollo. Las herramientas que se implementaran son:

- Limpieza de outliers
- Datos faltantes
- Selección de características y reducción de dimensiones
- Procesamiento de Lenguaje Natural
- Métodos de Machine y Deep Learning

Objetivo: Aprender el ciclo de vida de un sistema de modelado predictivo, a partir de técnicas de Machine y Deep Learning, para la solución de problemas.

DWTM4

Este es un taller destinado a especialistas en integración de datos durante el cual se presentarán Las diferentes arquitecturas y se desarrollara el modelo Active Data Hub.

Descubra como la arquitectura «Data Hub» le permite centralizar y distribuir sus datos de manera simple y eficaz:

- Mejorando e Incrementando la agilidad, flexibilidad y rendimiento para reaccionar ante cualquier evolución de las tecnologías.
- Aportando flexibilidad a sus flujos de datos: batch, tiempo real, etc.

Se expondrán consideraciones sobre plataformas unificadas y tecnologías evolutivas que permitan acelerar e industrializar la implementación de este tipo de arquitectura de integración de datos entre sus aplicaciones em modo semi-flujo.

Al final de este taller se presentará un caso práctico de sincronización en tiempo real una plataforma CRM en la nube con sus Sistemas de gestión.

Al concluir el asistente logrará:

- Reconocer la diferencia práctica entre un Datahub, Data Lake y Data Warehouse.
- Identificar la importancia de tener un Datahub y considerando éste como un gran paso en su camino hacia su estrategia de Big Data.
- Visualizar un plan inmediato de estrategia para poder implementar un Datahub en su organización.
- Considerar una fácil resolución a problemas reales de integración de fuentes de datos dispersas, diversas, estructuradas y semi-estructuradas para gestionar el activo más importante de la organización, el dato. No importando la tecnología (de hoy o mañana), la frecuencia o el lugar donde el cambio del dato suceda.

DWTV1

There is great promise in the ability to handle text. Text is found in many places - email, call centers, contracts, Internet, and many more places. Yet most organization do little or nothing with text. This presentation addresses the obstacles in dealing with text. taxonomies are discussed along with subjects such as sentiment analysis, document relationship analysis, and others. The value of visualizing the results is addressed as well.

DWTV2

Limpiar, preparar, transformar, explorar datos y modelar es lo que escuchamos todo el tiempo sobre ciencia de datos, y estos pasos pueden ser los más importantes. Pero eso no es lo único sobre la ciencia de datos, en esta charla aprenderás cómo la combinación de Apache Spark, Optimus, el ecosistema Python y Data Operations pueden formar un marco completo para la ciencia de datos que te permitirá a ti y a tu empresa ir más allá, y más allá del sentido común y la intuición para resolver problemas de negocio complejos.

DWTV3

Diagnóstico del estado de preparación de la organización para emprender una iniciativa de Gobierno de Datos exitosa. Recomendaciones de los expertos de Erwin y Dayset para, en función del diagnóstico anterior, emprender una iniciativa de Gobierno de Datos exitosa.

DWTV4

En este taller se abordarán conceptos asociados al gobierno de datos, desde una perspectiva práctica para su implementación dentro de una organización, mediante el uso del DMBOK2 como guía de implementación. Igualmente, se explorará su alineamiento con la gestión de calidad de datos, así como la puesta en marcha de métricas que permitan cerrar el ciclo de la administración de calidad bajo el contexto de un gobierno de datos previamente definido, a través de la elaboración de ejercicios y revisión de ejemplos.

PONENTES

BILL INMON

Científico informático estadounidense, reconocido por muchos como el padre del DATA WAREHOUSE. Inmon escribió el primer libro, realizó la primera conferencia (con Arnie Barnett), escribió la primera columna en una revista y fue el primero en ofrecer clases sobre almacenamiento de datos. Inmon creó la definición aceptada de lo que es un DATA WAREHOUSE: una recopilación de datos, orientada al tema, no volátil e integrada en el tiempo, que respalda las decisiones de la gerencia. En comparación con el enfoque del otro arquitecto pionero del almacenamiento de datos, Ralph Kimball, el enfoque de Inmon se caracteriza a menudo como un enfoque de arriba hacia abajo.

El libro más recientemente publicado de Bill es Data Architecture: A Primer for Data Scientist, Segunda edición.

Bill es fundador y CEO de Forest Rim Technology, en donde ha creado el primer software Textual ETL del mundo. Con el software Textual ETL puede leer cualquier forma de texto en cualquier idioma y convertirlo de manera significativa en cualquier sistema de gestión de bases de datos relacionales estándar.

DR. LAURA SEBASTIAN-COLEMAN

Laura Sebastian-Coleman (CDMP, IQCP), Líder de calidad de datos en Aetna / CVS Health, ha trabajado en calidad de datos en grandes almacenes de datos (Data Warehouse) analíticos de salud desde 2003. Laura ha implementado métricas e informes de calidad de datos, ha lanzado y facilitado grupos de trabajo de calidad de datos, y contribuyó a los programas de capacitación de consumidores de datos. Ha dirigido esfuerzos para establecer estándares de datos y administrar metadatos para grandes almacenes de datos analíticos y esfuerzos de gobierno de datos empresariales.

Autora de Navigating the Labyrinth (2018) y Measuring Data Quality for Ongoing Improvement (2013), Laura también fue editora de producción de DAMA-DMBOK2 (2017), por la que recibió el premio de DAMA International por sus contribuciones a la profesión de gestión de datos (2018). . Como asesora de DAMA New England (2018 a la fecha), se desempeñó como Oficial de Publicaciones de DAMA (2015-2018) y como Directora de Servicios para Miembros de IAIDQ (ahora IQ International) (2009-2011). En 2015, recibió el premio de Miembro Distinguido de IAIDQ.

KATHERINE O'KEEFE

Es una consultora líder, gobernadora de información y privacidad, ética y jefa de ética con Castlebridge. Desde su incorporación a Castlebridge en 2013, Katherine ha trabajado con clientes en los sectores de telecomunicaciones, transporte, educación y organizaciones sin fines de lucro y dirige el equipo de consultoría de empresas sociales y microempresas dentro de los proyectos de consultoría y asesoría representativos de Castlebridge que incluyen:

- Asesoramiento en protección de datos y apoyo de capacitación a uno de los mayores proveedores de educación secundaria y superior en Irlanda.
- Análisis y diseño de una hoja de ruta estratégica de Gobierno de Datos para una empresa de telecomunicaciones líder en Irlanda.
- Asesoría sobre Solicitud de acceso basado en temas para un cliente del sector educativo.
- Identificación y definición de prioridades de gobernabilidad de datos y desarrollo de una estrategia y hoja de ruta para una gran organización en el sector del transporte.
- Protección de datos y revisiones de cumplimiento para varios clientes en el sector irlandés sin fines de lucro.
- Diseño, desarrollo y entrega de una variedad de programas de capacitación de Gobierno de datos y protección de datos a medida para clientes en diversos sectores, utilizando métodos tradicionales de aula, e-learning y entrega combinada.

Katherine coescribió documentos de consulta pública sobre la gobernabilidad de datos en el sector público, un capítulo sobre el uso de la gobernabilidad de datos ágil en un entorno regulatorio y documentos técnicos sobre las solicitudes de acceso de los sujetos, las implicaciones de privacidad de los mensajes instantáneos en un contexto médico y la gobernanza de la información. Estrategia para la implementación de identificadores de salud. En colaboración con Daragh O'Brien, ha escrito Gestión de información y datos éticos: conceptos, herramientas y métodos.

En su calidad de Vicepresidenta de Desarrollo Profesional en DAMA International, Katherine supervisó la reurbanización de la Certificación CDMP y es contribuyente a la 2ª Edición de DAMA DMBOK.

FAVIO VÁZQUEZ

Físico e ingeniero informático con una maestría en física. Trabajando en ciencia de datos y cosmología computacional. Tiene una pasión por la ciencia, la filosofía, la programación y la música. En este momento está trabajando en ciencia de datos, aprendizaje automático y big data como Senior Data Scientist en Raken Data Group. Además, es el fundador y jefe de Data Science, una empresa de Data Science y educación en español. Le encantan los nuevos desafíos, trabajar con un buen equipo y le apasiona tener problemas interesantes para resolver, también

ha aplicando sus conocimientos y experiencia en ciencia, análisis de datos, visualización y aprendizaje automático para ayudar al mundo a convertirse en un lugar mejor.

DR. FABIAN TORRES ROBLES

Investigador Posdoctoral de la Facultad de Ingeniería C.U.-UNAM. Es egresado del Instituto Tecnológico y de Estudios Superiores de Monterrey como Ingeniero en Sistemas Electrónicos, cuenta con una maestría en ciencias y mención honorífica en el posgrado de instrumentación eléctrica de la ESIME-Zacatenco IPN; posteriormente, obtuvo el grado de Doctor en Ingeniería en el Instituto de Ciencias Aplicadas y Tecnología de la UNAM, colaborando y realizando una estancia de investigación en la Universidad Nacional de Taiwan. Su investigación se basa en el análisis, implementación y desarrollo de modelos predictivos mediante técnicas de Machine y Deep Learning para el reconocimiento de patrones y el procesamiento de datos e imágenes, con aplicaciones en diferentes áreas como medicina y seguridad.

DRA. JIMENA OLVERES

Es profesora en la Universidad Nacional Autónoma de México (UNAM). Se recibió como Ingeniera biomédica en la Universidad Iberoamericana en la ciudad de México, Maestría en ingeniería realizando estancias de investigación en la Universidad de Washington, Seattle, y Ph.D. en Ciencias de la Computación dentro de la UNAM. Sus áreas de investigación incluyen modelos computacionales para el procesamiento digital de imágenes, visión computacional, aprendizaje de máquina y aprendizaje profundo, aplicado al análisis de imágenes médicas y cardíacas.

M. en C. BLANCA VÁZQUEZ

Maestra en Ciencias de la Computación por el Centro Nacional de Investigación y Desarrollo Tecnológico (Cenidet) con mención honorífica. Ha colaborado en proyectos de investigación en la Fundación Bruno Kessler en Trento, Italia, en la Benemérita Universidad Autónoma de Puebla (BUAP) y en la Universidad Tecnológica de la Mixteca (UTM). Actualmente estudia un posgrado en Ciencia e Ingeniería de la Computación en el Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas, UNAM. Sus áreas de investigación son aprendizaje máquina en el área de salud, procesamiento de lenguaje natural y modelos secuenciales.

DR. CARLOS GERSHENSON GARCÍA

Carlos Gershenson es investigador definitivo de tiempo completo del Departamento de Ciencias de la Computación del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas de la Universidad Nacional Autónoma de México (UNAM), donde es líder del Laboratorio de Sistemas Auto-organizantes. También es investigador asociado del Centro de Ciencias de la Complejidad de la UNAM. Es coordinador del Laboratorio Nacional de Ciencias de la Complejidad. Realizó una estancia sabática en MIT y Northeastern University en 2015-2016. Realizó una estancia postdoctoral en el Instituto de Sistemas Complejos de Nueva Inglaterra. Es doctor en ciencias summa cum laude por la Universidad Libre de Bruselas en Bélgica, donde su tesis fue sobre el "Diseño y Control de Sistemas Auto-organizantes". Tiene una maestría en sistemas evolutivos y adaptativos por la Universidad de Sussex en Inglaterra y es ingeniero en computación por la Fundación Arturo Rosenblueth en la Ciudad de México. Estudió cinco semestres de filosofía en la UNAM.

Ha sido un investigador activo desde 1997, trabajando en el Instituto de Química de la UNAM y un verano en el Instituto Weizmann de Ciencia en Israel. Tiene más de cien publicaciones en libros, revistas y memorias de congresos, las cuales han sido citadas en más de cuatro mil ocasiones. Ha dado más de doscientas presentaciones en diversos foros.

Tiene una gran variedad de intereses académicos, incluyendo sistemas auto-organizantes, complejidad, urbanismo, evolución, vida artificial, información, cognición, sociedades artificiales, y filosofía.

Es editor en jefe de Complexity Digest y miembro del consejo de asesores de Scientific American. Colabora con artículos de opinión en el diario Reforma. Ha trabajado como maestro a nivel de bachillerato y de posgrado, consultor, desarrollador y divulgador. Entre los premios que ha recibido se encuentran el Reconocimiento Distinción Universidad Nacional para Jóvenes Académicos, una Cátedra de Investigación Marcos Moshinsky para Jóvenes Científicos, un Google Research Award in Latin America y el Audi Urban Future Award. Es miembro de la Academia Mexicana de Ciencias y de la Academia Mexicana de Informática.

J o r g e Z a m o r a

Consultor principal con más de diez años de experiencia en proyectos de gestión de datos así como en el asesoramiento e implementación de estrategias y modelos de gobierno y gestión de datos en diferentes industrias – Manufactura, Gobierno, Servicios Financieros y en diferentes latitudes, como México, Perú y Colombia, entre otros. Certificado CDMP.

Mtro. José Luis León Gómez

Profesional de TI con más de 30 años de experiencia que van desde la automatización de las bibliotecas de la UNAM y sus catálogos, hasta proyectos de integración de datos estructurados y no estructurados de más de 1,200 millones de registros para detectar patrones y entidades. Ing. en Computación por la UNAM, MBA, y actualmente inscrito en el Doctorado en Dirección de Organizaciones. Docente de posgrado en universidades privadas. Ha dirigido operaciones y proyectos de TI, y actualmente está a cargo de los productos y servicios de Gestión de Datos en DAYSET.

J o a q u í n G a r c í a

Consultor principal con más de diez años de experiencia en la ejecución e implementación de proyectos de Calidad de datos, en múltiples industrias, desde la perspectiva empresarial,

considerando conceptos claves de gobierno de datos, para su ejecución e implementación efectiva dentro de las organizaciones.

