



Edgar Oswaldo Díaz
edgar.diaz@ucags.edu.mx

Abstract

CONACYT CVU 399463 | Professional Certificate 6684741
Researcher and Architect in Information Technologies for the National Institute of Statistics and Geography (INEGI), Experience in Auditing for Systems Engineering based on DevSecOps, as well as in processes oriented to DataSecOps for the Laboratory of Institutional Data Science, on Risk Management, Information Security and Privacy {Ethical Specialties (Hacking-Forensics) – Data Science, Big-Small-Fast Data Analytics, Mobile, Cloud Computing, IoT}.

Certifications:

- SGSI (ISO/IEC-27001:2013) | UNAM-CERT
- ITIL IT Service Management | AXELOS
- Certified Ethical Hacker - CEH | EC-Council
- Computer Hacking Forensic Investigator - CHFI | EC-Council
- Chief Information Security Officer (CISO) | EC-Council

Author of the following international refereed publications

- Springer | DevOps: Fundamentals and implementation in a Data Center DOI: 10.1007/978-3-319-65082-1_10
- IEEE | Reinforcing the DevOps approach with security and risk management: an implementation experience in a data center of a Mexican organization. DOI: 10.1109/CIMPS.2017.8169957
- IEEE | Responsive and highly available infrastructure with cybersecurity for automated DevSecOps processes. DOI: 10.1109/CIMPS49236.2019.9082439
- IGI Global | Strategy to carry out critical projects in a data center using the DevSecOps approach and risk management. DOI: 10.4018/IJITSA.2020010104
- IEEE | Data Lake Strategy for Data Science Workflows. DOI: 10.1109/CIMPS57786.2022.10035694.

Contributions

- Mora, M., Wang, F., Gómez, J.M., Díaz, O. (2020). A Comparative Review on the Agile Tenets in the IT Service Management and the Software Engineering Domains. In: Mejia, J., Muñoz, M., Rocha, Á., A. Calvo-Manzano, J. (eds) Trends and Applications in Software Engineering. CIMPS 2019. Advances in Intelligent Systems and Computing, vol 1071. Springer, Cham. DOI: 10.1007/978-3-030-33547-2_9
- Mora, M., Adelakun, Díaz, O., Reyes-Delgado, P.Y. et al. AVS_FD_MVITS: an agile IT service design workflow for small data centers. J Supercomput (2023). DOI: 10.1007/s11227-023-05244-w

https://www.researchgate.net/profile/Edgar_Diaz26

Calendario

- Agosto (12,19,26)
- Septiembre (2,9,23,30)
- Octubre (7,14,21)
- Noviembre (11,18,25)
- Diciembre (2,9,16)

Mes	Día	Actividad
Agosto	12	Bienvenida e Introducción
Agosto	12	1.1 Revisión de especificación de requisitos.
Agosto	12	1.1.1 Norma IEEE830
Agosto	19	1.1.2 Trazabilidad de requisitos
Agosto	19	1.2 Descripción de procesos actuales
Agosto	26	Mesa de debate modelos basados en ingeniería de datos
Septiembre	2	Primera evaluación parcial
Septiembre	9	1.3 Diagramas UML
Septiembre	9	1.4 Estudio de Factibilidad
Septiembre	23	1.5 Análisis Costo-Beneficio
Septiembre	23	2.1 Diseño de procesos propuestos
Septiembre	30	2.1.1 Herramientas CASE para diseño
Octubre	7	2.2 Diseño arquitectónico
Octubre	7	2.3 Diseño de datos
Octubre	14	Segunda evaluación parcial
Octubre	21	2.4 Diseño de interfaz de usuario
Octubre	21	3.1 Lenguajes para el desarrollo de software
Noviembre	11	3.2 Manejadores de bases de datos
Noviembre	11	3.3 Construcción del sistema
Noviembre	18	4.1 Diseño de caso de prueba.
Noviembre	18	4.2 Pruebas de componentes
Noviembre	25	4.3 Pruebas del sistema
Diciembre	2	4.4 Documentación de resultados de las pruebas.
Diciembre	2	4.5 Entrega del sistema y Capacitación a usuarios
Diciembre	2	4.6 Entrega de documentación técnica y de usuario del sistema.
Diciembre	9	Tercera evaluación parcial
Diciembre	9	Asesoría para el proyecto
Diciembre	16	Entrega y presentación del proyecto



Colaboradores

- | | |
|------------------------------------|------------------------------------------------------------------------------------------------|
| • Mitsiu Alejandro Carreño Sarabia | mitsiu.carreno@alumnos.ucags.edu.mx |
| • Kevin Gustavo Domínguez Lopez | kevin.dominguez@alumnos.ucags.edu.mx |
| • Omar Franco Ramirez | omar.franco@alumnos.ucags.edu.mx |
| • Andrea Graciela Lopez Segura | andrea.lopez@alumnos.ucags.edu.mx |
| • Ulises Paz Vidales | ulises.paz@alumnos.ucags.edu.mx |
| • Andres Daniel Pérez Rodriguez | andres.perez@alumnos.ucags.edu.mx |
| • Marco Antonio Rodriguez Rangel | marco.rodriguez@alumnos.ucags.edu.mx |

Actividad

- Porque es importante establecer una estrategia para realizar actividades de ingeniería de datos.
- Mencionar en la vida cotidiana como puede ayudar la ingeniería de datos.
- Mencionar un caso de uso relacionado con ingeniería de datos.
- Narrar que es business intelligence.
- Narrar que es business analytics.

<https://moodle.ucags.edu.mx/course/view.php?id=2082>



La **ingeniería de datos** es la práctica que consiste en diseñar y construir sistemas de recolección, almacenamiento, procesamiento y análisis de datos a gran escala.

Las instituciones / organizaciones, tienen la necesidad de recopilar cantidades masivas de datos, contar con el personal, la tecnología adecuada para garantizar que la información esté disponible para que los científicos y analistas de datos realicen sus actividades.

“Es un campo amplio con aplicaciones en casi todos los sectores”.

El **perfil de ingeniería de datos** trabaja en una variedad de entornos para construir sistemas que recopilan, gestionan y convierten los datos iniciales en información utilizable para que los científicos de datos y los analistas la interpreten.

Hacer que los datos sean accesibles para generar indicadores cuantitativos y cualitativos para la toma de decisiones.



Tareas más comunes.

- Adquirir conjuntos de datos que se ajusten al requerimiento.
- Desarrollar sistemas para transformar los datos en información útil y procesable.
- Construir, probar y mantener arquitecturas de canalización en bases de datos.
- Crear nuevos métodos de validación, herramientas de análisis de datos.
- Garantizar el cumplimiento de las políticas en cumplimiento de seguridad y gobernanza de datos.

Estrategia

- Implementar técnicas de búsqueda y uso de información: datos heterogéneos, temporales, espaciales, ética en el uso de datos.
- Bases de datos relacionales y NoSQL: Spark, MongoDB, GraphDB, entre otras.
- Entendiendo los datos: análisis exploratorio, reducción e integración de grandes volúmenes de datos.
- Presentación y visualización de grandes volúmenes de datos: interfaces visuales, analítica visual.



Habilidades

- Codificación de datos.
- Bases de datos relacionales y no relacionales.
- Sistemas ETL- ELT para transformación de datos.
- Automatización y scripting de datos.
- Herramientas para utilizar grandes volúmenes de datos.
- Computación de alto desempeño en procesamiento, analítica o ciencia de datos .
- Privacidad, seguridad de los datos.

Objetivo

- Conocer y aplicar diferentes **técnicas estadísticas** para el resumen, descripción y visualización de datos.
- Construir y validar **modelos estadísticos**.
- Desarrollar la habilidad de modelar procesos de adquisición de datos de manera **probabilística**, desde un enfoque tanto **frecuentista** como **Bayesiano**.
- Analizar situaciones reales que permitan adquirir experiencia en la generación, prueba e interpretación de **hipótesis estadísticas**.



Técnicas Estadísticas (nivel de medida)

Nominal: Cada valor de una variable nominal se corresponde con una categoría de la variable, este emparejamiento es por lo general arbitrario.

Ordinal: Cada valor representa la ordenación o el ranking.

Intervalo: En variables de intervalo un incremento de una unidad en el valor numérico representa el mismo cambio en la magnitud medida, con independencia de donde ocurra en la escala.

Razón: Las variables de Razón tienen las mismas propiedades que las de intervalo, pero además tienen un punto cero significativo, dicho punto representa una ausencia completa de la característica medida.

Modelos estadísticos

Los modelos estadísticos utilizan ecuaciones matemáticas para codificar información extraída de los datos. En algunos casos, las técnicas de modelado estadístico pueden proporcionar modelos adecuados de forma ágil. Incluso en el caso de problemas en los que las técnicas más flexibles de aprendizaje de las máquinas (como redes neuronales) pueden ofrecer mejores resultados, es posible usar algunos modelos estadísticos como modelos predictivos de línea base para evaluar el rendimiento de técnicas más avanzadas.



Modelos estadísticos

- Los modelos de **regresión lineal** predicen un objetivo continuo tomando como base las relaciones lineales entre el destino y uno o más predictores.
- La **regresión logística** es una técnica de estadístico para clasificar los registros en función los valores de los campos de entrada. Es análoga a la regresión lineal, pero toma un campo objetivo categórico en lugar de uno numérico.
- El nodo **PCA/Factorial** proporciona técnicas eficaces de reducción de datos para reducir la complejidad de los datos. Análisis de componentes principales (PCA) busca combinaciones lineales de los campos de entrada que realizan el mejor trabajo a la hora de capturar la varianza en todo el conjunto de campos, en el que los componentes son ortogonales (perpendiculares) entre ellos. Análisis factorial intenta identificar factores subyacentes que expliquen el patrón de correlaciones dentro de un conjunto de campos observados. Para los dos métodos, el objetivo es encontrar un número pequeño de campos derivados que resuma de forma eficaz la información del conjunto original de campos.

Modelos estadísticos

- El **análisis discriminante** realiza más supuestos rigurosos que regresiones logísticas, pero puede ser una alternativa o un suplemento valioso al análisis de regresión logística si se cumplen dichos supuestos.
- El **modelo lineal generalizado** amplía el modelo lineal general, de manera que la variable dependiente está relacionada linealmente con los factores y las covariables mediante una determinada función de enlace. Además, el modelo permite que la variable dependiente tenga una distribución no normal, incluye la regresión lineal, regresión logística.



Modelos estadísticos

- Un **modelo lineal mixto generalizado** (GLMM) amplía el modelo lineal de modo que el objetivo pueda tener una distribución no normal, esté linealmente relacionado con los factores y covariables mediante una función de enlace especificada y las observaciones se puedan correlacionar. Los modelos lineales mixtos generalizados cubren una amplia variedad de modelos, desde modelos de regresión lineal simple hasta modelos multinivel complejos para datos longitudinales no normales.
- El **nodo Regresión de Cox** permite crear un modelo de supervisado para datos de tiempo hasta el evento en presencia de registros censurados. El modelo produce una función supervisada que predice la probabilidad de que el evento de interés se haya producido en el momento dado (t) para valores determinados de las variables de entrada.

Un **enfoque frecuentista** se basa en la probabilidad como una expresión matemática que predice en qué medida es verosímil que ciertos eventos ocurran basándose en el patrón observado hasta este momento. (*Ronald Fisher*)

La **interpretación bayesiana** se basa en un conocimiento limitado de las cosas. Afirma que sólo asocias una probabilidad a un evento porque hay incertidumbre sobre el mismo, es decir, porque no conoces todos los hechos. (*Thomas Bayes*).



Hipótesis estadística

- **Causales:** tipo de hipótesis que tienen el objetivo de explicar los factores de relación – causa, existentes entre dos o más variables estadísticas.
- **Correlacionales:** Estas hipótesis tratan de establecer qué tipo de relación existe entre dos variables.
- **Diferencia de grupos:** Tienen el objetivo de crear una distinción entre dos o más grupos estudiados en función de las características de los mismos.
- **Descriptivas:** Son aquellas hipótesis que tienen la función de informar acerca de la relación existente entre dos o más variables.

Método científico “pasos”:

- Realizar una observación o formular una pregunta.
- Formar una hipótesis.
- Probar la hipótesis a través de un experimento.
- Revisar la hipótesis y, si es necesario, volver al punto anterior para mejorar la hipótesis.
- Obtener conclusiones y publicar el resultado para que otros también lo comprueben.



1. Análisis

1.1 Revisión de especificación de requisitos.

1.1.1 Norma IEEE830.

1.1.2 Trazabilidad de requisitos.

1.2 Descripción de procesos actuales.

1.3 Diagramas UML.

1.4 Estudio de Factibilidad.

1.5 Análisis Costo-Beneficio.

Revisión de especificación de requisitos.

- Ayuda a garantizar que todas las partes interesadas tengan un entendimiento común del sistema que se va a desarrollar. Esto evita cualquier malentendido durante las etapas posteriores del desarrollo.
- Sirve como punto de referencia para todas las partes interesadas durante el proceso de desarrollo.
- Ayuda a identificar cualquier brecha en los requisitos en una etapa temprana.
- Reduce el costo general y el tiempo de desarrollo, ya que evita la repetición del trabajo debido a cambios en los requisitos.



Tipos de Requisitos Especificaciones

Especificaciones de requisitos funcionales: (FRS) es un documento que captura las funciones que debe realizar un sistema. Incluye todas las funcionalidades, instalaciones, medidas de seguridad y otra información relevante.

Tipos de Requisitos Especificaciones

Especificaciones de requisitos de rendimiento: (PRS) es un documento que captura todos los aspectos relacionados con el rendimiento de un sistema. Esto incluye el tiempo de respuesta, el rendimiento de datos, la eficiencia, la escalabilidad, entre otras variables.



Tipos de Requisitos Especificaciones

Especificación de requisitos de configuraciones: (CRS) es un documento que captura toda la información relacionada con la configuración de un sistema. Esto incluye detalles como plataformas compatibles, dependencias de software/hardware/firmware, requisitos básicos para la funcionalidad del sistema.

Tipos de Requisitos Especificaciones

Especificaciones de requisitos comerciales: (BRS) es un documento que captura todos los aspectos relacionados con el negocio de un sistema. Esto incluye características tales como administración de usuarios, seguridad, integridad, calidad, gobernanza de datos.



Tipos de Requisitos Especificaciones

Especificaciones de requisitos de confiabilidad: (RRF) es un documento que captura toda la información relacionada con la confiabilidad de un sistema. Esto incluye aspectos como el tiempo de actividad, el tiempo de recuperación, los umbrales cuantitativos y cualitativos con base en niveles de servicio (SLA's).

Tipos de Requisitos Especificaciones

Especificaciones de requisitos de compatibilidad: (CRF) es un documento que captura toda la información relacionada con la compatibilidad de un sistema. Esto incluye aspectos como plataformas compatibles, dependencias de software/hardware/firmware, de los componentes involucrados en el sistema.



Tipos de Requisitos Especificaciones

Especificaciones de requisitos del sistema: SRS es un documento que captura todos los aspectos relacionados con el sistema. Esto incluye aspectos como la funcionalidad, el rendimiento, la escalabilidad, entre otras variables.

Características de un documento de especificación

- **Preciso** – El objetivo de un documento es que sea fácil de comprender. Nada debe quedar poco claro, por lo que no hay diferencia de opiniones entre las partes interesadas.
- **Mensurable** – Los requisitos en el documento deben ser medibles, por lo tanto, el artefacto terminado puede validarse y certificarse según los requisitos. El documento debe incluir suficiente información para que su equipo de desarrollo y evaluadores completen el artefacto y garanticen que cumple con los requisitos del usuario sin eventualidades.
- **Factible** – Los requisitos deben reflejar el estado real de los roles, procesos, tecnología en el cronograma de trabajo establecido por ambas partes.
- **Verificable** – Todos los involucrados en el equipo de desarrollo deben tener acceso al documento para que puedan consultarlo tantas veces como sea necesario.
- **Consistente** – Los requisitos deben ser compatibles. No debe haber ninguna contradicción entre las partes del documento dentro del entorno de la especificación de requisitos. El documento debe estar estructurado de manera consistente y la terminología debe usarse de la misma manera en todo el documento.
- **Preciso** – Los requisitos especificados en el documento deben ser muy precisos para evitar cualquier tipo de confusión. No deben tener lagunas, ambigüedades, subjetividad, superlativos o comparaciones.

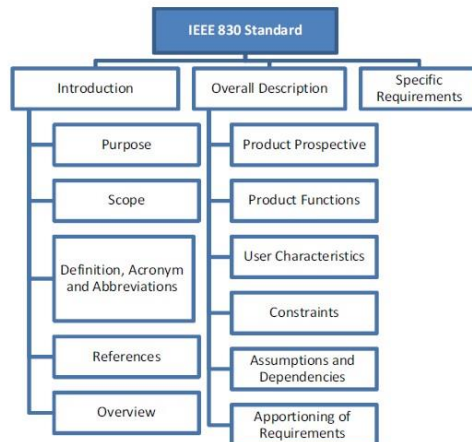


Componentes esenciales

- **Impulsores del negocio.**
- **Modelo de negocio.**
- **Requisitos funcionales y del ecosistema.**
- **Casos de uso del ecosistema** – diagramas basados en el lenguaje de modelado unificado llamado (UML), que explica todas las entidades externas clave que interactuarán con los componentes involucrados en el ecosistema.
- **Requerimientos Técnicos** – esta sección analiza todos los requisitos no funcionales que conforman el entorno técnico y las limitaciones técnicas involucradas en el ecosistema.
- **Cualidades** – definir confiabilidad, facilidad, seguridad, escalabilidad, disponibilidad y la mantenibilidad, de los componentes involucrados en el ecosistema.
- **Limitaciones y supuestos** – esta sección describe todas las limitaciones impuestas al diseño de los componentes involucrados en el ecosistema considerando el equipo de trabajo colaborativo.
- **Criterios de aceptación** – En esta sección son analizados los cumplimientos normativos establecidos considerando los componentes que interactúan en el ecosistema.

Estructura propuesta para un documento

1. Introducción.
 - 1.1. Propósito.
 - 1.2. Público objetivo.
 - 1.3. Uso previsto.
 - 1.4. Alcance.
 - 1.5 Definiciones y siglas.
2. Descripción general.
 - 2.1 Necesidades.
 - 2.2 Supuestos y Dependencias.
3. Características y requisitos del sistema.
 - 3.1 Requisitos funcionales.
 - 3.2 Requisitos de la interfaz externa.
 - 3.3 Requisitos del sistema.
 - 3.4 Requisitos no funcionales.



Especificación de Requisitos según el estándar de IEEE 830

IEEE Std. 830-1998

22 de Octubre de 2008

Resumen

Este documento presenta, en castellano, el formato de Especificación de Requisitos Software (ERS) según la última versión del estándar IEEE 830. Según IEEE, un buen Documento de Requisitos, pese a no ser obligatorio que siga estrictamente la organización y el formato dados en el estándar 830, si deberá incluir, de una forma o de otra, toda la información presentada en dicho estándar. El estándar de IEEE 830 no está libre de defectos ni de prejuicios, y por ello ha sido justamente criticado por múltiples autores y desde múltiples puntos de vista, llegándose a cuestionar incluso si es realmente un estándar en el sentido habitual que tiene el término en otras ingenierías. El presente documento no pretende pronunciarse ni a favor ni en contra de unos u otros: tan sólo reproduce, con propósitos fundamentalmente docentes, cómo se organizaría un Documento de Requisitos según el estándar IEEE 830.

<https://moodle.ucags.edu.mx/course/view.php?id=2082§ion=2>

Norma IEEE 830 - 2009

1. Introducción

- Propósito
- Alcance
- Definición, Acrónimos y Abreviaturas
- Referencias
- Descripción general

2. Descripción general

- Perspectiva del producto
- Funciones del producto
- Características del usuario
- Restricciones
- Suposiciones y dependencias
- Prorrato de requisitos

3. Requisitos específicos

- Interfaces externas
- Funciones
- Requisitos de rendimiento
- Requisitos de la base de datos lógica
- Restricciones de diseño
- Atributos de calidad del sistema
- Organizar los requisitos específicos

4. Información de apoyo

- Tabla de contenido
- Índice
- Apéndices

Shoub, Khaled & Carvalho, Marco & Bond, Walter. (2017). Recommended Practices for the Specification of Multi-Agent Systems Requirements. 10.1109/UEMCON.2017.8249021.



Referencias

- Renz, Kai. (2022). Using OpenSky Data for Teaching Software Engineering to Undergraduates. *Engineering Proceedings*. 28. 3. 10.3390/engproc2022028003.
- Koch, Christian & Stadi, Markus & Berle, Lukas. (2023). From Data Engineering to Prompt Engineering: Solving data preparation tasks with ChatGPT.
- Abeykoon, Vibhatha & Fox, Geoffrey. (2023). Trends in High Performance Data Engineering for Data Analytics. 10.5772/intechopen.1001458.
- Tamburri, Damian & Heuvel, Willem-Jan. (2023). Big Data Engineering. 10.1007/978-3-031-19554-9_2.
- Hussain, Mumtaz & Siddiqui, Samrina & Islam, Noman. (2023). Social Engineering and Data Privacy. 10.4018/978-1-6684-6581-3.ch010.
- Popovych, Valerii & Drlik, Martin. (2023). Towards Development of Data Architecture for Learning Analytics Projects Using Data Engineering Approach. 10.1007/978-981-99-1479-1_38.
- Users, Panel & Statistics, Committee & Education, Division & Sciences, Division & Council, National. (2012). Communicating Science and Engineering Data in the Information Age. 10.17226/13282.
- Sydorov, N.O. & Sydorova, N.M.. (2022). Software engineering and big data software. *PROBLEMS IN PROGRAMMING*. 69-72. 10.15407/pp2022.03-04.069.
- Cascavilla, Giuseppe & Dalla Palma, Stefano & Driessen, Stefan & Heuvel, Willem-Jan & Pascale, Daniel & Sangiovanni, Mirella & Schouten, Gerard. (2023). Data Engineering in Action. 10.1007/978-3-031-19554-9_5.
- Ke, Dandan & Dai, Jingyi. (2023). Multidimensional analysis of engineering cost database based on descriptive data mining. *Soft Computing*. 1-8. 10.1007/s00500-023-07992-6.
- Du, Shuiting & Liu, Shaobo & Xu, Peng & Zhang, Jianfeng. (2023). Engineering archive management model based on big data analysis and deep learning model. *Applied Mathematics and Nonlinear Sciences*. 10.2478/amns.2023.1.00212.
- Bosamiya, Nidhi & Patel, Shivangi. (2022). Symbiosis of Data Science and Software Engineering.
- Chicco, Davide & Oneto, Luca & Tavazzi, Erica. (2022). Eleven quick tips for data cleaning and feature engineering. *PLoS computational biology*. 18. e1010718. 10.1371/journal.pcbi.1010718.
- Safdar, Mutahar & Lamouche, Guy & Paul, Padma & Wood, Gentry & Zhao, Yaoyao. (2023). Engineering of Additive Manufacturing Features for Data-Driven Solutions: Sources, Techniques, Pipelines, and Applications. 10.1007/978-3-031-32154-2.