

## Quality Assurance – QA en Ingeniería de Datos.

### Definición:

La QA en Ingeniería de Datos se refiere a las prácticas y procesos destinados a garantizar la calidad de los datos en todas las etapas de su ciclo de vida, desde la adquisición hasta el análisis y la presentación. Implica la implementación de estándares, procesos y herramientas que aseguren la integridad, consistencia y exactitud de los datos, así como la identificación y mitigación de posibles problemas.

### Importancia:

- **Toma de Decisiones Precisas:** La toma de decisiones basada en datos requiere que estos sean precisos y confiables. La QA asegura que los datos utilizados para la toma de decisiones estén libres de errores y sean representativos de la realidad.
- **Eficiencia Operativa:** Los procesos de QA optimizan la eficiencia operativa al reducir el tiempo y los recursos dedicados a la resolución de problemas causados por datos defectuosos. Identificar y corregir problemas temprano evita costosos retrabajos.
- **Cumplimiento Normativo:** En muchos sectores, existen regulaciones estrictas sobre la calidad y privacidad de los datos. La QA garantiza que las organizaciones cumplan con estas regulaciones, evitando sanciones y protegiendo la reputación de la empresa.
- **Confianza del Usuario:** La confianza del usuario en los sistemas y productos de datos depende directamente de la calidad de los datos. La implementación de prácticas de QA fortalece la confianza de los usuarios al garantizar la coherencia y precisión de la información presentada.

### Desafíos:

- **Volumen y Variedad de Datos:** El crecimiento exponencial de datos, junto con su diversidad, plantea desafíos únicos. La QA debe adaptarse a diferentes tipos de datos, desde estructurados hasta no estructurados, y enfrentar la complejidad de grandes conjuntos de datos.
- **Integración de Datos:** En entornos empresariales complejos, los datos provienen de diversas fuentes y sistemas. La integración efectiva de datos requiere una QA que aborde la coherencia entre estas fuentes, evitando inconsistencias y redundancias.
- **Escalabilidad:** A medida que las organizaciones crecen, la infraestructura de datos debe escalar para manejar grandes volúmenes de información. La QA debe ser escalable para mantener la calidad de los datos a medida que crece la infraestructura.

## Estándares internacionales

**ISO/IEC 25012:2008 (anteriormente ISO 9126):** Este estándar internacional se centra en la calidad del producto de software y define un modelo de calidad que aborda características como la funcionalidad, confiabilidad, usabilidad, eficiencia, mantenibilidad y portabilidad. Aunque no está específicamente diseñado para Ingeniería de Datos, muchos de sus principios pueden aplicarse a la gestión de la calidad de los datos.

Las herramientas utilizadas en ISO/IEC 25012:2008 conocido como SQuaRE (Software Quality Requirements and Evaluation) incluyen:

- **Revisión documental:** La revisión documental es el proceso de examinar documentos relacionados con la GI para determinar su conformidad con los requisitos de la norma.
- **Entrevistas:** Las entrevistas son el proceso de hablar con las personas que están involucradas en la gestión de la GI para obtener información sobre su implementación.
- **Observación:** La observación es el proceso de observar las actividades relacionadas con la gestión de la GI para determinar su conformidad con los requisitos de la norma.
- **Pruebas:** Las pruebas son el proceso de ejecutar pruebas para determinar si un sistema de GI cumple con los requisitos de la norma.

**ISO 8000-61:2019:** Este estándar se centra en la calidad de los datos y proporciona un marco para la especificación de requisitos de calidad de datos. Define un conjunto de atributos y sus valores que se pueden utilizar para describir la calidad de los datos.

Las herramientas utilizadas en ISO 8000-61:2019 incluyen:

- **Análisis de riesgos:** El análisis de riesgos es el proceso de identificar, evaluar y mitigar los riesgos relacionados con la seguridad de la información.
- **Controles de seguridad:** Los controles de seguridad son las medidas que se implementan para mitigar los riesgos relacionados con la seguridad de la información.
- **Procesos de seguridad:** Los procesos de seguridad son los procesos que se utilizan para gestionar la seguridad de la información.
- **Requisitos legales y reglamentarios:** Los requisitos legales y reglamentarios son los requisitos que deben cumplirse para cumplir con las leyes y regulaciones aplicables.

**DAMA-DMBOK (Data Management Body of Knowledge):** Aunque no es un estándar ISO, el DAMA-DMBOK es una guía de mejores prácticas en la gestión de datos. Proporciona una estructura para la gestión de datos y aborda aspectos como la calidad de los datos, la gobernanza de datos y la arquitectura de datos.

Las herramientas utilizadas en el DAMA-DMBOK incluyen:

- Estrategia de datos: La estrategia de datos es un plan para gestionar los datos de una organización.
- Plan de datos: El plan de datos es un documento que describe cómo se implementará la estrategia de datos.
- Arquitectura de datos: La arquitectura de datos es la estructura de los datos de una organización.
- Gestión de la calidad de los datos: La gestión de la calidad de los datos es el proceso de garantizar la calidad de los datos de una organización.

El DAMA-DMBOK (Data Management Body of Knowledge) es un marco de trabajo que proporciona una guía integral para la gestión de datos. Aunque no es una norma o estándar, se utiliza comúnmente como una referencia en la disciplina de gestión de datos. Aquí hay algunas áreas y aplicaciones en las que el DAMA-DMBOK puede ser utilizado:

- Gestión de Datos Empresariales: Organizaciones que buscan establecer y mejorar sus prácticas de gestión de datos pueden utilizar el DAMA-DMBOK como una guía integral para desarrollar un enfoque estructurado y coherente.
- Desarrollo de Estrategias de Gestión de Datos: Al desarrollar estrategias para la gestión de datos, las empresas pueden referirse al DAMA-DMBOK para obtener orientación sobre la planificación estratégica, la gobernanza de datos y la gestión de activos de información.
- Gobierno de Datos: El DAMA-DMBOK proporciona principios y prácticas para establecer un marco de gobierno de datos efectivo, asegurando la responsabilidad y la toma de decisiones relacionadas con los datos en toda la organización.
- Calidad de Datos: En el ámbito de la calidad de datos, el DAMA-DMBOK ofrece directrices sobre cómo definir, medir y mejorar la calidad de los datos, asegurando que los datos sean precisos, completos y confiables.
- Integración de Datos: Para proyectos de integración de datos, el DAMA-DMBOK puede proporcionar principios para la identificación, transformación y carga (ETL) de datos, asegurando la coherencia y la integridad de los datos.
- Arquitectura de Datos: En el desarrollo de la arquitectura de datos, el DAMA-DMBOK ofrece pautas para el diseño y la implementación de estructuras de datos que respalden las necesidades empresariales y faciliten la integración y la interoperabilidad.

- Seguridad de Datos: Para abordar la seguridad de los datos, el DAMA-DMBOK ofrece directrices para la clasificación de datos, la gestión de acceso y la protección de la privacidad de la información.
- Metadata Management (Gestión de Metadatos): En proyectos de gestión de metadatos, el DAMA-DMBOK puede proporcionar pautas para la identificación, almacenamiento y gestión de metadatos que describen y contextualizan los datos.
- Cumplimiento Normativo: En entornos donde el cumplimiento normativo es crítico, el DAMA-DMBOK puede ser utilizado para establecer prácticas que aseguren la conformidad con regulaciones relacionadas con la privacidad, la seguridad y la calidad de los datos.
- Desarrollo de Competencias en Gestión de Datos: Para desarrollar habilidades y competencias en gestión de datos, el DAMA-DMBOK proporciona un conjunto estructurado de conocimientos y habilidades que son esenciales para los profesionales de datos.

**ISO/IEC 27001 e ISO/IEC 27002:** Estos estándares se centran en la seguridad de la información, pero la calidad de los datos es un componente crucial de la seguridad de la información. La protección de la calidad de los datos contra amenazas y riesgos es esencial en el ámbito de la Ingeniería de Datos.

Las herramientas utilizadas en ISO/IEC 27001 e ISO/IEC 27002 incluyen:

- Análisis de riesgos: El análisis de riesgos es el proceso de identificar, evaluar y mitigar los riesgos relacionados con la seguridad de la información.
- Controles de seguridad: Los controles de seguridad son las medidas que se implementan para mitigar los riesgos relacionados con la seguridad de la información.
- Procesos de seguridad: Los procesos de seguridad son los procesos que se utilizan para gestionar la seguridad de la información.
- Requisitos legales y reglamentarios: Los requisitos legales y reglamentarios son los requisitos que deben cumplirse para cumplir con las leyes y regulaciones aplicables.

## Referencias

- Mordinyi, Richard & Pacha, Alexander & Biffl, Stefan. (2011). Quality assurance for data from low-tech participants in distributed automation engineering environments. IEEE International Conference on Emerging Technologies and Factory Automation, ETFA. 1-4. 10.1109/ETFA.2011.6059149.
- Staegemann, Daniel & Volk, Matthias & Turowski, Klaus. (2021). Quality Assurance in Big Data Engineering - A Metareview. Complex Systems Informatics and Modeling Quarterly. 1-14. 10.7250/csimq.2021-28.01.