



BCIT in **Data Management**



Objetivos

El objetivo general de la formación consiste en preparar a los alumnos de forma intensiva en el área de integración de datos. Esta área de conocimiento está enmarcada en el área Data Management y persigue obtener un conjunto de datos dentro de la organización con los requisitos especificados por negocio.

Este proceso de obtención de datos no es un simple transporte desde diferentes orígenes sino que permite implementar conceptos como trazabilidad, calidad uniformidad, historificación, etc.

En concreto, los objetivos específicos de la formación son:

- Adquirir conocimientos sobre integración de datos y sus procesos de captura, extracción, manipulación y conversión de datos en diferentes entornos.
- Aplicar conocimientos adquiridos a casos prácticos
- Conocer y aplicar herramientas del área de Data Quality:
- Introducir conceptos de arquitectura (Data Architecture)
- Introducir conceptos de diseño de bases de datos (Data Modeling).
- Comprender la problemática de no usar metodologías de desarrollo.

Requisitos

Es necesario tener nociones básicas en lenguaje de consultas SQL.

Metodología

En esta formación, el aprendizaje se llevará a cabo de **una forma dinámica** con una enseñanza teórica y práctica compaginando ambas mediante la realización de ejercicios que permitan hacer uso de todo lo aprendido buscando asentar el conocimiento.

Programa

El programa se estructura en los bloques que se describen a continuación:

Bloque 1: Introducción a Data Management

- Evolución de mundo data 1960-2020
- Necesidades que motivan un cambio en la infraestructura: cambios estratégicos en las empresas, crecimiento en el volumen de información, cambio en las tecnologías, desarrollo de nuevas aplicaciones, etc.
- Un aliado en la metodología: DAMA y su definición de las distintas operaciones que se relacionan con los datos.
- Puntos frontera con otras áreas tecnológicas: BigData, DevOps, DataScience y Programación
- Data Architecture
- Data Storage & Ops
- Data Integration (Interoperability)
- Data Security
- Data Modeling & Design
- Documents & Content
- Data Warehousing & BI
- Data Quality
- Metadata Management
- Master Data Management & Reference Data Management
- Data Governance

Bloque 2: Structured Query Language (SQL)

- Introducción
- Consultas simples (WHERE, DISTINCT, TOP, ORDER...)
- Consultas multitabla (JOIN, UNION, INTERSECT...)
- Consultas de resumen (GROUP BY, HAVING, COUNT, SUM...)

- Subconsultas y anidamiento
- Funciones ventana
- Actualización de datos (INSERT,DELETE, UPDATE)
- Creación de objetos en base de datos (CREATE, ALTER, DROP...)

Bloque 3: Introducción a la integración de datos

- Integración de datos usando ETLs
- La necesidad de asegurar la calidad de los datos.
- La necesidad de asegurar la trazabilidad. Concepto de linaje
- La necesidad de buenas prácticas: tipado y naming correctos en columnas de BBDD
- Opciones tecnológicas disponibles: SSIS (vehicular en la formación), Pentaho, Data Stage, PowerCenter, etc.
- Otras opciones tecnológicas que no son ETLs y que nos podemos encontrar: IDQ, Collibra, IBM Cognos, etc.

Bloque 4: Conceptos de diseño de un Data Warehouse

- La importancia de un buen diseño en base de datos: integridad, rendimiento y legibilidad de los datos
- Claves: Primarias | Foráneas | Compuestas
- Normalización: 1a FN | 2a FN | 3a FN.
- Relaciones: 1:1 | 1:n | N:n
- Arquitecturas tipo en el diseño de Data Warehouses
- STG -> TRF -> DWH -> HIST
- Tipos de tablas: maestros, tablas relacionales, lookups
- Formas normales, diseño desagregado vs diseño agregado
- Gestión de datos históricos: cargas incrementales vs cargas completas

Bloque 5: SSIS

- Instalación

- Orígenes de Datos
- ControlFlow/Workflow
- DataFlow/Mapping
- Parámetros / Variables de usuario/Variables de entorno
- Historificación
- Data Quality / Validaciones
- Despliegues / Implementación en servidor
- Seguridad
- Ejecución ETL
- Automatización carga (Jobs). Revisión de ejecuciones
- Control versiones (Git o Azure DevOps)
- Buenas prácticas

Bloque 6: Especialización

En este bloque se llevará a cabo una especialización en una de las siguientes herramientas:

- Pentaho
- Informática PowerCenter
- Talend
- IBM Data Stage
- SAP Data Services
- Azure Data Analytics
 - Data Factory
 - Synapse
 - Azure DevOps



binaia.es