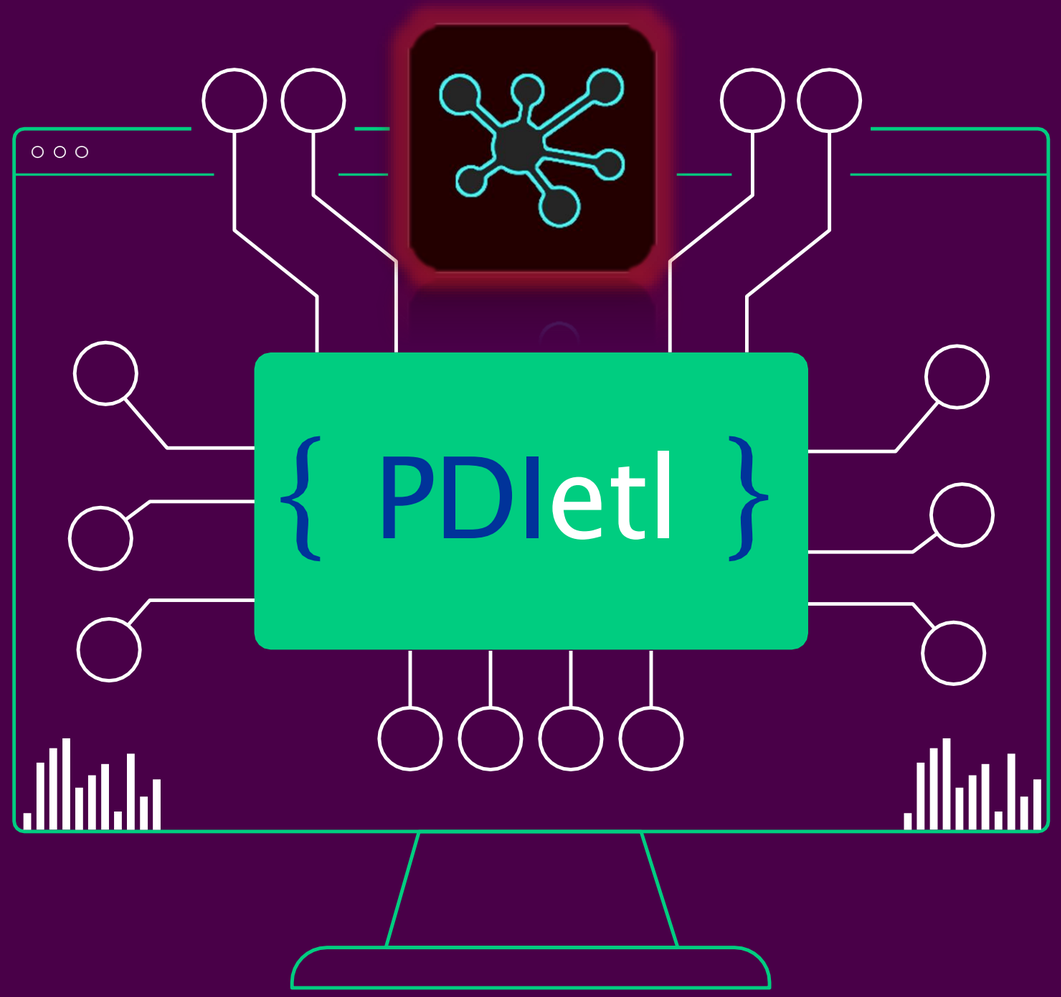
Pack Virtual

CURSO

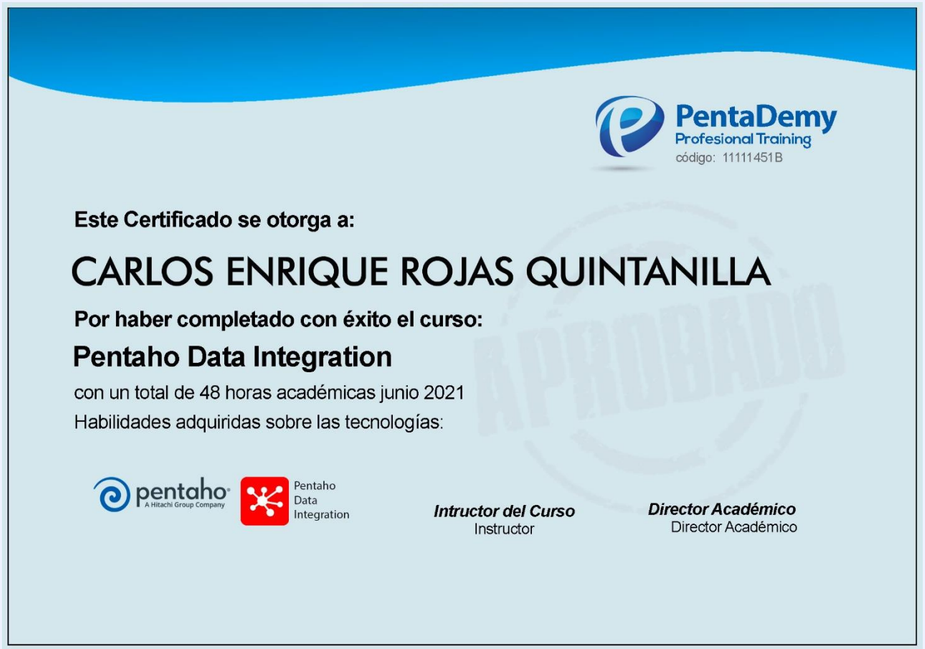
PENTAHO DATA INTEGRATION BA3000 INGENIERÍA DE DATOS

Desarrollando procesos ETL con Pentaho Data Integration



Certificación Validada

- ✓ **PENTAHO DATA INTEGRATION**


PentaDemy
Profesional Training
código: 11111451B

Este Certificado se otorga a:

CARLOS ENRIQUE ROJAS QUINTANILLA

Por haber completado con éxito el curso:

Pentaho Data Integration

con un total de 48 horas académicas junio 2021

Habilidades adquiridas sobre las tecnologías:

pentaho A Hitachi Group Company — Pentaho Data Integration

Intructor del Curso Instructor

Director Académico Director Académico

APROBADO

Certificado validez internacional

Nuestros certificados tiene validez en todos los países de Latinoamérica (a excepción de Brasil), código QR y validez en Linkedin con lo cual podrás compartir tu certificado en

LinkedIn

Pack
Virtual

RESUMEN

Aprenderás de forma progresiva y práctica los conceptos clave, para utilizar **Pentaho Data Integration** en diversas arquitecturas para **Business Intelligeence y Big Data**, verás de forma gradual cada una de las opciones de Spoon, trabajarás con los steps más utilizados en los Jobs/Transformations. Aprenderás cómo cargar y actualizar un DW, interactuando con una variedad de orígenes de datos, **estructurados, semi-estructurados y no estructurados,** Al finalizar el curso podrás crear Jobs/Transformations altamente parametrizables y adaptadas a tu contexto..

OBJETIVO

Formar profesionales que deseen aumentar sus oportunidades laborales y enriquecer su perfil profesional con un elemento diferenciador y de gran demanda actualmente, como lo es el uso de **Pentaho Data Integration** como un orquestador en proyectos de mediana y gran escala entorno al **Business Intelligence y Big Data**.

METODOLOGÍA

- Exposición teórica de los temas
- Desarrollo de casos prácticos
- Acceso a las clases grabadas
- Acceso al material exclusivo

REQUISITOS

- Portar una laptop personal para las clases de mínimo 4 GB de RAM.

TECNOLOGÍAS

- Pentaho Data Integration
- Pentaho Report Designer
- Pentaho User Console
- MySQL,
- PostgreSQL
- Arquitectura SOA / Rest

PentaDemy

Pack
Virtual

PLATAFORMA MODERNA DE APRENDIZAJE | E-LEARNING

PentaDemy

Pack Virtual

Impartido por:

Ing. Pablo Valdivia

Chief Data Architect at GIS

Chief Executive Officer in EGS GROUP

http://www.egs.pe

ACERCA DEL EXPOSITOR:

Ingeniero de Sistemas, realizó sus estudios de ingeniería en la UNAC, complementando con estudios en administración empresarial en la PUCP, Pablo es ejecutivo senior en tecnologías de la información, con más de 20 años de experiencia como consultor nacional e internacional en proyectos de Business Analytics y Big Data, así como en la dirección de proyectos & gerencia en tecnologías de la información, es asesor empresarial y especialista en Gobierno Electrónico, con dominio de tecnologías emergentes en Cloud con proveedores tales como AWS, Azure y GCP, es especialista e instructor en soluciones de clase mundial como Pentaho, IDempiere, Odoo, SuiteCRM e instructor en tecnologías privativas como Power BI, Microstrategy, Tableau, con dominio de lenguajes R, Python, Java y dominio de sistemas Linux y Unix, posee diversas especializaciones en seguridad informática, Big Data, DevOps, Pentaho, AWS, Azure y GCP. Desde 1993, es un activista del Software Libre en proyectos como Pentaho, IDempiere, Odoo, entre otros, actualmente se desempeña como **Chief Data Architect at GIS y Chief Executive Officer in EGS GROUP**

- Ex-Director de Tecnologías TIC en la empresa transnacional Carvajal S.A.
- Ex-Director de Tecnologías TIC en el Instituto del Mar del Perú – IMARPE
- Fue asesor en la hoy Secretaría de Gobierno Digital de la Presidencia del Consejo de Ministros (ex-ONGEI) – Perú.
- Ha brindado consultorías a diversas empresas nacionales e internacionales, entre las cuales destacan: El Grupo El Comercio, AJE Group, Premier Motors, Rural Telecom, Ministerio de Crédito y Hacienda en Nicaragua, entre otras.

PENTAHO DATA INTEGRATION
DESARROLLO DE PROCESOS ETL – BA3000

PentaDemy

SESIONES

PentaDemy

01

PENTAHO DATA INTEGRATION (PDI)

- Características
- Definición y uso de integración de datos
- Licencia
- Ejemplificación de tareas de integración de datos
- Configuración de variables de entorno | Descarga | Instalación | Configuración de Driver JDBC de MySQL
- Tipos y utilización de Repositorio: Conexión con Repositorio de Pentaho BA | Repositorio en Base de Datos | Repositorio en sistema de archivos | Opción Repository Manager | Metadata
- Características y diferencias entre Transformations y Jobs

02

TRANSFORMATIONS, PANEL EXECUTION, PANEL EXECUTION RESULTS

- Características y funcionamiento de las Transformations
- Panel Execute, desplegado antes de ejecutar las transformationes/Jobs: Environment Type | Log Level | Parameters
- Descripción y ejemplificación del Panel Execution Results
- Descripción y análisis de las opciones más importantes de sus Tabs:
  - Execution History, Botón SQL
  - Logging, Step Metrics
  - Performance Graph, Metrics
  - Preview data
- Laboratorio:
  - Transformacion con cálculos lógicos y matemáticos

SESIONES

PentaDemy

03

VARIABLES DE ENTORNO, PARAMETERS, ARGUMENTS

- Descripción y uso de las Variables de Entorno
- Ejemplos y notación de las Variables de Entorno
- Descripción y uso de los Parámetros
- Modos de creación de Parámetros
- Descripción, definición y uso de Argumentos
- Descripción y uso de la opción Preview
- Práctico: creación de Transformación cuyos valores obtenidos dependa de los Parámetros asignados en la ejecución
- Práctico: creación de Transformación que obtenga valores de Argumentos, ejecute una función JavaScript y genere un documento HTML

04

EXPRESIONES REGULARES (REGEX), JAVASCRIPT (JS)

- Aplicación y ejemplos de RegEx
- Documentación y patrones más utilizados de las RegEx
- Laboratorio: Obteniendo los nombres de las librerías presentes en PDI y que mediante RegEx separe sintácticamente su nombre, extensión y versión
- Descripción y documentación de JS, Aplicación de JS en PDI
- Ejemplificación y aplicación avanzada de Step Modified Java Script Value:
  - Transform Scripts | Functions
  - Transform Constants: SKIP | ERROR | CONTINUE
  - Opciones: Position | Compatibility mode | Optimization level
  - Configuración de la Grilla Fields para obtener dataset de salida
  - Añadir, modificar y configurar distintos tipos de Script: Transform | Start | End

SESIONES

05

VARIABLES GLOBALES:

- Descripción, uso, ejemplos
- Administración de Variables Globales
- Práctico: creación de Transformation que realice las siguientes tareas: utilizar Variables de Entorno para establecer URL y nombres de archivos; trabajar con datos en formato XML; convertir filas en columnas; comparar dos flujos de datos por aproximación utilizando algoritmo Levenshtein; obtener valores mínimos y máximos; trabajar con datos JSON

DATAFLOW I

- Práctico: creación de Transfomation que realice las siguientes tareas: análisis, distribución, mapeo, clasificación, aplicación de rangos, aplicación de secuencia condicionada, conversiones
- Manejo del Dataflow, Unión básica de Datasets

06

DATAFLOW II

- Unión de Datasets con diferente Metadata
  - Unión de Datasets estableciendo condición de relación
  - Unión de Datasets de forma secuencial
  - Dividir Dataset entre diversos Steps
  - Compartir Dataset completo
  - Compartir Dataset de forma distributiva
- Laboratorio: Convertir de filas a columnas, convertir de columnas a filas, unir Datasets, mapeo y distribución de Datasets, aplicación de fórmulas avanzadas, compartir Dataset

SESIONES

07

HOPS

- Descripción y administración de Hops de Transformations y Jobs
- Configuración avanzada de Hops de Transformations: Habilitar/Deshabilitar | Cambiar dirección | Condición | Borrar | Bulk Change
- Configuración avanzada de Hops de Jobs y análisis de Status: Incondicional | Exito | Fracaso | Habilitar/Deshabilitar
- Descripción de Notas en Transformations/Jobs
- Descripción de las opciones de Grilla

08

SHARE OBJECTS

- Descripción y tipos de Objetos Compartidos
- Administración, ejemplificación y utilización de Objetos Compartidos
- Configuración de Metadata de Objetos Compartidos
- Práctico: creación de Transformation que realice las siguientes tareas; obtener diferentes archivos de salida dependiendo de condiciones establecidas en el flujo de datos; comparar flujos de datos identificando elementos nuevos, eliminados y modificados; utilizar Variables de Entorno y RegEx
- Práctico: creación de Transformations y Jobs para ejemplificar las diferentes utilizaciones de Result Filenames

SESIONES

09

JOBS

Descripción, características y principales usos

- Comportamiento y modo de funcionamiento de los Jobs
- Configuración para ejecución de Steps en paralelo
- Configuración para ejecución de Transformations por cada fila analizada del Dataset
- Análisis y explicación de Ruta de Ejecución de los Steps de Jobs
- Práctico: creación de un Job que realice las siguientes tareas; controle el workflow de ejecución de dos Transformations; evalúe la salida de status de los diferentes Steps
- Laboratorio: Generar un Dataset; guardar el Dataset en la lista Result rows; ejecutar una segunda Transformation que obtenga el Dataset de la lista Result rows; configurar salidas de log y analizar los resultados

10

BASES DE DATOS

- Descripción, uso y realización de acciones avanzadas sobre Bases de Datos:
  - Obtener Dataset
  - Insertar registros
  - Actualizar registros
  - Borrar registros
  - Añadir columna
  - Ejecutar Script SQL
- Utilización y configuración avanzada de error handling
- Definición y utilización de opción Clear Cache Database
- Práctico: creación de Job que realice múltiples tipos de acciones sobre bases de datos.

SESIONES

11

E-MAIL & WEB:

- Ejemplificación, uso y configuración avanzada de envío de e-mails
- Utilización de diferentes protocolos: POP3 | IMAP | MBOX
- Práctico: creación de Transformations y Jobs que realicen las siguientes tareas; obtener de un archivo CSV una lista de URLs web con los discos de artistas de rock; obtener el documento HTML de cada URL web; filtrar de cada documento HTML la sección dedicada a la lista de canciones de cada disco; generar un archivo CSV por cada disco con la información de sus respectivas canciones.
- Práctico: creación de un Job que realice las siguientes tareas: utilizar Variables de Entorno y RegEx para obtener una lista de archivos; validar direcciones de e-mail; enviar e-mail que contenga como adjuntos los archivos obtenidos

12

PAN & KITCHEN

- Descripción de las principales herramientas PDI: Spoon | Pan | Kitchen | Carte
- Opciones avanzadas ejecución de Transformations o Jobs por líneas de comandos
  - Parámetros
  - Argumentos
  - Registro Log

SCHEDULING

- Descripción, ejemplificación y uso de Calendarización de ejecución de Transformations y Jobs
- Calendarización utilizando Cron
- Calendarización utilizando Task Scheduler

SESIONES

13

MARKETPLACE

- Descripción y características del Marketplace de PDI
- Instalación de plugins: Weka, DataCleaner

TRANSFORMACIONES COMO DATA SOURCES:

- Utilización de Transformation como Datasource para Dashboards (CDE)
- Utilización de Transformation como Datasource para Reporting (PRD)

Bonus Track: Delivery PRD

- Configuración y ejecución de reportes PRD en Transformation PDI