

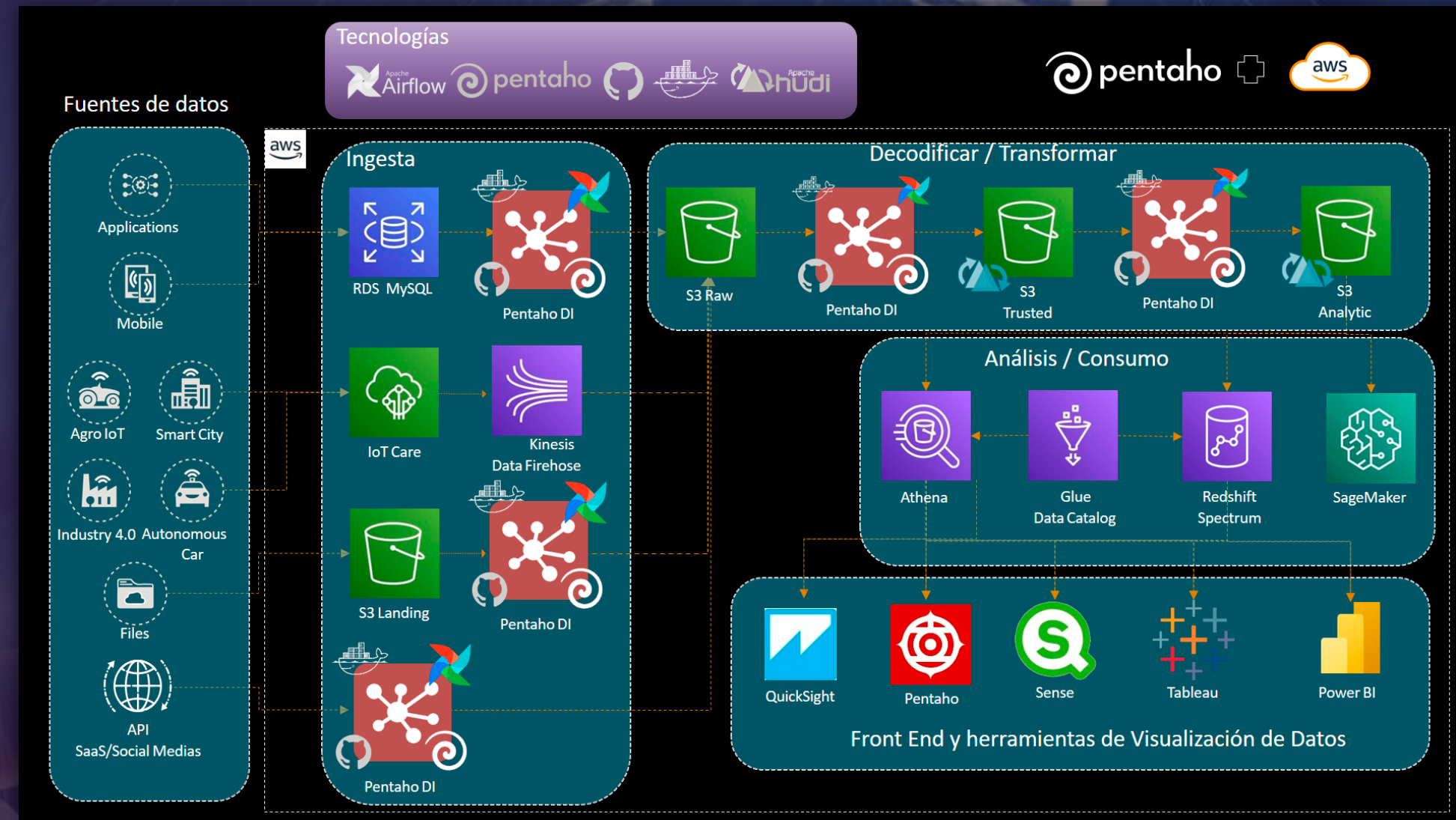


Pack
Virtual

C
U
R
S
O

Big Data con AWS y Orquestación con Pentaho BA

Curso 100% en español con más de 25 laboratorios de casos reales y certificación en Big Data en AWS con Pentaho

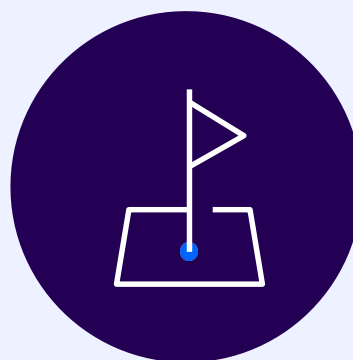




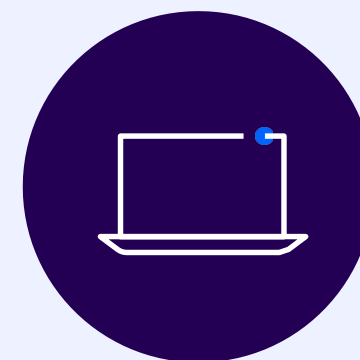
BIG DATA: INGENIERÍA DE DATOS CON AWS & PENTAHO



Inicio:
17 DE MARZO



Finalización:
14 DE ABRIL



44 HORAS
académicas



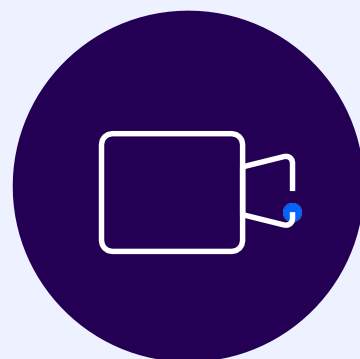
Martes y Jueves
**De 07:00 pm a
10:00 pm (GMT-5)**



Soporte
TÉCNICO



Plataforma
DIGITAL



Aplicativo
Google Meet



**CERTIFICACIÓN
POR NIVEL**



**DIPLOMAS
DIGITALES DEL
PROGRAMA**



Certificación



✓ Big Data Architec con Pentaho y AWS



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





RESUMEN

El equipo académico de **PentaDemy** cuenta con profesionales que se dedican a desarrollar proyectos TIC en empresas líderes de diversos sectores económicos, enfrentándose a todo tipo de retos. Gracias a esta experiencia, **PentaDemy** propone diferentes cursos que ayudarán a los expertos a incrementar sus skills profesionales permitiéndoles así, abordar proyectos de Big , ya sean **Real Time, near Real Time** o procesamiento **Batch**, domina desde la ingestión de datos, el procesamiento, el delivery y el descubrimiento con la analítica, sorbe AWS con Pentaho

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 los servicios de Big Data Cloud sobre AWS.

METODOLOGÍA

- ➔ Exposición teórica de los temas
- ➔ Desarrollo de casos prácticos
- ➔ Acceso a las clases grabadas
- ➔ Acceso al material exclusivo
- ➔ Acceso a una Maquina Virtual con Clusters reales de Big Data



REQUISITOS

- ➔ Conocimientos básicos de SQL
- ➔ Portar una laptop personal para las clases de mínimo 1GB de RAM para el uso de Clúster de 128 GB en la nube



TECNOLOGÍAS

- ➔ AWS – Amazon Web Services
- ➔ KINESIS FIREHOSE
- ➔ FINESIS DATA STREAM
- ➔ DMS
- ➔ SHOWBALL
- ➔ RDS
- ➔ S3
- ➔ DYNAMODB
- ➔ REDSHIFT
- ➔ AWS LOT
- ➔ EMR,HADOOP, SPARK
- ➔ ATHENEA, ELASTIC SEARCH
- ➔ KMS, CLOUD HSM, CLOUD WATCH, VPC



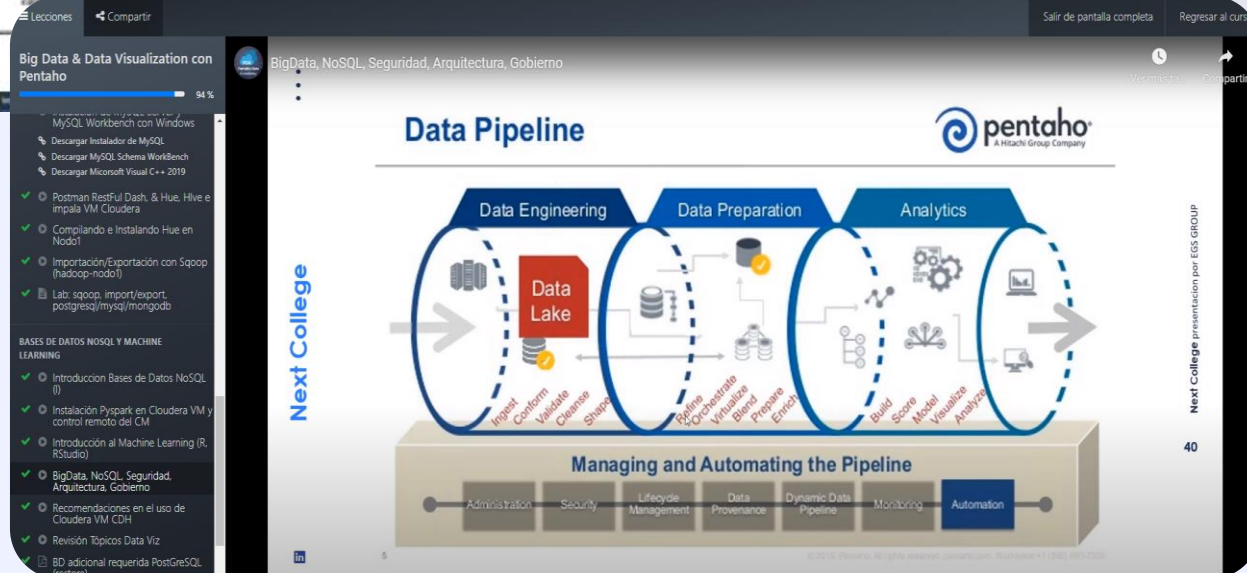
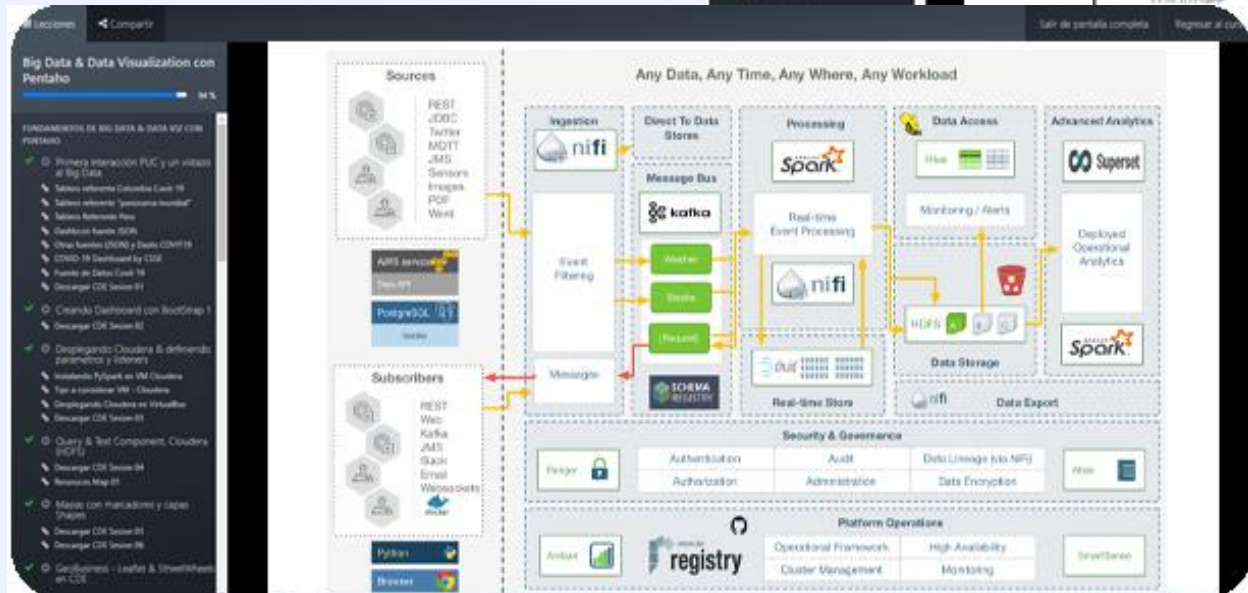
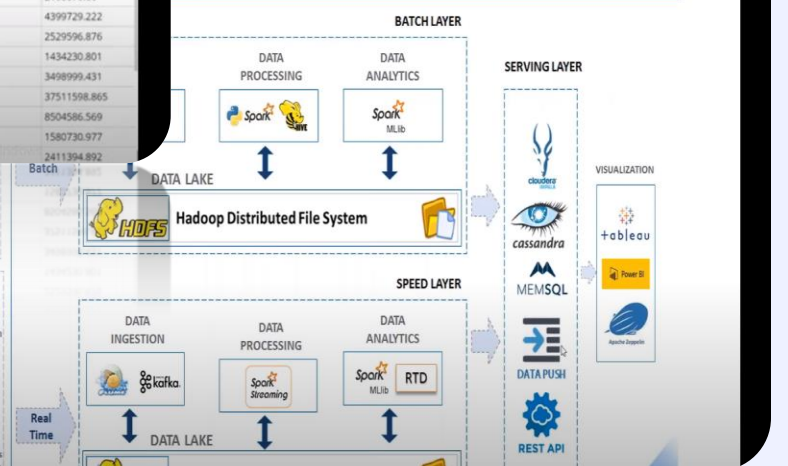
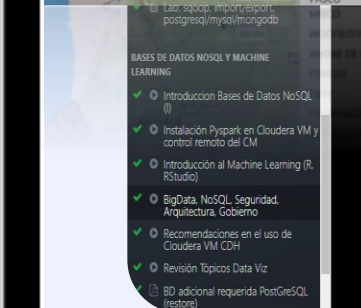
Big Data & Data Visualization con Pentaho

Abarca desde los fundamentos de Hadoop, manejo de los lenguajes de programación python, R, uso de spark y streaming, aprenderás los fundamentos en base a laboratorios de machine learning, así como a realizar la ingesta y tratamiento de datos con Pentaho.



- El diagrama ilustra el flujo de datos en un ecosistema de Big Data y Visualización, dividido en seis etapas principales:

 - Producers:** Fuentes de datos como IoT, Raspberry Pi y Minijiff.
 - Ingestion:** Servicios de ingestión como Event HUB, Event Streaming y Google Pub/Sub.
 - Stream Processing:** Servicios de procesamiento de flujo como Stream Analytics, Kinesis Analytics y Cloud Dataflow.
 - Storage:** Servicios de almacenamiento como Azure Cosmos DB, EKV DynamoDB y BigQuery.
 - Presentation:** Servicios de visualización como Power BI, Amazon QuickSight y Cloud Datavise Data Studio.
 - Architectures:** Arquitecturas específicas para Azure IoT, AWS IoT y Google IoT.





Impartido por:

Ing. Pablo Valdivia

Chief Data Architect at GIS

Chief Executive Officer in EGS GROUP

<http://www.egs.pe>

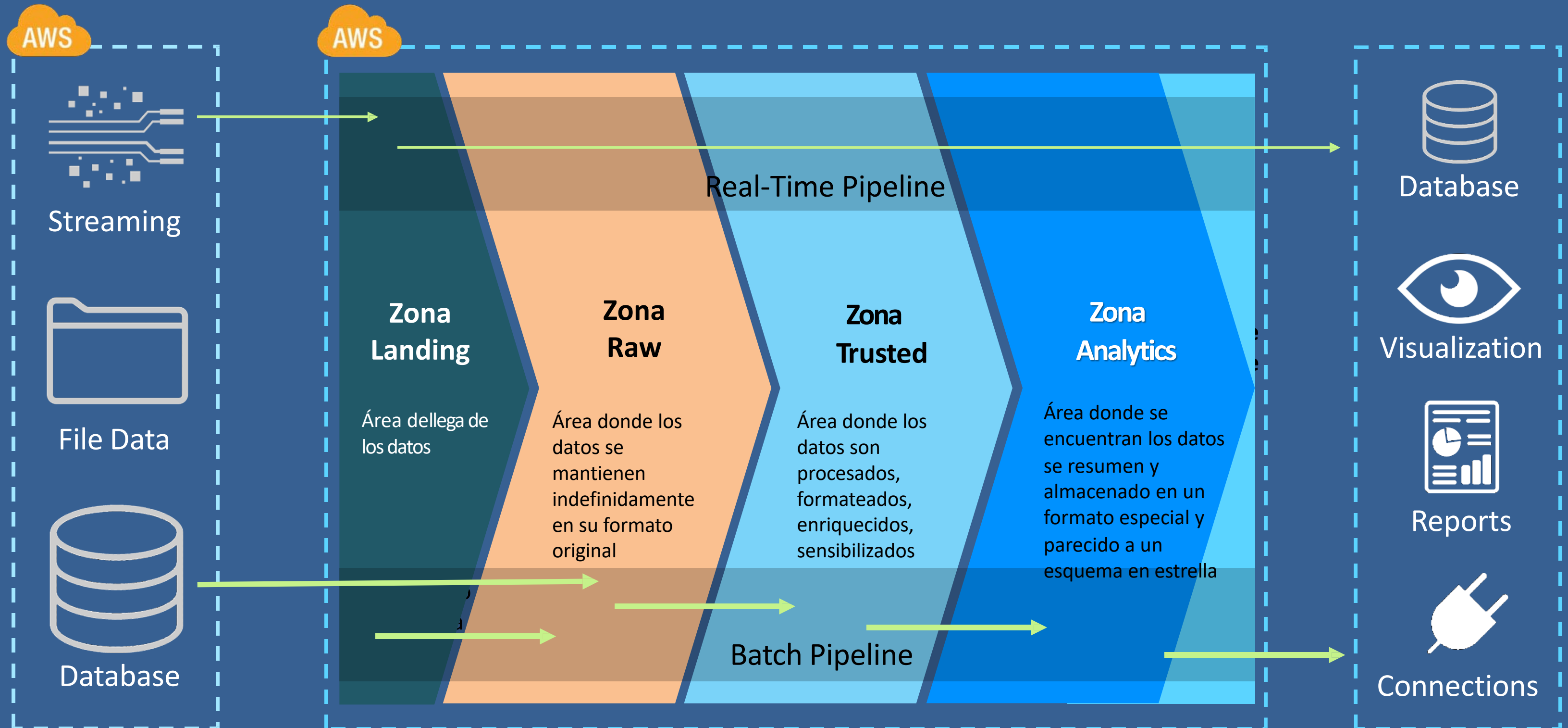
ACERCA DEL EXPOSITOR:

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.

FUNDAMENTOS DE BIG DATA EN AWS

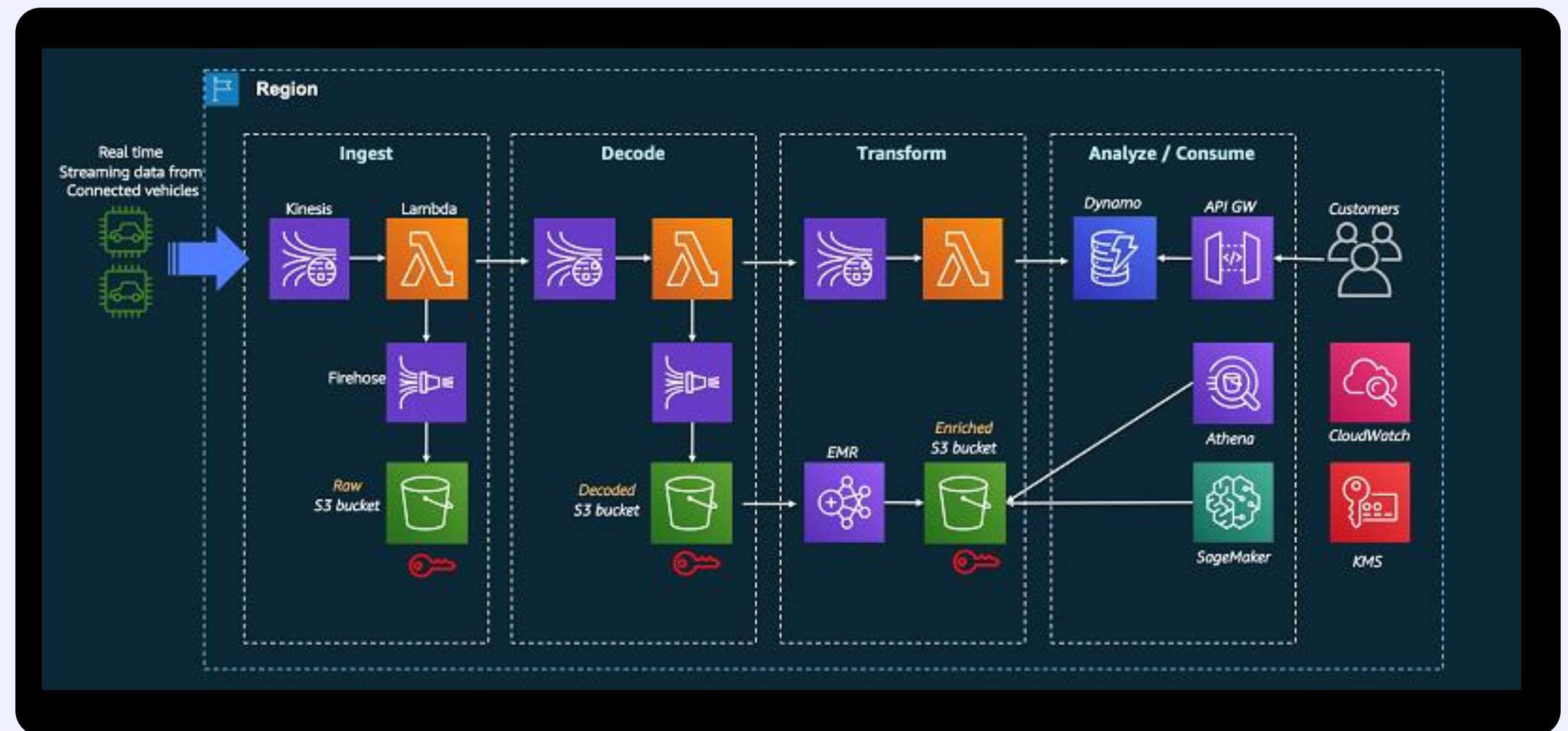
PENTAHO – DATALAKE – AMBIENTE DE TRABAJO



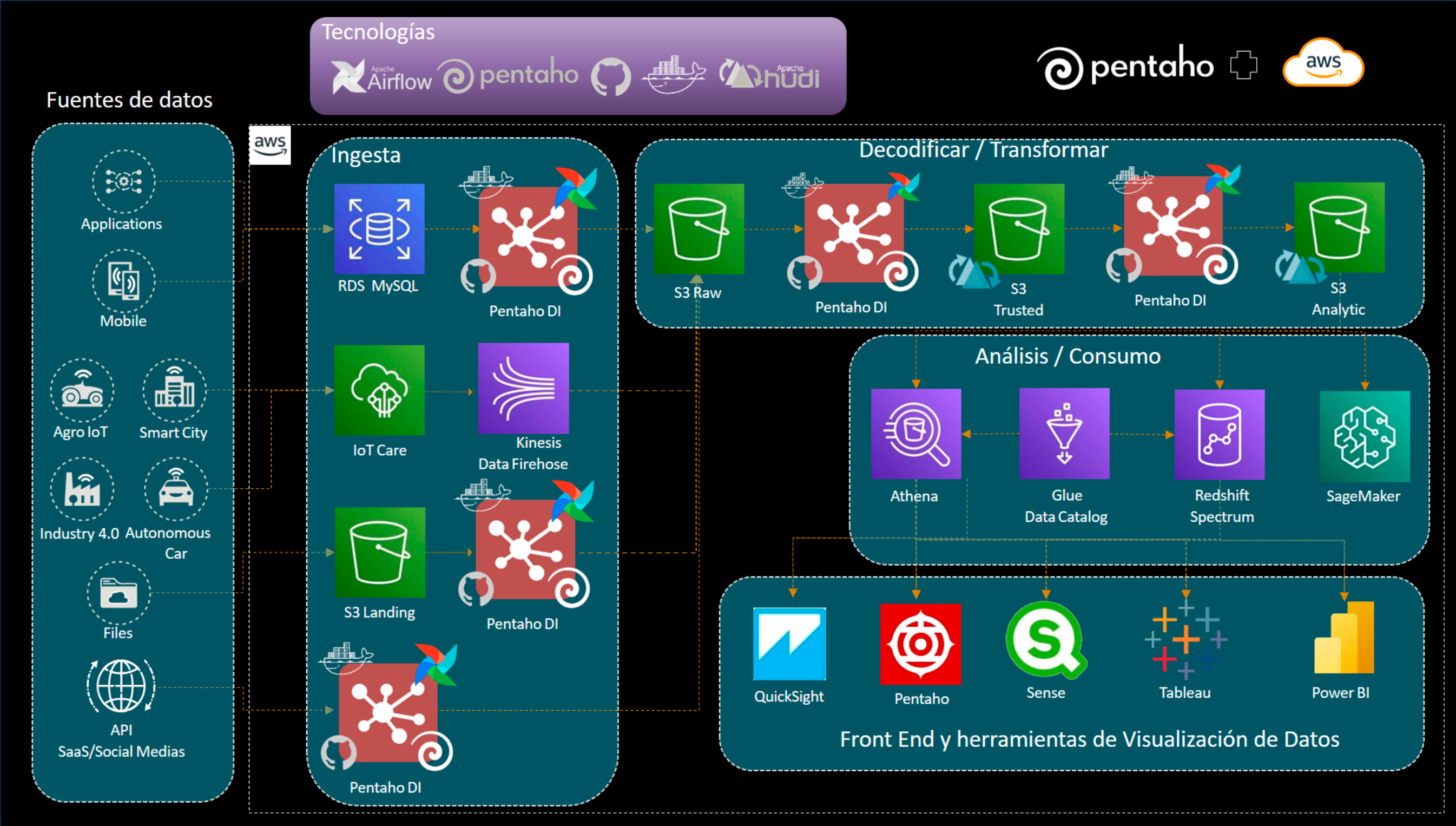
MÓDULO 01

INTRODUCCIÓN A AWS

- Presentación de conceptos de Big Data, Analytics y Cloud.
- Fases de una arquitectura Big Data Cloud.
- Presentación de mix tecnológicas (storage, processing, analytics, visualization, seguridad) existentes.
- Revisando estructura Pentaho +AWS
- Preparando ambiente: desplegando base de datos
- Preparando ambiente: desplegando Pentaho Data Integration
- Construcción de un Datalake usando AWS
- Lab01: Construyendo tu DataLake en AWS



FUNDAMENTOS DE BIG DATA EN AWS & PENTAHO – PREPARANDO AMBIENTE



MÓDULO 02

COLECCIÓN DE DATOS

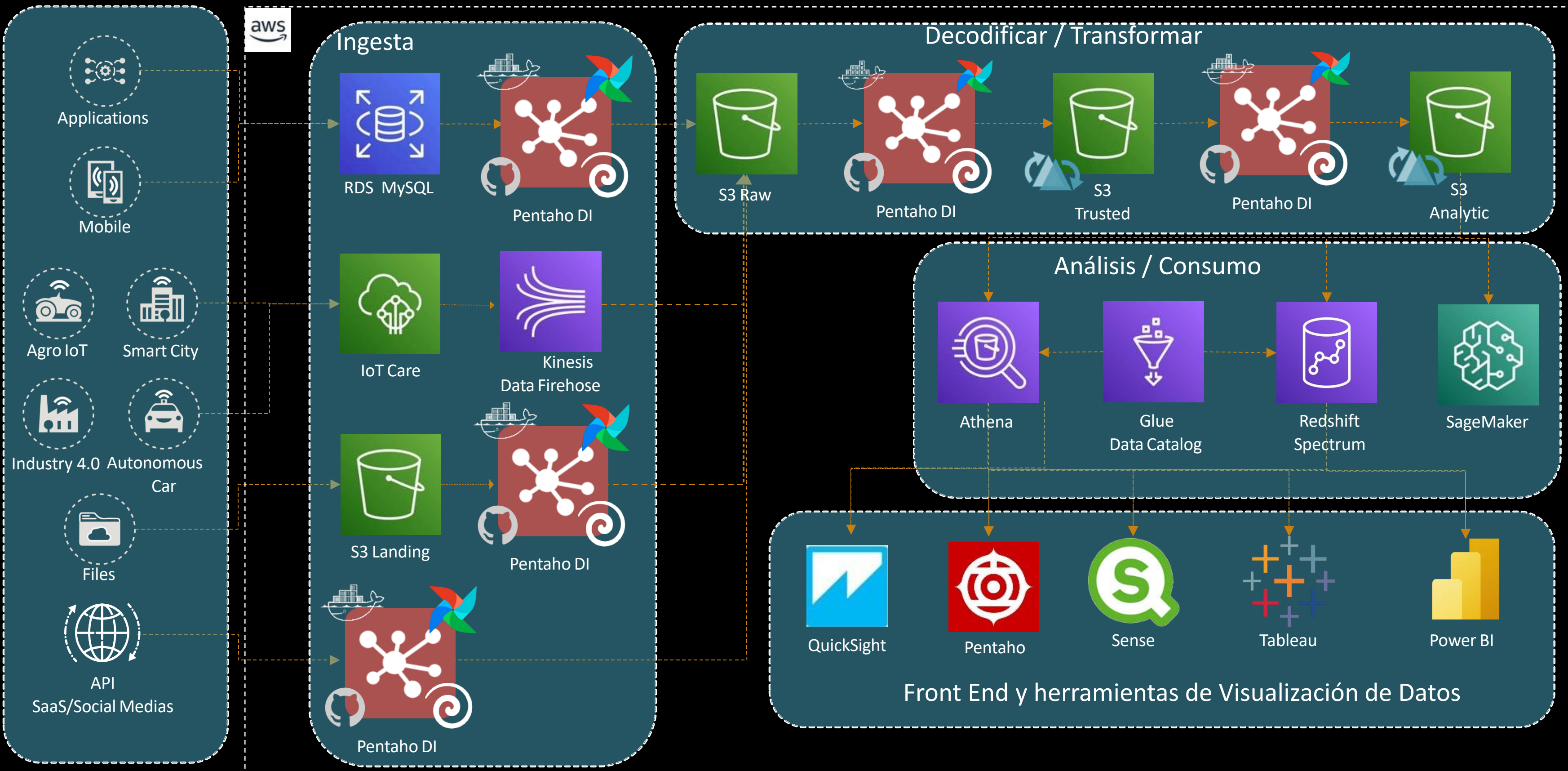
- Introducción a varios métodos de recolección de datos.
- Kinesis Firehose (Lab02: Desarrollo de un flujo de datos near real time extrayendo data de instagram).
- Kinesis Firehose (Lab02: Desarrollo de un flujo de datos near real time extrayendo data de Twitter).
- DMS (Lab04: Migración de base de datos AWS).
- Arquitectura Big Data y Pipeline con Pentaho.
- Entendiendo la ingesta de datos en Big Data con Pentaho.
- Construyendo un ETL empresarial para ingesta de datos en un entorno AWS
- Pipeline (Lab05: Construye tu primer pipeline).
- Mejores prácticas para Kinesis Data Stream.



Tecnologías



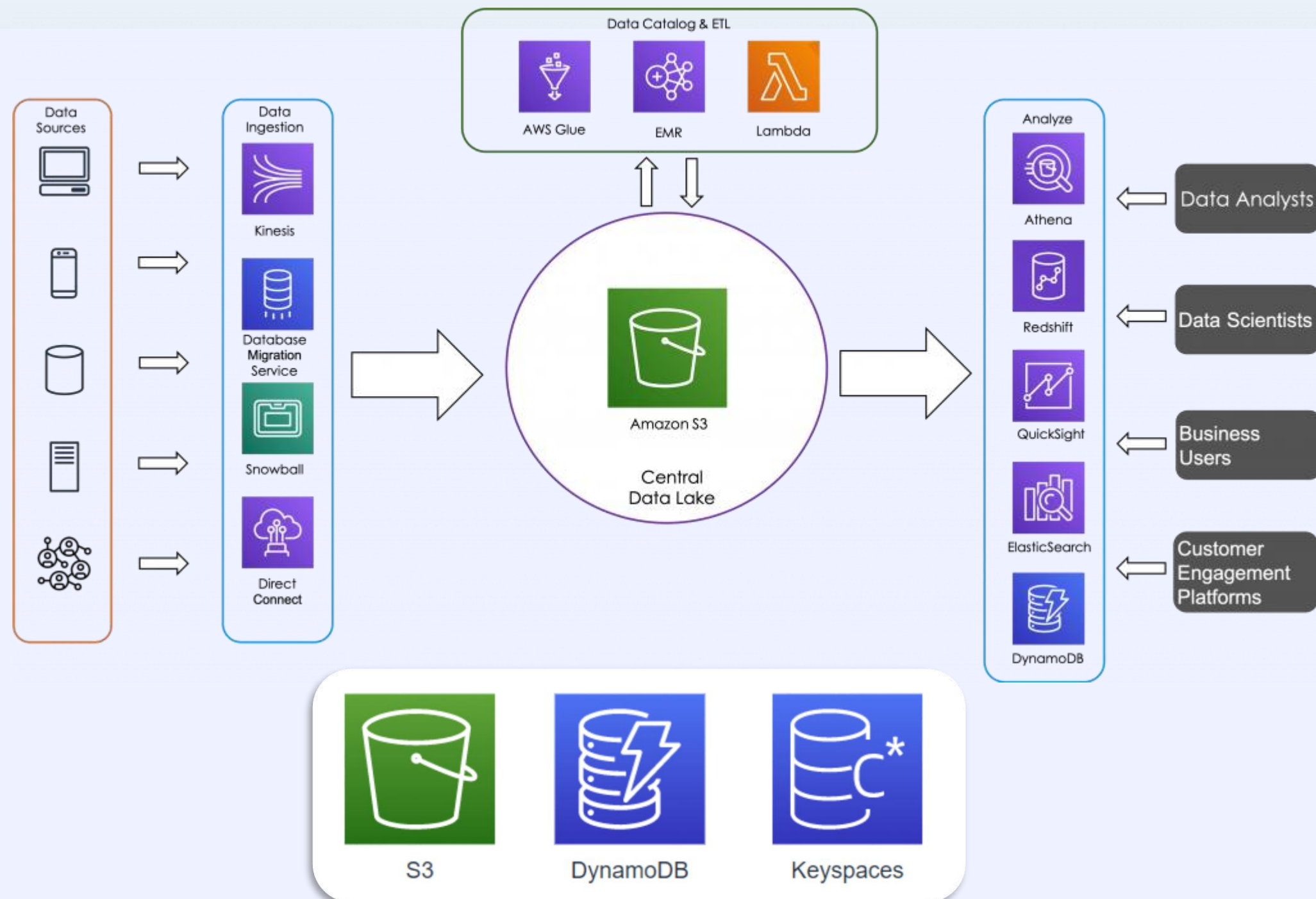
Fuentes de datos



MÓDULO 03

ALMACENAMIENTO Y GESTIÓN DE DATOS

- S3 (LAB06: Creación de bucket, reglas del ciclo de vida de S3) y Lab07: Creación de web estáticas.
- Niveles de almacenamiento S3.
- DynamoDB y WCU, particiones, DAX y Consumo por API.
- DynamoDB (Lab08: CRUD DynamoDB)
- Lab09: Integración de Kinesis Firehose y DynamoDB.
- RDS, Lab10: Creando una base de datos
- Redshift (Lab11: Configurando tu DW).
- Redshift Spectrum (Lab12: analizando datos con Redshift Spectrum).
- Data Sync (Lab13: Cargando archivos locales a AWS)
- AWS IoT (Lab14: Obtención de datos).
- AWS Greengrass (Lab15: configurando Greengrass)



MÓDULO 04

PROCESAMIENTO DE DATOS

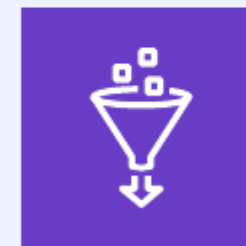
- Lambda (Workshop: Implementado una arquitectura de ingesta en real time con alertas por correo, usando Kinesis Data Streams, Lambda, DynamoDB, SNS, IAM, CloudFormation)
- Glue (Workshop: Construyendo un Data Lake y una arquitectura ETL Serverless)
- EMR (Workshop: Creando un clúster Big Data en EMR, construyendo un procesamiento de datos realtime con Spark Streaming, usando AWS IoT Core, Kinesis Data Streams, Kinesis Data Firehose, DynamoDB, Lambda, SNS, S3 y CloudFormation)
- Workshop : Implementación de un Data Lakehouse con Apache Hudi (DMS, RDS, EMR, S3, CloudFormation)
- Glue DataBrew (Workshop: Transformando datos de partidas de ajedrez con Glue DataBrew)
- Step Functions (Workshop: Orquestando un pipeline de ingesta y transformación de datos usando Glue DataBrew, Athena y SNS)
- Workshop : Creando pipeline de datos con MWAA – Airflow



Lambda



EMR



Glue



Step Functions



MÓDULO 05

ANÁLISIS Y VISUALIZACIÓN DE DATOS

- Athena (Workshop: Creando esquema y tablas de schema, tablas en formato textfile, parquet y particionadas)
- Kinesis Data Analytics (Workshop: Implementando una arquitectura de detección de fraude en tiempo real con alertas, usando Kinesis Data Analytics, Kinesis Data Streams, Lambda, DynamoDB y CloudFormation)
- SageMaker
- Redshift (Workshop: Creando tablas en un clúster de RedShift, ingestado datos desde S3 y exportando datos de RedShift a S3 en formato parquet)
- ElasticSearch (Workshop: Implementando una arquitectura de análisis de sentimientos de tweets en tiempo real, usando ElasticSearch, Kibana, Lambda, Kinesis Data Streams, DynamoDB, SNS, Comprehend y CloudFormation)
- Comprehend



Athena



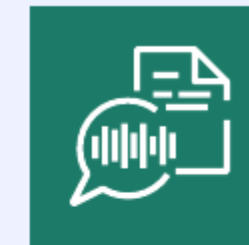
Kinesis Data Analytics



Comprehend



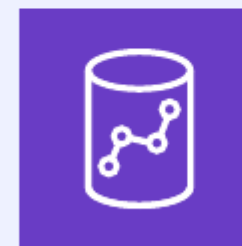
SageMaker



Transcribe



Elasticsearch



Redshift



QuickSight



Rekognition

- Rekognition (Workshop: Construyendo una arquitectura serverless para la detección del uso de mascarillas en imágenes, usando Rekognition, Lambda, DynamoDB, SNS, S3 y CloudFormation)
- Transcribe (Workshop : Análisis de sentimientos de llamadas de un callcenter, usando Transcribe, Lambda, SNS, EventBridge, Comprehend).
- QuickSight (Workshop: Creando un Dashboard)
- Apache Superset (Workshop : Visualizando datos con Superset)

MÓDULO 06

SEGURIDAD Y OTROS

- IAM (Workshop: Administrando los accesos de grupos y usuarios en AWS, manejo permisos)
- KMS (Workshop: Encriptando objetos en S3 con KMS, integración con Lambda para lectura de datos cifrados)
- Secret Manager (Workshop: Creando y obteniendo secretos, integración con Kinesis Data Streams, Lambda, DynamoDB y CloudFormation)
- CloudFormation (En todos los workshop se manejará el concepto de Infraestructura como Código, para desplegar arquitecturas Big Data en cuestión de minutos)

