

Modulo Tratamiento de Datos

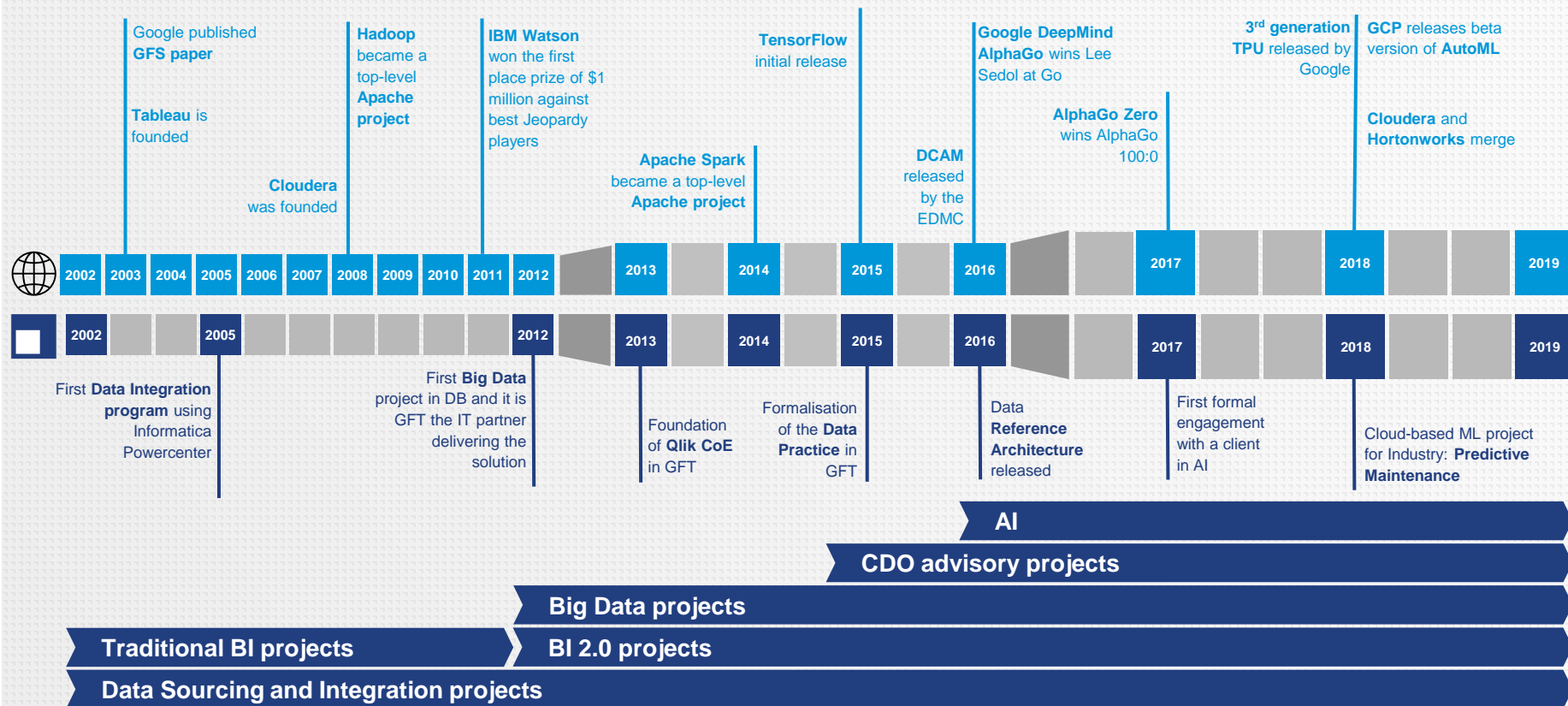
GFT

Pedro Nieto

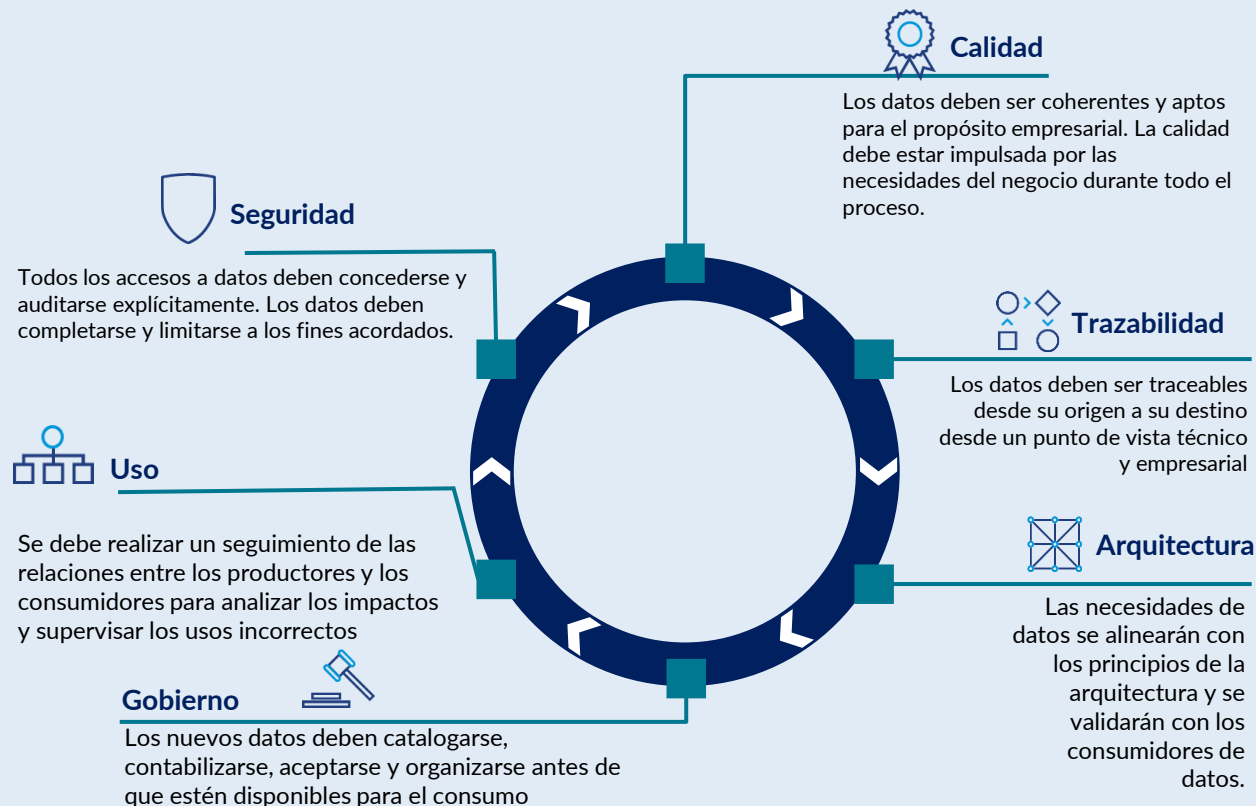


Hablemos de datos...

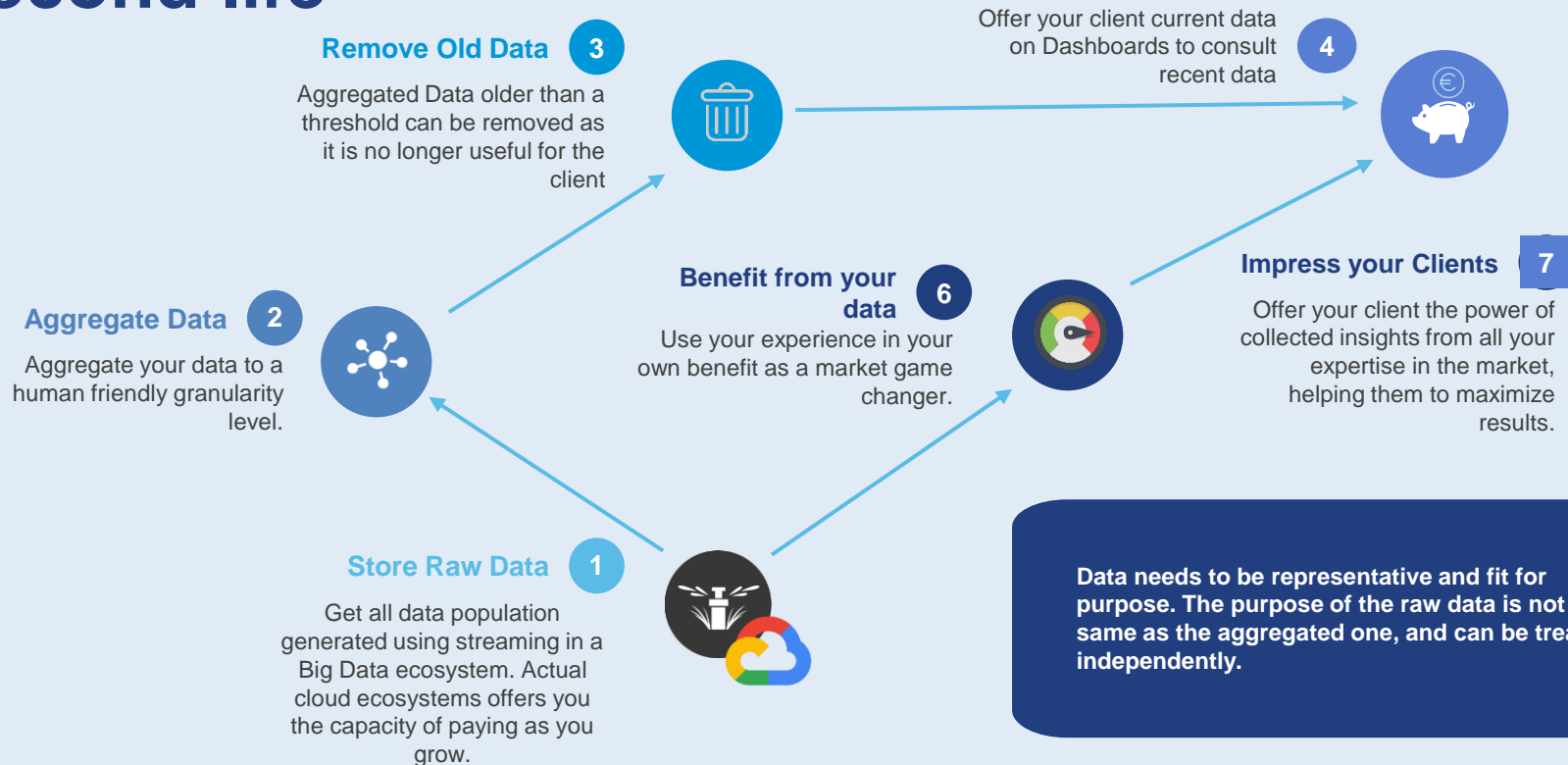
Our journey in Data



No son solo los datos...



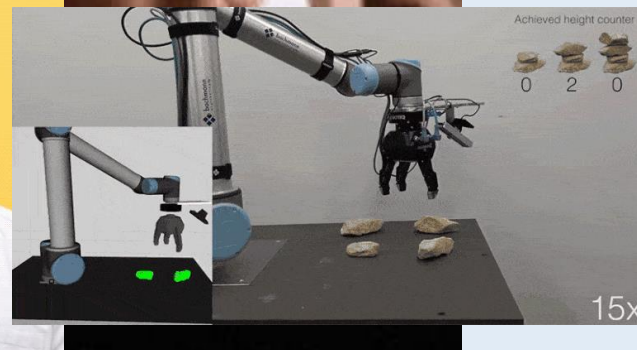
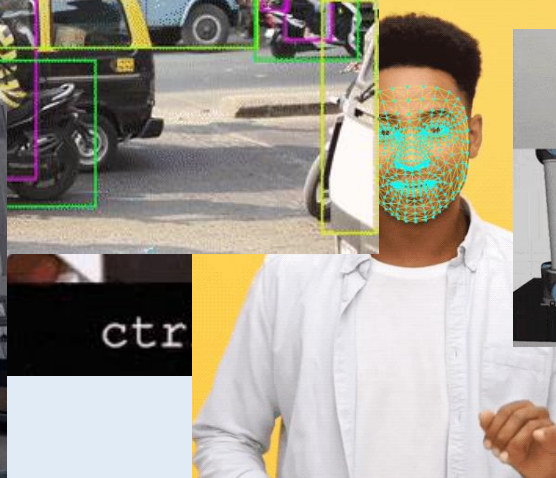
Give your data a second life



Everywhere...



What all of
them have
in
common?



Data as an Asset

Puzzle starts to fit



Data Evolution

Where are you?

Companies are evolving to embrace smart industry, as an IT company we can help you succeed in the process.



Data

Environments usually on premise store small amounts of data for transactional purposes

- Optimize Transactional Environment
- Identify weak points in your technical approach



Information

Historical Data and innovative data sources start to gain attention in the ecosystem and are stored.

- Platform Strategy (Cloud)
- IoT Integration
- Reporting Dashboards
- Process Optimization



Knowledge

Use your experience in your own benefit as a market game changer. Squeeze your data with ML and get valuable insights.

- Business Rules
- Decision Support Systems
- Machine Learning Models
- Predictive Analysis



Wisdom

Offer your new view to your clients and predict or suggest new actions.

- Model Prediction
- Process Automation
- DAAS (Data as a Service)

LinkedIn Learning

The Skills Companies Need Most in 2020



Top 5 Soft Skills

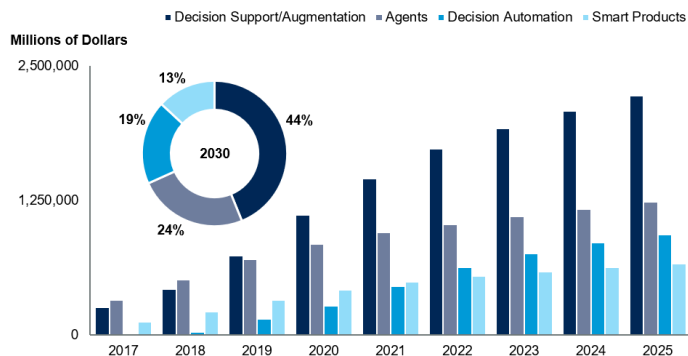
- 1 Creativity
- 2 Persuasion
- 3 Collaboration
- 4 Adaptability
- 5 Emotional intelligence



Top 10 Hard Skills

- 1 Blockchain
- 2 Cloud computing
- 3 Analytical reasoning
- 4 Artificial intelligence
- 5 UX design
- 6 Business analysis
- 7 Affiliate marketing
- 8 Sales
- 9 Scientific computing
- 10 Video production

Business Value Forecast by AI Type



Source: Gartner
ID: 386366

Las profesiones más buscadas en 2020

Sector

El más buscado

'Telecos'	Arquitecto de computación en nube	> 40.000
Tec. de la información	Desarrollador Java	> 40.000
Marketing	Jefe de producto	35.000-50.000
Comercial	Director de desarrollo empresarial	45.000-65.000
Industria	Ingeniero de automatización	35.000-45.000
Salud	Médico del trabajo	50.000-60.000
Sector financiero	Responsable de control financiero	45.000-70.000

CincoDías

LAS 15 PROFESIONES CON MÁS FUTURO EN 2020 SEGÚN LINKEDIN

Ranking en función del incremento anual de la demanda, según el "Informe Empleos Emergentes 2020" en España, elaborado por LinkedIn.



Triunfo con
LinkedIn
Consultores de Social Selling

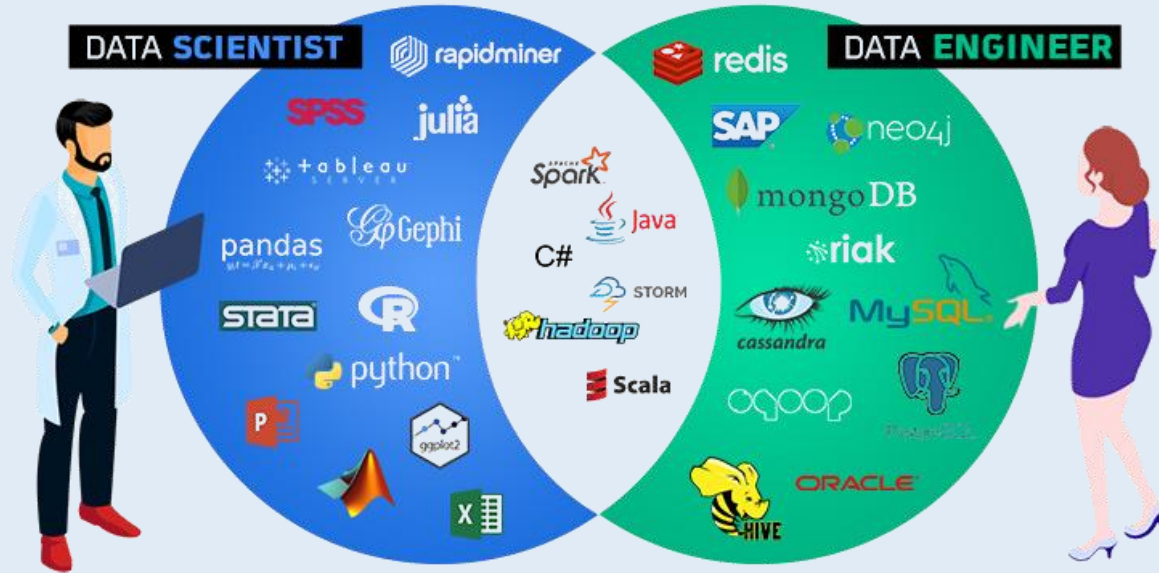
What do we expect from you?

Data Scientist also known as Data Managers, statisticians.	Data Engineers also known as database administrators and data architects.	Data Analysts also known as business Analysts.
		
A data scientist will be able to take data science projects from end to end. They can help store large amounts of data, create predictive modelling processes and present the findings.	They are versatile generalists who use computer science to help process large datasets. They typically focus on coding, cleaning up data sets, and implementing requests that come from data scientists.	They typically help people from across the company understand specific queries with charts.
Skills: Mathematics, Programming, Communication	Skills: Programming, Mathematics, Big data	Skills: Statistics, Communication, Business knowledge
  	  	  
Will use programmes such as: SQL, Python, R	Will use programmes such as: Hadoop, NoSQL, and Python	Will use programmes such as: Excel, Tableau, SQL

Data Engineers

What Will you learn?

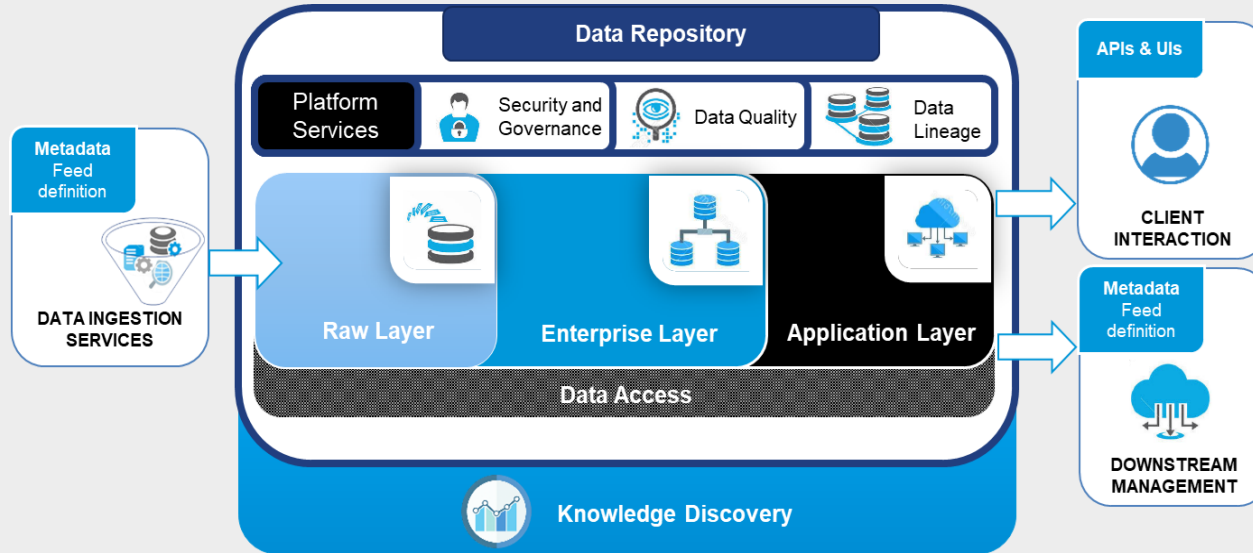
LANGUAGES, TOOLS AND SOFTWARE





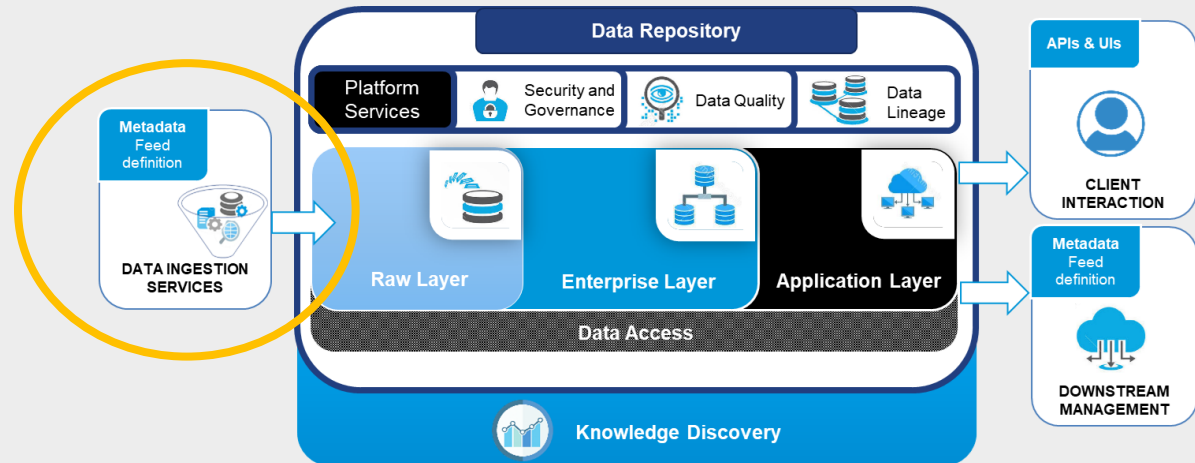
Módulo 1

Data Architecture example



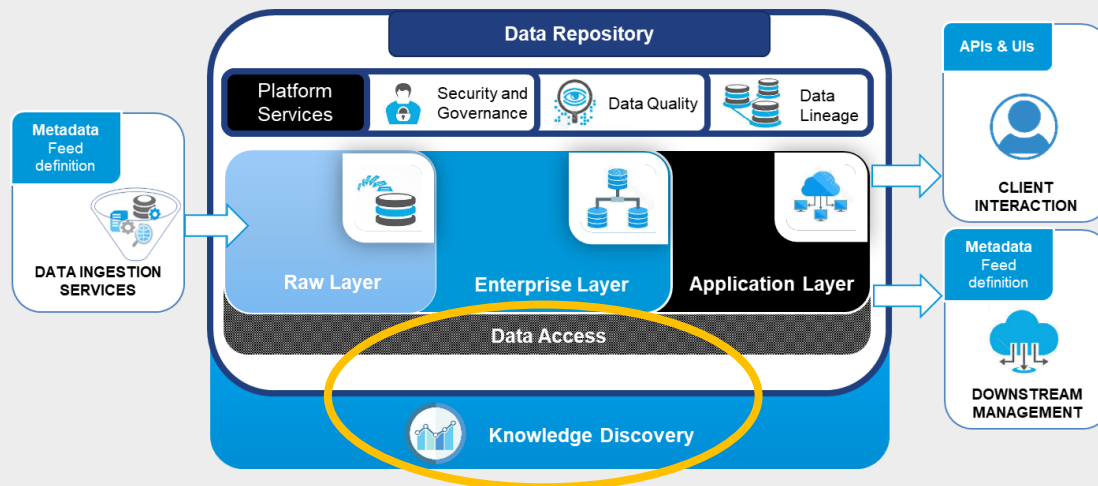
Get the data

- Data should be easily obtained, and frequently updated
- Data can come in any format or shape
- Be ready for all speeds (from batch to real-time)
- Data as a service



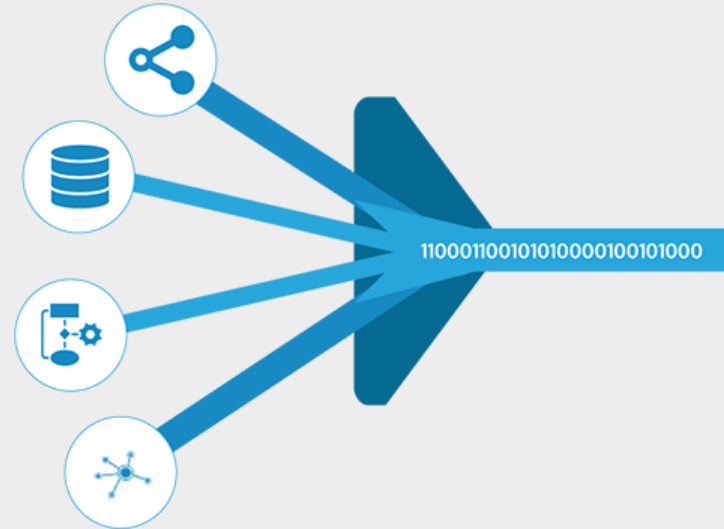
Data discovery

- We want to get not only the data, but the right data, and from the right source
- What data do we have?
- Is it relevant and enough?
- Is external or reference data required?
- Is it updated?



Captura y Recogida Datos

- **Unidad 1: Introducción a los Datos**
- **Unidad 2: Tipos de Datos**
 - Estructurados
 - No Estructurados
 - Semi-estructurados
- **Unidad 3: Integración de Fuentes**
 - Internal Data Sources
 - External Data Sources
 - Data Sources Access
- **Unidad 4: Frecuencia de Creación**
 - Batch Sources
 - Micro-batching
 - Streaming Sources



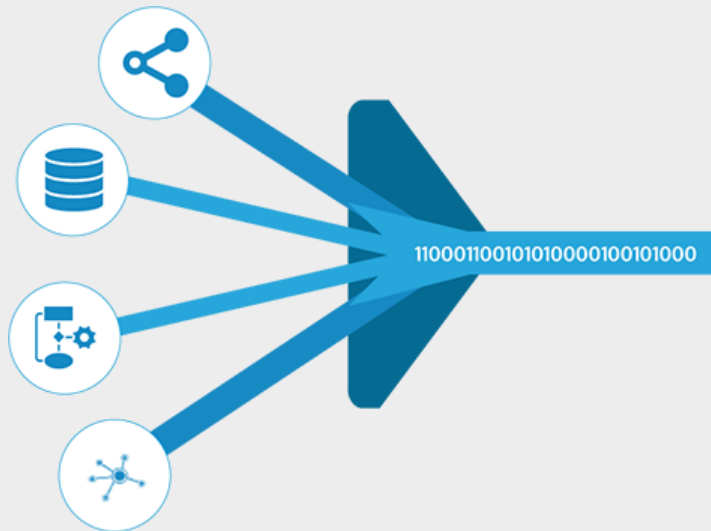
Captura y Recogida Datos

- **Unidad 5: Ingestión de Datos**

- ETL
- Ingestión Pura
- Basada en Eventos

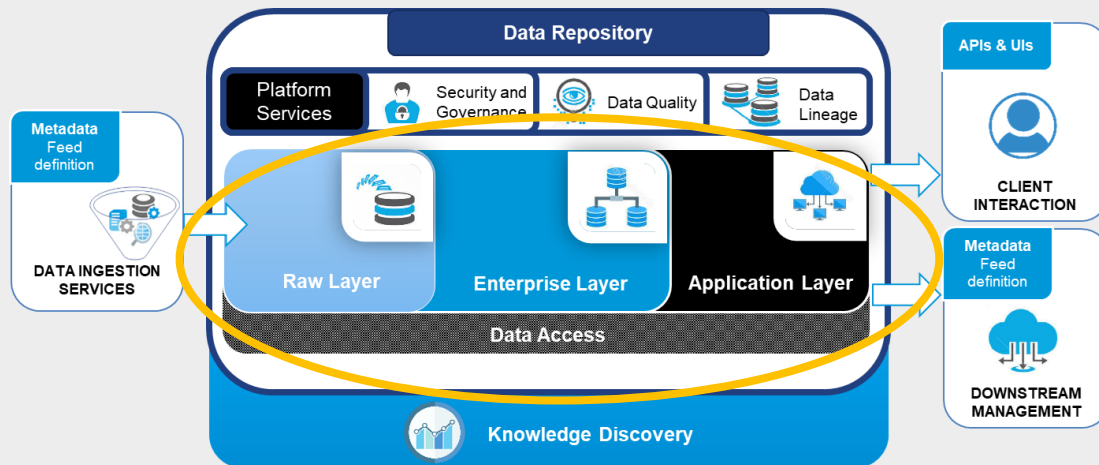
- **Unidad 6: Tipos de Fuentes**

- Open Data
- Public DataSets
- Data as a Service



Prepare the data

- Maximize analysts time by having the data ready
- Data transformations
- Filtering
- Data quality checks
- Enrichment
- Use a common repository



Almacenamiento y procesamiento

- Unidad 1: Arquitecturas de Datos
 - Batch Architectures
 - Real Time Architectures
 - Lambda & Kappa Architectures
- Unidad 2: Tipos de almacenamiento
 - BBDD Relacionales
 - BBDD No-SQL
 - Documentos Indexados
 - Sistemas almacenamiento columnar
 - Almacenamiento distribuido
 - Almacenamiento serializado



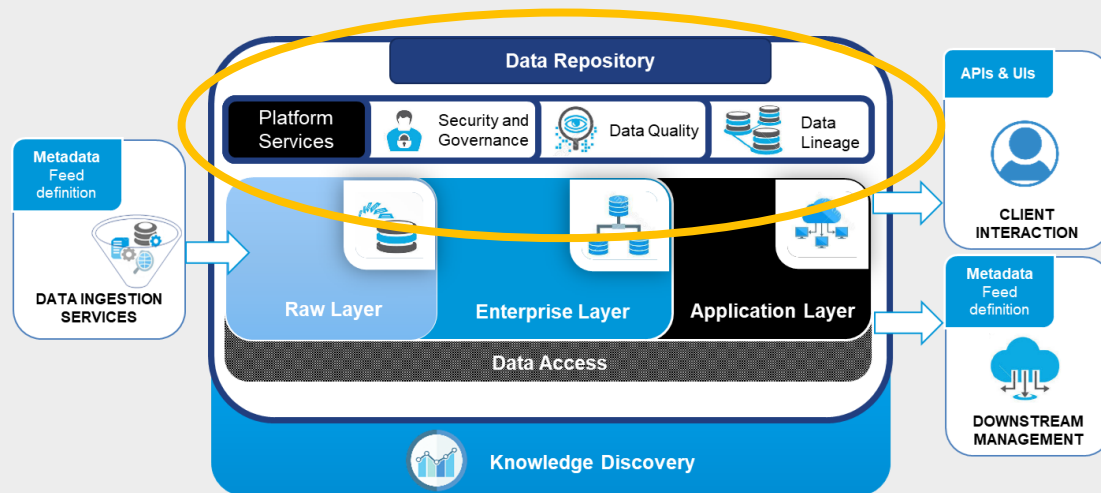
Almacenamiento y procesamiento

- Unidad 3: Procesamiento de Datos
 - Procesamiento en modo Batch
 - Procesamiento en Streaming
- Unidad 4: Privacidad y protección de Datos
- Unidad 5: Tipos de Infraestructuras
 - Cloud (IAAS, PAAS, SAAS)
 - On premises
 - Híbridas



Organize the data

- Separate and properly catalog the data
- Tag and complete with metadata
- Data stewardship



Gobierno del Dato

- Unidad 1: Origen del problema
 - Crecimiento Orgánico
 - Ciclo de vida de los datos
 - Data Oriented Architecture
- Unidad 2: Estándares de Mercado
- Unidad 3: Data Governance
 - Roles y Responsabilidades
 - Data Owner
 - Data Stewardship
 - Data Stakeholders
 - Comités y Funcionamiento
 - Herramientas de Mercado



Gobierno del Dato

- Unidad 4: Data Quality
 - Data profiling
 - Data curation
 - Reglas de calidad
 - Herramientas de mercado
- Unidad 5: Data lineage and Traceability
 - Glosario de términos & diccionario
 - Capa semántica
 - Linaje y trazabilidad
 - Flujos de trabajo
 - Herramientas de mercado
 - Data Stakeholders
 - Comités y Funcionamiento
 - Herramientas de Mercado



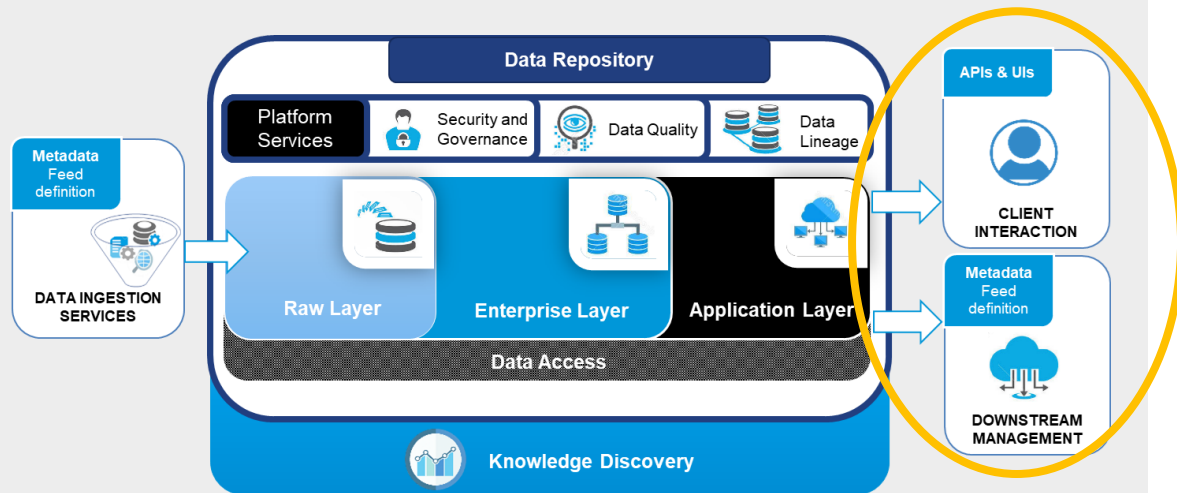
Gobierno del Dato

- Unidad 6: Éticas del dato y Seguridad
 - Ecosistema legal (GDPR)
 - Data masking y cifrado de datos
 - Datos sensibles
- Unidad 7: MDM y Reference Data
 - Metadatos
 - Visión 360
 - Reglas de supervivencia
 - Reglas de consolidación
- Unidad 8: Nuevos escenarios de datos
 - Datos como servicio (Daas)
 - Catálogos de datos
 - Casos Reales



Use the data

- Build reports
 - Help on decision support
- Visual dashboards
 - Control and monitor operations
- Build Machine Learning Model
 - Offer new products
 - Improve processes
 - Optimize production

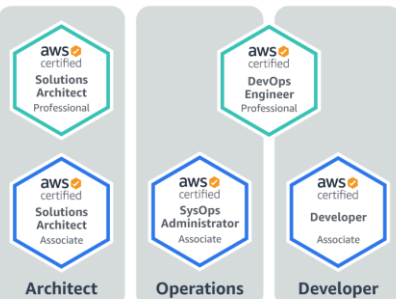


The Cloud

Available AWS Certifications

Professional

Two years of comprehensive experience designing, operating, and troubleshooting solutions using the AWS Cloud



Associate

One year of experience solving problems and implementing solutions using the AWS Cloud



Foundational

Six months of fundamental AWS Cloud and industry knowledge



aws certified
Updated May 2019

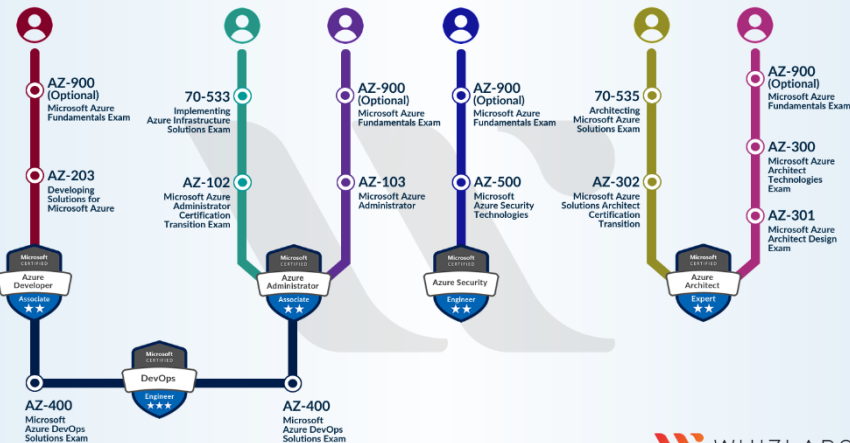
Specialty

Technical AWS Cloud experience in the Specialty of
exam guide

aws certified
Advanced Network Speciality

aws certified
Machine Learning Speciality

New Role-based Microsoft Azure Certification Path



WHIZLABS

Tecnologías:

