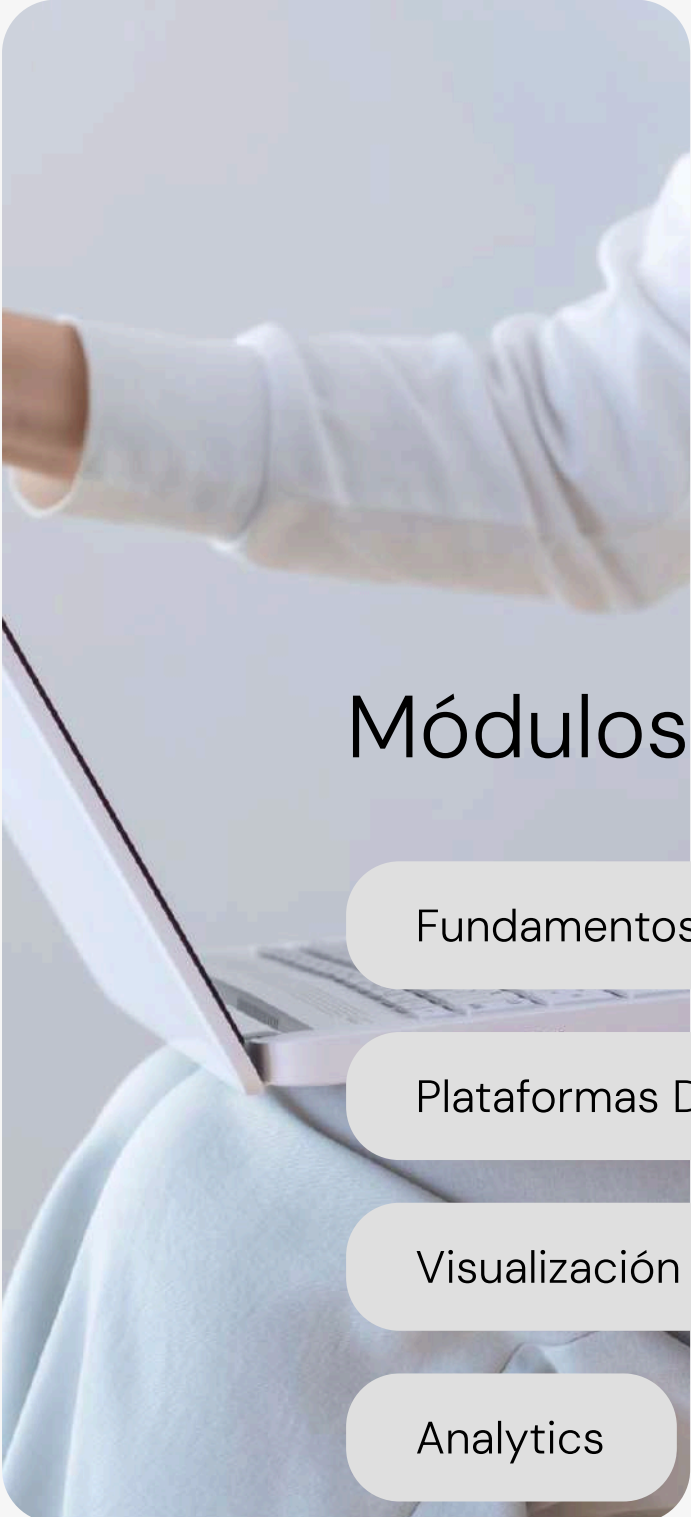


Catálogo formación oficial DATA





Módulos catálogo

Fundamentos Data & Cultura del Dato

Plataformas Data

Visualización & Cuadros de Mando

Analytics

IA & Agentes

QA del Dato & Calidad

MOD 0

Fundamentos Data & Cultura del Dato

Público objetivo

Todos los perfiles de la organización: técnicos y negocio. Módulo abierto y accesible sin conocimientos previos.

Contenidos

- Ecosistema de datos y arquitecturas modernas
- Tipos de datos, fuentes y flujos ETL/ELT
- Gobernanza, roles y gobierno del dato
- Cultura del dato: de los silos a la colaboración
- Casos de uso sectoriales: retail, banca, salud, industria

20 horas

Teórico-práctico · Workshops colaborativos

Sesiones en vivo + asíncronas

Problemas en organizaciones reales

/01

Silos de datos entre áreas sin ninguna visión global ni estrategia

/02

Resistencia cultural al cambio y a basar decisiones en datos

/03

KPIs no definidos o con resultados contradictorios entre departamentos

/04

"Los datos son problema de IT", no una responsabilidad de negocio

Evaluación y seguimiento

Quiz conceptual final · Análisis de caso real · Autoevaluación de madurez data organizacional

Detalle de contenidos



Ecosistema de datos y arquitecturas modernas

- Data Lake, Data Warehouse, Lakehouse y Data Mesh: cuándo y por qué usar cada arquitectura
- Procesamiento batch vs streaming según el caso de uso real
- Tendencias actuales: datos descentralizados, data fabric y el modelo de dominio de datos



Tipos de datos, fuentes y flujos ETL/ELT

- Datos estructurados, semiestructurados y no estructurados: formatos y diferencias prácticas
- Fuentes habituales: BBDD relacionales, APIs REST, IoT, logs de sistema y redes sociales
- ETL vs ELT: diferencias operativas, herramientas del mercado y cuándo aplica cada modelo



Gobernanza, roles y gobierno del dato

- Catálogos de datos, linaje y gestión de metadatos técnicos y de negocio
- Roles clave: Data Owner, Data Steward y Data Custodian — responsabilidades y diferencias
- Marcos de referencia: DAMA-DMBOK y un modelo de gobierno adaptable a la organización



Cultura del dato: de los silos a la colaboración

- Diagnóstico de madurez data: del nivel 1 reactivo al nivel 5 data-driven predictivo
- Modelos organizativos: centralizado, federado y descentralizado — ventajas y riesgos
- Change management: cómo impulsar la adopción del dato sin resistencia organizativa



Casos de uso sectoriales por industria

- Retail: personalización de oferta, previsión de demanda y optimización de inventario
- Banca y seguros: detección de fraude en tiempo real, scoring y customer lifetime value
- Salud e industria: IoT industrial, predicción de fallos en maquinaria y calidad de producto

MOD 1

Plataformas Data

Público objetivo

Data Engineers, Arquitectos de Datos, DevOps con orientación data. Perfil técnico con experiencia en cloud.

Contenidos

- Google Cloud Platform: visión general y servicios data
- BigQuery: arquitectura, modelado, optimización y gestión de costes
- Cloud Storage, Dataflow y Pub/Sub en arquitecturas reales
- Arquitectura Lakehouse en GCP: patrones y mejores prácticas
- Seguridad, IAM y gobierno en entornos cloud data

30 horas

Hands-on labs en GCP

Sandbox real de cliente

Problemas en organizaciones reales

/01

Plataformas sobredimensionadas y caras para el caso de uso real del cliente

/02

Vendor lock-in no evaluado correctamente en la decisión inicial de plataforma

/03

Costes cloud descontrolados por falta de optimización y gobernanza financiera

/04

Migraciones mal planificadas que generan deuda técnica difícil de revertir

Evaluación y seguimiento

Proyecto de arquitectura data en GCP · Preparación certificación GCP Professional Data Engineer

MOD 1

Detalle de contenidos



Google cloud platform: servicios data

- Arquitectura de servicios GCP: compute, storage, analytics y ML — visión completa
- Comparativa con AWS y Azure: posicionamiento, diferenciales y decisión de plataforma
- Modelo de facturación GCP: on-demand, compromisos y estrategias de reducción de costes



Bigquery: diseño, optimización y costes

- Arquitectura columnar y separación compute/storage: implicaciones en rendimiento y precio
- Particionado, clustering y reescritura de queries para reducir bytes procesados
- Reservas, slots y FinOps en BigQuery: control de gasto en organizaciones grandes



Cloud storage, dataflow y PUB/SUB

- Cloud Storage: tipos de bucket, políticas de ciclo de vida y replicación multi-región
- Dataflow: pipelines batch y streaming con Apache Beam — patrones y operadores clave
- Pub/Sub: mensajería asíncrona desacoplada y patrones event-driven en tiempo real



Arquitectura lakehouse en GCP

- Medallion architecture sobre GCS: capas bronze, silver y gold — flujo y responsabilidades
- Formatos abiertos: Parquet, Avro y Delta Lake sobre Cloud Storage con BigQuery externo
- Patrones de ingesta, transformación con Dataform y consumo desde BI o ML



Seguridad, IAM y gobierno en Cloud

- IAM en GCP: roles predefinidos vs personalizados, principio de mínimo privilegio
- VPC Service Controls y Private Google Access para entornos regulados
- Cloud Audit Logs, Data Catalog y cumplimiento normativo: GDPR, ISO 27001

MOD 2

Visualización & Cuadros de Mando

Público objetivo

Analistas de Negocio, BI Developers, Product Managers con necesidades de reporting y toma de decisiones.

Contenidos

- Panorama de herramientas: Power BI, Tableau, Looker, QlikSense — fortalezas y debilidades
- Gobernanza, publicación y control de accesos en Power BI Service
- Power BI como herramienta central: modelo de datos y DAX avanzado
- Conectores, fuentes de datos y actualización automática
- Diseño de dashboards orientado a decisiones de negocio

25 horas

Talleres prácticos · Proyecto real de dashboard

Revisión de casos de cliente

Problemas en organizaciones reales

/01

Dashboards técnicamente correctos que nadie del negocio consulta ni entiende

/02

Métricas duplicadas con definiciones y resultados distintos según el equipo

/03

Inconsistencias entre informes: la misma KPI con diferente valor dependiendo de quién la mira

/04

Exceso de indicadores sin jerarquía, sin narrativa y sin orientación a decisión

Evaluación y seguimiento

Dashboard funcional sobre datos reales · Certificación Microsoft Power BI Data Analyst

MOD 2

Detalle de contenidos



Panorama de herramientas de visualización

- Power BI: ecosistema Microsoft, precio competitivo, bueno en datos corporativos
- Tableau: máxima flexibilidad visual y análisis exploratorio; precio elevado
- Looker (LookML) y QlikSense: gobernanza semántica vs motor asociativo — pros y contras
- Otras...



Power BI como herramienta central del curso

- Power BI Desktop, Service y Mobile: el ecosistema completo y sus diferencias
- Modos de conexión: Import, DirectQuery y Composite — cuándo usar cada uno
- Integración con Azure, Microsoft Teams y Microsoft 365 en entornos corporativos



Modelo de datos y DAX avanzado

- Modelado en estrella y copo de nieve: buenas prácticas y antipatrones habituales
- DAX avanzado: CALCULATE, contextos de fila y filtro, time intelligence functions
- Optimización con DAX Studio: identificar medidas lentas y reescribir para rendimiento



Diseño de dashboards orientado a decisiones

- Principios de diseño visual: jerarquía de información, contraste y alineación
- Elección del gráfico correcto: barras, líneas, scatter, treemap, KPI cards — cuándo y por qué
- UX para dashboards ejecutivos: navegación, filtros de contexto y drillthrough



Gobernanza y publicación en power BI service

- Workspaces, apps y control de accesos: roles de visualizador, miembro y administrador
- Certified datasets, Row-Level Security (RLS) y protección de datos sensibles
- Gateways, actualización automática y monitorización del uso de informes

MOD 3

Analytics

Público objetivo

Data Analysts, Data Scientists junior/mid, Ingenieros interesados en análisis avanzado y modelado.

Contenidos

- SQL avanzado: window functions, CTEs, optimización de consultas
- Python para análisis de datos: pandas, numpy, matplotlib, seaborn
- Estadística aplicada: distribuciones, tests de hipótesis, correlaciones
- Modelado predictivo y machine learning aplicado a casos reales
- Storytelling con datos: comunicar hallazgos a negocio

30 horas

· Labs de código sobre datos reales · Proyectos de análisis

Análisis de casos de negocio

Problemas en organizaciones reales

/01

• Modelos técnicamente correctos que no tienen ningún impacto en decisiones de negocio

/03

Precisión técnica excesiva que nadie del negocio puede interpretar ni aprovechar

/02

Análisis ad-hoc sin gobernanza, sin reproducibilidad y sin documentación

/04

Datos mal preparados que invalidan completamente el análisis realizado

Evaluación y seguimiento

Caso de negocio analítico end-to-end · Defensa de conclusiones ante panel de evaluación

MOD 3

Detalle de contenidos

**SQL avanzado y optimización de consultas SQL**

- Window functions: ROW_NUMBER, RANK, LAG/LEAD y agregados particionados
- CTEs (WITH), consultas recursivas y subconsultas para lógica compleja
- Optimización: EXPLAIN/EXPLAIN ANALYZE, índices y reescritura de queries lentas

**PYTHON para análisis de datos**

- pandas: merge, groupby, pivot_table, resample y reshaping de DataFrames
- numpy: operaciones vectoriales, álgebra lineal y generación de datos sintéticos
- Visualización analítica: matplotlib, seaborn y plotly para exploración de datos

**Estadística aplicada al análisis**

- Distribuciones de probabilidad y su aplicación práctica en datos empresariales
- Tests de hipótesis: t-test, chi-cuadrado, ANOVA — interpretación y toma de decisiones
- Correlación, regresión lineal y validación de supuestos del modelo estadístico

**Modelado predictivo y machine learning**

- Árboles de decisión, Random Forest y Gradient Boosting: cuándo usar cada familia
- Validación cruzada, métricas (AUC, RMSE, F1) y gestión del overfitting
- Interpretabilidad con SHAP values y feature importance para comunicar a negocio

**Storytelling con datos**

- Estructura narrativa: contexto → hallazgo → impacto → recomendación accionable
- Elección del gráfico según el mensaje: comparación, tendencia, distribución, relación
- Técnicas para comunicar resultados a audiencias no técnicas sin perder rigor

MOD 4

IA & Agentes

Público objetivo

Desarrolladores senior, Data Scientists, Arquitectos con base en ML. Perfil técnico avanzado.

Contenidos

- Fundamentos de LLMs: arquitecturas Transformer y modelos generativos
- Prompt engineering avanzado y fine-tuning con datos propios
- Retrieval Augmented Generation (RAG): diseño e implementación
- Construcción de agentes IA: LangChain, CrewAI, AutoGen
- IA responsable: sesgos, alucinaciones, privacidad y cumplimiento

35 horas

Labs intensivos · Proyecto de agente en producción

Evaluación de modelos en escenarios reales

Problemas en organizaciones reales

/01

IA implementada sin caso de uso claro, sin ROI definido ni métricas de éxito

/02

Alucinaciones de modelos en producción no detectadas ni gestionadas adecuadamente

/03

Deuda técnica generada por adopción de IA sin arquitectura ni gobierno definido

/04

Resistencia interna por falta de explicabilidad: "¿por qué dice eso el modelo?"

Evaluación y seguimiento

Agente IA funcional sobre caso de negocio real · Análisis de riesgos y mitigaciones documentado

MOD 4

Detalle de contenidos



Fundamentos de LLMS y modelos generativos

- Arquitectura Transformer: mecanismo de atención, encoders y decoders en detalle
- Modelos open-source vs propietarios: GPT-4o, Claude, Llama 3, Mistral — comparativa
- Tokenización, embeddings y el espacio vectorial semántico de los modelos



Prompt engineering y fine-tuning

- Técnicas: zero-shot, few-shot, chain-of-thought, ReAct y structured outputs
- Function calling y herramientas externas: conectar LLMs con APIs y datos propios
- Fine-tuning con LoRA/QLoRA: cuándo merece la pena frente a RAG o prompting



Retrieval augmented generation (RAG)

- Pipeline RAG: chunking, embedding, vector store y búsqueda semántica con ANN
- Técnicas avanzadas: reranking, HyDE, query expansion y multi-vector retrieval
- Evaluación de RAG: faithfulness, relevance, context recall y answer correctness



Diseño y construcción de agentes IA

- Patrones de agentes: ReAct, Plan-and-Execute y arquitecturas Multi-agent
- Frameworks comparados: LangChain, CrewAI y AutoGen — casos de uso de cada uno
- Memory, tools y orquestación: cómo construir agentes autónomos en producción



IA responsable: riesgos y mitigación

- Tipos de alucinaciones y técnicas de detección: groundedness, self-consistency
- Sesgos en datos de entrenamiento y su impacto en outputs y decisiones automatizadas
- Privacidad, GDPR y cumplimiento normativo al desplegar LLMs con datos corporativos

MOD 5

QA del Dato & Calidad

Público objetivo

Data Engineers, QA Engineers, Responsables de Gobierno del Dato y calidad en pipelines.

Contenidos

- Dimensiones de calidad: completitud, exactitud, consistencia, unicidad
- Data lineage y trazabilidad end-to-end del dato
- Google DataPlex como herramienta de laboratorio y validación
- Frameworks y estándares: DAMA-DMBOK, ISO 8000
- Testing de pipelines: validaciones automáticas y contratos de datos

20 horas

Laboratorio DataPlex · Tests automatizados

Auditoría sobre pipeline real de cliente

Problemas en organizaciones reales

/01

Datos erróneos llegando a producción sin ningún mecanismo de detección activo

/02

Pipelines sin validaciones que generan errores en cascada a lo largo de semanas

/03

Métricas de negocio incorrectas derivadas de datos de calidad deficiente aguas arriba

/04

Ausencia de linaje: imposible rastrear el origen de un dato incorrecto en producción

Evaluación y seguimiento

Suite de tests de calidad sobre pipeline real · Informe de auditoría del dato en formato profesional

MOD 5

Detalle de contenidos



Dimensiones de calidad del dato

- Complejidad: detección y tratamiento sistemático de nulos y valores faltantes
- Exactitud y consistencia: validaciones cruzadas entre sistemas y fuentes de verdad
- Unicidad y vigencia: deduplicación de registros y frescura del dato en tiempo real



Google Dataplex como herramienta de laboratorio

- Arquitectura DataPlex: lakes, zones y assets — organización lógica del dato
- Data quality rules: creación de reglas, ejecución programada y monitorización
- Integración nativa con BigQuery y Cloud Storage: validaciones en tiempo real



Testing de pipelines y validaciones automatizadas

- Contratos de datos: definición de schema, tipos, rangos y restricciones de negocio
- dbt tests: schema tests, singular tests, custom tests y data freshness checks
- Great Expectations: expectation suites, checkpoints y resultados en CI/CD pipelines



Data Lineage y trazabilidad end-to-end

- Linaje técnico vs linaje de negocio: diferencias y complementariedad
- Herramientas: OpenLineage, Marquez y DataPlex Lineage — comparativa práctica
- Trazabilidad completa: desde la fuente de origen hasta el dashboard de negocio



Frameworks y estándares de calidad

- DAMA-DMBOK: las 11 áreas de conocimiento en data management y su aplicación
- ISO 8000: estándar internacional de calidad del dato — requisitos y certificación
- Data quality scorecards y reporting: cómo medir y comunicar la calidad en la org.

Formación bonificable a través de FUNDAE

Las formaciones de tecdata pueden bonificarse con el crédito de formación de FUNDAE (Fundación Estatal para la Formación en el Empleo). La bonificación la tramita la propia empresa, directamente o a través de su gestoría o entidad organizadora, y tecdata facilita toda la documentación que necesita.

Cómo se bonifica

/01 Consultar el crédito

En la aplicación de FUNDAE, con certificado digital, o a través de la gestoría.

/02 Informar a la RLT

Si hay representación legal de los trabajadores, se le informa antes de empezar. Dispone de 15 días hábiles para emitir su informe.

/03 Comunicar el inicio

Se da de alta la acción formativa y se comunica el inicio de cada grupo con al menos 2 días naturales de antelación.

/04 Comunicar la finalización

Al terminar, se notifica la finalización con los participantes y los costes, y se entregan los diplomas.

/05 Aplicar la bonificación

Se descuenta en los seguros sociales, como muy tarde en el boletín de cotización de diciembre del mismo ejercicio.

Qué facilita tecdata

01 Programa y contenidos detallados de cada acción formativa.

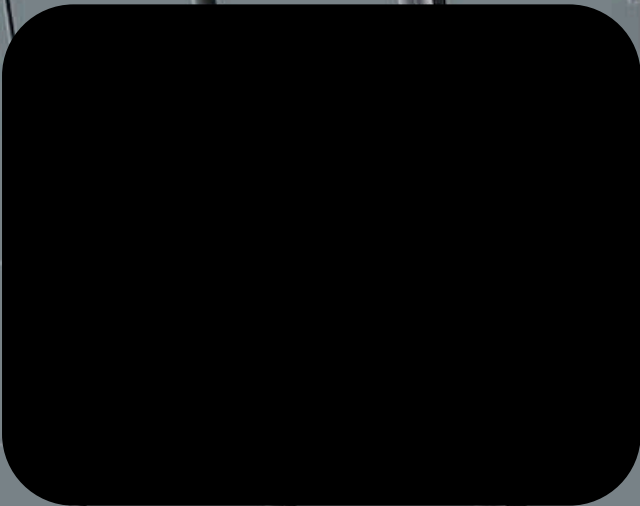
02 Control de asistencia de los participantes.

03 Diplomas o certificados para los participantes.

04 Fechas, horario, lugar y formadores para la comunicación de inicio.

05 Factura desglosada con el detalle de la formación.

06 Apoyo para resolver dudas durante la tramitación.



www.tecdata.es

C/ Julián Camarillo 4,
Madrid – España