

DATA PIPELINE ENGINEERING

Service Definition
G-Cloud 15

REPLY

38 Grosvenor Gardens - London SW1W 0EB - United Kingdom - tel +44 (0) 207 730 6000 - fax +44 (0) 207 259 8600
info@reply.com

www.reply.com



TABLE of CONTENTS

SERVICE DEFINITION	1
TABLE OF CONTENTS	2
SERVICE DESCRIPTION	3
ONBOARDING	7
OFFBOARDING	8
ACCESS TO DATA	8
SECURITY	9
PERSONNEL SECURITY	9
INFORMATION SECURITY	9
PHYSICAL SECURITY	9
TRAINING & KNOWLEDGE TRANSFER	10
ORDERING AND INVOICING PROCESS	11
CUSTOMER RESPONSIBILITIES	12
ABOUT REPLY	14



SERVICE DESCRIPTION

Fragile spreadsheet workflows, manual data extracts, and aging ETL tools create a brittle foundation that fails when organisations need data most. Reply Data Pipeline Engineering service designed for G-Cloud 15, replaces these unreliable processes with production-grade, automated data flows built using modern orchestration frameworks, transformation tools, and quality controls that deliver trusted data consistently and at scale. We engineer pipelines using modern and stable platforms like Dagster, Apache Airflow, dbt, and Great Expectations deployed across AWS, Azure, or GCP to ingest data from any source (e.g. databases, APIs, SaaS applications, files, streaming platforms) transform it with testable, logic under version control, and deliver analytics-ready datasets that power self-service insights, AI/ML models, and real time operational intelligence. Unlike legacy approaches where pipelines are black boxes maintained through dangerous production queries, we treat data infrastructure as software products with comprehensive testing, monitoring, version control, and CI/CD automation that enables safe evolution and eliminates firefighting. Whether you are drowning in manual processes, suffering repeated data quality incidents, or unable to support modern analytics and AI initiatives due to infrastructure limitations, we deliver the reliable, scalable pipeline foundations that transform data from a liability into a strategic asset.

FEATURES

Our pipeline engineering combines modern orchestration and transformation frameworks with rigorous quality controls and DevOps practices, creating data infrastructure that operates treating data as products, like in software development.

- **Automated ingestion from databases, APIs, SaaS applications, files, and streaming sources (Boomi, Informatica, and ad-hoc ingestion patterns):** We build flexible ingestion frameworks that extract data from any source system (e.g. polling databases with CDC, calling REST APIs with OAuth and pagination, retrieving files from SFTP/S3/SharePoint, consuming streaming events from Kafka, etc) using both commercial tools like Boomi and open-source patterns. This source-agnostic approach means no data remains trapped in legacy systems or isolated SaaS applications.
- **Modern orchestration using Dagster, Apache Airflow or Prefect, with dependency management:** We implement sophisticated workflow orchestration that manages pipeline dependencies, schedules, retries, and alerting using platforms built for data workflows rather than general purpose schedulers. These orchestrators provide DAG dependency management ensuring downstream processes only run when upstream data is ready, automatic retries with exponential backoff, and rich monitoring showing exactly where failures occurred.
- **SQL (e.g. DBT) or Python based transformation ensuring testable, documented, version-controlled analytics logic. Treating Data as a Product:** We implement transformation logic using dbt for SQL-based analytics engineering or Python for complex algorithmic processing, both following software engineering best practices including automated testing, peer-reviewed version control, and comprehensive documentation. Building data as a product treats transformation code as maintainable software rather than undocumented queries, enabling safe evolution and knowledge transfer.
- **Multi-cloud deployment across AWS, Azure, GCP supporting any data platform preference:** We deploy pipeline infrastructure across any cloud provider using infrastructure-as-code patterns that make deployments repeatable and portable, avoiding vendor lock-in. Whether you are on AWS due to existing commitments, Azure, or GCP for specific capabilities, our implementations work consistently across platforms.



- **Target platform flexibility: Snowflake, BigQuery, Databricks, Redshift, Synapse, Fabric, EDB:** We write pipelines that load data into any modern data platform, using native features where beneficial but maintaining portability through standard SQL and APIs. This flexibility means your pipeline investments survive platform migrations and enable more than one platform architectures where appropriate.
- **Embedded data quality using Great Expectations, Soda, dbt tests preventing data quality issues:** We integrate quality validation directly into pipeline execution using frameworks that automatically check completeness, accuracy, referential integrity, and business rules at each stage. Pipelines fail fast when quality thresholds are breached, preventing bad data from propagating downstream rather than discovering issues days later in reports.
- **Infrastructure-as-code with Terraform, CloudFormation enabling auditable, repeatable deployment across environments:** We define all pipeline infrastructure (e.g. compute resources, storage, networking, permissions) as version-controlled code that can be peer-reviewed, tested, and deployed consistently across development, staging, and production environments. This IaC approach eliminates manual configuration drift and provides complete audit trails of infrastructure changes.
- **Real-time and batch processing supporting streaming (Kafka, Kinesis) and scheduled workloads:** We build pipelines that handle both batch processing for periodic analytics refreshes and stream processing for real-time operational intelligence, using appropriate patterns for each. This versatility enables everything from overnight warehouse loads to sub-second event processing for operational dashboards.
- **Complete observability with lineage tracking, monitoring dashboards, SLA alerting, and incident response:** We instrument pipelines with comprehensive monitoring capturing execution times, data volumes, quality metrics, and resource utilisation, with automated alerting when SLAs are breached. Lineage tracking shows exactly how data flows from sources through transformations to consumption, enabling rapid impact analysis during incidents.
- **Production-ready CI/CD pipelines with automated testing, deployment, and rollback capabilities built-in:** We implement continuous integration and deployment workflows that automatically test pipeline changes, deploy them through staged environments, and provide easy rollback if issues emerge. This automation reduces deployment risk and enables frequent, safe updates that would be impossibly risky with manual processes.

These capabilities combine to create data infrastructure that operates with the reliability users expect from critical production systems, eliminating the uncertainty and firefighting legacy approaches currently need.

BENEFITS

Modern pipeline engineering delivers immediate operational improvements while establishing the foundation for advanced capabilities that drive transformational business outcomes.

- **Deliver trusted, analysis-ready data to business users by enabling self-service:** Automated pipelines consistently produce clean, validated, documented datasets that business users can confidently query without technical gatekeepers validating extracts or transforming data manually.



This self-service capability democratises data access and accelerates insight generation from weeks to minutes.

- **Accelerate AI/ML initiatives with reliable feature pipelines and model training datasets:** ML models require consistent, quality-assured feature pipelines that deliver training data on predictable schedules. Our engineering provides this foundation, reducing ML development time from months to weeks. Automated feature generation and validation eliminate the manual preparation bottlenecks that stall AI initiatives.
- **Power real-time dashboards and operational intelligence with fresh, validated data feeds:** Modern processing pipelines enable operational dashboards showing live case management status, service delivery metrics, or resource utilisation with sub-minute latency. This real-time visibility enables proactive intervention rather than reactive responses to problems discovered in outdated daily reports.
- **Eliminate manual data wrangling saving analysts 60-80% of time currently spent preparing data:** When pipelines deliver analysis-ready data, analysts spend time generating insights rather than writing single extraction queries, cleaning data in Excel, or reconciling conflicting sources. This efficiency multiplier transforms analytical capacity without hiring additional staff.
- **Reduce data incidents by 90% through automated validation catching issues before consumption:** Quality checks integrated into pipeline execution catch schema changes, completeness drops, and business rule violations before they reach reports that inform decisions. This prevention eliminates the damage control, apologies, and rework that follow incidents where bad data reaches stakeholders.
- **Support evidence-based policymaking with timely, accurate data replacing outdated monthly reports:** Automated pipelines refresh analytical datasets daily or hourly rather than monthly, providing policymakers with current information rather than stale snapshots. This timeliness enables responsive policy adjustments based on emerging trends.
- **Meet audit requirements with complete lineage, version control, and change tracking built-in:** Version controlled transformation logic, infrastructure-as-code, and automated lineage provide the audit trail regulators and internal auditors require, showing exactly what happened to data, when, and why. This documentation eliminates scrambles to reconstruct processing for compliance reviews.
- **Future-proof architecture using open-source standards adopted across government and commercial sectors:** Pipelines built on Apache Airflow, dbt, and open-source orchestration frameworks align with industry-standard patterns being adopted across progressive organisations, avoiding proprietary lock-in. This alignment ensures long-term viability, skills availability, and community support.
- **Free technical teams from firefighting to focus on delivering analytical value consistently:** Reliable automation with comprehensive monitoring and alerting eliminates the constant firefighting where engineers manually rerun failed jobs, investigate data quality issues, and patch broken processes. This operational stability frees engineers to build new capabilities rather than maintaining broken legacy systems.
- **Treating data as a software product, stopping data engineers from having to run dangerous production queries:** Development workflows with testing, staging environments,



and automated deployment eliminate the risky practice of engineers running individual queries directly against production systems. This software engineering discipline prevents outages, data corruption, and compliance violations that result from manual production changes.

These benefits compound over time. Reliable pipelines enable more sophisticated use cases, which justify further investment in quality and automation, creating a virtuous cycle of increasing data capability and organisational maturity.



ONBOARDING

We will hit the ground running from day one. Our structured onboarding process eliminates the confusion and slow starts that plague consulting engagements, ensuring we're delivering value within weeks, not months.

What We Do:

- **Intensive Kick-off Workshop:** Bring together your executive sponsor, technical leads, and key stakeholders to establish crystal-clear objectives, success metrics, and non-negotiable delivery milestones with complete alignment across all parties.
- **Immediate Communication Infrastructure:** Establish dedicated channels (Slack/Teams), weekly governance meetings with executive visibility, and escalation paths ensuring blockers get resolved in hours, not weeks.
- **Rapid Discovery:** Interview stakeholders, review existing systems, identify data sources, document constraints, and surface risks that could derail delivery before they become problems.
- **Concrete Delivery Plan:** Deliver a sequenced plan showing exactly what gets built when, what resources are required, what decisions need making, and what risks could impact timelines. No theoretical documents, just actionable roadmaps.
- **Quick Wins Identification:** Identify and deliver tangible outputs early that demonstrate immediate value and build organisational confidence in our approach and your investment.
- **Access Secured:** Obtain all necessary system access, credentials, environments, and documentation during the first week, because access delays are the single biggest cause of timeline delays.

Our onboarding doesn't end until we've delivered our first tangible output, proving our team is productive and your investment is already generating returns. You'll know within the first month whether we're the right partner. We move fast, communicate clearly, and deliver evidence of progress.



OFFBOARDING

We do not always believe in traditional "handover" periods where knowledge gets dumped on your team at the end. Instead, we work with you throughout the engagement, training your staff while the work is ongoing and creating experts along the way. Our approach ensures your team is not inheriting unfamiliar solutions but has been building and operating them alongside us from the start. However, if your plan is to have us "do it for you", we will generate documentation on the way, as part of any decision we make.

What We Do:

- **Continuous Knowledge Building (Throughout Engagement):** Your team works directly with our engineers from week one: pair programming, collaborative design, joint troubleshooting. So they understand not just what was built but why, developing expertise incrementally rather than through rushed final handover sessions.
- **Structured Transition Plan:** Formalise the knowledge transfer that's been happening continuously, identifying any remaining gaps in credentials, documentation, or procedures requiring explicit handover with clear ownership assignments.
- **Validated Competency Demonstrations:** Ensure your team independently executes key operational tasks (e.g. deploying changes, responding to incidents, onboarding new data sources, etc.) under our observation and coaching, proving they are genuinely ready for independent operations.
- **Production-Ready Documentation:** Deliver architectural decision records explaining design choices, operational runbooks covering daily operations, incident response procedures with real examples, and video walkthroughs of complex tasks for ongoing reference and future team onboarding.
- **Active Hypercare Period:** Proactively monitor for issues, conduct weekly check-ins reviewing operational metrics, and provide rapid response to problems during the critical transition to fully independent operations.
- **90-Day Post-Delivery Review:** Schedule a structured review assessing what is working, what needs adjustment, and what additional support might prevent future issues, ensuring long-term sustainability and addressing any emerging challenges.

Your success operating our deliverables determines whether we get repeat business and referrals. We have zero interest in building solutions that collapse after we leave. By training your staff throughout delivery rather than at the end, we ensure you're genuinely independent and confident, with deep expertise to maintain and evolve what we've built for years to come.

Access to Data

Following formal offboarding, all data, code, documentation, and intellectual property developed during the engagement transfers entirely to your ownership with appropriate licensing. We retain no copies of your production data, credentials, or proprietary information beyond what's legally required for contract compliance and professional indemnity purposes. Any sample data or anonymised examples used in training materials are sanitised to remove sensitive information. If you require our support post-engagement through separate maintenance contracts or ad-hoc consulting, we will request fresh access grants rather than assuming continued access rights. Your data sovereignty and security remain paramount. Our engagement ends cleanly with your full control over all assets and access.



SECURITY

Personnel Security

As a Specialist Cloud Service, the capability being offered is not limited to specific Impact levels (as it is not infrastructure, software or a platform) and can be used, subject to personal Security Clearance levels. Reply consultants are mostly Security Cleared (SC), some have higher-level Developed Vetting (DV) clearances. We also have a pool of NPPV3 Consultants for Policing work. The majority of our work for both public and private sector clients is at IL2 but we work in the Official, Secret and Top-Secret domains. Our hard and soft information security processes have been designed and approved by independent CESG CLAS accredited consultants.

Information Security

We are ISO27001, Cyber Essentials certified and our Quality Assurance processes are based upon and compliant with our ISO 9001 accreditation. Our Information Security processes are directly guided by our ISO27001 accreditation. We shall adhere to local information and other security policies and will apply local Security Operating Procedures (SyOPs) as may exist. If such do not exist we shall apply our own SyOPs.

Physical Security

The Main Reply Office (London, UK) is a Police Accredited Secure Facility (PASF) with access controls for each point of entry. All Laptops/Phones must be stored securely overnight. When working on a client-site our consultants adhere to client security policies.



TRAINING & KNOWLEDGE TRANSFER

We transfer knowledge continuously throughout delivery because cramming everything into final handover sessions doesn't work and we refuse to set you up for failure. From week one, your team members work directly alongside our engineers (e.g. pair programming on pipeline code, collaborating on dashboard designs, participating in architecture decisions) so they understand not just what was built but why specific approaches were chosen and how to extend them. We run weekly "lunch and learn" sessions covering specific technical topics, tools, and patterns we're using, building your team's capability incrementally rather than overwhelming them at the end. Every two weeks, we conduct show-and-tell sessions where we demonstrate progress and your team asks questions, tries features, and provides feedback. We create role-specific training:

- Technical deep dives for your data engineers covering orchestration platforms, transformation logic, and infrastructure-as-code
- Operational training for support teams covering monitoring dashboards, alert response, and troubleshooting procedures
- User training for business analysts covering BI tools, report building, and data interpretation
- Executive briefings for leadership covering governance frameworks, strategic capabilities, and ROI realisation

All training includes hands-on exercises using your actual data and real use cases, not generic examples that do not translate to your context. We provide comprehensive documentation: architecture diagrams showing system components and interactions, code repositories with detailed README files and inline comments, video recordings of training sessions and complex procedures, searchable knowledge base answering common questions. Most critically, we validate knowledge transfer has worked. Your team must successfully complete practical assessments demonstrating they can maintain, troubleshoot, and extend solutions before we consider training complete. We measure success by your team's independence, not hours of training delivered.



ORDERING AND INVOICING PROCESS

To begin the process, please contact glue.frameworks@reply.com with details of your organisation, role and high-level requirements.

Our Framework Manager will connect you to our experts and from here we will organise an introductory discussion to go into more detail and shape a proposed engagement that suits your needs.

We invoice for both fixed price and time and materials engagements one month in arrears. For fixed price engagements, this takes place upon completion of the deliverable(s), or in stages, if the size of the engagement is greater than £40,000 excluding VAT.

Payment terms are 30 days net of receipt of invoice.



CUSTOMER RESPONSIBILITIES

We deliver exceptional results, but we cannot succeed alone. While we bring expertise, tools, and delivery capability, certain contributions from your organisation are essential to ensure we are building solutions that meet your needs and can be sustained after our engagement. However, we recognise that organisations often engage us precisely because some of these capabilities are missing and we will be there filling those gaps.

What we need from you:

- **Access and Environment Provisioning:** We need access to relevant systems, data sources, and environments to understand your landscape and build appropriate solutions. We understand access can take time due to security procedures and approvals. We will work with you to navigate these processes and can begin with limited access while fuller permissions are secured. If production-like environments don't exist, we can help you establish them as part of our engagement.
- **Stakeholder Identification and Availability:** Help us identify the right people to speak with: subject matter experts who understand your data, processes, and business needs. We know everyone is busy, so we will be flexible with scheduling and efficient with their time. If key knowledge exists only in people's heads rather than documentation, that's fine. We will capture and document it as part of our work.
- **Decision-Making and Governance:** We need someone with authority to make decisions on direction, scope, and trade-offs when choices arise. This doesn't require daily involvement. Typically, weekly or fortnightly governance meetings are sufficient. If decision-making processes are unclear or complex, we can help you establish appropriate governance structures as part of the engagement. Also, if authority figures are missing (C-levels, heads of department, technical leads) we can provide experienced leadership to fill those gaps in order to drive decision-making and maintain momentum until permanent appointments are made or internal capability is established.
- **Representative Data and Use Cases:** Access to realistic data and actual business scenarios helps us build solutions that work for your real-world needs, not theoretical examples. We understand data may be messy, incomplete, or sensitive. We will work within those constraints, help you improve quality where needed, and ensure appropriate security controls. If you are unsure what data to provide, we'll guide you through that discovery.
- **Engagement in Knowledge Transfer:** Your team's participation in training and knowledge building activities is crucial for sustainability after we leave. We understand competing priorities and will design training that's practical, relevant, and respectful of time constraints. If you lack technical staff to receive handover, we can discuss options including extended support arrangements, recruiting assistance, or phased capability building.
- **Honest Conversation About Constraints:** Tell us early about limitations, challenges, or concerns: technical constraints, budget pressures, political sensitivities, data quality issues everyone knows about. Early honesty allows us to plan appropriately and propose realistic solutions. Late surprises create problems, but we would rather know about difficulties upfront than discover them mid-delivery.

We're Here to Help: We recognise that many organisations engage consultants precisely because they lack some of these capabilities internally. Limited technical expertise, no spare capacity for project work,



unclear governance, poor documentation. That is completely normal and exactly why we exist. Where you have gaps, we will work with you to fill them, whether that's establishing governance frameworks, creating documentation that does not exist, building technical capability in your team, or providing extended support while you recruit permanent staff. Our goal is your success, and we will adapt our approach to your reality rather than expecting you to meet some idealised client profile. The key is open communication about what you can provide, what you need help with, and what constraints we are working within. Together, we will figure out how to make it work.



ABOUT REPLY

Reply is a company that specialises in Consulting, Systems Integration and Digital Services with a focus on the conception, design and implementation of solutions based on the new communication channels and digital media.

Reply partners with key industrial groups in defining and developing business models made possible by the new technological and communication paradigms such as Artificial Intelligence, Big Data, Cloud Computing, Digital Communication, the Internet of Things and Mobile and Social Networking. In so doing, it aims to optimise and integrate processes, applications and devices.

Reply's offer is aimed at fostering the success of its customers through the introduction of innovation along the whole economic digital chain. Given its knowledge of specific solutions and due to a consolidated experience, Reply addresses the main core issues of the various industrial sectors.

Through its network of specialist companies, Reply supports some of Europe's leading industrial groups in Telco & Media, Industry & Services, Banks & Insurance, and Public Administration to define and develop business models, suited to the new paradigms of Artificial Intelligence, Big Data, Cloud Computing, Digital Media and the Internet of Things.

We make innovation happen.

We started with a small team and a purpose: to help the digital revolution happen. Today, we are a team of more than 14,600 people in 16 different countries but we still have the same DNA, made up of agile, vertical task forces and an inner passion for innovation.

We are a decentralised network of specialised companies.

Among Replyers, you will find passionate geeks, visionary strategists, and creative minds, each with sharp skills in a specific business, who cooperate together and never stop learning from each other.

Make forward, act sustainably.

We know a sustainable future is possible and technology can be a strong asset to reach it. As leaders in digital transformation, we push for change and operate in full accordance with the highest ethical standards and with respect for the rights of future generations