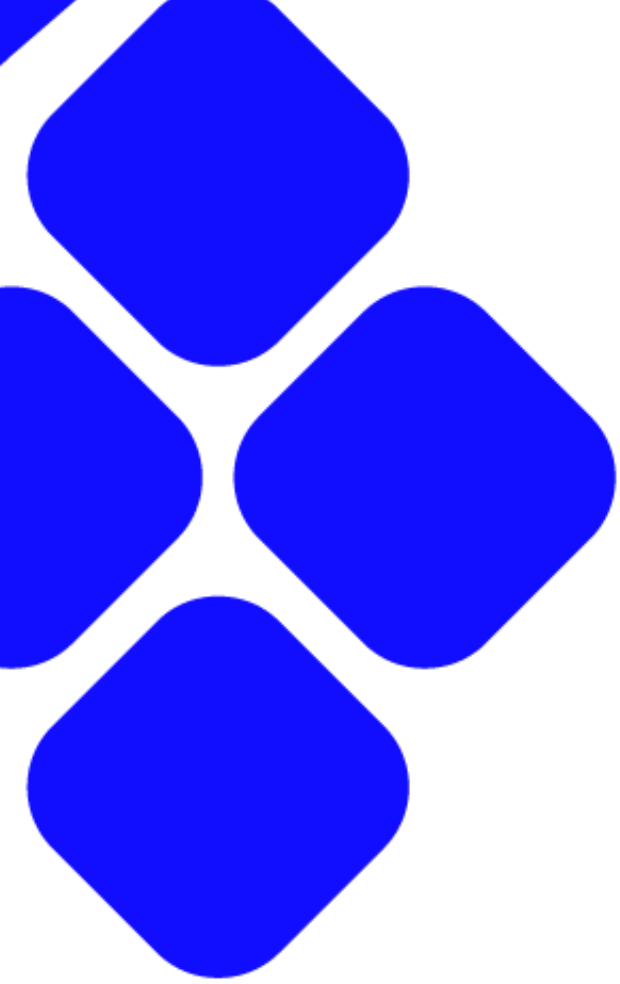


DRIVE UNIFIED DATA PLATFORM

Proposal

Disclaimer

While we have made every attempt to ensure that the information in this document has been obtained from reliable sources, Bluepineapple is not responsible for any errors or omissions, or for the results obtained from the use of this information. All information in this document is provided “as is”, with no guarantee of completeness, accuracy, timeliness or of the results obtained from the use of this information, and without warranty of any kind, express or implied, including but not limited to warranties of performance, merchantability and fitness for a particular purpose. In no event will Bluepineapple or its related corporations be liable for any decision or action taken in reliance on the information in this document for any consequential, special or similar damages, even if advised of the possibility of such damages.

All product names, trademarks and registered trademarks are property of their respective owners. All company, product and service names used in this website are for identification purposes only. Use of these names, trademarks and brands does not imply endorsement

Version	Date	Comments
1.0	14th Aug 2025	External Version for Drive.com.au

1. Executive Summary

This document outlines our proposed solution for the Drive Unified Data Platform. The approach covers phase-wise ingesting data from Marketplace, various content sources, and campaign marketing tools for actionable insights.

The architecture focuses strongly on

- Cost-effectiveness
- Ease of adoption
- Future-proofing technology for extensibility
- Strong governance.

This will enable Drive's business units and partners (e.g., vehicle brands and Dealers) to maximise value from consolidated data and insights, while reducing operational overheads and manual efforts in drawing reports and analysing metrics such as Cost per Lead and Revenue per Lead.

This approach supports rapid stakeholder onboarding, privacy management and governance, and a pathway for easy addition of new data sources— to keep pace with Drive's evolving technology ecosystem.

1.1. Business Objectives

The Unified Data Platform's primary objectives are around derivation of business value via Reporting and Analytics. The initial objectives are to consolidate and standardise data across:

- Campaign Ads (delivered via Google Ads)
- Content – Editorial content monetised via advertising
- Marketplace – Monetised via a Revenue Per Lead (RPL) model

The key goals are

- Faster access to clean data
- Reduced repetitive effort in data collation
- Cross-domain reporting
- Standardised and exploratory analytics
- Extensibility & Future proof capabilities
- Low TCO
- Governance and security requirements

1.2. Functional Objectives

The key functions that are targeted in this proposal are:

- Onboarding of core data sources (**Google Ads, Salesforce (both instances), Insider, WordPress, Redbook, Vehicle DBs, CMS Content**).
- Consolidated, Productised reporting (campaign performance, lead allocation, content/dealer engagement, ROI).
- Enablement of Ignition use cases: dealers gain dashboarded insights into engagement + leads, supporting business decisions.

- Provisions for leveraging AI tooling for use-cases such as narrative reporting or accurate lead scoring.

1.3. Efficiency Gains

Another key (architectural) objective is to ensure efficiency gains via this project by providing a **pathway for consolidation of current custom Drive data ingestion pipelines** into the UDP in later phases. This will streamline the various custom built ingestion infrastructure components into standard template structures.

While refactoring these existing custom pipelines (e.g. database loads) is not factored into the proposal, this solution expects to drive significant efficiency gains in infrastructure maintenance and ops for the engineering teams, while also allowing for consolidation of responsibilities of such maintenance to the common data product team.

Similarly, any system integration that requires structured/cleansed data from integrating data sources can consume such off the UDP (instead of legacy p2p pipelines)

1.4. Future Phases

With the initial data ingestion through to reporting pipelines in place, additional data sources can be onboarded for multi-dimensional analysis, as well as for surfacing these insights to different stakeholders.

This could potentially include the following for insights from newer data sources:

- AI driven (Agentic, multimodal, MCP) for simplifying query flows and enhancing insights
- Engagement metrics: (Additional Social feeds (including additional feeds from insider), newer tools for direct/email marketing etc.),
- Deeper insights, e.g., brand value reporting for partners
- Operational dashboards (Campaign planning, managing additional internal ad hoc/master data, spreadsheets).
- Consuming embedded external insights (Dealer/Partner portals) into the solution.
- Integration of more complex data (CDPs, Nine group warehouses and insights)

2. Drive x Bluepineapple

Drive.com.au has been steering a multi-year digital transformation to enhance dealer engagement, optimise lead management, and deliver an exceptional online experience for car buyers and sellers.

For over three years, Bluepineapple has been a trusted contributor to this journey — providing specialist engineering talent, agile delivery capability, and ongoing support to help bring Drive's strategic vision to life.

In close collaboration with Drive's leadership and product teams, Bluepineapple has:

- Supported the rewrite and optimisation of the Drive portal on AWS, enabling faster performance, streamlined architecture, and reduced infrastructure and licensing costs.
- Delivered front-end development aligned to Drive's evolving market requirements and user experience goals.
- Built and deployed a rule engine–driven Lead Management and CRM solution on Salesforce, empowering Drive to offer intelligent, accurate, and traceable lead routing for dealers.
- Provided ongoing managed support services, ensuring the portal remains reliable, scalable, and ready to adapt to new opportunities.

This strategic partnership has helped Drive.com.au maintain and strengthen its position as a market leader, with measurable gains in performance, operational efficiency, dealer engagement, and end-user satisfaction — while Bluepineapple continues to play a pivotal role in executing Drive's technology roadmap.

Looking ahead at Drive.com.au's next strategic initiative of a unified data platform — an intelligent, centralised hub that will seamlessly integrate data from its portal, lead management systems, and dealer engagement tools. This initiative is set to transform how Drive harnesses its data, enabling predictive analytics, real-time decision-making, and hyper-personalised experiences for both dealers and consumers.

The foundation laid through years of close collaboration, including shared architecture patterns, deep system knowledge, and a proven track record of delivery places Bluepineapple in the strategically preferred position to bring this vision to life. With an unparalleled understanding of Drive's digital ecosystem, Bluepineapple will apply its engineering expertise to ensure the platform becomes a defining competitive advantage, positioning Drive to lead the market well into the future.

3. Key Outcomes

The UDP will directly address long-standing operational pain points by replacing hours of manual data collation and fragmented analysis with ready-to-use, trusted, drill-down views for different user personas. These outcomes are designed to meet Drive's business goals of faster insights, cross-source reporting, and actionable decision support.

- Time Savings: Reduces manual data consolidation/analysis from hours to seconds.
- Accuracy: Removes spreadsheet error-risk with governed single source of truth.
- Actionability: All reports are interactive, filterable
- Extensibility: Framework supports adding new views, metrics, and sources without re-work.

Some examples are detailed below:

3.1. Outcome: Marketing Spend & Media Mix

Current ROI reporting for performance marketing is manually intensive, requiring campaign spend and revenue matching across channels in spreadsheets. No single view provides per-campaign profitability down to transaction level.

Outcome:

- Integrated Spend-to-Outcome Analysis: Links billing lead data, CRM revenue, and campaign investments within the warehouse.
- Media Mix & Channel Blend View – quickly assess spend allocation vs. revenue contribution across paid, organic, referral, and direct.
- Granular Profitability Metrics: Revenue per transaction, cost per transaction, and profit margin % at individual lead, campaign, and channel blend level.

Proposed Views/Reports:

- Marketing ROI Cockpit – combined metrics for campaign spend (filter by channel), billable leads, revenue, profit %, with channel blends.
- Channel Performance – channel wise (or blended) ROAS.
- Transaction-Level reports – for financial drill-down per lead/opportunity

3.2. Outcome: Unified Campaign & Engagement Intelligence

Currently, marketing and analytics teams spend hours aggregating data from Meta, Insider, Google Analytics, GAM, and campaign spreadsheets before they can answer even routine performance questions.

Outcome:

- Instant multi-source aggregation into consistent KPIs (impressions, clicks, CTR, spend, conversions, ROAS) updated hourly.
- Listings-Specific Churn Analysis:
    - View user churn rates segmented by listing type, category, or content publisher.
    - Drill down to the full pre-churn activity trail – pages visited, leads generated, content engaged with – across web and social touchpoints (Meta, Instagram, organic channels).

- Cross-segment by publisher/content nature to understand what type of content retains users longer.

Proposed Views/Reports:

- *Campaign Cross-Channel Performance* – unified Meta + Insider + GAM + GA4 metrics
- *User Churn Explorer* – pre-churn behavioural timelines, segmented by listing type and publisher
- *Publisher Engagement Report* – cross-platform engagement (social, portal, content) by publisher/content type

3.3. Outcome: Dealer & Data Publisher Resolution Dashboard

Sales/support staff servicing data publishers waste hours investigating “missing stock” queries (e.g., "I provided 500 listings, only 200 are live"), often needing technical assistance to trace ingestion or data validation failures.

Outcome:

- Data Publisher Health View – per-publisher metrics on:
  - Received vs. Accepted listings
  - Failures by cause (e.g., missing vehicle ID, schema mismatch, validation errors)
  - Historical failure rate trends to spot recurring data quality issues.
- Root-Cause Drill-Down – click into a failed listing to view exact rejection reason (e.g., invalid VIN, missing mandatory image).

Proposed Views/Reports:

- *Publisher Stock Pipeline Report* – end-to-end flow from file receipt/API pull → validation → live listing.
- *Top Rejection Reasons* – aggregated DQ errors by frequency and business impact.
- *Publisher SLA Tracking* – median time to resolve stock issues, grouped by source.

4. AI Use cases (Future Phases)

Drive uses multimodal content and data(structured data, text, images, videos etc.) with different semantic meanings and different use-case contexts. To navigate data across these different contexts and derive answers to complex questions, a MCP approach is key.

Note that these use-cases are not factored into the schedule and cost model in sections above.

Some example use-case scenarios are:

- Narrative reporting, (context aware querying): intelligent query co-pilot that interprets users questions and builds SQL from the UDP catalog
- Data Interrogation and Exploration: derive insights across raw data, report outcomes and other data feeds by combining these using multi-modal AI for richer narratives with visuals
- Anomaly exploration: trace anomalous behaviours and blips in trends across different datasets to determine root causes and impacts
- Reporting Chatbots: Agentic chatbots for self-service exploration and insights without needing to go to BI teams.

AI tooling can connect the data through its various pipeline stages from Ingestion through to BI. A more detailed set of potential use-cases is enumerated below:

#	Use Case	UDP Architecture Touchpoints	Role of Agentic AI	Role of MCP	Multimodal Input/Output
1	Guided, Context-Aware Query Building	Consumption Layer → BI / Ad-hoc SQL engine	Converts NLQs into optimised SQL on curated DW	Pulls schema, KPI definitions, lineage from Metadata Catalog	Text input → SQL query + charts/tables
2	Multimodal Data Exploration	BI Layer + Curated DW + Creative Store (external)	Combines performance data with uploaded image/text artefacts	Maps creative IDs from metadata → fact tables	Image/PDF/video input + charts, annotated visuals
3	Root-Cause & Impact Analysis	Curated DW + Metadata Catalog + Logs	Decides datasets to query, runs sequential analysis	Resolves dataset names, joins, lineage from zone mapping	Narrative text + time-series charts + linked creatives
4	Conversational Insight Agent in BI	BI Layer (Power BI/Looker Studio/Metabase)	Embedded Assistant answers follow-ups inside dashboards	Uses governed models and metrics via MCP API	Text/voice query → on-dashboard charts/tables
5	Automated Data Quality Fix Suggestions	Staging Zone + Metadata Catalog + DQ Engine	Diagnoses failed DQ checks, suggests	Finds impacted datasets and owner from metadata	Text summary + data profile tables +

			remediation scripts		missingness graphs
6	Cross-Source Narrative Generation	Curated DW + Ingestion metadata	Queries across multiple systems and builds slide-ready story	Pulls field descriptions, data lineage for commentary	Narrative text + embedded charts + asset snapshots
7	Real-Time KPI Alerts with Context	Streaming KPI monitor (optional future), Curated DW	Detects anomalies, auto-queries explainers	Pulls metric definitions and joins from MCP	Alert message + chart + screenshot of impacted creative
8	Zero-ETL Temporary Blending of Files	BI/Ad-hoc Query Engine + Local Renderer	Joins uploaded CSV/Excel file in-memory with warehouse data	Reads schemas of both datasets via MCP	File upload + instant blended chart/static table

Table 1. Sample AI Use cases

5. Reporting

From available information in the RFI and from our engagement with Drive stakeholders in ongoing projects, a presumptive data model and some sample reports are listed for a deeper (but non-exhaustive) analysis of the solution.

The initial phase is expected to deliver around 25 reports/dashboards covering a majority of the initial reporting needs.

These initial reports are expected to be derived from standardised report templates in line with the architectural strategy.

The data architecture supports large data volumes and growth, thereby ensuring support for analysis over 2-3 years or more for deriving trends or long-term metrics.

5.1. Illustrative Data Model

An illustrative data model is enumerated in the table below. This is derived from the standard dimensions available in Google Ads and Google Analytics data, with assumptions around campaign data models used in Drive.

With clear identification of Dimension and Fact tables, and with analysis of whether we ingest aggregated data or raw data from sources, we can plan the ingestion pipelines to load data to the reporting warehouse. These can then be used to derive meaningful reports.

Table	Type	Sample Cols
Campaign	Dimension	ID, Name, Channel, Start/End, Status
Date (Period)	Dimension	Date, Financial Period, Quarter, Year
Account	Dimension	ID, Name, Type, Region
Product	Dimension	ID, Make, Model, Trim etc.
Segment	Dimension	ID, Name, Definition
Lead	Fact	ID, CampaignID, AccountID, Date, Product, Segment, Chargeable, Revenue, CPA etc.
Spend	Fact	ID, CampaignID, Date Period, Amount
Engagement	Fact	ID, CampaignID, Impressions, Clicks, CTR, Conversions, Conversion Rate etc

Table 2. Sample Reporting Data Schema

5.2. Sample Reports

Based on the data sources and the RFI and subsequent clarifications, we can assume a few standard templates that are (potentially) representative of the reports desired by Business stakeholders. These are enumerated in the table below.

[Some example reports are elaborated upon in the Section: Key Outcomes. These include Marketing performance, Media Mix reports and churn reports.]

Template Name	Purpose	Key Measures	Dimensions
Campaign Performance Summary	Daily & monthly ad performance	Impressions, Clicks, CTR, CPA, Conversion Rate	Campaign, Date, Channel
Lead Funnel Report	Track lead conversion stages	Leads Received, Leads Qualified, Leads Chargeable	Campaign, Date, Lead Type
ROI Analysis Dashboard	Compare spend to revenue	ROI%, Revenue, Spend	Campaign, Date, Product
Audience Engagement	Content/marketplace engagement	Page Views, Avg Session Duration, Bounce Rate	Segment, Date, Geography
Inventory Availability	Stock levels and trends	Active Listings, New Listings, Sold Listings	Product, Dealer, Date
Revenue Per Lead	Overall Average Lead value tracking	Channel wise analysis of lead revenue averages	Campaign, Channel, Region, Product, Date

Table 3. Sample Reporting Templates

6. Program Approach, Schedule & Costs

The following section provides details in expected overall program schedule and resourcing and respective detailed resource loading, cost model for the initial <XX> month functional streams and a 12-month subsequent production management engagement.

6.1. Approach

Key message

Product Management team will be a key component of Governance to enable Drive to self-manage the UDP post program outcomes.

To ensure the successful implementation of program outcomes and to enable Drive to eventually self-manage the Unified Data Platform, we propose to have the program executed in 2 development phases and one support phase, enabling a clear governance model. The expected streams for the program are:

- Discovery Phase – This phase is proposed to define and design the solution, align it with the overall strategic program plan more clearly.
- Build Phase – this phase is proposed post the discovery phase to build the UDP solution in an agile fashion.
- Product (support) – This will be responsible for Governance, Enablement, Release Management, Management Enhancements backlog as well as tooling and daily BAU activities

Costing Approach

The discovery and development phases are modelled on a Time & Material engagement model. These phases will use a resource mix of onshore and offshore people to man the different required roles.

The production management phase is modelled as a pool of hours to be drawn down for any issue resolution or enhancements. These may be serviced by onshore or offshore team members as the need arises.

Periodically, the team will review and reprioritise requirements and review the structure and overall performance of the team.

Resourcing Considerations

The resourcing for Product Management and Project streams takes into consideration that there will be key resources provided from Drive that will take ownership of specific elements after initial implementation.

For the managed support period following production go-live, an identified set of resources with clear understanding of both the business requirements as well as the technical implementation will assist in on-going activities, including any fixes or enhancements.

In addition, time will be required from respective unit SMEs to provide detailed insight, requirements as well as user acceptance testing of developed and configured modules.

6.2. Deliverables

Discovery Phase

We propose a discovery phase of 4 weeks. This will help define the architecture and the solution model for the UDP and related integrations.

The key deliverables for this phase are:

- Prioritised stories (backlog)
- Wireframes & Reports UX model for Business Users
- Integration specifications (for Source Systems)
- Detailed solution design (incl. architecture & data model)
- Test strategy & high-level test paths
- Rollout phase planning (phase-wise lead types onboarding)
- Detailed Development Schedule & Sprint plan - resources, costs & timelines
- Environments & DevOps plan, including any automation
- Project Management plan

Build Phase

This will be planned and executed subsequent to the Discovery phase. Once the discovery deliverables are defined and a development plan is finalised, the development phase will be started.

We anticipate the development phase (based on current understanding and estimates) to span 5-6 months. (~25 weeks).

Development will be done by a mix of onshore and offshore resources. We will use an iterative development model based on agile principles, with stories planned and developed in sprints.

The key deliverables of the solution are:

- Implementation of UDP on the finalised technology infrastructure
- Pipeline Integrations with the source systems (listed in sections above)
- Unit testing of all program code
- System Testing of the complete solution
- Support to Drive team in User acceptance testing
- Lights on BAU Activities post phase deployments
- Data migration & loading (data setup) for go-live cutover.

6.3. Schedule

Illustrated below the high level proposed schedule for development of the UDP data warehouse solution. We propose the following structure for the project:

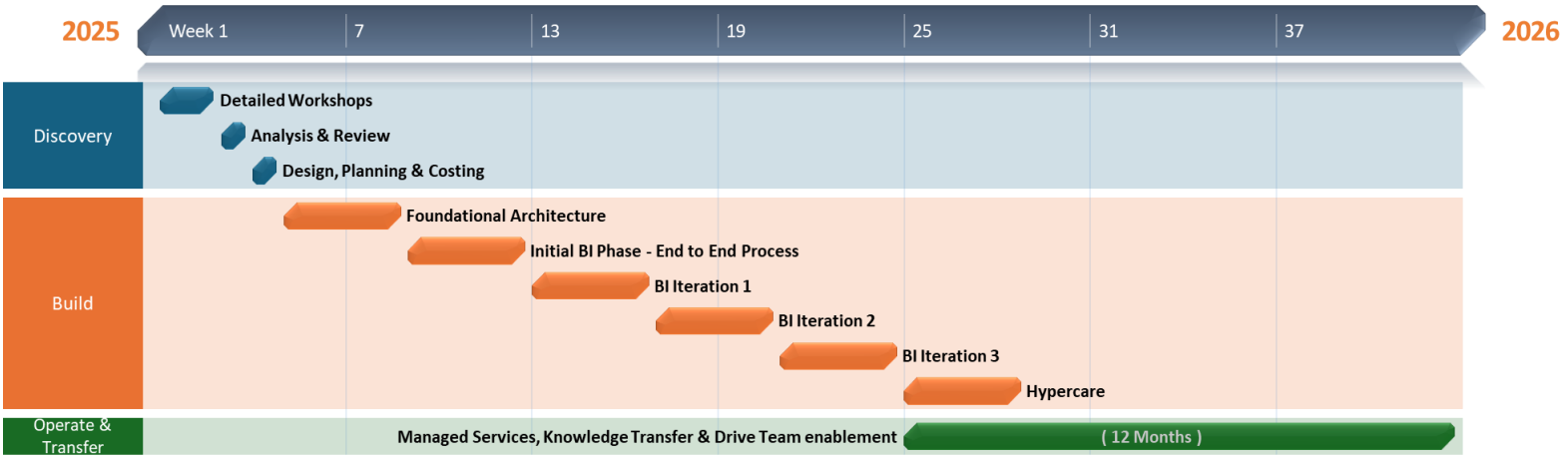


Figure 1. High Level Timeline Summary

Note: Resource availability and Public holidays have not been considered in timeline. This will need to be planned and may impact elapsed delivery times

Following the Foundation Phase, each BI Phase will follow the below iterative cycle for development, configuration and deployment.

Phase	Duration	Key Activities
A. Ingestion Build		Configure all connectors, RAW schemas, monitoring/alerting
B. Transformation Build		dbt staging/marts, incremental logic, snapshots, tests
C. Semantic Layer Setup		Define metrics in dbt SL or deploy Cube.js if needed
D. BI & Dashboarding		~25 reports/dashboards, governance, embedding/SSO
E. Testing & Validation		QA, performance, failover drills, RBAC validation
F. Deployment & Handoff		Prod cutover, runbooks, documentation
G. Post-Go-Live Support		Hypercare, tuning, minor enhancements

Table 4. Key Activities for Development

The schedule outline does not take into consideration impact of Drive change management activities. Detailed joint project and sprint planning to be carried out as part of the discovery phase.

Overall, we expect the project to be executed in about 4 weeks of discovery + 24 weeks (6 months) of development, following which we cut-over to production and hypercare.

bluepineapple

6.4. Implementation Cost Schedule

The project will be executed on a Time-and-Material basis. The schedule of resources planned for this project are listed in the table below:

These are divided across the Discovery and Development Phases.

Discovery Phase (~4 weeks)

Resources & Costs

Engagement: Drive UDP Discovery **Timeframe: 4 Weeks**

Role		Daily Rate	Effort Days	Week 01	Week 02	Week 03	Week 04		
				Discovery					
				1-Sep	8-Sep	15-Sep	22-Sep		
Offshore - Project Manager	$	800	4	1	1	1	1	$	3,200.00
Offshore - Technical Architect	$	1,000	16	4	4	4	4	$	16,000.00
Offshore - Senior Developer	$	625	12	3	3	3	3	$	7,500.00
Offshore - Developer	$	475	9		3	3	3	$	4,275.00
Offshore - Developer	$	475	9		3	3	3	$	4,275.00
Offshore - Senior Quality Analyst	$	625	9		3	3	3	$	5,625.00
							SUBTOTAL	$	40,875.00
							STRATEGIC PARTNERSHIP DISCOUNT	$	22,481.25
							TOTAL	**$**	**18,393.75**

Table 5. Discovery Phase Resourcing & Cost

Build Phase (~25 weeks)

Resources & Costs

Engagement: Drive UDP **Timeframe: 6 Month**

Role		Daily Rate	Effort Days	Week 01	Week 02	Week 03	Week 04	Week 05	Week 06	Week 07	Week 08	Week 09	Week 10	Week 11	Week 12	Week 13	Week 14	Week 15	Week 16	Week 17	Week 18	Week 19	Week 20	Week 21	Week 22	Week 23	Week 24		
				Foundation				BI Initial				BI 1				BI 2				BI 3				Hypercare					
				29-Sep	6-Oct	13-Oct	20-Oct	27-Oct	3-Nov	10-Nov	17-Nov	24-Nov	1-Dec	8-Dec	15-Dec	22-Dec	29-Dec	5-Jan	12-Jan	19-Jan	26-Jan	2-Feb	9-Feb	16-Feb	23-Feb	2-Mar	9-Mar		
Offshore - Project Manager	$	800	34	2	2	2	0	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	$	27,200.00
Offshore - Technical Architect	$	1,000	38	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2					$	38,000.00
Offshore - Senior Developer	$	625	102	4	5	5	0	5	5	5	5	5	5	5	5	5	3	5	5	5	3	5	5	3	3	3	3	$	63,750.00
Offshore - Developer	$	475	90	4	5	5	0	5	5	5	5	5	5	5	5	5	3	5	5	5	3	5	5					$	42,750.00
Offshore - Developer	$	475	102	4	5	5	0	5	5	5	5	5	5	5	5	5	3	5	5	5	3	5	5	3	3	3	3	$	48,450.00
Offshore - Senior Quality Analyst	$	625	90	4	5	5	0	5	5	5	5	5	5	5	5	5	3	5	5	5	3	5	5					$	56,250.00
Offshore - Quality Analyst	$	475	102	4	5	5	0	5	5	5	5	5	5	5	5	5	3	5	5	5	3	5	5	3	3	3	3	$	48,450.00
																											SUBTOTAL	$	324,850.00
																											STRATEGIC PARTNERSHIP DISCOUNT	$	64,970.00
																											TOTAL	**$**	**259,880.00**

Table 6. Development Phase Resourcing & Cost

Note: This cost model is only for the development effort and does not include any license fees or fees incurred towards infrastructure or hosting. The Development Phase costing is indicative and may be impacted by the outcome of Discovery & Solution Phase.

6.5. Managed Services - BAU Support

We propose managed services model for BAU support comprising of as a pool hours leveraging our innovation services offering.

We propose a pool totalling 750 hours to be consumed within 12 elapsed months from the beginning of the support period. The support is all-encompassing covering new development, enhancements, bug-fixes, administration, integration support, data loading, reporting etc. For additional summary on managed services please refer to Appendix I.

Support can be accessed by raising a case in the Bluepineapple innovation services support portal. Each support request will be estimated for and sent for approval. Once approved, the request will be processed, and the appropriate hours of support will be drawn down.

Based on a build operate transfer this will provide flexibility of utilisation of resources.

	Quantity	**Unit Price (AUD Ex GST)**	**Amount (AUD Ex GST)**
Innovation Services	750	$100.00	$75,000.00

Table 7.Innovation Services Costs

Alternatively, based on the outcome of both the discovery and build phases, if there is a specific plan of additional enhancements, a dedicated resource model based on specific scope of works can be tailored to suit Drive.com.au business initiatives and priorities.

7. Architecture & Approach

This section outlines the primary considerations and the architecture philosophy behind the proposed solution. As evidenced by the requirements in the RFI, the solution needs to be business friendly and should provide business value, while staying within manageable cost and operational complexity constraints. These and other constraints and principles are factored into the solution model.

7.1. Proposed Technology Stack

The proposed Unified Data Platform will leverage a modern, cloud-native, best-of-breed technology stack that balances cost efficiency, scalability, and ease of future expansion. All components are proven enterprise-grade and integrate seamlessly in an ELT-first architecture optimised for Snowflake.

The solution follows a Build-First baseline on AWS infrastructure — ensuring production-grade security, resilience, and operational controls from day one — while allowing future phases to scale by simply adding connectors, transformations, or reporting modules without re-engineering core components.

Technology	Role in Solution	Key Features / Benefits	Deployment Approach
Snowflake	Cloud Data Warehouse	Highly scalable, separation of compute & storage, multi-cluster concurrency, time travel, secure data sharing; supports semi-structured data (JSON, Parquet).	Hosted SaaS; available in AWS region closest to primary user base for low latency.
Apache Airflow (via AWS MWAA or ECS)	Workflow Orchestration	Manages ingestion, transformation, and quality-check pipelines; supports complex dependencies and scheduling; open-source with rich ecosystem.	Managed Airflow service (MWAA) or containerised on AWS ECS; integrated with AWS IAM and Secrets Manager.
Airbyte	Data Ingestion Framework	300+ pre-built connectors; supports both API and database sources; incremental sync; open-source with enterprise support option.	Containerised deployment on AWS ECS/Fargate with S3 as staging for file loads and direct Snowflake EL.
dbt (Data Build Tool)	In-Warehouse Transformations	SQL-based data modelling, tests, and documentation; supports incremental models; integrates with CI/CD; creates governed semantic layer.	dbt Cloud (hosted) for developer experience & job scheduling, or self-hosted containerised dbt Core on ECS.
Tableau	BI, Reporting & Visualisation	Rich interactive dashboards, deep analytical capabilities,	Tableau Server or Tableau Cloud; SSO integrated with

		role-based views; strong enterprise adoption.	corporate IdP for secure access.
AWS ECS	Compute & Container Orchestration	Flexible hosting of ingestion, orchestration, and transformation services; integrate with AWS networking, monitoring, IAM.	ECS/Fargate for scalable containerised microservices.

Table 8. Proposed Architecture Stack

The rationale for the proposed stack is explained in the subsequent subsections of this chapter.

7.2. Cost Effectiveness: Architecture & Ops

The key considerations in the choice of tech stack are OOTB functionality that satisfies the solution needs and , cost-effectiveness over a multi-year horizon with cost factoring infrastructure, licensing and resource expertise costs.

- **Pipeline Connectors and Transforms:** Airbyte (pipelines), Airflow (DAG), dbt(transformations) provide required connectivity to the primary sources, are well documented and stable.
- **Cloud Native Stack:** AWS primary (IaaS/storage), GCP for ad/marketing data, Snowflake (data warehouse pay-per-use), S3 (landing zone), dbt (transformation, open source), Airbyte (low-cost ingestion).
- **Phased Resourcing:** Infrastructure can be sized and scaled as needed in line with growing business needs.
- **Low-Code & Automation:** The use of OOTB connectors from chosen product stacks will reduce manual work in onboarding data sources. Similarly, the use of configurable data pipeline solutions reduces complexity in orchestration of data movement to the analytics warehouse.

7.3. Data as a Product

A key theme of the UDP is the ability to provide data and analytics as a product, whereby any downstream consumer (human or system) can connect to the UDP and derive insights tailored to their requirements, while staying within control limits of data privacy and sensitivity. Examples include:

1. **Productised Reporting:** Unified dashboards (Campaign Engagement → Lead Generation→ Dealer ROI) for business units and brands.
2. **Rapid Stakeholder Onboarding:** User-friendly dashboards, RBAC, and the ability to provision embedded analytics.
3. **Reduced Overheads:** Automation of ingestion and (future abilities to derive) narrative reporting can reduce expensive efforts from skilled team members.
4. **Embeddable metric dashboards:** Parameterised dashboards can be published for embedding to other first party or third party portals (e.g. Dealers' portals) for viewing their performance and other metric reports on their portals.

5. **Increased Data Freshness:** Purpose built sequence of data ingestion ensures low latency in data movement to reporting, ensuring insights are derived from fresh information, both raw and correlated. For specific datasets that required near real-time insights (e.g. Lead Events), these are designed to be streamed from the Event queues through to the UDP. In all cases, there will be a definitive view of how aged (fresh or stale) the datasets are in the views.

7.4. Ongoing Integration, Change Management, Flexibility

- Support for onboarding of new sources/platform APIs with minimal disruption or cost.
- Modular, persistent "Drive IDs" facilitate mapping as new entities, sources, and integrations are added. These could potentially be namespaced to identify source of truth/origin.
- Alignment to roadmap requirements (cross-sell, dealer/consumer insight embedding) in early phases.

7.5. Principles

The proposed solution is based on the following first principles to ensure alignment to strategy, high quality, solution maintainability while leveraging top-quadrant technologies for longer solution shelf life.

Principle	Description
Modularity & Extensibility	Add sources and features without breaking live workloads
OOTB & Low Code	Reduce build cost/time using OOTB products and open-source services
Quality Governance	Data validation rules where applicable, synthetic identifiers (drive ids), metadata enrichment
Secure by Design	RBAC, sensitive data protection
User Empowerment	Enable analysts to self-serve within governance boundaries
Agile	To be built over sprints to ensure business feedback is incorporated continuously over the build cycle.
Strategically Aligned	Aligned with medium to long term strategic goals for organisation process automation
Increased information quality	Fix sources of truth of data and model quality management processes for such data.

Table 9. Architecture principles

7.6. Key Stakeholders

The proposed solution primarily aims to provide analytics to Drive Business stakeholders and BI analysts for informed decision making on campaigns, for running more targeted/personalised campaigns, and more accurate understanding of KPI outcomes such as CPL, RPL metrics.

Additionally, this solution will be managed by Drive Technology teams in concert with Bluepineapple teams over a period in order to hand-over the technology reins for in-house management of enhancements and for day-to-day operations.

7.7. Additional Data Sources Coverage

The solution design considers that some **additional datasets as data sources may need to be staged in this phase**. Some of these ad-hoc datasets may need to be manually loaded into the data warehouse in the initial phase for correlated metrics.

- Internal/ad hoc Data: Campaign masters and logs, other master data, and business managed spreadsheets.
- Other social feeds (where relevant to prioritised reporting), Content/WordPress logs/database data: Standardised, mapped for cross-domain analytics.
- Redbook Identifiers & Other 3rd party Ids: Redbook Ids/Vehicle IDs for correlation with external datasets.

7.8. Operational Standards & Security

- RBAC: Implemented across all data, model, and report layers.
- Masking of Sensitive Data (e.g., PII): To be implemented where identified as necessary.
- Persistent Identifiers: Standardised Identifiers for all data to ensure traceable joins across various entities (dealer, lead, vehicle, stock) even as new data types/sources are added.

7.9. Parameters for Platform Choice

While the proposed solution does not consider a fully bespoke-built approach, it is a multi-technology construct. This solution leverages readily available products and/or frameworks that accelerate the development of required features, while staying true to the architecture principles in the previous section.

The stack choice is predicated on the rubric outlined in the RFP viz. Configurability, Cost, Low-Code/No-Code, Security, Gartner Position, Industry Standards Compliance, and OOTB Templates/Accelerators.

- 5 = Excellent / Fully Meets Requirement
- 4 = Strong / Meets Most Requirements
- 3 = Adequate / Meets Minimum Requirements
- 2 = Partial / Gaps Present
- 1 = Poor / Does Not Meet Requirement

Technology / Layer	Configur-ability (Ref. 1, 2)	Lowered Cost / TCO (Ref. 4)	Low-Code / No-Code (Ref. 4)	Security & Governance (Ref. 3)	Gartner MQ Position* (Ref. 1) *[1]	Standards Compliance (Ref. 2)	OOTB Templates (Ref. 5, 6)	Weighted Score (/35)
Data Warehouse								
Snowflake (DW)	5	4	4	5	5	5	4	32
Google BigQuery (DW)	5	5	4	5	5	5	4	33

*Gartner MQ Position from most recent Magic Quadrants per technology category.

Ref. 1 – Gartner Magic Quadrant reports: Cloud DBMS (2025), Analytics & BI Platforms (2025), Metadata Management (2024), plus Gartner Cool Vendors reports for open-source tools like dbt, Airflow.

Ref. 2 – Industry Standards: ANSI SQL (ISO/IEC 9075), Apache Parquet/ORC formats, OpenAPI Initiative, Kimball/Data Vault modelling frameworks.

Ref. 3 – Security & Governance benchmarks: NIST SP 800-57, ISO 27001 Annex A, NIST RBAC model, DAMA-DMBOK2 metadata governance best practices.

Ref. 4 – TCO & Low-Code benefits: AWS/Google/Snowflake cost calculators, Gartner Market Guide for Low-Code Platforms (2024).

Ref. 5 – OOTB Templates: Drive UDP RFI – "Template-driven report generation" requirement; vendor documentation (Fivetran, Airbyte, AWS Glue connectors).

Ref. 6 – Drive-specific: Key Project Goals in RFI – "OOTB tools / low-code / extensibility / reasonable TCO", "Template-driven reporting".

AWS Redshift (DW)	4	4	3	5	4	5	3	28
Data Ingestion & Orchestration								
Apache Airflow	5	5 *(OSS)*	3	4	4	5	3	29
Airbyte (Ingestion)	5	5	4	4	3	5	4	30
Fivetran (Ingestion)	4	3	5	5	4	5	5	31
dbt (Transformation)	5	5 *(OSS)*	4	4	4	5	4	31
Data Visualisation								
Metabase (BI)	4	5 *(OSS)*	5	4	3	4	4	29
Microsoft Power BI (BI)	5	5	5	5	5 (Leader)	5	5	35
Tableau	5	3	4	5	5 (Leader)	5	5	32
Looker Studio (BI)	4	5	5	4	4	5	4	31
DataHub (Catalog)	4	5 *(OSS)*	3	5	3	5	3	28
AWS Glue Data Catalog	4	4	3	5	4	5	3	28

Table 10. Tech choice scoring matrix (Considered options in highlighted green)

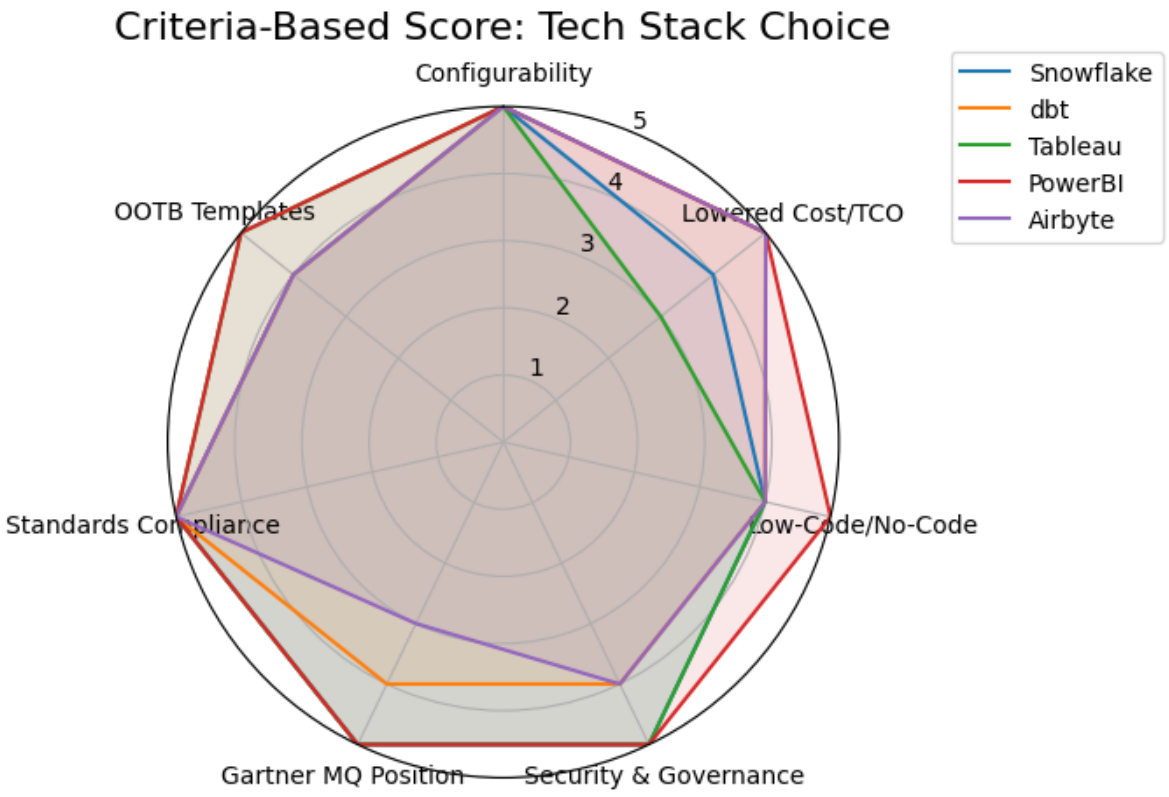


Figure 2. Radar Chart of Tech Stack scores from table

7.10. Platform Scalability & Extensibility

One of the imperatives in the UDP technology stack design is ensuring that the platform remains scalable as business needs evolve. To this extent, the following essential features are considered:

Modular & Extensible Architecture

- Distinct Landing, Staging and Processing/Curation zones for isolation of raw, intermediate and analytics datasets. This helps scale the different zones independently, while allowing easier onboarding of newer data sources.
- OOTB or modular connectors for Google Ads, Salesforce, databases etc allow faster integration of new data sources without major rewiring of the pipelines.
- Pipeline Layering for discrete processing of different stages of data transformation allows component reuse, as well as allows accommodating newer data sources efficiently.

Cloud Native Technologies

- Scalable cloud storage in Snowflake that supports growth in data volumes without disproportionate increase in cost.
- Elastic Compute resources scale up or down based on workloads (including snowflake warehouses with auto-suspension when idle to save costs)

Flexible Data Modelling

- Canonical Data Models to standardise Dimensions and metrics making it easier to add new sources and reporting templates
- Template based reporting to rapidly rollout newer reports through BI tools. Parameterised reports to reduce development efforts towards addressing changing business needs
- Exploratory analysis workbench tools leveraging direct SQL for self-serve capabilities without engineering involvement

Cost and Performance Monitoring

- Resource usage tracking for storage, compute and query costs to optimise for efficiency as data size grows
- Tuning of data models (e.g., addition of materialised/normalised datasets) for management of performance at scale

Example: New Data Source addition

- Standard sources: we use pre-built Airbyte connectors to rapidly ingest data
- Custom or proprietary systems: we will develop a custom connector, but the overall protocol remains the same – ingestion, dbt transformation, metric definition, BI update, QA, and deployment
- Considerations: Data format, connector availability, historical backfill needs, and metric alignment
- Ease: Most standard sources are onboarded quickly using proven connectors and templates; custom sources require extra effort for connector development but follow the same, repeatable process

8. Technical Analysis & Solution

The solution approach attempts to identify the discrete technical system elements that will be leveraged or will need to be custom developed to align with the high-level requirement/business functionality required. Additionally, the approach outlines some cardinal non-functional viewpoints that are critical to the system design and development lifecycle.

8.1. Data Source Inventory

The different data sources (and corresponding platforms) are listed in the table below, along with some calculated assumptions around the frequency volumes. This is indicative only and will be refined further during the discovery process. The frequency, volume and transformation will additionally depend on whether the datasets provide audit timestamps for *Change Data Capture (CDC) for delta loads.*

#	Source System	Data Type / Entities	Target Zone	Frequency	Volume	Protocol	Transformation Required	Error Handling
1	Google Ads	Campaign data, impressions, clicks, spend	Landing → Staging	Hourly	~500 MB/day	API / JSON	Yes – date normalisation, currency conversion	Retry logic, alert on failure
2	Google Analytics 4	Engagement events, sessions, conversion data	Landing → Staging	Hourly	~1 GB/day	API / JSON	Yes – time zone alignment, UTM scheme mapping	Retry + log
3	Salesforce (Marketplace + Advertising)	Leads, Accounts, Contacts, Opportunities etc.	Landing → Staging	Hourly	~250k rows/day	REST/ Bulk API	Yes – field mapping, ID reconciliation	Queue & retry
4	Insider	Campaign metrics, audience segments	Landing → Staging	Daily	~200 MB/day	API / CSV	Yes – segment ID mapping	Retry, fail log
5	Meta Business Manager	Ads, events, conversions	Landing → Staging	Hourly	~400 MB/day	Graph API	Yes – date mapping	Retry logic
6	Other Ad Platforms (?)	Campaigns & creative metrics	Landing → Staging	Daily	Variable	API / CSV	Normalise to canonical ad schema	Retry, QC check
7	JIRA / Monday.com	Tickets, status workflows	Landing → Staging	Daily	~50 MB/day	REST API / JSON	Map statuses to standard workflow dimension	Retry, incremental refresh
8	Lead Events	Lead Stream SNS Topics	Landing → Staging	Near real time	~5000 leads/day	REST/JSON	Normalise to canonical schema	Idempotent retry
9	Custom DB – MySQL/Postgres	Inventory, CMS, specs	Landing → Staging	Daily/ Hourly/Stream[2]	Variable	JDBC/ ODBC	Yes – schema align	Fail/ Alert
10	Flat Files via SFTP/Upload	CSV/JSON	Landing	Ad-hoc	Variable	SFTP	Convert to common format	Reject/Alert on schema mismatch

Table 11. Data Source Inventory

[2] Data load Events (file loads) to be triggers for data loading into UDP

8.2. L1 System View

At a high-level solution, the solution stack is composed of the different data sources (with differing formats and data), ingestion pipelines, staging and aggregation data platforms and analytics. Since each of these components are linked via defined and structured interfaces, these may be swapped out for alternative products with similar functionality.

Architecture Choice – Data Mart First vs Hybrid

The solution allows data ingestion and storage in both Datamart and Hybrid mechanisms. The requirements and design phase of the solution will drive the determination of whether pre-computed analytics are sufficient directly from the source platform, or whether raw data needs to be staged and processed. choosing the path that best meets product/analytics needs. (e.g., GA data)

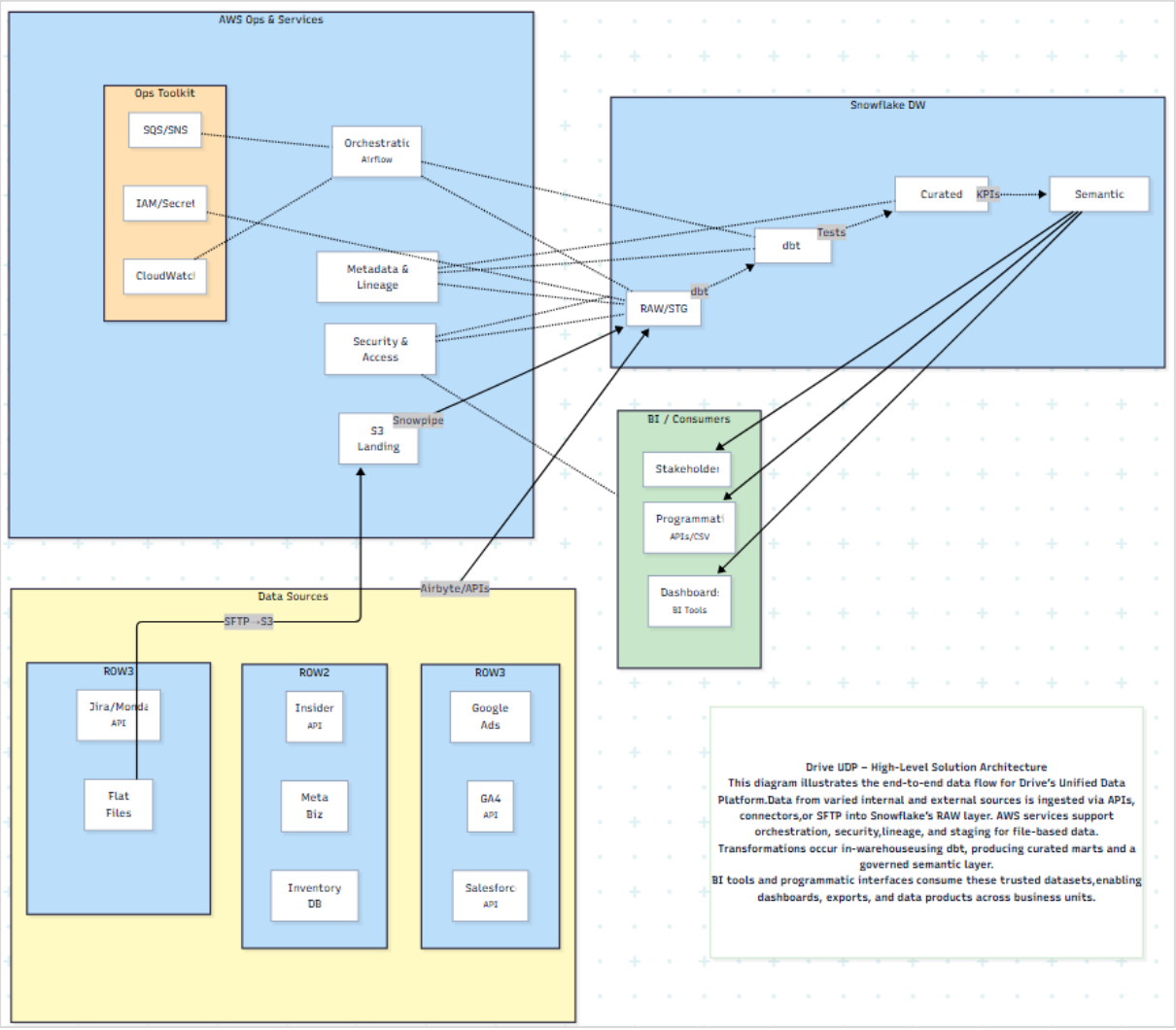


Figure 3. Solution Systems Architecture View L1

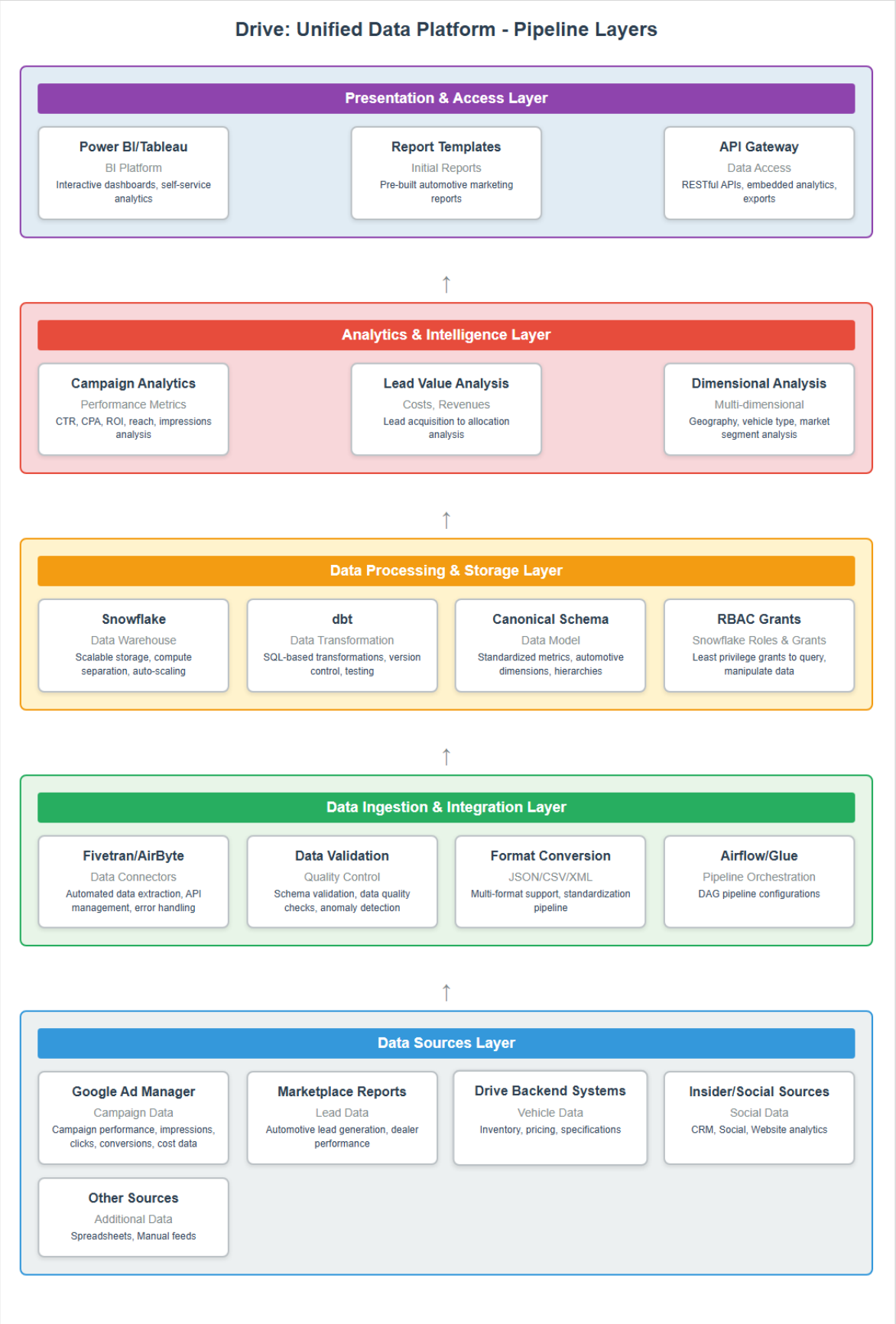


Figure 4. Solution Systems Pipeline Layers View L1

8.3. Pipeline Connectors

The UDP solution requires connectors for different data sources to ingest the data (periodically or real-time) into the staging data schemas and then through to the reporting warehouse. This involves orchestration of ingestion and transformation of these different datasets in different pipelines. Below table illustrates the list of potential pipelines and connectors anticipated for this solution.

#	Source System / Data Type	Connector Type	Example Tools / Options
1	Google Ads	Native API connector pulling Campaign, Click, Impression, Spend metrics	Airbyte "Google Ads" connector, Fivetran Google Ads, BigQuery Data Transfer Service
2	Google Analytics	GA4 API connector / streaming ingestion for hourly engagement data	Airbyte GA4, Fivetran GA4, Google BigQuery native connector
3	Salesforce	Bulk & REST API ingestion of Accounts, Leads, Contacts, Opportunities	Airbyte Salesforce, Fivetran Salesforce
4	Insider	REST API ingestion of campaign metrics, audience segments, attributes	Airbyte HTTP API (custom schema), Fivetran function connector
5	Meta Business Manager	Facebook Marketing API for Ads, Events, Conversions	Airbyte Facebook Marketing, Fivetran Facebook Ads
7	Ticketing Systems (JIRA, Monday.com)	REST API connectors	Airbyte JIRA & Monday connectors, Fivetran JIRA
8	Custom Databases (MySQL/Postgres)	JDBC/ODBC pull with CDC (optional)	Airbyte MySQL/Postgres, Fivetran, Debezium CDC
9	Flat Files (CSV, JSON, Parquet)	SFTP/Cloud Storage ingestion	Airbyte S3/Google Drive, AWS Glue crawler, GCP Storage Transfer
10	Future Data Sources (CRM add-ons, SEO tools, ad exchanges)	Modular plugin architecture	HTTP API connector, cloud storage loader

Table 12. Pipeline and Connectors

8.4. Pipeline Types

Based on the RFI's advice on having separated Landing -> Staging -> Curated data zones, our solution proposes three main pipeline layers. Data is orchestrated from source to reporting schemas using tools like Airflow.

8.4.1. Ingestion Pipelines

These pull data from the raw source through to Landing Zones. The frequency of runs and volume of data will be analysed during the discovery phase. Typically, OOTB Airbyte/Fivetran connectors can be configured for this data flow.

E.g., Google Analytics logs

8.4.2. Transformation Pipelines

Our solution will provision these in Staging zones for processing incoming data (cleansing, enriching via joins). This is typically in Snowflake leveraging tools like dbt and/or pure SQL joins.

E.g., normalisation of datetime formats, mapping external ids to canonical Drive Ids.

8.4.3. Loading Pipelines

These will load the datasets to the analytics warehouse for reporting tools. This will be achieved using dbt or Snowflake stored procedures (copying data from Staging schema to reporting).

e.g., Processed Fact tables such as Leads, Engagement metrics from Staging zones through to reporting.

8.4.4. Orchestration

A suitable orchestration tool (Airflow) will be leveraged if necessary to synchronise the data flows, with dedicated DAG configurations per source system, specialised DAGs for processing and loading. This can help improve loading performance (due to parallelism), as well as provide controls around error handling, retry attempts, logging etc.

Further, such an orchestration approach forces modular solution components, with easily repairable/replaceable DAG nodes.

8.4.5. Error Handling & Retry Mechanisms

The data pipelines incorporate idempotent processing for re-execution of failed tasks. This includes handling transient failures such as network or API timeouts. Errors will be logged for analysis and resolution along with their metadata for diagnostics.
Examples

Retrying Google Ads ingestion failures (Network timeout) with exponential backoff. This will retry the same dataset ranges (date period delta) with deduplication keys to avoid duplication of data in Landing zones (idempotency). If this fails again, this is logged for diagnostics and can be alerted for manual intervention.

Salesforce Governor Limits hit: each load is marked with checkpoint markers (load identifier, datetime offset). Loading Failures can resume from last checkpoint and error details are logged for tuning the load frequency , sizes etc.

9. Non-Functional Viewpoints

The following non-functional systems views are explored to ensure the system is maintainable, secure, and high quality.

9.1. Usability

Ease of use is a first-class design principle in this design solution. The solution is future proofed against evolving business user needs.

The system will provide an easy-to-use UI Platform from day one, while laying the foundation for additional features and self-service analytics in the future.

9.1.1. Phase 1 – Immediate Usability Features

Feature	Description
Single Sign-On (SSO)	Integrated with Drive identity provider (Azure AD) for one-click secure access to all analytics tools.
Unified Access Point	Central portal with role-based menus for dashboards, reports, and KPIs.
Pre-built Reports & Dashboards	Role-specific templates covering majority of reporting needs at launch.
Easy Filtering & Parameter Changes	Users can adjust date ranges, campaigns, regions, and other filters directly in the UI.
Drill-Down & Cross-Navigation	Interactive click-through to detailed transactional views for investigative analysis.
Data Export	One-click export to Excel formats
Responsive UI	Dashboards optimised for desktop and tablet for mobile usability.
Report Delivery	
Complex Visual	

Table 13. UX Feature

9.1.2. Future Phases – Scalability & Enhancements

Feature	Description
Self-Service Analytics Expansion	Users build ad-hoc reports using governed data models without IT intervention.
Natural Language Query & AI Assistance	Conversational analytics converts plain-English queries to visualised insights.
Multimodal Content	Combine images, creatives, and documents with reporting views for richer context.
Progressive Access to New Data	Seamless integration of new domain

Table 14. Future phase Features

9.2. Administrator Controls

For report parameterisation (defaults), error diagnostics, pipeline control configurations etc, superusers & administrators will ideally have a control panel that allows controlling key parameters without having to re-deploy pipelines or infrastructure. Along the lifecycle of this solution, this may include features such as:

Function*	Description
Credential & Access Management	Manage user accounts, roles, and group memberships; integrate with enterprise SSO (Azure AD) for authentication; enforce password/2FA policies for service accounts.
Role-Based Access Control (RBAC)	Define and manage permissions at dataset, dashboard, and pipeline levels to control access to sensitive information.
Default Parameter Settings	Configure global defaults – timezone, currency, fiscal calendar, default dashboard filters – applied across reports for consistency.
Pipeline Configuration	Create, schedule, and modify ingestion, transformation, and load jobs; manage dependencies and priority order in orchestration tools.
Source Connector Management	Add, update, or disable data source connectors via governed workflow; store API keys and DB credentials securely in secrets manager.
Data Quality Rule Management	Define, edit, and deploy data validation and cleansing rules across ingestion and transformation stages.
Audit & Logging Access	View detailed system logs, pipeline run history, change logs for configurations, and user activity logs.

Table 15. Potential Admin Control Panel Functions

** These features may not be provisioned during the initial phases of the project, may need to b be manually configured at the infrastructure level.*

9.3. Security & Controls

Access Control:

The system should provide configurable and role-based access control (RBAC) at datasets (schema/tables) and/or reporting levels. Depending on the user's privilege, information access should be curtailed or enhanced.

This will be achieved using a principle of Least Privilege and aligned using Active Directory Roles (or similar).

Authentication:

The solution will have the ability to provision SSO against the standard SSO configurations used in Drive. Further, it can enforce 2FA/MFA mechanisms for increased authentication safety.

Data Masking & Encryption:

Any catalog items identified as sensitive, that need to be masked/obfuscated or encoded to prevent leakage of sensitive information will be masked as appropriate, without hindering the ability of privileged users to query these or join these entities to other datasets. These will be outlined during the discovery phase.

9.4. Logging & Forensics

There will be an audit trail for changes to critical data catalog, structural changes, configuration changes etc. Critical process executions will be logged into a persistent log for forensic analysis of issues, performance etc.

All data entities will be tagged with audit data of created/last modified timestamps and users.

10. Development & DevOps

At a minimum, there are three logical environments in the solution:

Stage	Purpose	Key Notes
PROD	Live customer / reporting workload	High availability, DR, strict RBAC
UAT / PRE-PROD	Final testing sign-off, user acceptance, performance testing	Mirrors PROD setup
DEV / TEST	Iterative feature & pipeline dev, unit/integration tests	Lower cost scaling, relaxed access

Table 16. Environments

10.1. Strategy

Since this is a greenfield deployment, a tactical build-first strategy can be employed, where an initial Prod-grade base environment is built first, hardened to production grade security standards and tested. This can be designated as UAT for thorough testing, once ready.

All environments will be provisioned using Infrastructure as Code (IaC), [Terraform or similar]

Development and Hardening on Base Env

- Ingestion → Landing → Staging → Processing flow for first priority sources.
- Apply full RBAC, encryption, governance & metadata capture as if live.
- Run with synthetic or masked production-like data until UAT.opment & Hardening
- Phase wise deployment – initial greenfield phase vs subsequent phases after go-live

Flipping Base to Pre-Prod/UAT

- After priority backlog features are built, freeze feature deployment for UAT deployment.
- Tag the environment: set to UAT
- Load real UAT data, and start business user testing

PROD from UAT Clone

- Replicate UAT to a new Prod environment using IaC templates
- Point connectors & joins to Prod
- Configure env with Prod credentials

Smaller scale dev/test

- Set up Dev/Test using same IaC templates but to smaller scaled sizes.

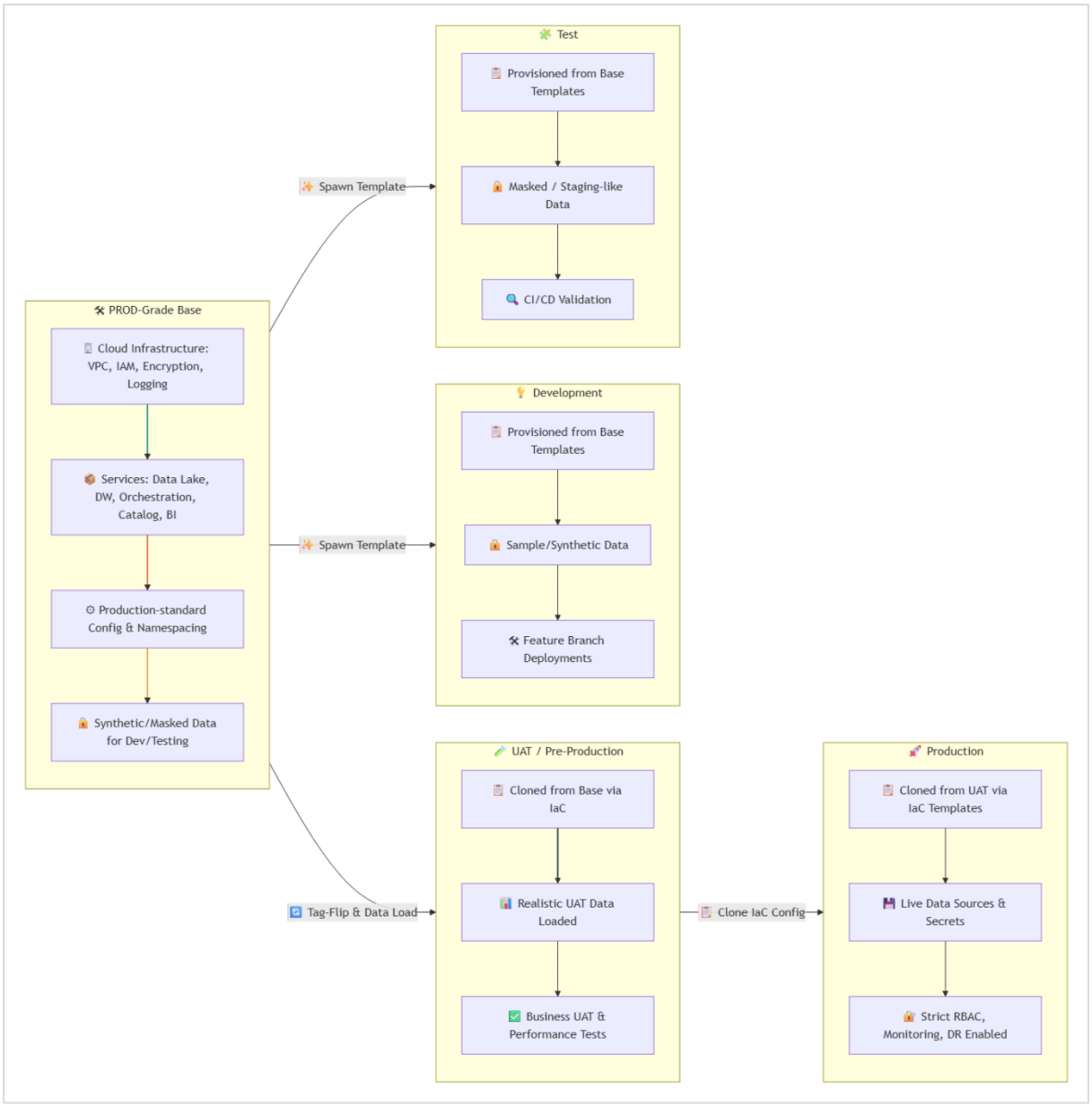


Figure 5. Environments Strategy

Note: This diagram is indicative and will evolve during the design and discovery phases

10.2. Benefits

The above approach allows for faster, lower risk production readiness since there is no specific production hardening phase required. Additionally, the following advantages accrue, helping smoothen the multi-environment management requirements.

- Consistency: IaC ensures UAT/PROD are identical (apart from volume & secrets).
- Reduced Rework: Governance, security, and scaling are tested early.
- Parallel Capability: DEV/TEST can be spun up without touching UAT/PROD stability.

Tactical Note

When building the base environment, all components (S3 buckets, Snowflake DBs, Airflow DAGs) as if they are in live PROD with the appropriate naming conventions derived from environment configurations. Swapping data connections at the end of the UAT will help “flip” from build → UAT → PROD in configuration, with a need for reconstruction of the environment fully from templates.

11. Licensing & TCO

This section details the assumed licensing & anticipated Total Cost of Owership over a three year period.

11.1. Application & Infrastructure Licensing

We recommend that Drive procures the appropriate number of platform licenses (Snowflake, Airbyte, dbt etc. , in this instance) before development commences to use these in development and to set up the governance and deployment model appropriately.

An Indicative list of infrastructure licenses needed. Costs indicated are based on available information on the respective sites online with estimated usage based on assumed data sizes and load frequencies listed in data sources table in earlier sections.

	Licence	Units/Users	Unit Price	Annual Cost
Snowflake	Standard	Multiple X-small WH @ usage of ~ 8 credits/day x 365 days (10hrs/day weekday + 2hrs/day weekend)	1 credit = USD $2.00-$2.75 – Australia (AP-Sydney) is typically USD $0.75 higher per credit than US pricing	AUD 12,000
Dbt	Cloud	1-2 users	100 USD/Dev/month	3,000
Airbyte	Cloud	10-20 connectors	2.5 USD/credit	AUD 12,000
Airflow	Self-hosted	1 Env @ 1 AUD/Hr	1 Env	AUD 9,000
Tableau	Creator, Viewer	5 Creator @USD70/mon, 25 Viewers @ USD15/mon	See calc	AUD 12,000
AWS Infra	EC2, S3, Networking etc	Estimated monthly spend AUD 4000	See calc	AUD 48,000

Table 17. Infrastructure & Application Licencing Costs

11.2. TCO Optimisation

To ensure that costs are predictable, are driven by actual usage, and only scale with increased usage, the licensing models here are without explicit lock-ins.

The TCO calculation uses the following sizing assumptions:

- ~25 dashboard consumers.
- <100 GB data ingested per month (initially).
- **Ingestion cadence:** various, highest is every hour.
- **dbt transformation cadence:** every 12 hours (triggered after the last EL batch in that window).
- **Report/dashboard freshness SLA:** ≤12 hours.
- CI/CD or lightweight orchestration.
- Security, RBAC, and compliance handled within each managed tool where possible.

While some of these frameworks have open-source, free to use libraries available for self-hosting, we recommend procuring licenses for these to both reduce the efforts and complexity of self-hosting as well as to avail of the upgrades and support made available via subscription.

Component	Year 1	Year 2	Year 3	Three-Year Total
AWS Infra (S3, EMR)	$48,000	$30,000	$30,000	$108,000
Snowflake Data Warehouse	$20,000	$20,000	$20,000	$60,000
Airbyte Support/Hosting	$12,000	$12,000	$12000	$36,000
dbt Team Support	$3,000	$3,000	$3,000	$9,000
Airflow (Orchestrator)	$9,000	$9,000	$9,000	$27,000
Tableau licenses	$12,000	$12,000	$12,000	$36,000
Misc/Contingency	$4,000	$3,500	$3,000	$10,500
Total	$108,000	$89,500	$89,000	**$286,500**

Table 18. Component Yearly costs in AUD

Further, for the highest cost elements in the infrastructure, i.e. AWS and Snowflake, we assume that the reporting can be derived using X-Small warehouses in all the environments (including Prod). These costs are anticipated to reduce after the initial development period burst usage.

Assumed Usage Scenario	WH Size	Credits/Hr	Est. Active Hours/Month	Est. Monthly Spend
Minimal (POC / light dev)	X-Small	1	120–180	$240 – $360
Steady prod (4-hr EL, 12-hr dbt)	Small	2	300–450	$1,200 – $1,800
Burst / concurrency spikes	Medium	4	200–300	$1,600 – $2,400

Table 19. Snowflake WH costs (indicative)

Further, this assumes tight auto-suspend (1-5 mins) after each job run, separating EL/transform WH from query WH to conserve costs.

- Time Travel/Fail-safe beyond defaults adds storage cost; monitor large transient tables – these are not assumed. For any critical data, we can provision backup to S3.

12. Case Study

12.1. Unified Data Platform for IAG Australian Insurance Carrier

Insurance Australia Group (IAG), a large Australian insurance carrier needed to integrate monthly bordereaux submissions covering written premium, claims, and paid losses from different underwriting agencies, each covering multiple products.

The data arrived in different formats (in Excel) with tables containing 100+ columns — including policy and risk identifiers, dates, geographies, coverage classes, and financial metrics. This data had inconsistent field names, coding standards, and validation rules created significant reporting delays and compliance risks.

Challenges

- **Heterogeneous Data**: Agencies used inconsistent schemas, terminologies, and currency/date formats.
- **Data Volume & Complexity**: Wide tables (>100 columns) needing efficient processing.
- **Data Quality Gaps**: Missing values, invalid codes, duplicate transactions, mismatched aggregations.
- **Canonical Model Requirement**: Unified schema across all agencies for portfolio-level analytics.
- **Regulatory & Management Reporting**: Timely, accurate bordereaux reconciliation for compliance, reserving, and performance monitoring.
- **Complex Reporting**: Complex queries for calculating performance, loss ratios, debtor ageing etc.

Cloud-native data platform (AWS/Snowflake) with automated **Multi-Source type ingestion**, data quality (DQ) remediation, normalisation, and analytics enablement.

Data Quality & Fix Rules

- Referential integrity checks across related datasets
- Detection and fixing issues (date parsing, default value substitution, code remapping, complex lookups)
- Configurable fix-rules for agency-specific quirks without hard-coding.

Analytics & Reporting

- High-Performance Processing
- Canonical Data Model for unified fact/dimension model covering: Policies, Claims and Payments
- Surfaced clean, canonical datasets into a BI layer – Solvefins Platform UI, Extracts for SaS
- Reports for Written, Claims and Paid analysis: Earned, Gross, Claims/Loss Ratio, Agency performance
- Regulatory bordereaux packs and management dashboards with drill-downs.

Outcomes

- Data Load Time Reduced: From 10-12 days manual to <8 hours automated load & transform across all agencies.
- DQ Issue Rate Cut: 70% fewer manual DQ exceptions due to embedded rules.
- Portfolio-Wide Analytics: Consistent reporting across agencies, enabling consolidated risk and profitability insights.
- Process & Compliance: Accurate reporting of risks and accounts to government regulators and to in-house financial teams

Key Success Factor: By building the canonical target model first, then iteratively mapping each agency's feed to it, the team reduced onboarding time for new agencies from months to weeks, with minimal re-engineering.

13. Delivery Methodology

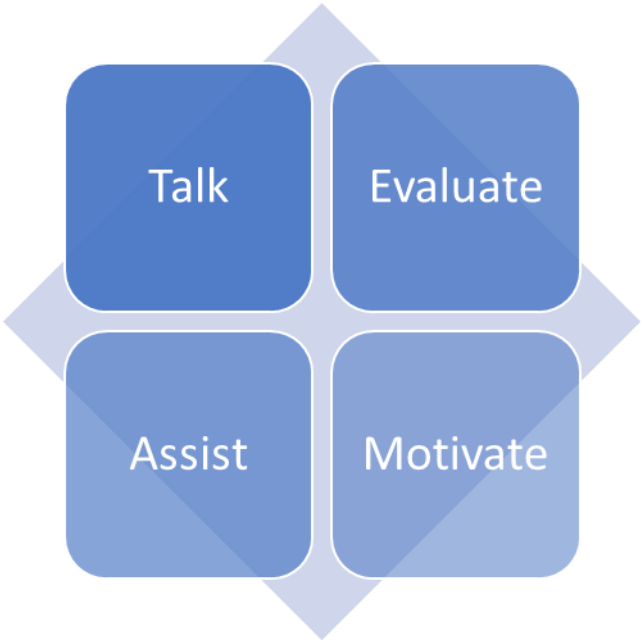
13.1. TEAM FIT

The Bluepineapple TEAM-FIT model is designed to ensure successful delivery of projects in distributed teams by leveraging an understanding of team composition and strengths and ensuring strong communication throughout project cycles.

Each element of the TEAM-FIT model is based on objective questions that lead to clear outcomes.

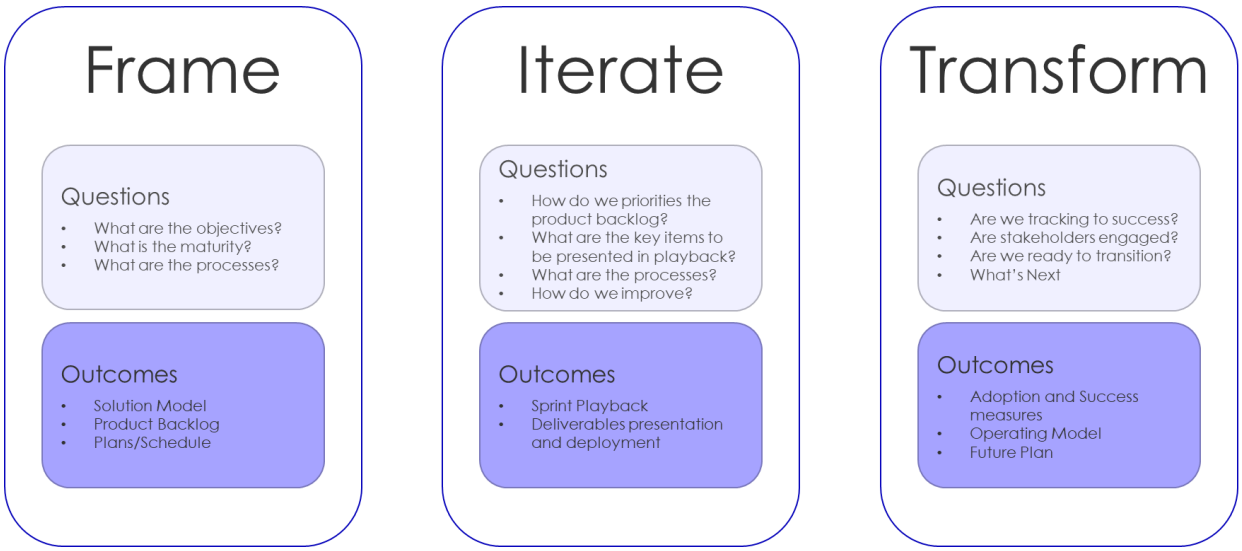
13.1.1. The TEAM Model

This is used to form and develop highly productive teams by managing behaviour. It is illustrated below at a high level.



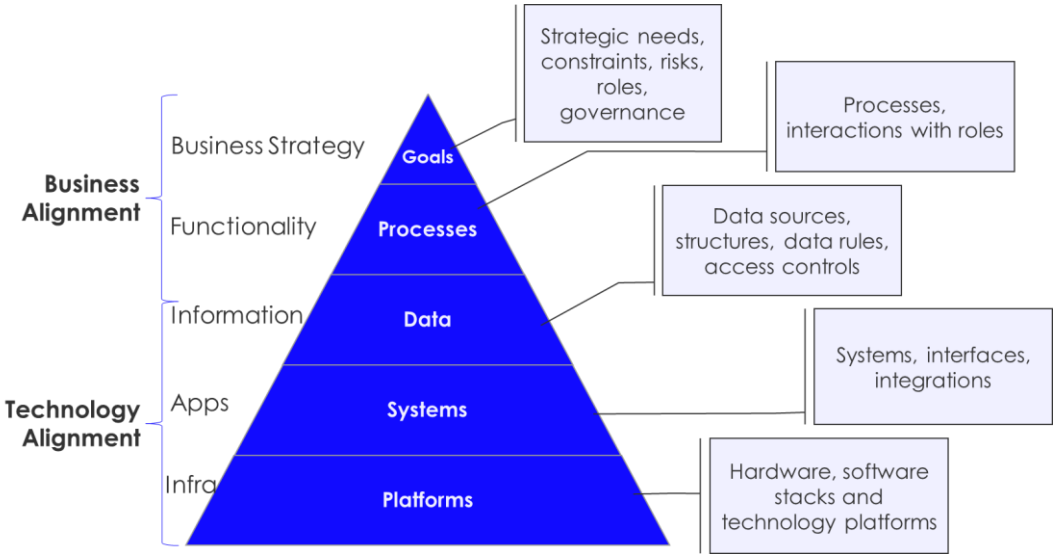
- The Talk aspect of this model lays out the team structure and the communication plan to ensures constant communication to help identify team roles, responsibilities and constraints.
- The Evaluate aspect focusses on team dynamics to help develop individual and team plans to set expectations and monitor results.
- The Assist part of this model helps keep team members productive by leveraging team expertise as well as by keeping them accountable by providing proactive feedback on ownership, deliverables and deadlines.
- The Motivate aspect of this model focusses on positive reinforcement of the team by identifying motivators and celebrating achievements.

13.1.2. FIT Model



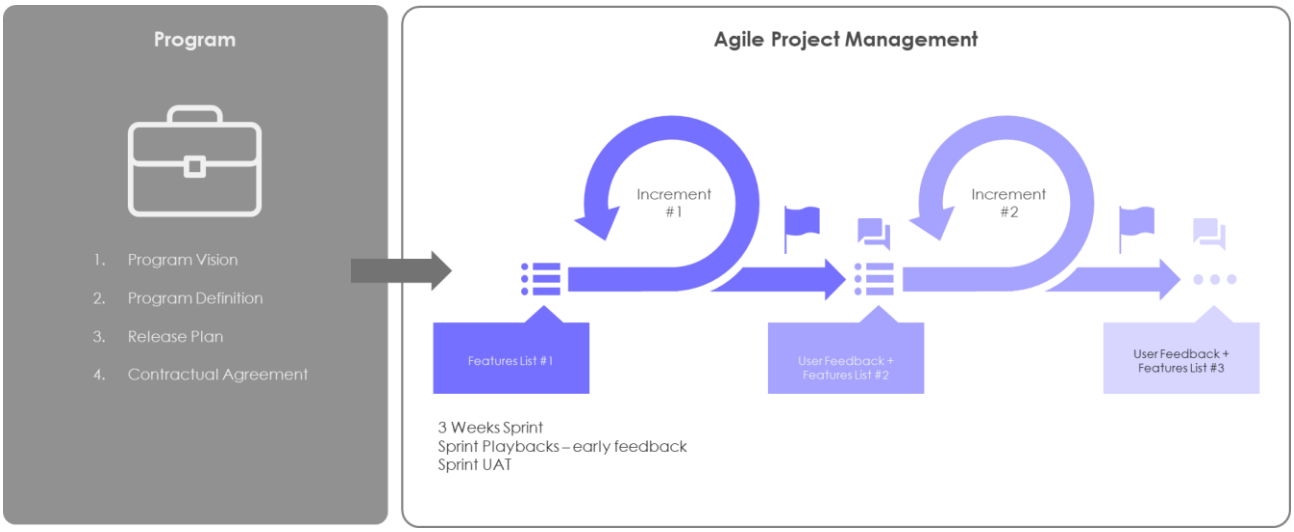
Frame

We use the Frame phase to define the project/program and establish context at all stakeholder levels. From initial discussions, through to problem definition, architecture and solution definition, this approach aims at structuring discussions, workshops, and approaches throughout the engagement lifecycle.



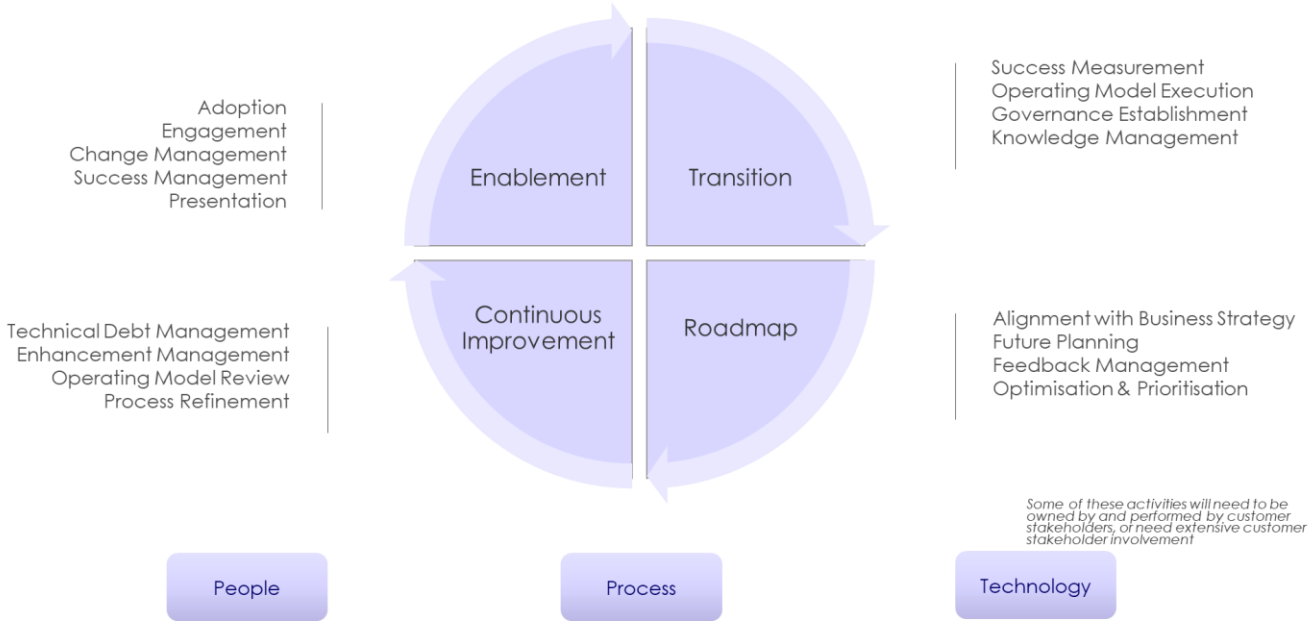
Iterate

We run development sprints along the Iterate phase to generate release artefacts. These are validated for quality and released to production as appropriate. As the name suggests, we iterate through and burn down the product backlog in multiple sprints across multiple releases.



Transform

The Transform phase spans the entire project and provides guidance to ensure project/program success. This includes defining adoption and change management measures and a continuity plan for ongoing operational improvement.



14. Quality Assurance

The following areas are critical to quality and will be monitored during the program

- Configuration management
- Testing. i.e. Quality Control

Configuration Management

A standard configuration management plan will be enforced to manage compliance guidelines for documentation, programming environments, tools and platform artefact versions. This includes release management using CI/CD tools.

A rigorous CM plan will be used to enforce gating around code/configuration quality. Automation, where available will be implemented to ensure efficiencies and instant feedback on code/configuration quality and adherence to guidelines.

Testing & Quality Control

Development and Testing will be executed in parallel during the Iterate cycles of the program. All system, UAT and regression testing will be performed in controlled, defined environments. Testing will be primarily of four types.

- Unit Testing – this will be performed by developers during development/refactoring.
- System Integration Testing – this will be performed by dedicated QA personnel on checked in, already unit tested branches of development. This testing is to ensure that parallelly developed aspects of the solution are correctly merged and functionally consistent.
- User Acceptance Testing – this will be performed by Drive teams in order to validate the readiness of the solution for end-users.
- Regression testing – this will predominantly be automated using tools and will span a subset of total functionality such that validation of the selected subset will provide assurance of continued quality of existing solution aspects.

A detailed quality plan will be prepared with entry/exit criteria for each release. These plans will cover relevant test-cases and identify potential test-cases for automation for regression.

14.1. BAU Quality Assurance

For BAU cases, each defect/issue raised will be triaged against a severity matrix to classify prioritisation.

These cases will then be estimated and resolved with inputs to both configuration management (release) teams as well as other development project streams.

High priority fixes that are deployed to production will be replicated/merged into existing development streams as required periodically to avoid larger conflict resolution complexity in downstream project cycles.

Appendix I: Innovation Services for BAU

The innovation services team will work on and assist in resolution of any issues raised by Drive. These include quick fixes to data and program logic as well as generation of reports. In essence, any identified system issues that are hampering business processes can be addressed by this BAU team including minor and medium enhancements

In order to structure ad hoc issue requests into a manageable framework, the BAU team will have defined processes for

- Capture and acknowledgement of raised issues
- Communication – meetings or email
- Estimation and timelines
- Approvals
- Testing and Quality Control
- Deployment and merge to larger program trunk.

Issue Severity matrix

Each of these issues will be tagged with a severity as per the below table

Severity	Description	Response Time
System Outage	• Showstopper • Major error that affects multiple areas of the business • Cannot go live until the defect has been resolved	1 Hours
Business Critical	• High impact • Major error that affects a single error of the business • Defect impacts key functionality and must be fixed and migrated in the next release	4 Hours
High	• A functional error that has a significant impact to an area of the • system but not the usability of the entire system • Can go live without resolution but exit criteria must be met • Workaround possible	8 Hours
Medium	• Cosmetic defect • Minor functional error that has minimum impact to the system • Defect does not prevent the business process from being completed	24 Hours
Low	• Enhancement functionality to originally agreed scope • Delivery to be deferred to a future release	Deferred to future releases

Enhancements should ideally be added to the enhancements backlog and prioritised into sprint plans.

Dev/Test environment management

To the extent possible, all development will be performed in dev environments – once these have been validated against requirements, the corresponding validated functionality will be migrated to Test/Staging environments for acceptance. These changes will then be deployed to production.

The development framework will be coupled with the chosen CI/CD tool and a source control repository to ensure strong controls over deployment candidates.

Responsibilities

The following table outlines the administrative and support activities that are covered by the innovation services team as part of the BAU Lights on activities.

Administration	• *User administration* • *Reconfiguration of parameters* • *CI/CD pipelines upkeep* • *Infrastructure monitoring* • *Cost monitoring*
Support	• Integration • Bug Fixes • Reports enhancement • Additional data source onboarding