

Question -1:**Data Governance**

Unity Catalog will be used to provide first line Data Governance on the platform. Has Deakin already aligned on the Data Governance Operating model and roles / processes that need to be implemented?

Deakin response-1:

Deakin's data governance framework and associated roles is currently under review. The Information Governance and Data protection team have specific deliverables will be actively contributing to this project. (Unity Catalog is not expected to be the Enterprise Data catalog).

Question -2:**BI Integration**

There is a reference to Tableau and TM1 being the in-scope analytics platforms / visualisation tools. Will integration with these tools be required to be in scope and is there a requirement for a business semantic layer to integrate into?

Deakin Response- 2:

Yes, Tableau and TM1 will need to be connected/integrated to databricks. The output of Use Case 1 is a BI ready dataset that is ready for dashboard development/analysis. Use case 2 requires integration, but the model is generated on/with TM1.

AJ: To integrate Tableau and TM1, we would seek guidance on appropriate drivers/connectivity setup such as OAuth via AzureAD and also advice on best practice for TM1 connectivity: ODBC, File-based (parquet/CSV) or Azure SQL DB.

Question-3:**FinOps & Reporting**

Does Deakin have a specific FinOps tool requiring integration or is configuration of Databricks native reports sufficient in the initial phase?

Deakin Response-3:

No, Databricks/Azure native reports

AJ: Built-in Usage Dashboard will be sufficient

Question-4:**Data Modelling**

There is a requirement for a medallion architecture within Databricks. Has Deakin already conducted data analysis and mapping of the source systems and the presentation layers or will this be required of the vendor?

Deakin Response-4:

Current/future state modelling is currently in progress. With future state models still in development, the teams may require some guidance/support, but we expect the majority of work to be completed internally.

Question-5:**Source Systems, Integration & Data Scope**

For Oracle as a source system, is there a requirement for a large scale data migration and data cleansing, or will it be point to point integration initially?

Deakin Response-5:

There will be a requirement to migrate historical data from the existing Oracle Data Warehouse to Databricks, this will be for both Callista and Workday data. Point to point integration will need to be

established for current and ongoing data.

Question-6:

Source Systems, Integration & Data Scope

The scope highlights integrating with some primary systems - Oracle, Workday, Callista. Do you have details of the integration patterns for each these specific sources (e.g. API, JDBC and whether they are batch, streaming, CDC)

Deakin Response-6:

Callista (Oracle) data is currently ingested as batch data via ODBC into the Oracle Warehouse. It is anticipated that we would employ streaming ingestion to allow for future use cases however this can be batch if streaming is infeasible. CDC is preferred.

Workday data will be ingested via API from Workday standard and custom reports. Data will need to be ingested several times a day to cater for journal and transaction changes in the source system, for reporting and analysis in TM1

Question-7:

Source Systems, Integration & Data Scope

The foundational setup requires implementing controls aligned with Deakin's Privacy by Design approach, including PII discovery and masking. Is there a specific tool or technique you wish for us to use or are Databricks native transformations sufficient?

Deakin Response-7:

The Information governance and Data Protection Team will be leading this exercise. The privacy by design approach is a deliverable for the team. There is no specific tool or technique currently, and Databricks native transformations are sufficient at this stage. Deakin's approach to Privacy by Design emphasises aligning data workspaces with the relevant privacy collection notices (example: Alumni data logically segregated to Student data, as each persona has distinct privacy statements).

Access to each logical workspace needs to be conditional upon users acknowledging and understanding their privacy data collection and terms of use obligations.

This design principle significantly reduces the need for additional overhead such as PII discovery, masking, and remediation.

As part of the project, the Deakin Information Governance and Data Protection team will provide guidance and instructions for the vendor to catalogue data at the point of ingestion. This proactive cataloguing process ensures that data is classified correctly from the start, further mitigating the need for downstream PII discovery and masking activities.

Our strategic intention is to build the platform with privacy controls embedded from inception, rather than relying on reactive measures.

Question-8:

Project Delivery

There is reference that this project is anticipated to take 12 months. Would you prefer the service provider to align to this timeline or are you open to the vendor driving more aggressive timelines if possible, based on the scope provided?

Deakin Response-8:

We are comfortable anchoring to a 12-month delivery timeline as the baseline. The reason for the longer duration is that we are aiming to build the standards and artefact for scalability and sustainability in parallel. However, we are open to partners proposing more efficient or accelerated delivery sequences if they are realistic, aligned to the scope, and do not compromise platform foundations, security, or adoption. Any proposed acceleration must factor in internal review gates.

Question-9:**Internal Team Involvement**

What are the anticipated roles, responsibilities, and availability of Deakin's internal Data Platform, Network, Cyber, Cloud Systems, and Enterprise Architecture teams across the key delivery phases (platform foundation, ingestion, security, testing), and which tasks are expected to be shared versus vendor-led?

Deakin Response-9:

AJ: Across the key phases:

Platform Foundation

Vendor-led: Platform deployment, workspace and UC config, CI/CD setup, environment hardening.

Shared: IAM design, networking, landing zones, patterns, cloud resource approvals.

Deakin roles involved: Cloud Systems (for landing zones), Cyber (IAM, policies), Networks (routing, private endpoints), Enterprise Architecture (review and approval).

Ingestion & Modelling

Vendor-led: Ingestion pipelines, medallion patterns, optimisation.

Shared: Source-to-target mapping, Business rule validation, semantic alignment, reference model validation.

Deakin roles involved: Data Engineering, MDM/Info Strategy, BI Platforms (where downstream impacts exist).

Security & Governance

Vendor-led: UC security model configurations, catalog/lineage templates, governance patterns.

Shared: RBAC policy design, data product standards, onboarding process development.

Deakin roles involved: Cyber, IGDP, Enterprise Architecture.

Testing & UAT

Vendor-led: Technical testing, unit/integration tests.

Shared: UAT execution and acceptance.

Deakin roles involved: Data Engineering, Business SMEs, BI/Finance/Student domain SMEs.

Overall availability: teams will contribute part-time, with peak involvement during design reviews, IAM/security tasks, and UAT. Vendor should assume core delivery is vendor-led.

Question-10:

What are the internal review and approval process timelines for technical artefacts within Deakin (e.g. Solution Design Document, Security & IAM Document) etc. This will directly contribute to our timings.

Deakin response-10:

Typical review timeframes:

- Solution Design Document: 1–2 weeks
- Security & IAM Design: 2–3 weeks (cyber assessment is the pacing item)
- Integration patterns / networking: 1–2 weeks
- Data models / ingestion patterns: 1 week
- Architecture governance review (ARB): Scheduled fortnightly; allow 2–4 weeks end-to-end

Vendors should build review windows into the delivery plan.

Question-11:

In reference to the Scope of Work requirement in the Data Architecture and Engineering section, is "Reference Use Case Build" an additional item on top of the two use cases captured in the Proposal Background?

Deakin Response-11:

No. The two use cases (Admissions/Enrolments and Finance Delivery) are the reference use cases. There is no additional third use case expected. Reference groups are referred to as governance

bodies to discuss/sign of business rules associated with the use cases. This process and ownership will be driven by the Info Gov and data protection team.

Question-12:

In reference to the Scope of Work requirement, does "establish a clear roadmap for platform adoption" refer to the 12-month duration of the engagement, or is there a requirement to provide a roadmap beyond that

Deakin Response-12:

The required roadmap should cover:

- The 12-month delivery horizon, tied to the initial use cases
- A forward-looking view (high level) of platform adoption beyond the 12 months, including future domains, operationalisation, and scaling patterns

This is not intended to be a large artefact – concise, directional guidance is sufficient.

Question-13:

In reference to the Scope of Work requirement "on-the-job training", what is the estimated size and composition of the team that will be upskilled

Deakin Response-13:

Formal funding approved for 6 team members

Assume approximately 10–12 people across:

- 4–5 Data Engineers
- 2–3 BI/Analytics engineers (for downstream adoption)
- 1–2 Architects (EA + Data Architecture)
- 1–2 MDM/Info Strategy staff

Training should mix shadowing, pair-build, and pattern handover.

Question-14:

In reference to the Scope of Work requirement "Validation and Transition" which party will be responsible for performing user acceptance testing? Should the vendor provide a quality assurance capability to the project?

Deakin Response-14:

- Deakin owns UAT and will provide business SMEs.
- Vendor should provide QA capability for technical validation, pipeline testing, and data reconciliation. Vendor-led quality assurance is expected as part of delivery.

Question-15:

In reference to the Scope of Work requirement "Deakin's Tooling and Platform details/Primary source systems" - beyond Oracle and Workday, are there any major SaaS or Databases that are considered for the scope?

Deakin Response-15:

Primary sources for this phase remain:

- Oracle (Callista)
- Workday (Finance)

Other potential sources (not necessarily in-scope for build, but relevant for patterns and roadmap):

- Salesforce for specific functions
- Student systems (Evaluate, Surveys)
- HR ancillary systems (Peoplesoft – Oracle on prem)
- Identity systems (Entra / IAM)

- Data currently in Cloudera (legacy)
- D2L Brightspace Learning Management System which is a cloud-based service

Question-16:

In reference to the Scope of Work requirement "Deakin's Tooling and Platform details/Integration services" - is there an existing architectural decision to utilise specific technologies like MuleSoft and Azure Data Factory (with existing operations established), or does the scope of the project include assessment of those technologies and providing complete or partial data architecture and integration design beyond Azure Databricks and immediate integrations.

Deakin Response-16:

we are hoping to get advice on the best patterns, tooling and approaches for integration and are not wedded to the existing tools. We would also like to use Databricks native tools if it makes sense.

Question-17:

Initial use cases

Would these use cases focus on the data ingestion, transformation and modelling within Azure Databricks, or would they also include producing insights (e.g. through Tableau)?

Deakin Response-17:

The focus is on the data ingestion, transformation and modelling within Azure Databricks and building the Lakehouse architecture and delivering usable, curated datasets. The focus is not to produce Tableau dashboards.

Question-18:

Do you already have a fixed scope for these use cases, or would your selected partner have the opportunity to work with you to jointly shape the scope during the project?

You mention "rebuilding the Admissions and Enrolments data sets". What is the current state of those data sets? For example:

Can you provide an overview of what currently exists, and what the key gaps are (if any)?

Deakin Response-18:

Scope is partly defined but the selected partner will co-shape details with Deakin.

Current state:

- Admissions and Enrolments datasets are extracted from Callista (Oracle) into Cloudera and Data Warehouse layers
- Models are fragmented, partially duplicated, and require consolidation

Key gaps include:

- inconsistent business rule application
- manually maintained enrichment layers
- limited governance and lineage
- latency issues
- lack of reusable medallion patterns
- no centralised security model (IAM/RBAC)

The rebuild is intended to standardise models, create Silver/Gold tables, and establish future-state data product patterns. this is likely to be the first year of a multi-year project.

Question-19:

Where are the data sets currently - are they in an existing / legacy data warehouse? If so, what technologies are involved and can you provide any volumetrics, e.g. number of tables stored, volume of data stored, volume of data ingested per ingestion load, number of users, etc.

Deakin Response-19:

Currently stored in:

- Oracle extracts processed through DataStage
- On-prem compute with ageing infrastructure

Indicative volumetrics:

- Hundreds of tables across Admissions/Enrolments
- Daily ingest volumes: tens of GB across domains (Callista and Workday)
- Users: approx 500+ downstream consumers (BI, analysts, operational teams)
- Data pipelines: largely batch-based, some manually triggered

Question-20:**For the Finance Delivery use case:**

It appears that there are existing workflows and schemas, but that you would like to improve the end-to-end automation, timeliness, consistency and robustness of the process (e.g. replacing SFTP-based ingestion, implementing MDM). Besides these, are there other key technical or business goals you wish to accomplish?

What technologies are involved with the current workflows?

Deakin Response-20:

You are correct that the goal is to uplift:

- End-to-end automation
- Timeliness
- Consistency
- Replace SFTP ingestion
- Improve business rule management
- Strengthen lineage and auditability
- Prepare for MDM alignment

Additional goals include:

- improving governance around reference data
- reducing pipeline fragmentation
- providing reusable ingestion/modelling patterns

Current workflow technologies:

- Workday Finance
- SFTP file drops
- DataStage ETL for ingestion
- Oracle Data Warehouse for staging
- TM1 for downstream planning

Question-21:

Can you provide any volumetrics, e.g. number of tables stored, volume of data stored, volume of data ingested per ingestion load, number of users, etc.

What is the maturity level of the target Finance MDM (Master Data Management) solution? Is it already defined, or is the vendor expected to help design the MDM schema then implement the solution?

Deakin Response-21:

Volumetrics:

- Finance domain contains approx 50 reference tables
- Daily/weekly loads ranging from MBs to low GBs
- High business sensitivity
- Approximately 200 regular Finance data consumers

MDM maturity:
- Early-stage.

Question-22:

Data platform and architecture

Are Oracle and Workday the primary data sources, or are there others that Deakin would like to integrate with (either as part of or beyond the two use cases), other than Callista?

Deakin Response-22:

Beyond the two use cases, other data sources we would look to integrate in the future are HR and Research, both Oracle-based as well as the D2L Brightspace Learning Management System which is a cloud-based service

Question-23:

Will Deakin require a bulk migration of data from existing data warehouses / platforms (e.g. this may be needed for historical data that doesn't exist in the source systems)? Or is the intention to ingest all data from the source systems (e.g. Oracle and Workday)?

Deakin Response-23:

There will be a requirement to migrate historical data from the existing Oracle Data Warehouse to Databricks, this will be for both Callista and Workday data. Point to point integration will need to be established for current and ongoing data.

Question-24:

Integration patterns for batch (SFTP), CDC, APIs, and streaming are mentioned in the platform details but not aligned to the specific use cases? Would the patterns needed align to the use cases only?

Deakin response-24:

Yes. Oracle (Callista + Finance) and Workday are the initial and primary sources.
Other sources may be phased in later but are not part of the current build scope.
Callista remains the major transactional system for the student domain.

Use Case 2

Where possible, CDC is the preferred approach. Future state is to have silver layer "near real time", <15 minutes, ODBC connection from Oracle copy will be more likely as CDC from Oracle will incur significant costs.

Ingestion from external files will be required

Use Case 3 - API connection to Workday. Extracted multiple times per day.

Question-25:

What is the current readiness and stability of the underlying Azure Landing Zone where Databricks will be deployed (e.g., are networking routes, VNET peering, and required subscription quotas already approved and operational)?

Deakin Response-25:

The Landing zone and express route are deployed and ready.

Question-26:

What are the defined data ownership boundaries between MuleSoft and Databricks? Which core data ingestion activities are mandated to run through MuleSoft versus being built natively in Databricks?

Deakin Response-26:

There is no mandate to use Mulesoft for integration - this information was provided for context.

Question-27:**Data governance**

Does Deakin have a specific reference implementation or interpretation of CAUDIT standards for data architecture that the Medallion layers must strictly adhere to?

Deakin Response-27:

Deakin is planning to build out the silver layer based on the MortarCAPS integration standards (currently v1.3). Gold layer will follow the current Kimball designs used in the Deakin Data Warehouse

Question-28:

Which specific technology is Deakin currently using as its enterprise data catalogue, and what level of API access or native connector support is required/available for the metadata synchronisation effort?

Deakin response-28:

Deakin does not currently have an Enterprise data catalogue. The assessment of requirements and subsequent decision about what platform to use will be part of this project. If a decision is reached separately, or earlier in the project, this will be assessed.

Question-29:

Has Deakin defined a specific PII classification framework and a definitive list of data fields requiring masking across the in-scope source systems (Callista, Workday, Oracle)?

Deakin Response-29:

- Deakin is currently refining our PII classification framework
- The list of fields exist and, although fairly comprehensive, it is not definitive. Deakin teams will be working on this in parallel.

Question-30:**Project governance and delivery**

Are you able to share details around the size of the Deakin Data Platform and Data Engineering and BI team/s that:

Will need to be trained and upskilled as part of the project?

Deakin Responses-30:

Assume approximately 10–12 people across:

- 4–5 Data Engineers
- 2–3 BI/Analytics engineers (for downstream adoption)
- 1–2 Architects (EA + Data Architecture)
- 1–2 MDM/Info Strategy staff

Training should mix shadowing, pair-build, and pattern handover.

Question-31:

Will be available to work with the vendor team for implementation and knowledge transfer throughout the 12-month period, as embedded members of the project team?

Deakin Response-31:

Yes - see above

Question-32:

What are the key internal sign-off milestones and gates required by Deakin (e.g., Enterprise Architecture review, Cyber sign-off) that may impact the 12-month delivery timeline?

Deakin Response-32:

Although the teams are already aligned and motivated to get the project over the line, the typical review timeframes are:

- Solution Design Document: 1–2 weeks
 - Security & IAM Design: 2–3 weeks (cyber assessment is the pacing item)
 - Integration patterns / networking: 1–2 weeks
 - Data models / ingestion patterns: 1 week
 - Architecture governance review (ARB): Scheduled fortnightly; allow 2–4 weeks end-to-end
- Vendors should build review windows into the delivery plan.