<Customer Name>
Oracle Digital Assistant

Solution Definition

October 2023 | Version 3.0

Contents

Document Control

1.1 Version Control

Version	Authors	Date	Comments
3.0	Maurits Dijkens	25 Oct 2023	Initial Document

1.2 Team

Name	Email	Role	Company
Martijn de Grunt	martijn.de.grunt@oracle.com	Specialist Generative AI & AI Services	Oracle
Maurits Dijkens	maurits.dijkens@oracle.com	Specialist Generative AI & AI Services	Oracle

1.3 Abbreviations and Acronyms

Acronym	Meaning
OCI	Oracle Cloud Infrastructure
VM	Virtual Machine
DR	Disaster Recovery
HA	High Availability
DRG	Dynamic Routing Gateway
GNN	Graph Neural Network
PGQL	Property Graph Query Language
ODA	Oracle Digital Assistant
FUSION	Human Capital Management
AI	Artificial Intelligence
RAG	Retrieval Augmented Generation

Document Purpose

This document provides a high-level solution definition for an Oracle Digital Assistant solution and aims at describing the high-level requirements and the to-be architecture.

The document may refer to a 'Workload', which summarizes the full technical solution for a customer during a single engagement. The Workload is described in chapter Workload Requirements and Architecture.

This is a living document; additional sections might get added as the engagement progresses resulting in a final workload architecture to be handed over at the end of the customer engagement.

Business Context

3.1 Executive Summary & Business Value

Customer requires a high-performance Digital Assistant(s) that can support the business for numerous use cases across the organization. These use cases include the following (not exhaustive)

- Analysis
- Integrations with Fusion Applications
- Access to FAQs
- Access to policy documents.

As an initial exploration customer has suggested a Pilot based upon the following scope, utilising Oracle's Limited Availability programme to take advantage of upcoming Artificial Intelligence (AI) features. The key activities of this Pilot would include:

- Creation of on Oracle Digital Assistant based on, an Artificial Intelligence, Retrieval Augmented Generation (RAG) pattern
- Build up a Pilot around on the topic of policies; add policy document and make sure the ODA can answer questions on these documents using the embedded RAG
- Help expose this feature in MS-Teams
- Showcase the out of the box skills for Fusion applications (explain concept and give guidelines)

Workload Requirements and Architecture

4.1 Overview

Currently, Customer does not have a Digital Assistant available to its employees to support the Fusion applications functions. The plan is to build an Oracle Digital Assistant.

Requirement	Description
Languages	English
Channels	MS Teams

4.2 ODA Functional & Non-Functional Requirements

The following sub chapters describe ODA required features and functions as known at the time of writing this Solution Definition document.

4.2.1 Overview

The ODA will provide following functionalities:

- Policy Questions - Capability to answer a set of question based upon PDF policy documents using a RAG model.
- Embedded within MS Teams
- Ability to integrate with Fusion applications

4.2.2 High Availability Requirements

- Oracle Digital Assistant Cloud Service is rated as 99.9% availability for a production shape which is sufficient for Customer's use cases

- For more information on all IaaS/PaaS service level agreements, please check Oracle PaaS and IaaS Service Level Objectives where you can find the Oracle Digital Assistant Service SLAs.

  Oracle PaaS and IaaS Public Cloud Services Pillar Document.pdf

4.2.3 Disaster Recovery Requirements

- Disaster Recovery is not part of the requirement

4.2.4 Security Requirements

The system will reside in Customer OCI Secure Landing Zone.

Customer chatbot will be internally available via Customer MS Teams. The web widget communicates with the chatbot Oracle Web Channel configured in ODA A user must authenticate themself within the Customer ODA dialog flow when he needs to interact with the Fusion systems. The authentication method is based on the OAuth 2.0 and OpenID Connect protocols and uses Customer Google Identity Authentication Service as the authorisation server.

4.2.5 User Communities

User's groups working with the ODA:

- ODA end-users will access the Digital Assistant features through MS Teams

- Internal ODA teams for development and management of the Digital Assistant project would be given the right to access as a group within Oracle Cloud to manage the ODA instance for the development of the Digital Assistant Project.

4.2.5.1 Identity and Access Management

- **Default IAM ID** Oracle Cloud Infrastructure has its Default Identity and Access Management Identity Domain (IAM ID) system to manage users, groups, and policies. All OCI administrators will be maintained and managed in Default IAM ID.

- **Customer ODA IAM ID** Customer ODA t admin and developer roles users are maintained in the Customer ODA environment specific IAM ID instance using the roles:

  - ServiceAdministrator: Full administrator access to the ODA instance.

  - ServiceDeveloper: Access to create and update Digital Assistants and Skills, with some restrictions in data purging and deletion of published artefacts.

  - ServiceBusinessUser: Basic access for accessing Insights and Analytics in ODA.

4.2.6 Environments

Oracle will provision ODA for Customer ODA in the following environments:

- 1 x ODA Development Environment used for all the Skills development efforts.

- 1 x ODA Production Environment used for all deployable production chatbots.

4.2.7 System Configuration Control Lifecycle

Oracle recommended the below approach to automate continuous delivery and updates for Digital Assistants built on OCI:

- Versioning of Digital Assistants is available on ODA Platform where you can have different versions of the same chatbot, and it is done through the console of the ODA platform as out-of-the-box functionality.

- It is recommended to do an Export (backup) of the Digital Assistant to ensure a proper recovery scenario periodically in case any disaster happens which is done also from the console of the ODA platform where Customer' team can perform this task manually.

4.2.8 Management and Monitoring

Management & Monitoring of the system will be done via:

- OCI console: to get access to the service instances supporting the architecture and be able to manage them after being granted the proper rights.

- ODA Platform: For versioning, backups, chatbots creation/update and ODA insights tab where you can view all the statistics and insights about the bot's performance, channels usage frequency, conversations' history, and top intents being asked about through the platform console of ODA.

- Customer will ensure and validate that the solution will be placed under the proper controls for ensuring business continuity, system availability, recoverability, security control, monitoring and management

4.3 Future State Architecture

A physical solution architecture is prepared for the easy understanding the Future State Deployment for ODA in OCI. It is also important that this architecture is used directly for the implementation phase. Below are different sections describing the Future Deployment architecture.

4.3.1 OCI Secure Landing Zone Architecture

The design considerations for an OCI Cloud Landing Zone relate to OCI and the implementation of industry architecture best practices, along with customer specific architecture requirements that reflect the Cloud Strategy (hybrid, multi-cloud, etc). An OCI Cloud Landing zone involves a variety of fundamental aspects that have a broad level of sophistication.

To run your workloads in Oracle Cloud, you need a secure environment that you can operate efficiently. The deployment build that will be added after customer acceptance of the solution definition is based on the security guidance prescribed in the Center for Internet Security (CIS) Oracle Cloud Infrastructure Foundations Benchmark. The CIS Secure Landing Zone reference architecture can be found here. The implementation of this as a landing zone for EBS is shown below, although not all OCI services are specifically needed or will be implemented for Graph.

Customer already have an extensible, CIS based, landing zone and it is proposed that ODA will be deployed into its own compartment structure within this Landing Zone.

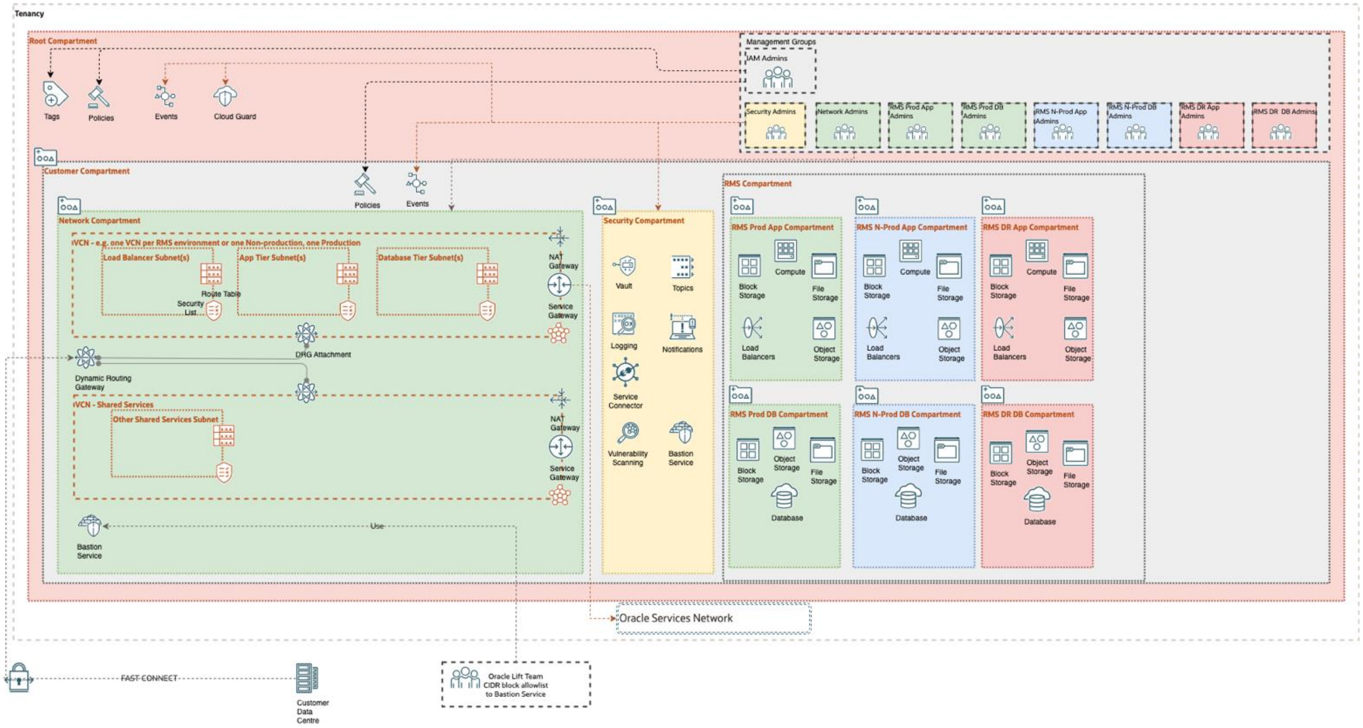

Figure 1: Secure Landing Zone

4.3.1.1 Naming Convention

A naming convention is an important part of any deployment to ensure consistency as well as security within your tenancy. Hence, we jointly agree on a naming convention, matching Oracle's best practices and Customer requirements.

As Customer already has an OCI deployment Oracle recommend using the same naming convention for ODA.

4.3.2 Physical Architecture

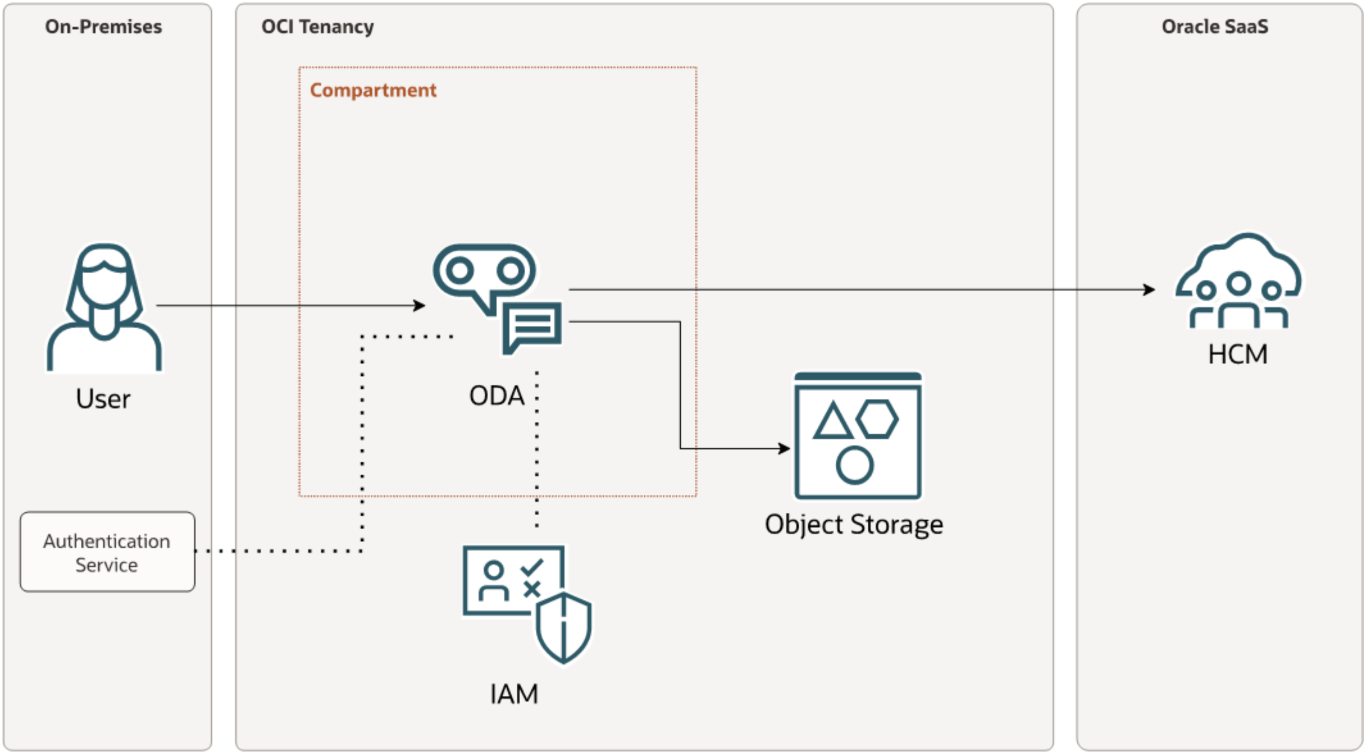

Figure 2: Future State Production Physical Architecture

4.3.2.1 On-Premise Connectivity

Customer will be using FastConnect for their connectivity to OCI via an Equinix Cloud Exchange. The customer is in the processes of deploying a second FastConnect endpoint to OCI to establish diversity of routes and high availability of their network connections. A FastConnect connection is a virtual circuit is terminated at Dynamic Routing Gateway (DRG) that connects to the Virtual Cloud Networks (VCN). It is assumed that there are no overlapping CIDRs between the existing Customer locations to allow for routing through the DRG.

Each VCN is connected to the DRG using a DRG Attachment. The Service Gateway in each VCN provides access to Oracle services such as Object Storage, using the Oracle backbone without needing Internet access. A DMZ Private Subnet already created should there be a need for the RMS Application to be exposed to the Internet via the Palo Alto next generation firewalls.

4.3.3 Networking

Below are the networking security modules deployed in the OCI as per the customer requirement:

4.3.3.1 Palo Alto Next Generation Firewall

As per the customer requirement pre-existing Palo-Alto Firewalls are deployed in both Primary and Secondary Availability Domains for connectivity from and to the internet.

4.4 Sizing and Bill of Materials

The following is an illustrative sizing for the future state architecture shown above. Sizing is based on the information received in response to our discovery questions and applying Oracle's Solution Center Graph Reference Architectures for user populations and internal sizing tools to this information. However, sizing tools by their very nature cannot capture all the nuances of a particular customer's implementation, in particular the use of customizations, performance of long running processes or integrations etc. or the exact customer business processes the application is supporting. Sizing is therefore indicative and does not remove the need for customer testing and resource adjustment to ensure that the OCI deployment meets their specific situation.

Product	Note	Qty
B90260 - Oracle Digital Assistant Cloud Service – Request Per Hour	Dev (50) ± Production (200)	250
B91628 - Oracle Cloud Infrastructure - Object Storage - Storage - Gigabyte Storage Capacity per Month	Backup & Documents	1,000
B91627 - Oracle Cloud Infrastructure - Object Storage - Requests - 10,000 Requests per Month	Backup & Documents	1 x 10,000
B91633 - Oracle Cloud Infrastructure - Archive Storage - Gigabyte Storage Capacity per Month	Backup	1,000

Figure3: BoM

ODA Pilot Implementation

In order to proceed with the ODA solution it has been agreed that a Pilot implementation of the Graph Solution be deployed within the Customer Tenancy.

This will be deployed by utilising skills and resources from both the Oracle Technology Specialist and Customer Data Teams.

5.1 Implementation Scope

This pilot proposal is based on the following assumed scope of services:

1. OCI CIS Landing Zone Extension (note: there is already an existing Landing Zone in the customer tenancy)

2. OCI Foundations:
    - Create VCNs and Subnets
    - Create Route Tables, Security Lists and NSGs
    - Create Compartment Structure

3. ODA Infrastructure
    - Create PaaS Service
    - Create Object Storage Bucket
    - Configure Backups

4. Test Cases and Success Criteria (joint input)
    - definition of test cases
    - ODA model design to support the test cases
    - include scaling/elasticity for training environ (both vertical and horizontal)
    - execution of the test cases including performance measurements
    - tracking test cases against defined success criteria
    - ad hoc queries

5. Quick-Start Training
    - Workshop to help analysts/data scientists to get started with Oracle Digital assistant including modelling a RAG and the fundamentals of Fusion applications integration.

5.2 Deliverables

- Solution definition and Solution design details

- OCI Foundations and Infrastructure for EBS implemented in accordance with the scope

- ODA Solution

5.3 Excluded Activities

The below section provides an insight into the scope exclusions. In general, anything not mentioned explicitly as being inside the scope of the project is being placed outside of the scope of the project. The below list provides additional guidance to this for clarification reasons only and should not be seen as a definitive scope exclusion list.

1. OEM Install and Setup, DB Vault, Data Masking, DB Security installation and Setup
2. IDM-SSO Setup (integration of OCI IAM with Google Identity)
3. On Premise VPN Tunnel, FastConnect, VPN Firewall Configuration, Routing
4. Load Testing, Performance benchmarking, testing & tuning of any component in the solution
5. Third Party Firewall implementation, Security tools, monitoring tools implementation
6. Third Party Backup tool implementation
7. Service Hardening, Audit certification implementation
8. Any Vulnerability Assessment and Penetration Testing
9. Setup DNS server
10. Trainings on deployed products and Cloud Services
11. Implementation of Production Workloads
12. Any other activity not listed under “Implementation Scope” section

5.4 Disclaimer

As part of the ODA Pilot Project, any scope needs to be agreed by both the customer and Oracle. A scope can change but must be confirmed again by both parties. Oracle can reject scope changes for any reason and may only design and implement a previously agreed scope. A change of scope can change any agreed times or deadlines and needs to be technically feasible.

5.5 Environments

- Non-Production ODA
- Production ODA

5.6 Success Criteria

XXX To be defined along-side Customer

5.7 Assumptions

For this Pilot proposal, we have made some key assumptions:

- All required contractual agreements between Oracle and the Customer are in place to ensure an uninterrupted execution of the project. A lead time of at-least two weeks will be required to kick-off the project after contracts are in place and the scope has been agreed upon between the Customer and Oracle.

- All work will be done remotely and within either central European time or India standard time normal office working hours.

- Unless explicated stated ongoing Database and Application upgrades are excluded from the scope of work.

- All required Oracle cloud technical resources are available for use during the duration of the project and that engineers involved have been granted the appropriate access to those technical resources by the customer prior to the start of the project.

- All required customer resources, and if applicable third-party resources, are available during the duration of the project to work in an open and collaborative manner to realise the project goals in an uninterrupted fashion.

- All required customer resources, and if applicable third-party resources, are aware of all technical and non-technical details of the as-is and to-be components, there intend, and technical working as far as is needed for the execution of the project.

- All required documentation, system details and access needed for the execution of the project can be given / granted to parties involved when and where deemed needed for the success of the project.

- The customer will have adequate licenses for all the products that may/will be used during the project and that appropriate support contracts for those products are in place where the customer will take the responsibility of managing any potential service request towards a support organisation to seek resolution of a problem.

- The customer will provide the appropriate level of information and guidance on rules and regulations which can directly and/or indirectly influence the project or the resulting deliverables. This includes, however not limited to, customer specific naming conventions, security implementation requirements, internal SLA requirements as well as details for legal and regulatory compliancy. It will be the responsibility of the customer to ensure that the solution will adhere to this.

- The customer will ensure and validate that the solution will be placed under the proper controls for ensuring business continuity, system availability, recoverability, security control and monitoring and management as part of a post project task.

- The customer will take responsibility on testing all functional and non-functional parts of the solution within the provided timeline and ensure a proper test report will be shared with the full team (including customer, Oracle and if applicable third party).

- Any requirement, deliverable or expectation which is not clearly defined as in-scope of the project will not be handled as part of the project and is placed under the responsibility of the customer to be handled outside of the project.

- The Customer will provide URLs and HTTPS Certificates for the application to be deployed.

5.8 RACI

Matrix responsibilities between Oracle and customer including customer obligations and prerequisites

Serial Number	Activity	Oracle ACS	Oracle Tech Specialist	Cus-tomer	TCS
1	Project kick-off	R	C	A	
2	Define Test Cases and Acceptance Criteria	I	R	AR	
3	Project Planning	R	C	A	
4	Provide Consultant Access	I	I	AR	
5	OCI Foundation Build	R	C	AC	
6	Provision ODA	AR	I	I	
7	Configure ODA	AR	I	I	
9	Provide enablement training workshop	I	AR	C	
10	Test the system	I	C	AR	
11	Project completion Sign-off	I	I	AR	

R- Responsible, I- Informed, A- Accountable, C- Consulted

Production Implementation

The Implementation of the Production ODA solution will be performed by XXX/Customer. Please refer to the migration services proposal submitted by XXX for the relevant sections related to the Implementation.

Annex

6.1 Security Guidelines

6.1.1 Oracle Security, Identity, and Compliance

Oracle Cloud Infrastructure (OCI) is designed to protect customer workloads with a security-first approach across compute, network, and storage – down to the hardware. It's complemented by essential security services to provide the required levels of security for your most business-critical workloads.

- Security Strategy – To create a successful security strategy and architecture for your deployments on OCI, it's helpful to understand Oracle's security principles and the OCI security services landscape.
- The security pillar capabilities pillar capabilities reflect fundamental security principles for architecture, deployment, and maintenance. The best practices in the security pillar help your organization to define a secure cloud architecture, identify and implement the right security controls, and monitor and prevent issues such as configuration drift.

6.1.1.1 References

- The Best Practices Framework for OCI provides architectural guidance about how to build OCI services in a secure fashion, based on recommendations in the Best practices framework for Oracle Cloud Infrastructure.
- Learn more about Oracle Cloud Security Practices.
- For detailed information about security responsibilities in Oracle Cloud Infrastructure, see the Oracle Cloud Infrastructure Security Guide.

6.1.2 Compliance and Regulations

Cloud computing is fundamentally different from traditionally on-premises computing. In the traditional model, organizations are typically in full control of their technology infrastructure located on-premises (e.g., physical control of the hardware, and full control over the technology stack in production). In the cloud, organizations leverage resources and practices that are under the control of the cloud service provider, while still retaining some control and responsibility over other components of their IT solution. As a result, managing security and privacy in the cloud is often a shared responsibility between the cloud customer and the cloud service provider. The distribution of responsibilities between the cloud service provider and customer also varies based on the nature of the cloud service (IaaS, PaaS, SaaS).

6.2 Additional Resources

- Oracle Cloud Compliance – Oracle is committed to helping customers operate globally in a fast-changing business environment and address the challenges of an ever more complex regulatory environment. This site is a primary reference for customers on Shared Management Model with Attestations and Advisories.
- Oracle Security Practices – Oracle's security practices are multidimensional, encompassing how the company develops and manages enterprise systems, and cloud and on-premises products and services.
- Oracle Cloud Security Practices documents.
- Contract Documents for Oracle Cloud Services.
- OCI Shared Security Model
- OCI Cloud Adoption Framework Security Strategy
- OCI Security Guide
- OCI Cloud Adoption Framework Security chapter