Architecture and Design Approach

For

Next Generation Sales Engine (NGSE) International   
Customer DNA

Approvals

Electronic signatures are appended at the end of this document. List the names of the approvers and which approval role they are filling below.

|  |  |
| --- | --- |
| **Approval Role** | **Name** |
| **System Owner** | Sharath Gundala |
| **Enterprise Architect or Line Architect** | Nalini Jayanthi Mayur Dongaonkar |
| **ZTD QA** | Belinda Conroy |

Table of Contents

[1. Purpose 3](#_Toc131172638)

[2. Scope 3](#_Toc131172639)

[2.1 Exclusions, Assumptions and Limitations 3](#_Toc131172640)

[3. Governance and References 4](#_Toc131172641)

[4. Terms and Definitions 4](#_Toc131172642)

[5. Roles and Responsibilities 4](#_Toc131172643)

[6. Architecture 5](#_Toc131172644)

[6.1 Solution Architecture 5](#_Toc131172645)

[6.1.1 Solution Description 5](#_Toc131172646)

[6.1.2 Solution Architecture Diagram 5](#_Toc131172647)

[6.1.3 Solution Architecture Diagram Details 5](#_Toc131172648)

[6.1.4 Alternative Solution Architecture Considered (Optional) 6](#_Toc131172649)

[6.1.5 Evolutionary Considerations 6](#_Toc131172650)

[6.1.6 Retirement Considerations 6](#_Toc131172651)

[6.2 Application Architecture 7](#_Toc131172652)

[6.2.1 Application Architecture Diagram 7](#_Toc131172653)

[6.2.2 Application Components 7](#_Toc131172654)

[6.2.3 Application Interface Details (Including one-time data migration) 7](#_Toc131172655)

[6.3 Data Architecture 8](#_Toc131172656)

[6.3.1 Data Flow Diagram 8](#_Toc131172657)

[6.3.2 Key Data Entities/Objects and their Data Sources (DNA-Data) 8](#_Toc131172658)

[6.3.3 ER Diagram (where relevant) 9](#_Toc131172659)

[6.3.4 CRUD Matrix 9](#_Toc131172660)

[6.3.5 Data Processing 10](#_Toc131172661)

[7. Technical Architecture (Mandatory) 13](#_Toc131172662)

[7.1 Technology Standards Mapping 13](#_Toc131172663)

[7.2 Physical Architecture 14](#_Toc131172664)

[7.3 Performance and Scalability 14](#_Toc131172665)

[7.4 Disaster Recover and High Availability 14](#_Toc131172666)

[8. Revision History 14](#_Toc131172667)

# Purpose

The purpose of this document is to present the architecture and design approach for Next Generation Sales Engine (NGSE) International Customer DNA which is meant for the non-US markets. This document will be maintained throughout the life of the solution configuration items in scope.

# Scope

The scope of this project is to build an enhancement of the Solution Selling Model with capabilities to support sales reps' day-to-day workflows to drive innovative growth and improve the effectiveness of sales and marketing activities. The capabilities include Territory Planning, Pre-Call Planning, Identification and Confirmation of customer needs, Developing tailored solution(s) and Post-Call Follow-up. The following list identifies the application master and deployment CI names that are in scope for this document. Note that there may be several documents to support the overall solution design for an Application Master CI and/or Deployment CI.

|  |  |
| --- | --- |
| **CI Name** | **CI Type (Master/Deployment)** |
| Production/Master (PROD) | GitHub |
| Quality (QA) | GitHub |
| Develop (DEV) | GitHub |

Links to the Configuration Management Database (CMDB) are mentioned below.

| Name | Version | Manufacturer | System Reference |
| --- | --- | --- | --- |
| [NGSE Int'l (DEV)](https://compaid.service-now.com/nav_to.do?uri=cmdb_ci_spkg.do?sys_id=fcb0634e1b670158d936a827bc4bcb21) | 1 | Zoetis | fcb0634e1b670158d936a827bc4bcb21 |
| [NGSE Int'l (MASTER)](https://compaid.service-now.com/nav_to.do?uri=cmdb_ci_spkg.do?sys_id=30b0634e1b670158d936a827bc4bcb99) | 1 | Zoetis | 30b0634e1b670158d936a827bc4bcb99 |
| [NGSE Int'l (PROD)](https://compaid.service-now.com/nav_to.do?uri=cmdb_ci_spkg.do?sys_id=b0b0634e1b670158d936a827bc4bcb97) | 1 | Zoetis | b0b0634e1b670158d936a827bc4bcb97 |
| [NGSE Int'l (QA)](https://compaid.service-now.com/nav_to.do?uri=cmdb_ci_spkg.do?sys_id=3cb0634e1b670158d936a827bc4bcb95) | 1 | Zoetis | 3cb0634e1b670158d936a827bc4bcb95 |

## Exclusions, Assumptions and Limitations

* Only data logged in Zoetis systems can be displayed in NGSE – e.g., if an upcoming call is not scheduled in Touchpoint, it is not displayed in NGSE.
* Data is refreshed two times daily (EAME – 1 am est, CLAR - 7 am est and APAC - 12:45 pm est) with all information provided in Zoetis systems at the time of refresh.
* Data is collected from all available Zoetis systems used in the market –Touchpoint, Salesforce marketing cloud, SAP, Hybris and DDP Foundation.

# Governance and References

This Plan is governed by *ITSO-00006 Risk-based Solution Delivery Lifecycle*. Other governing references particular to the Project and listed throughout this Plan can be found in the following Table.

| Document Number | Document Title |
| --- | --- |
| ITSO-00006 | Risk-based Solution Lifecycle |
| ITSO-00007 | Change Control |
| ITJA-00022 | Zoetis Tech & Digital Glossary |
| IT-26839 | [NGSE International - Architecture Diagram](https://zoetis.sharepoint.com/:p:/r/sites/NGSEInternational/Shared%20Documents/General/04.%20Design%20%26%20build%20phase/03.%20Tech%20Enablement/02.%20Key%20Outputs/02.%20Code%20architecture/NGSE%20International%20Architecture%20and%20Design/NGSE%20International%20-%20Architecture%20Diagrams.pptx?d=w331bb030ce6d4384b0712c9f03405951&csf=1&web=1&e=RpaOad) |
| IT-26840 | [NGSE International - Customer DNA Data Dictionary](https://zoetis.sharepoint.com/:x:/r/sites/NGSEInternational/Shared%20Documents/General/02.%20Key%20Outputs/02.%20Design%20%26%20build%20phase/Data_Dictionaries/Dictionary_Output_Files/NGSE%20International%20-%20Customer%20DNA%20Data%20Dictionary.xlsx?d=wea1420eb0e0840c28f23a7ed12ef511f&csf=1&web=1&e=OOaVHU) |
| IT-26841 | [NGSE International - Customer DNA ML Data Dictionary](https://zoetis.sharepoint.com/:x:/r/sites/NGSEInternational/Shared%20Documents/General/02.%20Key%20Outputs/02.%20Design%20%26%20build%20phase/Data_Dictionaries/Dictionary_Output_Files/NGSE%20International%20-%20Customer%20DNA%20ML%20Data%20Dictionary%20.xlsx?d=w342f4323ec0745d4b9ba2883ad22d092&csf=1&web=1&e=PxfuRy) |
| IT-26842 | [NGSE International - Postgres Data Dictionary](https://zoetis.sharepoint.com/:x:/r/sites/NGSEInternational/Shared%20Documents/General/02.%20Key%20Outputs/02.%20Design%20%26%20build%20phase/Data_Dictionaries/Dictionary_Output_Files/NGSE%20International%20-%20Postgres%20Data%20Dictionary.xlsm?d=wc579e6d0e83247e291c93aaf1aca0360&csf=1&web=1&e=bNzebG) |
| IT-26843 | [NGSE International - Postgres ER Diagram](https://zoetis.sharepoint.com/:i:/r/sites/NGSEInternational/Shared%20Documents/General/04.%20Design%20%26%20build%20phase/03.%20Tech%20Enablement/02.%20Key%20Outputs/02.%20Code%20architecture/NGSE%20International%20Architecture%20and%20Design/NGSE%20International%20-%20Postgres%20ER%20Diagram.png?csf=1&web=1&e=AKlhAP) |
| N/A | [NGSE International - List of Secrets](https://zoetis.sharepoint.com/:x:/r/sites/NGSEInternational/Shared%20Documents/General/04.%20Design%20%26%20build%20phase/03.%20Tech%20Enablement/02.%20Key%20Outputs/02.%20Code%20architecture/NGSE%20International%20Architecture%20and%20Design/NGSE%20International%20-%20List%20of%20Secrets.xlsx?d=w82a32137891949b1ab9879c38496f7d4&csf=1&web=1&e=6OQV9A) |

# Terms and Definitions

Common terms used in this document are defined in the Zoetis Tech & Digital Glossary. There are no additional terms and definitions specific to this document.

| Term or Acronym | Definition |
| --- | --- |
| NGSE | Next Generation Sales Engine |
| Customer DNA | Customer Data and Analytics |
| DDP | Data and Digital Platform |
| DDW | Dimensional Data Warehouse |
| ZIR | Zoetis Information Repository |
| ADLS Gen2 | Azure Data Lake Storage Gen2 |
| ADF | Azure Data Factory |
| API | Application Programming Interface |
| UI | User Interface |
| CSV | Comma Separated Values |

# Roles and Responsibilities

Roles and responsibilities for implementing the Solution using the SDLC process are as described in *ITSO-00006 Risk-based Solution Delivery Lifecycle*.

Additional and/or more detailed responsibilities relevant to this project or approved work are presented in the following Table.

| Role | Responsibility |
| --- | --- |
| Program Manager | Oversee fulfillment of NGSE requirements and co-ordinate activities between multiple project teams |
| Solution Architect | Designs hardware and software intended to solve identified business requirements for NGSE |
| Data Engineer | Design and build DNA pipeline ingesting data from sources systems to Azure Data Lake |
| Frontend Developer | Design and build NGSE web application UI |
| Backend Developer | Design and build API for NGSE web application and postprocessing data pipeline |
| Machine Learning Engineer | Design and build NGSE machine learning models |

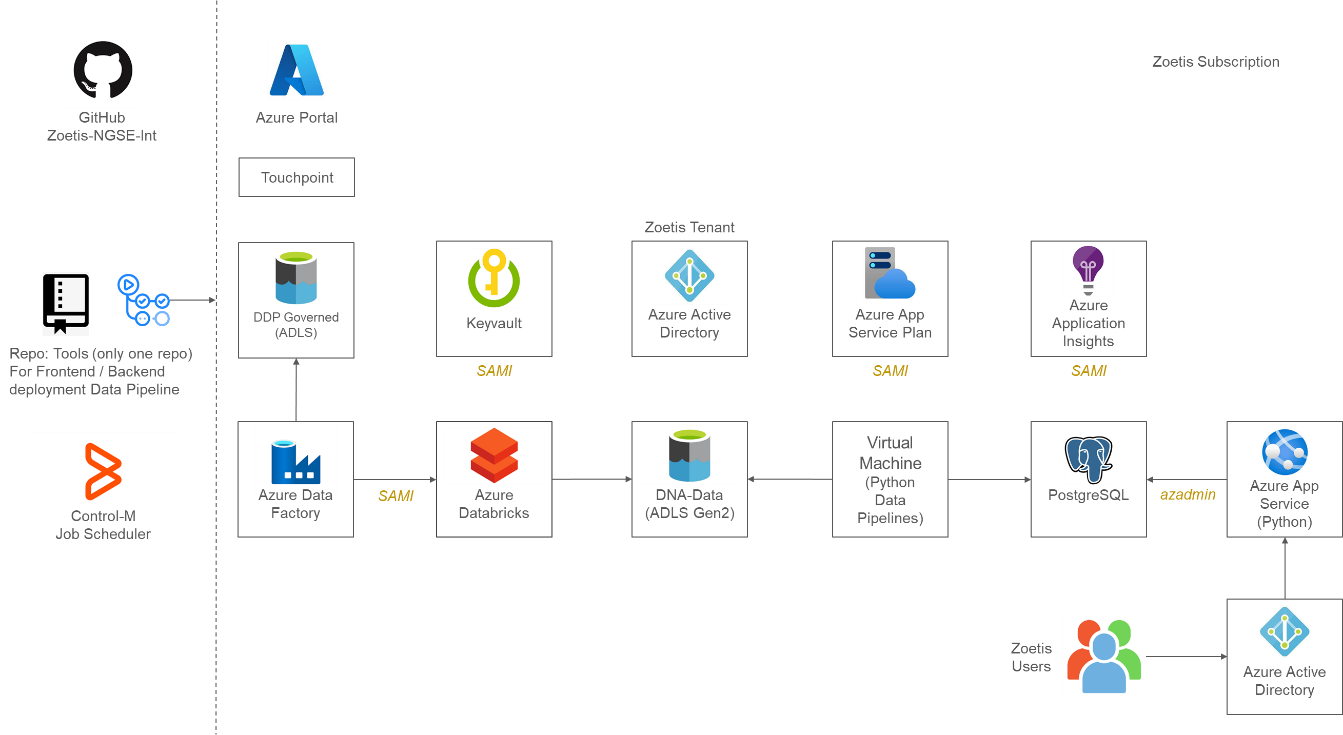
# Architecture

## Solution Architecture

### Solution Description

The solution for NGSE International involves data flows and transformations from DDP Transformed/Governed to Customer DNA, and Customer DNA to Backend Database. The transformations are implemented by a combination of Azure Data Factory (ADF) Pipelines, Azure Databricks Notebooks, Django Data Pipelines and Azure App Service (Python). Data is stored in ADLS Gen2 (DDP) and PostgreSQL.

### Solution Architecture Diagram



Refer section 6.3.5 for details on data flow and processing

### Solution Architecture Diagram Details

Key components of the solution by environment are listed below:

| Resources | Dev | QA | Prod |
| --- | --- | --- | --- |
| Subscription | Zoetis DEVELOPMENT | Zoetis QA | Zoetis PRODUCTION |
| Resource Group | dev-eastus2-ddp-rg dev-eastus2-ngsein-rg | qa-eastus2-ddp-rg qa-eastus2-ngsein-rg | prod-eastus2-ddp-rg prod-eastus2-ngsein-rg |
| Storage Account | azrdevngsein01adls  azrdlngseinvm01diag | azrqangsein01adls  azrqlngseinvm01diag | azrprodngsein01adls  azrplngseinvm01diag |
| PostgreSQL | postgresql-ngsein-dev01-eastus2 | qa-eastus2-ngsein-postgres-01 | prod-eastus2-ngsein-postgres-01 |
| Azure Databricks | devdatabricks01 | qa-eastus2-databricks-dbws | prod-eastus2-databricks-dbws |
| Azure Data Factory | dev-eastus2-ngsein-df6wxx | qa-eastus2-ngsein-dfc4bk | prod-eastus2-ngsein-dfu6om |
| Azure App Service Plan | dev-eastus2-ngsein-lin-plan | qa-eastus2-ngsein-lin-plan | prod-eastus2-ngsein-lin-plan |
| Azure App Service | dev-eastus2-ngsein-lin-apptct4 dev-eastus2-ngsein-lin-appkvxe | qa-eastus2-ngsein-lin-appnaxd | prod-eastus2-ngsein-lin-appzzbe |
| Application Insights | dev-eastus2-ngsein-insightsdlza | qa-eastus2-ngsein-insightsgcen | prod-eastus2-ngsein-insightssof1 |
| Virtual Machine | [azrdlngseinvm01](https://portal.azure.com/) | azrqlngseinvm01 | azrplngseinvm01 |
| Key Vault | deveastus2nginkv | qaeastus2ngseinkv | prodeastus2ngseinkv |

Interactions among the components are mentioned below:

1. Control-M executes Azure Data Factory and Python data pipelines.
2. Azure Data Factory pipeline executes Azure Databricks notebook “Driver” for data transformation according to list of tables in parameters file.
3. Transformation scripts read data from Transformed Zone.
4. Processed data is copied to Customer DNA tables in Governed Zone.
5. Machine Learning scripts read from, transform and write back enhanced data into Customer DNA tables.
6. Python Data Pipelines read data from Customer DNA tables and loads into PostgreSQL database.
7. Azure App Service (Django – API and React – UI) reads data from PostgreSQL and shows data in NGSE web application.

### Alternative Solution Architecture Considered (Optional)

Not applicable

### Evolutionary Considerations

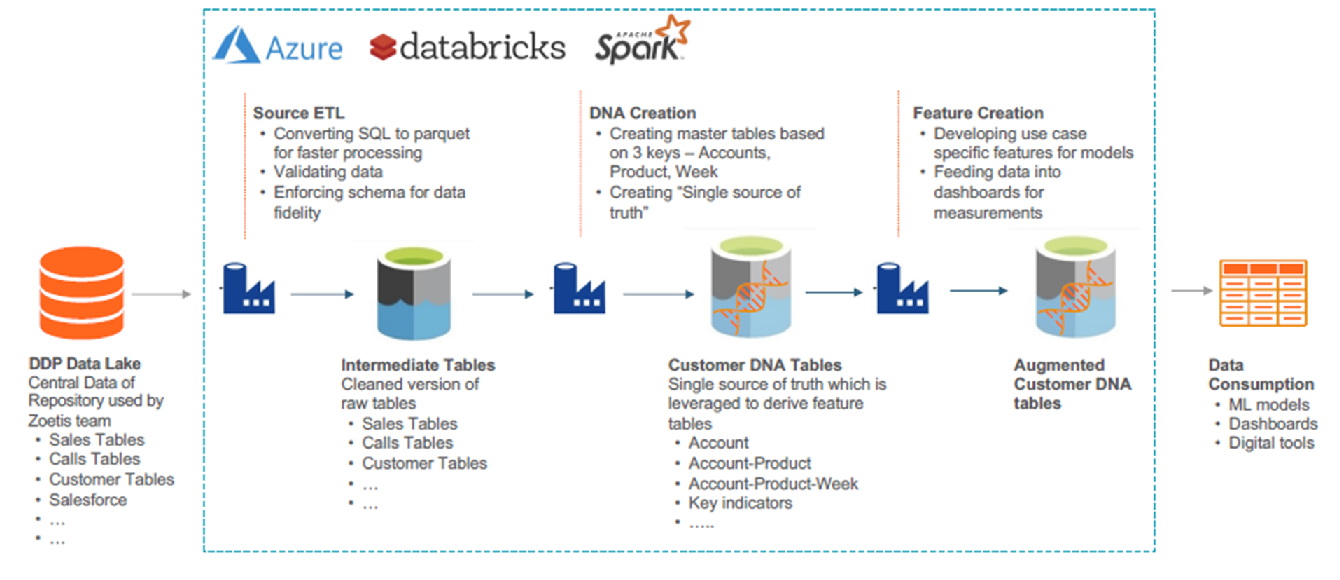
Not applicable

### Retirement Considerations

Not applicable

## Application Architecture

### Application Architecture Diagram



### Application Components

#### Web Application (React Application and Django API)

#### Database Layer (PostgreSQL)

#### Postprocessing Data Pipeline (Python)

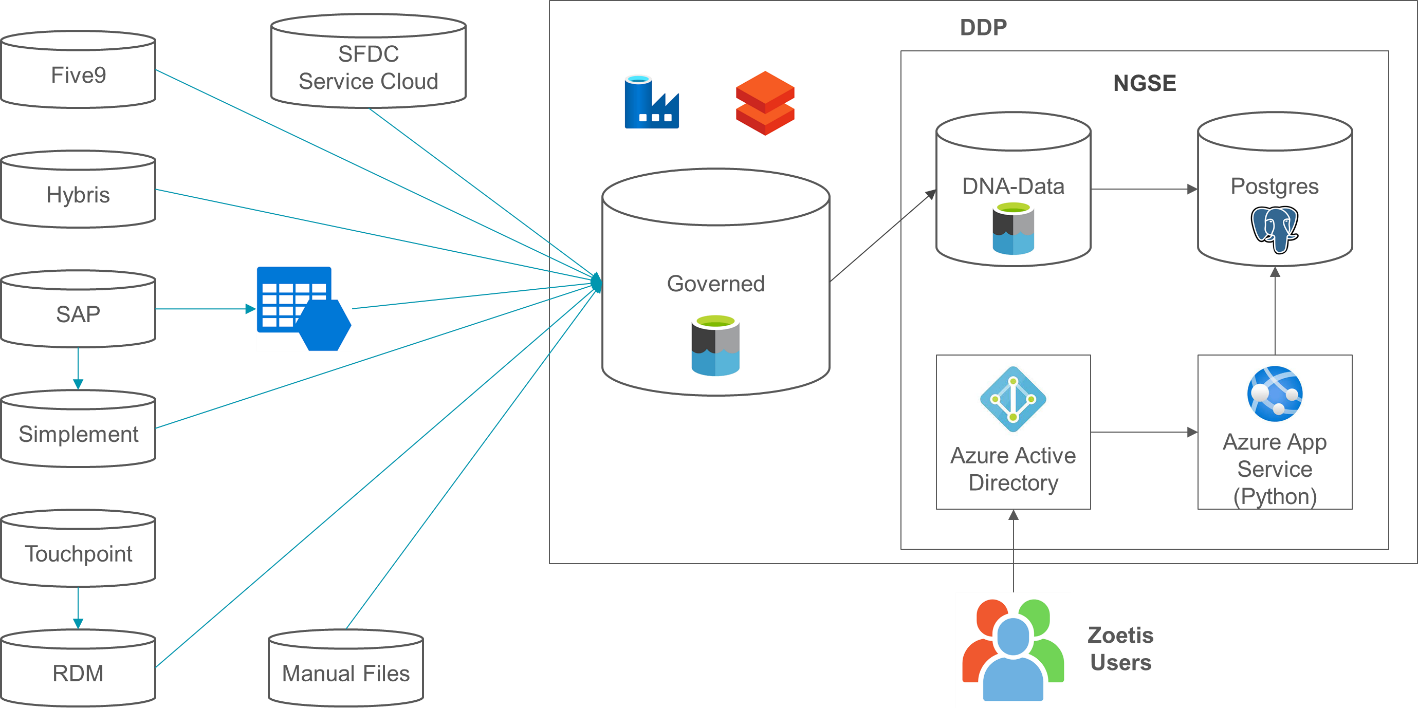
#### DNA Data Layer (Databricks, ADLS Gen2)

### Application Interface Details (Including one-time data migration)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Interface Name** | **Source** | **Target** | **Interface Details**  **(Type, Format, Freq, etc.)** | **Data & Data Processing Details** |
| DDP | DDP | NGSE DNA | Type: Batch Format: Delta, CSV, Database Views Frequency: Daily twice | Dimension and Fact data |
| Touchpoint | Touchpoint | NGSE DNA | Type: Batch  Format: Delta files Frequency: Daily twice | Fact data |

## Data Architecture

### Data Flow Diagram



### Key Data Entities/Objects and their Data Sources (DNA-Data)

| **Data Entity/Domain** | **Data Source** | **Comments** |
| --- | --- | --- |
| Account | Delta tables | Master Data |
| Product | Delta tables | Master Data |
| Account Product | Delta tables | Master Data |
| Account Product Discounts | Delta tables | Fact Data |
| Account Campaign | Delta tables | Fact Data |
| Account Day Sales | Delta tables | Fact Data |
| Account Day Calls | Delta tables | Fact Data |
| Rep Account | Delta tables | Master Data |
| Account Product Day | Delta tables | Master Data |
| Product Day | Delta tables | Master Data |
| Account Contact | Delta tables | Master Data |
| Exchange Rate | Delta tables | Fact Data |
| Rep | Delta tables | Fact Data |
| Account Potential | Delta tables | Fact Data |
| Account Day Cases | Delta tables | Fact Data |
| Account Nearby | Delta tables | Fact Data |
| Account Email Tracking | Delta tables | Fact Data |
| Zoetis Rewards Price | Delta tables | Fact Data |
| Account Contract | Delta tables | Fact Data |
| Account Product Backlist | Delta tables | Fact Data |
| Account Contract Rx | Delta tables | Fact Data |
| Product Line Chart | Delta tables | Fact Data |
| Survey | Delta tables | Fact Data |
| Rep Performance | Delta tables | Fact Data |

### ER Diagram (where relevant)

Refer to [IT-26843 NGSE International – Postgres ER Diagram](https://zoetis-qualitydocs.veevavault.com/ui/#doc_info/1216399/1/0=&bi=978979&bma=0&bmi=1&nk=1679332308656%3A1240730825&pi=978979&pma=0&pmi=1&idx=&pt=bdr&sm=&tvsl=JnY9b3V0bGluZVZpZXcmc3A9MTY1NDEyMDQ0NzU4MiZzPTE2NTQxMjA0NDc1ODImYmk9OTc4OTc5JmJtYT0wJmJtaT0x)

### CRUD Matrix

| **Data Object/Function** | **Application System** | **Create** | **Read** | **Update** | **Delete** |
| --- | --- | --- | --- | --- | --- |
| product\_line\_chart | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| coach\_dash | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| request\_history | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| account\_email\_tracking | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| account\_nba\_ranking | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| account\_nbas | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| contract\_content | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| upsell | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| cross\_sell | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| rewards\_contract\_tracker | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| new\_purchaser | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| defection\_risk | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| account\_contract | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| account\_days\_calls | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| account\_day\_cases | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| account\_days\_sales | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| account\_campaign | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| account\_product | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| account | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| product | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| rep\_details | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| rep\_account | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| account\_brand | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| ngse\_account | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| rewards\_price | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |
| account\_nearby | PostgreSQL | Checkmark with solid fill | Checkmark with solid fill | Close with solid fill | Checkmark with solid fill |

### Data Processing

Data is stored in Azure Storage (ADLS Gen2) in following containers.

| Container | Description |
| --- | --- |
| raw | Contains all the raw data from source without any cleanup or transformation |
| transform | Contains transformed data. No business logic applied. Basic cleanup and conversion to delta/parquet format |
| governed | Contains gold standard dimension and fact data from DDW and ZIR (to be replaced by DDP in future state) in delta/parquet and csv formats respectively |
| DNA-Data | Transformed gold standard data for business consumption |

Data from Sources is moved into Storage Account to “raw” and then to corresponding subfolder for each data source within “transform” folder. Data from Azure SQL Datawarehouse (DDW and ZIR) is moved to corresponding subfolder for each data source within “governed” folder.

Once the data is available in datalake storage, an ADF pipeline executes a set of ADF data flows, Databricks Notebooks, PySpark, and Python Machine Learning scripts first to load data into intermediate tables and then to transform and load data into Customer DNA tables. Subsequently Python Data Pipelines move data to destination in PostgreSQL Database.

#### Data Cleansing and Standardization

Not applicable

#### Data Deduplication and Consolidation

Not applicable

#### Data Transformation

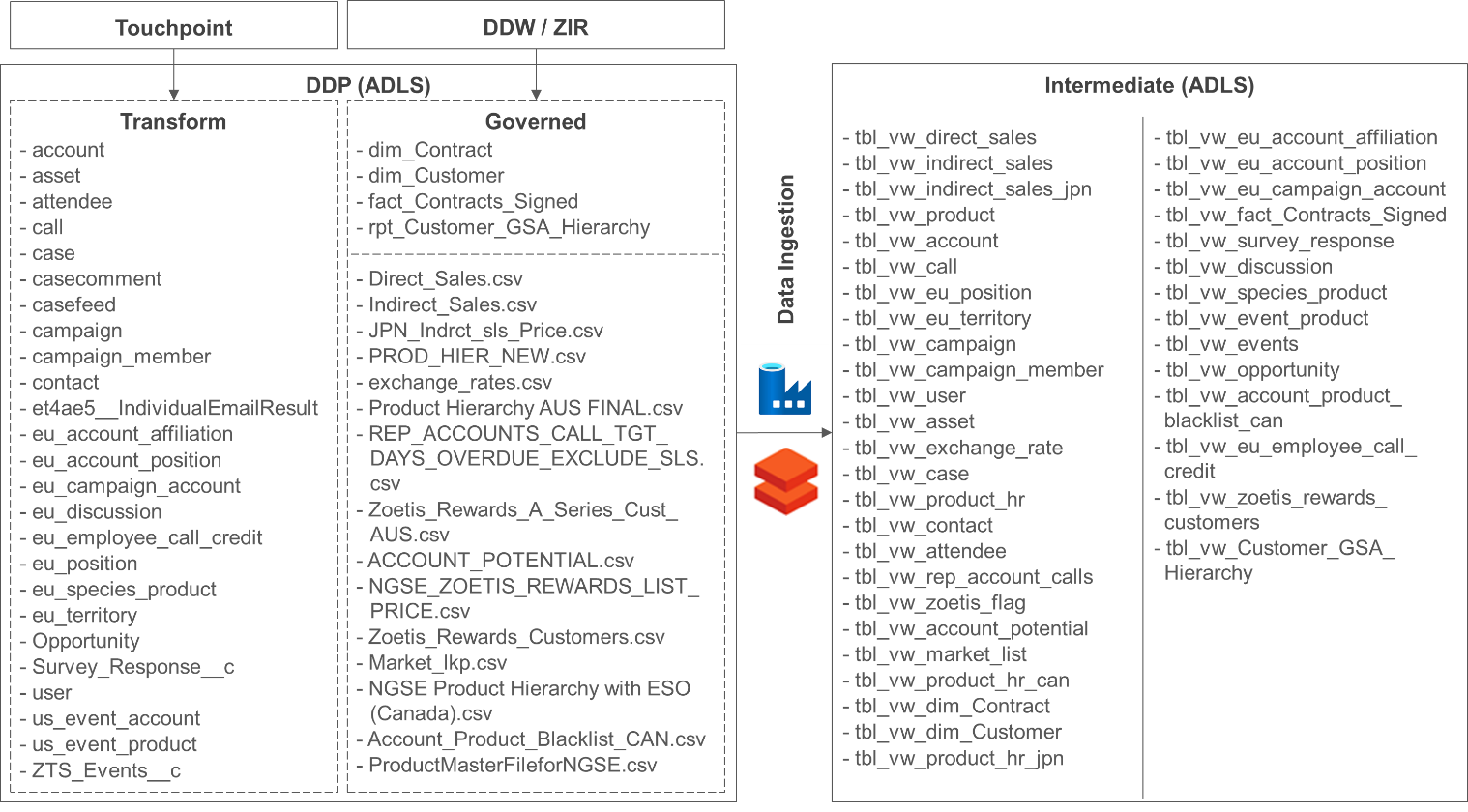
Refer to Source to Target Mappings:

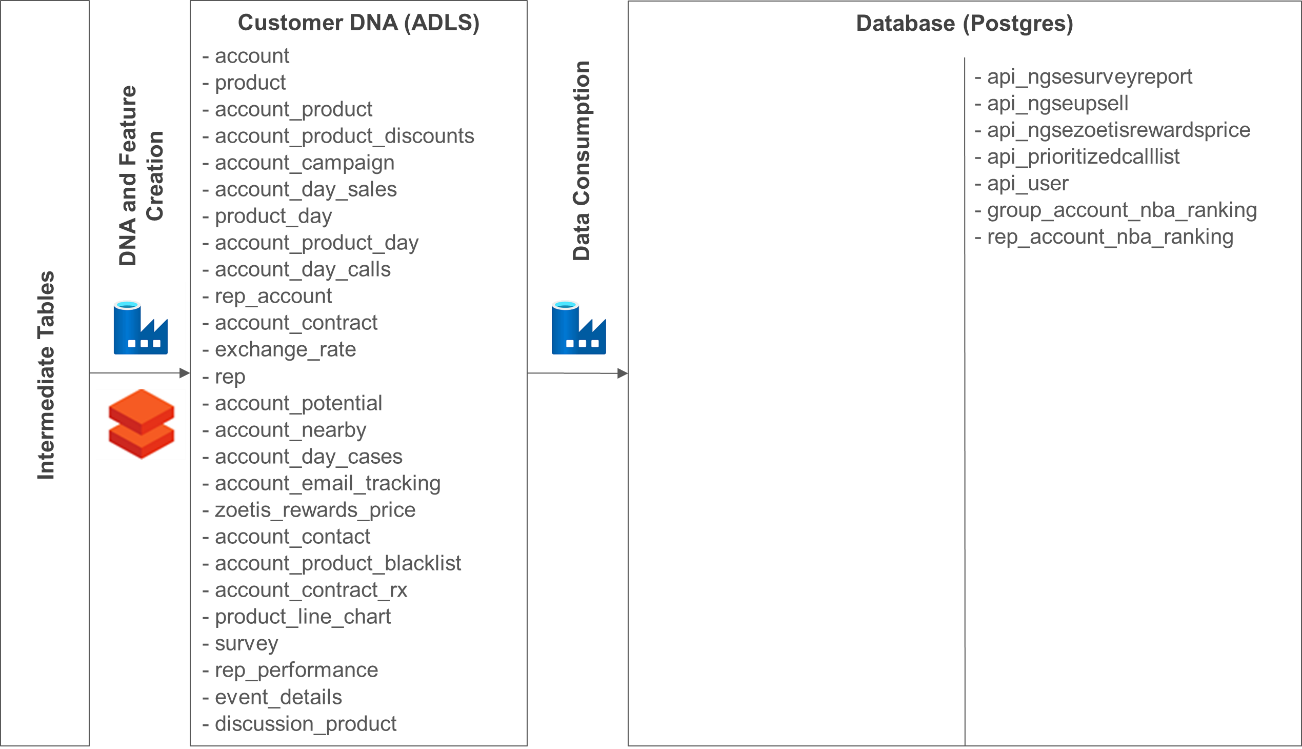
[IT-26840 NGSE International - Customer DNA Data Dictionary](https://zoetis-qualitydocs.veevavault.com/ui/#doc_info/1216393/1/0=&bi=978979&bma=0&bmi=1&nk=1679332198147%3A-1059578149&pi=978979&pma=0&pmi=1&idx=&pt=bdr&sm=&tvsl=JnY9b3V0bGluZVZpZXcmc3A9MTY1NDEyMDQ0NzU4MiZzPTE2NTQxMjA0NDc1ODImYmk9OTc4OTc5JmJtYT0wJmJtaT0x)

[IT-26841 NGSE International - Customer DNA ML Data Dictionary](https://zoetis-qualitydocs.veevavault.com/ui/#doc_info/1216396/1/0=&bi=978979&bma=0&bmi=1&nk=1679332234662%3A-2056907310&pi=978979&pma=0&pmi=1&idx=&pt=bdr&sm=&tvsl=JnY9b3V0bGluZVZpZXcmc3A9MTY1NDEyMDQ0NzU4MiZzPTE2NTQxMjA0NDc1ODImYmk9OTc4OTc5JmJtYT0wJmJtaT0x)

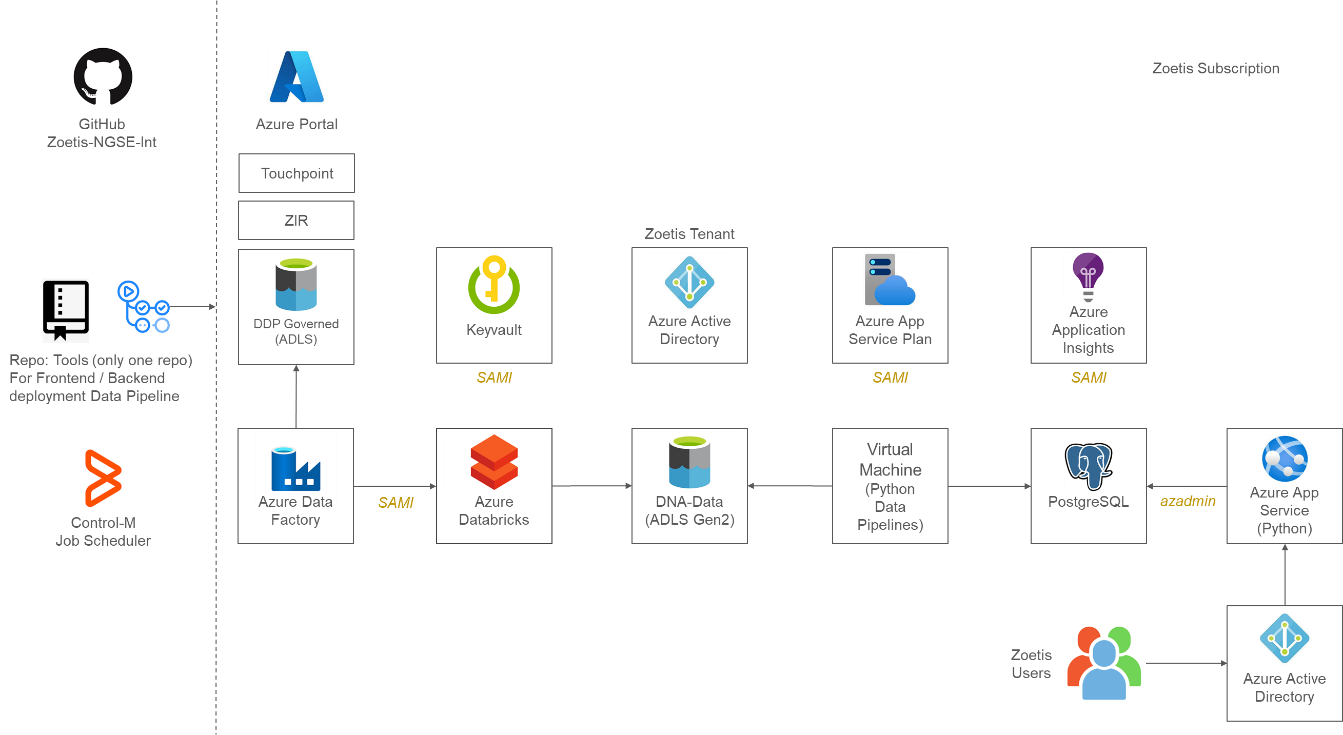
[IT-26842 NGSE International - Postgres Data Dictionary](https://zoetis-qualitydocs.veevavault.com/ui/#doc_info/1216397/1/0=&bi=978979&bma=0&bmi=1&nk=1679332267644%3A-957691062&pi=978979&pma=0&pmi=1&idx=&pt=bdr&sm=&tvsl=JnY9b3V0bGluZVZpZXcmc3A9MTY1NDEyMDQ0NzU4MiZzPTE2NTQxMjA0NDc1ODImYmk9OTc4OTc5JmJtYT0wJmJtaT0x)

The following diagrams depict Data processing and transformation flow starting from DDP to Postgres.





# Technical Architecture (Mandatory)



## Technology Standards Mapping

|  |  |
| --- | --- |
| **Component** | **Technology** |
| User Interface | React (JavaScript) |
| Application Layer/s | Django (Python) |
| Communication/Messaging Layer | REST API |
| ETL | Databricks, Data Factory, Python Data Pipeline |
| Data Quality / Data Consolidation | Databricks, Data Factory |
| Master Data Management | Not Applicable |
| Database Repository | PostgreSQL, ADLS Gen2 |
| Hardware / OS / Virtualization | Azure Virtual Machine |
| Security and Access Management | Azure Active Directory |
| Job Scheduling | Control-M |
| Reporting & Analytics | Power BI |
| Dashboards / Data Visualization | React (JavaScript) |
| Performance and Scalability | Refer Section 7.3 |
| Disaster Recovery / High Availability | Refer Section 7.4 |

## Physical Architecture

Not Applicable

## Performance and Scalability

Since all resources / services belong to Azure Topology, they are scalable to match demand to accommodate workload.

## Disaster Recover and High Availability

Since all resources / services belong to Azure Topology, the disaster recovery / high availability is supported by Microsoft.

# Revision History

| Version | Author | Date | Revisions |
| --- | --- | --- | --- |
| 1.0 | Ranjan Das | 21 Mar 2023 | Initial Version |