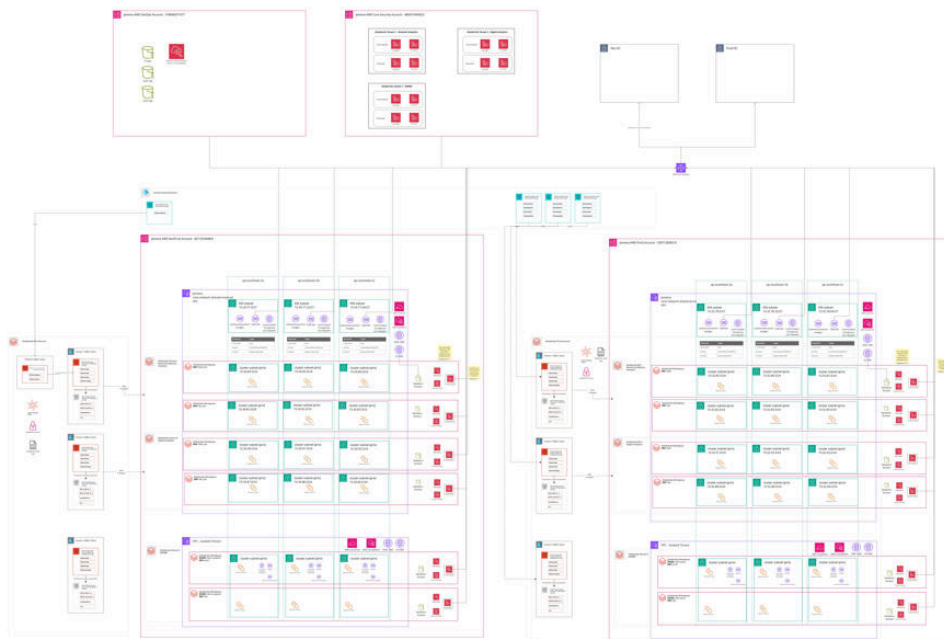


# Infrastructure Design

- 1 [E2E AWS Design](#)
- 2 [Design Features](#)
  - 2.1 [AWS Accounts](#)
  - 2.2 [VPC & Subnets](#)
    - 2.2.1 [Provisioned Dev/Nonprod Subnets Summary](#)
    - 2.2.2 [Provisioned Prod Subnets Summary](#)
- 3 [Control Plane & Data Plane Connectivity](#)
- 4 [Discovery FAQ Minutes](#)
- 5 [Additional Databricks Discovery Qs](#)

## E2E AWS Design

The end-to-end view AWS Design is below:



e2e Solution Design of Databricks Platform

## Design Features

### AWS Accounts

- Non-production stack (LHS of diagram above) to be deployed to `app-datahub-nonprod` (AWS account id 851725449831)

- Production stack (RHS of diagram above) to be deployed to `app-datahub-prod` (AWS account id 339712836516)

#### VPC & Subnets

⚠ **IMPORTANT:** Jemena's firewall must whitelist

`host=mdnrak3rme5y1c.c5f38tyb1fdu.ap-southeast-2.rds.amazonaws.com` on port 3306 to allow cluster connectivity to the RDS metastore. This must be done in NPD and PRD environments.

Contact: @David Hunter

- Use Jemena's `core-network-shared-prod` and `core-network-shared-nonprod` VPC; [Shared VPC](#)
  - [Shared VPC | Prod Subnets](#)
  - [Shared VPC | Nonprod Subnet Allocation Table](#)
- 2 workspaces per Business Unit: `lab` and `field`. Data product teams within each BU are expected to develop data products in `lab` workspace and promote to `field` workspace for automated pipelines. (previous version of diagram had `field` named as `prod`; which was a source of confusion)
- Isolated /24 subnet per databricks workspace
- 3 AZs per workspace
- This means  $256 * 3 \text{ AZ} * 2 \text{ ENV} = 1536$  private IP addresses used, or a footprint of 2.3% on Jemena's shared VPC, per Business Unit deployed. Please refer to [Unity Catalog Design](#) for further details on business units and workspaces.
- We expect to eventually have 5 distinct BUs on the platform; this would be an eventual footprint of ~12% on Jemena's shared VPC.
  - Electricity
  - Gas
  - Digital
  - Corporate
  - \$future

#### Provisioned Dev/Nonprod Subnets Summary

Subnet Name	IPv4 CIDR	IPv4 CIDR	Subnet ID	Workplace & Tenant	Designate Use Case
-------------	-----------	-----------	-----------	--------------------	--------------------

		combined to fit within SG rule limit			
app- datahub- dev- private- az1-02	10.34.81.0/24	10.34.81.0/24	subnet- 02b6b517365 02bd3d	Databricks Tenant 1 Workspace 1	Elec- Network BU Lab Workspace
app- datahub- dev- private- az2-02	10.34.82.0/24	10.34.82.0/24	subnet- 09607bd6c72 a58127	Databricks Tenant 1 Workspace 1	Elec- Network BU Lab Workspace
app- datahub- dev- private- az3-02	10.34.83.0/24	10.34.83.0/24	subnet- 06ec4ccef21d 9c53e	Databricks Tenant 1 Workspace 1	Elec- Network BU Lab Workspace
app- datahub- dev- private- az1-03	10.34.84.0/24	10.34.84.0/22	subnet- 083f1bfcad09 0a502	Databricks Tenant 1 Workspace 2	Elec- Network BU Field Workspace
app- datahub- dev- private- az2-03	10.34.85.0/24		subnet- 01080eb5f28 4fcaac	Databricks Tenant 1 Workspace 2	Elec- Network BU Field Workspace
app- datahub-	10.34.86.0/24		subnet- 064b721b9b0	Databricks Tenant 1	Elec- Network

dev-private-az3-03			55cce2	Workspace 2	BU Field Workspace
app-datahub-dev-private-az1-04	10.34.87.0/24		subnet-039d49c832c613766	Databricks Tenant 2 Workspace 1	Digital BU Lab Workspace
app-datahub-dev-private-az2-04	10.34.88.0/24	10.34.88.0/22	subnet-04264bb265a9b477f	Databricks Tenant 2 Workspace 1	Digital BU Lab Workspace
app-datahub-dev-private-az3-04	10.34.89.0/24		subnet-06c48e0b4be6a9b1c	Databricks Tenant 2 Workspace 1	Digital BU Lab Workspace
app-datahub-dev-private-az1-05	10.34.90.0/24		subnet-011451954924d667a	Databricks Tenant 2 Workspace 2	Digital BU Field Workspace
app-datahub-dev-private-az2-05	10.34.91.0/24		subnet-057a5a1f9b2d4d983	Databricks Tenant 2 Workspace 2	Digital BU Field Workspace

app- datahub- dev- private- az3-05	10.34.92.0/24	10.34.92.0/24	subnet- 0d360b9a1fdb 82ccb	Databricks Tenant 2 Workspace 2	Digital BU Field Workspace
app- datahub- dev- private- az1-06	10.34.98.0/24	10.34.98.0/24	subnet- 02f3853e265 64ca4c	Databricks Tenant 3 Workspace 1	corporate BU Lab Workspace
app- datahub- dev- private- az2-06	10.34.99.0/24	10.34.99.0/24	subnet- 038e0256435 f135fe	Databricks Tenant 3 Workspace 1	corporate BU Lab Workspace
app- datahub- dev- private- az3-06	10.34.100.0/24	10.34.100.0/22	subnet- 0a31eb73d0d 8ccc1a	Databricks Tenant 3 Workspace 1	corporate BU Lab Workspace
app- datahub- dev- private- az1-07	10.34.101.0/24		subnet- 069a07a4ac3 4ad07a	Databricks Tenant 3 Workspace 2	corporate BU Field Workspace
app- datahub- dev-	10.34.102.0/24		subnet- 047730e6dba 90306f	Databricks Tenant 3 Workspace 2	corporate BU Field Workspace

private-az2-07					
app-datahub-dev-private-az3-07	10.34.103.0/24		subnet-07cf7f35e1c8d9328	Databricks Tenant 3 Workspace 2	corporate BU Field Workspace

Provisioned Prod Subnets Summary

Subnet Name	IPv4 CIDR	IPv4 CIDR  combined to fit within SG rule limit	Subnet ID	Workplace & Tenants	Designated Use Cases
app-datahub-prod-private-az1-02	10.32.82.0/24	10.32.82.0/24	subnet-0b0abab900ea78847	Databricks Tenant 1 Workspace 1	elec-network BU Lab Workspace
app-datahub-prod-private-az2-02	10.32.83.0/24	10.32.83.0/24	subnet-08ca9b18ea75495ee	Databricks Tenant 1 Workspace 1	elec-network BU Lab Workspace
app-datahub-prod-private-az3-02	10.32.84.0/24	10.32.84.0/22	subnet-07a0a740dd1aebaa8	Databricks Tenant 1 Workspace 1	elec-network BU Lab Workspace

app- datahub- prod- private- az1-03	10.32.85.0/24		subnet- 050ded168eb 6d186a	Databricks Tenant 1 Workspace 2	elec- network BU Field Workspace
app- datahub- prod- private- az2-03	10.32.86.0/24		subnet- 017f2292c6ff4 a849	Databricks Tenant 1 Workspace 2	elec- network BU Field Workspace
app- datahub- prod- private- az3-03	10.32.87.0/24		subnet- 04e6b65974a e7f7a4	Databricks Tenant 1 Workspace 2	elec- network BU Field Workspace
app- datahub- prod- private- az1-04	10.32.88.0/24	10.32.88.0/22	subnet- 04cb88918f7 2952c5	Databricks Tenant 2 Workspace 1	Digital BU Lab Workspace
app- datahub- prod- private- az2-04	10.32.89.0/24		subnet- 0b7f474021a0 24b9c	Databricks Tenant 2 Workspace 1	Digital BU Lab Workspace
app- datahub- prod-	10.32.90.0/24		subnet- 00faaadbf234 97fea	Databricks Tenant 2 Workspace 1	Digital BU Lab Workspace

private-az3-04					
app-datahub-prod-private-az1-05	10.32.91.0/24		subnet-036cc7c3d19a3f831	Databricks Tenant 2 Workspace 2	Digital BU Field Workspace
app-datahub-prod-private-az2-05	10.32.92.0/24	10.32.92.0/24	subnet-0f256a5ba1ade1eb6	Databricks Tenant 2 Workspace 2	Digital BU Field Workspace
app-datahub-prod-private-az3-05	10.32.93.0/24	10.32.93.0/24	subnet-023920bd0e4406926	Databricks Tenant 2 Workspace 2	Digital BU Field Workspace
app-datahub-prod-private-az1-06	10.32.98.0/24	10.32.98.0/24	subnet-000e95795e48cbefb	Databricks Tenant 3 Workspace 1	corporate BU Lab Workspace
app-datahub-prod-private-az2-06	10.32.99.0/24	10.32.99.0/24	subnet-0ea30ee71d3aa8b86	Databricks Tenant 3 Workspace 1	corporate BU Lab Workspace
app-datahub-	10.32.100.0/24	10.32.100.0/22	subnet-0e40b54cef7	Databricks Tenant 3	corporate BU Lab

prod-private-az3-06			3551c7	Workspace 1	Workspace
app-datahub-prod-private-az1-07	10.32.101.0/24		subnet-00a96ead7e16456c1	Databricks Tenant 3 Workspace 2	corporate BU Field Workspace
app-datahub-prod-private-az2-07	10.32.102.0/24		subnet-06becb4238e4d61a6	Databricks Tenant 3 Workspace 2	corporate BU Field Workspace
app-datahub-prod-private-az3-07	10.32.103.0/24		subnet-02573ce4fdbb7da89	Databricks Tenant 3 Workspace 2	corporate BU Field Workspace

## Control Plane & Data Plane Connectivity

- Front End + Back End private link enabled as per [Databricks Reference Architecture](#) and requirement provided by @Sampath Jagannathan on July 25th 2024. [AWS PrivateLink for Control Plane](#)
- Shared VPC design already have private endpoints and private gateway endpoints (S3, DDB) to AWS services for **private connectivity**.

## Discovery FAQ Minutes

✓ What is the platform connectivity option with Jemena datacenter?

We utilize **AWS Direct Connect** for connectivity between AWS and Jemena's on premises data centres. Details on routing between **prod** and **nonprod** networks is found on the

same documentation page

✓ What is the platform connectivity option to the internet (if allowed)

All traffic is subject to **SSL inspection** as it passed through the firewalls. You'll need to add our root CA certificate to any process that makes a call out to the internet.

[E2E AWS Design](#)

[Design Features](#)

[AWS Accounts](#)

[VPC & Subnets](#)

[Provisioned Dev/Nonprod Subnets Summary](#)

[Provisioned Prod Subnets Summary](#)

[Control Plane & Data Plane Connectivity](#)

[Discovery FAQ Minutes](#)

[Additional Databricks Discovery Qs](#)

## Additional Databricks Discovery Qs

✓ The following question has been either clarified in the response dropdown above or the collective response dropdown below:

1. *How many environments (dev/qa/prod) do we need to provision ?*
2. *Do workspaces need to be mapped one to one to AWS accounts ?*
3. *Databricks Workspace design - options are single set of workspaces (dev/qa/prod) with multiple tenants or isolation by line of business (LOB) design as mentioned [here](#) which is dev/qa/prod X LOB (tenants).*
4. *Do we need to have network isolation between tenants ?*
5. *How isolated you want your data to be ? (S3 related)*
6. *Is there connectivity between aws non-prod and prod accounts to all Jemena On-prem data centers (dev and prod)?*
7. *Is there network isolation between data centers ?*
8. *Would you like a lower environment to test platform changes such as account Metastore, Catalog & Schema Permission (RBAC), Cluster policies or any Databricks workspace resources etc ? This will determine how many Databricks account do we need to provision.*

✓ Collective Responses

### Collective Responses - Chat with David Hunter and Sampath, 25th Jul 2024

- *Digital and Network Analytics tenants can share a VPC. They will have their own Databricks workspaces.*
- *David's team will be provisioning the AWS VPCs. We will be given the VPC configuration to parse to our Databricks modules*

- *We provide David with a list of VPC requirements for Databricks, which will then be created for us*
- *Jemena are open to best-practice advice from us on how to best deploy Databricks. Their process can be modified to adhere with our recommendations*

## **ACTIONS**

Create a sample architecture to show what multiple tenants on multiple vpcs will look like for Sampath and David Hunter – Separate workspaces at the VPC level (done)