Azure Databricks Security Architecture

databricks

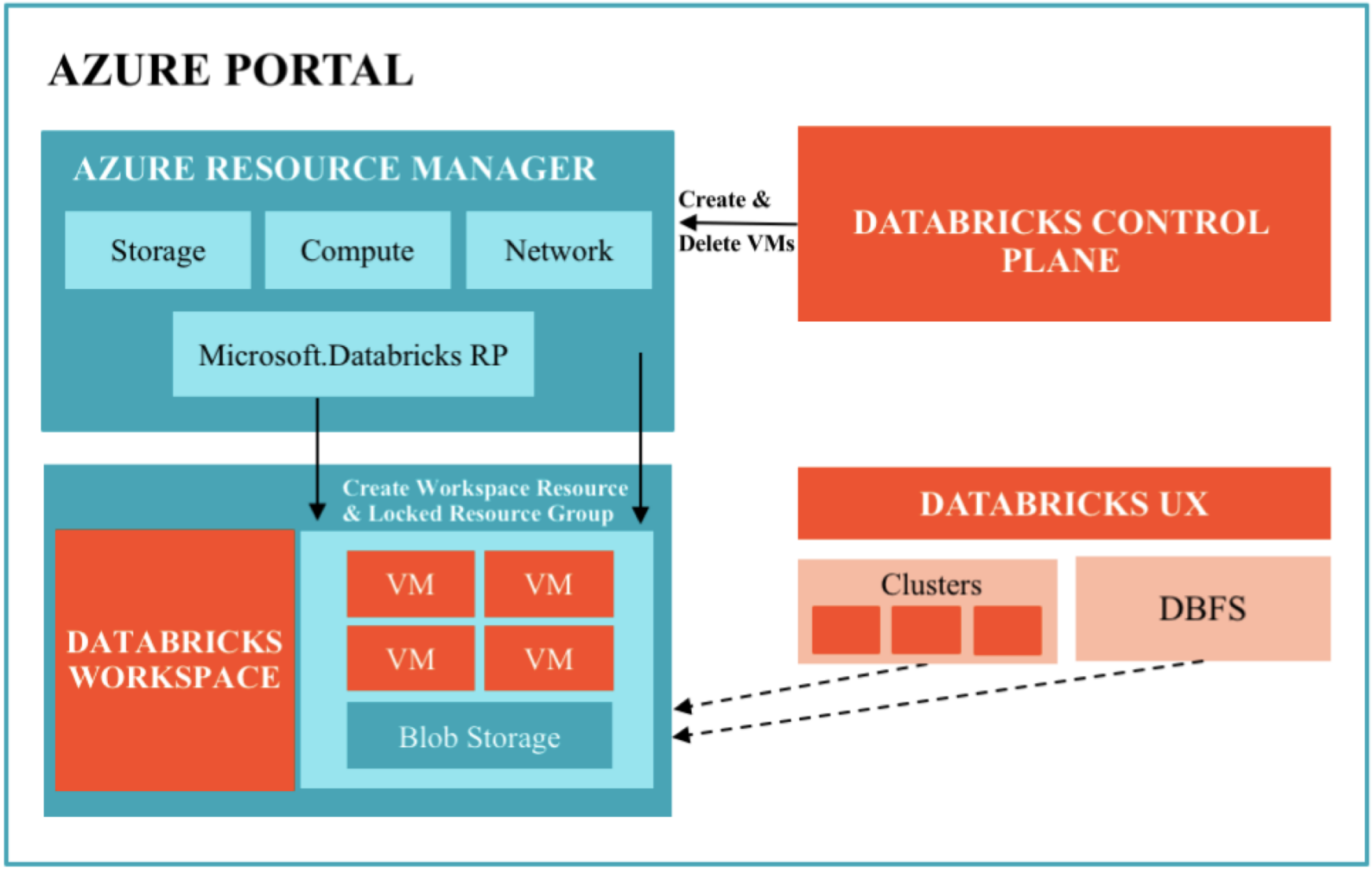
Azure Security Best Practices

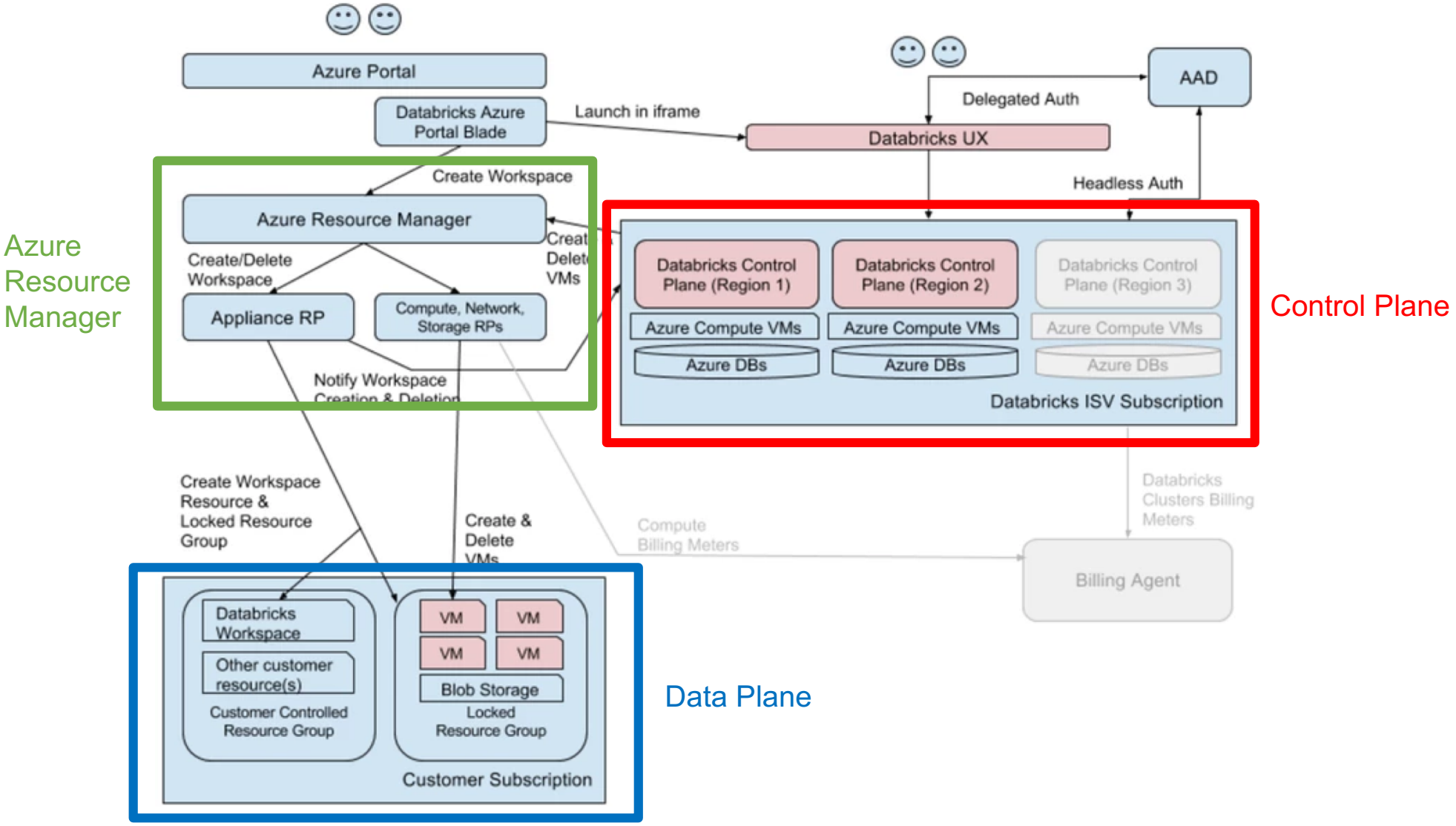
Databricks designed the security infrastructure and configurations based on Azure security best practices including but not limited to the following:

- Azure Management Console access is controlled with IdP (e.g. OKTA) SSO with Multifactor-Authentication.
- Data stores are encrypted at rest using native Azure storage encryption.
- Customers are isolated using managed resource groups and VNETs.

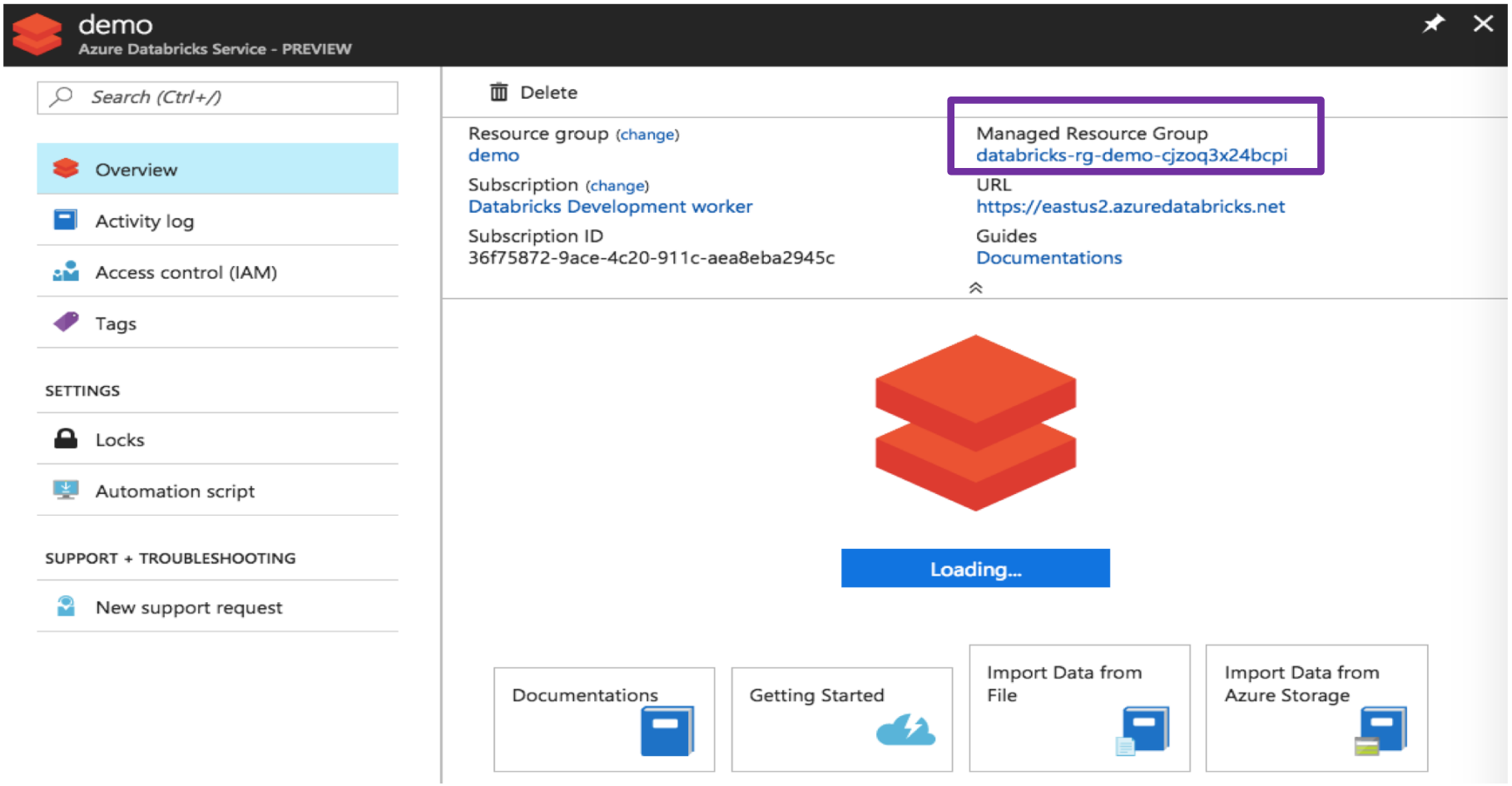
Architecture

- Azure Databricks is a multi-tenant deployment.
- Upon sign-up, the Databricks launches a Control Plane in Databricks' Azure account and the Data Plane will be launched in Customers' Azure account.
- The Control plane and the Data plane are be connected via vnet peering (VNET peering targeted for GA).

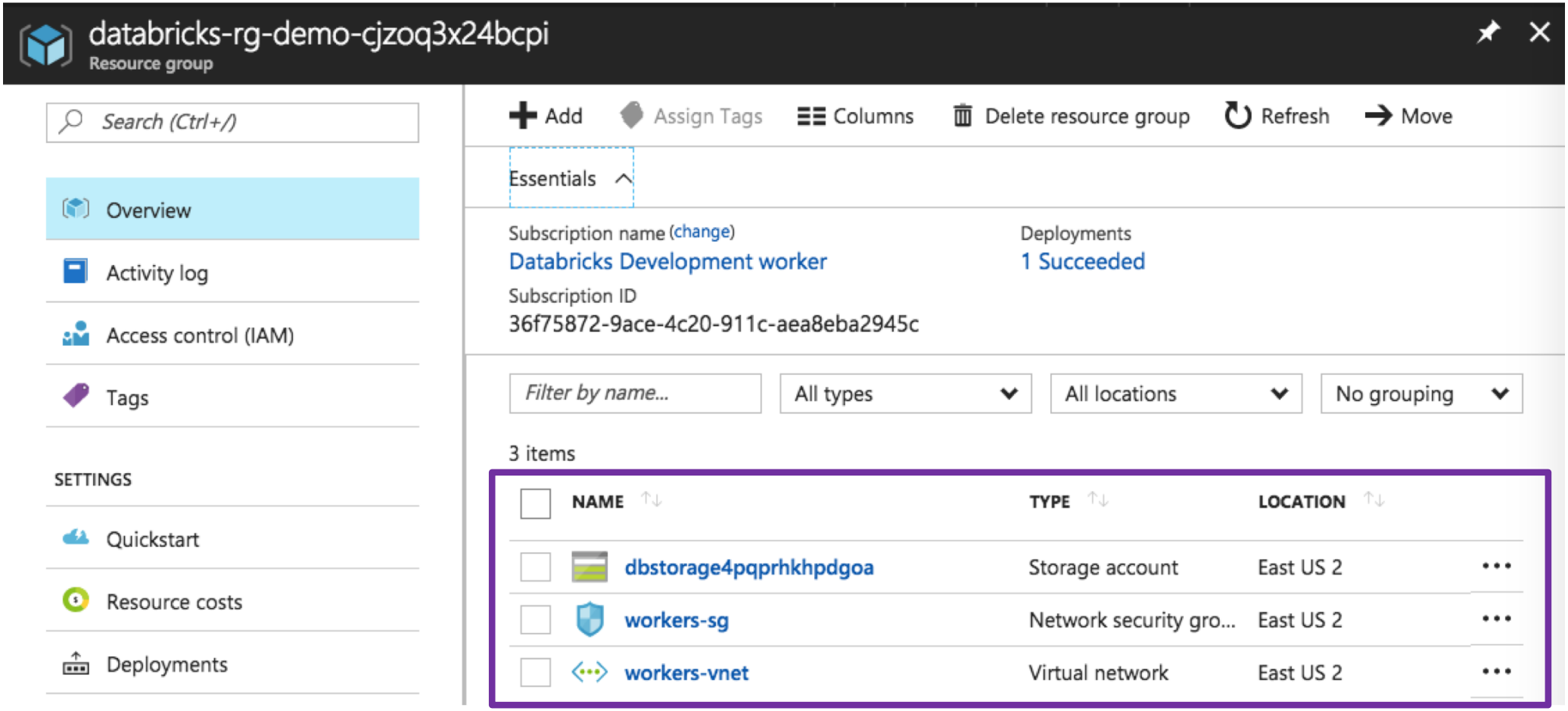

AZURE PORTAL
AZURE RESOURCE MANAGER
Storage
Compute
Network
Microsoft.Databricks RP
Create &
Delete VMs
DATABRICKS CONTROL PLANE
Create Workspace Resource & Locked Resource Group
DATABRICKS WORKSPACE
VM
VM
VM
VM
Blob Storage
DATABRICKS UX
Clusters
DBFS


Azure Portal
Databricks Azure Portal Blade
Launch in iframe
Databricks UX
Delegated Auth
AAD
Headless Auth
Create Workspace
Azure Resource Manager
Create/Delete Workspace
Appliance RP
Compute, Network, Storage RPs
Notify Workspace
Azure Resource Manager
Databricks Control Plane (Region 1)
Databricks Control Plane (Region 2)
Databricks Control Plane (Region 3)
Azure Compute VMs
Azure DBs
Databricks ISV Subscription
Control Plane
Create Workspace Resource & Locked Resource Group
Create & Delete VMs
Compute Billing Meters
Databricks Clusters Billing Meters
Billing Agent
Databricks Workspace
Other customer resource(s)
Customer Controlled Resource Group
VM
Blob Storage
Locked Resource Group
Customer Subscription
Data Plane

Azure Databricks Architecture



Azure Databricks Architecture



Control Plane

- Notebooks, jobs, clusters, users, and ACLs are managed by control plane services.
- These services store their data in dedicated databases in Databricks' Azure subscription.
- Access to Control Plane VNET is limited to Databricks Shared Services through security groups. These services are not Internet facing, and access is only provided via a proxy server.
- The web application UI and API is accessible via the Internet.
- Access to Cluster Manager is restricted to the webapp via security groups.

Data Plane

Spark Clusters are deployed in a customer Azure subscription. Each workspace and it's associated clusters reside within a dedicated VNET, which separates it from other workspaces.

- The customer deployment is isolated at a VNET level. In a VNET, public IPs are assigned to nodes, but access is restricted via Azure security groups.
- Workspaces are launched within managed security groups.
- Managed groups allow connections from the control plane and allow workers to communicate with each other.

Databricks File System (DBFS)

- The Databricks File System or DBFS is a distributed file system that comes installed on Spark Clusters in Databricks. It is a layer over Azure Blob Storage, which allows customers to Mount containers to make them available to users in your workspace.

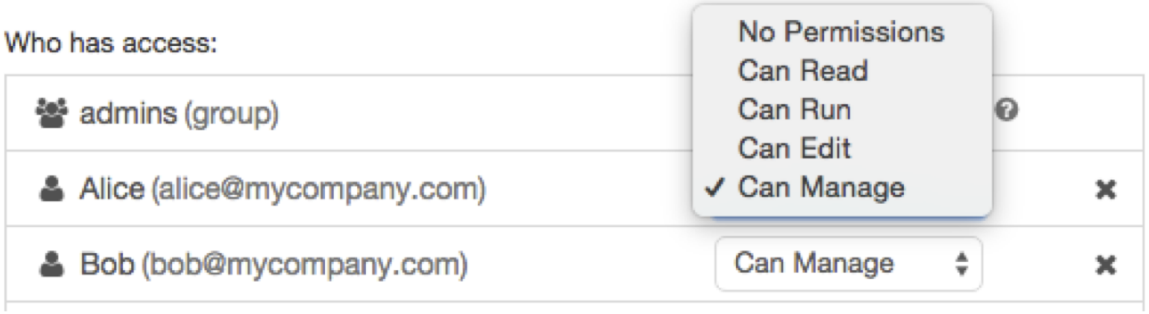
Application Security

Authentication

- Databricks authentication is delegated to Azure Active Directory (AAD).

Access Control and Permissions

Databricks has powerful permissions settings for stricter control over what users can perform what actions. This includes access controls for clusters and workspaces. These controls are quintessential for larger organizations with many users; Databricks makes managing permissions easy.



Cluster Access Control

- Using this feature, control which users can:
  - Attach notebooks to clusters
  - Terminate clusters
  - Start clusters
  - Restart clusters
  - Resize clusters
  - Modify permissions

Detailed doc here

Workspace Access Control

For each Notebook, control which users can:

- View Cells
- Comment
- Run Commands
- Attach or Detach notebooks
- Edit cells
- Change Permissions

- For each Folder, control which users can:
  - Create or Delete items
  - Move or Rename Items
  - Change Permissions

Detailed doc here

Abilities	No Permissions	Read	Run	Edit	Manage
View cells		✅	✅	✅	✅
Comment		✅	✅	✅	✅
Run Commands			✅	✅	✅
Attach/detach notebooks			✅	✅	✅
Edit cells				✅	✅
Change permissions					✅

Abilities	No Permissions	Read	Run	Edit	Manage
Create items					✅
Delete items					✅
Move/rename items					✅
Change permissions					✅

Jobs Access Control

For each scheduled job, control which users can:

- View job details and settings
- View results, Spark UI, logs of a job run
- Run now
- Cancel run
- Edit job settings
- Modify permissions
- Delete job
- Change owner

Detailed doc here

Jobs Access Control

Abilities	No Permissions	Can View	Can Manage Run	Is Owner	Can Manage (admin)
View job details and settings	x	x	x	x	x
View results, Spark UI, logs of a job run		x	x	x	x
Run now			x	x	x
Cancel run			x	x	x
Edit job settings				x	x
Modify permissions				x	x
Delete job				x	x
Change owner					x

Detailed doc here

Tables Access Control

Databricks supports fine-grained access control via the Spark SQL interface. In this context, access can be restricted for any securable objects, e.g., tables, views, databases or functions.

Fine-grained level access control (i.e. on rows or columns matching specific conditions) can be accomplished via access control on derived views that can contain arbitrary queries. These access control policies are enforced by the Databricks SQL query analyzer at runtime.

Detailed doc here

Structured Data (Tables) Access Control

Privileges

- `SELECT` privilege – gives read access to an object.
- `CREATE` privilege – gives ability to create an object (e.g., a table in a database).
- `MODIFY` privilege – gives ability to add/delete/modify data to/from an object (e.g., a table).
- `READ_METADATA` privilege – gives ability to view an object and its metadata.
- `CREATE_NAMED_FUNCTION` privilege – gives ability to create a named UDF in an existing catalog or database.
- `ALL PRIVILEGES` – gives all privileges (gets translated into all the above privileges).

Objects

These privileges apply to the following class of objects:

- `CATALOG` - controls access to the entire data catalog.
- `DATABASE` - controls access to a database.
- `TABLE` - controls access to a managed or external table.
- `VIEW` - controls access to SQL views.
- `FUNCTION` - controls access to a named function.
- `ANONYMOUS FUNCTION` - controls access to anonymous or temporary functions.
- `ANY FILE` - controls access to the underlying filesystem.

Security Features Demo


Thank you
databricks