

Create Deployment Architecture

Capstone Step 7

Decision Making Process

This process was fairly straightforward, as I now have experience using each of the tool required.

1. I did have to decide between using **Apache Airflow** vs **Azure Data Factory** for orchestration of ETL pipeline tasks

I chose **Azure Data Factory** because I knew it would ease any pain encountered when authenticating connections between services

2. I successfully imported and queried the data using both:

Databricks using a Pyspark kernel

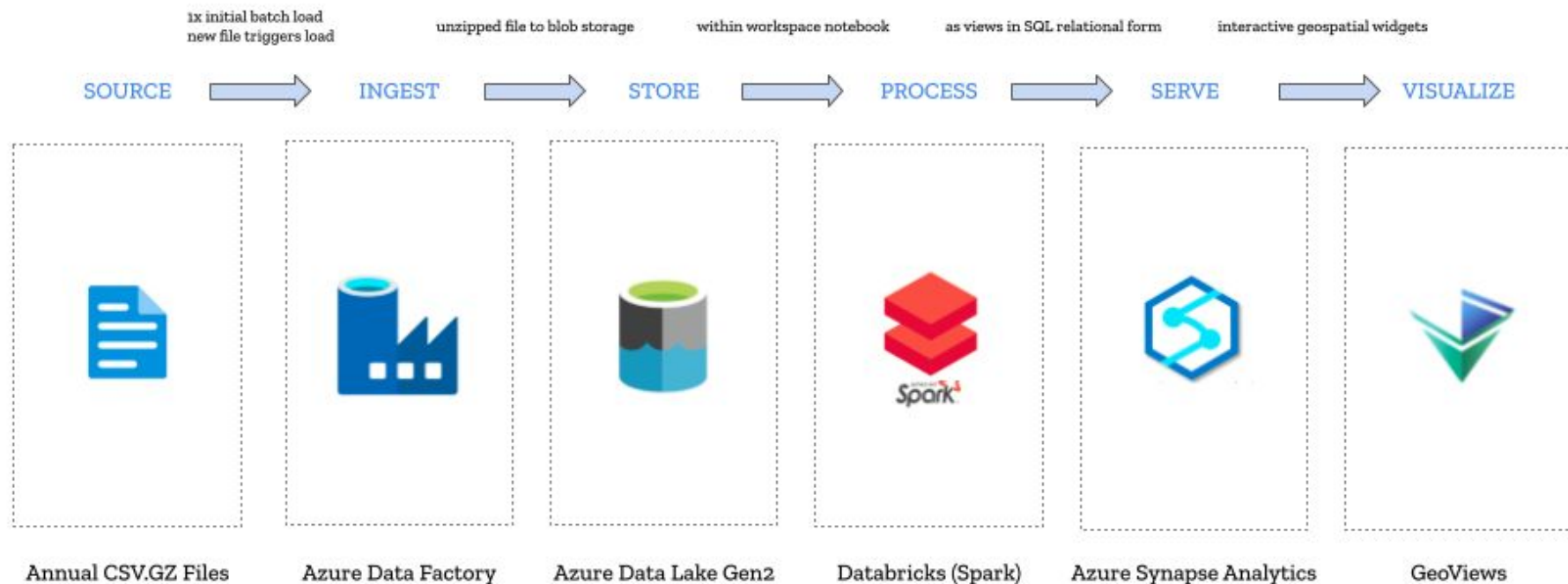
Azure Data Studio using a SQL kernel

The speeds were roughly the same, but VM core quota issues with **Databricks** frustrated me, and I chose to serve the data with **Azure Data Studio**

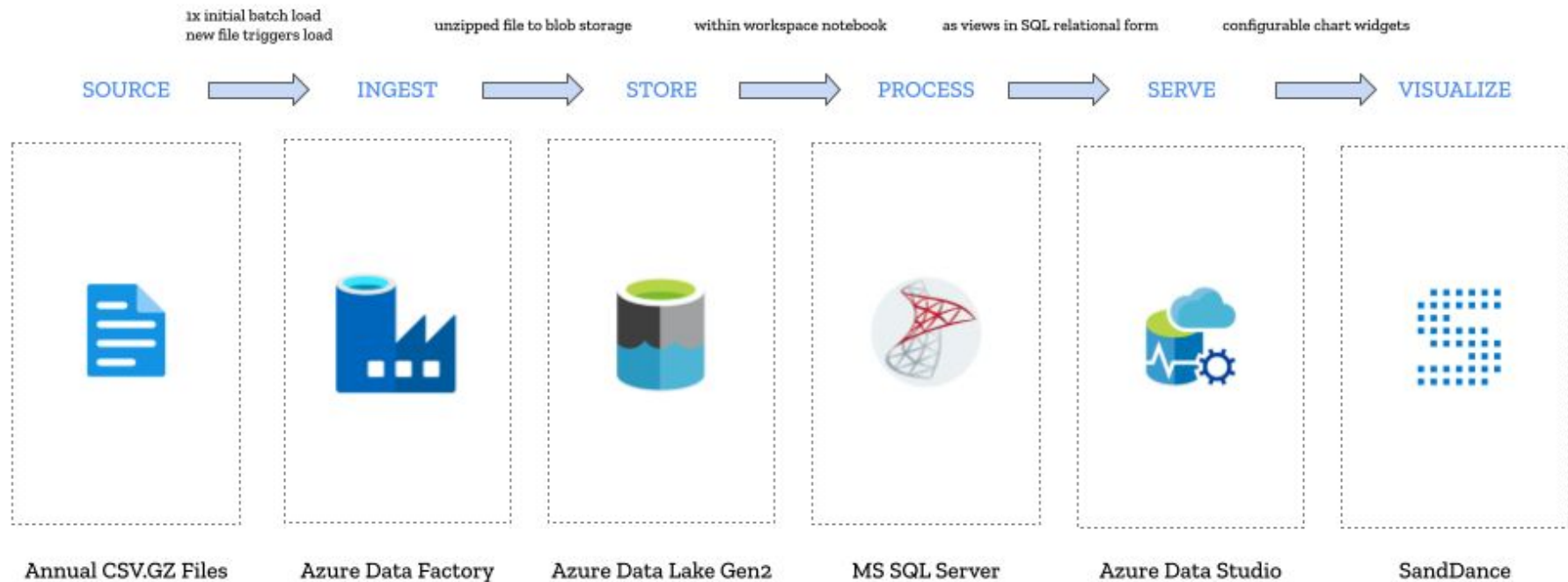
3. I was unable to create a **Microsoft Power BI** resource without using an organization email account

Provisioning business intelligence software is overkill for my requirements, as **Azure Data Studio** extension **SandDance** accommodates plotting.

Initial Architecture



Final Architecture



Resource Deployment Exports

The ability to deploy and export the resources was possible for all resources except **Azure Data Factory**, which lacks the feature.

```

databricks template.json
#
storage_template.json

$schema: 'https://schema.management.azure.com/schemas/2019-04-01/deploymentTemplate.json',
'contentVersion': '1.0.0.0',
parameters: {
  'workspaces databricksname': {
    'defaultValue': 'databricksname',
    'type': 'String'
  }
},
resources: [
  {
    'apiVersion': '2021-04-01-preview',
    'location': 'australiaeast',
    'name': '[parameters('workspaces databricksname')]',
    'properties': {
      'authorizations': [
        {
          'principalId': '0a7da0f0-8153-4348-988a-c267228be0d',
          'roleDefinitionId': '83af603f-48ff-443c-876c-2f6dc8e1635c'
        }
      ],
      'createdBy': {},
      'createdDateTime': '2022-06-15T12:09:33.656924Z',
      'managedResourceGroup': '
[concat('subscriptions/5f5807fa-cdb-44de-8357-76a176efc946/resourceGroups/databricks-rg-',
parameters('workspaces databricksname'), '-yafcdkhubis')]',
      'parameters': {
        'enablePublicIp': {
          'type': 'Bool',
          'value': false
        },
        'natGatewayName': {
          'type': 'String',
          'value': 'nat-gateway'
        },
        'prepareEncryption': {
          'type': 'Bool',
          'value': false
        },
        'publicIpName': {
          'type': 'String',
          'value': 'nat-gw-public-ip'
        },
        'requireInfrastructureEncryption': {
          'type': 'Bool',
          'value': false
        },
        'storageAccountName': {
          'type': 'String',
          'value': 'dbsstoragest2d7gprhsk'
        },
        'storageAccountSkusName': {
          'type': 'String',
          'value': 'Standard_GS'
        },
        'webAddressPrefix': {
          'type': 'String',
          'value': '10.139'
        }
      },
      'updatedBy': {}
    },
    'sku': {
      'name': 'Standard'
    },
    'type': 'Microsoft.Databricks/Workspaces'
  }
],
'variables': {}

```

```
storage_template.json                                     synapse_template.json
```

```
{  
  "name": "Microsoft.Storage/storageAccounts",  
  "publisher": "1.0.0-0",  
  "parameters": {},  
  "variables": {},  
  "resources": [],  
  "  
    {  
      "type": "Microsoft.Storage/storageAccounts",  
      "apiVersion": "2021-09-01",  
      "name": "pipelinestorageaccounts",  
      "location": "australiaeast",  
      "sku": {  
        "name": "Standard_LRS",  
        "tier": "Standard"  
      },  
      "kind": "StorageV2",  
      "properties": {  
        "blobEndpointType": "Standard",  
        "defaultToHttpsAuthentication": false,  
        "publicNetworkAccess": "enabled",  
        "allowCrossTenantReplication": true,  
        "minimumTlsVersion": "TLS1_2",  
        "allowBlobPublicAccess": true,  
        "allowSharedKeyAccess": true,  
        "isHnsEnabled": true,  
        "networkAcls": {  
          "bypass": "AzureServices",  
          "virtualNetworkRule": {  
            "ipRules": [],  
            "defaultAction": "Allow"  
          }  
        },  
        "supportTrustedDomains": true,  
        "encryption": {  
          "services": {  
            "file": {  
              "keyType": "Account",  
              "enabled": true  
            },  
            "blob": {  
              "keyType": "Account",  
              "enabled": true  
            }  
          },  
          "keySource": "Microsoft.Storage"  
        },  
        "accessTier": "Hot"  
      },  
      "  
    },  
    {  
      "type": "Microsoft.Storage/storageAccounts/blobServices",  
      "apiVersion": "2021-09-01",  
      "name": "pipelinestorageaccounts/default",  
      "dependsOn": [ "  
        \"resourceId\": \"Microsoft.Storage/storageAccounts\", \"pipelinestorageaccounts\" ]  
      ],  
      "sku": {  
        "name": "Standard_LRS",  
        "tier": "Standard"  
      },  
      "properties": {  
        "changeFeed": {  
          "enabled": false  
        },  
        "restorePolicy": {  
          "enabled": false  
        },  
        "containerDeleteRetentionPolicy": {  
          "enabled": false  
        },  
        "cors": {  
          "corRules": []  
        },  
        "deleteRetentionPolicy": {  
          "allowManagedDelete": false,  
          "enabled": false  
        }  
      },  
      "
```

```

synapse_template.json
{
  "schema": "https://schemas.microsoft.com/schemas/2019-04-01/azurevmtemplates.json",
  "contentVersion": "1.0.0.0",
  "parameters": {
    "virtualMachinesDefaultStorageContainerPath": {
      "type": "string",
      "workspaces_synapseaccount": {
        "defaultValue": "synapseeast",
        "type": "string"
      },
      "storageaccount_synapseaccount_externalid": {
        "defaultValue": "/subscriptions/8687fa-65a0-4463-8327-76a79efc446/resourcenames/storageaccounts/ps",
        "type": "string"
      }
    },
    "variables": {},
    "resources": [
      {
        "type": "Microsoft.Synapse/workspaces",
        "apiVersion": "2021-06-01",
        "name": "[parameters('workspaces_synapseaccount_name')]",
        "location": "australiaeast",
        "identity": {
          "type": "SystemAssigned"
        },
        "properties": {
          "securityInfo": {
            "resourceId": "[parameters('storageaccounts_synapseaccount_externalid')]",
            "createWorkspacePrivateEndpoint": true,
            "resourceUrl": "https://admlastest.synapseact.dfs.core.windows.net",
            "fileSystem": "pipelinecontainer"
          },
          "encryption": {},
          "connectivityInfo": {
            "web": "[concat('https://web.azureynapse.net/workspaces/subscriptions/8687fa-65a0-4463-8327-76a79efc446/workspaces/ps')]",
            "sql": "[concat('https://sql.azureynapse.net/workspaces/ps')]",
            "sqlDomain": "[concat(parameters('workspaces_synapseaccount_name'), '-domain.sql.azure')]",
            "sql": "[concat(parameters('workspaces_synapseaccount_name'), '-sql.azureynapse.net')]"
          },
          "managedResourceGroups": "synapsamanager",
          "administratorsInfo": {
            "privilegeEndpointConnections": {},
            "managedVirtualNetworkSettings": {
              "preventSelfAdministration": false,
              "allowLocalDefaultRouting": {}
            },
            "enableAzureArc": "Enable",
            "enableKubernetesAccess": {},
            "initialWorkspaceDefaultPath": "4f82926e-d586-40f0-8027-52106f20c6c4",
            "azureADOnlyAuthentication": false,
            "trustedServiceBypassEnabled": false
          }
        }
      },
      {
        "type": "Microsoft.Synapse/workspaces/auditingSettings",
        "apiVersion": "2021-06-01",
        "name": "[concat(parameters('workspaces_synapseaccount_name'), '/Default')]",
        "dependsOn": [
          "[resourceId('Microsoft.Synapse/workspaces', parameters('workspaces_synapseaccount_name'))]"
        ],
        "properties": {
          "retentionDays": 90,
          "selectColumnsNames": [],
          "isTargetSecondaryKeyIndex": false,
          "isQueryResultTargetEnabled": false,
          "state": "Disable",
          "storageAccountSubscriptionId": "00000000-0000-0000-0000-000000000000"
        }
      }
    ],
    "type": "Microsoft.Synapse/workspaces/azureADOnlyAuthentications",
    "apiVersion": "2021-06-01",
    "name": "[concat(parameters('workspaces_synapseaccount_name'), '/default')]",
    "dependsOn": [

```