

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**

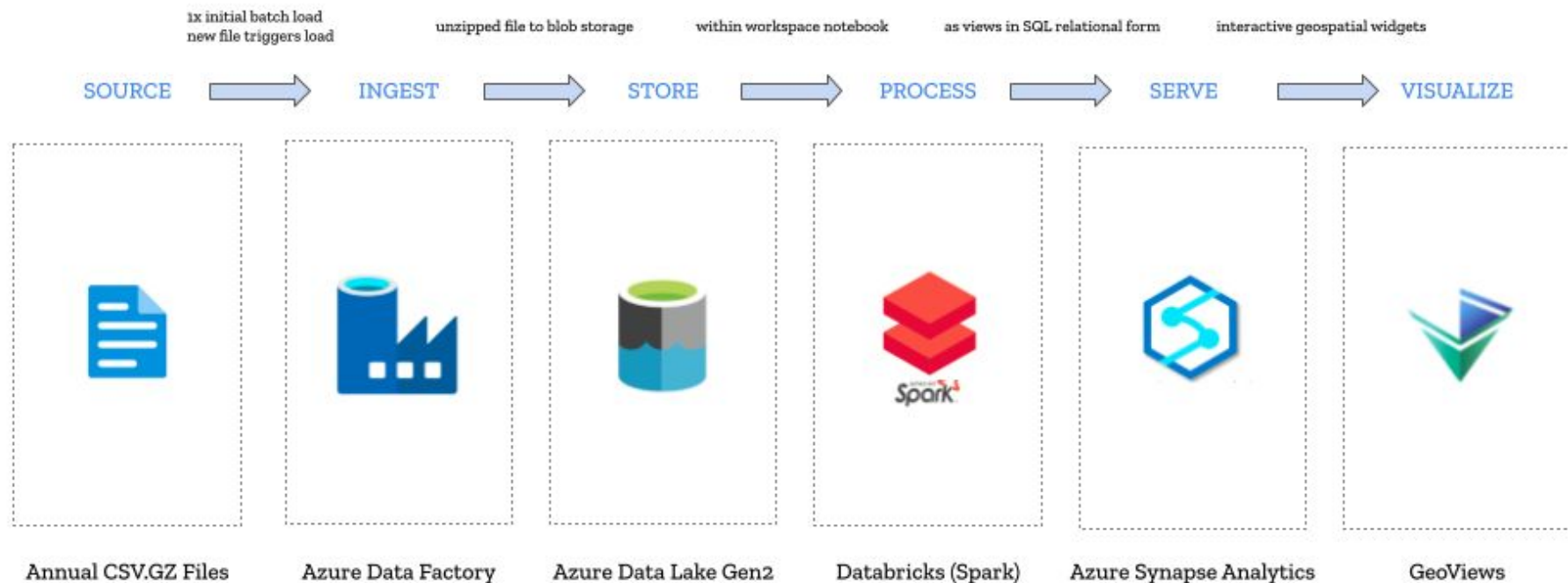
3. I was unable to create a **Microsoft Power BI** resource without using an organization email account

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

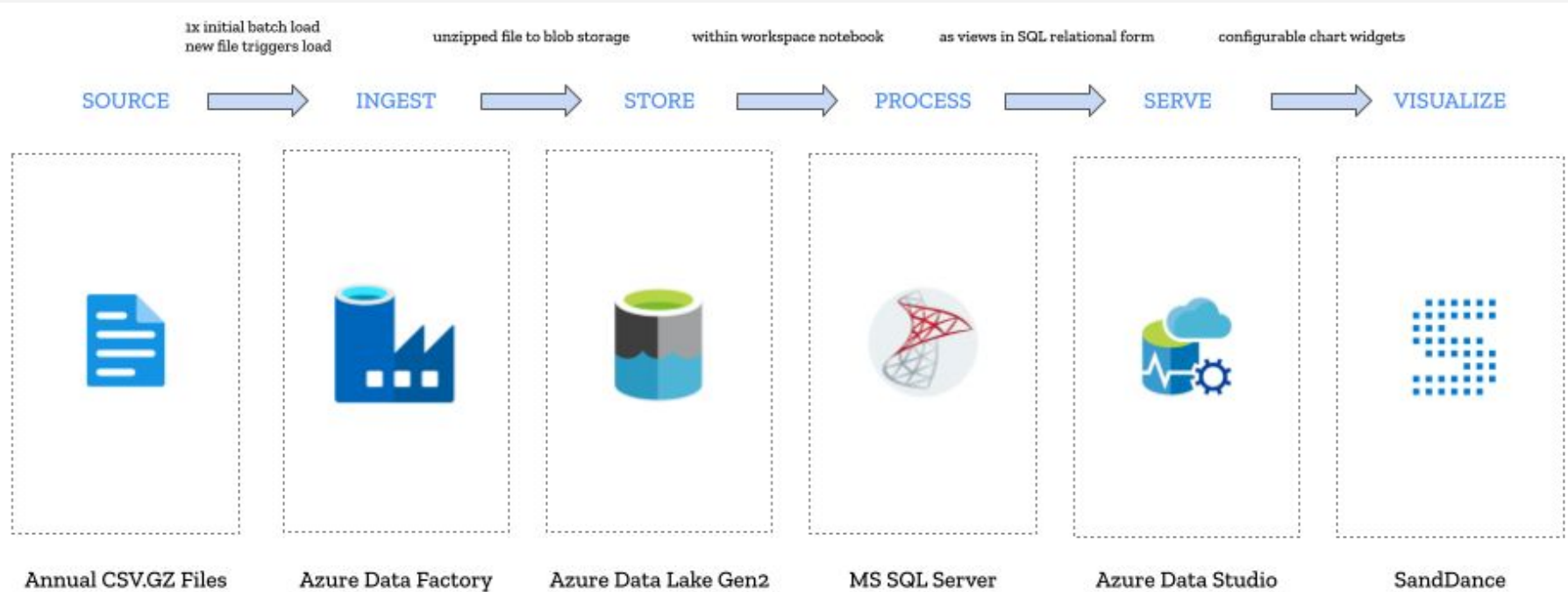
4. I then realized that if and when I want to move to a non-Azure infrastructure, I would want to be cloud-agnostic,

so I converted **Azure Data Studio** + **SandDance** to **MySQL** + **Jupyter**

Initial Architecture



Intermediate Architecture



Final Decision Architecture

1x initial batch load
new file triggers load

unzipped file to blob storage

within workspace notebook

as views in SQL relational form

plot using python libraries via notebook

SOURCE



INGEST



STORE



PROCESS



SERVE



VISUALIZE



Annual CSV.GZ Files



Azure Data Factory



Azure Data Lake Gen2



MySQL



MySQL Workbench



Jupyter Notebooks

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
{
  "$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": "6a7da9f6-d333-4348-988a-e267929b0eb0",
            "roleDefinitionId": "6a6a657f-a8ff-443c-a75c-2f68dcb43575"
          }
        ],
        "createdBy": {},
        "createdDateTime": "2022-06-13T12:09:33.6569284Z",
        "managedResourceGroup": "[concat('/subscriptions/55f807fa-c5eb-44dd-8357-76a176efc946/resourceGroups/databricks-rg-', parameters('workspaces.databricksname'), '-', yacddkhubois')]",
        "parameters": {
          "enablePublicIp": {
            "type": "Bool",
            "value": false
          },
          "natGatewayName": {
            "type": "String",
            "value": "nat-gateway"
          },
          "requireEncryption": {
            "type": "Bool",
            "value": false
          },
          "publicIpName": {
            "type": "String",
            "value": "nat-gw-public-ip"
          },
          "requireInfrastructureEncryption": {
            "type": "Bool",
            "value": false
          },
          "storageAccountName": {
            "type": "String",
            "value": "dstorage2cd76pvhsk"
          },
          "storageAccountSkuname": {
            "type": "String",
            "value": "Standard_GS"
          },
          "webAddressPrefix": {
            "type": "String",
            "value": "10.139"
          },
          "updatedBy": {}
        },
        "sku": {
          "name": "standard"
        },
        "type": "Microsoft.DataBricks/workspaces"
      },
      "variables": {}
    }
  ]
}
```

```
storage_template.json
{
  "$schema": "https://schema.management.azure.com/schemas/2019-04-01/deploymentTemplate.json#",
  "contentVersion": "1.0.0.0",
  "parameters": {
    "variables": {}
  },
  "resources": [
    {
      "type": "Microsoft.Storage/storageAccounts",
      "apiVersion": "2022-09-01",
      "name": "pipelinestorageaccount",
      "location": "australiaeast",
      "sku": {
        "name": "Standard_LRS",
        "tier": "Standard"
      },
      "kind": "StorageV2",
      "properties": {
        "endpointsType": "Standard",
        "defaultObjectEncryption": false,
        "publicNetworkAccess": "Enabled",
        "allowCrossTenantReplication": true,
        "minimumTlsVersion": "TLS1.2",
        "allowBlobPublicAccess": true,
        "allowSharedKeyAccess": true,
        "isInferencing": true,
        "networkDefault": {
          "type": "AzureServices",
          "virtualNetworkRules": [
            {
              "ipAddress": [
                "defaultAction": "Allow"
              ],
              "supportHttpTrafficOnly": true,
              "encryption": {
                "requireInfrastructureEncryption": false,
                "services": [
                  {
                    "keyType": "Account",
                    "enabled": true
                  },
                  {
                    "keyType": "Account",
                    "enabled": true
                  }
                ]
              },
              "blob": {
                "keyType": "Account",
                "enabled": true
              },
              "keySource": "Microsoft.Storage"
            }
          ],
          "accessTier": "Hot"
        },
        "type": "Microsoft.Storage/storageAccounts/blobServices",
        "apiVersion": "2022-09-01",
        "name": "pipelinestorageaccount/default",
        "dependsOn": [
          "[resourceId('Microsoft.Storage/storageAccounts', 'pipelinestorageaccount')]"
        ],
        "sku": {
          "name": "Standard_LRS",
          "tier": "Standard"
        },
        "properties": {
          "changeMetadata": {
            "enabled": false
          },
          "restorePolicy": {
            "enabled": false
          },
          "containerDeleteRetentionPolicy": {
            "enabled": false
          },
          "cors": {
            "corsRules": [
              {
                "deleteRetentionPolicy": {
                  "allowOverwrite": false,
                  "enabled": false
                }
              }
            ]
          }
        }
      }
    }
  ]
}
```

```
synapse_template.json
{
  "$schema": "https://schema.management.azure.com/schemas/2019-04-01/deploymentTemplate.json#",
  "contentVersion": "1.0.0.0",
  "parameters": {
    "workspaces.synapseaccountname": {
      "defaultValue": "synapseaccount",
      "type": "String"
    },
    "storageAccountName": {
      "defaultValue": "[concat('/subscriptions/55f807fa-c5eb-44dd-8357-76a176efc946/resourceGroups/storagexp-rg-', parameters('workspaces.synapseaccountname'))]",
      "type": "String"
    },
    "variables": {}
  },
  "resources": [
    {
      "type": "Microsoft.Synapse/workspaces",
      "apiVersion": "2021-06-01",
      "name": "[concat(parameters('workspaces.synapseaccountname'))]",
      "location": "australiaeast",
      "identity": {
        "type": "SystemAssigned"
      },
      "properties": {
        "defaultDataLakeStorage": {
          "resourceId": "[parameters('storageAccounts.pipelinestorageaccountname')]",
          "createOnDemand": true,
          "accountUrl": "https://pipelinestorageaccount.dfs.core.windows.net",
          "filesystem": "pipelinestorageaccount",
          "encryption": {
            "type": "Microsoft.Storage/storageAccounts",
            "apiVersion": "2021-06-01",
            "name": "[concat(parameters('workspaces.synapseaccountname'))]",
            "dev.azure.com": "[concat('https://', parameters('workspaces.synapseaccountname'), '.dev.azure.com')]",
            "sql": "[concat(parameters('workspaces.synapseaccountname'), '.sql.azure-synapse.net')]"
          },
          "managedResourceGroup": "synapsemanagedrg",
          "sqlAdministratorLogin": "sqladminuser",
          "privateEndpointConnections": [
            {
              "preventDataExfiltration": false,
              "allowDefaultTenantForLinking": [
                {
                  "type": "Microsoft.Synapse/workspaces",
                  "apiVersion": "2021-06-01",
                  "name": "[concat(parameters('workspaces.synapseaccountname'))]",
                  "dependsOn": [
                    "[resourceId('Microsoft.Synapse/workspaces', parameters('workspaces.synapseaccountname'))]"
                  ]
                }
              ]
            }
          ],
          "properties": {
            "retentionDays": 9,
            "auditActionGroups": [
              {
                "isStorageSecondaryKey": false,
                "isAzureMonitorExportEnabled": false,
                "state": "Disabled"
              }
            ],
            "storageAccountSubscriptionId": "00000000-0000-0000-0000-000000000000"
          }
        },
        "type": "Microsoft.Synapse/workspaces/azureOnlyAuthentications",
        "apiVersion": "2021-06-01",
        "name": "[concat(parameters('workspaces.synapseaccountname'))]",
        "dependsOn": [
          "[resourceId('Microsoft.Synapse/workspaces', parameters('workspaces.synapseaccountname'))]"
        ]
      }
    }
  ]
}
```