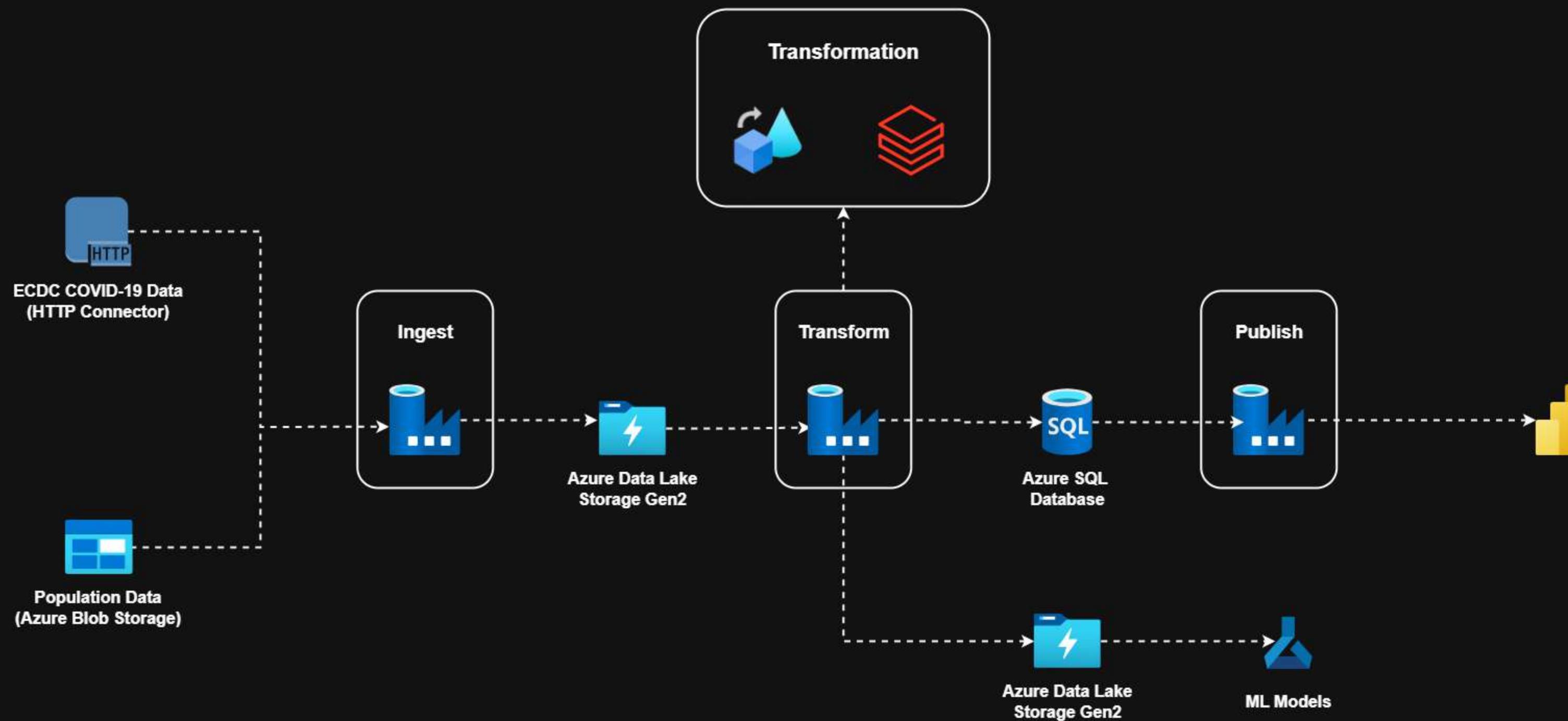


# Covid 19 Project Solution Architecture



» Data Factory Validate all Publish all Auto Save

Preview experience On

**Factory Resources** Filter resources by name +

**Pipelines** 10

- pl\_execute\_population\_pipelines
- ingest 2
  - pl\_ingest\_ecdc\_data
  - pl\_ingest\_population\_data
- processed 4
  - pl\_processed\_cases\_and\_deaths
  - pl\_processed\_hospital\_admissions
  - pl\_processed\_population\_data
  - pl\_processed\_testing\_data
- sql 3
  - pl\_sql\_cases\_and\_deaths
  - pl\_sql\_hospital\_admissions\_daily
  - pl\_sql\_testing\_data
- Change Data Capture (preview) 0
- Datasets 16
- Data flows 2
  - df\_transform\_cases\_deaths
  - df\_transform\_hospital\_admissions
- Power Query 0

**Activities** Validate Debug Trigger (1)

Search activities

> Move and transform

> Synapse

> Azure Data Explorer

> Azure Function

> Batch Service

> Databricks

> Data Lake Analytics

> General

> HDInsight

> Iteration & conditionals

> Machine Learning

> Power Query

**pl\_ingest\_ecdc\_data**

ForEach

Execute Copy for Each File

Lookup

Lookup ECDC File List

Activities

Copy ECDC Data

Parameters Variables Settings Output

+ New

```
graph LR; subgraph Loop [ForEach Execute Copy for Each File]; direction TB; L[Lookup ECDC File List] --> A[Activities]; end; A --> C[Copy ECDC Data]
```

» Data Factory Validate all Publish all Auto Save

Factory Resources Filter resources by name +

**Pipelines** 10

- pl\_execute\_population\_pipelines
- ingest 2
  - pl\_ingest\_ecdc\_data
  - pl\_ingest\_population\_data
- processed 4
  - pl\_processed\_cases\_and\_deaths
  - pl\_processed\_hospital\_admissions
  - pl\_processed\_population\_data
  - pl\_processed\_testing\_data
- sql 3
  - pl\_sql\_cases\_and\_deaths
  - pl\_sql\_hospital\_admissions\_daily
  - pl\_sql\_testing\_data

Change Data Capture (preview) 0

Datasets 16

Data flows 2

- df\_transform\_cases\_deaths
- df\_transform\_hospital\_admissions

Power Query 0

Activities Validate Debug Add trigger

Search activities

> Move and transform

> Synapse

> Azure Data Explorer

> Azure Function

> Batch Service

> Databricks

> Data Lake Analytics

> General

> HDInsight

> Iteration & conditionals

> Machine Learning

> Power Query

Validation Check if the file exists → Get Metadata Get File Metadata

If Condition If Column Count Matches

True Copy Population... → Delete Source File

False Column Match... → + New

Parameters Variables Settings Output

+ New

```
graph LR; Validation[Validation: Check if the file exists] --> GetMetadata[Get Metadata: Get File Metadata]; GetMetadata --> IfCondition[If Condition: If Column Count Matches]; IfCondition -- True --> CopyPopulation[Copy Population...]; CopyPopulation --> DeleteSourceFile[Delete Source File]; IfCondition -- False --> ColumnMatch[Column Match...];
```

» Data Factory Validate all Publish all Auto Save Preview experience On

Factory Resources Filter resources by name +

- Pipelines 10
  - pl\_execute\_population\_pipelines
  - ingest 2
    - pl\_ingest\_ecdc\_data
    - pl\_ingest\_population\_data
  - processed 4
    - pl\_processed\_cases\_and\_deaths
    - pl\_processed\_hospital\_admissions
    - pl\_processed\_population\_data
    - pl\_processed\_testing\_data
  - sql 3
    - pl\_sql\_cases\_and\_deaths
    - pl\_sql\_hospital\_admissions\_daily
    - pl\_sql\_testing\_data
- Change Data Capture (preview) 0
- Datasets 16
- Data flows 2
  - df\_transform\_cases\_deaths
  - df\_transform\_hospital\_admissions
- Power Query 0

Activities Validate Debug Trigger (1) Data flow debug

Search activities

- > Move and transform
- > Synapse
- > Azure Data Explorer
- > Azure Function
- > Batch Service
- > Databricks
- > Data Lake Analytics
- > General
- > HDInsight
- > Iteration & conditionals
- > Machine Learning
- > Power Query

Data flow df\_transform\_cases\_deaths

Parameters Variables Settings Output

+ New

Validate all  Auto Save  Preview experience  On

Factory Resources   +

Pipelines  10

- pl\_execute\_population\_pipelines
- ingest  2
  - pl\_ingest\_ecdc\_data
  - pl\_ingest\_population\_data
- processed  4
  - pl\_processed\_cases\_and\_deaths
  - pl\_processed\_hospital\_admissions
  - pl\_processed\_population\_data
  - pl\_processed\_testing\_data
- sql  3
  - pl\_sql\_cases\_and\_deaths
  - pl\_sql\_hospital\_admissions\_daily
  - pl\_sql\_testing\_data
- Change Data Capture (preview)  0
- Datasets  16
- Data flows  2
  - df\_transform\_cases\_deaths
  - df\_transform\_hospital\_admissions
- Power Query  0

df transform\_cases\_deaths

Validate  Data flow debug

The data flow diagram consists of the following stages:

- CasesAndDeathSource**: Import data from ds\_raw\_cases\_and\_deaths.
- FilterEuropeOnly**: Filtering rows using expressions on columns 'continent', 'country\_code'.
- SelectOnlyRequiredField**: Renaming FilterEuropeOnly to SelectOnlyRequiredField with columns 'country', 'country\_code', 'population'.
- PivotCounts**: Pivots row values into columns, groups columns and aggregates data.
- LookupCountry**: Lookup on 'PivotCounts' from 'CountryLookup'.
- SelectForSink**: Renaming LookupCountry to SelectForSink with columns 'country', 'country\_code\_2\_digit', 'country\_code\_3\_digit'.
- CasesAndDeathSink**: Export data to ds\_processed\_cases\_and\_deaths.

Add Source

Parameters  Settings

» Data Factory Validate all Publish all Auto Save Preview experience On

### Factory Resources

Filter resources by name +

- Pipelines (10)
  - pl\_execute\_population\_pipelines
- ingest (2)
  - pl\_ingest\_ecdc\_data
  - pl\_ingest\_population\_data
- processed (4)
  - pl\_processed\_cases\_and\_deaths
  - pl\_processed\_hospital\_admissions
  - pl\_processed\_population\_data
  - pl\_processed\_testing\_data
- sql (3)
  - pl\_sql\_cases\_and\_deaths
  - pl\_sql\_hospital\_admissions\_daily
  - pl\_sql\_testing\_data
- Change Data Capture (preview) (0)
- Datasets (16)
- Data flows (2)
  - df\_transform\_cases\_deaths
  - df\_transform\_hospital\_admissions
- Power Query (0)

### Activities

Validate Debug Trigger (1) Data flow debug

Search activities

- > Move and transform
- > Synapse
- > Azure Data Explorer
- > Azure Function
- > Batch Service
- > Databricks
- > Data Lake Analytics
- > General
- > HDInsight
- > Iteration & conditionals
- > Machine Learning
- > Power Query

Data flow

df\_transform\_hospital admissions

Parameters Variables Settings Output

New

» Data Factory Validate all Publish all Auto Save Preview experience On

### Factory Resources

Filter resources by name +

- Pipelines (10)
  - pl\_execute\_population\_pipelines
- ingest (2)
  - pl\_ingest\_ecdc\_data
  - pl\_ingest\_population\_data
- processed (4)
  - pl\_processed\_cases\_and\_deaths
  - pl\_processed\_hospital\_admissions
  - pl\_processed\_population\_data
  - pl\_processed\_testing\_data
- sql (3)
  - pl\_sql\_cases\_and\_deaths
  - pl\_sql\_hospital\_admissions\_daily
  - pl\_sql\_testing\_data
- Change Data Capture (preview) (0)
- Datasets (16)
- Data flows (2)
  - df\_transform\_cases\_deaths
  - df\_transform\_hospital\_admissi...
    - ...
- Power Query (0)

Validate Data flow debug

pl\_ingest\_ecdc\_data pl\_ingest\_population... pl\_processed\_cases... df\_transform\_cases... df\_transform\_hospital... df\_transform\_hospital... X

Validate Data flow debug

HospitalAdmissionS... SelectRequiredFields LookupCountry SelectRequiredFields2 Weekly JoinWithDimDate PivotWeeklyCounts SelectRequiredFields3 SortWeekly WeeklyDataSink

Import data from 'ds\_new\_hospital\_admissions' Renaming HospitalAdmissionSource to 'SelectRequiredFields' with column 'country, indicator' Lookup on 'SelectRequiredFields' from 'CountryLookup' Renaming 'LookupCountry' to 'SelectRequiredFields2' with column 'country, country\_code\_2\_digit' Conditionally distributing the data in indicator groups, based on columns 'IV' Inner join on 'DailyAndWeeklySplit@Weekly' and 'AggDimDate' Visits row values into columns, groups columns and aggregates data Renaming 'PivotWeeklyCounts' to 'SelectRequiredFields3' with column 'country, country\_code\_2\_digit' Sorting rows on column 'reported\_year\_week\_country' Export data to 'ds\_processed\_hospital\_admissions\_weekly'

CountryLookup AggDimDate Daily PivotDailyCounts SelectRequiredFields4 SortDaily DailyDataSink

Import data from 'ds\_country\_lookup' Conditionally distributing the data in indicator groups, based on columns 'IV' Visits row values into columns, groups columns and aggregates data Renaming 'PivotDailyCounts' to 'SelectRequiredFields4' with column 'country, country\_code\_2\_digit' Sorting rows on columns 'reported\_date, country' Export data to 'ds\_processed\_hospital\_admissions\_daily'

CountryLookup DimDateLookup AddECDCYearWeekField AggDimDate

Import data from 'ds\_dim\_date\_lookup' Creating/updating the columns date\_key, date, year, month, day, day\_name, day\_of\_year, week\_of\_month, week\_of\_year, week\_start\_date, week\_end\_date Aggregating data by 'week, year, week' producing columns 'week\_start\_date', 'week\_end\_date'

Add Source

Parameters Settings

New

» Data Factory Validate all Publish all Auto Save Preview experience On

### Factory Resources

Filter resources by name +

- Pipelines (10)
  - pl\_execute\_population\_pipelines
  - ingest (2)
    - pl\_ingest\_ecdc\_data
    - pl\_ingest\_population\_data
  - processed (4)
    - pl\_processed\_cases\_and\_deaths
    - pl\_processed\_hospital\_admissions
    - pl\_processed\_population\_data
    - pl\_processed\_testing\_data
  - sql (3)
    - pl\_sql\_cases\_and\_deaths
    - pl\_sql\_hospital\_admissions\_daily
    - pl\_sql\_testing\_data
  - Change Data Capture (preview) (0)
  - Datasets (16)
  - Data flows (2)
    - df\_transform\_cases\_deaths
    - df\_transform\_hospital\_admissions
  - Power Query (0)

### Activities

Validate Debug Trigger (1)

Search activities

  - > Move and transform
  - > Synapse
  - > Azure Data Explorer
  - > Azure Function
  - > Batch Service
  - > Databricks
  - > Data Lake Analytics
  - > General
  - > HDInsight
  - > Iteration & conditionals
  - > Machine Learning
  - > Power Query

Notebook

ExecuteTestingData Transformation ✓

Parameters Variables Settings Output

New

» Data Factory Validate all Publish all Auto Save Preview experience On

Factory Resources Filter resources by name +

Pipelines 10

- pl\_execute\_population\_pipelines
- ingest 2
  - pl\_ingest\_ecdc\_data
  - pl\_ingest\_population\_data
- processed 4
  - pl\_processed\_cases\_and\_deaths
  - pl\_processed\_hospital\_admissions
  - pl\_processed\_population\_data
  - pl\_processed\_testing\_data
- sql 3
  - pl\_sql\_cases\_and\_deaths
  - pl\_sql\_hospital\_admissions\_daily
  - pl\_sql\_testing\_data
- Change Data Capture (preview) 0
- Datasets 16
- Data flows 2
  - df\_transform\_cases\_deaths
  - df\_transform\_hospital\_admissions
- Power Query 0

Activities Validate Debug Add trigger

Search activities

- Move and transform
- Synapse
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

Notebook ExecutePopulationDataTransformation ✓

Parameters Variables Settings Output

+ New

» Data Factory Validate all Publish all Auto Save Preview experience On

Factory Resources < < data pl\_ ingest\_population... pl\_ processed\_cases... df\_ transform\_cases... pl\_ processed\_hospit... df\_ transform\_hospita... pl\_ processed\_testin... pl\_ processed\_popu... pl\_ sql\_cases\_and\_d... X

Filter resources by name +

Pipelines 10

- pl\_execute\_population\_pipelines
- ingest 2
  - pl\_ ingest\_ecdc\_data
  - pl\_ ingest\_population\_data
- processed 4
  - pl\_ processed\_cases\_and\_deaths
  - pl\_ processed\_hospital\_admissions
  - pl\_ processed\_population\_data
  - pl\_ processed\_testing\_data
- sql 3
  - pl\_ sql\_cases\_and\_deaths
  - pl\_ sql\_hospital\_admissions\_daily
  - pl\_ sql\_testing\_data
- Change Data Capture (preview) 0
- Datasets 16
- Data flows 2
  - df\_ transform\_cases\_deaths
  - df\_ transform\_hospital\_admissions
- Power Query 0

Activities < < Validate Debug Trigger (1)

Search activities

- > Move and transform
- > Synapse
- > Azure Data Explorer
- > Azure Function
- > Batch Service
- > Databricks
- > Data Lake Analytics
- > General
- > HDInsight
- > Iteration & conditionals
- > Machine Learning
- > Power Query

Copy data

CopyCasesAndDeaths

Parameters Variables Settings Output

+ New

» Data Factory Validate all Publish all Auto Save Preview experience On

Factory Resources < atio... 000 pl\_processed\_cases... 000 df\_transform\_cases... 000 pl\_processed\_hospit... 000 df\_transform\_hospita... 000 pl\_processed\_testin... 000 pl\_processed\_popu... 000 pl\_sql\_cases\_and\_d... 000 pl\_sql\_hospital\_ad... X

Filter resources by name +

Pipelines 10

- 000 pl\_execute\_population\_pipelines
- ingest 2
  - 000 pl\_ingest\_ecdc\_data
  - 000 pl\_ingest\_population\_data
- processed 4
  - 000 pl\_processed\_cases\_and\_deaths
  - 000 pl\_processed\_hospital\_admissions
  - 000 pl\_processed\_population\_data
  - 000 pl\_processed\_testing\_data
- sql 3
  - 000 pl\_sql\_cases\_and\_deaths
  - 000 pl\_sql\_hospital\_admissions\_daily
  - 000 pl\_sql\_testing\_data
- Change Data Capture (preview) 0
- Datasets 16
- Data flows 2
  - 000 df\_transform\_cases\_deaths
  - 000 df\_transform\_hospital\_admissions
- Power Query 0

Activities < Validate Debug Trigger (1)

Search activities

- > Move and transform
- > Synapse
- > Azure Data Explorer
- > Azure Function
- > Batch Service
- > Databricks
- > Data Lake Analytics
- > General
- > HDInsight
- > Iteration & conditionals
- > Machine Learning
- > Power Query

Copy data

CopyHospitalAdmissionsDaily

Parameters Variables Settings Output

+ New

» Data Factory Validate all Publish all Auto Save Preview experience On

Factory Resources < < Datasets... df\_transform\_cases... df\_processed\_hosp... df\_transform\_hospita... df\_processed\_testin... df\_processed\_popu... df\_sql\_cases\_and\_d... df\_sql\_hospital\_ad... df\_sql\_testing\_data X ...

Filter resources by name +

Pipelines 10

- pl\_execute\_population\_pipelines
- ingest 2
  - pl\_ingest\_ecdc\_data
  - pl\_ingest\_population\_data
- processed 4
  - pl\_processed\_cases\_and\_deaths
  - pl\_processed\_hospital\_admissions
  - pl\_processed\_population\_data
  - pl\_processed\_testing\_data
- sql 3
  - pl\_sql\_cases\_and\_deaths
  - pl\_sql\_hospital\_admissions\_daily
  - pl\_sql\_testing\_data
- Change Data Capture (preview) 0
- Datasets 16
- Data flows 2
  - df\_transform\_cases\_deaths
  - df\_transform\_hospital\_admissions
- Power Query 0

Activities < < Validate Debug Trigger (1)

Search activities

> Move and transform

> Synapse

> Azure Data Explorer

> Azure Function

> Batch Service

> Databricks

> Data Lake Analytics

> General

> HDInsight

> Iteration & conditionals

> Machine Learning

> Power Query

Copy data

CopyTestingData

Parameters Variables Settings Output

+ New