

# Dataiku.

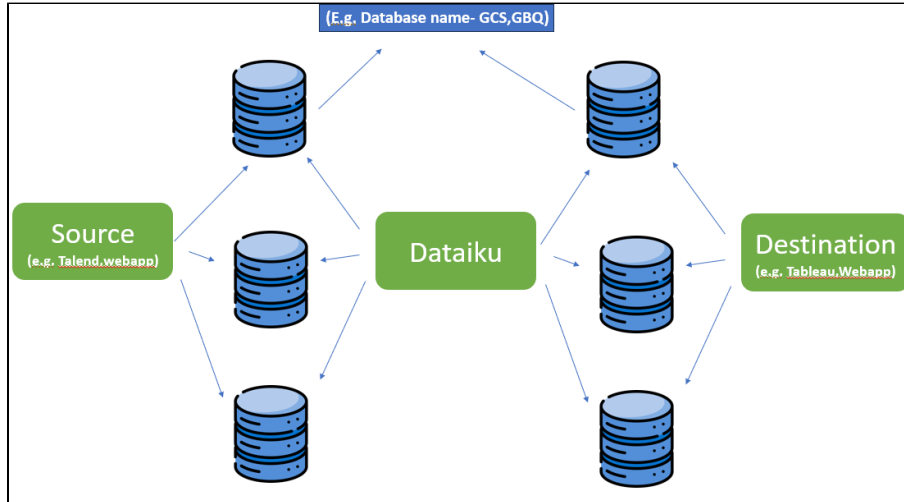
Note: - Please create similar document if target project has dependency on other projects

## Data Flow :-

Upload the Architecture diagram of the project(Data Flow)

For Ex: -

i.

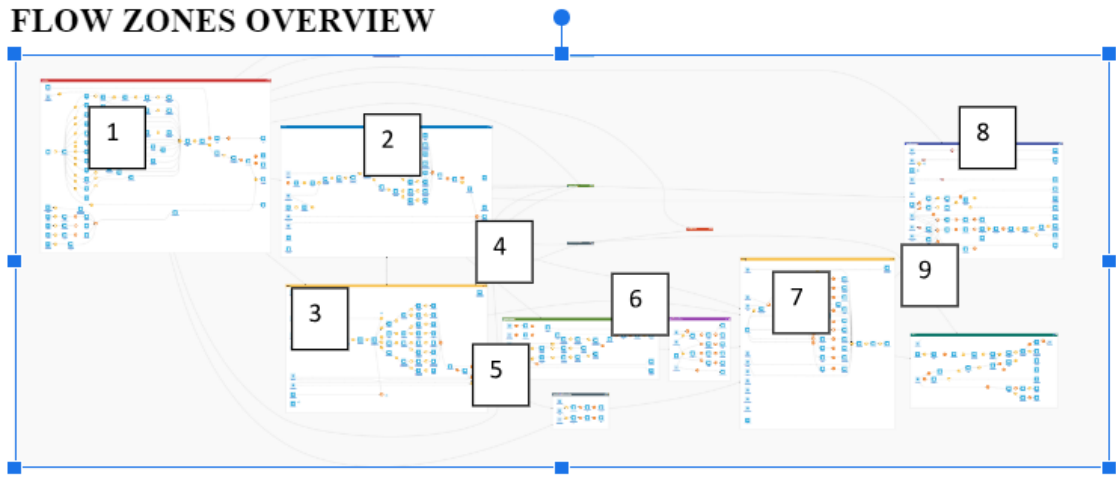


### List down all upstream and downstream dependencies

- a. Upstream Dependencies
  - i. Upstream Dependencies-1
    1. Detail about dependency
      - a. How data getting generated.?
      - b. What is the source of the data.?
      - c. What is the output of the data.?
      - d. How frequently data getting loaded to the output.?
      - e. Mention Point of contact/Team to reach out incase any issue.
    2. Dataset-1 (Which is input source to dataiku).
      - a. List if any intermediate data transformation is happening.
      - b. Detail about data transformation(e.g. Function, Calculation..)
    3. Dataset -2
      - a. Mention similar as Dataset-1
    4. Dependent Scenario-
      - a. Add the scenario name which is been triggered by upstream dependency and mention its time .
      - b. If the dependent scenario trigger is required by dataiku team or not.
  - ii. Upstream Dependencies-2
    1. Mention same details if any other dependencies available
- b. Downstream Dependencies
  - i. Downstream Dependencies-1
    1. Detail about Dependency.
      - a. Where data getting loaded (i.e. Data source, Any platform like Talend, Tableau)?
      - b. How frequently data should be loaded ?
      - c. If any additional reporter needs to added(i.e. Who will receive email on success or failure)?
      - d. No. of user impacted
      - e. Mention Point of contact/Team to reach out incase any issue.
    2. Dataset-1
      - a. Mention if any quality/specific details need to maintain or needs to check.?
      - b. Potential issues/Concerns if any data loss.
      - c. Mention if any backup/ Rollback options available
    3. Dataset - 2
      - a. Mention similar as Dataset -1 if any other dataset connected.
  - ii. Downstream Dependencies-2
    1. Mention same details if any other dependencies available..?

# Procedures : -

1. Project Name
2. URL
3. Short description about the project.(Overview)
4. Diagram of complete flow zone of project.



1. List No of Zones.
  - a. Zone1
  - b. Zone2
  - c. Zone3
2. List No. of Recipes in the project.
  - a. No. of code recipes
    - i. no. of SQL recipes
    - ii. no. of Python recipes
  - b. No. of visual recipes
  - c. Provide tabular data of all recipe details

Recipe Name	Recipe Type	Environment Name
Visual recipe -1	Recipe -Type -1(e.g. Sync, Join.. etc)	Visual recipe
Visual Recipe -2	Recipe -Type -1(e.g. Sync, Join.. etc)	Visual recipe
Sql Recipe -1	SQL	Sql Recipe
Python Recipe -1	Python	Env Name -1

3. List No. Datasets in the project.
  - a. List all the dataset and details.

Datasets Name	Connection Type	Connection Name	Dataset Type
Dataset 1	connection type(e.g. GCS,BQ)	Connection name	Dataset type( e.g. Base/Input/Source, Transformation, Output)
Dataset 2	connection type(e.g. GCS,BQ)	Connection name	Dataset type( e.g. Base/Input/Source, Transformation, Output)
Dataset 2	connection type(e.g. GCS,BQ)	Connection name	Dataset type( e.g. Base/Input/Source, Transformation, Output)

4. Description about all the zones
  - a. Zone 1
    - i. Snapshot of Zone 1
    - ii. List of all sources to the Zone.(e.g. Dataset name.)
    - iii. Short description about how source data is generated and from where its generated.
    - iv. Mention if any query used in datasets.
    - v. Provide detail description about data transformation and data flow in the specific zone.
    - vi. Mention each recipe and problem statement.
    - vii. Mention the queries used in the code recipes(e.g. BW queries, sql queries)
    - viii. Where the destination/output dataset is used.?
    - ix. Mention if any append used in any dataset and backup of dataset.
  - b. Zone -2
    - i. Provide similar description about the zone -1 if any other zones available.
5. Detail about Library
  - a. Library -1

- i. Mention if any python code available in Library
    - ii. Detail description about the python code.
    - iii. Mention if maintenance needed from Dataops team.
    - iv. Mention the changes need to updated.
    - v. Where does this library used.
  - b. Library -2
    - i. Mention details if any library available.
6. Detail about the API.
  - a. Mention if any API created in API designer.
  - b. Short description about the API.
  - c. Mention if any maintenance/code changes needed as per instance.(e.g. URL change, variable change)
7. Project/ Scenario variable
  - a. Mention if any variables stored in project variable.
  - b. Mention if any data needs to be maintained in variables.
  - c. Reason of variable and how its used..
8. Detail about Webapp/Dashboard.
  - a. Webapp/Dashboard -1
    - i. Description about the webapp/dashboard.
    - ii. List of datasets connected to webapp/ dashboard.
    - iii. Mention if any scenario/Job is updating/loading the webapp/dashboard.
    - iv. Frequency of webapp/dashboard to be updated.
    - v. etc.
  - b. Webapp/Dashboard -2
    - i. Mention as similar to webapp-1 if any other webapp/dashboard is available.
9. Plugin info.
  - a. Mention if any plugin used.
  - b. Detail description about the plugin.
  - c. Mention if the plugin is installed from dataiku store or self developed.
  - d. Mention the environment details if self developed.
10. resource consumption details.
  - a. Memory consumption
  - b. CPU consumption.

## Scheduling and Monitoring: -

1. Detail about the scenario.
  - a. Scenario -1
    - i. Short description about the scenario(What it does.?)
    - ii. Mention the steps.
    - iii. Mention the datasets which is building.
    - iv. Mention the trigger details.
    - v. Mention if any custom codes available in steps/triggers.
    - vi. Mention if any further changes needed in scenario.
  - b. Scenario -2
    - i. Mention same details as Scenario 1 if any other scenario available.

## Error Handling: -

1. Potential Problems/Concerns
    - a. Problem -1
      - i. Any recurring and potential issues for the project.
      - ii. Point of contact in case of any issues.
      - iii. Related documentation
    - b. Problem -2
      - i. Mention the same details as problem 1
- Link to the additional documents if you have(Dataiku projects)
    - Dataset related.
    - PPTs
    - Docs
  - Point of contacts for any issue in dataiku project from DT team and Business Team
  - No. of impacted users.

