VDM Data Object Definition

# Source Description

The following information can be captured and stored when identifying a source file or table to be ingested into the “Raw” area.

* ID
* Version
* Name
* Type (File, Table)
* Description
* For files:
  + Host Name
  + Source Path
  + File Format (Fixed Width or Delimited)
  + Delimiter (if delimited)
* For tables
  + Database Name
  + Schema Name
  + Table Name

## JSON Document Example

{    
   **"SourceDesc"**:{    
      **"ID"**:1,  
      **"Version"**:1,  
      **"Name"**:"Customers.csv",  
      **"Description"**:"Customer file from Acme Company",  
      **"Type"**:"File",  
      **"File"**:{    
         **"Hostname"**:"xyz",  
         **"SourcePath"**:"/home/user/data",  
         **"FileFormat"**:"Delimited",  
         **"Delimiter"**:","  
      },  
      **"database"**:{    
         **"Database"**:"CoB",  
         **"Schema"**:" ",  
         **"Table"**:"Customers"  
      }  
   }  
}

# Raw File

The following information can be captured and stored when identifying a file in the Raw area

* ID
* Version
* Name
* Description
* Raw Path Location
* File Format (Fixed Width or Delimited)
* Delimiter (if delimited)
* Status (Inactive or Active)
* Source ID

## JSON Document Example

{    
   **"RawFile"**:{    
      **"ID"**:12,  
      **"Version"**:1,  
      **"Name"**:"Customers.csv",  
      **"Description"**:"Customer file from Acme Company",  
      **"Location"**:"/home/user/data",  
      **"FileFormat"**:"Delimited",  
      **"Delimiter"**:",",  
      **"Status"**:"Active",

**"SourceID"**:1  
   }  
}

# Raw Table

The following information can be captured and stored when identifying a table in the Raw area

* ID\*
* Version\*
* Name\*
* Description
* Database Connection\*
* Status\* (Inactive or Active)
* Source ID

## JSON Document Example

{    
   **"RawTable"**:{    
      **"ID"**:1,  
      **"Version"**:1,  
      **"Name"**:"Customers.csv",  
      **"Description"**:"Customer file from Acme Company",  
      **"Database"**:"CoB",  
      **"Table"**:"Customer",  
      **"Status"**:"Active",

**"SourceObjectID":1**  
   }  
}

# Ingestion Job Processing Information

During ingestion execution the following information is captured to be used and/or displayed later.

* Job Execution Start Time
* Job Execution End Time
* Job Status (Completed or Failed)
* Number rows read from source
* Number rows written to target

# Data Element

The following information is captured and stored when identifying a specific data element.

* Data Element ID
* Name
* Description
* Type
* Length
* Precision
* Constraints[constraint, constraint]
* Raw Source ID (File ID or Table ID)
* Known As (Conformed Object Name)

## JSON Document Example

{    
   **"DataElement"**:{    
      **"DataElementID"**:1,  
      **"Name"**:"cust\_nm",  
      **"Description"**:"Customer Name",  
      **"Type"**:"String",  
      **"Length"**:25,  
      **"Contraints"**:[    
         "No Spaces Allowed"  
      ],  
      **"KnownAs"**:"CustomerName",  
      **"RawSourceID"**:12  
   }  
}

# File Transformation

The following information is captured and stored for a specific file oriented transformation. File transformation are steps like Nest, Pivot, Splitrows, Unnest, Unpivot, Extractlist, Extractkv

* Transform ID
* Transformation
  + Transform Type
  + Sources[ {ID, Name}, …]
  + Targets[{ID, Name}, …]
  + Parameters[{Name, Value}, …]

## JSON Document Example

{    
   **"FileXtms"**:{    
      **"ID"**:1,  
      **"TransformType"**:"Splitrows",  
      **"Sources"**:[    
         {    
            **"ID"**:1,  
            **"Name"**:"column1"  
         }  
      ],  
      **"Parameters"**:[    
         {    
            **"Name"**:"on",  
            **"Value"**:"\r\n"  
         }  
      ]  
   }  
}

# Data Element Transformation

The following information is captured and stored for a specific data element (column) oriented transformation.

* Transform ID
* Transformation
  + Transform Type
  + Sources[ {ID, Name}, …]
  + Targets[{ID, Name}, …]
  + Parameters[{Name, Value}, …]

## JSON Document Example

{    
   **"DataElementXtms"**:{    
      **"ID"**:1,  
      **"TransformType"**:"Split",  
      **"Sources"**:[    
         {    
            **"ID"**:1,  
            **"Name"**:"cust\_nm"  
         }  
      ],  
      **"Targets"**:[    
         {    
            **"ID"**:2,  
            **"Name"**:"first\_nm"  
         },  
         {    
            **"ID"**:3,  
            **"Name"**:"last\_nm"  
         }  
      ],  
      **"Parameters"**:[    
         {    
            **"Name"**:"On",  
            **"Value"**:","  
         },  
         {    
            **"Name"**:"Limit",  
            **"Value"**:"3"  
         }  
      ]  
   }  
}

# Conformed Data Element

The following information is captured and stored when defining a conformed data element.

* Object ID
* Name
* Description
* Definition
* Sources[[{ID, Name}, …] (array of Data Element IDs and associated applicable Rules)
* Preferred Source
* Status (Draft or Approved)

## JSON Document Example

{    
   **"ConformedDataElement"**:{    
      **"ObjectID"**:8,  
      **"Name"**:"CustomerFirstName",

**"Description"**:"Customer …",

**"Definition"**:"Customer …",  
      **"Sources"**:[    
         {    
            **"ID"**:1,  
            **"Name"**:"cust\_nm",  
            **"Rules"**:[    
               {    
                  **"RuleID"**:1,  
                  **"TransformType"**:"Split"  
               }  
            ]  
         },  
         {    
            **"ID"**:9,  
            **"Name"**:"first\_name"  
         }  
      ],  
      **"PreferredSource"**:{    
         **"ID"**:9,  
         **"Name"**:"first\_name"  
      },  
      **"Status"**:"Approved"  
   }  
}

# Conformed Object

The following information is captured and stored when defining a conformed data object.

**Note:** A conformed data object can only contain Conformed Data Elements which are in Approved status.

* Object ID
* Name
* Description
* Conformed Data Elements[]
* Status (Draft or Approved)

## JSON Document Example

{    
   **"ConformedDataObject"**:{    
      **"ObjectID"**:11,  
      **"Name"**:"Customer",

**"Description"**:"Customer …",  
      **"Sources"**:[    
         {    
            **"ID"**:8,  
            **"Name"**:"CustomerFirstName"  
         },  
         {    
            **"ID"**:45,  
            **"Name"**:"CustomerLastName"  
         }  
      ],  
      **"Status"**:"Draft"  
   }  
}

# Job

* Job ID
* Name
* Description
* Type (Batch, Streaming, Ad-Hoc)
* Layer (Source-To-Raw or Raw-To-Conformed)
* Source Objects[]
* Target Objects[]
* If Type is Batch
  + Frequency (Hourly, Daily, Weekly, Monthly)
    - For Weekly specify which day of week (Sun, Mon, Tue, Wed, Thu, Fri, Sat)
    - For Monthly specify which day(s) of month (1 – 31)
  + Time (HH:MM AM/PM)
    - For Hourly only specify minutes after the hour (0-59)
* Dependencies[]
* Active Indicator (Y or N)
* Run History[{StartTime, EndTime, RecordsRead, RecordsWritten}, …]

## JSON Document Example

{    
   **"Job"**:{    
      **"JobID"**:1,  
      **"Name"**:"Ingest Cubstomer",  
      **"Description"**:"Ingest Customer data from …",  
      **"Type"**:"Batch",

**"Layer"**:"Source-To-Raw",  
      **"Sources"**:[    
         {    
            **"ID"**:1,  
            **"Name"**:" Customers.csv"  
         }  
      ],  
      **"Targets"**:[    
         {    
            **"ID"**:2,  
            **"Name"**:"Customers.csv"  
         }  
      ],  
      **"Schedule"**:{    
         **"Frequency"**:"Weekly",  
         **"DayOfWeek"**:"Fri",  
         **"Time"**:"23:00"  
      },  
      **"ActiveIndicator"**:"Y",  
      **"RunHistory"**:[    
         {    
            **"StartTime"**:"1/1/2018 23:00",  
            **"EndTime"**:"1/1/2018 23:30",  
            **"Status"**:"Completed",  
            **"Records Read"**:100,  
            **"Records Written"**:100  
         },  
         {    
            **"StartTime"**:"1/8/2018 23:00",  
            **"EndTime"**:"1/8/2018 23:30",  
            **"Status"**:"Failed",  
            **"Records Read"**:100,  
            **"Records Written"**:90  
         }  
      ]  
   }  
}