***Sample Power BI Report***

***Design Document Fragment***

***Data***

|  |  |
| --- | --- |
| *Date:* | *2025-03-10* |
| *Version:* | *0.3* |
| *Author:* | *Greg Philps* |

***Table of Contents***

[1. Introduction 3](#_Toc192515092)

[2. Scope of Work 3](#_Toc192515093)

[3. Workflow 3](#_Toc192515094)

[4. Issues 3](#_Toc192515095)

[5. Business Rules 3](#_Toc192515096)

[6. Data 3](#_Toc192515097)

[6.1. Environments 3](#_Toc192515098)

[6.2. Data Sources 3](#_Toc192515099)

[6.2.1. Sources 4](#_Toc192515100)

[6.2.2. Credentials 5](#_Toc192515101)

[6.2.3. Volume 6](#_Toc192515102)

[6.2.4. Refresh 6](#_Toc192515103)

[6.3. Data Connections 6](#_Toc192515104)

[6.3.1. Direct 7](#_Toc192515105)

[6.3.2. Gateway 8](#_Toc192515106)

[6.3.3. Cloud 9](#_Toc192515107)

[6.4. Dataflows 10](#_Toc192515108)

[6.5. Semantic Model 10](#_Toc192515109)

# Introduction

*… see sample fragment 01 …*

# Scope of Work

*… see sample fragment 01 …*

# Workflow

*… see sample fragment 02 …*

# Issues

*… see sample fragment 02 …*

# Business Rules

*… see sample fragment 02 …*

# Data

## Environments

*Describe the different separate environments to be used in the project. (For example, in a 3-tier layout, separate instances of the source application(s), databases, and target report sharing platform will be available and used [DEV, TEST, PROD]. In a 4-tier layout, the environments may be named DEV, TEST, STAGE, and PROD.)*

*Note whether dedicated source applications are available in each environment, or if it is necessary, for example, to use the PROD environment for development and acceptance testing.*

*If separate instances of source applications are available, note the data characteristics and whether they are representative of the PROD environment (e.g., network performance, data security, data volume, data variability, data refresh, etc.).*

## Data Sources

*Describe the characteristics of accessing the data for the report(s) including the source systems, access credentials, volume of data (and available level of detail [or granularity]), and the frequency of update.*

*Operational systems will ideally not be used as data sources for direct reporting purposes; rather dedicated data warehouses or replicas will hopefully be the sources from which to do reporting.*

*If operational systems are in fact the reporting data source, then the performance impact of the data transfer processes must be acceptable. One possible approach to minimize the impact is to do all reporting on* ***yesterday's*** *data (i.e., only refresh once per day, after hours, when performance impacts are acceptable).*

### Sources

*Assign an [ID] to each data source, describe it, and use the [Status] column to note if it is new or existing (e.g., EXISTING, NEW, IN DEVELOPMENT, APPROVED (by whom), VERIFIED (by whom), etc.) and date (use a consistent and unambiguous date format [e.g., yyyy-mm-dd, dd-mmm-yyyy, etc.]).*

*Include all data sources that will be used in the data flows and semantic models serving the reports that are existing and those included in this document.*

| **ID** | **Name** | **Description** | **Status** |
| --- | --- | --- | --- |
| *S-1* | *Customer Data Source* | *Microsoft Dynamics (on-premises)* | *NEW (2025-02-14):*   1. *A request was submitted to [Specific Person Name] in the IT department for implementation of a new Power BI Gateway and service account to access all customer data*   *IN DEVELOPMENT (2025-02-28):*   1. *The [Specific Name] gateway is currently being implemented with the [Specific Name] service account to access all customer data in the DEV environment* 2. *The gateways have not been implemented yet nor service accounts created for the TEST and PROD environments* |
| *S-2* | *Invoice Data Source* | *SAP Financials (on-premises)* | *EXISTING (2025-02-27):*   1. *The [Specific Gateway Name] Power BI Gateway can be used to access to all financial data using the [Specific Service Account] account credentials in the PROD environment* 2. *There are currently no instances of SAP Financials in the DEV or TEST environments, nor are there any plans to implement them* |

### Credentials

*Assign an [ID] to each credential set that will be used to access each data source in each environment and the IP address/URL, account name, and password that will be used to connect to it, ideally service accounts created and used exclusively for data access by other services.*

*NOTE: These credentials should refer to* ***system*** *accounts and not personal accounts, as personal accounts can be subject to changes by others outside of the reporting system without notice (e.g., reassignment, promotion/demotion, retirement, leaving, etc.).*

| **ID** | **Name** | **Environment** | **Credentials** |
| --- | --- | --- | --- |
| *C-11* | *Customer Data Source Credentials* | *DEV* | *IP/URL: 192.1.1.1*  *Account name: serviceDYN-DEV@abc.ca*  *Password=(retrieve from Keepass)*   * *Entry=serviceDYN-DEV* * *Vault=* *Insert Key Vault Name]* * *Secret=[Insert Secret Name]* |
| *C-12* | *Customer Data Source Credentials* | *TEST* | *IP/URL: dynamics-TEST.abc.com*  *Account name: serviceDYN-TEST@abc.ca*  *Password=(retrieve from Keepass)*   * *Entry=serviceDYN-TEST* * *Vault=* *Insert Key Vault Name]* * *Secret=[Insert Secret Name]* |
| *C-13* | *Customer Data Source Credentials* | *PROD* | *IP/URL: dynamics.abc.com*  *Account name: serviceDYN@abc.ca*  *Password=(retrieve from Keepass)*   * *Entry=serviceDYN* * *Vault=* *Insert Key Vault Name]* * *Secret=[Insert Secret Name]* |
| *C-21* | *Invoice Data Source Credentials* | *DEV* | *IP/URL: 192.2.2.2*  *Account name: serviceSAP-DEV@abc.ca*  *Password=(retrieve from Keepass)*   * *Entry=serviceSAP-DEV* * *Vault=* *Insert Key Vault Name]* * *Secret=[Insert Secret Name]* |
| *C-22* | *Invoice Data Source Credentials* | *TEST* | *IP/URL: sap-TEST.abc.com*  *Account name: serviceSAP-TEST@abc.ca*  *Password=(retrieve from Keepass)*   * *Entry=serviceSAP-TEST* * *Vault=* *Insert Key Vault Name]* * *Secret=[Insert Secret Name]* |
| *C-23* | *Invoice Data Source Credentials* | *PROD* | *IP/URL: sap.abc.com*  *Account name: serviceSAP@abc.ca*  *Password=(retrieve from Keepass)*   * *Entry=serviceSAP* * *Vault=* *Insert Key Vault Name]* * *Secret=[Insert Secret Name]* |

### Volume

*Assign an [ID] to the data source and describe the data volume required to answer the report(s) questions, the grain of the data source and its’ use in the report(s), and the aggregation that will be applied to each.*

| **ID** | **Name** | **Volume** | **Grain/Aggregation** |
| --- | --- | --- | --- |
| *V-1* | *Customer Data Refresh* | *Monthly* | *Source/G=n/a*  *Source/A=none*  *Report/G=n/a*  *Report/A=none* |
| *V-2* | *Invoice Data Refresh* | *Daily* | *Source/G=millisecond*  *Source/A=none*  *Report/G=day*  *Report/A=rollup to day* |

### Refresh

*Assign an [ID] to the refresh of the data source and describe the frequency, procedure, target, monitoring, and responsibility for each.*

| **ID** | **Name** | **Frequency** | **Procedure / Target /**  **Monitoring / Responsibility** |
| --- | --- | --- | --- |
| *R-1* | *Customer Data Refresh* | *Monthly* | *P=Manual via Excel file*  *T=Dedicated SharePoint folder*  *M=Sue Jones (IT department)*  *R=Jane Doe (sales department)* |
| *R-2* | *Invoice Data Refresh* | *Daily* | *P=Automatic via scheduled dataflow*  *T=Dedicated SharePoint folder*  *M=Sue Jones (IT department)*  *R=John Smith (accounting department)* |

## Data Connections

Power BI can connect to data in 3 ways:

* Direct
* Gateway
* Cloud connections

Direct connections are internal sources you can see from your computer (i.e., those you can connect to directly)

* File sources (e.g., Excel, CSV, Text, JSON, OneDrive, SharePoint, etc.)
* Database sources (e.g., SQL Server, Oracle, MySQL, IBM DB2, IBM Netezza, etc.)

Gateway connections are internal sources you cannot see from your computer (i.e., those you cannot connect to without a “bridge”)

* On-premises data sources that are inaccessible to the internet (e.g., behind a firewall on a network drive, etc.)
* Use internally stored and managed credentials

Cloud Connections are external (Internet, or Web) sources hosted on cloud platforms

* Cloud sources (e.g., Azure SQL, Snowflake, Databricks, etc.)
* Use externally stored and managed credentials

### Direct

*Describe each internal data connection that will be directly accessed, including ID, type, location, folder, responsibility, and filename.*

Here’s an example:

| **ID** | **Name** | **Description** | **Type / Location / Folder / Responsibility / Filename** |
| --- | --- | --- | --- |
| *CD-1* | *Aging Buckets* | *Excel worksheet with five buckets for categorizing invoice data*   * *0-30 days* * *31-60 days* * *61-90 days* * *91-180 days* * *181 days and over*   *To be used as the data source for the [Aging] supporting table* | *T=Excel*  *L=SharePoint*  *Folder=Accounting/Master Data/*  *R=Jennifer Smith/Accounting*  *Filename=Invoice Aging.xlsx* |

### Gateway

*Describe any gateways that will allow indirect connection to on-premises data sources such that they can be accessed by the Power BI Service, including ID, type (new, existing, standby), and status (online, offline, etc.), environment, availability, use, responsibility, verification, and any other notes.*

*Ideally all gateways will use service accounts with full permission to read all source records (i.e., no security restrictions).*

*(For all gateways, describe the environment, availability [clustered; standalone] and use [e.g., dedicated to a single gateway; shared by multiple gateways; used by other services, etc.]).*

*(For any standby gateways, describe the required takeover verification intervals and any characteristics that differ from the primary gateway.)*

Here’s an example:

| **ID** | **Name** | **Description** | **Type / Status / Environment / Availability / Use /**  **Responsibility / Verification / Other** |
| --- | --- | --- | --- |
| *CG-1* | *On-premises HR System (Sole)* | *Existing Power BI Gateway already permitting read access by the Power BI Service to Customer data*   * *(List of tables exposed and time horizon for each)* * *Customers* * *Contacts* * *Addresses (current)* * *Addresses (historical)* * *Regions* | *T=Existing*  *S=Online*  *E=PROD (no DEV or TEST gateway)*  *A=Power Platform Gateway Cluster GC-1)*  *U=Dedicated/single*  *R=centrally monitored and managed by corporate tenant administrators*  *V=n/a*  *O=none* |
| *CG-2* | *On-premises Financial System (Primary)* | *New Power BI Gateway permitting read access by the Power BI Service*   * *(List of tables exposed and time horizon for each)* * *Invoices (current fiscal year)* * *Invoices (previous fiscal year)* * *Invoices (historical [that is, before previous fiscal year])* * *Payments (current fiscal year, previous fiscal year)* | *T=New*  *S=Offline*  *E=PROD (no DEV or TEST gateway)*  *A=Standalone*  *U=Shared/multiple*  *R=locally monitored and managed by accounting department*  *V=n/a*  *O=none* |
| *CG-3* | *On-premises Financial System (Standby)* | *Standby Power BI Gateway to provide the same access as the primary financial system Power BI Gateway (CG-2)* | *T=Standby*  *S=Offline*  *E=PROD (no DEV or TEST gateway)*  *A=Standalone*  *U=Shared/multiple (common standby gateway that provides failover protection for multiple primary gateways)*  *R=locally monitored and managed by accounting department*  *V=monthly for 24 hours (on first Monday on or after the 15th of the month; full takeover [disable primary, enable standby])*  *O=memory (16 GB vs. 256 GB for CG-2)* |

### Cloud

Describe all details necessary to configure the connection, including type, credential type, credentials, authentication method, etc.

Here are a few best practices to consider:

* Use shareable connections
  + Support multiple connection to the same data source
  + Security and permissions can be adjusted to suit:
    - Can assign different tables their own separate connections
    - Can grant User permission to enable others to use the connection to bind to the data source
* Use service principals (preferred) or service accounts (good) rather than personal accounts (poor)
  + Service Principals: No Power BI license needed, centralized security, governance
  + Service Accounts: Power BI license needed, distributed security (credentials must be shared), no-notice password expiry
  + Personal Accounts: Power BI license needed, security managed by others, no-notice credential activation and permission changes (e.g., reassignment, promotion/demotion, retirement, leaving, etc.)

## Dataflows

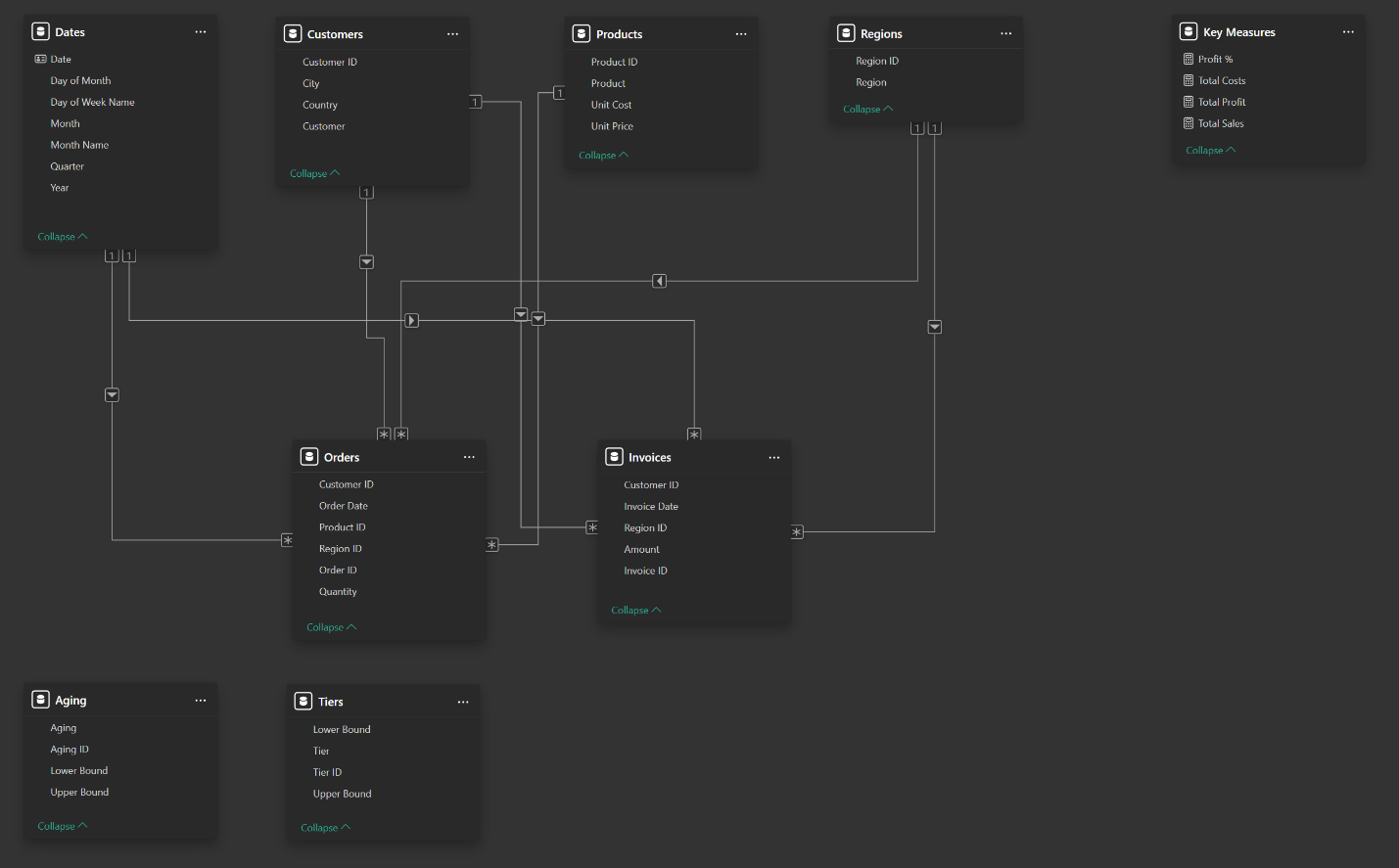
*Describe the design of all dataflows that will be developed for this (series) of reports. Include the type (e.g., staging, fact, dimension, supporting, etc.) and all filters and transformations that will be applied to the dataflow*

*NOTE: Staging tables are ideally used for connection and data transfer only and should contain no transformations.*

| **ID** | **Name (Example)** | **Type** | **Source / Filters / Transformations** |
| --- | --- | --- | --- |
| *D-1* | *RAW Customers* | *Staging* | *S=operational system*  *F=none*  *T=none* |
| *D-2* | *RAW Invoices* | *Staging* | *S=operational system*  *F=only last 2 fiscal years*  *T=none* |
| *D-3* | *Customers* | *Dimension* | *S=reference of D-1 (RAW Customers)*  *F=none*  *T=*   * *columns renamed to Proper Case* * *column data types verified* |
| *D-4* | *Invoices* | *Fact* | *S=reference of D-2 (RAW Invoices)*  *F=none*  *T=*   * *columns renamed to Proper Case* * *column data types verified* |
| *D-5* | *Regions* | *Dimension* | *S=reference of D-1 (RAW Customers)*  *F=none*  *T=*   * *columns renamed to Proper Case* * *column data types verified* * *duplicates removed* * *index added* |
| *D-6* | *Aging Buckets (0-30d, 31-60d, 61-90d, 91-180d, 181d+)* | *Supporting* | *S=Excel spreadsheet*  *F=none*  *T=none* |

## Semantic Model

*Describe the design of the semantic model, including all fact tables, dimension (or lookup) tables, and supporting tables.*

**

*Assign an [ID] to each relationship and fully describe it, including the fields/columns that will be used to link tables, the cardinality, and the directionality of each relationship.*

*NOTE: This is the design intent, not the implementation; the DAX INFO functions will be used in the appendices [a subsequent issue] to extract the actual relationships from the developed model.*

| **ID** | **From (table[column])** | **To (table[column])** | **Cardinality / Directionality** |
| --- | --- | --- | --- |
| *CR-1* | *Dates[Date]* | *Invoices[Date]* | *C=one-to-many*  *D=single* |
| *CR-2* | *Customers[Customer ID]* | *Invoices[Customer ID]* | *C=one-to-many*  *D=single* |

*-- end of fragment*