EASYNEO

LEAD WITH DATA™ Qlik

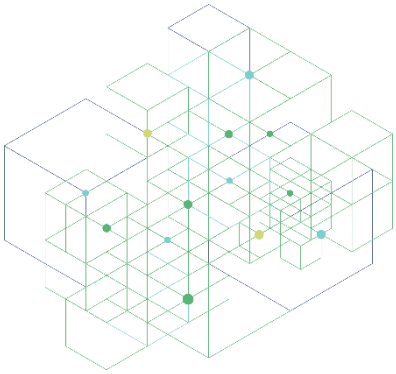
Mainframe Modernization

Unlock the value of your Mainframe data, save time and reduce costs with Qlik Data Integration Platform

WHITE PAPER

Contents

Summary



- Unlock the value in your Mainframe data and accelerate your data analytics
- Automate the creation of analytics-ready data structures with continuous data delivery
- Minimize the impact and cost to replicate from key production systems

Qlik Data Integration Platform

As organizations look to modernize their analytics environments to enable digital transformation, they are embracing a DataOps approach, which requires IT and business alignment along with a modern data strategy and architecture. Qlik's Data Integration Platform accelerates the discovery and availability of analytics-ready data by automating real-time data streaming, refinement, cataloging and publishing. It enables DataOps and drives agility in the analytics process through automated data pipelines that stream from transactional systems, warehouses, or data lakes to create actionable data on-demand.

- **Real-Time Data for Faster, Better Insights**
  Using Change Data Capture (CDC) technology, Qlik's Data Integration Platform can intelligently deliver data from a wide range of systems to ensure users are always analyzing the latest information.

- **Agile Data Delivery**
  To fulfill ever-changing business needs and meet the demand for increasingly diverse data sets, Qlik's Data Integration Platform automates the creation of analytics-ready data pipelines to empower users to operate at the speed of business

- **Trusted, Enterprise-Ready Data**
  Business users need to quickly find and be confident that the data they analyze is accurate, safe and verifiable. Qlik's Data Integration Platform includes a robust set of enterprise-scale quality, governance and collaboration capabilities to streamline DataOps processes, making sure that all data is completely validated, protected and secured from data ingest through delivery.

The Qlik Data Integration Platform efficiently delivers large volumes of real-time, analytics-ready data into streaming and cloud platforms, data warehouses, and data lakes.

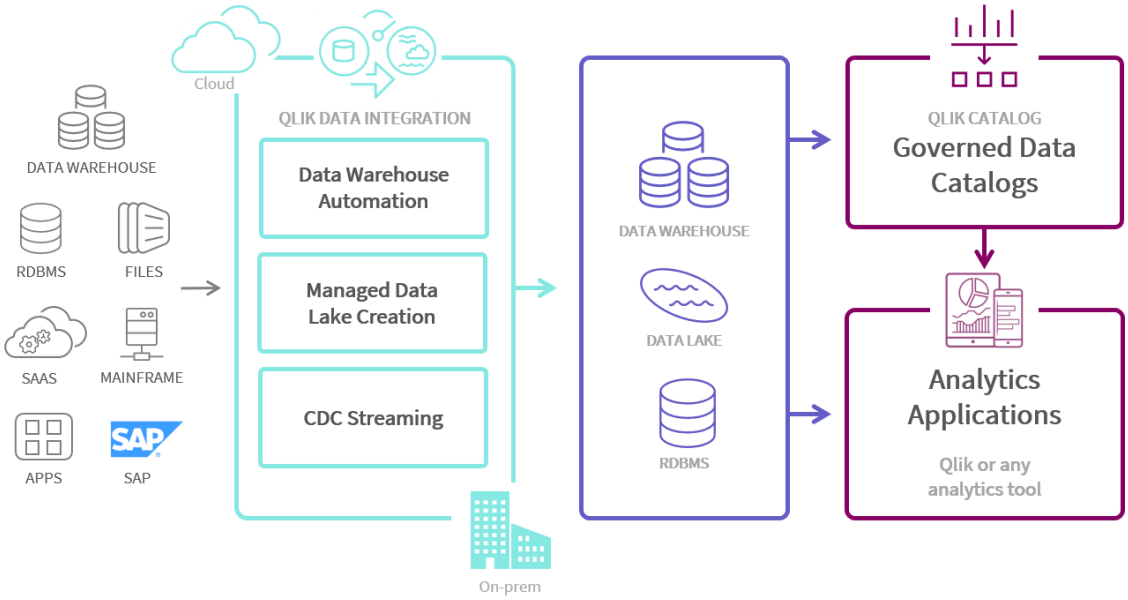


Figure 1

In a hyper-competitive business climate where real-time insights and decisions are critical, the need for agile analytics is driving new data architecture and integration requirements. Unlike traditional batch movement and ETL scripting approaches which are slow, inflexible and labor-intensive, Qlik's Data Integration Platform can automate the creation of data streams from core transactional systems, efficiently move it to the cloud and data lakes, and refine it to make it immediately available via an Amazon-like marketplace experience. By quickly delivering data to the user without typical business friction, Qlik's Data Integration Platform enables the agility necessary to drive greater business value.

Learn more at ***easyneo.fr/integration.html***

Transforming Mainframe Data with Qlik Data Integration

For many organizations, your Mainframe is mission critical containing massive amounts of core-business data. However there are challenges in accessing that valuable data locked inside your Mainframes and making it accessible in secure and governed ways. Qlik Data Integration can help with:

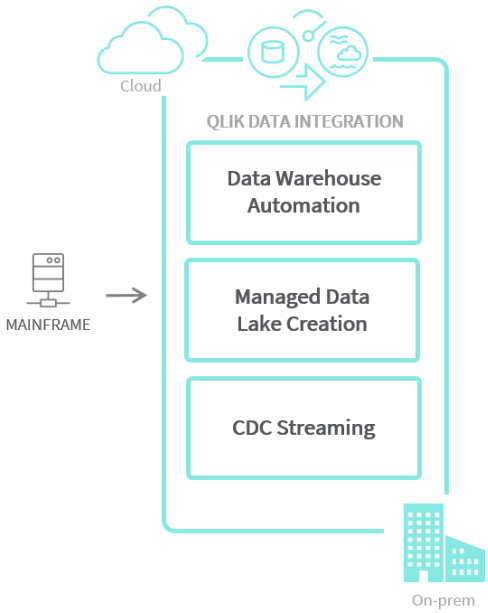


Figure 2

Continuous Replication to Offload Processing the Qlik Way

Because it replicates data continuously and automatically propagates to one or more targets, Qlik Replicate™ cuts the need to move data in periodic batches, keeping that data coordinated and supplying real-time access to the data in other platforms.

Reducing the MIPS Processing Overhead

Qlik Replicate supplies log-based CDC for DB2 as well as for VSAM and IMS, rather than performing repeated brute-force queries into the data. This supplies a minimal impact approach to capturing changes which does not incur the hefty MIPS (Millions of Instructions Per Second) price tag that results from direct querying.

Optimizing Data Transfer to Cloud

Qlik Replicate supports direct endpoints (connectors) to all the major cloud platforms. Each endpoint is optimized for the cloud environment that it supports. When near real-time performance is critical, multiple Qlik Replicate instances can be leveraged (e.g. one on-prem, one in the cloud).

Efficiently Capture Once, Deliver to Many Targets

Qlik Replicate provides a unique Log Stream capability that saves data changes from the transaction log of a single source database and enables them to be applied to multiple targets, without the overhead of reading the logs for each target separately.

Automating the data pipeline for faster time to insights

The Qlik Data Integration platform saves data engineers valuable time by enabling them to automate the availability of accurate, trusted data sets and transform it to analytics-ready data for their business, automating the entire data warehouse lifecycle and the creation of managed data lakes without coding.

Overcoming Challenges of Traditional ETL Solutions

- **Batch File Transfer** - *involving the need for complex custom code that takes a lot of time and resources to maintain and run to create large flat files that often are not fresh*
- **Direct Database Query** - *a brute force approach, querying directly into the Mainframe system to access the data, increasing the cost of your Mainframe MIPS bill. It negatively impacts production systems and the overall end-user experience dramatically*
- **Mainframe Real-Time Data Streaming** - *without the correct tools a significant amount of manual tuning and optimization is required to support broad, deep, and fast analysis demanded by enterprises today, taking away your precious time from other essential business tasks.*

The rest of this section looks at Qlik Replicate and Mainframe source endpoints for DB2 on z/OS and iSeries, IMS and VSAM examining the architecture, metadata and some things to consider.

Qlik Replicate for DB2 on z/OS

After a short setup on the Mainframe, which establishes a special DB2 User Defined Table Function (UDTF) along with required security grants, you can manage all replication design and execution tasks through a modern and friendly web user interface served from a Linux or Windows server machine.

The Qlik Replicate Mainframe component is lightweight in size (less than 300KB) and enables DB2 transaction log records to be read from a Qlik Replicate task. All parsing and transaction sorting, operations that take CPU and memory, occur on the Qlik Replicate platform. This contrasts with alternative solutions that perform the parsing and sorting on the Mainframe side which involves significantly more administrative and increases processing resources (MIPS, storage) on the Mainframe.

Architecture

The architecture in figure 3 describes the use of IBM DB2 for z/OS as a source endpoint within Qlik Replicate:

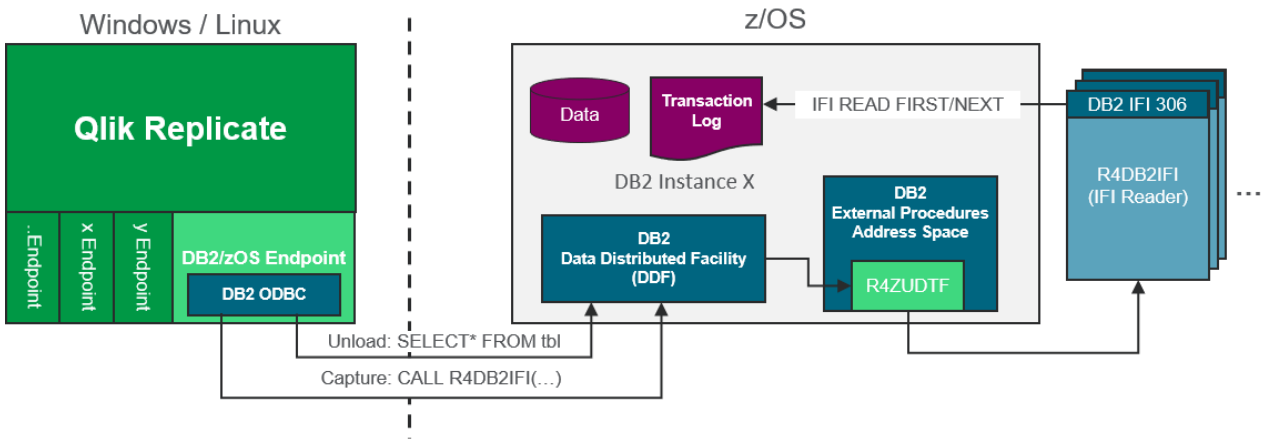


Figure 3

When working with DB2 on z/OS as a source endpoint, connectivity to DB2 is achieved through ODBC/CLI from the Qlik Replicate server. A small User Defined Table Function (UDTF) will be installed in DB2 for the purpose of capturing change data (transaction logs). Everything is performed through one or more ODBC connections using IBM's DB2 client driver "IBM Data Server Driver for ODBC and CLI" for DB2.

Metadata

Qlik Replicate uses a “task” to define a unit of work. Each task definition includes a reference to one or more DB2 tables. The DB2 catalog is queried interactively to present the designer of a Qlik Replicate task with the table selection. Wild cards are supported, making the inclusion of entire schemas or subsets of tables easy.

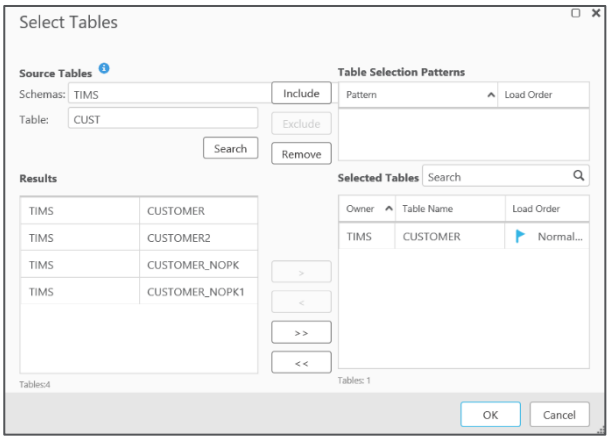


Figure 4

Things to Consider

ODBC/CLI client access to DB2 on z/OS may include additional licensing requirements through IBM. Please refer to the following IBM Knowledge Center link for details:

www.ibm.com/support/knowledgecenter/en/SSEPGG_11.1.0/com.ibm.db2.luw.apdv.cli.doc/doc/r0024162.html

Qlik Replicate for DB2 on iSeries

Architecture

The architecture in figure 5 describes the use of IBM DB2 for iSeries as a source endpoint within Qlik Replicate:

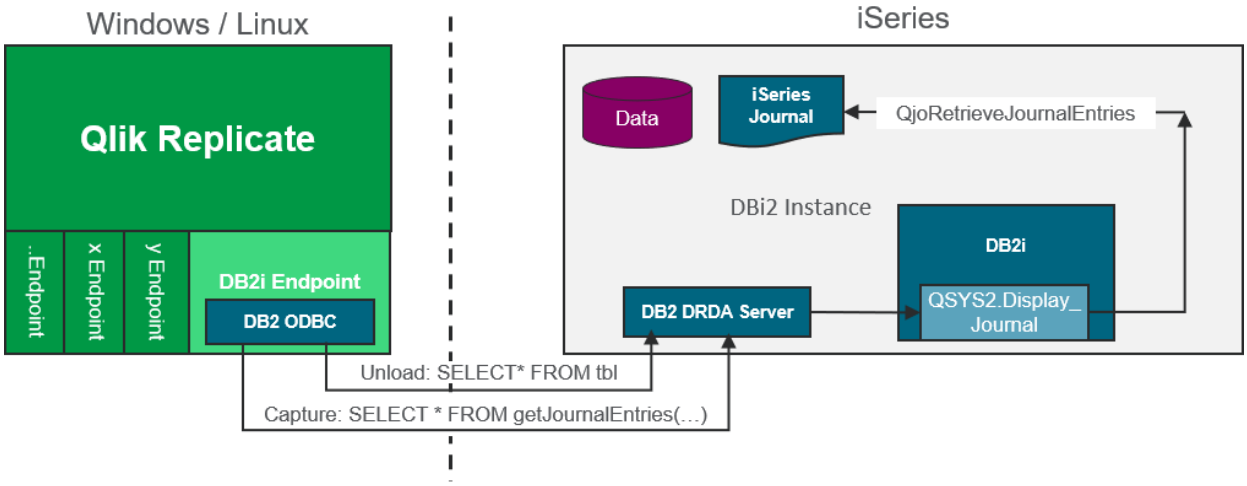


Figure 5

When working with DB2 on iSeries, Qlik Replicate uses the IBM ODBC drivers from the IBM i Access Client Solution package. Once configured Qlik Replicate will use ODBC connections for full load and change data capture. Qlik Replicate also uses the built-in Display_Journal UDTF for the purpose of receiving journal (log) data for change data capture purposes.

Metadata

As with all relational endpoints the source database catalog is queried interactively to allow the Qlik Replicate task designer to include tables in the task definition.

Things to Consider

You may need to download and install the IBM i Access Client Solution from IBM Entitled System Support (ESS).

Relative Record Number (RRN) is supported and can be used as the table's primary key.

Both Long names and System names are supported.

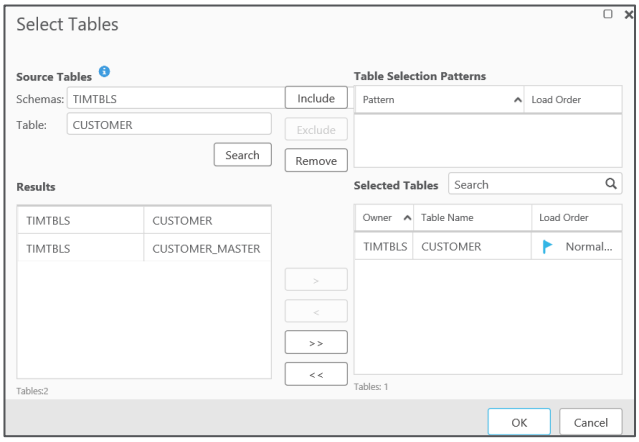


Figure 6

Qlik Replicate for VSAM and IMS on z/OS

The architecture in figure 7 describes the use of IBM ISM/VSAM for z/OS as a source endpoint within Qlik Replicate:

Architecture

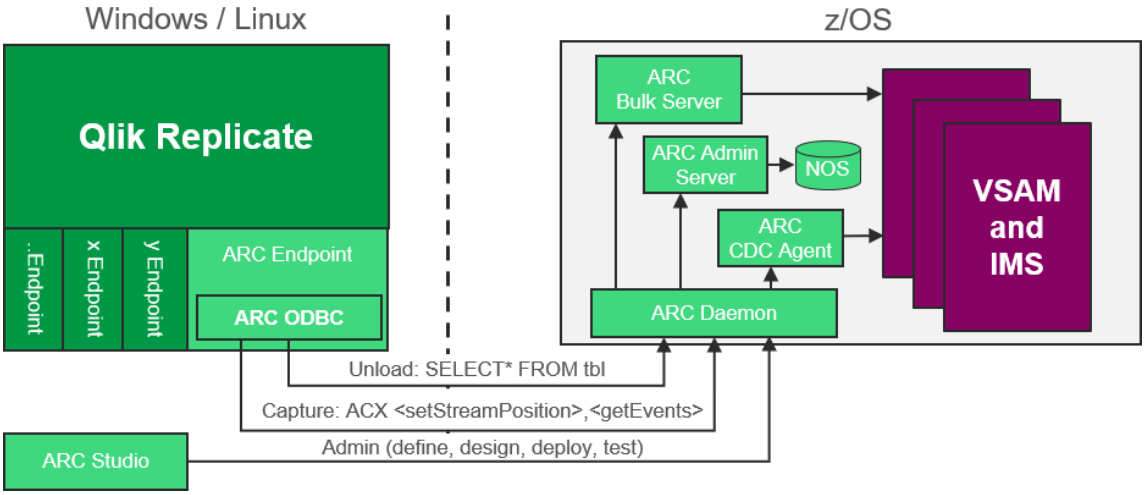


Figure 7

In the case of VSAM and IMS on z/OS Qlik Replicate utilizes an agent-based architecture through the installation of a Qlik component, Replicate Connect (ARC), on the z/OS platform. The Qlik Replicate ARC endpoint uses its own ODBC driver to connect to ARC. ARC in turn will read the VSAM and IMS

datasets when performing full load. For change data capture (CDC) it leverages VSAM and IMS user exits, configured in advance for this purpose.

Metadata

One of the goals of the metadata layer in Qlik Replicate is to achieve a relational model for these inherently non-relational sources of data. The ARC Studio is a design-time component, that includes import wizards for VSAM and for IMS data sources. In both cases COBOL copybooks are expected and in the case of IMS the Database Descriptor (DBD) and Program Specification Block (PSB) definitions are also used.

The wizards support various models including sequential flattening (VSAM and IMS), parent/child relationships (IMS), and virtual tables (VSAM and IMS). The latter is particularly powerful when considering the handling of OCCURS since the OCCURS structure itself can be mapped relationally to appear as a child table to the parent.

When flattening would result in too many columns in the table for the replication target or overly complicated SQL processing of the flattened data (aggregation for example) then virtual table mapping is a great choice. Additional copybook details including redefinitions and conditional names are also handled during the import process.

Things to Consider

Best practice is that copybooks and related supporting files should contain an accurate description of the underlying data. However, ARC Studio does provide a testing feature and dynamic changes to the metadata can be made in cases where the underlying data structure has changed.

Data quality should be considered and evaluated in advance when considering IMS and VSAM data sources as part of a data replication project.

Qlik Catalog: Mainframe Democratization

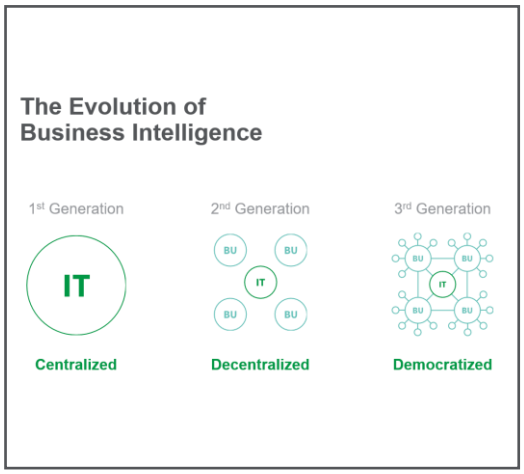


Figure 8

Democratization of Data is the ability to have all data, including combinations of this data, accessible to all users within a governance framework that provides security without limiting agility. Data has limited value if only a few, highly technical people can access, understand and utilize the data source.

Qlik Catalog™ (formerly Qlik Data Catalyst) offers a secure, enterprise-scale repository and catalog of all the data assets your business has available for analytics. This gives your data consumers a single, go-to destination to find, understand, and gain insights from all enterprise data sources. It includes data preparation and metadata tools that streamline the transformation of raw data into analytics-ready assets, and a Smart Data Catalog that helps people easily discover and choose whatever data they need. Built on a platform with hardened data security and governance, Qlik Catalog can be easily integrated with other data management tools for enterprise-grade scalability, reliability, and performance. Qlik Catalog can be deployed on premises or in the cloud and includes publishing directly to your analytics product of choice, including Qlik Sense.

Qlik Catalog for DB2 on z/OS, DB2 on iSeries, and Other JDBC Data Sources

Any Mainframe data source that provides JDBC access can participate as a data source within the Qlik Catalog framework. This “universal” connector is easily leveraged for DB2 on z/OS as the preferred approach to relational Mainframe data. Likewise, for the iSeries platform JDBC access to DB2/i brings the power of democratization directly to the database. IBM and third-party vendors provide other JDBC connectors and drivers for popular data sources on these platforms making it possible for additional data sources to be considered.

Architecture

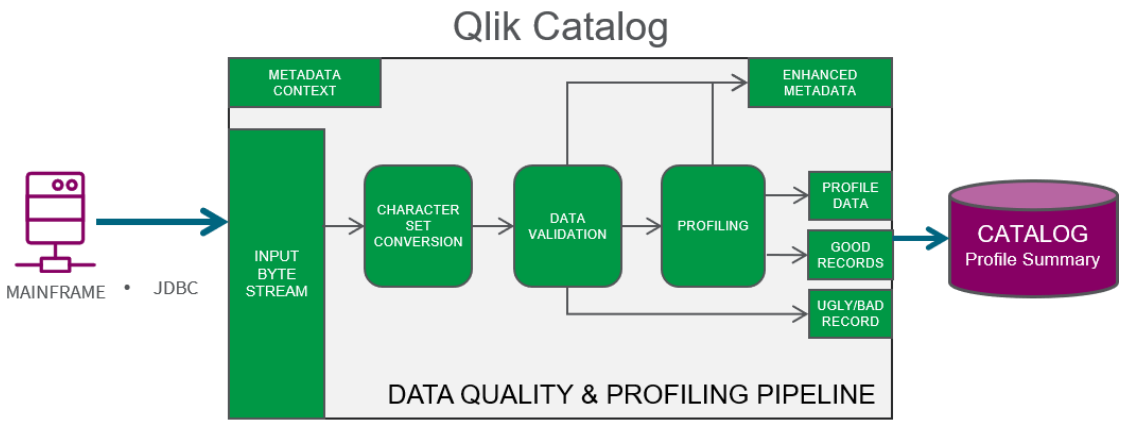


Figure 9

Metadata

As part of the standard JDBC environment Qlik Catalog can dynamically present the source catalog in a way that allows for selection of tables and data.

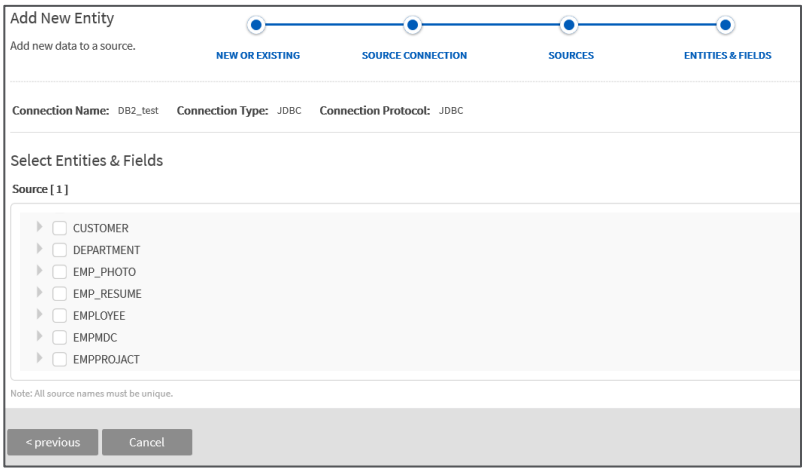


Figure 10

Qlik Catalog for Non-relational Data on z/OS

The primary method for accessing non-relational (flat) data on z/OS is by way of files. Files can be moved via FTP or shared with Qlik Catalog via NFS or any other means supported between the Qlik Catalog server and the Mainframe for file sharing and access.

Architecture

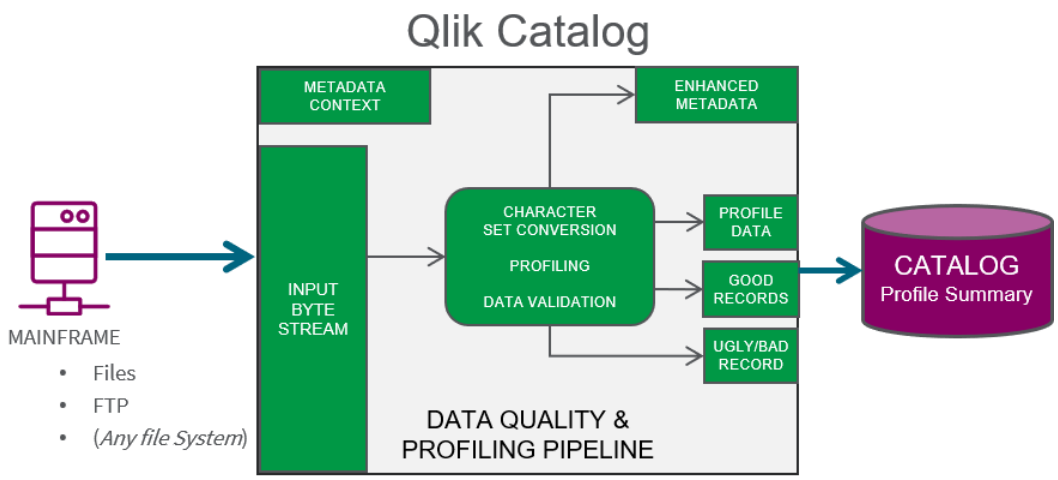


Figure 11

Metadata

Qlik Catalog expects an EBCDIC file and a COBOL copybook when using Mainframe files as a source. Extraction from the file will transform the data according to specifications in the copybook.

The CUSTOMERS record is shown as defined by a copybook (figure 12), an EBCDIC file (figure 13), a JDBC view (figure 14) and finally how it looks in Qlik Catalog (figure15).

COBOL copybook

```
01 CUSTOMERS.
        05 CUST_KEY              PIC 9(4).
        05 NAME                  PIC X(25).
        05 ADDRESS               PIC X(40).
        05 NATION_KEY    PIC 9(4).
        05 PHONE                 PIC X(15).
        05 ACCT_BAL              PIC 9(8).
        05 MKT_SEGMENT   PIC X(10).
        05 COMMENT               PIC X(114).
```

Figure 12

EBCDIC file

```
00000000h: F0 F0 F0 F1 D1 96 88 95 40 C6 99 89 85 84 94 81 | 0001John Friedma
00000010h: 95 40 40 40 40 40 40 40 40 40 40 40 40 F1 40 D9 | n           1 R
00000020h: 96 84 85 96 40 C4 99 89 A5 85 40 40 40 40 40 40 | odeo Drive
00000030h: 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 |
00000040h: 40 40 40 40 40 F0 F0 F1 F5 F2 F5 60 F9 F8 F9 60 |     001525-989-
00000050h: F7 F4 F1 60 F2 F9 F8 F8 F0 F0 F0 F0 F0 F7 F1 F2 | 741-298800000712
00000060h: C2 E4 C9 D3 C4 C9 D5 C7 40 40 D7 99 85 94 89 85 | BUILDING  Premie
00000070h: 99 40 83 A4 A2 A3 96 94 85 99 40 40 40 40 40 40 | r customer
00000080h: 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 |
00000090h: 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 |
000000a0h: 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 |
000000b0h: 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 |
000000c0h: 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 |
000000d0h: 40 40 40 40 40 40 40 40 40 40 40 40 F0 F0 F0 F2 |                0002
```

Figure 13

JDBC view

TABLE_NAME	COLUMN_NAME	DATA_TYPE	TYPE_NAME	PRECISION	LENGTH	SCALE
CUSTOMERS	CUST_KEY	2	NUMERIC	4	6	0
CUSTOMERS	NAME	1	CHAR	25	25	(null)
CUSTOMERS	ADDRESS	1	CHAR	40	40	(null)
CUSTOMERS	NATION_KEY	2	NUMERIC	4	6	0
CUSTOMERS	PHONE	1	CHAR	15	15	(null)
CUSTOMERS	ACCT_BAL	2	NUMERIC	8	10	0
CUSTOMERS	MKT_SEGMENT	1	CHAR	10	10	(null)
CUSTOMERS	COMMENT	1	CHAR	114	114	(null)

Figure 14

Qlik Catalog Discovery View

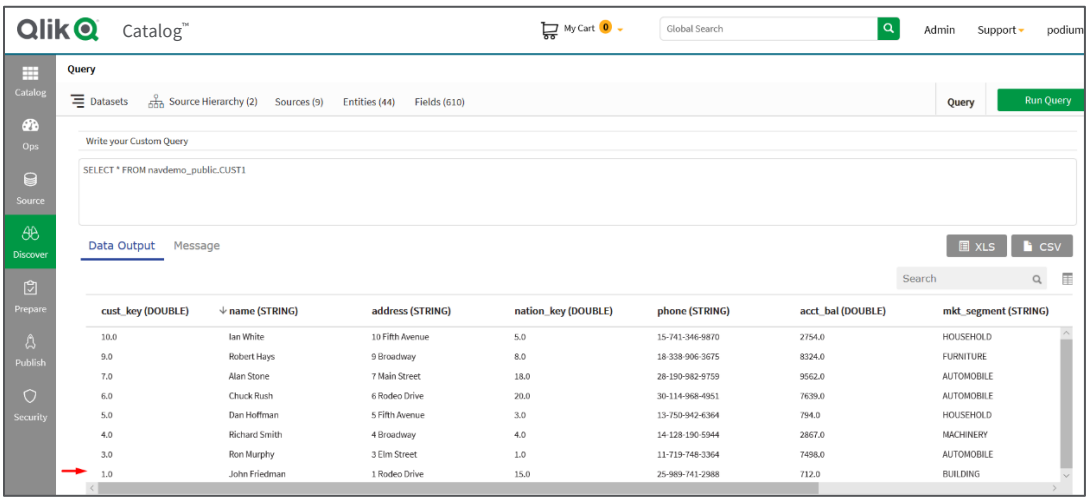


Figure 15

Considerations

When using FTP to share files, the file should be moved in BINARY format.

Conclusion

With the Qlik Data Integration Platform you can:

- Unlock the value of your data in your legacy sources, including DB2 z/OS, IMS and VSAM.
- Enable major data integration projects by ingesting incremental datasets continuously, from many transactional sources, into your data lake and data warehouse environments delivering up-to-date data with enterprise-class log-based change data capture (CDC).
- Increase your business agility and flexibility by aligning IT and business operations and enable “re platforming” by migrating your legacy data to innovative cloud alternatives.
- Minimize the impact to your production systems and reduce costly Mainframe resources by eliminating direct queries and capturing changes once while delivering to multiple targets.
- Save time with an automated, no code approach to data pipeline creation.
- Gain better visibility of your data landscape through enterprise-wide secure and governed catalogs and safely democratize your data across every line of business.
- Reduce back office IT workload and support your line of businesses with the flexibility they need while still maintaining data security.

EASYNEO

À PROPOS DE EASYNEO

EASYNEO vous accompagne par son expertise pour aller plus loin et plus vite grâce aux solutions d'analyse et d'intégration de données de bout en bout. Devenez une entreprise data-driven.

INTÉGRATION DE DONNÉES

Déployez les DataOps pour l'analytics afin de livrer en temps quasi réel des données fiables et prêtes à l'emploi.

ANALYSE DE DONNÉES

Placez les informations exploitables au cœur de chaque décision grâce à la plateforme de BI la plus complète du marché.

Pour en savoir plus,rendez-vous sur easyneo.fr

Qlik LEAD WITH DATA

About Qlik

Qlik's vision is a data-literate world, one where everyone can use data to improve decision-making and solve their most challenging problems. Only Qlik offers end-to-end, real-time data integration and analytics solutions that help organizations access and transform all their data into value. Qlik helps companies lead with data to see more deeply into customer behavior, reinvent business processes, discover new revenue streams, and balance risk and reward. Qlik does business in more than 100 countries and serves over 50,000 customers around the world.