This document describes the procedures required for an authorised Partner on the myGaru platform to deploy and launch the client-side component of the myGaru Data Clean Room (“DCR Client”). The myGaru Data Clean Room operates on a federated model in which each data vendor’s data is stored within its own premises. Accordingly, the DCR Client must be installed on the Partner’s premises (server or cloud) to preserve the Partner’s full control over its data, including its physical location. Installing and starting the DCR Client enables the Partner to collect first-party signals from the Partner’s digital properties (websites and apps) using myGaru ID, and to onboard first-party datasets gathered outside myGaru into the myGaru Data Clean Room. Use of the myGaru Platform without the DCR Client remains possible but with limited functionality, as the Partner’s customer base and associated behavioural data remain isolated from the platform.

This document covers:

- **Network requirements** for the server on which the DCR Client will be installed.
- **Resource requirements** for that server (CPU, memory, storage, OS/runtime).
- **Deployment procedure** for installing and launching the DCR Client.

1. Introduction

DCR Client is a Linux application, provided as an installation image in the form of Bash script. The script will install:

- DCR Client binary itself
- All required certificates:
  - Client certificate for Mutual TLS Authentication
  - Certificate for Data Manager UI
  - Certificate for Signals Collection API
- configuration and runtime management files
- systemd service file to run DCR Client automatically upon server restart
- co-located PostgreSQL database for storing first-party data

It is tested on a Debian-based (Debian, Ubuntu) or Redhat-based (Oracle, Rocky, Alma) distributions.

DCR Client itself is a stateless application - it stores all the data and states in the PostgreSQL database; it can be restarted anytime and will continue to work.

The DCR Client’s resource requirements depend on first-party data size, but there are no special requirements to the hardware - DCR Client can run as a Virtual Machine. In most cases it can be even better because if both DCR Client and co-located PostgreSQL will use the same shared block storage, then any hardware failure will just trigger start of the applications on another host, preserving the data in the database.

PostgreSQL note

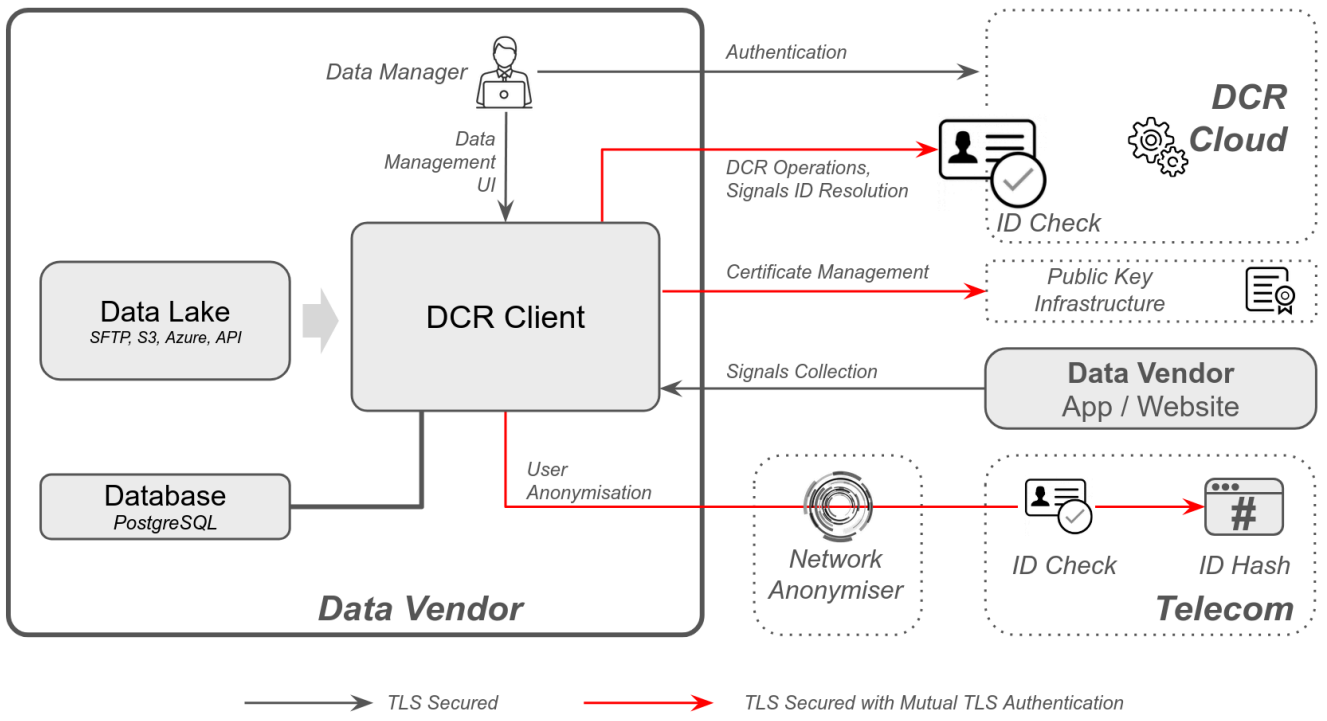
As said above, installation script will install PostgreSQL locally, co-locating it with the DCR Client, but:

- It does not matter, where PostgreSQL will be installed. Any corporate (if present) PostgreSQL instance / cluster can be used as the backend for the DCR Client - the database connection configuration stored in the /opt/dcr/etc/base.ini file.
- It is **not** a myGaru responsibility to ensure PostgreSQL high availability or backup strategy - it highly depends on the corporate infrastructure and procedures and cannot be covered by generic installation procedure.
- If Data Vendor will choose to use other PostgreSQL instance / cluster, then locally installed (co-located) one can be safely removed.

While myGaru can assist with these topics in all possible ways, in general it is the Partner's responsibility to ensure data will not be lost in case of any failure.

2. Network requirements

The overall network connectivity model is presented on the following scheme:



The following table describes the external connections and ports engaged in DCR Client operations. "Direction" defined from the DCR Client's point of view.

Operation	Direction	DCR Client Port	Remote Side Endpont	Remote Side Port	Auth Type	Comment
DCR Cloud operations, Signals ID Resolution	Outgoing	N/A	DCR Cloud	443	mTLS	Different endpoints used for this, configured in base.ini file
Certificates Management (renewals)	Outgoing	N/A	PKI	443	mTLS	
Users anonymisation	Outgoing	N/A	Telecom	443	mTLS	Few endpoints - one per every Telecom, connected to the Platform
Data Management	Incoming	443	Data Manager workplace	N/A	OAuth	
Signals Collection	Incoming	443	User Browsers Mobile Apps	N/A	N/A	Internet-facing endpoint. When enabled through the portal, must be accessible from the Internet.

Operation	Direction	DCR Client Port	Remote Side Endpont	Remote Side Port	Auth Type	Comment
Load data from Data Lake	Outgoing	N/A	Data Vendor's Storage	22, 443	Defined by Data Vendor	This can be remote SFTP, S3, Azure, etc and security configuration depends. Connection initiator is DCR Client
Database exchange	Outgoing	N/A	PostgreSQL	5432	Defined by Data Vendor	Co-located PostgreSQL by default use local socket and username/password authentication. When chosen external PostgreSQL, configuration will be defined by Data Vendor.
Operating System Maintenance	Outgoing	N/A	OS packages repository	443	N/A	While not shown on the network scheme above, it is a regular operation system maintenance to keep the system stable and safe with security and maintenance updates.

As it was said above, all connections are protected with TLS v1.2+ . Authentication done in few ways:

- Mutual TLS Authentication (mTLS) for all automated calls.
- OAuth for DCR Client Administrator authentication.
- Service-specific authentication for accessing Data Lake and PostgreSQL.

Other notes on connectivity:

- Collection of engagement signals works as a regular Web server, which serves thousands of clients, so no authentication is possible.
- It is better to make Data Management UI available from the Data Vendor perimeter (corporate network including VPN), but since everything is encrypted, this requirement is neither critical nor mandatory.
- To be able to perform operations, the DCR Client needs to have at least 100Mbps network connection.

Before deploying a DCR Client, make sure it is accessible by allowed remote sides and can access allowed remote sides by respective IP addresses / hostnames:

- it is accessible from the Data Manager workplace;
- if can access the Internet in order to:
  - install / update operating system service (depends on selected OS);
  - access myGaru Platform Components (cloud.mygaru.com, ca.mygaru.com, configurator.mygaru.com);
  - access Telecom anonymisation endpoints.

3. Resources requirements

The DCR Client is developed to be used on AMD64-based hardware, either on physical servers or in a virtualised environment under the Linux-based operating system. It is recommended to use so-called LTS releases, which are Ubuntu LTS, Debian, Redhat Enterprise Linux or derivatives (Rocky, Alma or Oracle Linux for example).

For proper functioning, the minimal server parameters are:

- 2C/4T CPU (or 4 vCPU for virtualised environment) at 2+ GHz;

- 8Gb of RAM;
- 100G Solid-State Disk (SSD);
- 100Mbps network connectivity,

but depending on the amount of stored data and other factors, resources requirements can change to bigger.

4. Deployment procedure

Deployment procedure consists of two steps:

1. Download DCR Client software through the DCR Cloud Web portal
2. Copy and install the DCR Client on the target server

Prepare DCR Client installation

To install the DCR Client software, the following steps are required:

- login at the myGaru portal (https://platform.mygaru.com), press “Download DCR Client Software” and follow the on-screen instructions;
- copy the downloaded file to the server, where it will be installed;
- run the downloaded file as user ‘root’.

This will create the user/group ‘dcr’ / ’dcr’ for running the software, install the DCR Client software in the directory /opt/dcr, install and configure PostgreSQL database with all required dependencies.

Check the status of the running DCR Client

One can always check the status of DCR Client, accessing the /status/ page of the Admin UI. The status page will show plenty of diagnostic information including (but not limited to) :

- whether used certificates are configured and valid;
- whether remote network endpoints are accessible;
- whether Datastream API enabled and is accessible from the Public Internet;
- the percentage of available to DCR Client resources

which will help to identify and fix the issues with the DCR Client configuration.