CyberCoach Data Use and Storage

This document describes the data of your organization and people that CyberCoach service contains. There are several data flow diagrams at the end.

CyberCoach is a cyber security awareness service that provides training and advice to your people on related matters. It is implemented with Microsoft Bot Framework on the Azure cloud computing platform. We try to collect as little data about your organization and people as possible to provide the service.

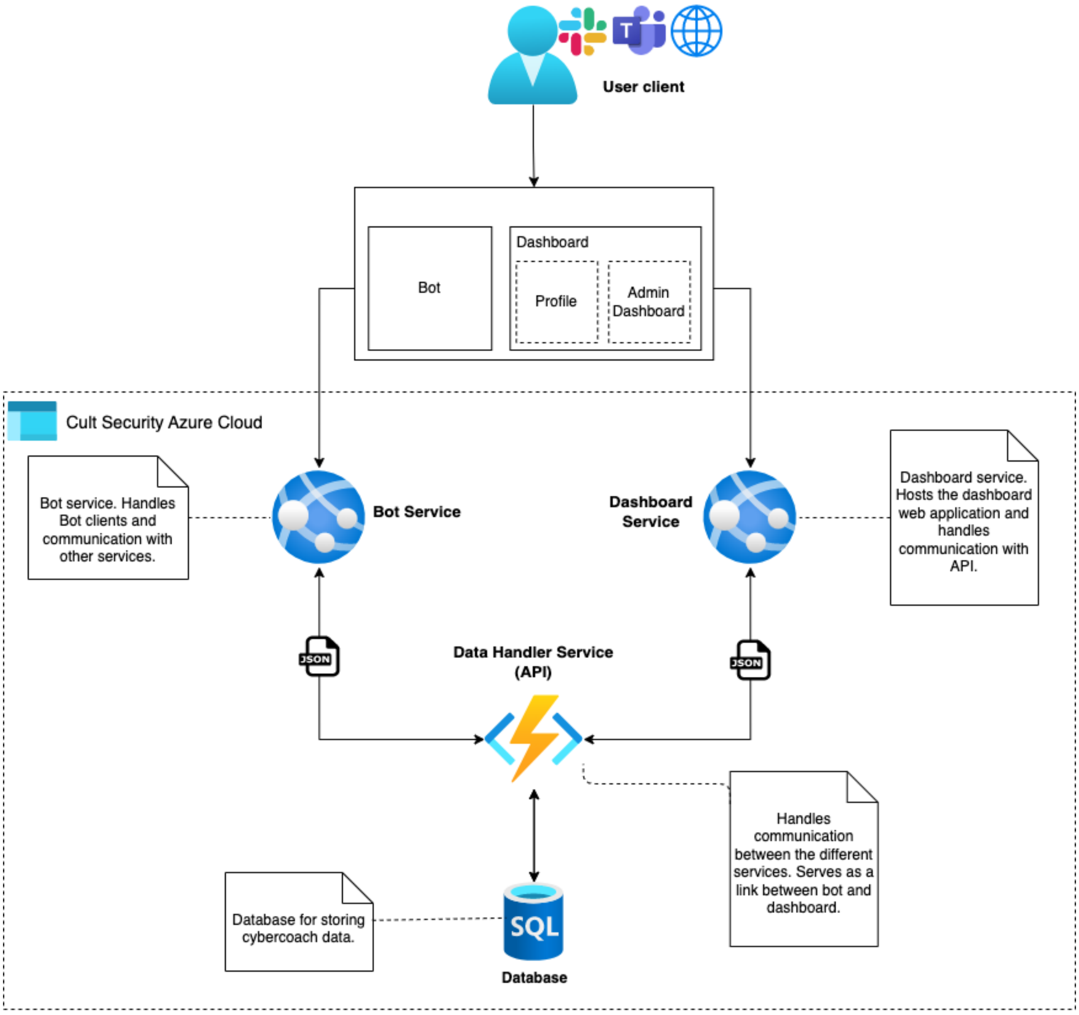

User client
Bot
Dashboard
Profile
Admin Dashboard
Cult Security Azure Cloud
Bot service. Handles Bot clients and communication with other services.
Bot Service
Dashboard Service
Dashboard service. Hosts the dashboard web application and handles communication with API.
JSON
Data Handler Service (API)
JSON
Handles communication between the different services. Serves as a link between bot and dashboard.
Database for storing cybercoach data.
SQL
Database

Figure 1: Logical architecture of CyberCoach

IDENTIFYING YOUR ORGANIZATION

First, we have a lonely unused license in our database. Then we give it to you.

You also receive a way to access our service. This may be:

- Teams app, which you install in your Teams
- Slack app installation link, which you access and choose a Workspace to install it in
- Web interface, made available to you in some manner

When you access the service, we receive your organization's ID in the first request:

- In Teams this is your Tenant ID, ie. abf988bf-86f1-41af-91ab-2d7cd011db46
- In Slack this is your Team ID, ie. T0Q4MKEA2
- On web it depends how identity management has been set up

These are globally unique identifiers (GUID). It should not be possible to resolve the name of your organization, team name, or domain from them.

We do a look up to our database and discover that your organization is new to CyberCoach. This is where your adventure ends, unless you have a valid license key from us.

When you enter the license key, we check its validity. Then we associate the license with your organization's ID. We call this external ID in our database, and give your organization an internal orgId, so that we don't need to store your actual tenant ID or team ID across our database.

IDENTIFYING YOUR PEOPLE

Since this is a new organization, you are also a new user. We get your user ID in the same request:

- In Teams this is of format 1980e8fc-4c8e-46bc-a283-7e54dcd81d91, it's called aadObjectId, and it is unique even across all tenants
- In Slack this is of format D0140HUJUTG, it’s called user ID, and it is only unique to the team (the organization) containing them

- On web it depends how identity management has been set up

We figure out the user is new, and like with organization, we add you to our users table, associate you with your organization, and only store this external ID for a user there. Elsewhere we use our internal userID.

Now the service can be used. When another user sends a request the organization is already there with a valid license, and we just add the user.

At this point what can be known about you by examining our database is:

- You are identified with a particular external ID
- You belong to an organization that is identified with a particular external ID

There's no name, or any contact information, or any way really to find those out, based on these IDs we have.

SERVICE DATA RELATED TO A USER

In addition, there is some service use related data that we store:

- If you select a language for yourself
- If you are associated with a learning path (*)
- If you have completed trainings and agreed to store the results

(*) Learning paths are subsets of the training content we have. For instance, there's a module Secure Development, which continues a training for Cryptography. This is likely not relevant to everyone in your organization, so you might create a "Developers" learning path, and users can be associated with that. By examining our database, one could then perhaps deduct that this particular external ID likely works with something technical. But again, not who this person actually is.

We provide a personal profile dashboard for the user that they can access:

- In Teams, it's a web app embedded in the dashboard tab in the CyberCoach application.
- In Slack, it's our web app made available through using Slack for the identity management

- On web it's like Slack, but access to it depends on how identity management has been set up

This does not add to the data we collect. It visualizes completed trainings and allows the user to choose a learning path and change language.

SERVICE DATA RELATED TO AN ORGANIZATION

We will store a bunch of things about your organization:

Licensing information:

- Your organization's external ID as mentioned above
- Your CyberCoach service license's seat count (user count), and duration
- Whether you use Teams, Slack, or something else (can be deducted from the format of the external ID)

Custom information by your preferences:

- The name you have chosen to be displayed for your organization
- The languages you have chosen to have supported, and the languages your users have selected to use
- What learning paths there are for the service, and how your users a divided between them
- Your cyber security policies and some guidelines in brief snippets, or links to where they can be found
- If added by you, a link to a logo for your organization that is displayed in the app

User and usage information:

- How many users (of this service) there are
- How active your users are in this service
- How many training completions your users have decided to store, and how well they have done
- Users can choose to give feedback and if they do, we save which org ID and user ID it’s from
- Activity logging keeps track of which advice has been accessed and which trainings have been taken on a “started”, “cancelled”, “ended” level of abstraction

- Suspected incidents – when user asks for advice and the dialog results in CyberCoach suggesting filing a cyber security incident report, we log this

EXTERNAL GROUPS LINKING

To support targeted analysis of completed trainings in your organization, CyberCoach associates users with external groups.

- In Teams, this is Microsoft AD groups that users belong to
- In Slack, this is Slack user groups
- On web this is currently not supported

To enable this, we link the user to 1-n groups and store the group belonging in the database. As with the organization ID and user ID we do not store any name of the groups.

- In Teams, we store the GUID of the AD group(s). Format is 02ce3874-dd86-41ba-bddc-013f3401997
- In Slack, we store the ID of the group. Format is S0614TZR7

Retrieving the Microsoft or Slack IDs for the group is enabled by their respective APIs. Microsoft provides Graph API and Slack has its own API. Access is provided by access tokens CyberCoach receives once your user logs in to the service. API is permission scoped to only give access to certain information and can only be used within your organization.

The APIs also enable CyberCoach to retrieve the actual name of groups based on their IDs. This enables the dashboard to render your AD/Slack group names in the service, for your administrator users, without us ever having to store the names in our database.

PROVIDER DATA RELATED TO A USER

Provider data here refers to data CyberCoach doesn’t store but still is present in the user interface. Provider data is also retrieved using either Microsoft or Slack API.

This data is:

- User name
- User profile picture

Through the APIs this data can be visible in the UI, but it is never stored in our database.

ADMINISTRATOR REPORT CREATION

For compliancy and awareness management reasons, it is possible to extract a CSV or Excel report of who has done which training. This comes complete with the names. How?

As described above, we have the external IDs for your users:

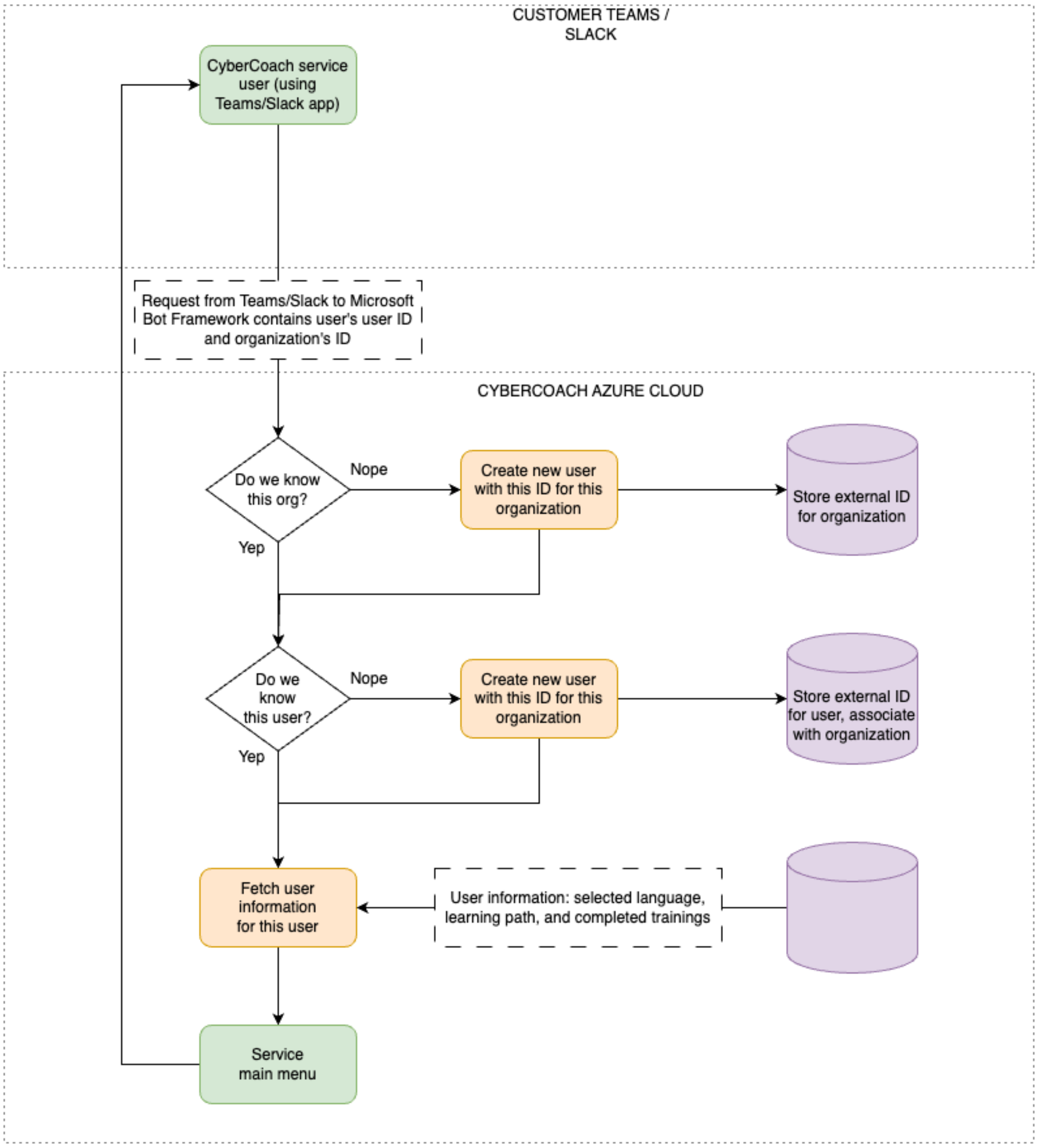
- In Teams, this is of format 1980e8fc-4c8e-46bc-a283-7e54dcd81d91, it's called aadObjectId, and it is unique even across all tenants
- In Slack, this is of format D0140HUJUTG, it’s called user ID, and it is only unique to the team (the organization) containing them
- On web, it depends how identity management has been set up

We need to map these with names to provide you with the report. As described in section “External groups linking” CyberCoach has, through APIs, access to some data stored by either Microsoft or Slack. One of the APIs enables retrieving the name based on external IDs.

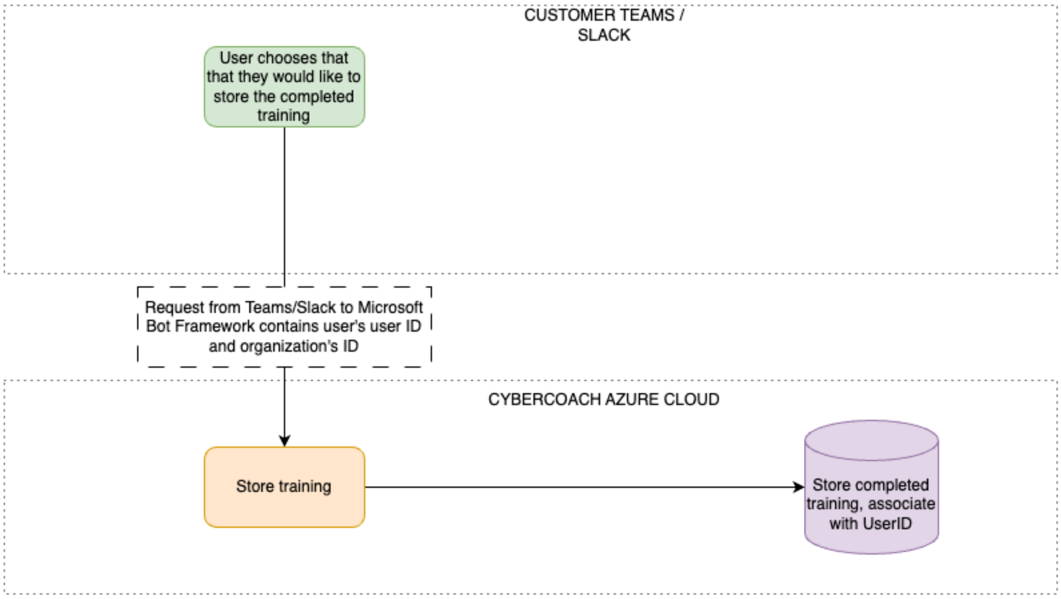
Reports to your administrator users are then generated in the following way:

1. Our database is queried for trainings completed with your given parameters
2. Database returns a list of trainings completed for each external ID in the selection
3. CyberCoach asks the API for real names of the users, using the external IDs
4. Report is generated with the real names along with training completion information
5. Report is provided to your administrator user as CSV
6. We do not store or log the names, and after the process we still only have the external IDs

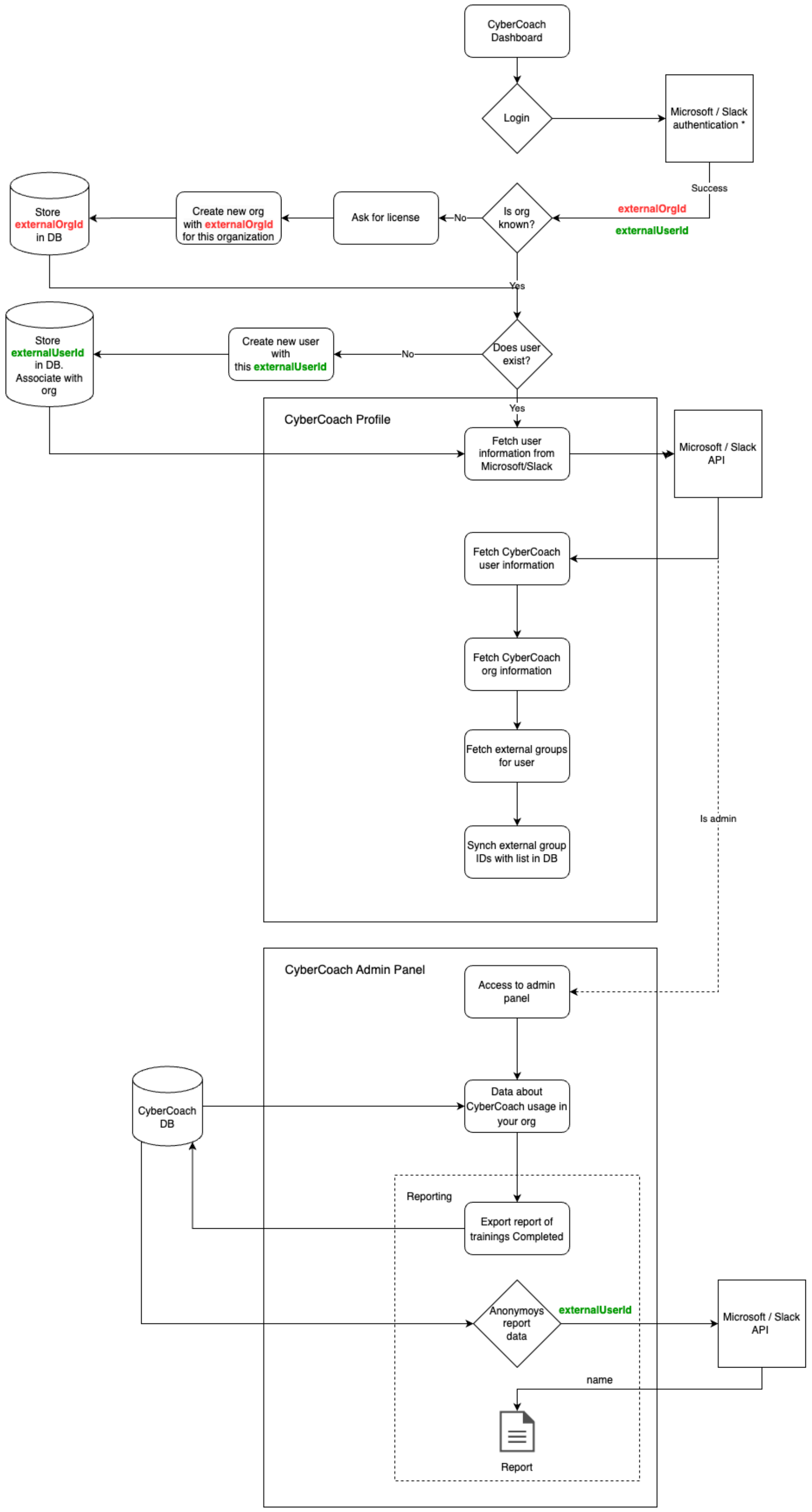
DFD: New session started with CyberCoach


CUSTOMER TEAMS / SLACK
CyberCoach service user (using Teams/Slack app)
Request from Teams/Slack to Microsoft Bot Framework contains user's user ID and organization's ID
CYBERCOACH AZURE CLOUD
Do we know this org?
Nope
Create new user with this ID for this organization
Store external ID for organization
Yep
Do we know this user?
Nope
Create new user with this ID for this organization
Store external ID for user, associate with organization
Yep
Fetch user information for this user
User information: selected language, learning path, and completed trainings
Service main menu

DFD: Training completed and stored


CUSTOMER TEAMS / SLACK
User chooses that that they would like to store the completed training
Request from Teams/Slack to Microsoft Bot Framework contains user's user ID and organization's ID
CYBERCOACH AZURE CLOUD
Store training
Store completed training, associate with UserID

DFD: Dashboard operations


CyberCoach Dashboard
Login
Microsoft / Slack authentication *
Success
externalOrgId
externalUserId
Is org known?
No
Ask for license
Create new org with externalOrgId for this organization
Store externalOrgId in DB
Yes
Does user exist?
No
Create new user with this externalUserId
Store externalUserId in DB. Associate with org
Yes
CyberCoach Profile
Fetch user information from Microsoft/Slack
Microsoft / Slack API
Fetch CyberCoach user information
Fetch CyberCoach org information
Fetch external groups for user
Synch external group IDs with list in DB
Is admin
CyberCoach Admin Panel
Access to admin panel
Data about CyberCoach usage in your org
CyberCoach DB
Reporting
Export report of trainings Completed
Anonymoys report data
externalUserId
Microsoft / Slack API
name
Report