DotAlign Cloud – Architecture, Information Security

Table of contents

Summary

DotAlign Cloud is a solution that analyzes email data (messages, calendar entries and contact cards) and provides up-to-date contact information and relationship intelligence based on that data. The solution is deployed on your Azure tenant and all data storage and processing happen there. No user data leaves your Azure tenant[i]. Data is encrypted both at rest and in motion as it moves through the various parts of the solution. Access to the application and the API requires authentication with the tenant's Azure Active Directory.

Data

What kind of data does DotAlign Cloud deal with?

DotAlign Cloud primarily deals with mailbox data. This includes email messages, calendar entries and contact cards. From these sources are extracted contact information, work experiences, people and company relationships and other such data and inferences. The data is considered extremely sensitive, and so the model DotAlign Cloud operates on is "on-premises", where the data is hosted and processed entirely on the customer's cloud infrastructure. No user data makes its way back to DotAlign, Inc., and no user data can be viewed by employees of DotAlign, Inc.

How is the database secured?

The high-level summary is that data is encrypted both at rest and in motion, and access to the database server is restricted via IP address restrictions. The following security features come into play w.r.t the database:

1. **SSL Transport Encryption** – Azure SQL Server supports the standard SSL encryption protocol to provide encryption in motion.

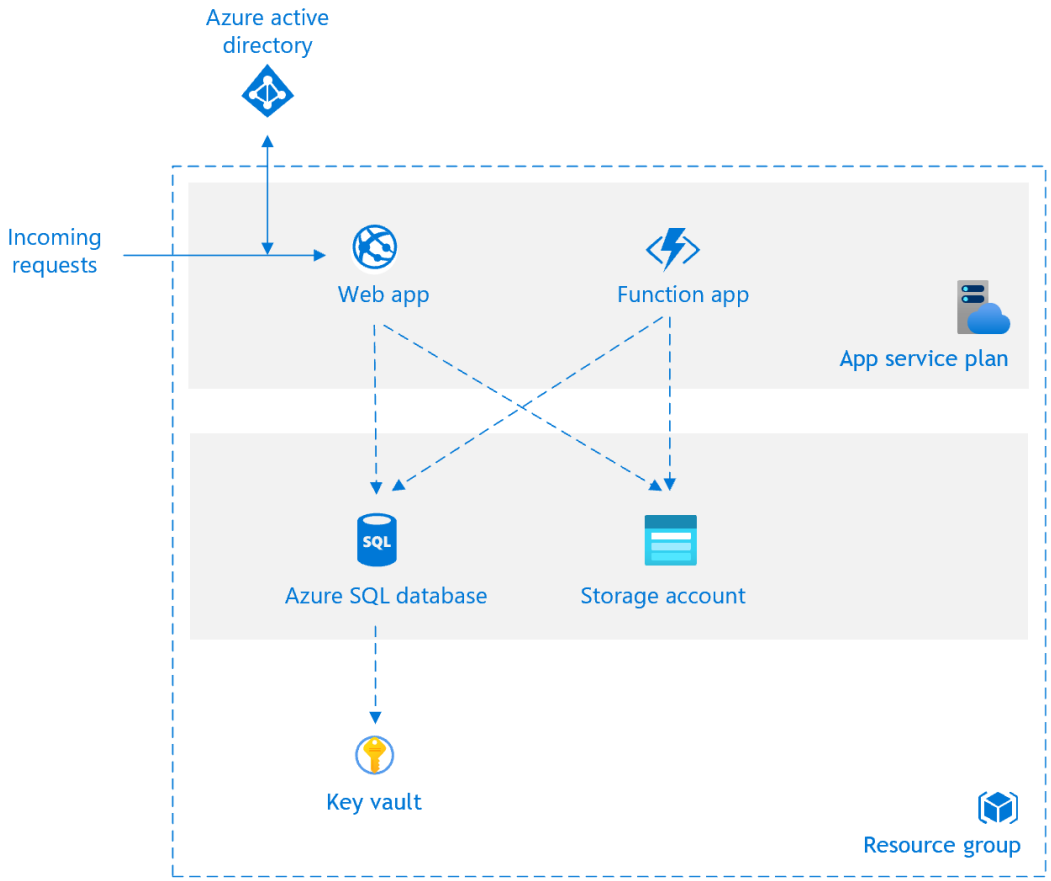
2. **Transparent Data Encryption (TDE) –** TDE is a form of file level encryption provided by Azure SQL Server. It provides encryption at rest. Data files cannot be accessed without the encryption key, which is stored separately in the Azure Key Vault.

3. **Dynamic Data Masking (DDM)** – DotAlign enables dynamic data masking on database columns in Azure SQL Server to hide sensitive data from non-privileged users, to provide an additional layer of security.

4. **Firewall** – Azure SQL Server can be configured to restrict access to only a specific list of IP addresses. In this case, only infrastructure components related to DotAlign Cloud are on that list.

Solution Architecture

The following is DotAlign Cloud's high-level architecture. All parts of the solution are deployed to your Azure tenant.



DotAlign Cloud solution architecture

Application Registration

An application registration needs to be made with Azure Active Directory. This allows DotAlign Cloud to communicate with the Microsoft Graph API. This following specific permissions are required. Detailed information about these permissions can be found in the MS Graph permission reference.

a. Mail.Read
b. Contacts.Read
c. Calendar.Read
d. Directory.Read.All
e. Group.Read.All
f. User.Read
g. User.Read.All
h. openid
i. profile

This application registration is relevant and applicable only for your Azure tenant, and no external system ever gains the ability to use it.

Data processing pipeline

The data processing pipeline is a series of Azure Functions, each of which do a specific part of the processing that leads to the data and insights that DotAlign Cloud provides via its API.

Included in the data processing pipeline are high availability persisted queues which are used to trigger the function. The data in the queues is encrypted by default via Azure Store Service Encryption. This provides encryption at rest and adds an additional layer of security, preventing any unauthorized access to raw data as it is being processed.

Any access to the queues themselves is via an SSL connection, which provides encryption in motion.

Web application

The API/Web application is a standard Azure Web App, which relies on SSL for encryption and Azure AD for authentication and authorization. It provides a secure API for individuals or applications to consume.

The web application has a built-in Swagger console where all API endpoints can be viewed and accessed. All data available in the front end functionality can be accessed via the API.

Authentication

Every request to the web application's endpoints is authenticated and authorized via Azure AD. Incoming request may have the security context of an Azure AD user (as in a specific employee) or that of the "team actor".

To further clarify this, let's assume that a team called "NY ventures" has been created in DotAlign and it has 10 members. Each member has a set of data, people and companies, that have been extracted from their mailbox. Furthermore, users can mark specific relationships as "private", while sharing others. The data that they share is pooled together, aligned, and made available as the combined team's data.

If a request is made to an endpoint, say "`/people`", with the security context of an individual user, the set of people records returned will include all of the people that have been extracted from that user's mailbox (including the people that the user may have marked as private), and data that is shared with that user by virtue of being in the team.

If a request is made to the "`/people`" endpoint with the security context of the team actor, the set of people records returned will be the pooled records from all team members. Relationships marked as private will not appear.

Surfacing DotAlign data in an in-house app or location

If you are trying to surface DotAlign data in an in-house app or location, it is important to consider what security context you are going to use. If the in-house app has the same set of users that have been enabled in DotAlign, the security context of the user may be the right one to use. If the users of the in-house app are not meant to be the same as those enabled in DotAlign, or the in-house app has a non-user context (i.e. it is run with some sort of global context), then the team actor context may be the right one to use.

Admin group and custom domain

An Azure AD group can be specified as being the admin group for DotAlign Cloud. Members of that group will be allowed to create teams, manage members and perform other administrative tasks.

Name	Description	Scope
Admin AD group	The AD group whose members are considered admins by DotAlign Cloud	The members of this group will be able to define teams inside DotAlign Cloud and assign users to those teams.

If a custom web URL (for example, “dotalign.yourcompany.com”), instead of the Azure provided default web URL, is needed, an SSL certificate must also be provided for securing the API/Web application. More details can be found here.

Contingency/backout plan

Since DotAlign Cloud is deployed on your Azure tenant, any contingency plans you already have may be relevant and applicable to it. Additionally, the following steps can be followed to back out DotAlign Cloud from your Azure tenant.

1. **Remove application registration** – You can go to Azure Active Directory, find the application registration that corresponds to DotAlign Cloud and delete it, or remove all access granted to it.
2. **Delete resource group** – All Azure resources set up as a part of DotAlign Cloud are grouped together in a resource group that you get to specify while setting up a deployment. That resource group can be deleted, and that will automatically cause all the underlying resources to be deleted.

Related links

1. **Azure SQL Database service, an overview**
   https://docs.microsoft.com/en-us/azure/sql-database/sql-database-technical-overview

2. **Azure Functions, an overview**
   https://docs.microsoft.com/en-us/azure/azure-functions/functions-overview

3. **Azure Web Apps, an overview**
   https://docs.microsoft.com/en-us/azure/app-service

4. **Azure Storage Service Encryption, an overview**
   https://docs.microsoft.com/en-us/azure/storage/common/storage-service-encryption

5. **Microsoft Graph API, an overview**
   https://docs.microsoft.com/en-us/graph/overview

6. **Microsoft Graph access on behalf of a user**
   This article explains how Microsoft Graph can be accessed on behalf of a user, without requiring the user to explicitly sign in. Importantly, it also describes the admin consent flow.
   https://developer.microsoft.com/en-us/graph/docs/concepts/auth_v2_service

7. **Azure AD authentication scenarios**
   https://docs.microsoft.com/en-us/azure/active-directory/develop/authentication-scenarios

8. **Transparent Data Encryption (TDE) – aka, encryption at rest**
   https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/transparent-data-encryption?view=sql-server-2017

9. **Managing TDE keys in the Azure portal**
   https://docs.microsoft.com/en-us/azure/sql-database/transparent-data-encryption-azure-sql#manage-transparent-data-encryption-in-the-azure-portal

10. **Encrypted connections, aka, encryption in motion**
    https://docs.microsoft.com/en-us/sql/database-engine/configure-windows/enable-encrypted-connections-to-the-database-engine?view=sql-server-2017

11. **Dynamic Data Masking (DDM)**
    https://docs.microsoft.com/en-us/sql/relational-databases/security/dynamic-data-masking?view=sql-server-2017

12. **Azure SQL firewall rules**
    https://docs.microsoft.com/en-us/azure/sql-database/sql-database-firewall-configure

13. **Azure SQL security, an overview**
    https://docs.microsoft.com/en-us/azure/sql-database/sql-database-security-overview

14. **SSL certificate for an Azure App Service**
    https://docs.microsoft.com/en-us/azure/app-service/web-sites-purchase-ssl-web-site

15. **Microsoft Graph API permissions**
    https://docs.microsoft.com/en-us/graph/permissions-reference

[i] By default, system logs (which do not include any user data), are sent to a DotAlign account with a 3rd party logging service called Loggly.