

Summit Sales Case study

17 August 2021
12:21

About Customer:

Summit Sales Solutions is a marketing and sales consulting firm located in Kansas City MO. Summit Sales Solutions (S3) focuses on connecting the innovative and growing Kansas City business community with many fortune 500 companies. S3's commitment is to provide the highest quality customer experience paired with professionalism and value.

Summit Sales Solutions is a leader provider of training solutions for sales professionals throughout the United States and beyond. By providing a strategic combination of training tools and speakers from various professional

Problem Statement:

The customer wants to analyse his product performance based on the data that will capture from the Amazon site.

The data Capture time will be daily, monthly, or weekly.

The data will be stored in an Excel File. and some data will be in MS Access Database.

Customer what to get the data from MS Access and Excel file to one place for analyse

Also Customer wants to track the date time in a new column when the excel file is append in data base

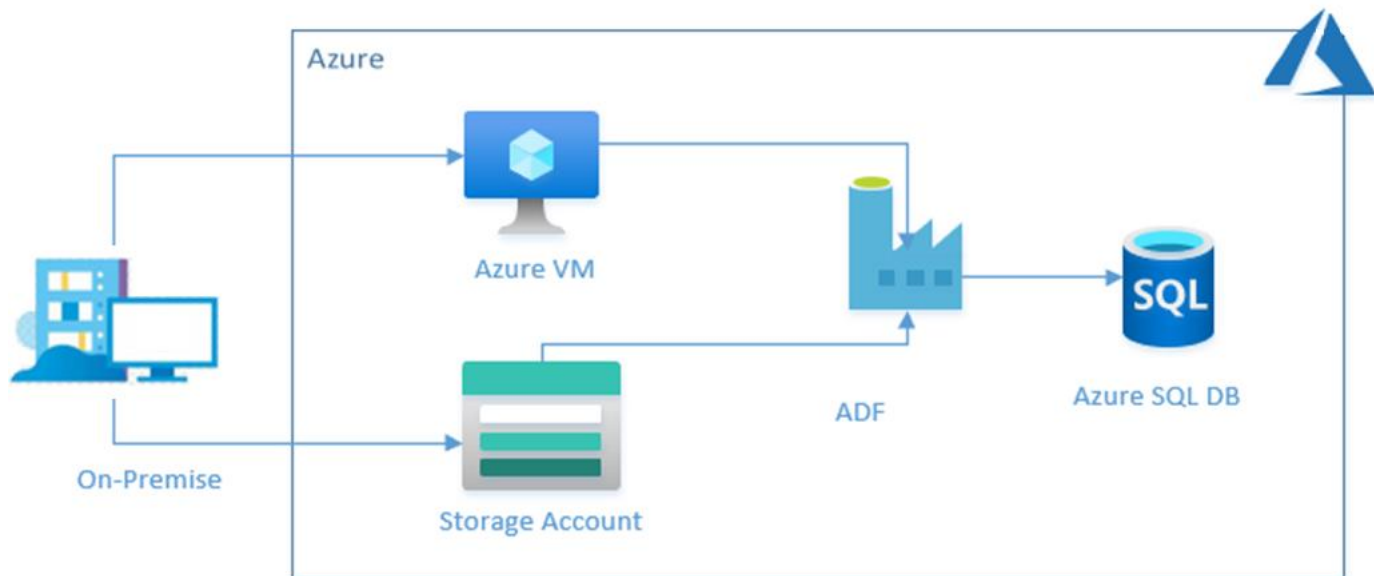
Solution:

- we will get the raw data from customer
- we need to store the excel files in a Storage Account
- we need a Azure VM to Get data from MS Access On-Premises To ADF using Self-Hosted IR
- we need a Azure SQL to store data witch we will get from storage account and MS Access on premises using ADF
- We need a ADF For data Flow and Data ingest

Visio Diagram:

1st part: Azure VM -> On-Premises MS Access DB -> On-Premises install self-hosted integration runtime -> ADF -> Azure SQL DB

2nd part: Azure Storage account -> ADF -> Azure SQL DB



Services Details:

Resource information:

Azure VM: Azure Virtual Machines are image service instances that on-demand and scalable computing resources

Azure Storage Account: Azure Blob storage is Microsoft's object storage solution for the cloud. Blob storage is optimized for storing massive amounts of unstructured data, such as text or binary data.

Azure Data Factory: Azure Data Factory is a **cloud-based data integration service** that allows you to create data-driven workflows in the cloud for orchestrating and automating data movement and data transformation. Azure Data Factory does not store any data itself

Azure SQL Database: Microsoft Azure SQL Database is a managed cloud database provided as part of Microsoft Azure. A cloud database is a database that runs on a cloud computing platform, and access to it is provided as a service. Managed database services take care of scalability, backup, and high availability of the database

Sub-Services Component Description.

Azure VM: We need to install **MS Access DB, self-hosted integration runtime** in VM.

MS Access DB: Microsoft Access is a database management system from Microsoft that combines the relational Microsoft Jet Database Engine with a graphical user interface and software development tools

Self-Hosted Integration Runtime: A self-hosted integration runtime can run copy activities between a cloud data store and a data store in a private network. It also can dispatch transform activities against compute resources in an on-

premises network or an Azure virtual network. The installation of a self-hosted integration runtime needs an on-premises machine or a virtual machine inside a private network.

Azure Storage Account:

Blob: Azure Blob storage is Microsoft's object storage solution for the cloud. Blob storage is optimized for storing massive amounts of unstructured data. Unstructured data is data that does not adhere to a particular data model or definition, such as text or binary data.

Container: A container organizes a set of blobs, similar to a directory in a file system. A storage account can include an unlimited number of containers, and a container can store an unlimited number of blobs.

Azure Data Factory:

Pipeline: A Data Factory or Synapse Workspace can have one or more pipelines. A pipeline is a logical grouping of activities that together perform a task. For example, a pipeline could contain a set of activities that ingest and clean log data, and then kick off a mapping data flow to analyse the log data. The pipeline allows you to manage the activities as a set instead of each one individually.

Trigger: Triggers are another way that you can execute a pipeline run. Triggers represent a unit of processing that determines when a pipeline execution needs to be kicked off. Currently, the service supports three types of triggers:

Schedule Trigger: A trigger that invokes a pipeline on a wall-clock schedule can be based on day, week, or months.

Architecture:

Step 1 Create Resource Azure VM, Azure SQL DB, Storage Account, ADF.

Step 2 Set up

Azure VM: Install MS Access & Self Hosted IR

Storage Account: Create 2 Containers with the same file hierarchy level

ADF: Configure ADF With Azure VM Using Self Hosted IR

Step 3 Build a Pipeline to Get data From MS Access To Azure SQL DB. **Note:**(This Pipeline will Run Only one Time)

3.1 Use Copy Activity For Copy data From Ms. Access to Azure SQL DB

Step 4 Build Pipelines for Moving The Excel data from Storage Account to Azure SQL DB.

4.1 There Will be multiple Container For Different Types of excel files For that we need to create multiple Pipelines.

4.2 For each ADF Pipeline There Will be a data flow that will record date time in Azure SQL Table in a new column.

4.3 Copy data from X Container and move the data to Azure SQL DB & Y Container and Delete from X Container at the end.

Step 5 Schedule the pipeline Daly

5.1 Schedule a Trigger in ADF That will run automate Pipeline Daly

Challenges:

- Challenges are discovering the right checks and uncheck in MS Access before connecting it to ADF.
- Challenge is to find the right way to fetch dates out of filename and store them in columns in a table.
- Stored procedures are acting in unusual ways where they don't execute at all sometimes.

Conclusion:

Summit Sales