



Prepared for WW Public Sector 31/10/2022

Contributors: Ashvith Bangera, Daylan Drummond, Raul Fores Palacios, Lon Tierney, George Earl, Vivek Warrier

Prepared byWW Public Sector Solutions



MICROSOFT MAKES NO WARRANTIES, EXPRESS OR IMPLIED, IN THIS DOCUMENT.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in, or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, our provision of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

The descriptions of other companies' products in this document, if any, are provided only as a convenience to you. Any such references should not be considered an endorsement or support by Microsoft. Microsoft cannot guarantee their accuracy, and the products may change over time. Also, the descriptions are intended as brief highlights to aid understanding, rather than as thorough coverage. For authoritative descriptions of these products, please consult their respective manufacturers.

© 2022 Microsoft Corporation. All rights reserved. Any use or distribution of these materials without express authorization of Microsoft Corp. is strictly prohibited.

Microsoft and Windows are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

The names of actual companies and products mentioned herein may be the trademarks of their respective owners.

EiAD Accelerator – Architecture Document © 2022 Microsoft Corporation.

Contents

1	Exe	cutive Summary	. 4
1	.1	Objectives	. 5
1	.2	Challenges	. 6
2	Intr	oduction	. 7
2	.1	Purpose	. 7
2	.2	Audience	. 7
3	Solu	rtion Architecture	.8
3	.1	High Level Architecture	. 8
3	.2	Workflow	. 9
4	Solu	ition Deployment	.11
4	.1	Pre-Requisites	11
4	.2	Deployment	11
4	.3	Solution Estimator and Licensing	11
5	Арр	endix A: Azure Configuration	12
5 5		endix A: Azure Configuration	
-	.1	-	12
5	.1 App	Summary	12 13
5 6	.1 App .1	Summaryendix B: Azure Services	12 13 13
5 6 6	.1 App .1	Summary endix B: Azure Services Application Insights	12 13 13 13
5 6 6	.1 App .1 6.1.1	Summary	12 13 13 13
5 6 6	.1 App .1 6.1.1	Summary	12 13 13 13 14
5 6 6	.1 App .1 6.1.1 .2 6.2.1	Summary	12 13 13 13 14 14
5 6 6 6	.1 App .1 6.1.1 .2 6.2.1	Summary	12 13 13 13 14 14 15
5 6 6 6	.1 App .1 6.1.1 .2 6.2.1 .3 6.3.1	Summary	12 13 13 14 14 15 15
5 6 6 6	.1 App .1 6.1.1 .2 6.2.1 .3 6.3.1	Summary	12 13 13 14 14 15 15
5 6 6 6	.1 App .1 6.1.1 .2 6.2.1 .3 6.3.1 .4 6.4.1	Summary	12 13 13 14 14 15 16 16

1 Executive Summary

The eIAD accelerator will provide a starting point in the identification of potentially anomalous invoices from the huge volumes of e-invoices that are submitted to an in-country Tax Agency in compliance with existing laws and regulations.

The life cycle of electronic invoice includes the process of issue, storage, verification, reimbursement, etc. In this process, there may be abnormalities such as over-limit/type issuance of e-invoice, repetitive/false invoice reimbursement, fake system connection, multiple attempts passwords, etc. In order to ensure the normal operation of the electronic invoice systems, strict supervision is required for the above abnormalities. However, the traditional manual supervision method will waste a large amount of manpower and material resources, and there is no guarantee that abnormal behaviors can be detected in time and handled jointly.

The main outcome from this project is a collection of technical assets that will help a tax agency to deploy a baseline model along with a pipeline that will facilitate the movement and transformation of data from primary sources to the model itself and publish results to appropriate endpoints. These technical assets are the final deliverable from the current project and all subsequent activities are driven by individual tax agencies.

Firstly, the tax agency will use this data pipeline and baseline model contained in the eIAD accelerator to train specific to its ecosystem of invoicing data. The training will incorporate defined priorities and choices specific to the tax agency in identifying anomalies.

The results are accompanied by a level of explain ability for those results providing the tax agency experts a basis to review the reasoning behind them.

Thereafter, based on an iterative process of experimentation and validation the model will be further optimised to perform best in the invoicing data portfolio specific to the country.

The tax agency will then move that optimised model and pipeline to its production infrastructure and incorporate its operation to business-as-usual (BAU) processes.

Over time, the tax agency will continuously improve the model through further learning, experimentation, and validation to achieve greater efficiency and productivity in the anomaly detection process.

Objectives

The Microsoft E-IAD, the "Electronic Invoicing Anomaly Detection" Solution Accelerator will help Tax Agencies in the Civilian Government that have implemented electronic invoicing to detect invoicing anomalies and check large data sets for consistency to help improve revenue collection. The E-IAD, provides capabilities to process tabular e-Invoicing transactional data at scale and detect anomalies in taxpayers' invoicing operations. E-IAD will provide additional tools to:

- Score anomalies and identify variables that influence the score.
- Visualize and do descriptive analytics.
- Measure the quality of raw data.

The primary use cases are around:

- Reduce data processing complexity for analytics to detect anomalies quickly and solve problems faster.
- Measure e-Invoicing data quality with a rapid e-Invoicing analysis that helps tax agencies identify policy gaps in collecting e-Invoicing documents.
- Represent B2B, B2C, and B2G commercial transactions in the economy, that help tax agencies to describe and analyze issuing patterns and commercial behaviors.
- Detect e-Invoicing anomalies in the tax agency's interest to identify issuing errors requiring taxpayer correction, detect unexpected issuing patterns, or detect issuing operations requiring further investigation for potential incorrect taxpayer conduct.
- Provide an anomaly index and score to increase productivity and intelligence of the tax agency investigation department.

The e-IAD output will feed the tax agency analysis and investigation system. Investigation and qualification of anomalies are out of the scope of this accelerator. The processed output files will allow agencies to use their visualization solution of preference.

Some examples of anomaly detection capabilities are:

- Detecting the industrial activities where a company is doing commercial transactions vs. the industrial activities
 reported to the tax agency in the registration process.; potentially represents a miss-classification of the industrial
 activity.
- Detect commercial transactions with an irregular pattern in the size of the total voucher, total tax, total discounts, etc.; detect commercial transactions with an irregular pattern in the seller's supply chain.

--- Microsoft

Challenges

The main challenges for effective adoption of Al/ML on e-Invoicing data are the ability to process large datasets at scale and reduce non-time series data's multidimensionality.

The well-known risk model approach to selecting tax documents or taxpayers for analysis and investigation also suffers from the challenges of big sets of data.

The e-IAD aims to address these typical challenges and, most importantly, enable an automated ML pipeline. This pipeline will facilitate the data loading, calculate the data quality, enrich, and expand the data features to do ML, and then apply an ML algorithm for anomaly detection and produce meaningful results that support the tax agency investigation process. A comprehensive set of visualizations will help with the interpretation of results and the analysis of data along the entire pipeline.

These functions enable revenue agents to recognize patterns that could indicate abnormalities in electronic invoices or similar data sources and then flag those issues for investigation.

From a non-functional perspective, the e-IAD will help tax agencies adopt AI/ML practices in the Azure Cloud, using best-in-class practices for continuous integration/delivery in a fully automated data processing pipeline.

2 Introduction

Purpose

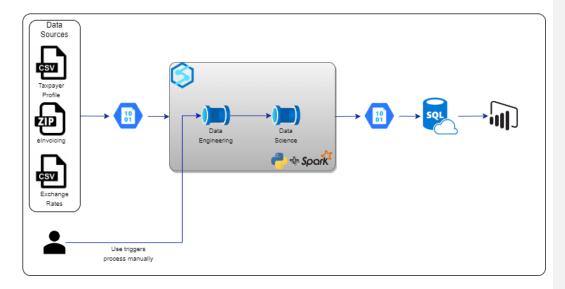
The purpose of this document is to provide an architectural overview of the accelerator with a focus on how it has been assembled and can be deployed.

Audience

This document is targeted primarily at Individuals working with public finance who would like to deploy the industry sales accelerator. Roles such as the ATU (Industry Execs, Account Technology Strategists), STU (Azure Specialists), CSU (Cloud Solution Architects), GBB (Global Black Belts) and NTO (National Technology Officers) and NSO (National Security Officers) will find this document particularly relevant.

3 Solution Architecture

High Level Architecture



Components

The architecture consists of the following components.

Azure Storage Account: Blob storage is used as a staging area for the source data before loading it into Azure Synapse.

<u>Azure Synapse</u>: is a distributed system designed to perform analytics on large data. It supports massive parallel processing (MPP), which makes it suitable for running high-performance analytics. Azure synapses pull data from a wide variety of data sources, both on-premises and in the cloud. Ingest data from file-based sources containing CSV or JSON files.

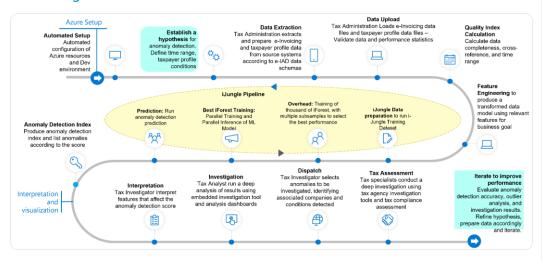
Call REST APIs provided by SaaS applications that will function as your data source for the pipeline. Azure Key Vault: Used to store and protect keys and credentials in a secure place and monitor operations and access to it.

<u>Azure Monitor and Log Analytics Workspace</u>. Used to monitor the environment, diagnostics, performance, audit logs, vulnerability scans, and traffic flows and enable the platform to send events for critical issues.

Power Bl: Allows import or direct query datasets to connect to data services. Power Bl is a suite of business analytics tools to analyze data for business insights.

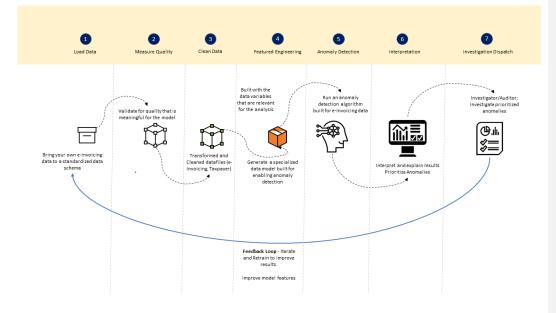
Azure Workbook: Workbooks provide a flexible canvas for data analysis and the creation of rich visual reports within the Azure portal. Workbooks combine text, <u>log queries</u>, metrics, and parameters into rich interactive reports.

High Level Workflow



Dataflow

The e-IAD Pipeline implemented integration pipelines, notebooks, logs, and output files according to the following visual structure:



4 Solution Deployment

4.1 Pre-Requisites

To find the latest pre-requisites for the EiAD Accelerator, view our <u>pre-requisites on GitHub.</u>

4.2 Deployment

A detailed set of deployment instructions and scripts can be found on our GitHub Project microsoft/EiAD (github.com)

Solution Estimator and Licensing

The proposed cost estimation is prepared based Sandbox environment; estimation may vary from customer to customer.

eIAD/costestimator.md at main · microsoft/eIAD (github.com)

5 Appendix A: Azure Configuration

Summary

The following Azure data centers were used in the accelerator deployment:



Location	Number of resources
global	1
eastus2	7

The following components are used in your deployment:

Trend	Category	Component Type	Number
	ETL (Extract, Transform, Load)	Application Insights Component	1
	ETL (Extract, Transform, Load)	Log Analytic Workspace	1
	Organizational Unit	Resource Group	1
	Security	Vault	1

Commented [A1]: Some of these values are not right.. 835 policies, 64 subscriptions??? [Mention was removed] please check

6 Appendix B: Azure Services

Application Insights



eiad-insights-XXXXX

Settings

Instrumentation Key	Hidden
Resource group	eiad-shared
Location	eastus2
Creation	2022-06-25 19:12:59.00 PM
Provisioning State	Succeeded
Kind	web
Application Type	web
Application ID	eiad-insights-XXXXX
Flow Type	
Sampling Percentage	100

Tag Key	Tag Value
ProjectName	elAD

Key Vaults



eiad-keyvault-XXXXX

Settings

Name	eiad-keyvault-XXXXX
Vault Uri	https://eiad-keyvault-XXXXX.vault.azure.net/
Enable Soft Delete	True
Enabled for Deployment	False
Enabled for Disk Encryption	False
Enabled for Template Deployment	False
Permission model	Vault access policy
Sku Name	Standard
Location	eastus2

Tag Key	Tag Value	
ProjectName	eIAD	

Storage Account



eiaddataXXXXX



The storage eiaddataXXXXX has the following settings:

_	
Resource group	eiad-shared
Status	Available
Location	eastus2
Replication	Standard Geo Replicated Storage
Account kind	StorageV2
Created date	2022-06-25 19:12:24.00 PM
SKU	Standard_GRS
Provisioning state	Succeeded
Used capacity (Avg)	
Blob EndPoint	https://eiaddataXXXXX.blob.core.windows.net/
File EndPoint	https://eiaddataXXXXX.file.core.windows.net/
Queue EndPoint	https://eiaddataXXXXX.queue.core.windows.net/
Table EndPoint	https://eiaddataXXXXX.table.core.windows.net/
Geo Primary Region	eastus2
Geo Secondary	centralus
Region	Certualus



Tag Key	Tag Value
ProjectName	eIAD

Log Analytics Workspaces



eiad-monitor-XXXXX



Resource group	eiad-shared
Provisioning State	Succeeded
Location	eastus2
Pricing tier	



Tag Key	Tag Value	
ProjectName	eIAD	



Resource groups



eiad-shared





Resources

Name	Location	Туре
<u>eiaddataXXXXX</u>	eastus2	Microsoft.Storage/storageAccounts
eiad-monitor-XXXXX	eastus2	Microsoft.OperationalInsights/worksp aces
eiad-synapse-XXXXX	eastus2	Microsoft.Synapse/workspaces
eiad-keyvault-XXXXX	eastus2	Microsoft.KeyVault/vaults
eiad-insights-XXXXX	eastus2	Microsoft.Insights/components
b0b86160-79ac-fde2-a0cd- 0c0c7ef060b1	eastus2	microsoft.insights/workbooks
eiad-synapse- XXXXX/eiadsparkpool	eastus2	Microsoft.Synapse/workspaces/bigDa taPools



Tag Key	Tag Value
BuildNumber	20220808.1
ProjectName	eIAD

7 Appendix C: Resources

Topics	Resource
Learning	Microsoft Al School Get started with artificial intelligence on Azure - Learn Microsoft Docs Explore Natural Language Processing in Microsoft Azure - Learn Microsoft Docs Microsoft Graph tutorials Microsoft Graph Fundamentals
Certifications	Microsoft Certified: Azure Al Fundamentals Microsoft Certified: Azure Al Engineer Associate Microsoft Certified: Azure Data Scientist Associate
Microsoft Docs	Overview of Microsoft Graph - Microsoft Graph Microsoft Docs
Reference Architecture	Azure reference Architectures Automated enterprise BI - Azure Architecture Center Microsoft Docs