



Konsolidovaná analytická vrstva (opis štandardov pre integráciu)

Využitie dát pre zlepšenie fungovania inštitúcií verejnej správy - data driven government

Cieľom tejto metodiky a štandardou je popísať základné možnosti integrácie zdrojových dát/údajov do systému KAV (Konsolidovaná Analytická Vrstva).

Tento systém umožňuje prácu so dátovými zdrojmi, získavanie, transformáciu, integráciu a následné poskytovanie zmysluplne pospájaných dát a dátových výstupov analytickým jednotkám a verejnosti. Systém je pripravený na načítanie dát z databázových systémov, aplikačných rozhraní (API) a dátových súborov.

Pre štandardný vývojový cyklus sa KAV pripravuje v prostrediach DEV, TEST a PROD. Toto treba zohľadniť pri integrácii a urobiť správnu voľbu.

Podporované formáty súborov

Systém KAV používa na prístup k zdrojom údajov konektory Azure Data Factory, takže na načítanie súborov môžete nakonfigurovať nasledujúce konektory: Amazon S3; Amazon S3 Compatible Storage; Azure Blob; Azure Data Lake Storage Gen1+Gen2; Azure Files; FTP; Google Cloud Storage; HDFS; http; Oracle Cloud Storage a SFTP.

Následne je možné s týmito formátmi pracovať aj v komprimovanej (zazipovanej) podobe: Avro; CSV; Excel; JSON; ORC; Parquet; XML; Podporované databázové systémy

Okrem prenosu súborov je možné aj priame pripojenie k databázovým systémom:

Microsoft SQL Server; Azure SQL Database; Azure SQL Data Warehouse; Oracle Database; MySQL; PostgreSQL; MongoDB; Cassandra; Amazon Redshift; Teradata; Sybase; MOngoDB; Cassandra; Snowflake-

Toto sú len najbežnejšie zdroje databáz. Úplný zoznam dostupných možností vrátane všetkých obmedzení nájdete v dokumentácii Azure Data Factory (<https://learn.microsoft.com/cs-cz/azure/data-factory/connector-overview>).

Ďalšie rozhrania

Okrem prenosu údajov prostredníctvom súborov alebo priamo z databázových systémov je možné pristupovať k údajom aj pomocou štandardizovaných aplikačných rozhraní (REST, SOAP) alebo dokonca len prostredníctvom všeobecného servera http. Možnosti manuálnej konfigurácie pripojenia k zdroju údajov sú široké, takže aj v prípade neštandardného rozhrania/zdroja je vysoká pravdepodobnosť, že zdroj údajov možno po dohode s tímom KAV integrovať.

Referencie na dokumenty

Identifikácia	Názov dokumentu	ID dokumentu/Zdroj
ADF	Dokumentácia Azure Data Factory	https://learn.microsoft.com/en-us/azure/data-factory/
Terraform	Dokumentácia Terraform	https://developer.hashicorp.com/terraform/docs
Azure DevOps	Dokumentácia Azure Devops	https://learn.microsoft.com/en-us/azure/devops/?view=azure-devops

EXTERNÝ PORTÁL:

[KAV - Konsolidovaná analytická vrstva | Ministerstvo investícií, regionálneho rozvoja a informatizácie SR \(gov.sk\)](#)

**INTERNÝ PORTÁL:**[KAV - Interný Portál – Home \(sharepoint.com\)](#)**Skratky a pojmy**

Skratka	Vysvetlenie
ADF	Azure Data Factory je cloudová služba od spoločnosti Microsoft, ktorá umožňuje orchestráciu, plánovanie a automatizáciu tokov údajov a procesov ETL (Extract, Transform, Load) v hybridných alebo čisto cloudových dátových prostrediach.
ARM	Azure Resource Manager (ARM) je správca zdrojov v cloude Microsoft Azure, ktorý vám umožňuje deplovať, spravovať a monitorovať zdroje Azure.
Azure Devops	Azure DevOps je balík nástrojov a služieb od spoločnosti Microsoft, ktorý poskytuje kompletné riešenie na plánovanie, sledovanie a správu vývoja a dodávky softvéru vrátane správy kódu, automatizácie zostavovania a nasadenia a sledovania celého vývojového cyklu.
CICD	CICD (Continuous Integration and Continuous Deployment/Delivery) je metodika vývoja softvéru, ktorá automatizuje procesy testovania, integrácie a nasadenia, čo umožňuje rýchle a pravidelné dodávanie kvalitného softvéru.
Deployment/Nasadenie	Umiestnenie a konfigurácia softvéru alebo infraštruktúry v cieľovom prostredí.
Hub and Spoke	V architektúre Azure Hub and Spoke predstavuje "hub" rozbočovač centrálnu miestu, kde sa zhromažďujú zdieľané zdroje a spravujú centrálnu služby, zatiaľ čo "spokes" predstavujú špecializované časti alebo prostredia pripojené k centrálnemu HUB, čo umožňuje efektívnu správu a distribúciu zdrojov v cloude.
IaC	Prístup v softvérovom inžinierstve, ktorý umožňuje spravovať a nasadzovať infraštruktúru prostredníctvom kódu, čo uľahčuje automatizáciu, opakovanie a správu zdrojov infraštruktúry.
Jump Server	Bezpečný server, ktorý umožňuje bezpečný vzdialený prístup k sieťovým zariadeniam alebo prostrediam, ktorý zvyšuje bezpečnosť obmedzením priameho prístupu k citlivým častiam siete.
Pipeline	Automatizovaný proces buď v rámci CICD alebo v rámci ADF.
Power BI	Power BI je platforma od spoločnosti Microsoft, ktorá umožňuje užívateľom vizualizovať a analyzovať svoje dáta, vytvárať interaktívne dashboards a sdieľať informácie v rámci organizácie.
Repository/Repozitár/Úložisko	Úložisko, kde sa zdrojový kód projektu a súbory zhromažďujú, ukladajú a spravujú. Uľahčuje sledovanie zmien a spoluprácu medzi členmi tímu pri vývoji softvéru alebo iných projektov.
SHIR/Gateway	Azure Data Factory Self Hosted Integration Runtime, počítač nakonfigurovaný na prístup k zdrojom údajov a na ňom je nainštalovaný softvér nakonfigurovaný pre zvolené ADF.
Storage Account	Storage Account v Azure je služba, ktorá poskytuje škálovateľné a zabezpečené úložisko pre rôzne typy údajov vrátane súborov, objektov blob, tabuliek a frontov.
Tableau	Tableau je vizualizačná a business intelligence platforma, ktorá umožňuje používateľom analyzovať, vizualizovať a zdieľať údaje prostredníctvom interaktívnych a esteticky príťažlivých vizualizácií.
Terraform	Nástroj na správu infraštruktúry, ktorý vám umožňuje deklaratívne definovať a spravovať infraštruktúru ako kód, čo zahŕňa vytváranie, zmenu a odstraňovanie cloudových zdrojov od rôznych poskytovateľov vrátane AWS, Azure a ďalších.
VM	Virtual Machine, virtuálny PC v cloude.

**Integrácia prostredníctvom Azure Storage**

Ak sú zdrojové údaje vo forme súborov, je možné sa integrovať priamo do takzvaného verejného úložiska (storage account with public endpoint /úložný účet s verejným koncovým bodom), kde môžu byť súbory po správnej konfigurácii uložené.

Autentifikácia:

Na overenie prístupu k účtu úložiska možno použiť nasledujúce možnosti (uvedené v poradí od najlepšej po najhoršiu):

RBAC (Entra/AAD)

- Role-Based Access Control, je mechanizmus riadenia prístupu v Azure, ktorý poskytuje flexibilné a granulárne riadenie prístupu k službám Azure.
- RBAC sa používa na definovanie rolí a priradenie týchto rolí používateľom alebo skupinám na riadenie činností, ktoré môžu vykonávať v prostredí Azure.
- Je možné použiť užívateľské účty alebo tzv. service principály
- Obmedzené iba na Mirri tenante, externí používatelia musia byť prizvaní.

Shared Access Signatures (SAS tokeny)/ Spoločné podpisy pre prístup

- SAS tokeny poskytujú obmedzený a časovo limitovaný prístup k prostriedkom Azure Storage bez potreby zdieľania prístupových kľúčov.
- Umožňujú definovať špecifické oprávnenia (čítanie, zápis, mazanie atď.) a platnosť (časový rámec) prístupu k úložisku.
- Tokeny SAS možno použiť na udelenie prístupu tretím stranám bez toho, aby bolo potrebné zdieľať celé prístupové kľúče.
- Tokeny SAS možno použiť na prístup k celému účtu, kontajnerom alebo jednotlivým objektom (blokom, súborom atď.).

Access Keys (Prístupové kľúče)

- Každý účet Azure Storage má dva primárne a dva sekundárne prístupové kľúče. Tieto kľúče sa používajú na overenie identity pri prístupe k účtu alebo pri manipulácii s údajmi v úložisku.
- Vždy umožňujú prístup k celému účtu, takže nie sú vhodné pre integračné scenáre s viacerými partnermi a jedným úložiskom.

Po dohode s tímom KAV sa zvolí spôsob overovania a povolí sa prístup k vyhradenému priečinku, do ktorého možno nahrávať súbory.

Jednotlivé Rozhrania:**Azure CLI**

Najjednoduchší spôsob nahrávania súborov je použiť AZ CLI, rozhranie príkazového riadka Azure. Týmto spôsobom je možné integrovať nahrávanie napr. do shell skriptov:

```
az storage blob upload --account-name <storage_account_name> --container-name <container_name> -
name <blob_name> --type block --file <local_file_path>
```

PowerShell

```
$storageAccountName = "<storage_account_name>"
$containerName = "<container_name>"
$blobName = "<blob_name>"
$localFilePath = "<local_file_path>"

$ctx = New-AzStorageContext -StorageAccountName $storageAccountName -UseConnectedAccount $true
Set-AzStorageBlobContent -Container $containerName -Blob $blobName -File $localFilePath -Context $ctx
```



Financovaný
Európskou úniou



PROGRAM
SLOVENSKO



MINISTERSTVO
INVESTÍCIÍ, REGIONÁLNEHO ROZVOJA
A INFORMATIZÁCIE
SLOVENSKEJ REPUBLIKY

REST API

Na integráciu priamo do aplikácie je vhodnejšia implementácia rozhrania Azure Storage REST API.

Dokumentácia vrátane odkazov na moduly použiteľné v rôznych prostrediach je k dispozícii v dokumentácii Azure:

<https://learn.microsoft.com/en-us/rest/api/storageservices/blob-service-rest-api>