

SkillSync

AI powered candidate matching system

TEAM 5

Yash Sethi

Nehal Joshi

Berly Biju

Ashraf Elrufaie

Table Of Contents

- **Approach & Methodology**
- **Algorithms Used**
- **Scoring Method**
- **Limitations**
- **UI**
- **Cloud Architecture**

Approach & Methodology

Compound AI Solution

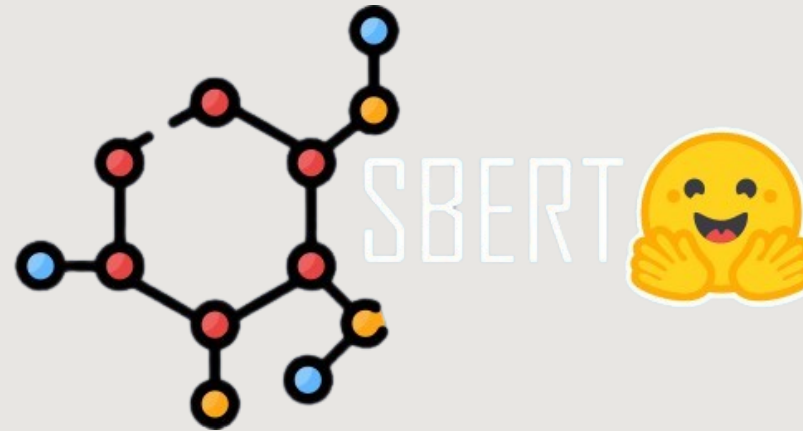
Leveraging and combining the power of AI into a well-harmonized candidate scoring solution.



GPT 3.5 Turbo



Deep Translate



Sentence Transformers



OCR



TEAM 5



3

Algorithm

From Text to Structure: Automating Job Description Analysis with Large Language Models

1

OCR and Translate

We used EasyOCR and Deep Translate to read the Job Description files. We use batch processing to parallelize and reduce latency.

2

LLM

We used GPT 3.5 Turbo to read this unstructured data and create semi-structured JSON output with information like required skills, preferred skills, language skills, experience required, job responsibilities and skill importance scores.

Algorithm

Cosine Match: Aligning Experience to Responsibilities

1

Translate

We used Deep Translate to translate the Mission Experience of each consultant.

2

Embedding

We used all-MiniLM-L6-v2 model to create vector embeddings for each experience they have and each job responsibility.

3

Average Cosine Similarity

We calculated the average cosine similarity of each sentence in the experience and each job responsibility to avoid outliers that are a perfect match. We also min max scaled the values.

Algorithm

Skill Match: Candidate Skills Scoring Matrix

1

Levenshtien Fuzzy Matching

We use Levenshtien distance to match the consultant's skills to the skills in the job requirement.

2

LLM Skill Priority Scoring

We used GPT 3.5 to analyze the job descriptions to give a priority level to each skill in the job description.

3

Final Weighted Skills Scoring Matrix

Using levels as weights we create the final skill matrix consisting of the matched required and preferred skills.

Algorithm

Data-Informed, Customizable Logic-Powered Scoring

1

Skills Match

Using the scoring matrix, we assign custom weights for each skill matched.

2

Cosine Match

Using a dynamic scoring logic to prioritize cosine similarity in certain scenarios by increasing its impact on the score.

3

Avalibility Scoring

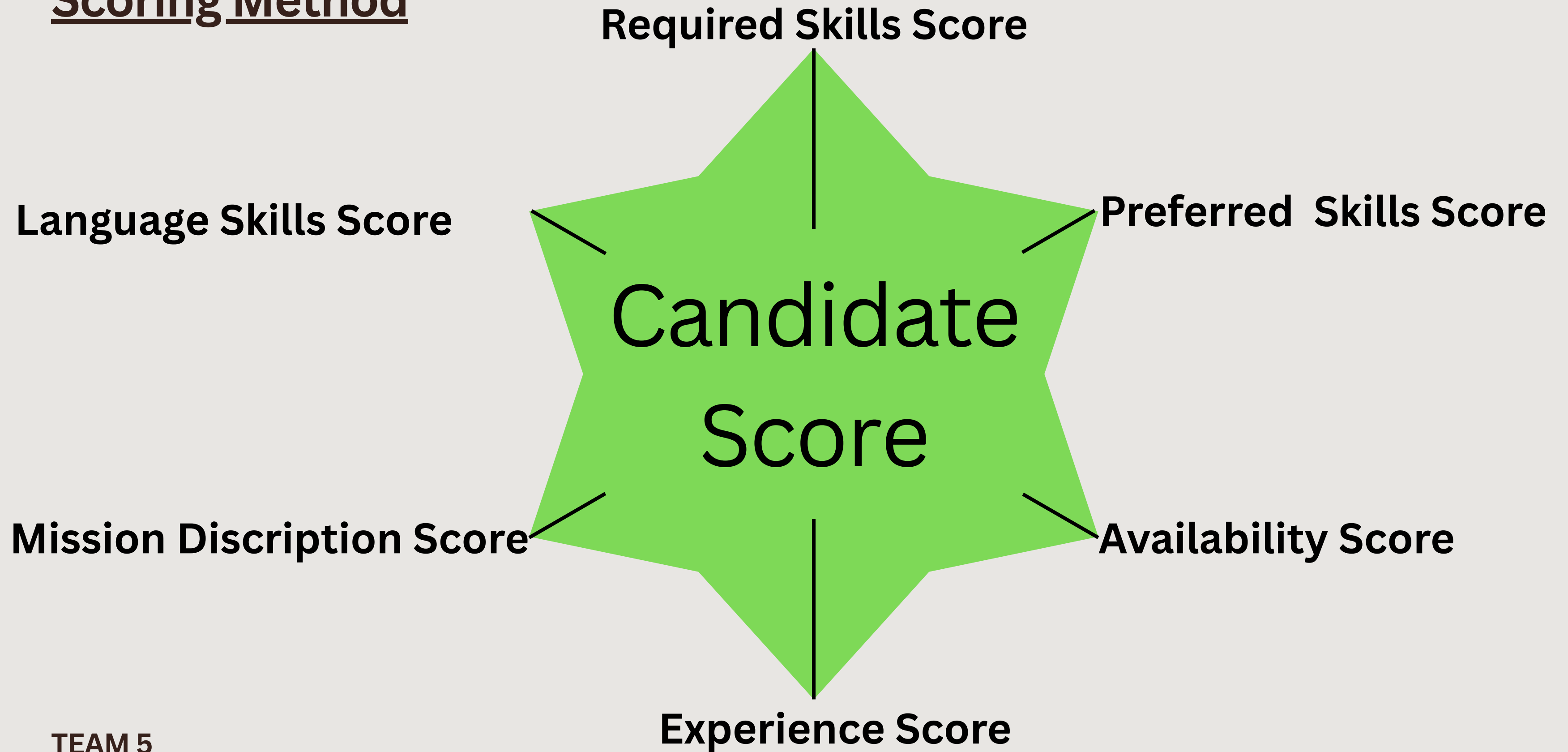
Using custom logic to check and filter consultants that will not be available for the duration of the job while assigning a score based on availability.

4

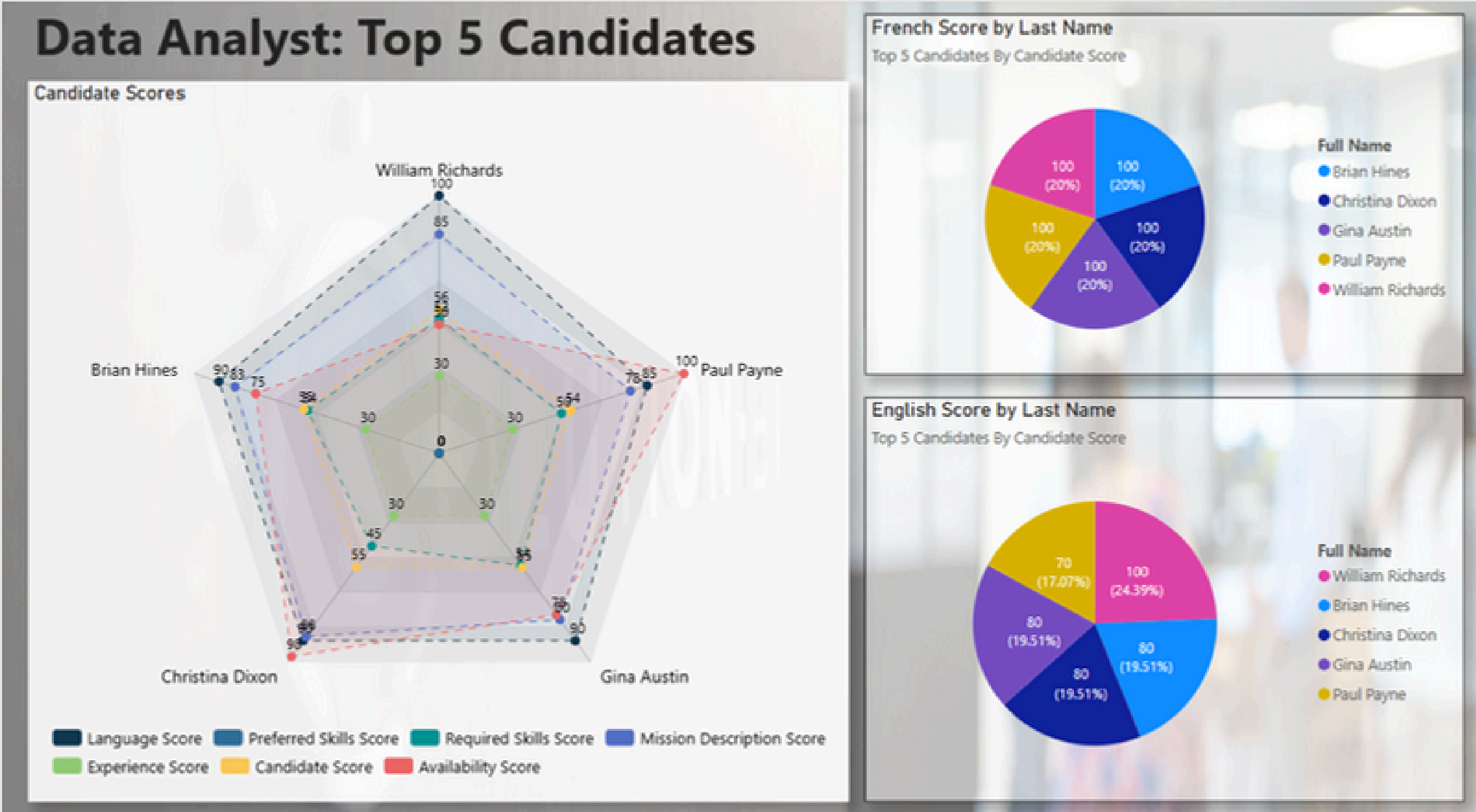
Experience Match

Using the required experience extracted via GPT 3.5 from the job description and creating a custom score for the match of the years of experience of each consultant.

Scoring Method



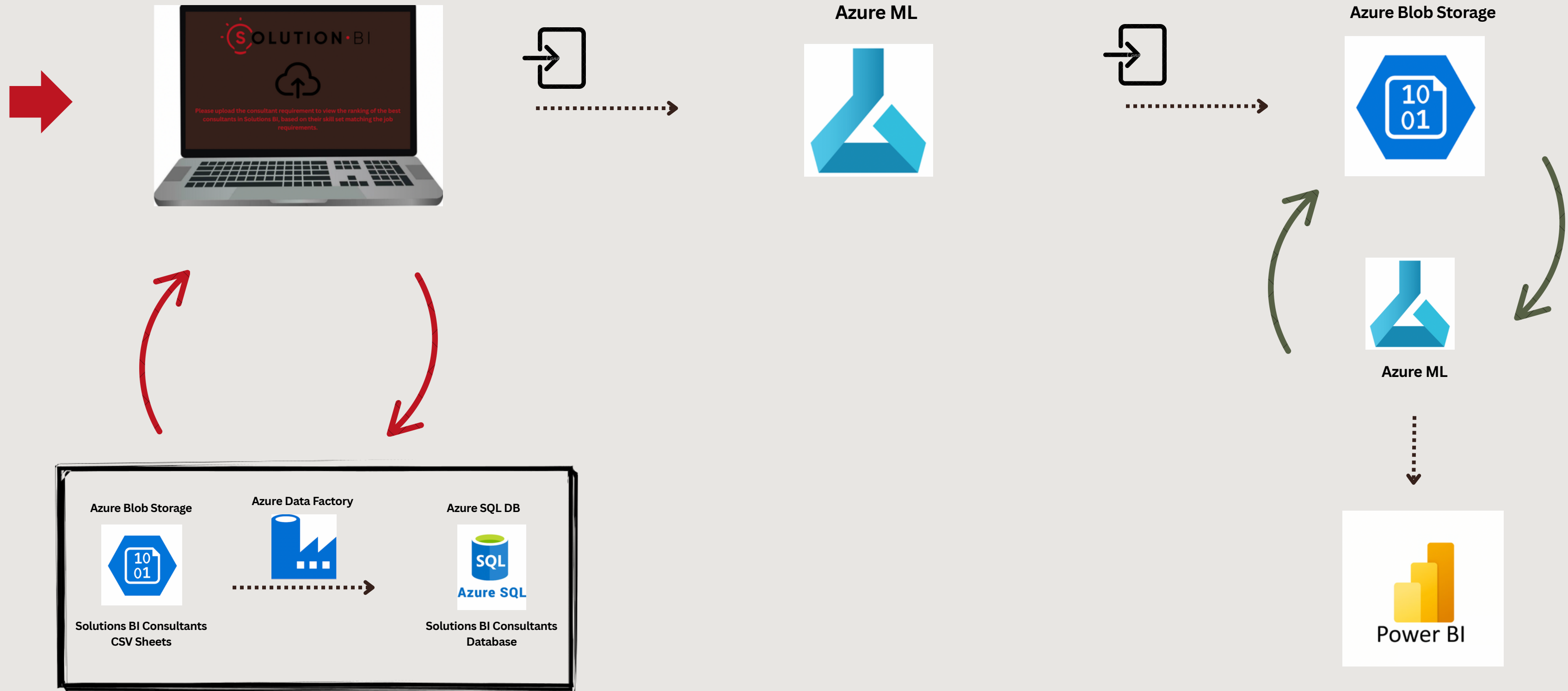
Dashboard



USER INTERFACE



Architecture Flow - Microsoft Azure



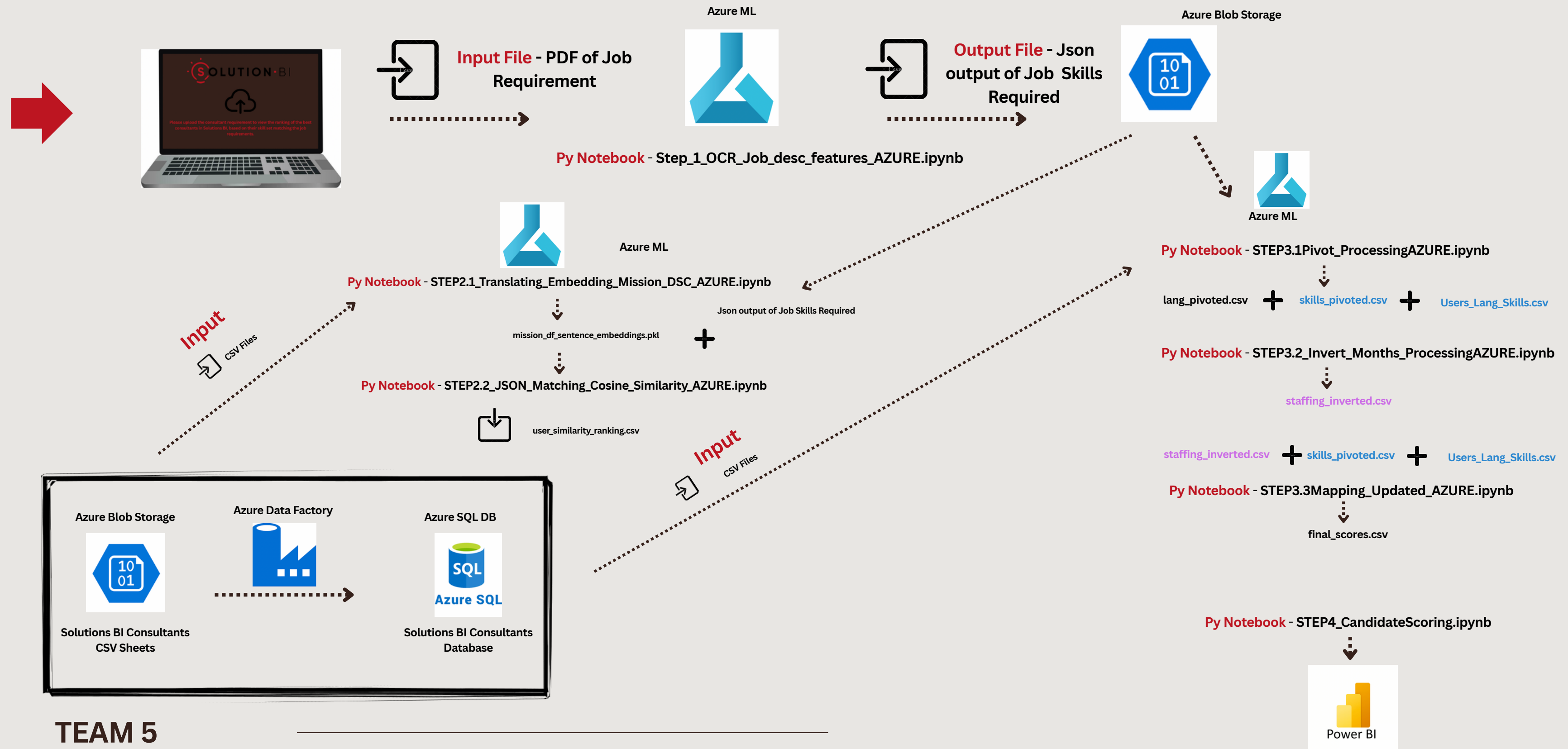
THANK YOU

For Listening

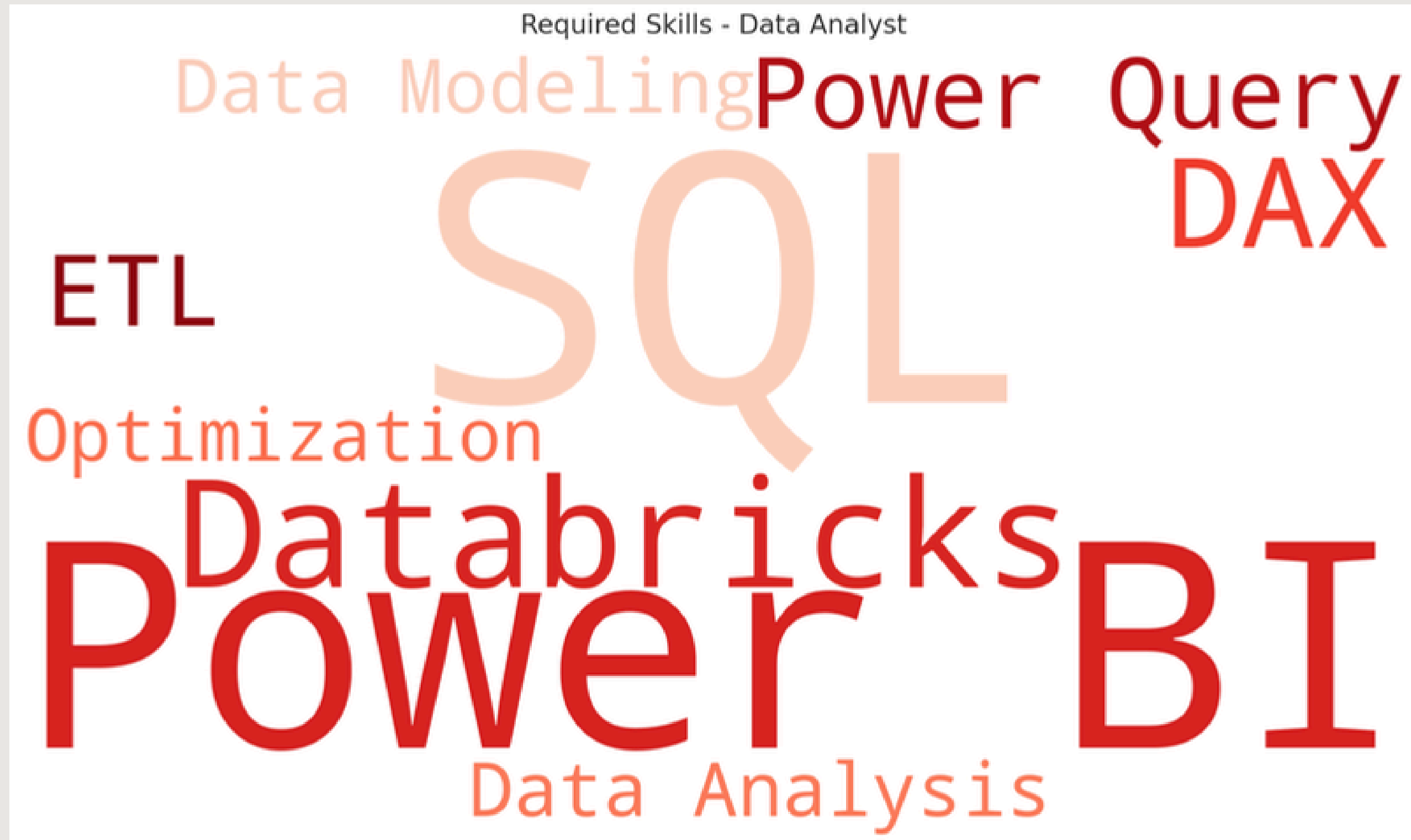
APPENDIX

Architecture Flow - Microsoft Azure - Detailed

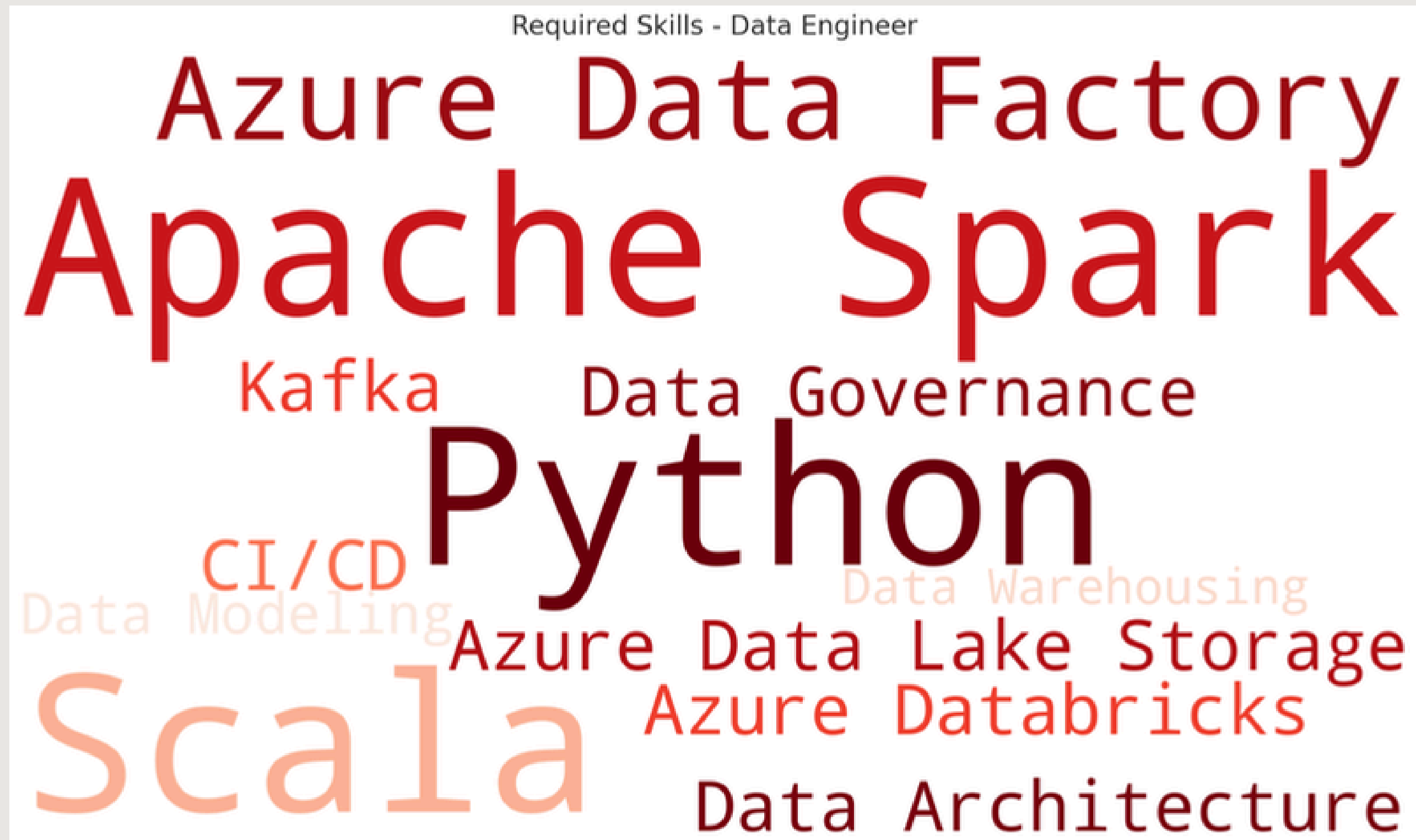
POC Phase - Considering only 3 Job Roles (Data Engineer, Data Analyst and Scrum Master)



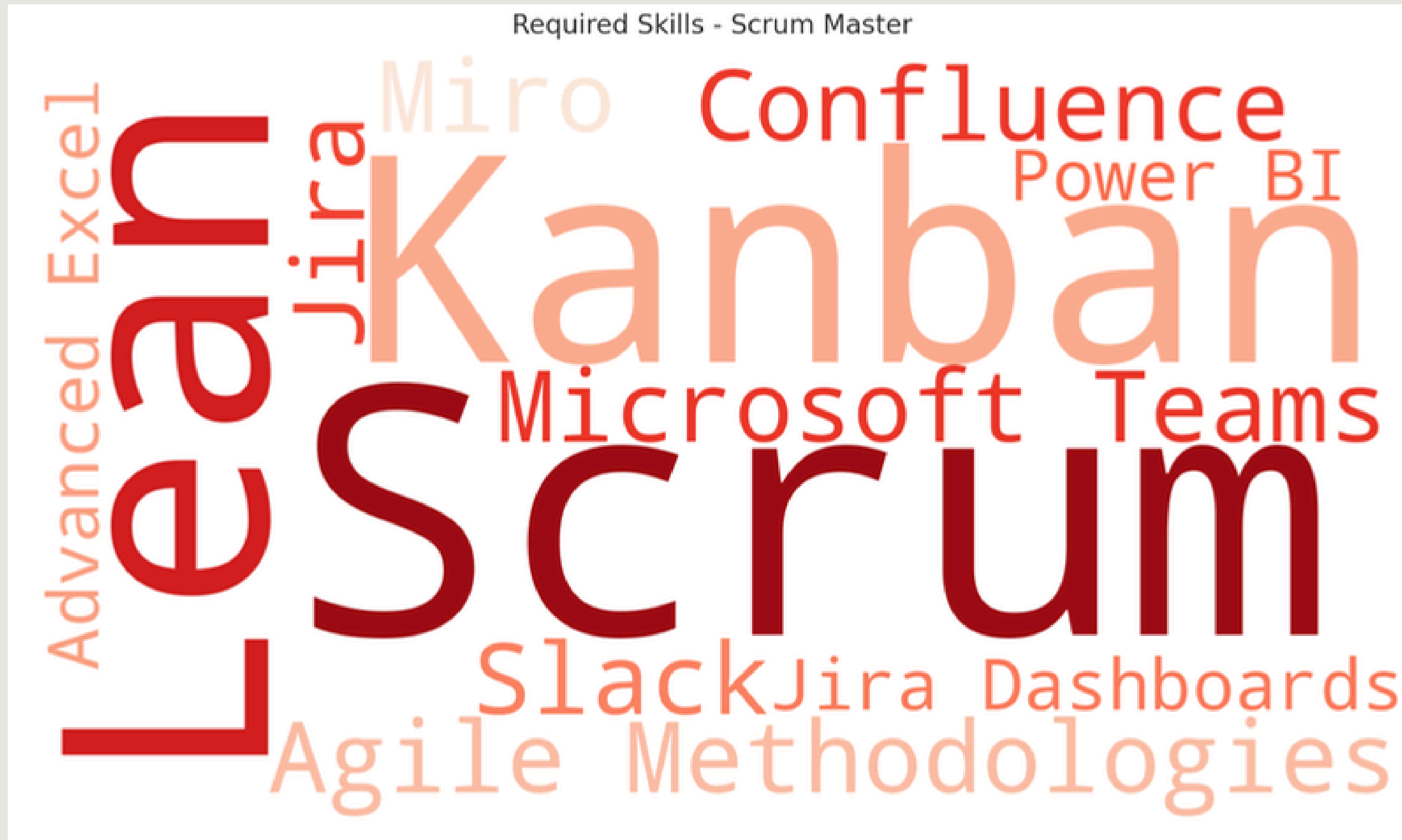
WORD CLOUD - DATA ANALYST



WORD CLOUD - DATA ENGINEER



WORD CLOUD - SCRUM MASTER



Top Candidates for Data Engineer



William Richards

Years of experience: 2.5

SQL
Microsoft Power BI
Databricks
DAX
Matillion ETL
Cloudera Data Platform

Languages: French, English



Gina Austin

Years of experience: 2.5

dax
informatica
intelligent cloud services
powercenter
microsoft power bi
microsoft sql server
python
sql
snowflake
data cloud

Languages: French, English



Christina Dixon

Years of experience: 2.5

airbyte amazon web services atlassian
bitbucket atlassian jira software dbt data galaxy
databricks lakehouse platform dataiku denodo
platform fivetran github enterprise gitlab
google bigquery ibm cognos analytics ibm
datastage ibm db2 ibm netezza performance
server informatica intelligent cloud services
informatica powercenter lookout matillion etl
matlab microsoft azure microsoft azure data
factory microsoft azure data lake store
microsoft azure devops services microsoft
azure sql database microsoft azure synapse
analytics microsoft fabric microsoft power bi
microsoft sql server microstrategy oracle
business intelligence enterprise edition oracle
database pl/sql postgresql project management
python sql snaplogic intelligent integration
platform snowflake data cloud teradata
database terraform

Languages: French, English



Payne Paul

Years of experience: 2.5

dax
databricks l
akehouse
platform
microsoft
power bi
sql
snowflake
data
cloud

Languages: French, English

Top Candidates for Data Analyst



Nielsan Nancy

Years of experience: 2.5

atlassian jira
software microsoft power bi
microstrategy postgresql
python r sql snowflake data
cloud tableau desktop and
online teradata database
teradata vantage

Languages: French, English



Shelby green

Years of experience: 1.5

dax
informatica
intelligent cloud services
powercenter
microsoft power bi
microsoft sql server
python
sql
snowflake
data cloud

Languages: French, English



David Campbell

Years of experience: 2.0

atlassian jira software dax denodo
platform github enterprise google
bigquery google cloud platform google
vertex ai javascript lookout mdx matlab
microsoft analytics platform system
(ssas) microsoft azure microsoft azure
data factory microsoft azure devops
services microsoft azure sql database
microsoft power bi microsoft sql server
microsoft sql server integration services
(ssis) microsoft sql server reporting
services (ssrs) oracle database python r
sap businessobjects bi snowflake data
cloud t-sql vba

Languages: French, English



Cynthia Reeves

Years of experience: 2.0

atlassian jira software dax google
bigquery informatica powercenter
microsoft azure data lake store
microsoft azure devops services
microsoft power bi microsoft sql
server microstrategy postgresql
python qlik sense r sql snowflake data
cloud t-sql tableau desktop and online
talend open studio

Languages: French, English



Hines Brian

Years of experience: 2.0

amazon web services dax databricks
lakehouse platform domo github
enterprise google cloud platform
microsoft azure microsoft azure data
factory microsoft azure devops
services microsoft azure sql database
microsoft fabric microsoft power bi
microsoft sql server microsoft sql
server integration services (ssis)
project management python r sap
businessobjects bi sql snowflake data
cloud tableau desktop and online

Languages: French, English

Architecture - Microsoft Azure - Screenshots

RESOURCE GROUP

The screenshot shows the Microsoft Azure portal interface for a resource group named 'hec_case'. The left sidebar contains navigation links for Home, Overview, Activity log, Access control (IAM), Tags, Resource visualizer, Events, Settings, Cost Management, Monitoring, Automation, and Help. The main content area displays the 'Overview' tab for the 'hec_case' resource group. At the top, there is a search bar and a toolbar with actions like Create, Manage view, Delete resource group, Refresh, Export to CSV, Open query, Assign tags, Move, Delete, Export template, and Open in mobile. Below this, the 'Essentials' section shows the Subscription (Azure for Students), Subscription ID (2d5d09ea-1c88-4712-9d56-d8f22e4900ea), Tags (Add tags), Deployments (1 Failed, 4 Succeeded), and Location (Canada Central). The 'Resources' section shows a list of 8 resources, with filters for Name, Type, and Location. The resources are sorted by Name and are all located in Canada Central.

Name	Type	Location
ai7gguq7jod4y	Log Analytics workspace	Canada Central
solutionsbi	Application Insights	Canada Central
solutionsbi	Key vault	Canada Central
SolutionsBIConsultants (solutionsbiserver/SolutionsBIConsultants)	SQL database	Canada Central
solutionsbi	Storage account	Canada Central
solutionsbidata	Data factory (V2)	Canada Central
solutionsbidatafactory	SQL server	Canada Central
solutionsbiserver	Azure Machine Learning workspace	Canada Central

BLOB STORAGE

Microsoft Azure

Search resources, services, and docs (G+/I)

Copilot

berly.bijur@mail.mcgill.ca
MCGILL UNIVERSITY (PAC001) (ENL)

Home > hec_case > solutionsbidata

solutionsbidata | Containers

Storage account

Search

+ Container Change access level Restore containers Refresh Delete Give feedback

Search containers by prefix

Show deleted containers

Name	Last modified	Anonymous access level	Lease state
<input type="checkbox"/> logs	3/28/2025, 8:52:05 AM	Private	Available
<input type="checkbox"/> azureml	3/28/2025, 3:24:57 PM	Private	Available
<input type="checkbox"/> azureml-blobstore-32bd3f1e-e3af-4f48-b010-854294513efc	3/28/2025, 3:24:57 PM	Private	Available
<input type="checkbox"/> insights-logs-audit-event	3/28/2025, 3:26:35 PM	Private	Available
<input type="checkbox"/> insights-metrics-pt1m	3/28/2025, 3:31:05 PM	Private	Available
<input type="checkbox"/> revisions	3/28/2025, 3:38:40 PM	Private	Available
<input type="checkbox"/> snapshots	3/28/2025, 3:38:40 PM	Private	Available
<input type="checkbox"/> snapshotlogs	3/28/2025, 3:38:40 PM	Private	Available
<input type="checkbox"/> solutionsbiconsultantsdata	3/28/2025, 8:53:02 AM	Private	Available

Containers

File shares

solutionsbiconsultantsdata - site where each input and output file is processed in python notebooks in Azure ML

SQL DB

Microsoft Azure

Search resources, services, and docs (G+I)

Copilot

berly.biju@mail.mcgill.ca
MCGILL UNIVERSITY (MCGILL.ORG)

Home > hec_case > solutionsbiserver > SolutionsBIConsultants (solutionsbiserver/SolutionsBIConsultants)

SolutionsBIConsultants (solutionsbiserver/SolutionsBIConsultants) | Query editor (preview)

SQL database

Search

Login + New Query + Open query Feedback Getting started

Overview
Activity log
Tags
Diagnose and solve problems
Query editor (preview)
Mirror database in Fabric (preview)
Resource visualizer
Settings
Compute + storage
Connection strings
Properties
Locks
Data management
Replicas
Sync to other databases
Integrations
Power Platform
Security
Auditing
Ledger
Data Discovery & Classification
Dynamic Data Masking
Microsoft Defender for Cloud
Identity

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables
dbo.HEC_HEC_LANG
dbo.HEC_HEC_SKILLS
dbo.HEC_HEC_STAFFING
dbo.HEC_HEC_USER
dbo.HEC_HEC_XP
Views
Stored Procedures

Query 1 x Query 2 x

Run Cancel query Save query Export data as Show only Editor

```
1  
2  
3 SELECT * FROM HEC_HEC_USER WHERE ANNEES_XP >= 1;  
4  
5
```

Results Messages

Search to filter items...

USER_ID	LAST_NAME	FIRST_NAME	ANNEES_XP
2432194	Middleton	Allen	2.5
2433083	Welch	Cynthia	2.5
2433092	Navarro	Lauren	2.5
2433099	Lozano	Nicole	2.5
2433109	Haley	Catherine	2.5
2433111	Richards	William	2.5
2433112	Dixon	Christina	2.5
2433114	Hardin	Melissa	2.5
2433117	Black	Clayton	2.5

Azure Data Factory (movement of data from csv to consultant database)

Microsoft Azure | Data Factory | solutionsbidatafactory

Search factory and documentation

berly.biju@mail.mcgill.ca

Validate all | Publish all

Preview experience

Factory Resources

Filter resources by name

Pipelines

move_csv_to_sql

Change Data Capture (preview)

Datasets

HEC_EXP_CSV

HEC_EXP_SQL

HEC_LANG_CSV

HEC_LANG_SQL

HEC_SKILL_SQL

HEC_SKILLS_CSV

HEC_STAFFING_CSV

HEC_STAFFING_SQL

HEC_USER_CSV

HEC_USER_SQL

Data flows

Power Query

Activities

Search activities

Move and transform

Synapse

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Power Query

move_csv_to_sql

Validate | Debug | Add trigger

Copy data

skills_copy

Copy data

user_copy

Copy data

lang_copy

Copy data

staffing_copy

Copy data

exp_copy

Parameters

Variables

Settings

Output

Pipeline run ID: a3473017-5a0e-4bee-b85f-0b3ff03acc7

Pipeline status: Failed

View debug run consumption

All status

Showing 1 - 5 of 5 items

Activity name	Activity st...	Activ...	Run start	Duration	Integration runtime	User prop...	Activity run ID
exp_copy	Failed	Copy data	3/29/2025, 10:56:50 PM	16s	AutoResolveIntegrationRuntime (Canada Central)		cb1a273a-8b3f-4832-b
staffing_copy	Succeeded	Copy data	3/29/2025, 10:56:50 PM	17s	AutoResolveIntegrationRuntime (Canada Central)		7c1877fe-ee9f-4b42-b
skills_copy	Succeeded	Copy data	3/29/2025, 10:56:50 PM	14s	AutoResolveIntegrationRuntime (Canada Central)		6aa5b92b-49c6-4742-f
user_copy	Succeeded	Copy data	3/29/2025, 10:56:50 PM	14s	AutoResolveIntegrationRuntime (Canada Central)		968f59c5-f56c-4ed4-8f
lang_copy	Succeeded	Copy data	3/29/2025, 10:56:50 PM	16s	AutoResolveIntegrationRuntime (Canada Central)		a49feedf-e294-42f5-af

Azure ML - Site of all our Python notebooks

The screenshot displays the Azure Machine Learning Studio interface. The top navigation bar shows 'Azure AI | Machine Learning Studio' and 'Azure for Students SolutionsBIWorkspace'. The left sidebar contains a 'Notebooks' section with a file explorer showing a list of notebooks and files. The main area shows a notebook titled 'STEP3.3Mapping_Upc' with a code editor and a compute instance dropdown menu.

Notebooks

Files Samples

berlybiju

- job_features.json
- pdfs
- STEP2.1_Translating_Embedding_Mission_DSC_AZURE.ipynb
- STEP2.1_Translating_Embedding_Mission_DSC_LOCAL.ipynb
- STEP2.2_JSON_Matching_Cosine_Similarity_AZURE.ipynb
- STEP2.2_JSON_Matching_Cosine_Similarity_LOCAL.ipynb
- STEP3.1Pivot_ProcessingAZURE.ipynb
- STEP3.1Pivot_ProcessingLOCAL.ipynb
- STEP3.2_Invert_Months_ProcessingAZURE.ipynb
- STEP3.2_Invert_Months_ProcessingLOCAL.ipynb
- STEP3.3Mapping_Updated_AZURE.ipynb
- STEP3.3Mapping_Updated_LOCAL.ipynb
- Step_1_OCR_Job_desc_features_AZURE.ipynb
- Step_1_OCR_Job_desc_features_LOCAL.ipynb

tempcompute · Kernel idle CPU 20% RAM 32% Last saved a few seconds ago

4 0.0 300.0

1 final_scores.head()

(36) ✓

	USER_ID	weight_required_skills	weight_preferred_skills	weight_language	req_R_score	req_Atlasian JIRA Software_score	req_Microsoft Power BI_score	lang_French_score	lan
0	2843838	9	0	5	0.0	0.0	0.0	300.0	
1	2479537	9	0	5	0.0	0.0	0.0	300.0	
2	2533337	9	0	5	0.0	0.0	0.0	300.0	
3	2446382	9	0	5	80.0	0.0	160.0	300.0	
4	2433124	9	0	5	0.0	0.0	0.0	300.0	

Compute instances

- tempcompute - Running
2 Cores, 14 GB (RAM), 28 GB (Disk), \$0.18/hr
- Azure Machine Learning Serverless Sp...
- Serverless Spark Compute - Available