

# SkillSync

*Matching the Right Talent with AI Precision*

**TEAM 5**

Yash Sethi

Nehal Joshi

Berly Biju

Ashraf Elrufaie

# Table Of Contents

- Approach & Methodology
- Algorithms Used
- Scoring Method
- Cloud Architecture
- Dashboard
- Limitations

# Technology & Algorithms Used

## *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



# Algorithm

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

1

### OCR Engine

We used EasyOCR 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

### Levenshtein 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

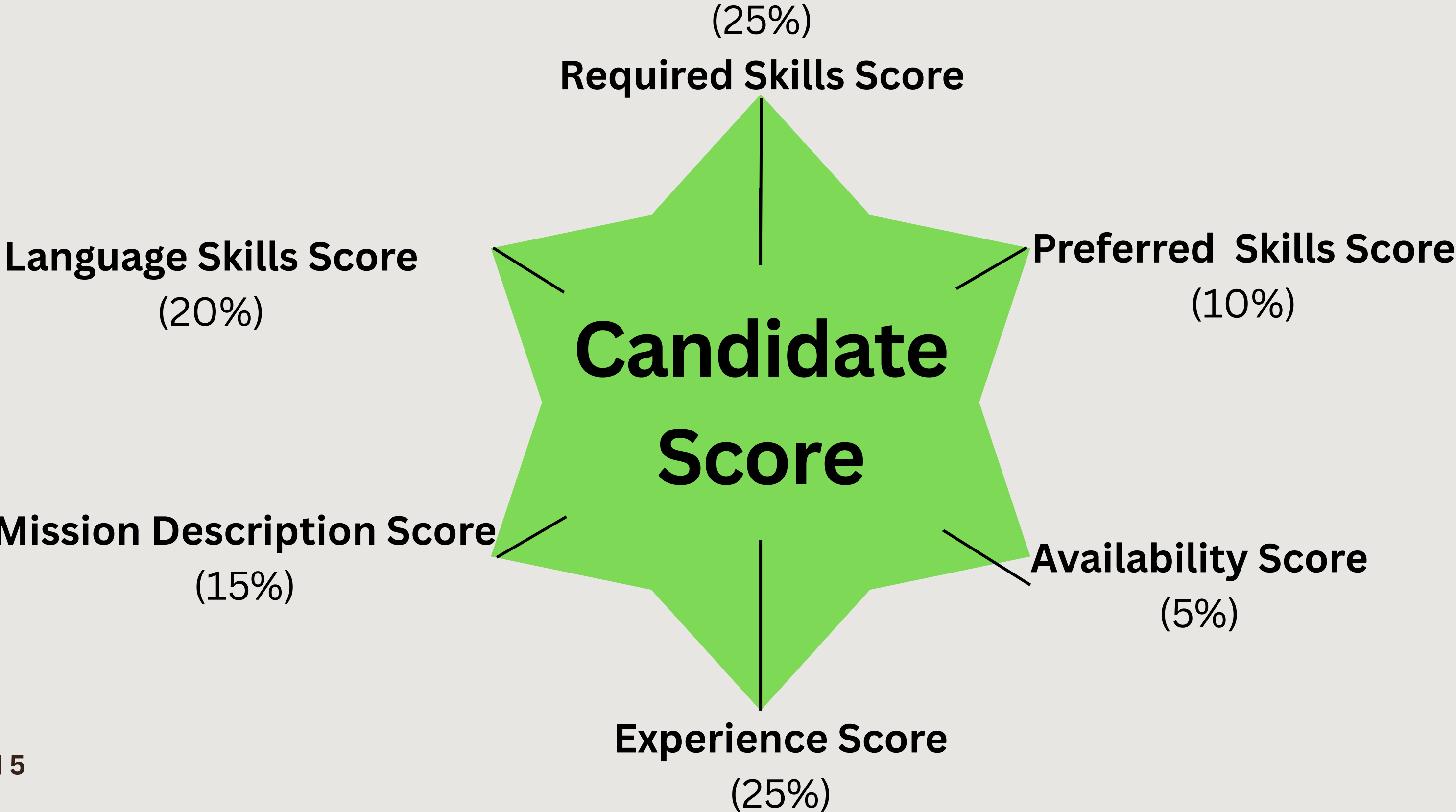
Using custom logic to check and filter consultants that will not be available for the duration of the job while assigning a score based on availability.

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



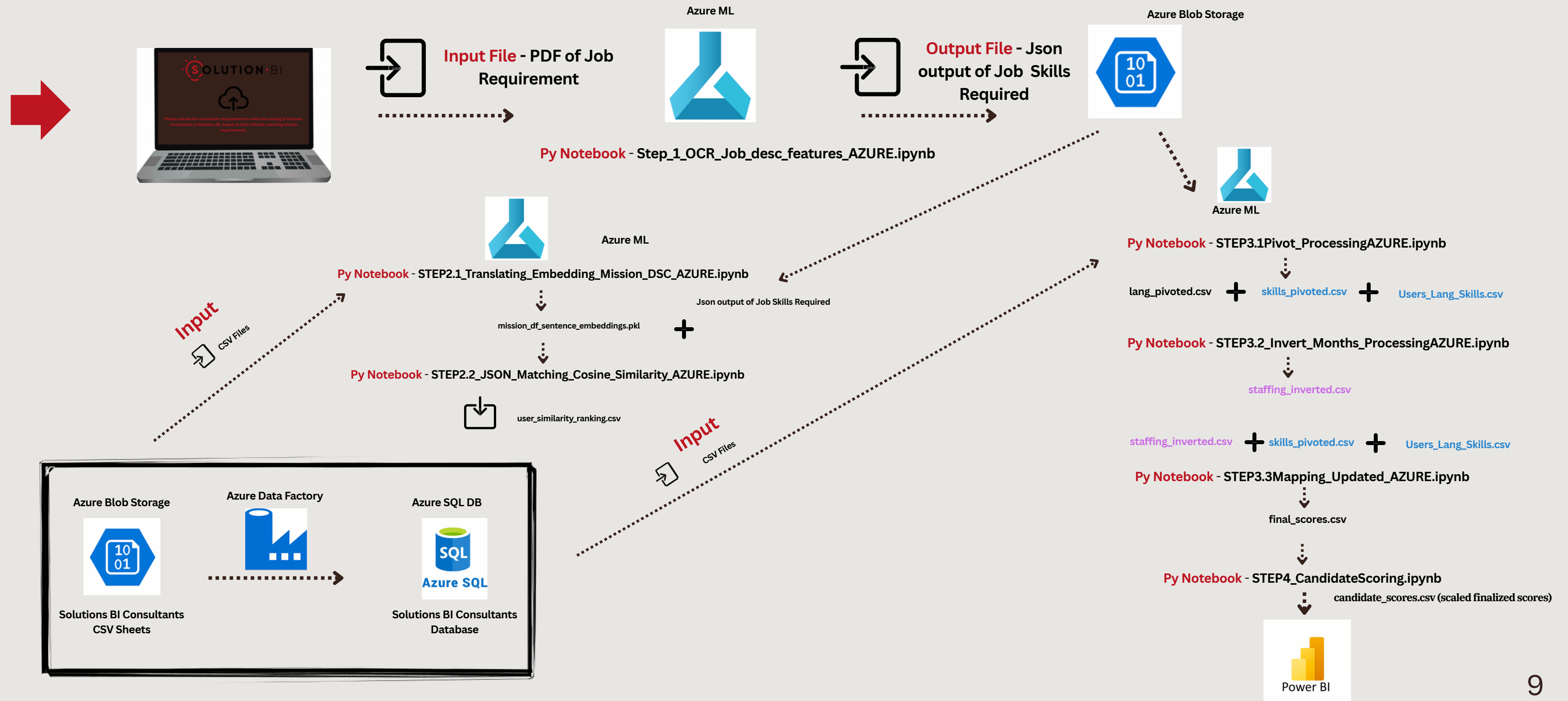


## USER INTERFACE

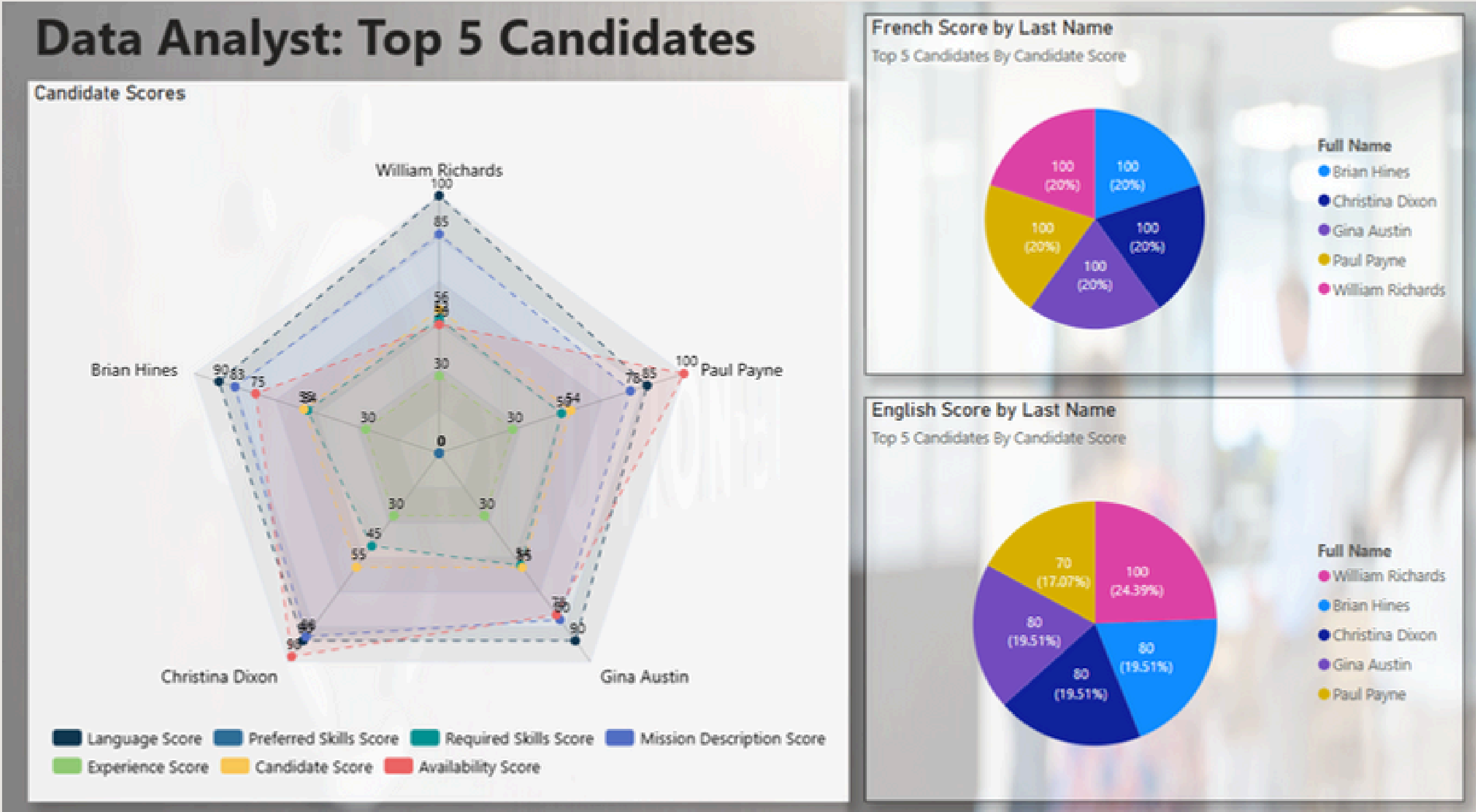


# Architecture Flow - Microsoft Azure - Detailed

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



# Dashboard



# Top Candidates for Data Analyst/Data Engineer



**William Richards**

**Years of experience: 2.0**

Atlassian Jira Software, DAX, Denodo Platform, GitHub Enterprise, Google BigQuery, Google Cloud Platform, Google Vertex AI, JavaScript, Looker, 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**



**Brian Hines**

**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**



**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, Looker, 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**



**Gina Austin**

**Years of experience: 1.5**

DAX, Informatica Intelligent Cloud Services, Informatica PowerCenter, Microsoft Power BI, Microsoft SQL Server, Python, SQL, Snowflake Data Cloud.

**Languages: French, English**



**Paul Payne**

**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**

# Top Candidates for Scrum



**Nielsan Nancy**

**Years of experience: 2.5**

Atlassian Jira Ssoftware  
Microsoft Power BI  
Microstrategy  
Postgresql  
Python R SQL  
Snowflake data cloud  
Tableau desktop  
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, Looker, 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 Business objects

**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, QlikSense, R, SQL, Snowflake  
data cloud, t-sql, Tableau desktop.

**Languages: French, English**



**Hines Brian**

**Years of experience: 2.0**

Amazon Web Services, DA 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**

# Limitations

## Scrum Master

Scrum Master's skills do not exist in the skills table

- Fix: Dynamically increase the weightage on the Mission Description similarity score

## Domain Description

- The Domain Description Column of the Skills Table was not used in the formulation of the score
- The Domain Description could have helped us generate a better score for Job Descriptions that do not have specific skills that overlap in our skills table

## Double Matching

- The similarity score generated from the Mission Description might include skills that also exist in the skills table, leading to a double count for the duplicated skills.
- Our goal for the similarity score was to capture skills that do not exist in the Skills table



# THANK YOU

*For Listening*

# APPENDIX

# Prompt-Engineering

```
You are an expert in job description analysis. Extract structured information from the following job description and categorize skills and languages using a 1-3 scale as per the given guidelines.

### Job Description:
{job_description}

### Guidelines for Scaling (1-3)

- **Required & Preferred Skills**:
  - **3** = Critical expertise (e.g., "Expert in Python", "2+ years experience in Kubernetes")
  - **2** = Required but not expert level (e.g., "Required: Java, C++")
  - **1** = Mentioned but not explicitly required (e.g., "Nice to have: AWS")

- **Languages**:
  - **3** = Critical requirement (e.g., "Fluency in French is essential")
  - **2** = Important but secondary (e.g., "Functional English required")
  - **1** = Nice to have (e.g., "Basic German knowledge preferred")

### Output Format (Strict JSON):
{{
  "Job Title": "",
  "Required Skills": [{"skill": "", "level": <1|2|3>}],
  "Preferred Skills": [{"skill": "", "level": <1|2|3>}],
  "Experience Required": ,
  "Languages": [{"language": "", "level": <1|2|3>}],
  "Responsibilities": ["", "..."],
  "Location": "",
  "Salary": "",
  "Additional Notes": {
    "Duration": "",
    "Type": "",
    "Mode of work": ""
  }
}}
```

Ensure the JSON is properly formatted, strictly follows the structure, and only includes relevant data.

## GPT 3.5 Turbo

```
Analyze the following job description and extract structured information with precise values:

Job Description:
{job_description}

Extracted Features:
- Job Title: (Extract only the exact job title)
- Required Skills: (List only the required skills as an array, e.g., ["Python", "SQL", "Azure"])
- Preferred Skills: (List only preferred skills as an array, e.g., ["Machine Learning", "AWS"])
- Experience Required: (Extract only the numeric value of years, e.g., 5.0 if the job states "5+ years of experience")
- Languages: (Extract languages explicitly mentioned in the job description, e.g., ["English", "French"])
- Responsibilities: (Extract job responsibilities in a structured bullet-point list)
- Location: (Extract exact location, if mentioned)
- Salary: (Extract exact salary range or numeric value if mentioned, otherwise leave empty)
- Additional Notes: (Any other relevant information without filler text)

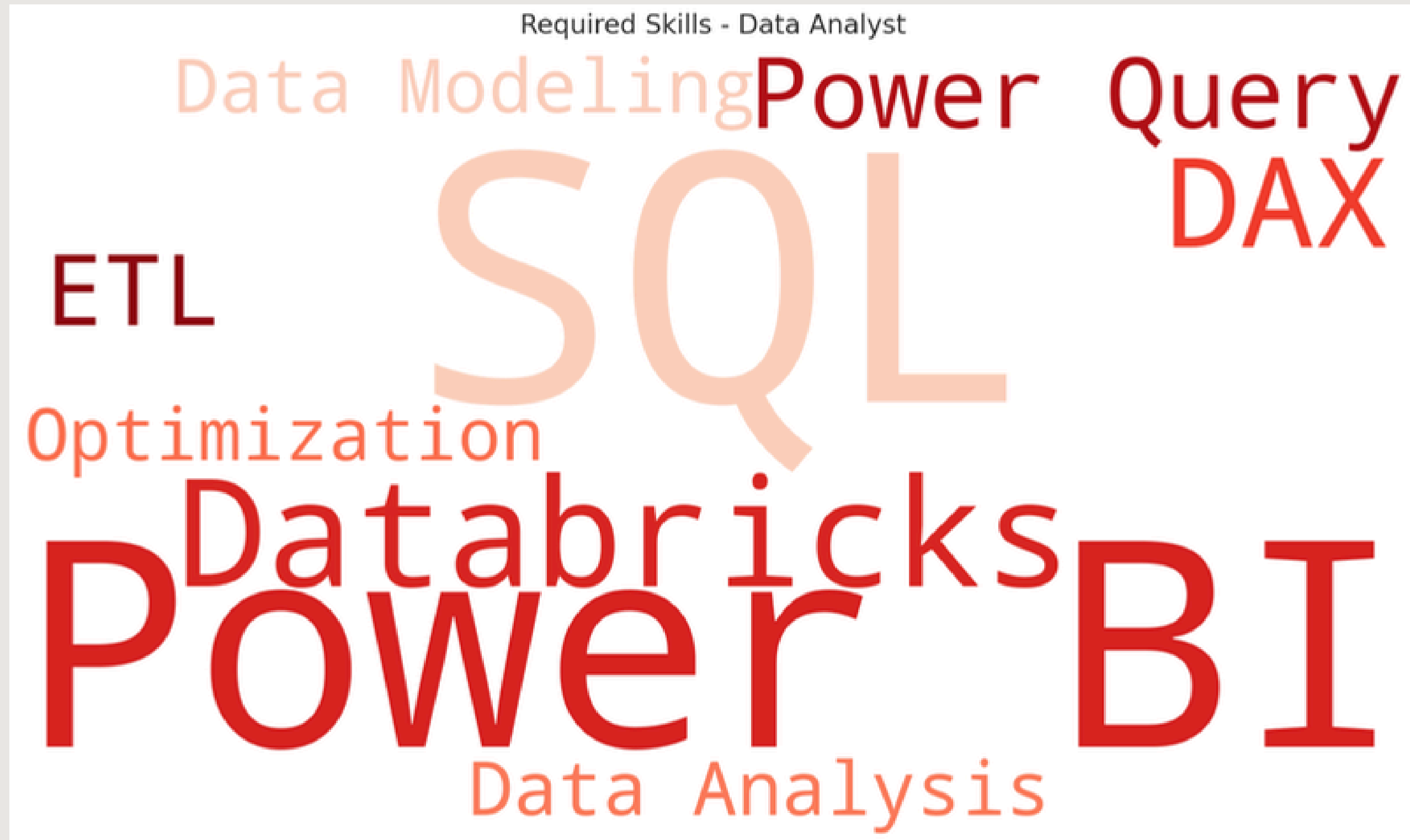
Provide the extracted information strictly in JSON format.
```

## ollama-mistral

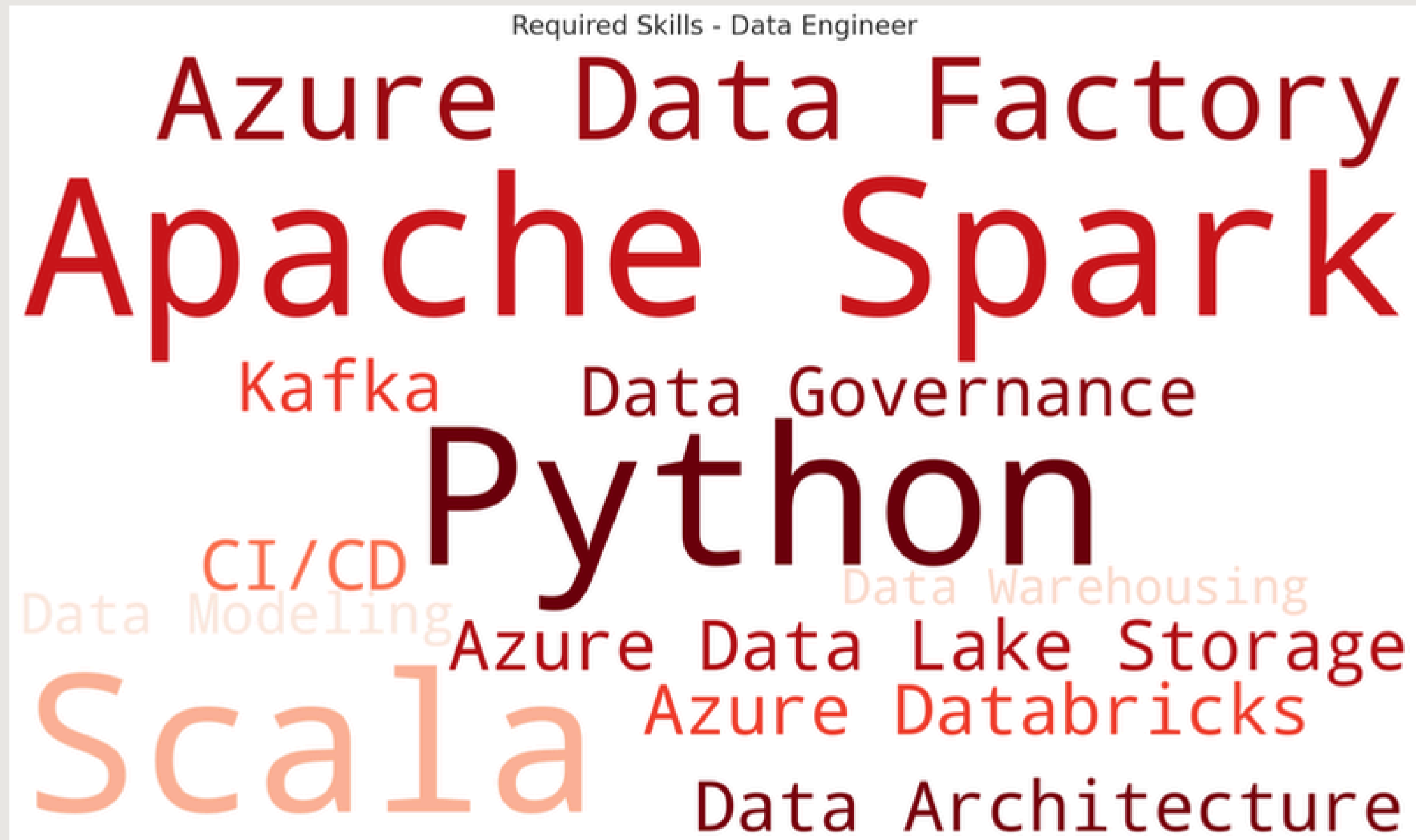
## GPT 3.5 Turbo - Output

```
{
  "Job Title": "Data Analyst - Marketing Analytics",
  "Required Skills": [
    { "skill": "SQL", "level": 3 },
    { "skill": "Power BI", "level": 3 },
    { "skill": "Databricks", "level": 3 },
    { "skill": "DAX", "level": 2 },
    { "skill": "Power Query", "level": 2 },
    { "skill": "ETL", "level": 2 },
    { "skill": "Data Analysis", "level": 1 },
    { "skill": "Data Modeling", "level": 1 },
    { "skill": "Optimization", "level": 1 }
  ],
  "Preferred Skills": [
    { "skill": "Marketing Data", "level": 1 }
  ],
  "Experience Required": 5.0,
  "Languages": [
    { "language": "English", "level": 3 },
    { "language": "French", "level": 3 }
  ],
  "Responsibilities": [
    "Manage and improve marketing datasets.",
    "Develop dashboards and reports in Power BI.",
    "Optimize the performance of Power BI datasets.",
    "Integrate data from various platforms (Google Analytics, CRM, marketing tools).",
    "Write efficient SQL, DAX, and Power Query code.",
    "Collaborate closely with analysts to understand their needs.",
    "Create queries in Databricks and SSMS.",
    "Ensure data quality and reliability.",
    "Support colleagues through code reviews and technical support.",
    "Design data models tailored to business objectives."
  ],
  "Location": "Montreal",
  "Additional Notes": {
    "Duration": "10 months (renewable)",
    "Type": "Full-time",
    "Mode of work": "Hybrid (2 days per week in the office in Montreal)"
  }
}
```

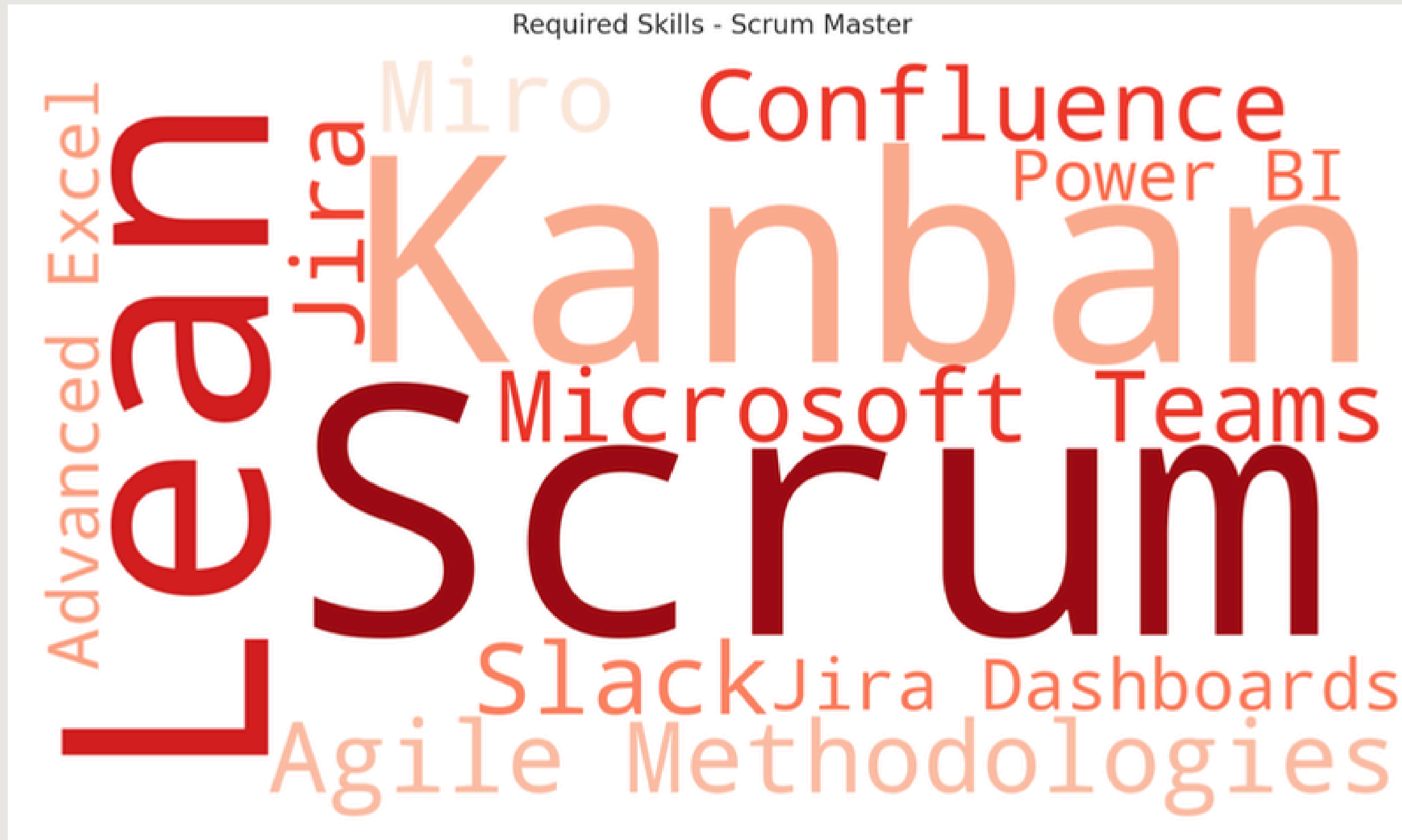
## WORD CLOUD - DATA ANALYST



## WORD CLOUD - DATA ENGINEER

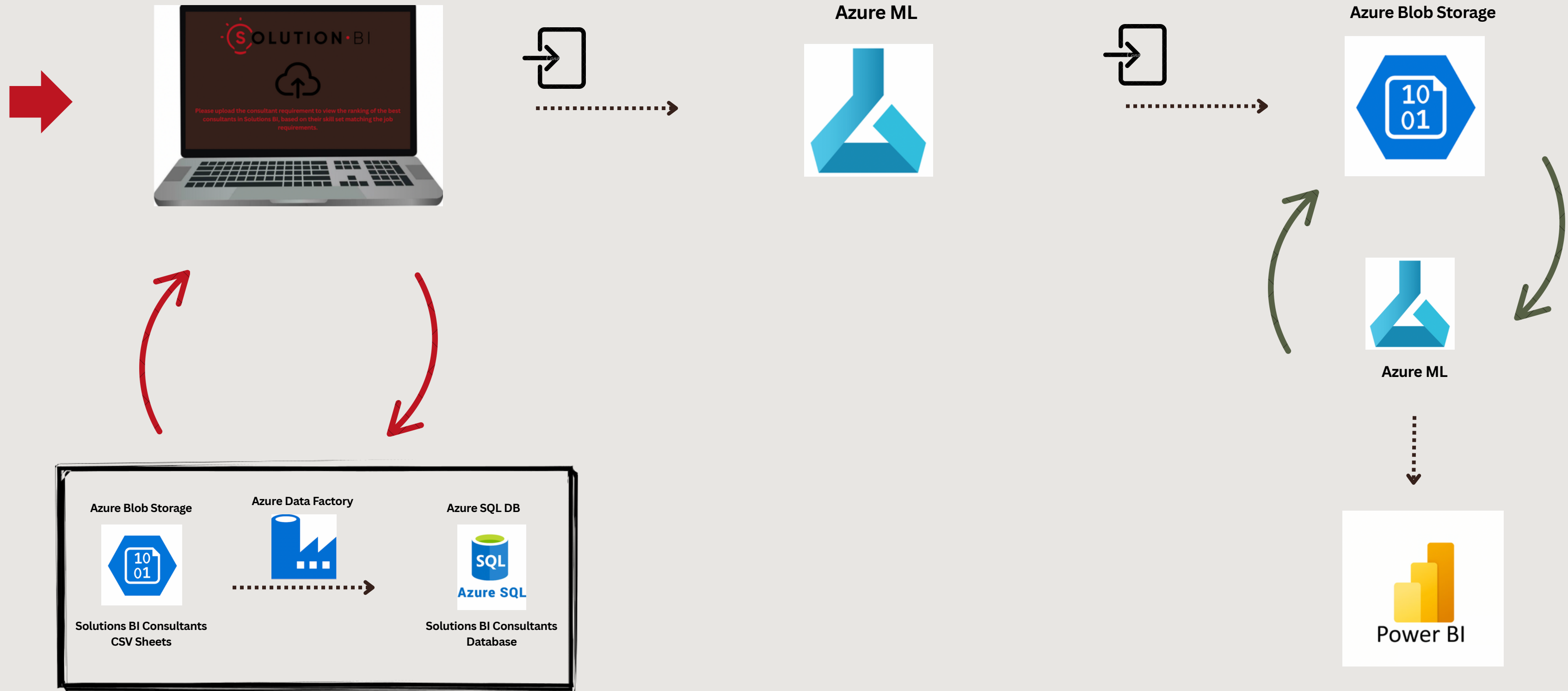


## WORD CLOUD - SCRUM MASTER





# Architecture Flow - Microsoft Azure



# 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)@MCL

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"/> snapshotslips	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 Query 2

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\_CSV

HEC\_SKILL\_SQL

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

Azure AI | Machine Learning Studio

McGill University > SolutionsBIWorkspace > Notebooks

## 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

STEP3.3Mapping\_Updated\_AZURE.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_Atllassian 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

Python 3 (ipykernel)