

DS684  
Cloud Computing  
Week 10

# Regarding Labs and Assignments

- Class participation means more than Zoom attendance. You must actively participate in the discussion and labs, and answer questions.
- Must hit Submit button, otherwise no grade
- If you need extension in time, must send written request (**email**). Otherwise no grade and no makeup. Requests sent over Zoom chat do not count.
- For any technical difficulty (installation, Azure access, etc), you must send written explanation (**email**) before the deadline. Otherwise no grade and no makeup.

# Regarding Labs and Assignments

Please submit all your labs and assignments **BEFORE** final presentation begins!

No make-up submissions after this point

# Teaching Schedule

Week 7: Azure Synapse Analytics Part I: Data Warehouse

Week 8: Azure Synapse Analytics Part II: Data Engineering

Week 9: Visualization using Power BI

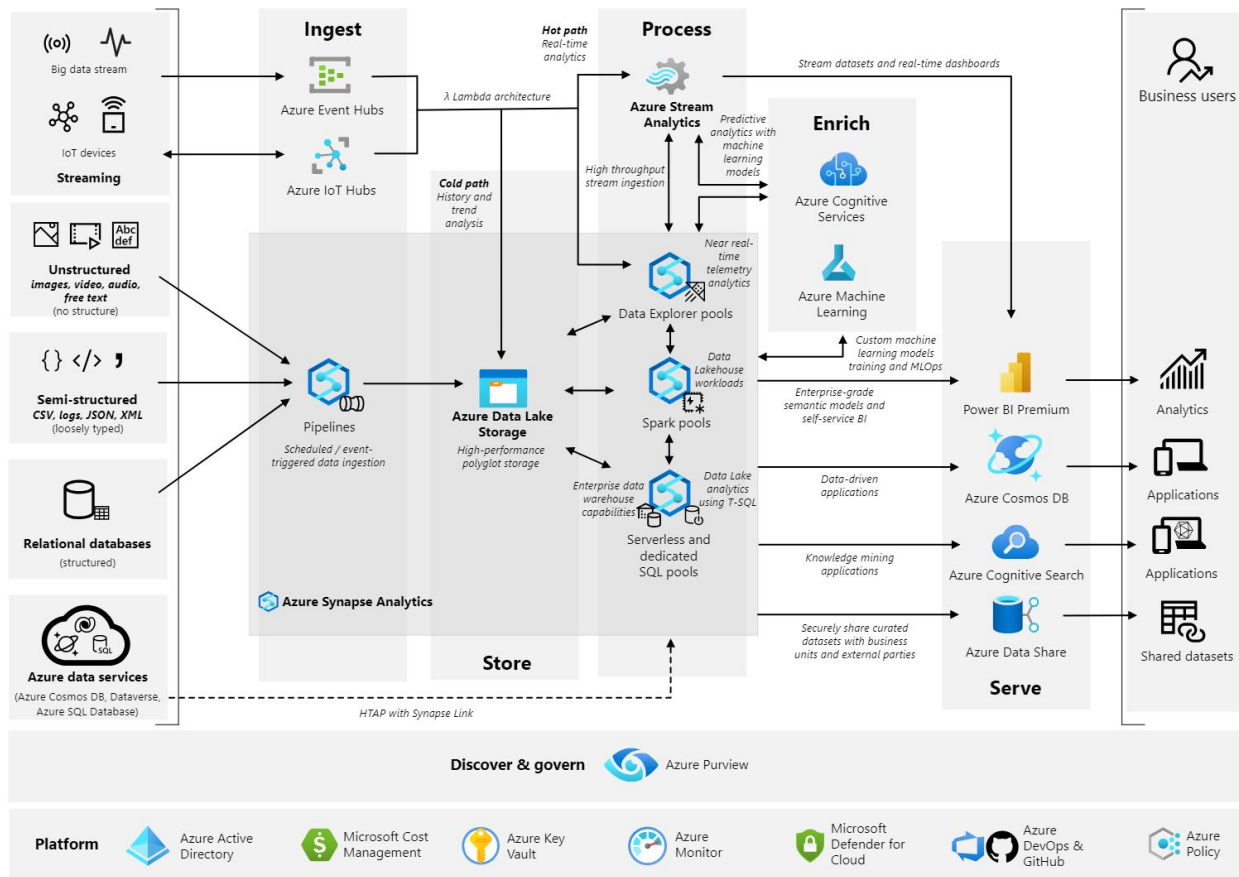
Week 10: Machine Learning and Artificial Intelligence

Week 11: Final project presentation

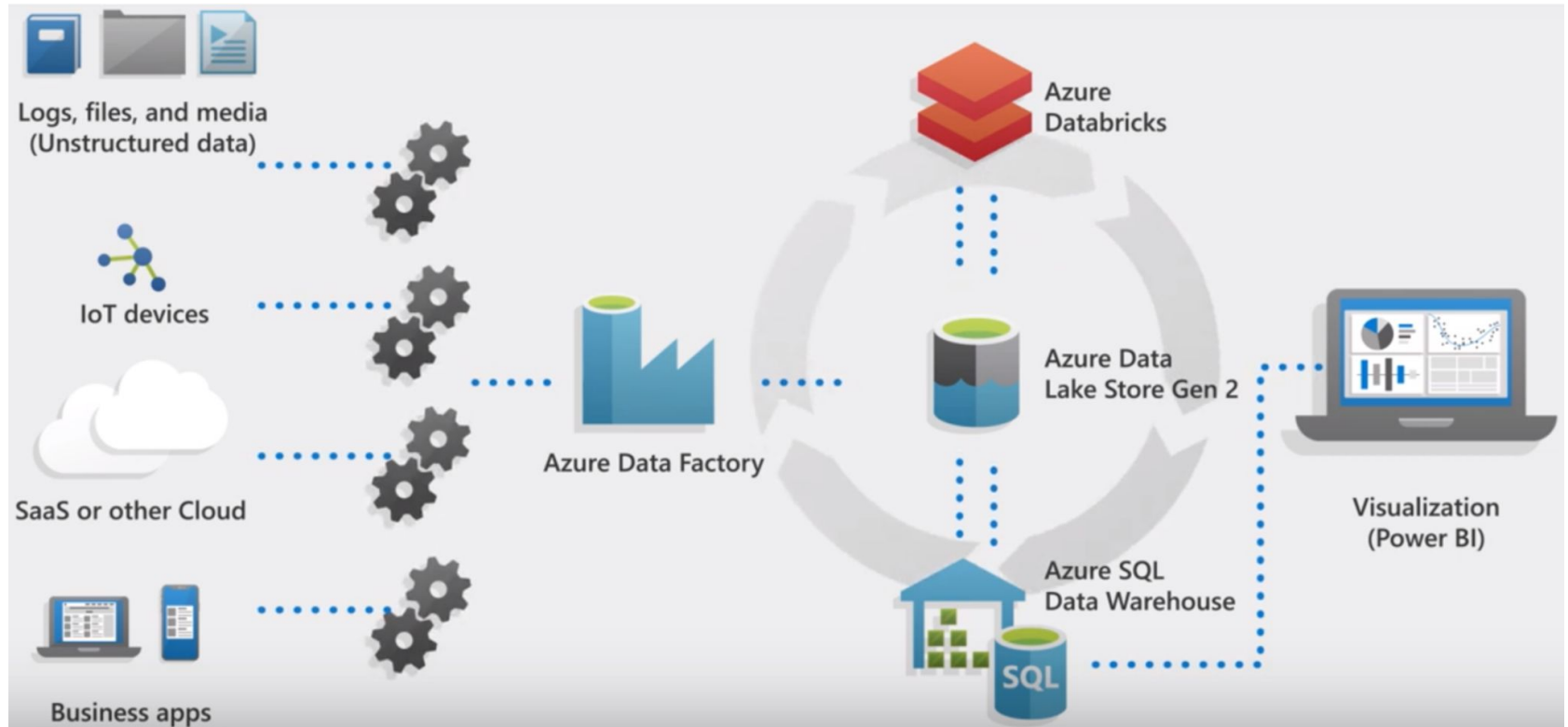
# Agenda

- Surviving Different Microsoft Data Services
  - Fabric demo
- Microsoft AI
  - Large Language Model (LLM) and Retrieval Augmented Generation (RAG)
  - Copilot
  - Azure AI Services
  - Azure Infrastructure
  - Azure Machine Learning
- Cloud Certifications

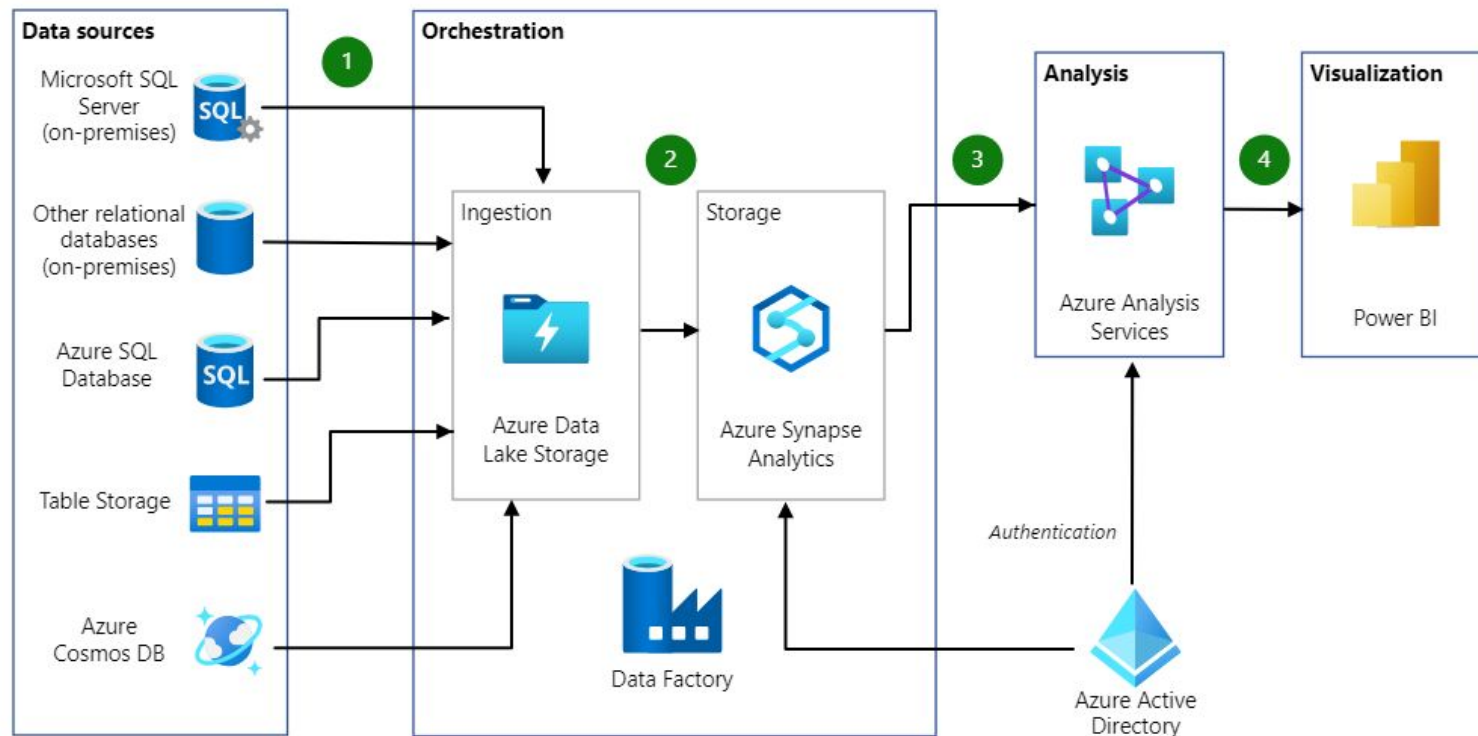
# Cloud Data Warehouse/ Lakehouse



# Back in 2017

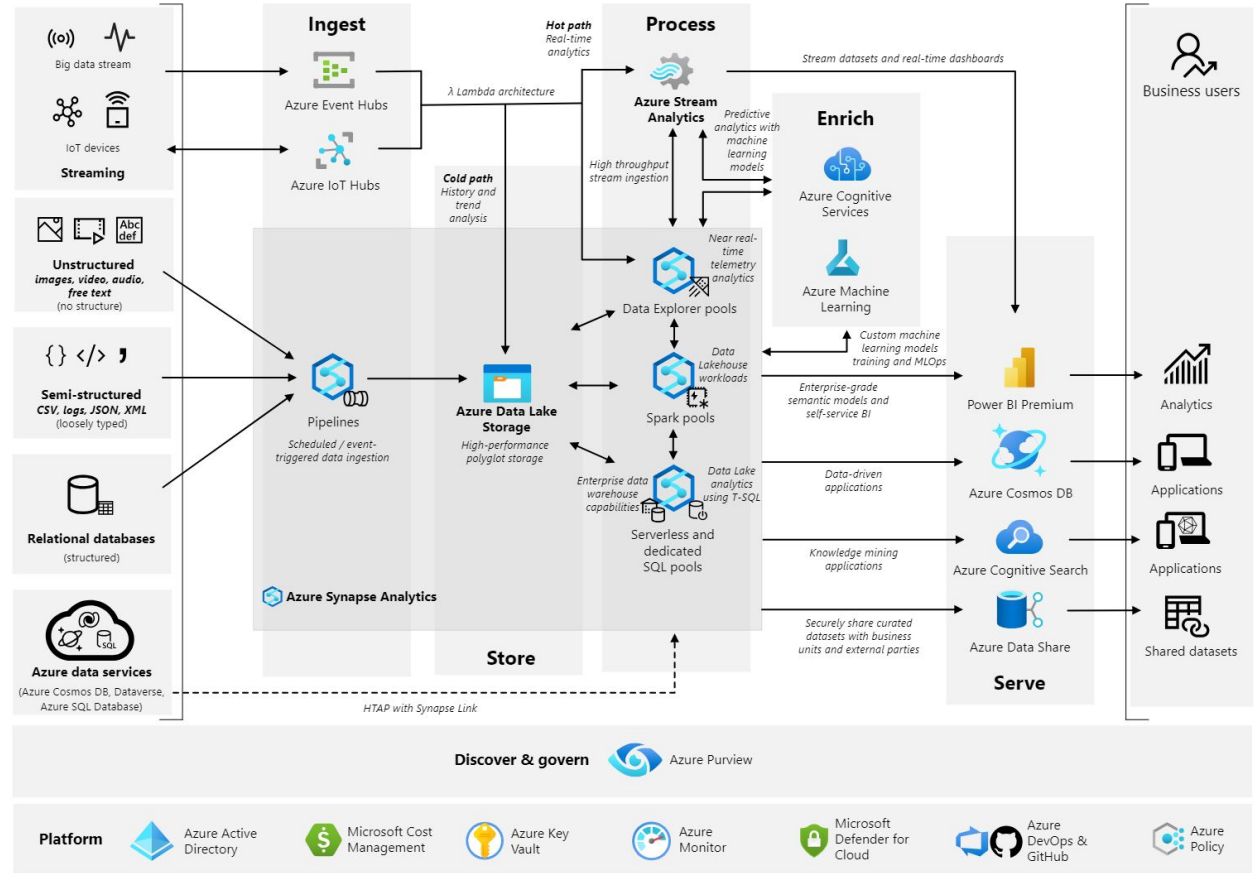


# How It Looked Like In 2021





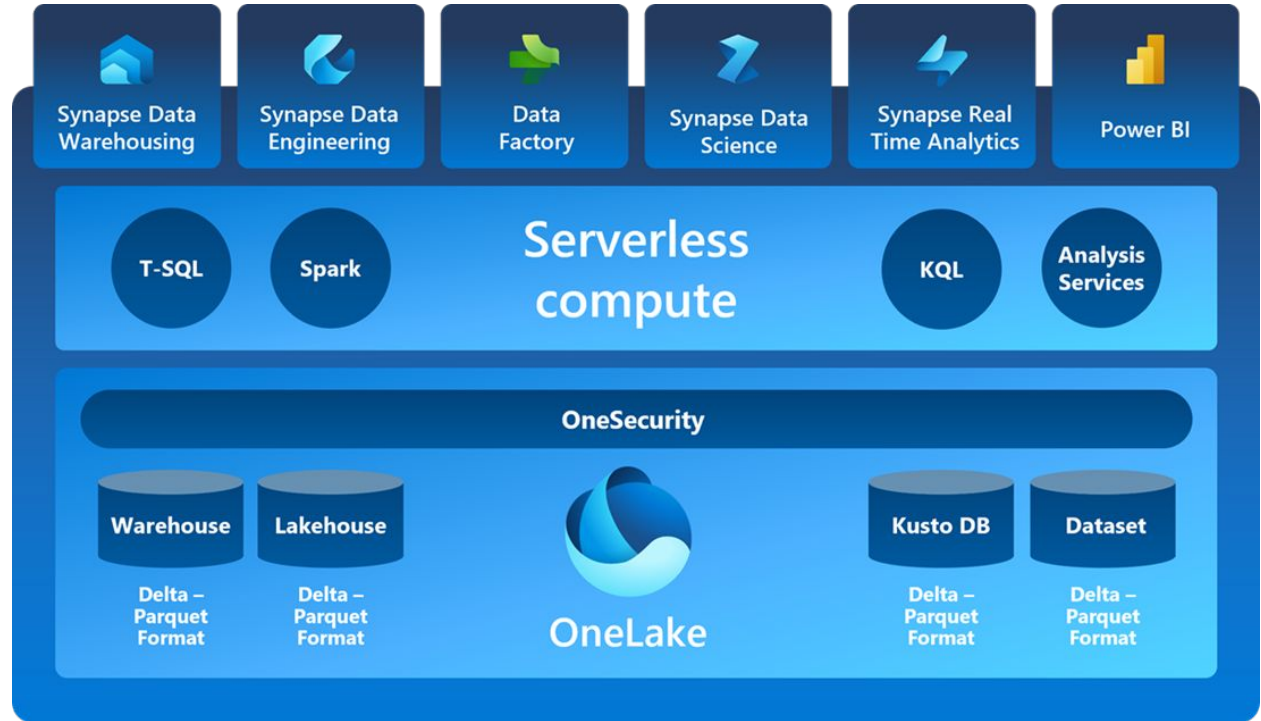
# Synapse Analytics



# Microsoft Fabric

A new product that is still not fully matured (released in Nov 2023).

A combination of some most popular Synapse features and PowerBI features



# Azure Data Factory

- Azure Data Factory: SSIS -> ADF -> Synapse Pipeline -> ADF
  - 95% similar to Synapse data pipeline, with some differences

There are some mixed marketing messages from Microsoft. At one point it seems that Microsoft will phase out Azure Data Factory. But now it seems that Microsoft will phase out Synapse Pipelines instead.

However since they pretty much share the same core, learning either one is fine.

We will use Synapse Pipelines in this class because it is a part of Synapse and will require less setup.

# Agenda

- Surviving Different Microsoft Data Services
  - Fabric demo
- Microsoft AI
  - Large Language Model (LLM) and Retrieval Augmented Generation (RAG)
  - Copilot
  - Azure AI Services
  - Azure Infrastructure
  - Azure Machine Learning
- Cloud Certifications

# Demos

Fabric Demo:

Different UI, same components, same functionalities

Azure Data Factory Demo

# But fundamentally...

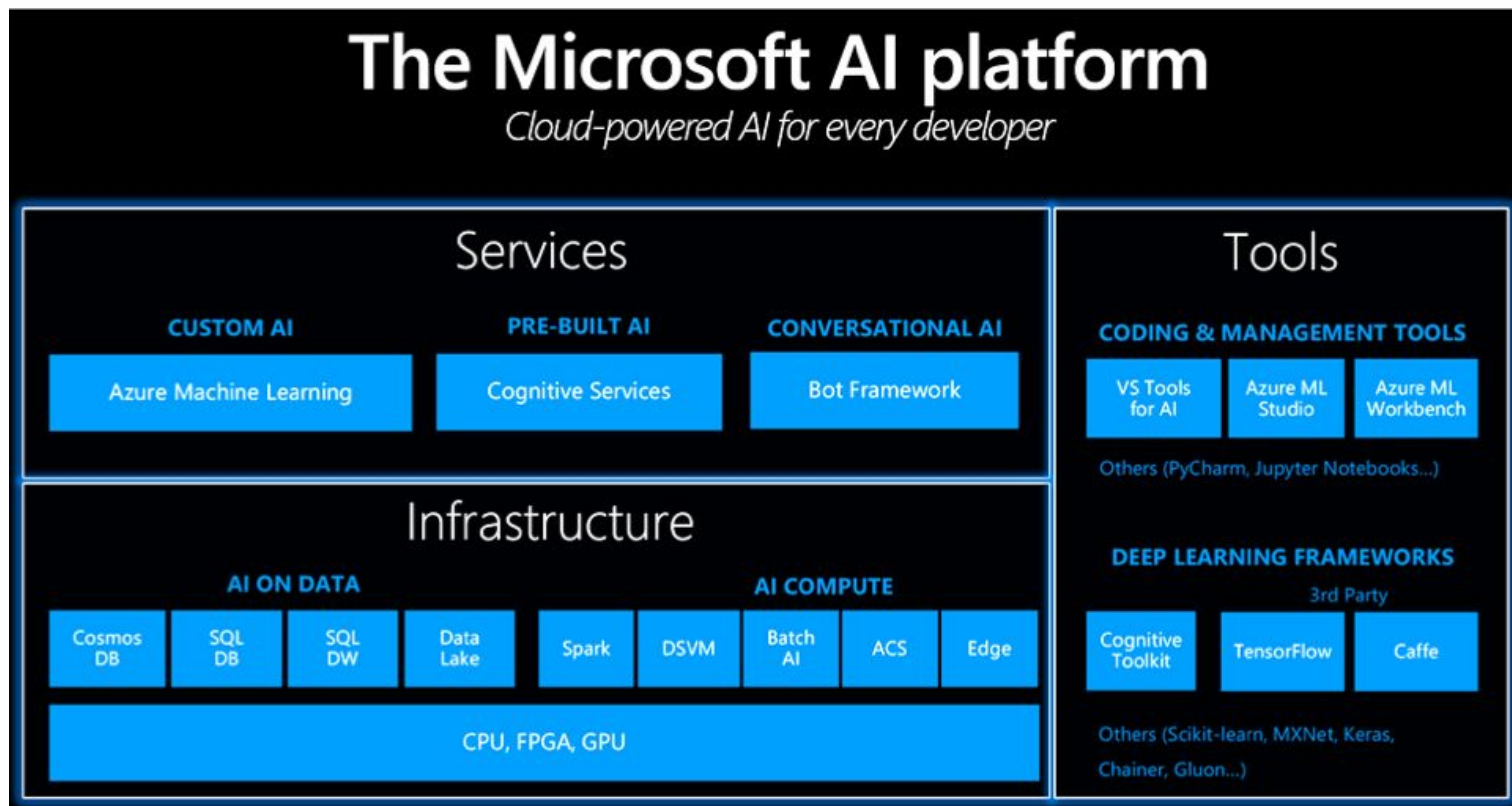
## Compute and Storage

- What do you want to do to the data, extracting (reading), transforming, loading (storing)?
  - Virtual machines
  - Azure data processing services
  - Your own SQL/python code
- Where are the datasets stored at, in what format?
  - Relational databases
  - Data lake files
  - NoSQL databases

# Agenda

- Surviving Different Microsoft Data Services
  - Fabric demo
- Microsoft AI
  - Large Language Model (LLM) and Retrieval Augmented Generation (RAG)
  - Copilot
  - Azure AI Services
  - Azure Infrastructure
  - Azure Machine Learning
- Cloud Certifications

# Microsoft AI Offerings





# Microsoft AI Offerings

Not only Azure, but the whole Microsoft

Microsoft AI: <https://www.microsoft.com/en-us/ai> > Solutions

- Many different AI products offered by different Microsoft products: Copilot
- Azure AI: <https://azure.microsoft.com/en-us/solutions/ai/>
  - Azure AI Services
  - Azure Machine Learning
  - Azure AI infrastructure
- OpenAI: LLM

# Agenda

- Surviving Different Microsoft Data Services
  - Fabric demo
- Microsoft AI
  - Large Language Model (LLM) and Retrieval Augmented Generation (RAG)
  - Copilot
  - Azure AI Services
  - Azure Infrastructure
  - Azure Machine Learning
- Cloud Certifications

# LLM and RAG

- Large Language Model (LLM)
  - Machine learning models that can comprehend and generate human language text. They work by analyzing massive data sets of language.
- Retrieval Augmented Generation (RAG)
  - Optimizing the output of a large language model, so it references an authoritative knowledge base outside of its training data sources before generating a response

# LLM AI as a Microsoft Offerings

LLM and OpenAI: <https://platform.openai.com/>

- Integrating LLM into existing frameworks
  - LLMs' capability and limit: Hallucination
  - Chunking and Embedding: customization for LLM answers
  - Similarity processing and vector database

# LLM Limitations

JU

what does this course teach me



I'd be happy to help you understand what a course teaches, but I need more specific information about the course you're referring to. Courses can cover a wide range of subjects and topics, so if you could provide the name of the course or some details about its content, I can give you a more accurate answer.

# Enhancing LLM with Customized Information

Find related information in your knowledge base, and add them as a part of your question to LLMs.



what does this course teach me The desired outcome of the course is the student's ability to put conceptual knowledge to practical use. Whether you are taking this course for future academic research, for work in industry, or for an innovative startup idea, this course should help you master the fundamentals of cloud computing.



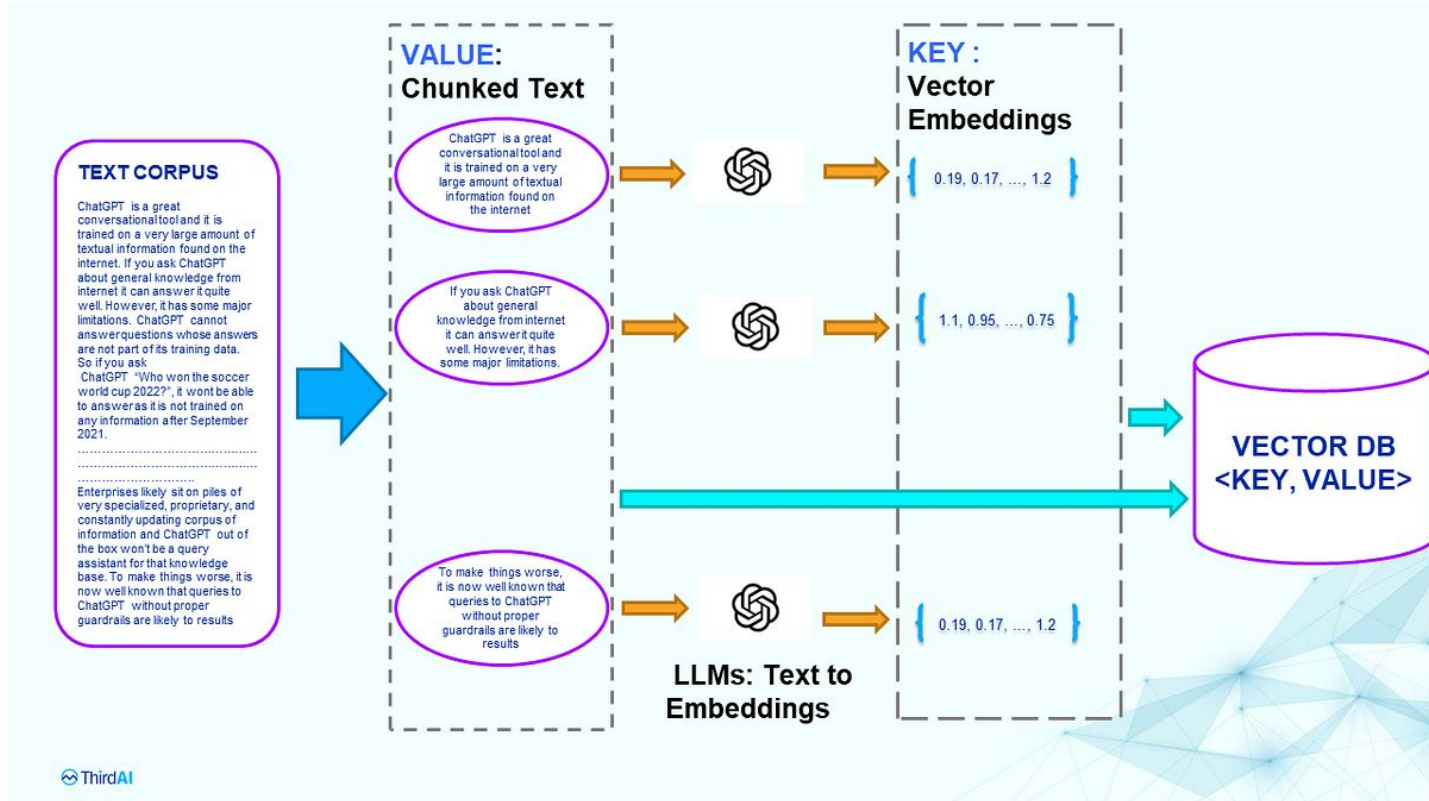
The course you described aims to teach students how to apply conceptual knowledge of cloud computing to practical situations. It has the following goals and intended outcomes:



1. **Conceptual Knowledge:** The course will provide you with a solid understanding of the theoretical and conceptual aspects of cloud computing. This might include topics like cloud architecture, virtualization, cloud service models (IaaS, PaaS, SaaS), security, scalability, and more.
2. **Practical Application:** The primary objective of the course is to equip you with the skills and knowledge necessary to apply what you've learned in real-world scenarios. This could involve tasks such as setting up and managing cloud-based infrastructure, deploying applications in the cloud, and optimizing cloud resources.

Regener

# Chunking and Embedding

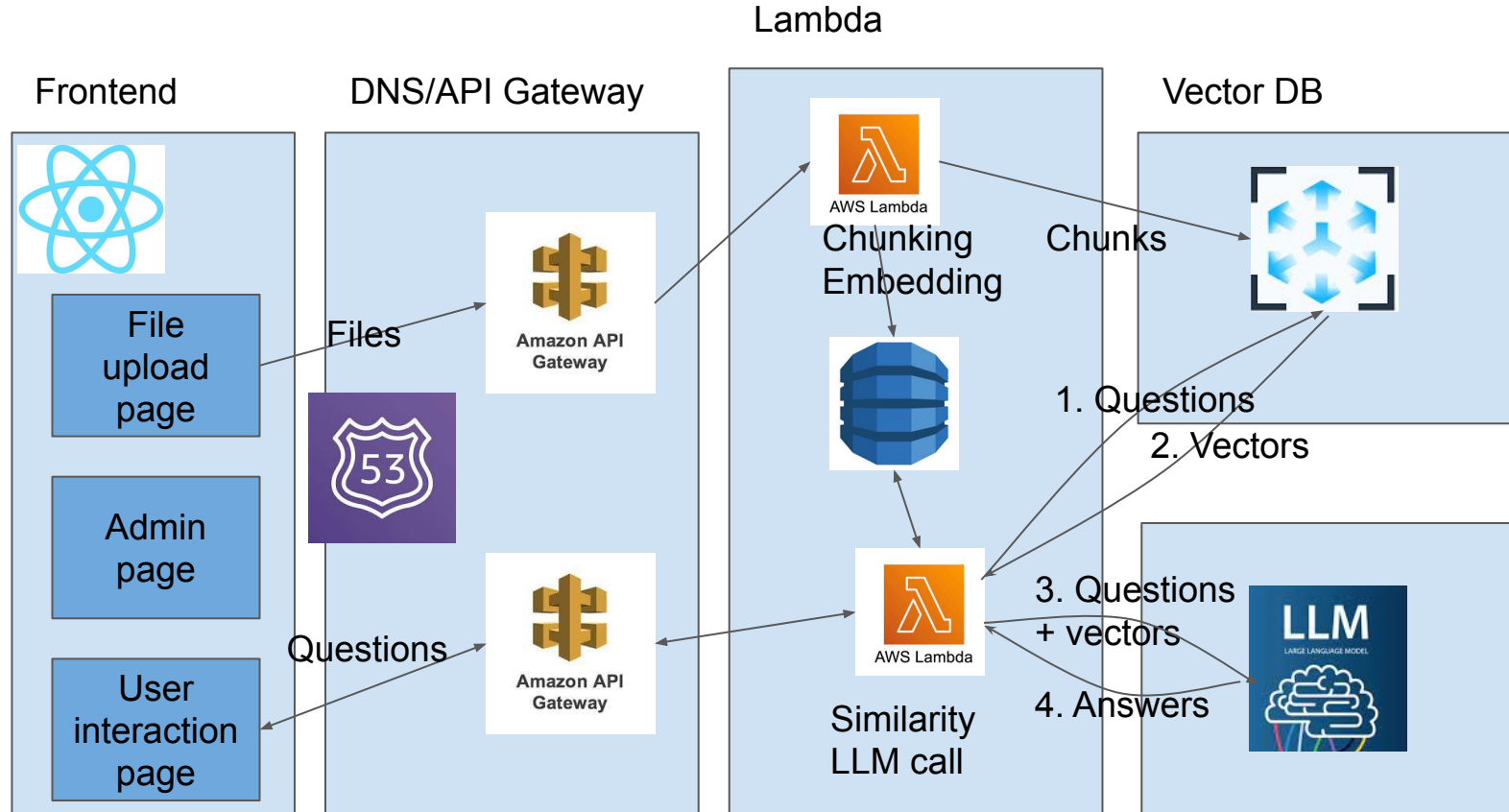


# Similarity Processing and LLM Invoking

- Similarity processing: Finding the Top n (3) vectors that are closest to your question
- Invoking LLM: The original question + the most similar vectors



# System Diagram Example



# Agenda

- Surviving Different Microsoft Data Services
  - Fabric demo
- Microsoft AI
  - Large Language Model (LLM) and Retrieval Augmented Generation (RAG)
  - Copilot
  - Azure AI Services
  - Azure Infrastructure
  - Azure Machine Learning
- Cloud Certifications

# Copilot

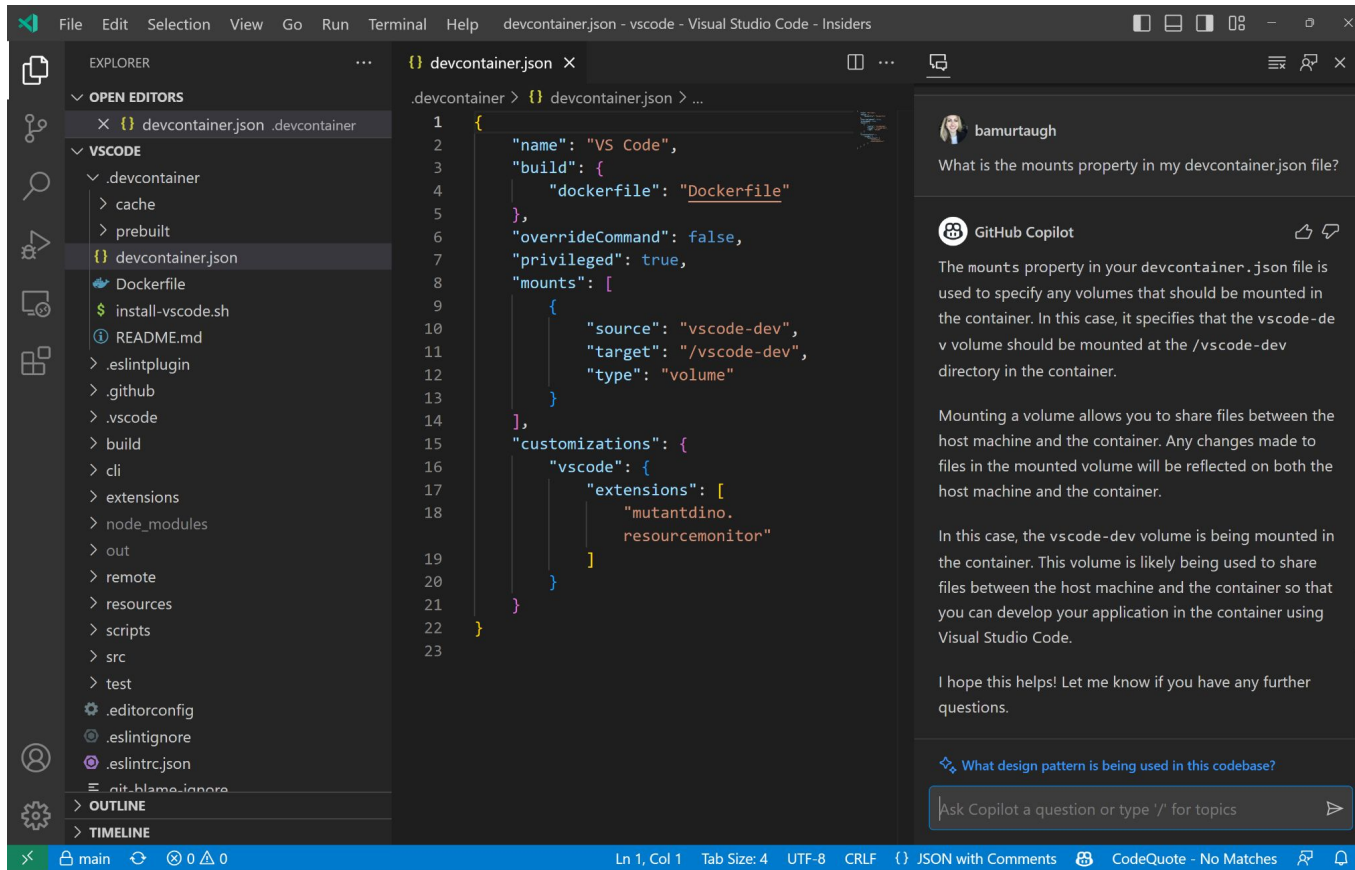
<https://adoption.microsoft.com/en-us/copilot/>

Example:

- Prepare a meeting by researching on previous emails
- Audio to text conversion
- Summarize meeting minutes, and create action items
- Summarize action results (Excel spreadsheet, Word doc, Github code) and
- Write a follow-up email

# Copilot

Provide help



The screenshot shows the Visual Studio Code editor with the `devcontainer.json` file open. The file content is as follows:

```
1 {
2   "name": "VS Code",
3   "build": {
4     "dockerfile": "Dockerfile"
5   },
6   "overrideCommand": false,
7   "privileged": true,
8   "mounts": [
9     {
10      "source": "vscode-dev",
11      "target": "/vscode-dev",
12      "type": "volume"
13    }
14  ],
15  "customizations": {
16    "vscode": {
17      "extensions": [
18        "mutantdino.resourcemonitor"
19      ]
20    }
21  }
22 }
23
```

On the right side, the GitHub Copilot chat interface is visible. It shows a conversation with a user named `bamurtaugh` asking: "What is the mounts property in my devcontainer.json file?". The Copilot response explains that the `mounts` property is used to specify volumes to be mounted in the container, specifically mentioning the `vscode-dev` volume and its target `/vscode-dev` directory. It also notes that mounting a volume allows sharing files between the host machine and the container.


At the bottom of the chat interface, there is a prompt: "What design pattern is being used in this codebase?" and a text input field with the placeholder "Ask Copilot a question or type '/' for topics".

# Copilot

Provide coding support

```
3
4 function bubbleSort(arr: number[]): number[] {
5     const n = arr.length;
6     for (let i = 0; i < n - 1; i++) {
7         for (let j = 0; j < n - i - 1; j++) {
8             if (arr[j] > arr[j + 1]) {
9                 const temp = arr[j];
10                arr[j] = arr[j + 1];
11                arr[j + 1] = temp;
12            }
13        }
14    }
15    return arr;
16 }
```

convert this to use quick sort

 Copilot generated code may be incorrect

# Office Copilot

Use CoPilot with Excel or Word or other office products

Streamline workflow  
(multiple meetings, e.g.)

The screenshot displays an Excel spreadsheet with three columns: Manufacturing Price, Sale Price, and Gross Sales. The data is organized into rows, with alternating light blue and white background colors. The Office Copilot sidebar is open on the right, showing a green header with the Copilot logo and a close button. Below the header, a green box contains the text "results and summarize three key trends". A large green-bordered box highlights the "Key trends:" section, which lists three bullet points: "Sales are trending up for all customers, except Proseware which is down this quarter.", "Manufacturing costs are down this quarter after an increase in Q3.", and "Low discount products made up nearly half of total sales last month." Below the trends, there is an "Explain" button with a magnifying glass icon and two thumbs-up/down icons. At the bottom of the sidebar, there are two buttons: "Filter Discount Band to Low" and "Add a margin column" with a refresh icon.

Manufacturing Price	Sale Price	Gross Sales
\$10	\$352	\$1,5
\$10	\$352	\$1,4
\$10	\$352	\$1,4
\$10	\$352	\$1,3
\$10	\$127	\$62
\$5	\$127	\$62
\$270	\$302	\$59
\$220	\$302	\$60
\$3	\$380	\$47
\$224	\$302	\$44
\$120	\$302	\$42
\$224	\$302	\$41
\$3	\$380	\$39
\$210	\$280	\$67

**Copilot**

results and summarize three key trends

Key trends:

- Sales are trending up for all customers, except Proseware which is down this quarter.
- Manufacturing costs are down this quarter after an increase in Q3.
- Low discount products made up nearly half of total sales last month.

Explain

Filter Discount Band to Low

Add a margin column

# Agenda

- Surviving Different Microsoft Data Services
  - Fabric demo
- Microsoft AI
  - Large Language Model (LLM) and Retrieval Augmented Generation (RAG)
  - Copilot
  - Azure AI Services
  - Azure Infrastructure
  - Azure Machine Learning
- Cloud Certifications

# Microsoft AI Offerings

Not only Azure, but the whole Microsoft

Microsoft AI: <https://www.microsoft.com/en-us/ai> > Solutions

- Many different AI products offered by different Microsoft products: Copilot
- Azure AI: <https://azure.microsoft.com/en-us/solutions/ai/>
  - Azure AI Services
  - Azure Machine Learning
  - Azure AI infrastructure



# Azure AI Services

Prepackaged AI Services that you can utilize: model trained by Microsoft

<https://azure.microsoft.com/en-us/products/ai-services/>

- Vision
- Speech
- Language
- Translator
- Documents
- Audio and video
- Decision
- Metrics Advisor

# AI Services

Usage example 1: Online customer support (real estate management)

- Image recognition (text labeling)
- Large Language Model text processing

Usage example 2: Online monitoring of customer support quality (healthcare)

- Speech to text
- Text sentiment analysis

# Agenda

- Surviving Different Microsoft Data Services
  - Fabric demo
- Microsoft AI
  - Large Language Model (LLM) and Retrieval Augmented Generation (RAG)
  - Copilot
  - Azure AI Services
  - Azure AI Infrastructure
  - Azure Machine Learning
- Cloud Certifications

# Azure AI Infrastructure

You run your own ML/AI, with Azure infrastructure

<https://azure.microsoft.com/en-us/solutions/high-performance-computing/ai-infrastructure/>

- VMs, with pre-configured images and GPUs

# Agenda

- Surviving Different Microsoft Data Services
  - Fabric demo
- Microsoft AI
  - Large Language Model (LLM) and Retrieval Augmented Generation (RAG)
  - Copilot
  - Azure AI Services
  - Azure Infrastructure
  - Azure Machine Learning
- Cloud Certifications

# Azure Machine Learning

You run your own AI/ML, with tools from Microsoft

<https://azure.microsoft.com/en-us/products/machine-learning/>

- Prepare data
- Build and train AI and ML models
- Validate and deploy
- Manage and monitor

# Azure Machine Learning Workspace Demo

<https://ml.azure.com/>

- Data: Data assets, Datastores
- Notebook
- Pipeline:
  - Data
  - Components: Data, Transformation, etc.
- Models and Jobs

# Agenda

- Surviving Different Microsoft Data Services
  - Fabric demo
- Microsoft AI
  - Large Language Model (LLM) and Retrieval Augmented Generation (RAG)
  - Copilot
  - Azure AI Services
  - Azure Infrastructure
  - Azure Machine Learning
- Cloud Certifications



# Microsoft Certifications

There are many certifications. The following slides only listed what I feel can be useful for your career.

Each certification costs time and money. So choose the path carefully

# Microsoft Certifications

## Fundamentals

Exam PL-900: Microsoft Power Platform Fundamentals
Exam AI-900: Microsoft Azure AI Fundamentals
Exam DP-900: Microsoft Azure Data Fundamentals
Exam AZ-900: Microsoft Azure Fundamentals

# Microsoft Certifications

Intermediate (data track) - Not official Microsoft term

Exam PL-300: Microsoft Power BI Data Analyst
Exam DP-100: Designing and Implementing a Data Science Solution on Azure
Exam DP-203: Data Engineering on Microsoft Azure
Exam DP-300: Administering Microsoft Azure SQL Solutions

# Microsoft Certifications

Intermediate (AI track) - Not official Microsoft term

Exam AI-102: Designing and Implementing a Microsoft Azure AI Solution

# Microsoft Certifications

Intermediate (Azure engineer track) - Not official Microsoft term

Exam AZ-104: Microsoft Azure Administrator
Exam AZ-204: Developing Solutions for Microsoft Azure

# Microsoft Certifications

Intermediate (Specialty track) - Not official Microsoft term

Exam DP-420: Designing and Implementing Cloud-Native Applications Using Microsoft Azure Cosmos DB
Exam AZ-500: Microsoft Azure Security Technologies
Exam AZ-700: Designing and Implementing Microsoft Azure Networking Solutions

# Microsoft Certifications

Somewhere between Intermediate and Advanced

Exam DP-500: Designing and Implementing Enterprise-Scale Analytics Solutions  
Using Microsoft Azure and Microsoft Power BI

# Microsoft Certifications

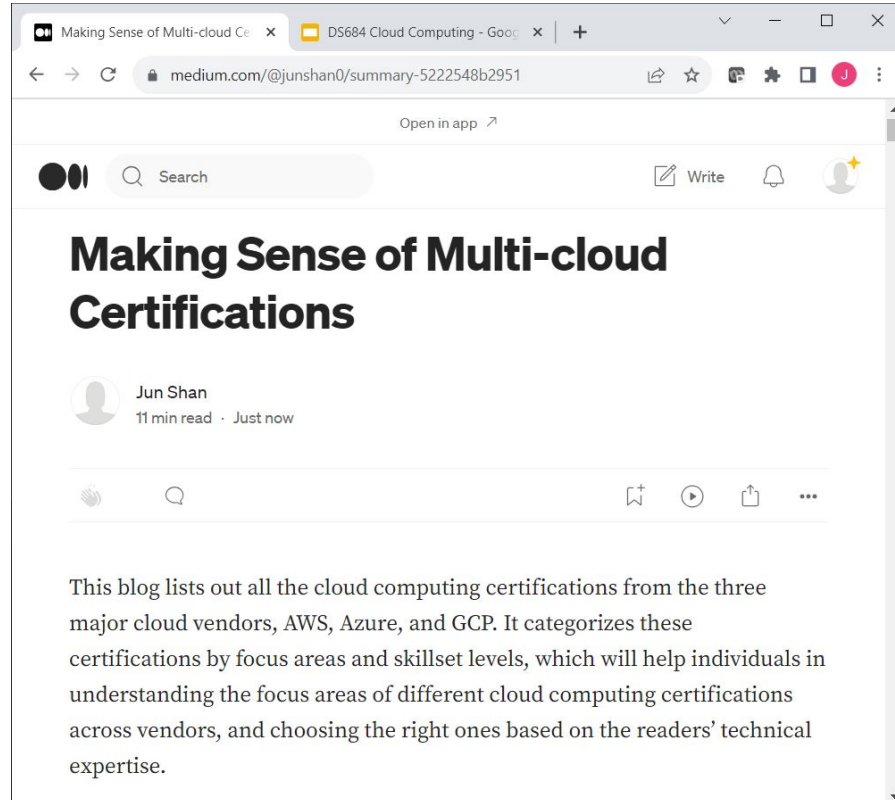
## Advanced

Exam AZ-305: Designing Microsoft Azure Infrastructure Solutions
Exam AZ-400: Designing and Implementing Microsoft DevOps Solutions
Exam PL-600: Microsoft Power Platform Solution Architect



# Multi-cloud Certifications

[https://medium.com/  
@junshan0/summary  
-5222548b2951](https://medium.com/@junshan0/summary-5222548b2951)



# Plan for Final Project Presentation

Expectations

Time slots