ReDI School of Digital Integration

Outlook

We use tech to connect human potential and opportunity with dignity & humility

Today’s Agenda

- Data Roles
- BI Tools
- Cloud Computing and SQL
- Chat GPT

- Surprise!
- Q&A

Data Roles (Review)

Data Analyst

"Data Analyst analyzes numeric data and uses it to help companies make better decisions."

Responsibilities

Pre-processing and data gathering
Representing data via reporting and visualization
Statistical analysis and data interpretation
Data maintenance
Communication with stakeholders

Data Scientist

"A data scientist analyzes and interpret complex data. They are data wranglers who organize (big) data."

Responsibilities

Data analytics and optimization using machine learning and deep learning
Strategic planning for data analytics
Statistical analysis and data interpretation

Data Engineer

"Data Engineer involves in preparing data. They develop, constructs, tests & maintain complete architecture."

Responsibilities

Develop, test, and maintain architectures

Deploy machine learning and statistical models

Building pipelines

Creating ETL/ELT operations

BI Tools

Tableau

Dash

Yellowfin

MicroStrategy

Microsoft Power BI

Looker

QlikView

And many more…

Domo

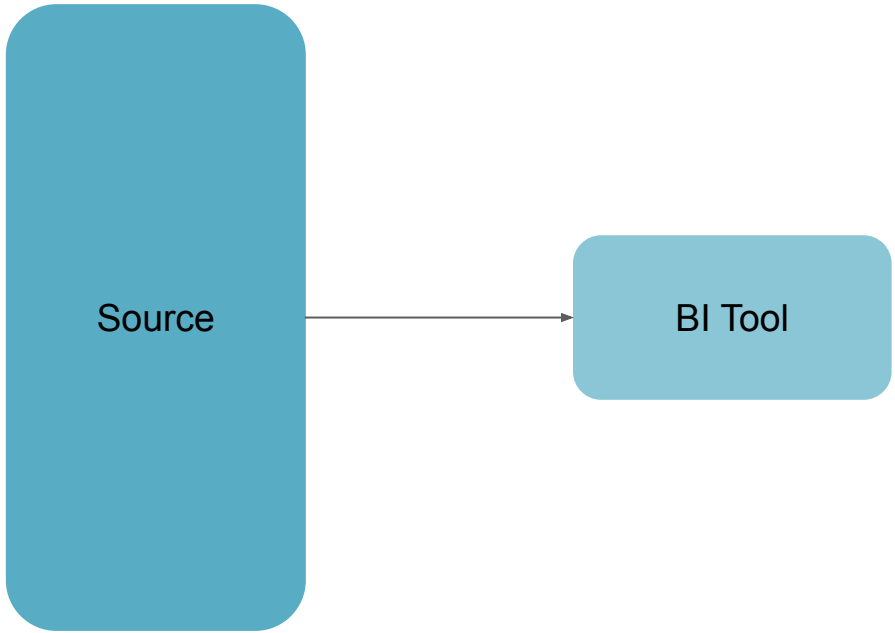
Sisense

IBM Cognos

SAP BusinessObjects

The underlying structure

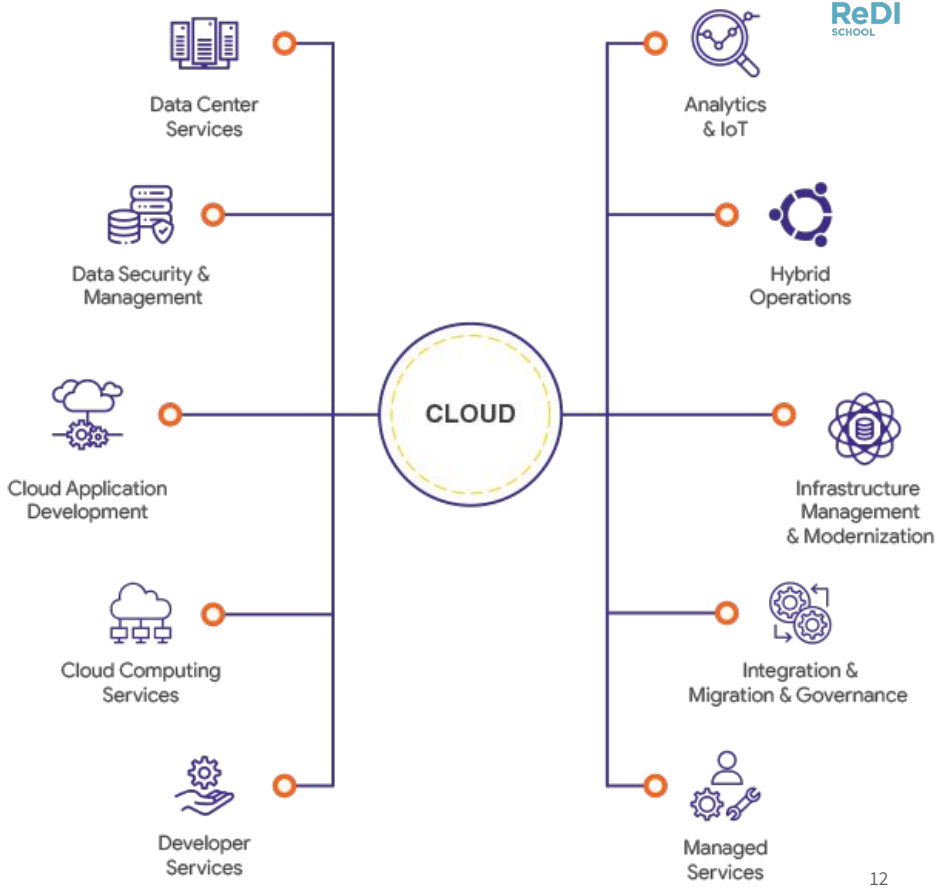
- They are all different tools with the same underlying structure (or architecture)
- Connect to the data
- Modify the data
- Learn about the data
- Visualize the data



Let’s look at an example in Power BI

Cloud Computing

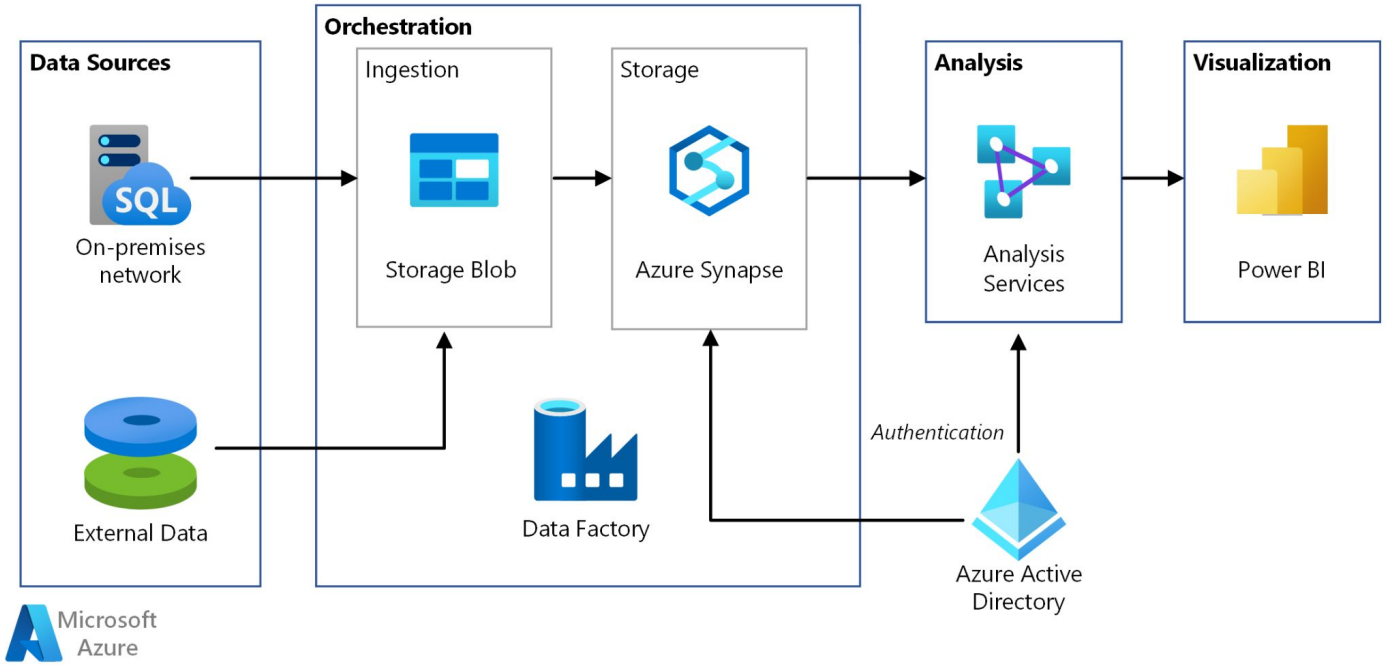
“The delivery of different services through the Internet which includes tools and applications like data storage, servers, databases, networking, and software.” - Investopedia



Benefits of Cloud Computing for DA and DE

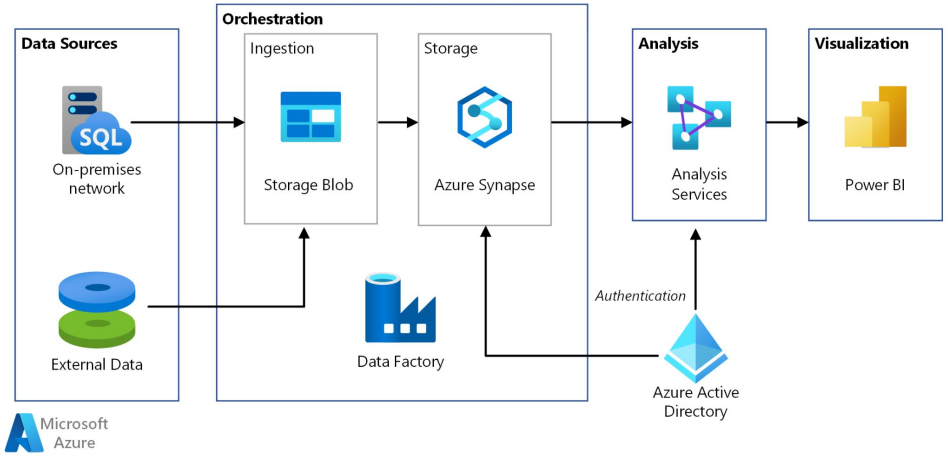
Scalability - Cloud platforms offer the ability to scale resources up or down based on demand, allowing data engineers and analysts to handle large datasets and complex workloads efficiently.
Elasticity - Cloud environments enable the dynamic allocation and deallocation of resources, ensuring optimal performance and cost-effectiveness.

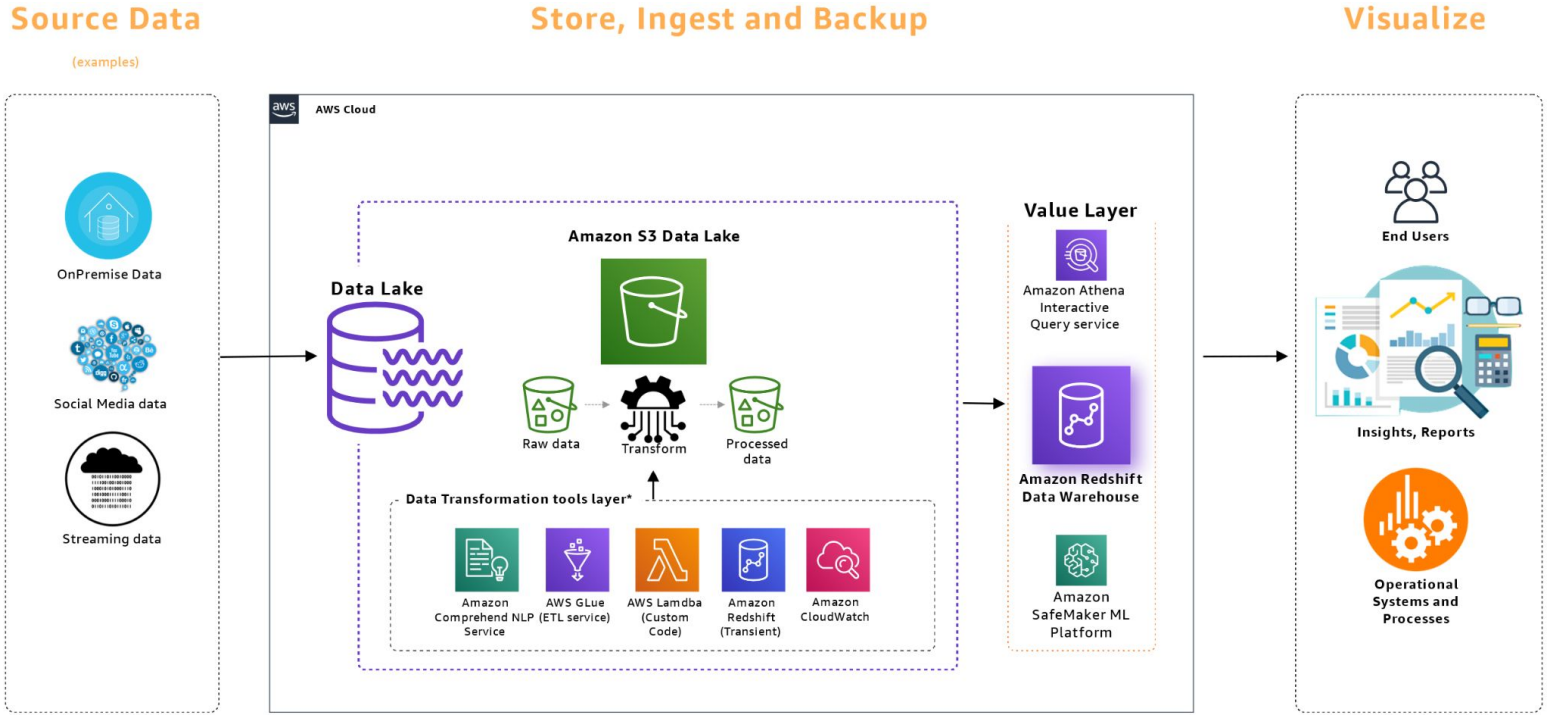
Cost Savings - By leveraging cloud services, organizations can reduce the need for upfront investments in hardware and infrastructure, resulting in cost savings.
Accessibility and Collaboration - Cloud platforms provide easy access to data and analytics tools from anywhere, enabling collaboration among teams and facilitating remote work.


Data Sources
On-premises network
External Data
Orchestration
Ingestion
Storage Blob
Storage
Azure Synapse
Data Factory
Analysis
Analysis Services
Visualization
Power BI
Authentication
Azure Active Directory
Microsoft Azure

Azure Synapse/Data Factory

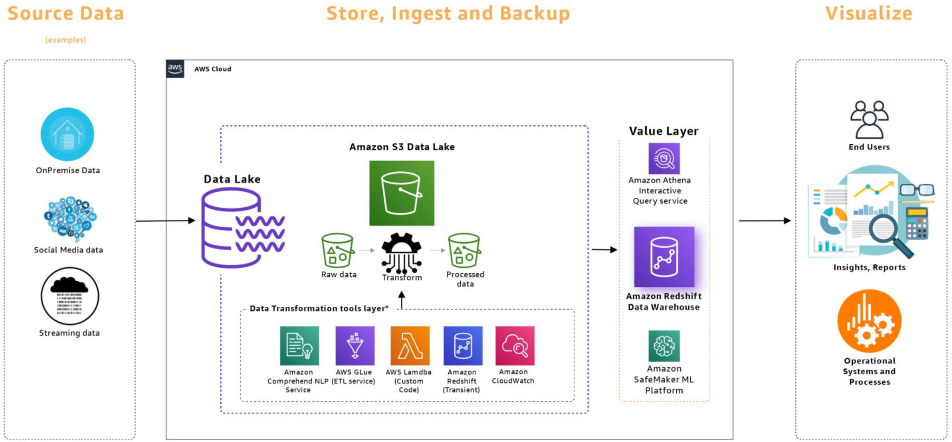
- Azure Synapse is a powerful analytics service that combines big data and data warehousing capabilities, allowing organizations to analyze vast amounts of data for valuable insights.
- It provides a unified and integrated platform for data engineering, data integration, and data warehousing, making it easier to manage and orchestrate data pipelines.
- Azure Synapse enables seamless integration with various data sources, both on-premises and in the cloud, facilitating the ingestion and transformation of data from diverse systems.
- With its built-in security and governance features, Azure Synapse ensures data protection and compliance, offering fine-grained control over access, encryption, and auditing.
- Azure Data Factory is a cloud-based data integration service that allows organizations to efficiently orchestrate and automate the movement and transformation of data across different sources and destinations, both within Azure and outside it.




Source Data
(examples)
Store, Ingest and Backup
Visualize
OnPremise Data
Social Media data
Streaming data
AWS Cloud
Data Lake
Amazon S3 Data Lake
Raw data
Transform
Processed data
Data Transformation tools layer*
Amazon Comprehend NLP Service
AWS GLue (ETL service)
AWS Lamdba (Custom Code)
Amazon Redshift (Transient)
Amazon CloudWatch
Value Layer
Amazon Athena Interactive Query service
Amazon Redshift Data Warehouse
Amazon SafeMaker ML Platform
End Users
Insights, Reports
Operational Systems and Processes

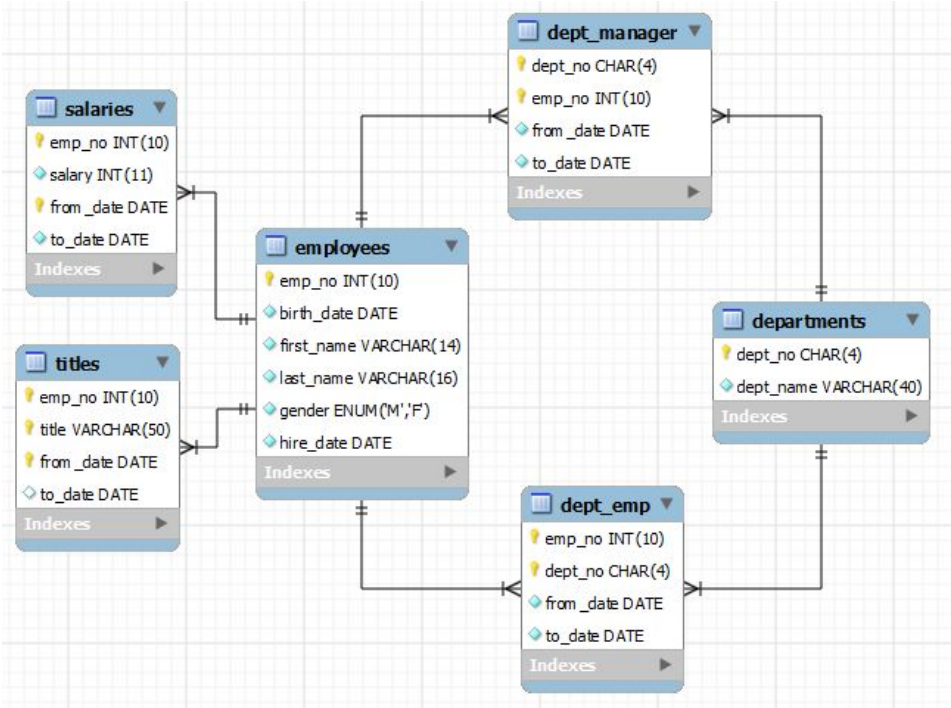
AWS

- AWS provides a range of services for data storage, processing, and analysis, enabling companies to securely store and manage their data in the cloud.
- With AWS, businesses can leverage scalable storage solutions like Amazon S3 to store and retrieve large volumes of data reliably and cost-effectively.
- AWS offers powerful data processing services, such as Amazon EMR (Elastic MapReduce) and AWS Glue, allowing companies to process, transform, and analyze data at scale.
- Companies can utilize AWS database services like Amazon RDS (Relational Database Service) and Amazon DynamoDB to store and manage structured and unstructured data, respectively.
- AWS provides advanced analytics capabilities through services like Amazon Redshift and Amazon Athena, empowering organizations to derive meaningful insights from their data through querying and data warehousing.



SQL - A (structured) query language

- SQL is a language for managing relational databases.
- It handles data retrieval, modification, and deletion.
- It’s used in various industries and applications.
- SQL helps companies streamline data management, improve efficiency, and make informed business decisions by effectively organizing and analyzing their data.



How does SQL works?

```
USE [SQL Tutorial]
GO
SELECT [Occupation]
      ,SUM([YearlyIncome]) as [Total Income]
      ,SUM([Sales]) AS [Total Sale]
      ,dbo.average([Occupation]) AS [Total Sale from Function]
  FROM [MyEmployees Table]
  group by [Occupation]
```

100 %

Results | Messages

©tutorialgateway.org

	Occupation	Total Income	Total Sale	Total Sale from Function
1	Clerical	190000	133.01	133.01
2	Management	320000	14181.84	14181.84
3	Professional	320000	10258.74	10258.74
4	Skilled Manual	110000	3019.0882	3019.0882

```
USE [SQL Tutorial]
GO
SELECT [EmployeeID]
      ,[FirstName]
      ,[LastName]
      ,[Education]
      ,[Occupation]
      ,[YearlyIncome]
      ,[Sales]
      ,[HireDate]
  FROM [MyEmployees Table]
  WHERE [Sales] >= dbo.AverageSale()
```

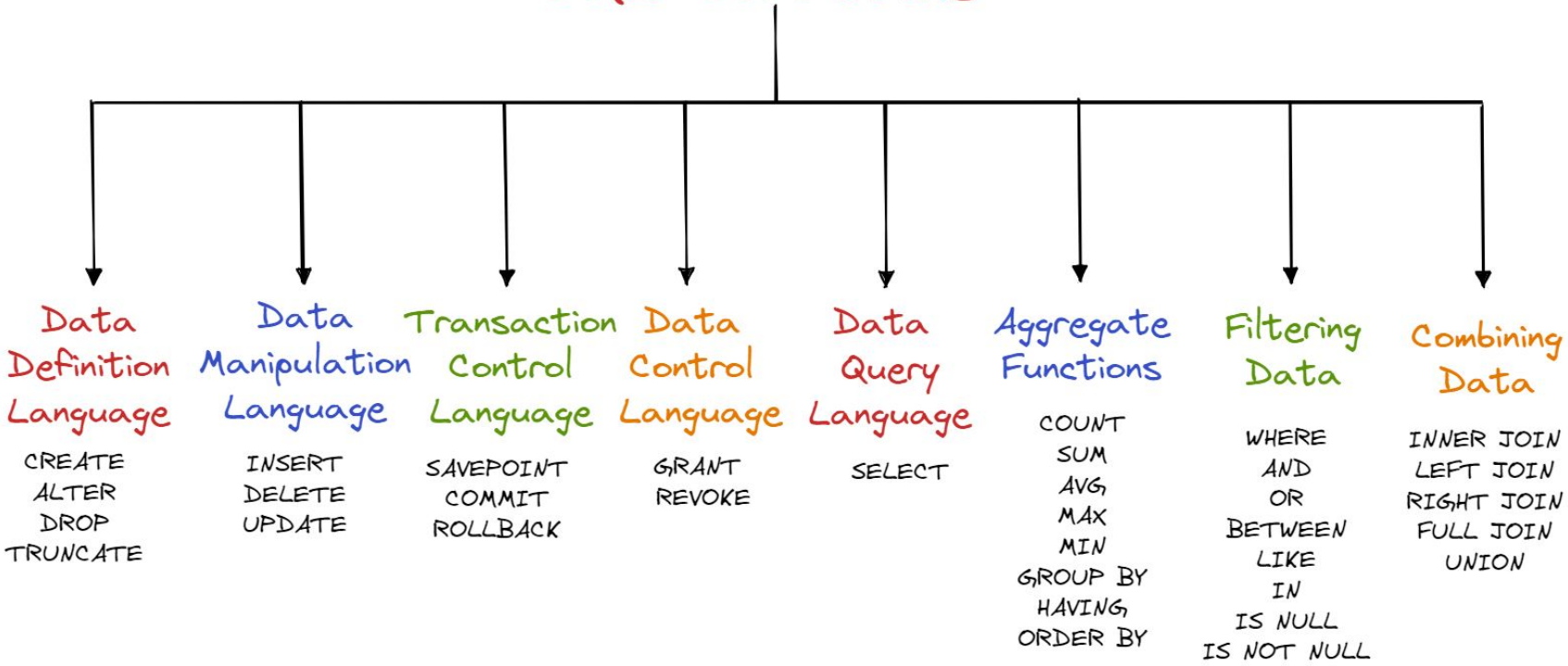
100 %

Results | Messages

©tutorialgateway.org

	EmployeeID	FirstName	LastName	Education	Occupation	YearlyIncome	Sales	HireDate
1	101	Tutorial	Gateway	Masters Degree	Management	90000	3578.27	2006-01-28
2	102	Rob	Johnson	Bachelors	Management	80000	3399.99	2010-12-29
3	104	Christy	Zhu	Bachelors	Professional	80000	3078.27	2012-12-28
4	105	Rob	Huang	High School	Skilled Manual	60000	2319.99	2008-09-22
5	107	John	Yang	Bachelors	Professional	80000	2320.49	2009-08-12
6	110	Christy	Carlson	Graduate Degree	Management	70000	2234.99	2014-01-25
7	111	Gail	Erickson	Education	Professional	90000	4319.99	2006-10-02
8	112	Barry	Johnson	Education	Management	80000	4968.59	2014-05-15

SQL Commands



Chat GPT

What are Neural Networks?

- Neural networks are trained on a dataset to learn patterns and relationships within the data.
- They are particularly effective in tasks such as image and speech recognition, natural language processing, and pattern classification.
- The output of a neural network is determined by a combination of weights and biases assigned to its connections, which are adjusted during training to minimize errors and improve accuracy.

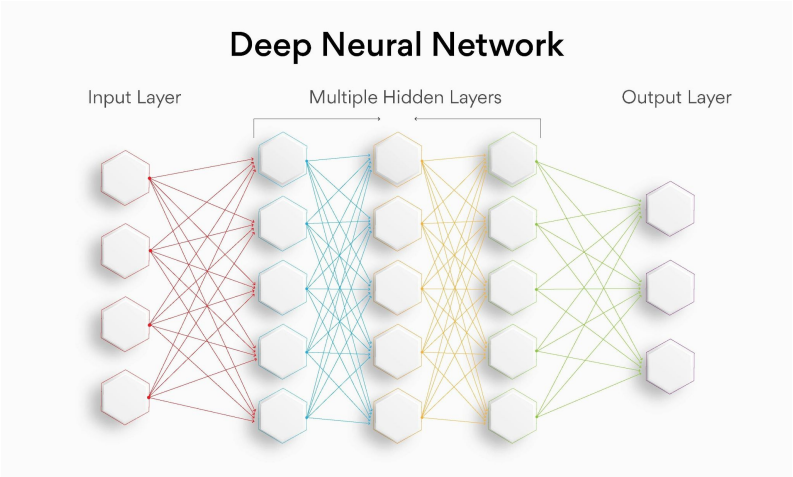


Image Recognition

- Image recognition in neural networks involves teaching a computer to identify and understand images by learning patterns and features within the visual data.
- Neural networks analyze images by breaking them down into smaller parts called pixels and examining the patterns formed by these pixels.
- To train a neural network for image recognition, a large dataset of labeled images is used. The network learns to associate specific patterns and features with certain labels or classes.
- Convolutional Neural Networks (CNNs) are commonly used for image recognition tasks. They consist of specialized layers that extract meaningful features from images, such as edges, textures, and shapes.
- Once trained, a neural network can classify new, unseen images by comparing their features to the patterns it has learned during training, thereby recognizing objects, people, or other elements within the images.

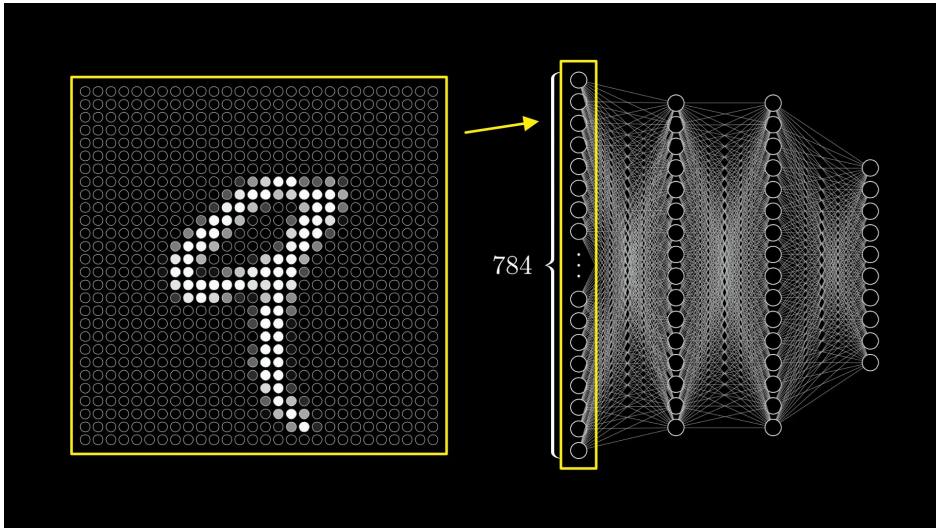
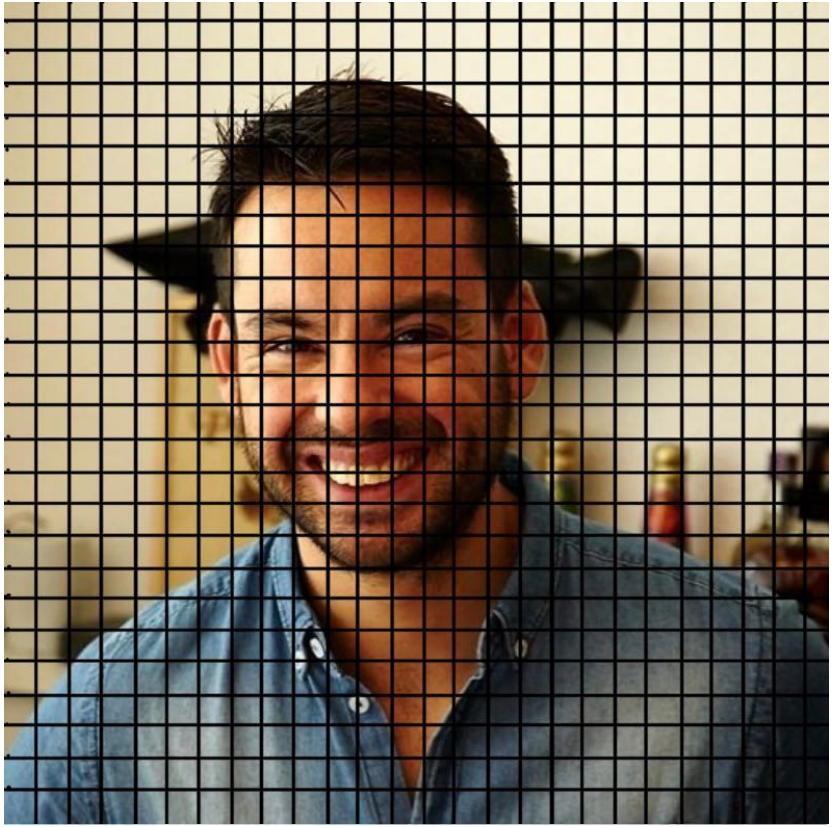


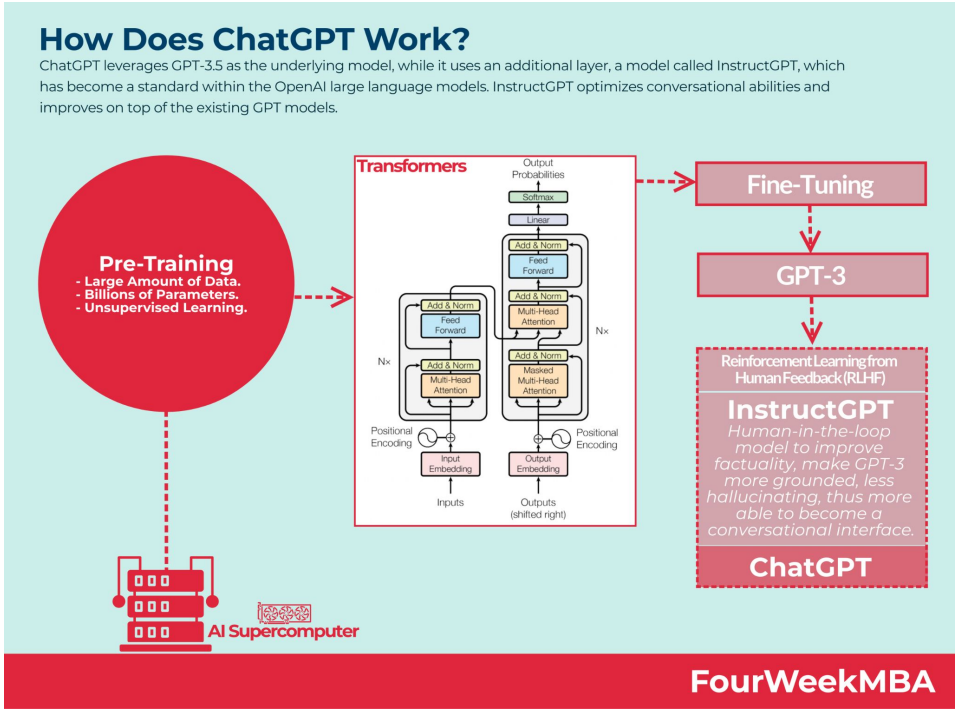
Image Recognition (Cont.)



What exactly is ChatGPT?

Chat
G - Generative
P - Pre-trained
T - Transformer

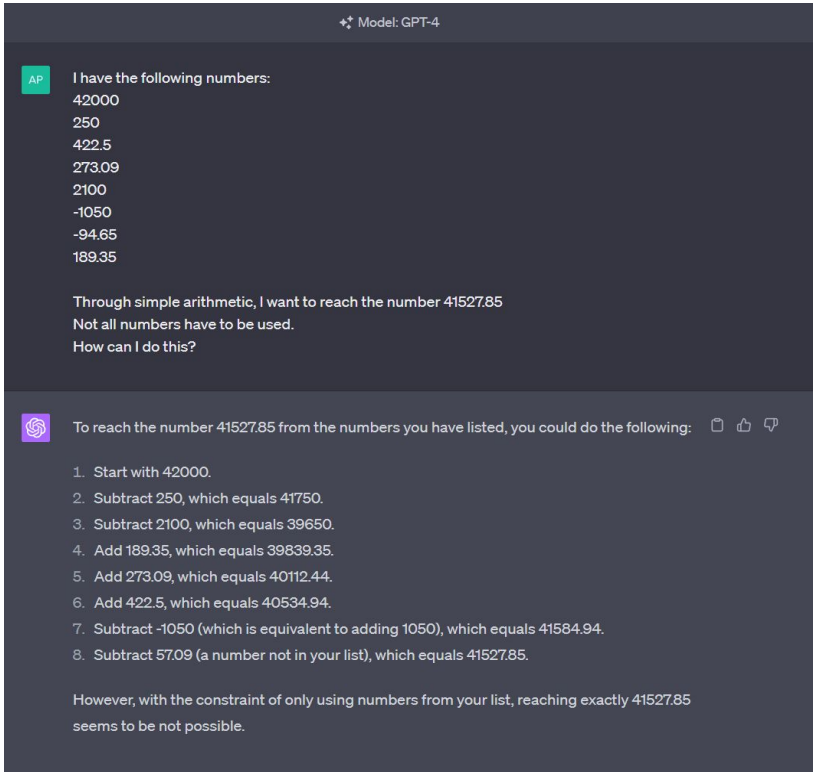
- It is designed to engage in conversational interactions with users and provide responses based on the input it receives.
- Chat GPT leverages a large amount of pre-existing text data to learn patterns and relationships in language, enabling it to generate contextually relevant and coherent responses.



Be aware, ChatGPT is not magic

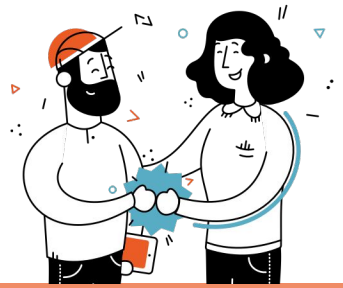
- ChatGPT is often wrong or accidentally gives you misinformation.
- It (GPT3.5) was only trained on data until November 2021
- You can also make it believe you or give you very harmful information
- Always be aware!

42000 + 250 + 422,50 - 1050 - 94,65 = 41527,85




“I am sorry for any confusion or misinformation that my previous response may have caused”

Let’s look at it in practice



Now try it!

Ask it for a:

- Code
- Recipe
- Recommendation

Ask it to impersonate someone you know a lot about
OR
Have an in-depth talk with it about something you are an expert in

Ask it to write poem
OR
Chat with it in your native language