Cognitive Computing
UCS420

Lecture 5 and 6

Recap

- **Which of the following is NOT a key component of a cognitive system?**
  a) Perception
  b) Reasoning
  c) Sensory input
  d) Predefined output

Answer: d) Predefined output

Recap

- **In which type of system can two identical inputs produce different outputs?**
  a) Deterministic
  b) Probabilistic
  c) Linear
  d) Static

Answer: b) Probabilistic

Recap

- **What is the primary goal of regression analysis?**
  a) To minimize the error between observed and predicted values.
  b) To predict future outcomes based on independent variables.
  c) To establish a correlation between variables.
  d) Both a and b.

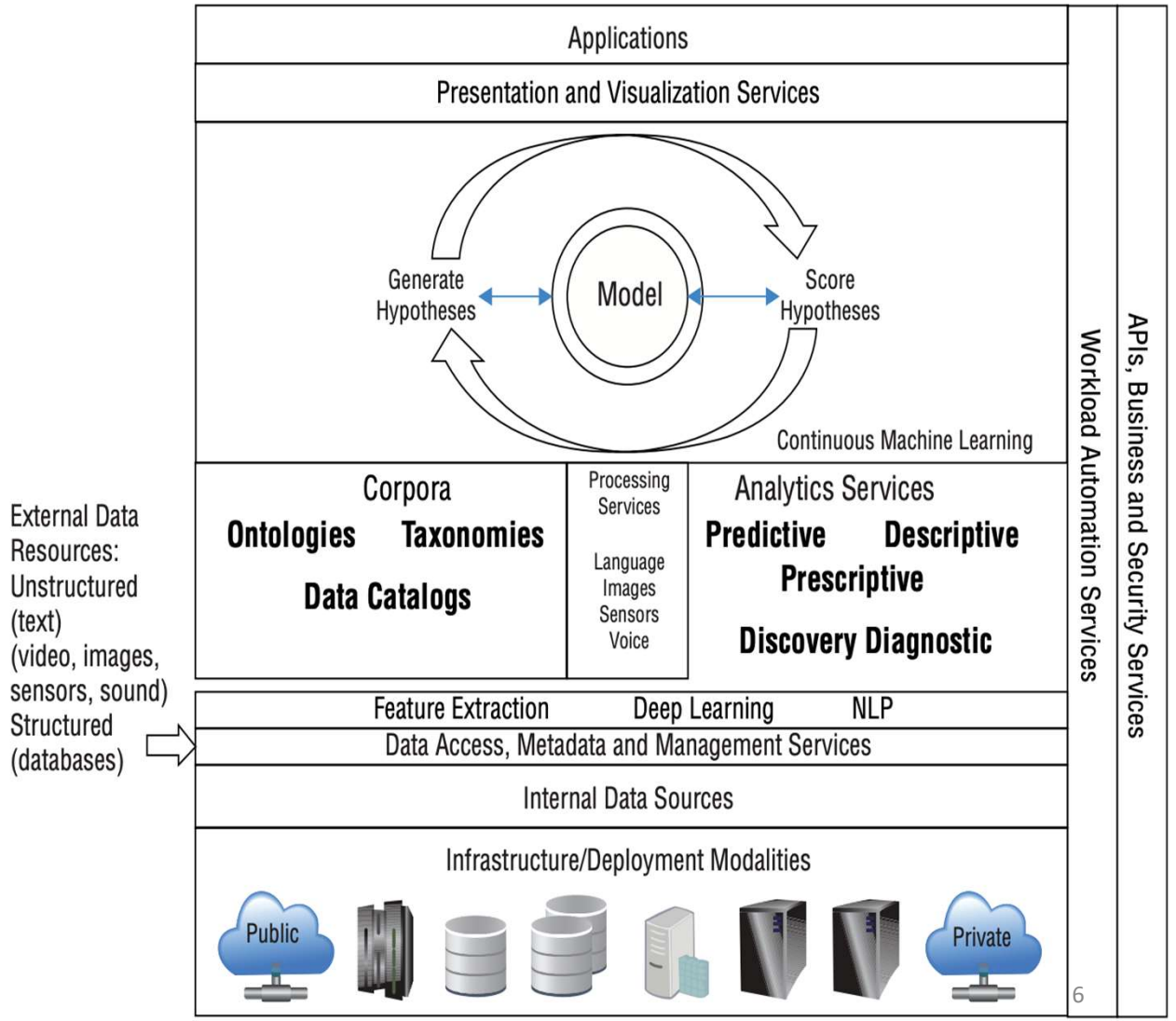
Answer: d) Both a and b

Recap

- **Which of the following is an independent variable in a study that examines the effect of advertising budget on sales revenue?**
  a) Sales revenue
  b) Advertising budget
  c) Customer satisfaction
  d) Product price

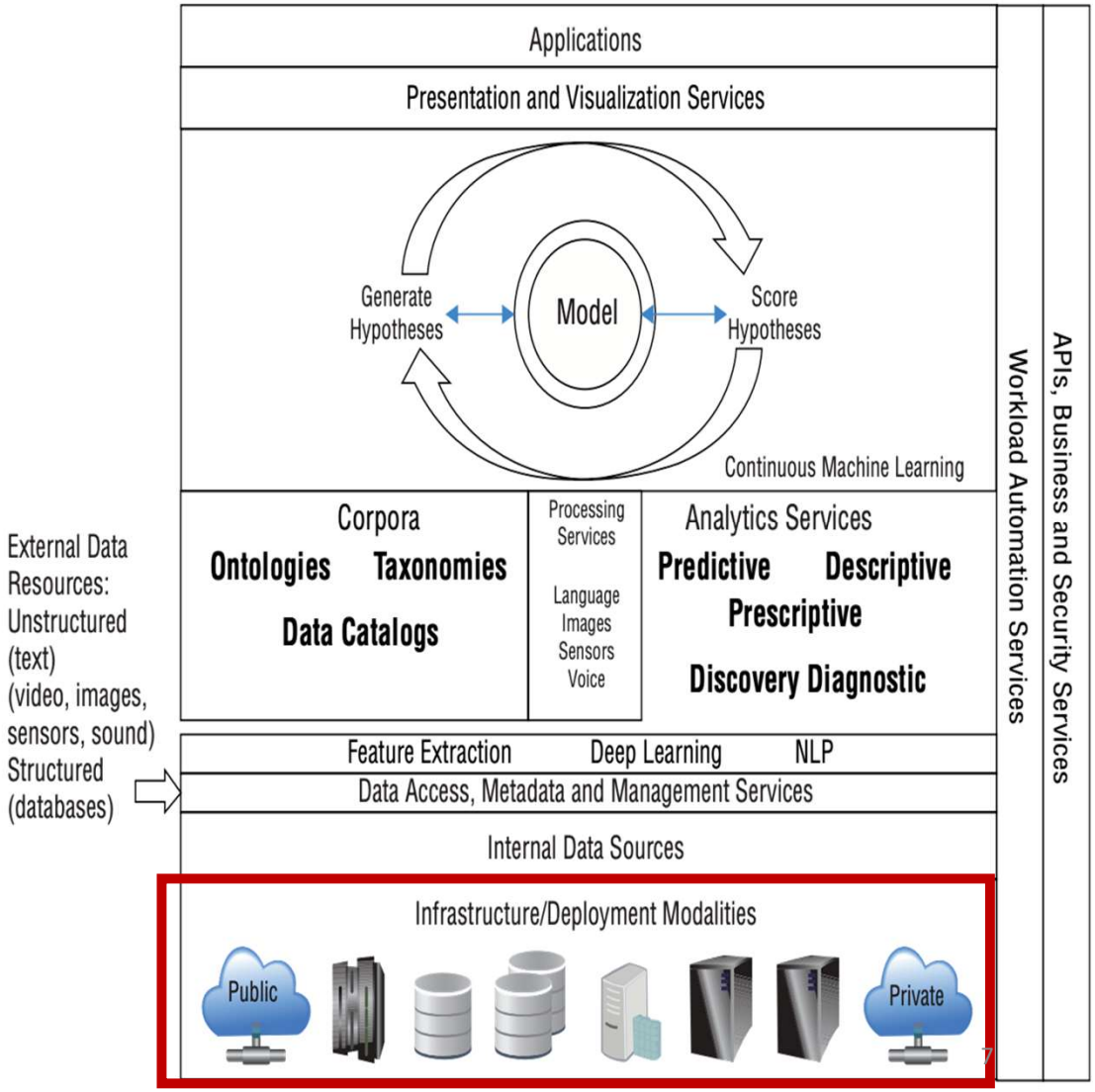
Answer: b)Advertising budget

Elements of Cognitive System

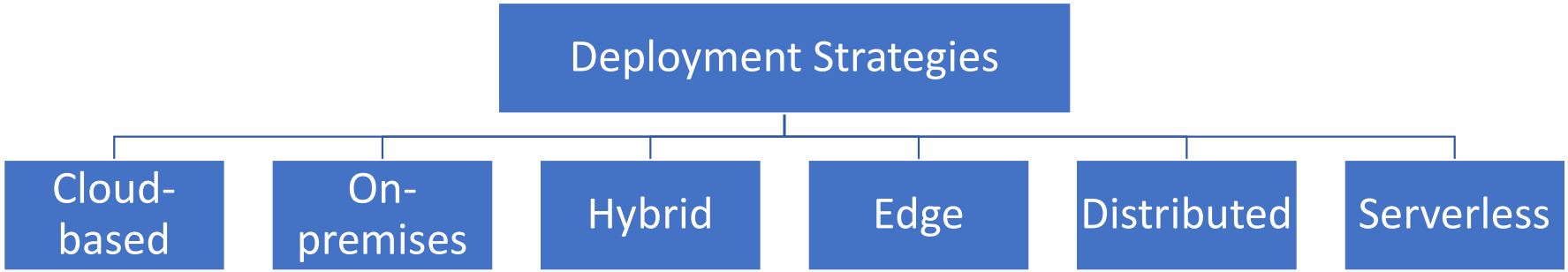


Infrastructure/Deployment Modalities

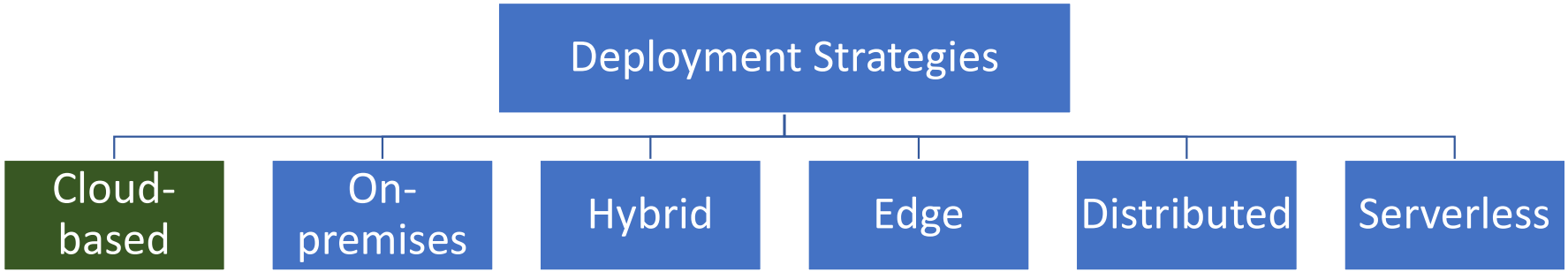
- To deploy cognitive computing systems effectively, appropriate infrastructure and deployment modalities are essential.
- These elements define the technical setup, environment, and strategies for implementing cognitive systems.
- Flexible and Agile infrastructure.
- Management of public and private data.
- Highly parallelized and distributed environment, including computational and storage cloud services.



Infrastructure/Deployment Modalities



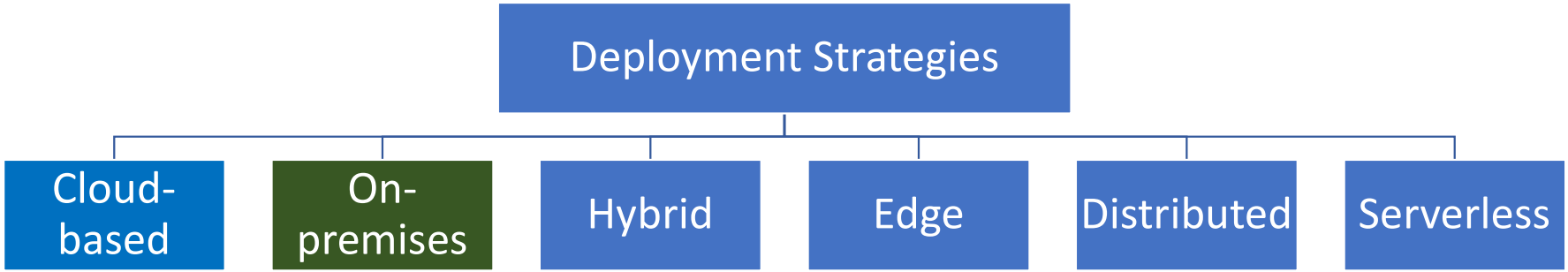
Infrastructure/Deployment Modalities



1. Cloud-Based Deployment

- Cognitive computing systems are deployed on cloud platforms, allowing for scalability, flexibility, and easy access.
- Leverage Software as a Service (SaaS) applications and services for cloud deployment.
- **Examples:** IBM Watson, Google Cloud AI, Amazon SageMaker.

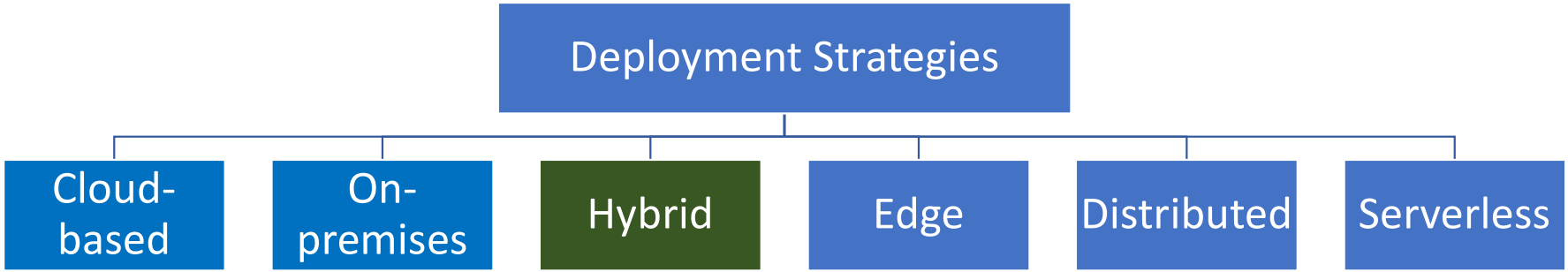
Infrastructure/Deployment Modalities



2. On-Premises Deployment

- Systems are hosted within an organization's private data centers for greater control and security.
- **Example Use Case:** Banks deploying cognitive fraud detection tools in-house.

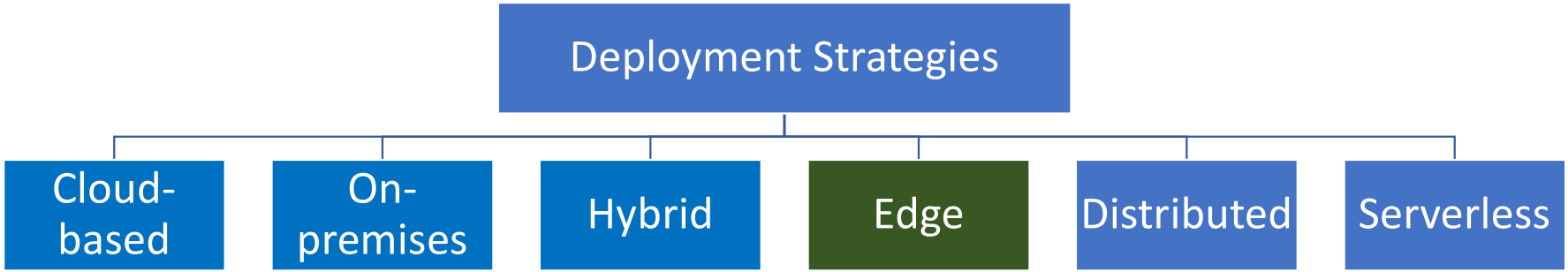
Infrastructure/Deployment Modalities



3. Hybrid Deployment

- Combines cloud-based and on-premises models, enabling data processing and storage across both environments.
- **Example Use Case:** A retail company using on-premises servers for transaction data and cloud for customer analytics.

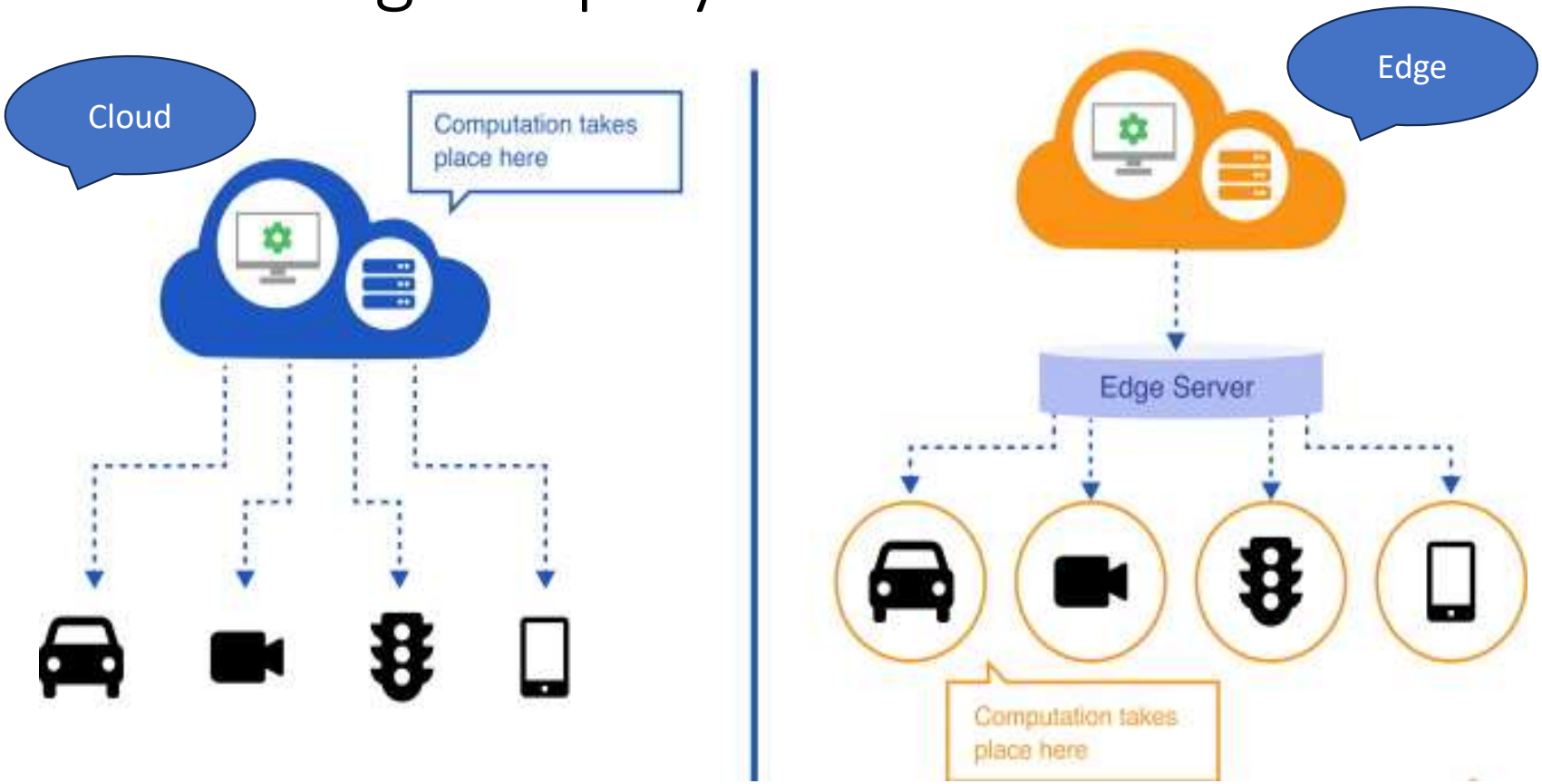
Infrastructure/Deployment Modalities



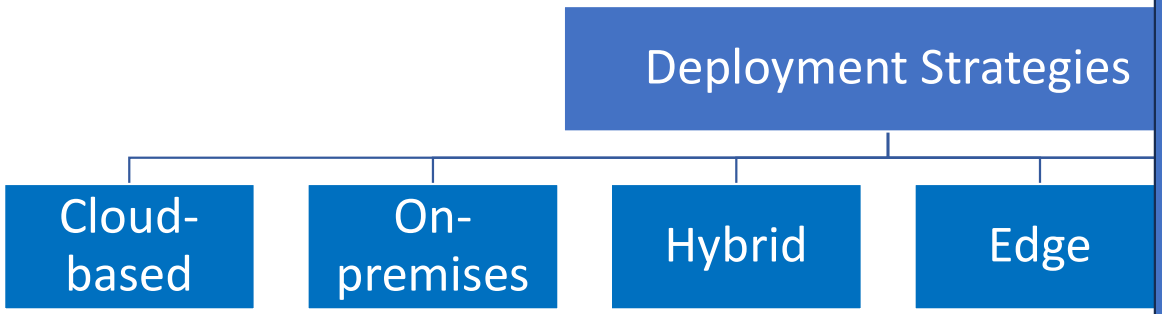
4. Edge Computing Deployment

- Cognitive systems are deployed closer to data sources (e.g., IoT devices) for real-time processing.
- **Example Use Case:** Smart home devices using NLP at the edge for voice recognition.

Cloud vs Edge Deployment



Infrastructure/Deployment Modalities



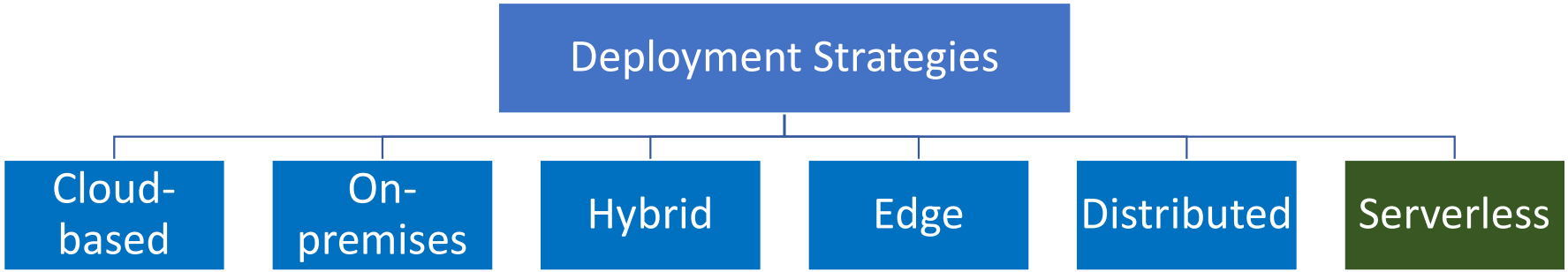
What is Apache Spark?

- **Apache Spark** is an open-source **distributed computing** system that provides a unified analytics engine for big data processing. It is widely used for data processing, real-time analytics, and machine learning on large dataset.
- Spark distributes data and computation across multiple nodes, allowing it to handle massive datasets that cannot fit into a single machine's memory.

5. Distributed Systems:

- Leverages multiple interconnected systems to distribute cognitive computing workloads.
- **Example Use Case:** Distributed AI frameworks like **Apache Spark** for large-scale machine learning tasks.

Infrastructure/Deployment Modalities



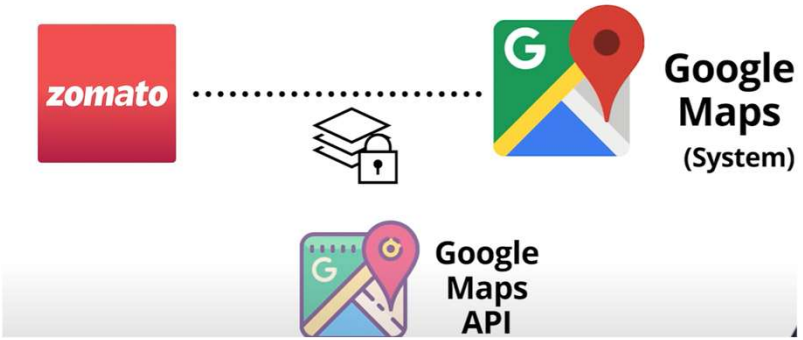
6. Serverless Architecture

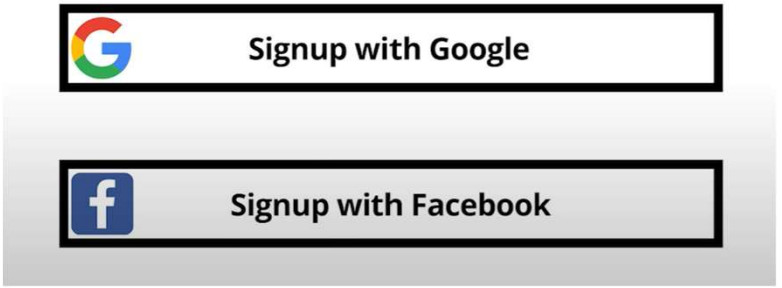
- Applications are deployed in a serverless computing environment where the infrastructure is abstracted, and the system scales automatically based on workload.
- **Example Use Case**: Cognitive chatbots powered by serverless **backend APIs**.

Infrastructure/Deployment Modalities

What is API ?

- API stands for **Application Programming Interface**.
- It's a set of rules and protocols that allow software applications to communicate with each other.
- APIs are a key part of modern technology and business infrastructure.
- APIs Uses key for security.
- Example-
  - Zomato uses the **Google Maps API** to track delivery person.
  - GitHub uses "**Signup with Google API**", to validate users.

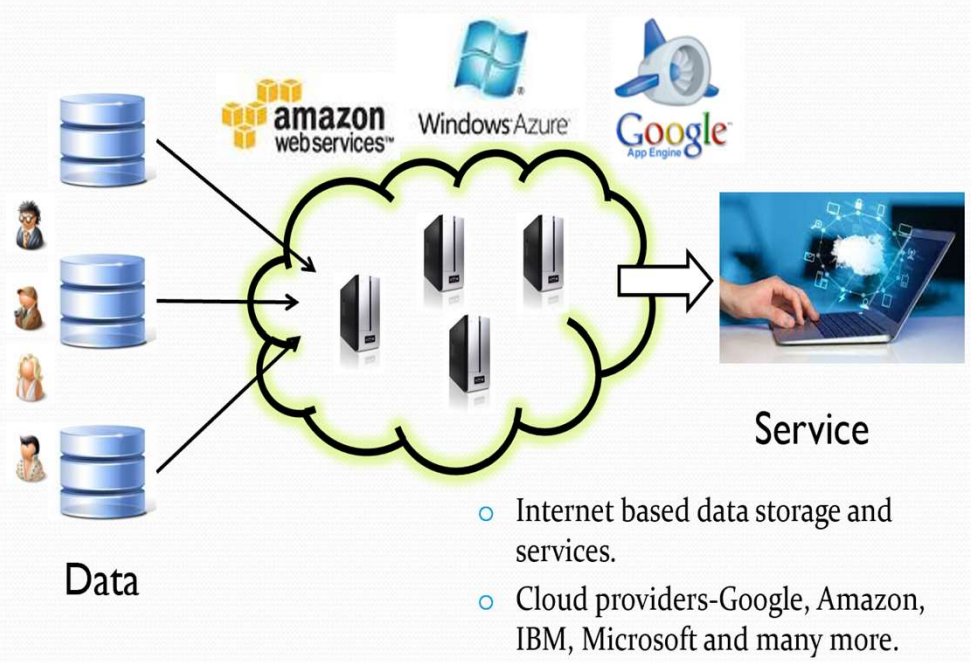




Infrastructure/Deployment Modalities

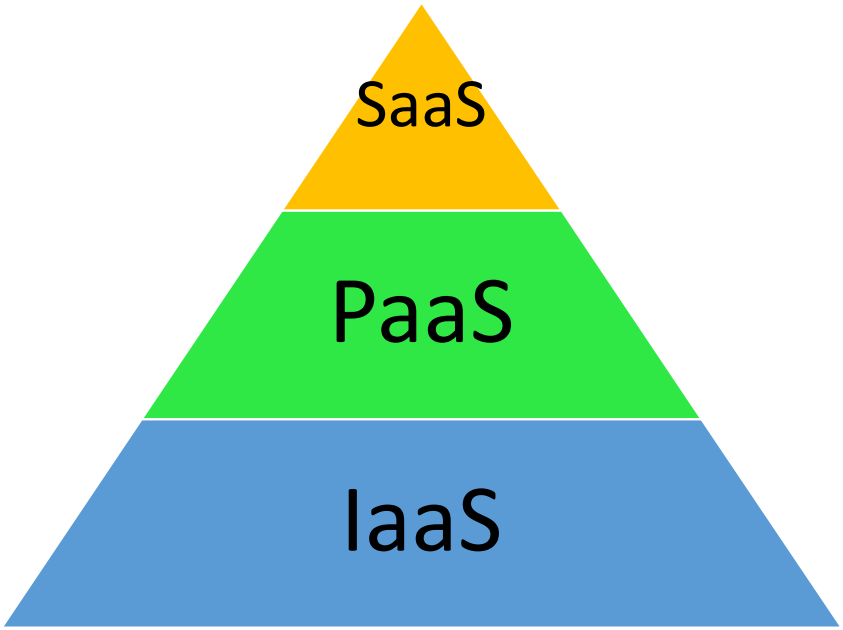
What is Cloud Computing?

- **Cloud computing** is an emerging computing technology that uses the internet and central remote servers to maintain data and applications.
- It refers to manipulating, configuring and accessing the applications online.
- It offers online data storage, infrastructure and applications.
- It is both a combination of software and hardware based computing resources delivered as a network service.



Infrastructure/Deployment Modalities

- **Service Model in Cloud Computing:**
  1. Software as a Service (SaaS)
  2. Platform as a Service (PaaS)
  3. Infrastructure as a Service (IaaS)



Infrastructure/Deployment Modalities

Software as a Service (SaaS)

- Sometimes referred to as "on-demand software".
- Application is used as an on-demand service. Often provided via the Internet.
- No hardware or software to manage.
- Example: Google Drive App (**online Office, Google Docs**), Dropbox

Infrastructure/Deployment Modalities

Platform as a Service (PaaS)

- It is a delivery of a computing platform over the web.
- Provides runtime environment for applications, development and deployment tools.
- It create web applications quickly without the cost and complexity of buying and managing the underlying software/hardware.
- Example:
  - Google Cloud Platform (GCP): App Engine
  - **Google Colab**

Infrastructure/Deployment Modalities

Infrastructure as a Service (IaaS)

- Application will be executed on a **virtual computer**.
- The basic computing infrastructure of servers, software, and network equipment is provided as an on-demand service upon which a platform to develop and execute applications can be founded.
- It can be used to avoid buying, housing, and managing the basic hardware and software infrastructure components.
- Example:
  - **Amazon Web Services (AWS): Elastic Compute Cloud (EC2)**
  - Oracle Cloud Infrastructure (OCI): Compute, Object Storage
  - IBM Cloud: Virtual Servers

Infrastructure/Deployment Modalities

- **Infrastructure as a Service (IaaS)**
- **Example-** Amazon Web Services (AWS) allow users to create and manage virtual machines (VMs) with the operating system of their choice.

1. Virtual Computer:

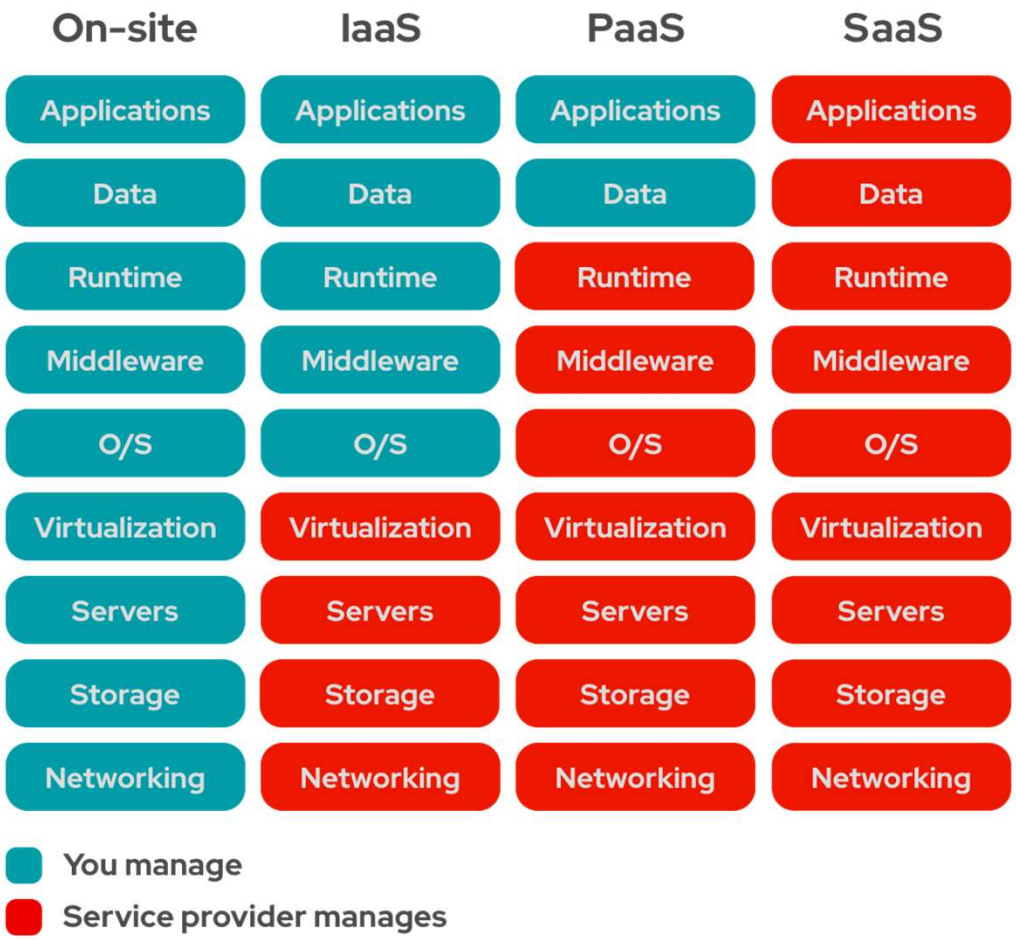
1. Suppose, a user needs a Linux environment to develop and execute a program, but their laptop runs Windows.
2. AWS provides **virtual computers** (via services like Amazon EC2) that can run different operating systems, such as Linux or Windows, independent of the user's physical machine.

2. AWS Compute Services (Amazon EC2):

1. With **Amazon Elastic Compute Cloud (EC2)**, users can create virtual instances of Linux, Windows, or other OSs.
2. These instances operate as independent virtual machines, and users have full control over the configuration, installation of software, and execution of programs.

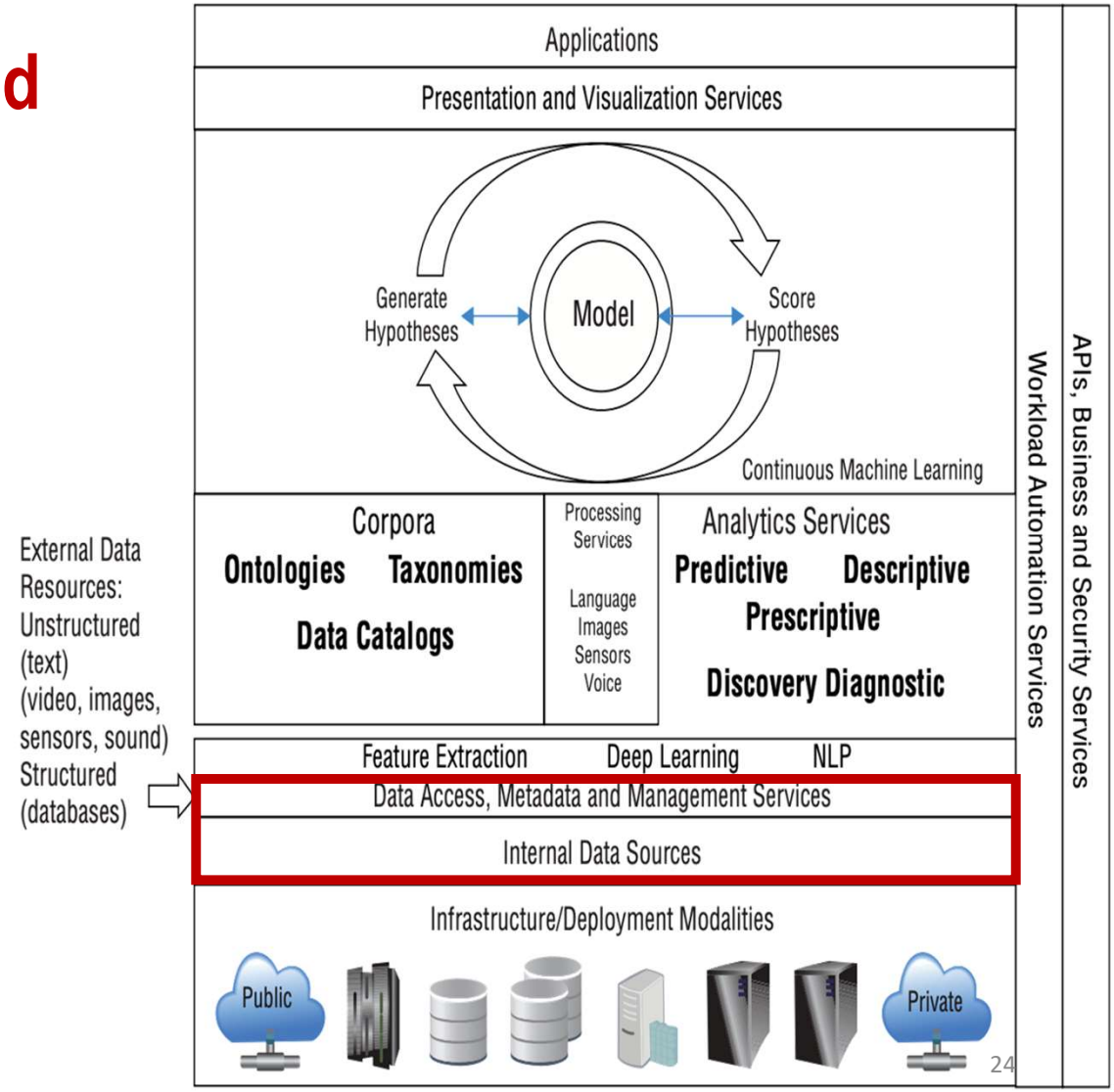
3. Why It's IaaS:

1. AWS provides the underlying infrastructure: virtual machines, storage, and networking.
2. Users are responsible for managing the OS, installing necessary software, and configuring the environment.
3. AWS abstracts the physical hardware, but the control of the virtual environment is left to the user.


On-site
IaaS
PaaS
SaaS
Applications
Data
Runtime
Middleware
O/S
Virtualization
Servers
Storage
Networking
You manage
Service provider manages

Data Access, Metadata, and Management Services

- **Data – Backbone of a cognitive system**
- In cognitive computing systems, **data access, metadata, and management services** form critical components that ensure efficient handling, organization, and retrieval of data.
- These services enable the system to process and analyze data effectively, ensuring meaningful insights while maintaining data integrity and security.

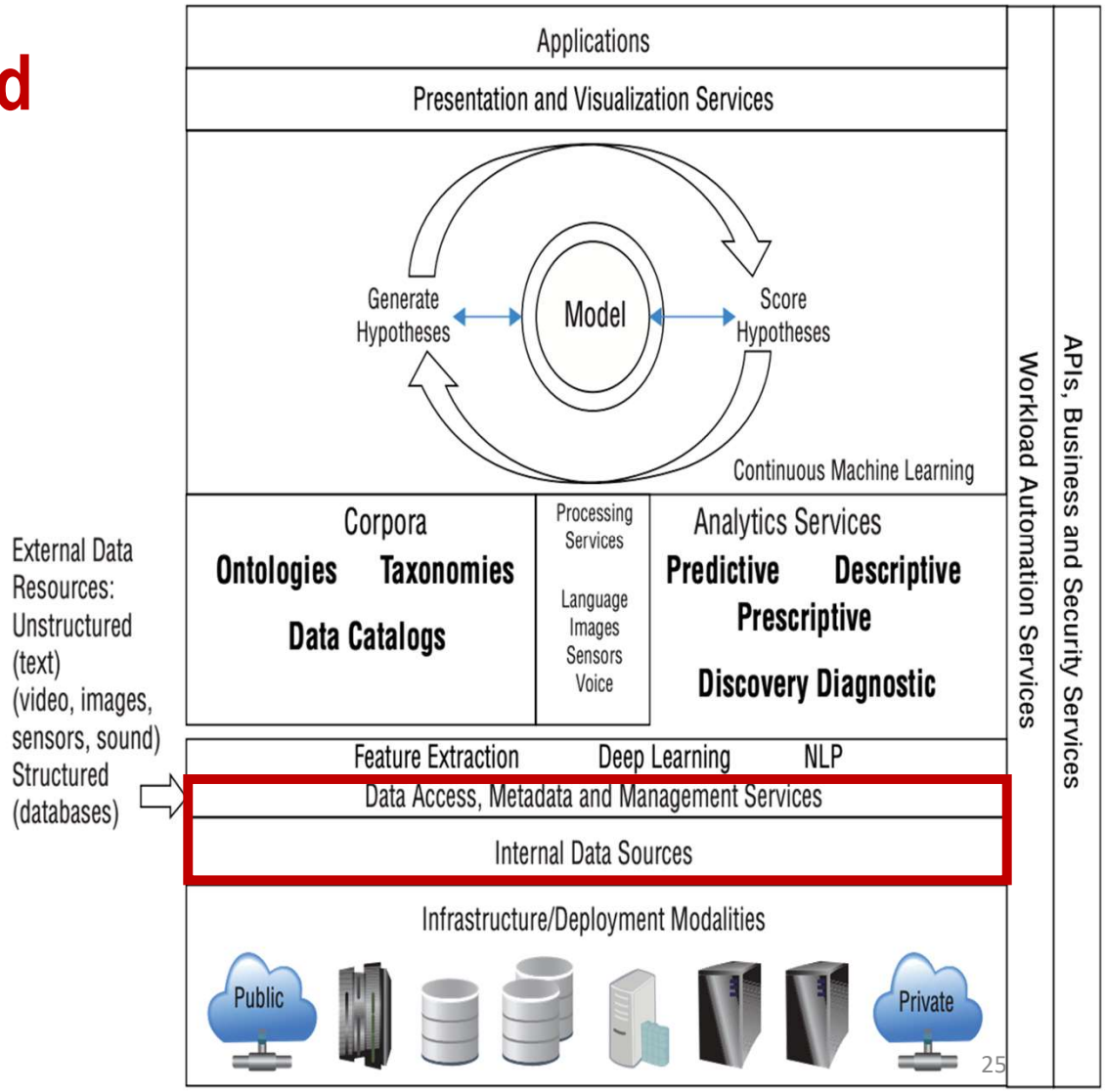


Data Access, Metadata, and Management Services

- **Data Access Services:** These services allow the cognitive system to access structured, semi-structured, and unstructured data from various sources.

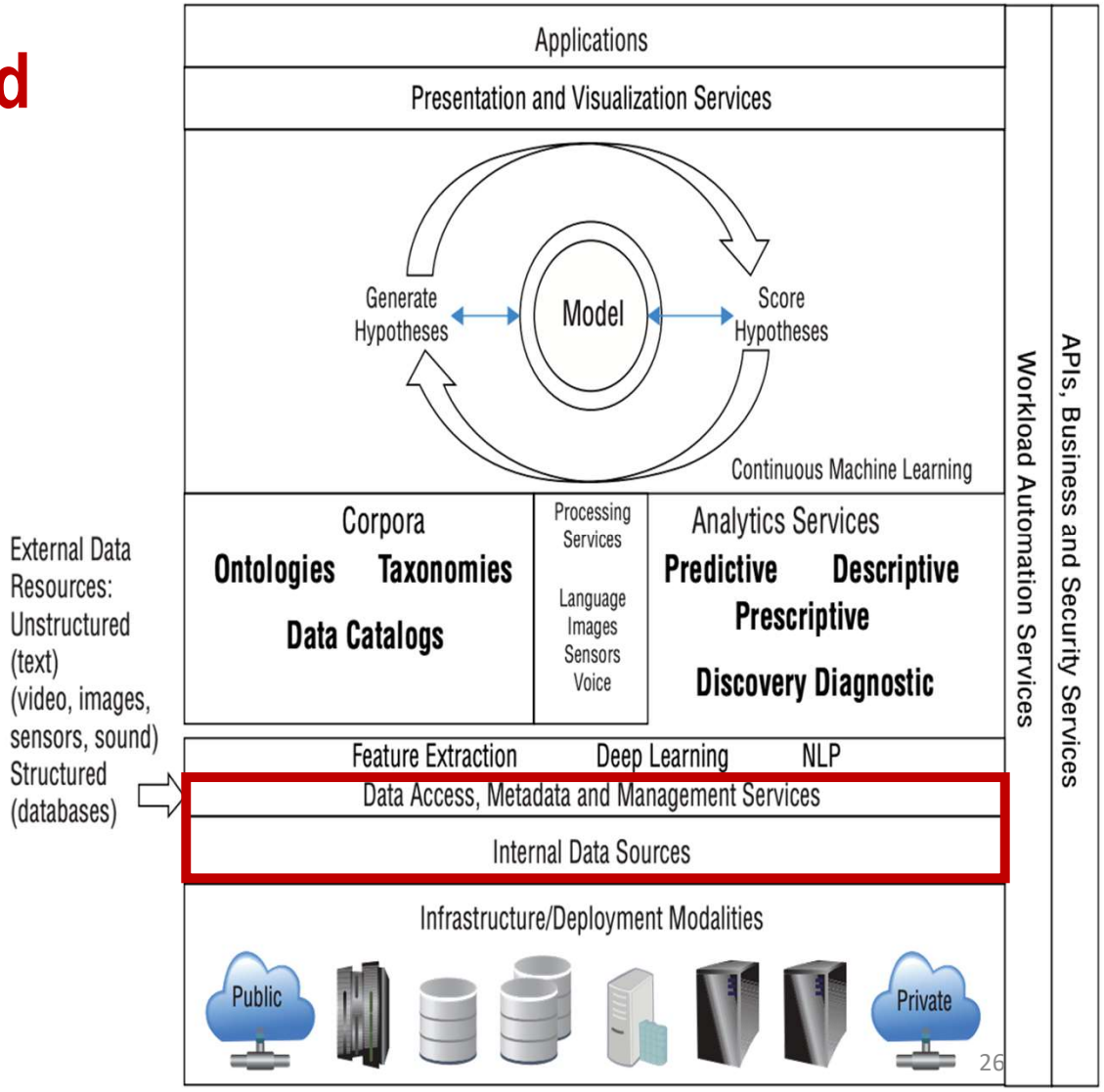
Examples:

- Extracting data from customer interaction logs for personalized recommendations.
- Accessing sensor data in IoT-based healthcare systems.



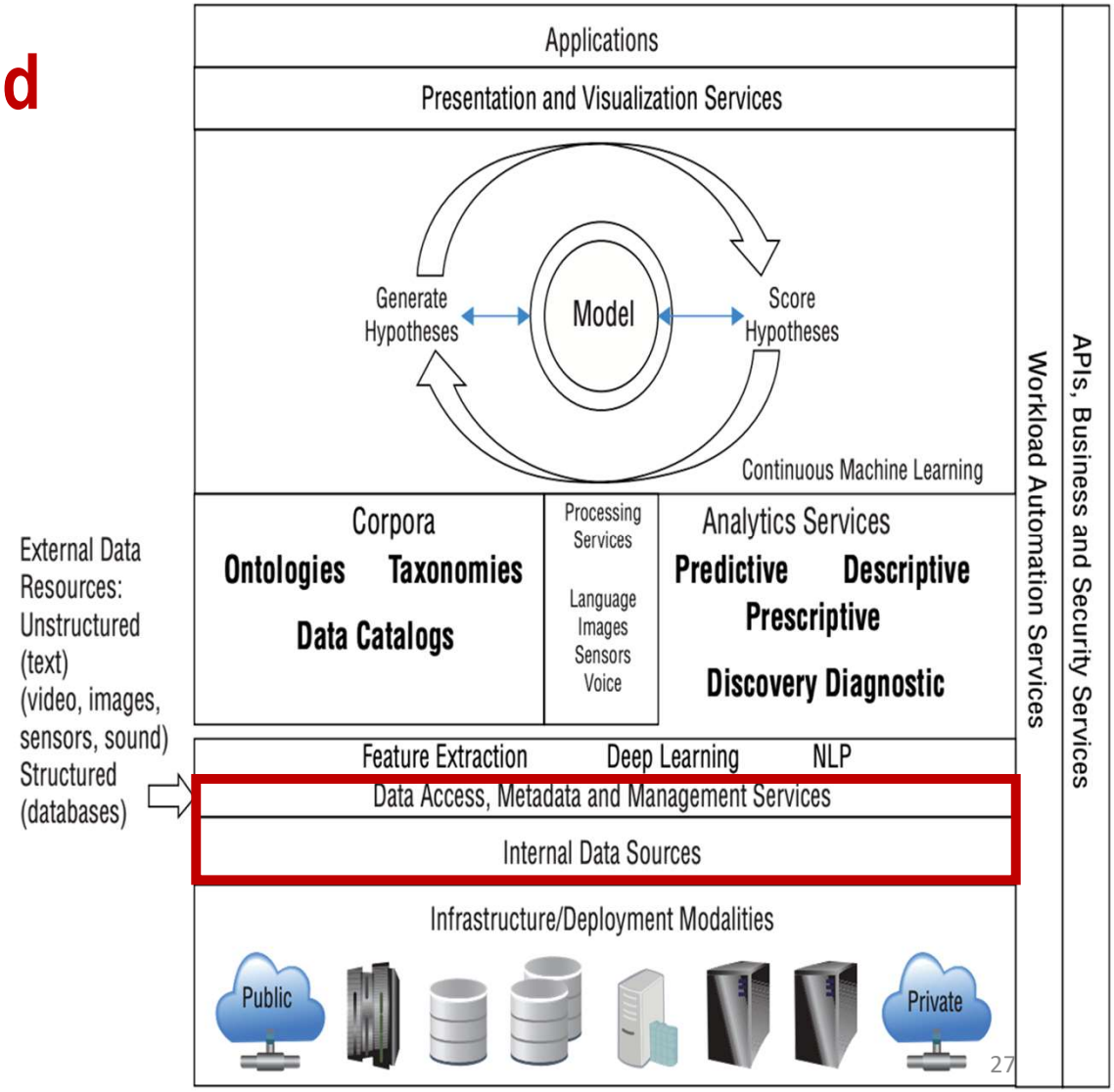
Data Access, Metadata, and Management Services

- **Metadata Services:** Metadata refers to data about data, describing its origin, structure, context, and usage. Metadata services manage and leverage this information to improve system performance and decision-making.
- **Examples:**
  - Annotating data in a healthcare system to identify patient records by disease type.



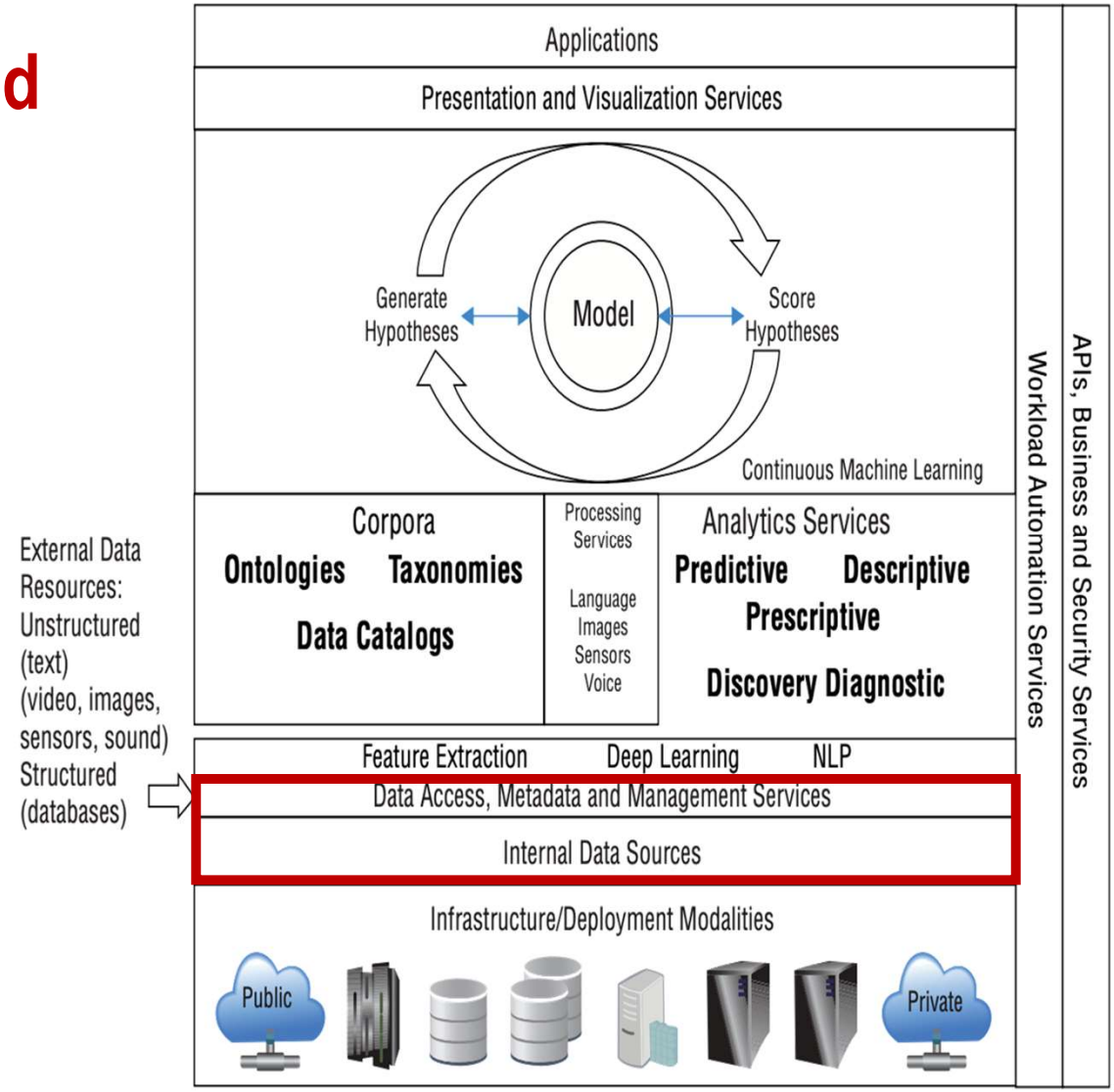
Data Access, Metadata, and Management Services

- **Data Management Services:** These services ensure the proper storage, organization, and maintenance of data throughout its lifecycle.
- **Examples:**
  - Managing a database of customer interactions in a retail cognitive system.
  - Ensuring secure storage of sensitive medical images in a cognitive healthcare application.



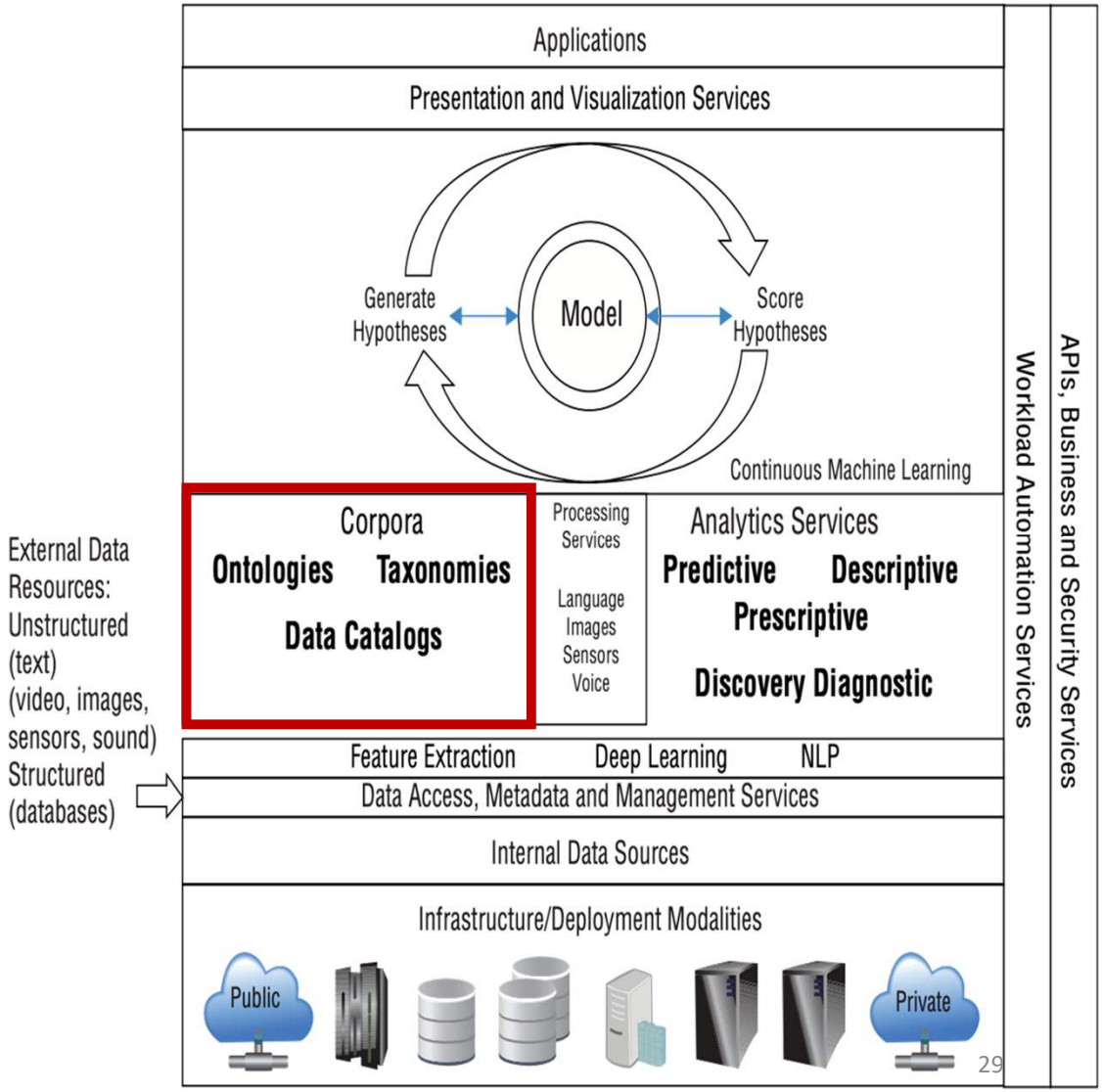
Data Access, Metadata, and Management Services

- Internal & External Data Sources
- Data Preparation
  - *Cleaning*
  - *Validating*
  - *Monitoring*



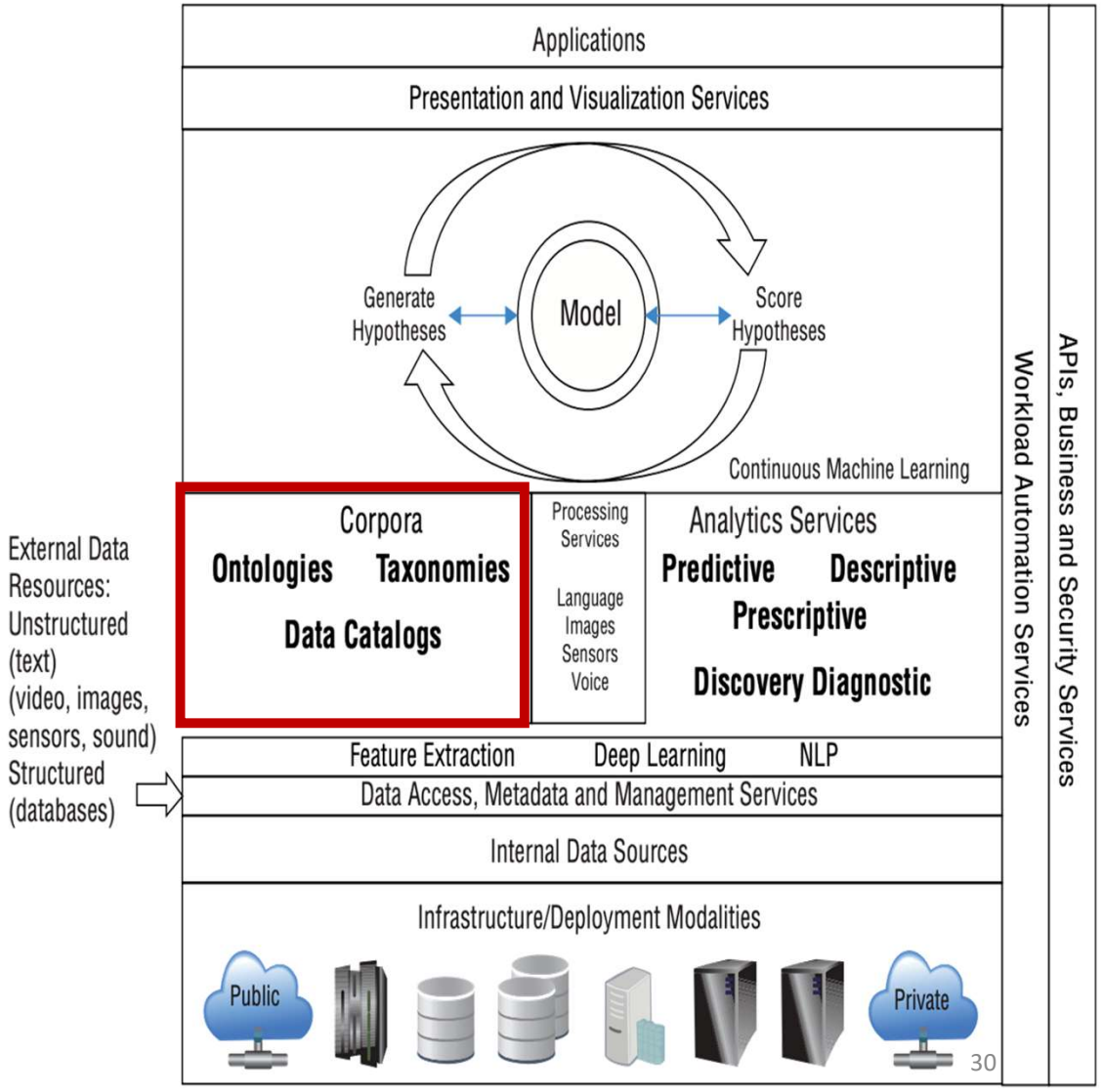
The Corpus, Taxonomies, and Data Catalogs

- In cognitive computing, **Corpus**, **Taxonomies**, and **Data Catalogs** are foundational components for managing, organizing, and making sense of vast amounts of data.
- They are essential for enabling cognitive systems to retrieve, analyze, and process information effectively.
- Each element has a unique role in enhancing the system's ability to simulate human cognitive processes.



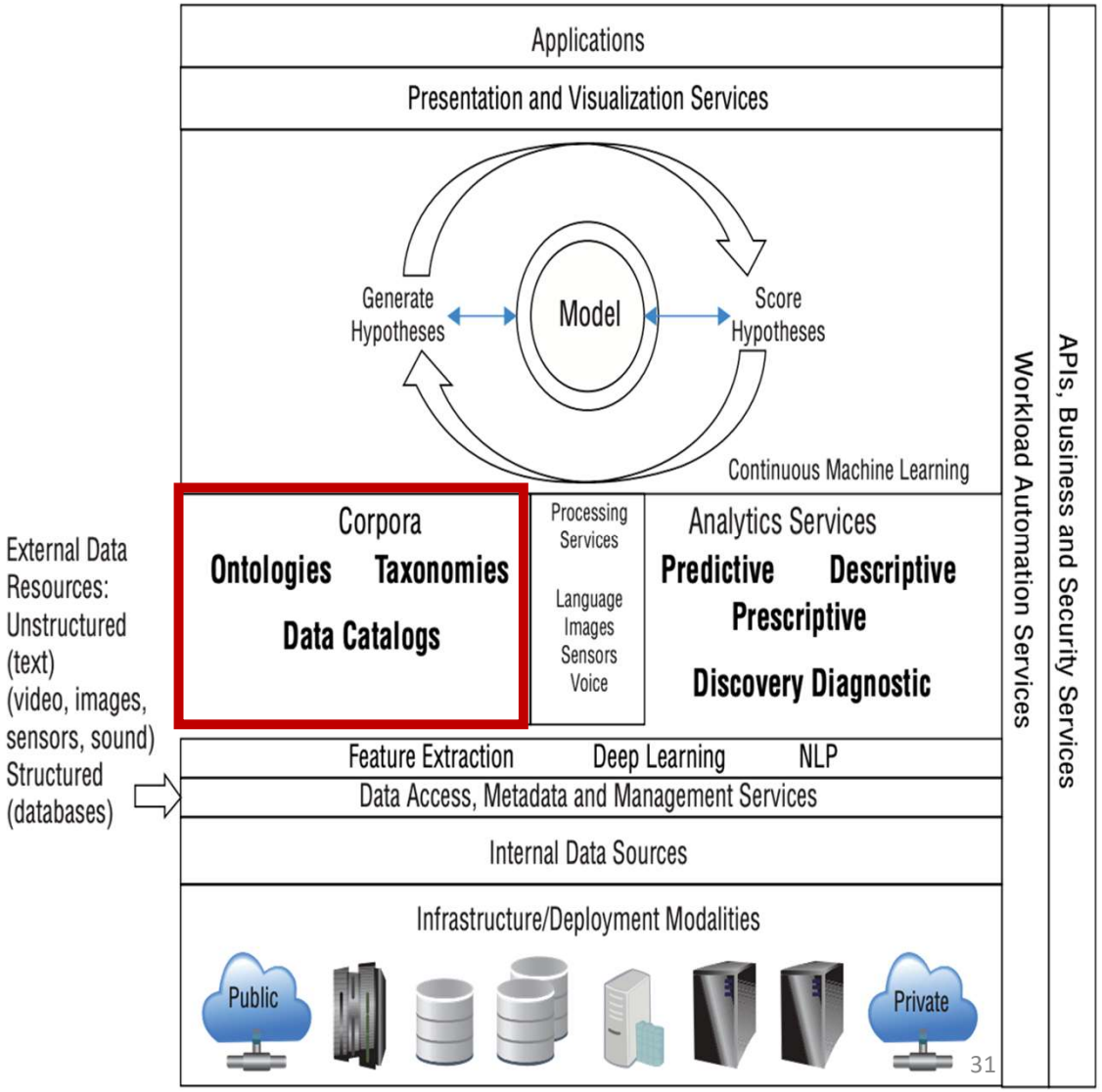
The Corpus, Taxonomies, and Data Catalogs

- **Corpus-** ***A corpus is a knowledge base of ingested data used for analysis and establishing the domain and contextual understanding for the system.***
- It forms the foundational knowledge base that the system relies on to generate insights.
- **Content Variety:** Can include structured, semi-structured and unstructured data.
- **Domain-Specific:** Often tailored to the specific domain of application (e.g., medical, legal, financial).
- **Dynamic Nature:** Can be continuously updated to include new information and remain relevant.



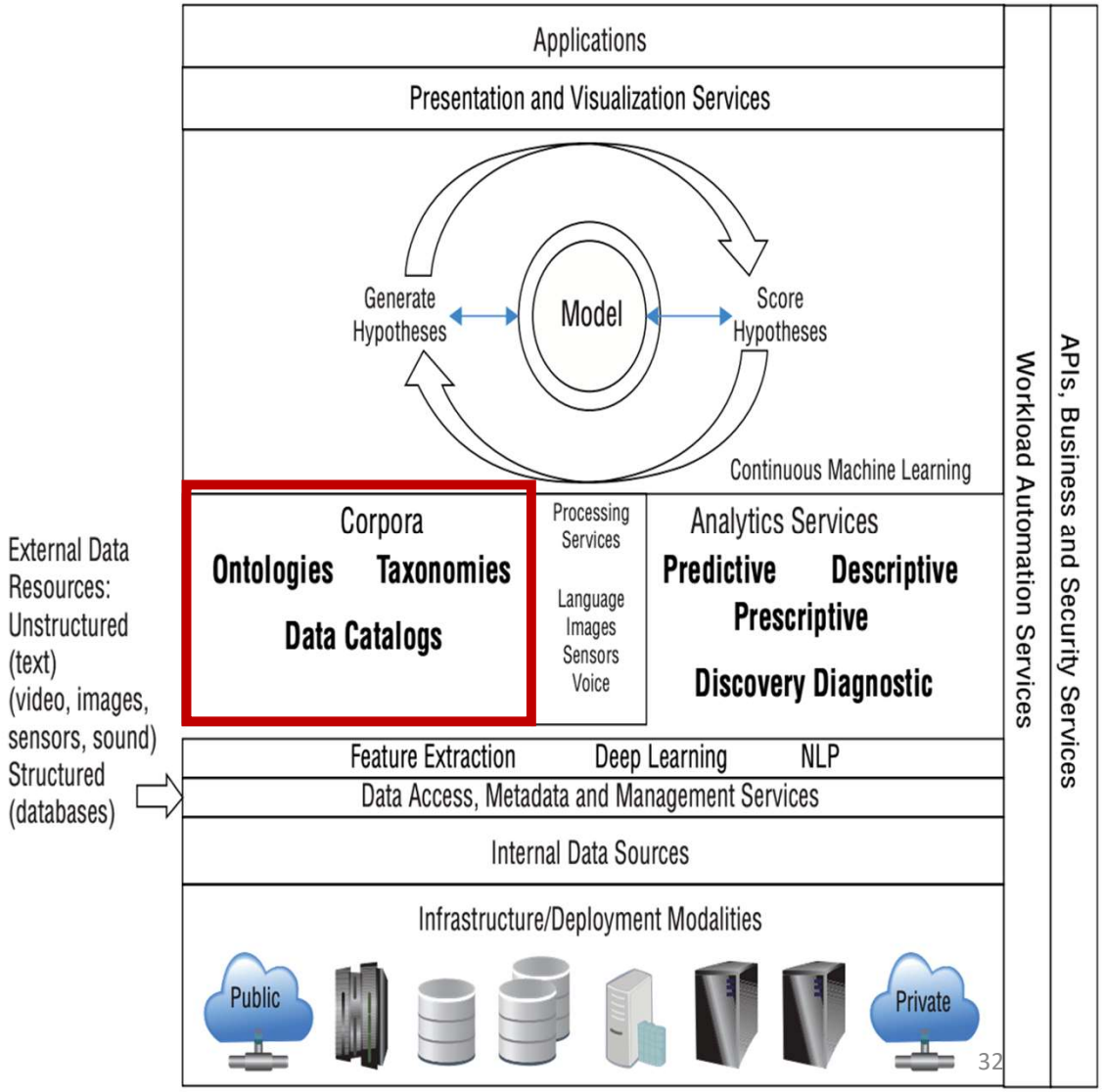
The Corpus, Taxonomies, and Data Catalogs

- ***Ontologies*** *- **Concepts in a domain that shows their properties and the relations between them.***
- *Use a subset of an industry-based ontology to include only the data that pertains to the focus of the cognitive system.*



The Corpus, Taxonomies, and Data Catalogs

- ***Taxonomies*** **- *Hierarchical classifications of entities.*** *OR*
- A hierarchical classification system used to organize information into categories, subcategories, and relationships.
- It provides a structured way to define and understand data relationships.
- *Provides context within ontology.*



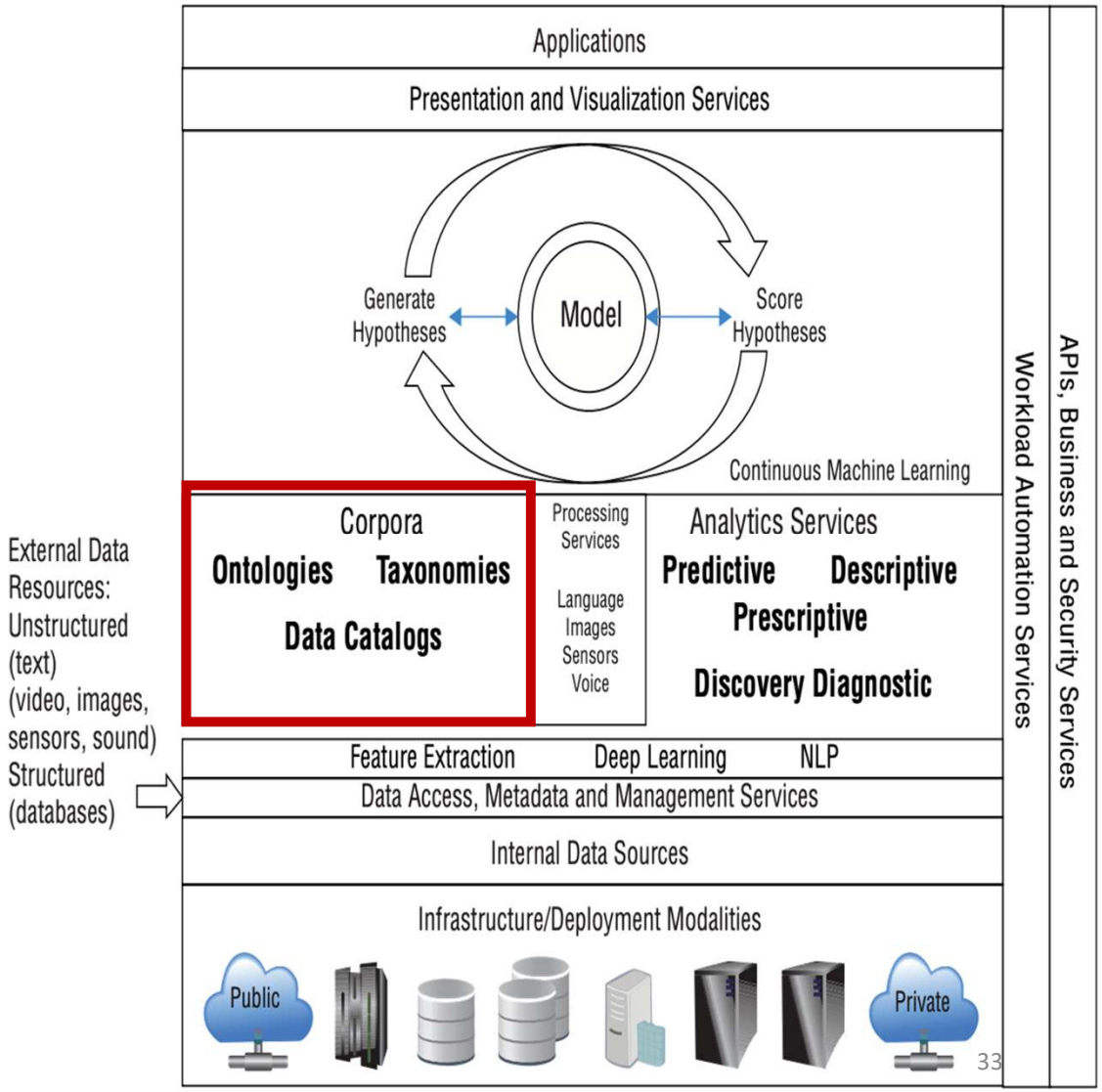
The Corpus, Taxonomies, and Data Catalogs

Data Catalogs-

- A data catalog is an organized inventory of data assets within an organization.
- It provides metadata about datasets, making it easier to discover, understand, and utilize data.

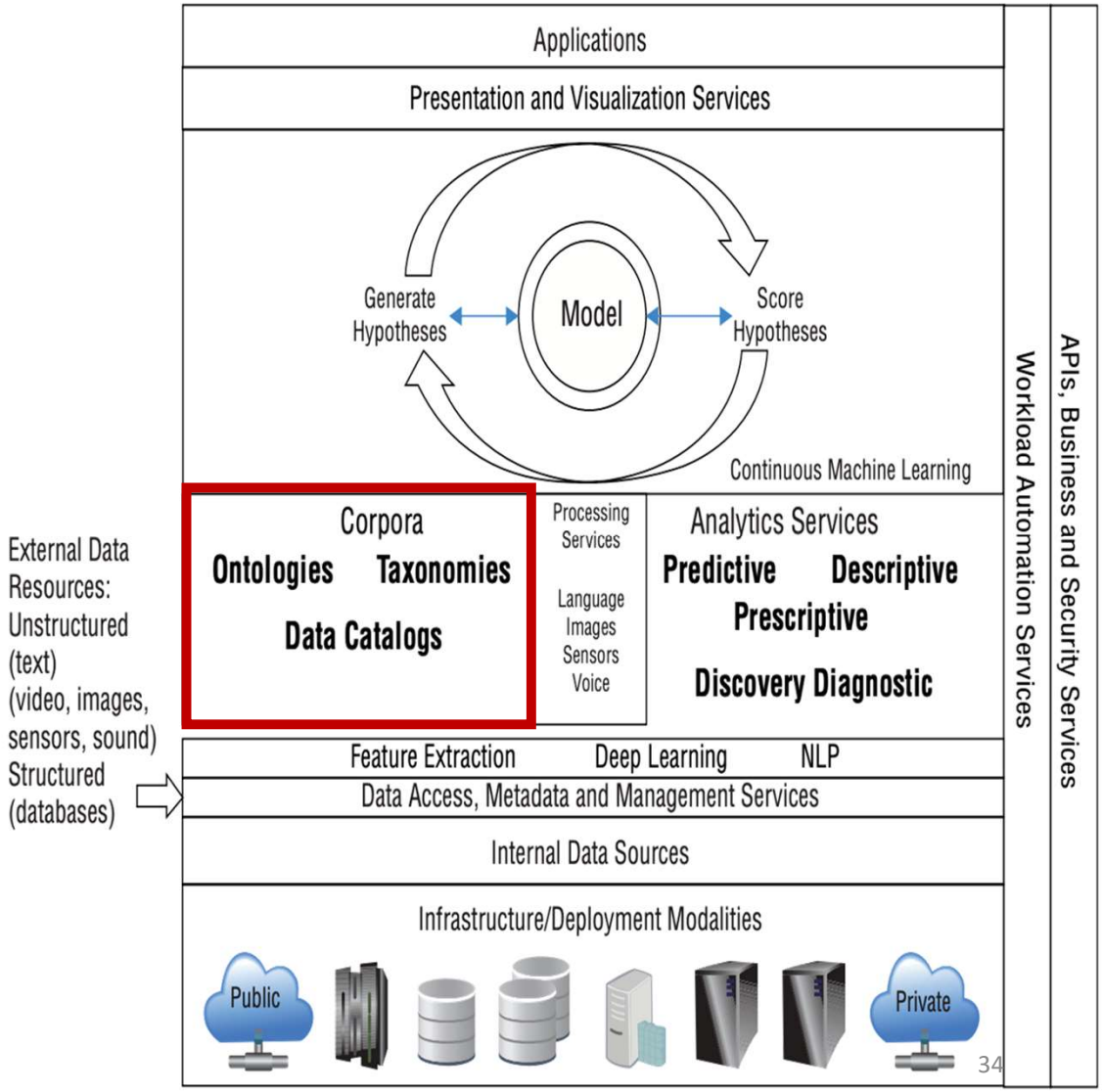
Examples:

- **Public Data Catalogs:** Government open data platforms or academic repositories.



The Corpus, Taxonomies, and Data Catalogs *(contd..)*

- **Eg: Bank - Financial vs Environment**
  - **Financial**
    - Financial Institutions > Commercial Banks > Retail Banking > State Bank of India
    - Provides "Loans," accepts "Deposits," and interacts with "Customers
  - **Environment**
    - Natural Features > River Banks > Erosion-Prone Areas > Amazon River Bank
    - "River Bank" supports "Flora and Fauna
- **Cell in Biology vs Telecommunications**
- **Node in Computer networks vs social networks**

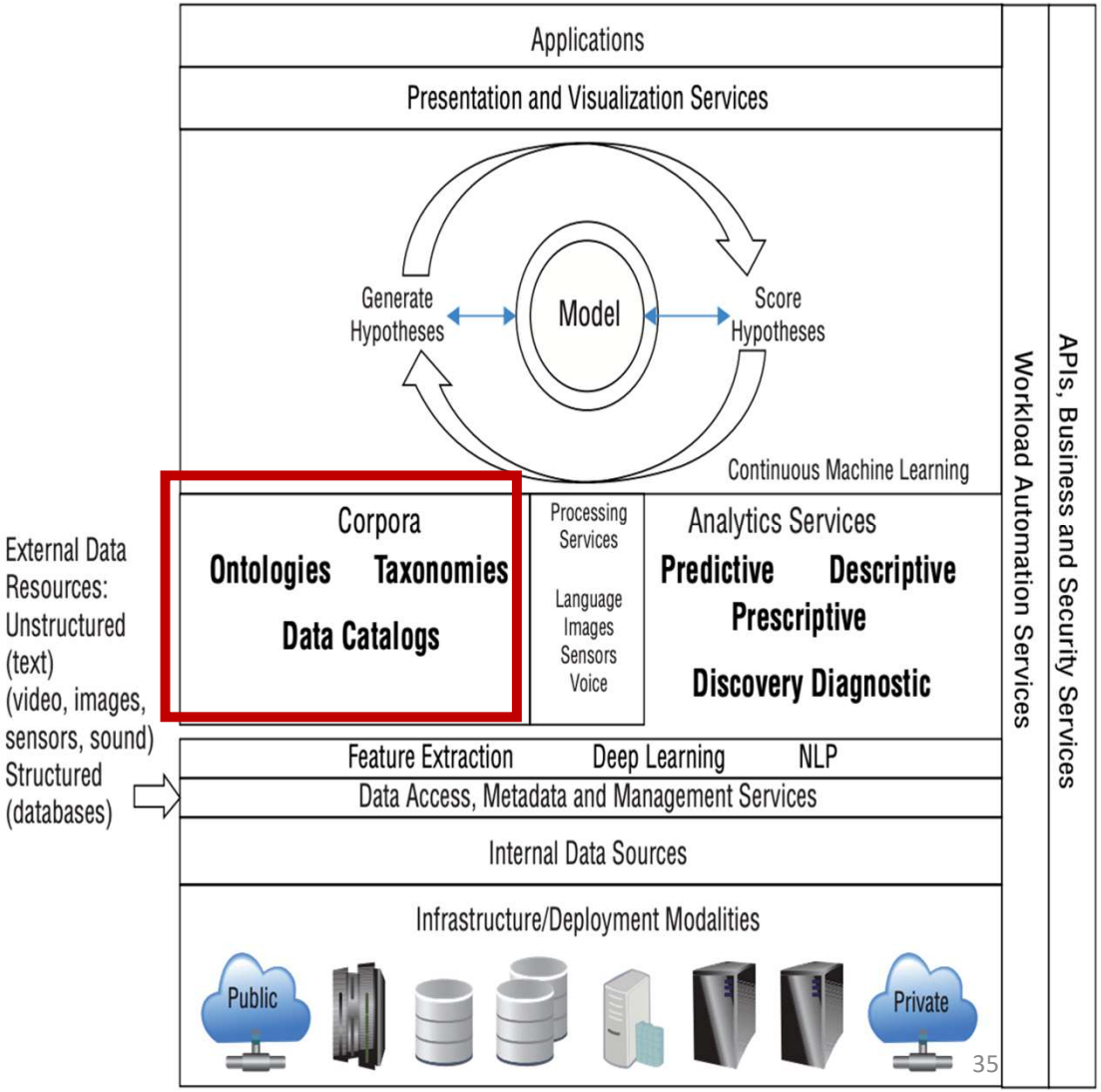


The Corpus, Taxonomies, and Data Catalogs *(contd..)*

Relationship Between Corpus, Taxonomies, and Data Catalogs

Think of a library system:

- The **Corpus** is the collection of all books and materials in the library.
- The **Taxonomy** is the Dewey Decimal System or other classification used to organize the books.
- The **Data Catalog** is the library's catalog, which helps users locate books by title, author, or subject.

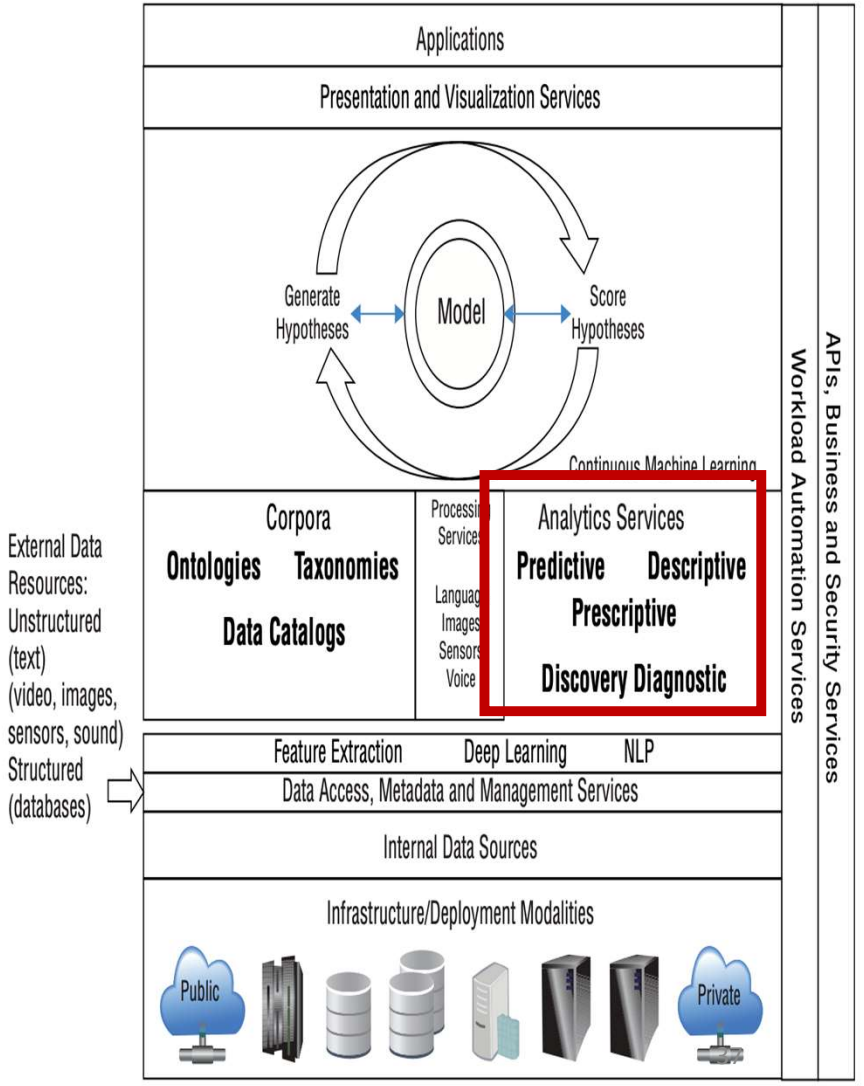


The Corpus, Taxonomies, and Data Catalogs

Concept	Corpus	Taxonomy	Data Catalog
Purpose	Understanding word meanings in context	Structuring knowledge hierarchically	Organizing datasets for easy access
Example for "Cell"	NLP model understands "cell" in medical vs telecom documents	"Cell" in **Biology** → **Animal Cell** / "Cell" in **Telecom** → **Base Station**	Datasets on **human cells** vs **telecom networks**
Example for "Node"	AI differentiates "node" in social vs computer networks	"Node" in **Computer Networks** → **Routers** / "Node" in **Social Networks** → **Users**	Network topology datasets vs social media datasets

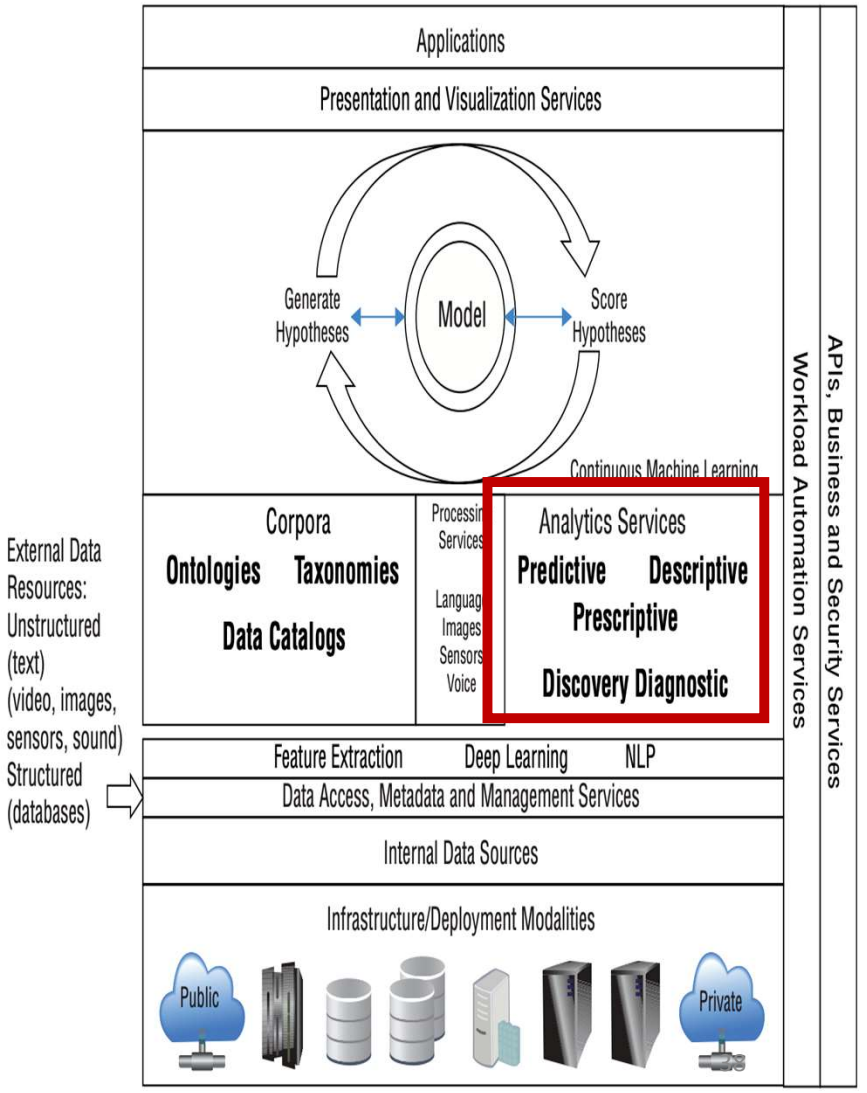
Analytics Services

- Techniques used to gain an understanding of the data ingested and managed within the corpus.
- Users can advantage of structured, unstructured, and semi-structured data that has been ingested and begin to use sophisticated algorithms to predict outcomes, discover patterns, or determine next best actions.
- Continuously access new data from the data access layer and pull data from the corpus.
- Number of advanced algorithms are applied to develop the model for the cognitive system.



Analytics Services

- Analytic approach depends on question being asked
- Five common types of questions:
    1. Descriptive
    2. Diagnostic
    3. Predictive
    4. Prescriptive
    5. Classification



1. Descriptive Questions: “What is the current status?”

Question: "What is the current status of our sales?"

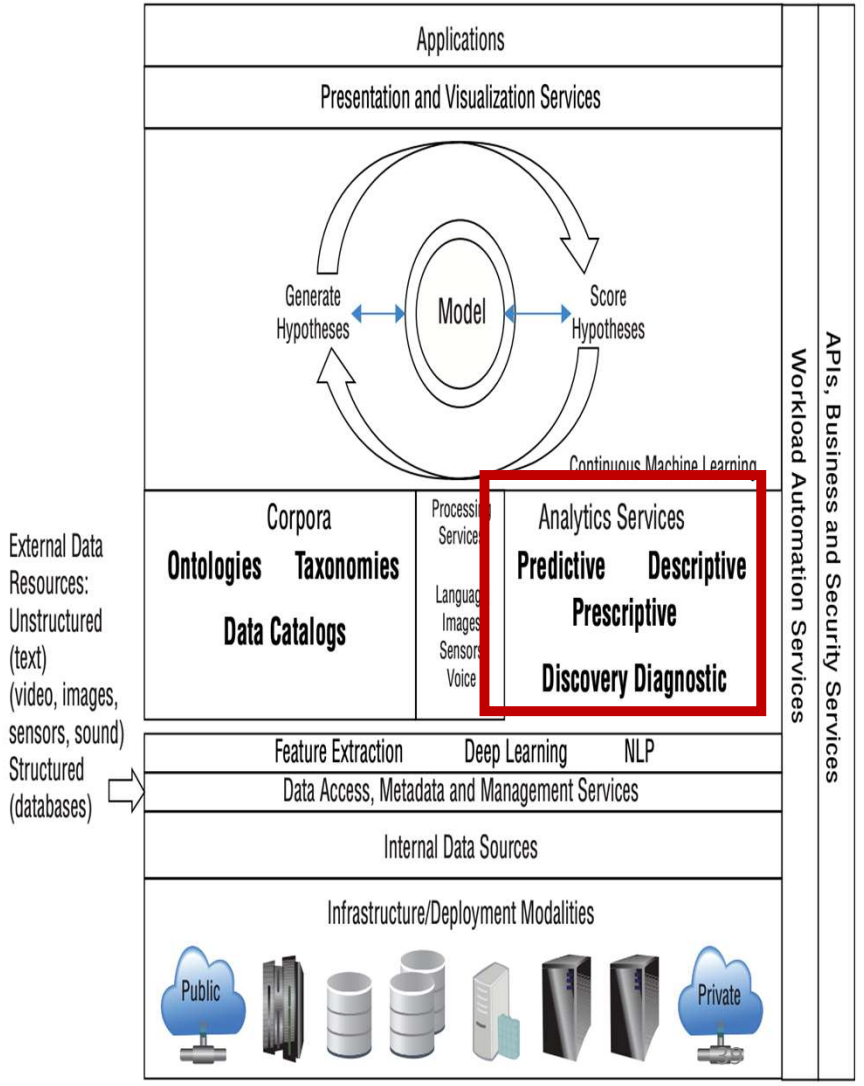
Approach: Descriptive Analytics

Techniques:

- **Data aggregation:** Combining data from various sources into a unified view.
- **Data mining:** Extracting useful information from large datasets.
- **Data visualization:** Using visual tools to present data in an easily understandable format.

Examples:

- Summarizing sales data
- Creating dashboards
- Generating reports



2. Diagnostic Questions: “Why did it happen?”

Question: "Why did our sales decline in the last quarter?"

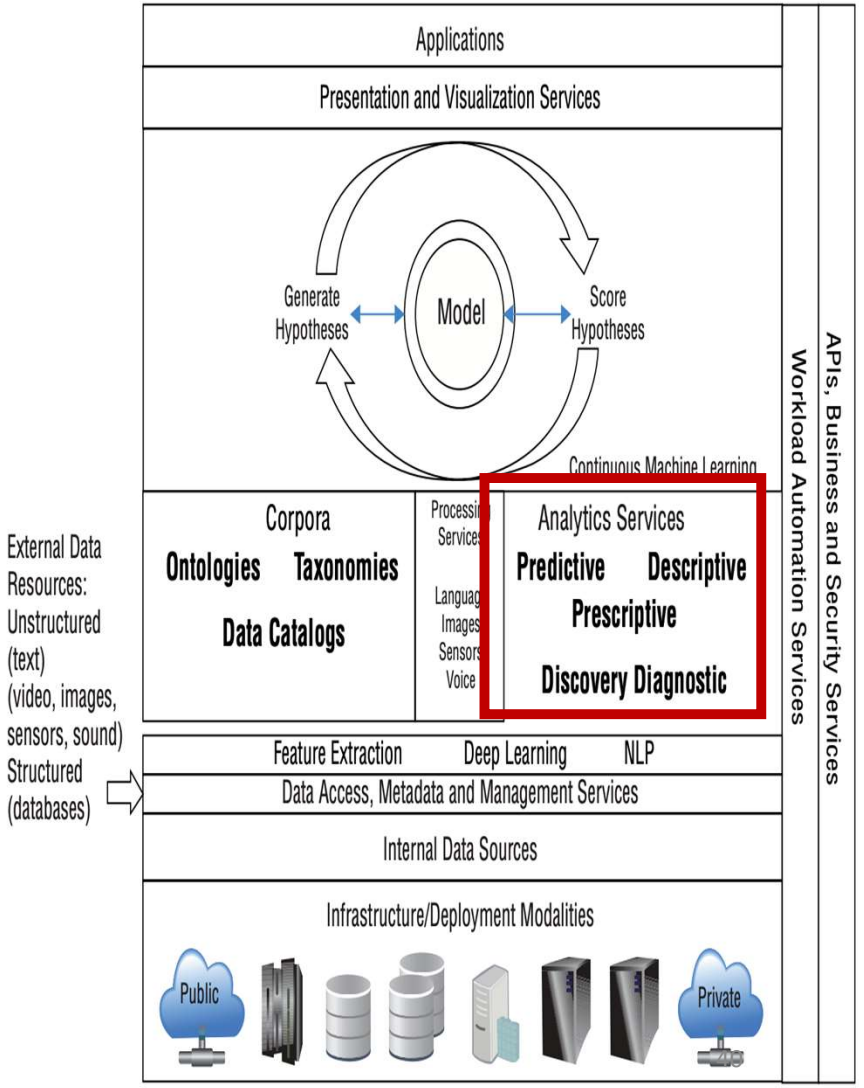
Approach: Diagnostic Analytics

Techniques:

- **Drill-down:** Exploring detailed data to find underlying causes.
- **Data discovery:** Identifying patterns and relationships in data.
- **Correlation analysis:** Assessing the relationship between different variables.

Examples:

- Identifying root causes of sales decline
- Analyzing customer complaints
- Understanding failure points in a process



3. Predictive Questions: “What is likely to happen?”

Question: "What is our sales forecast for the next year?"

Approach: Predictive Analytics

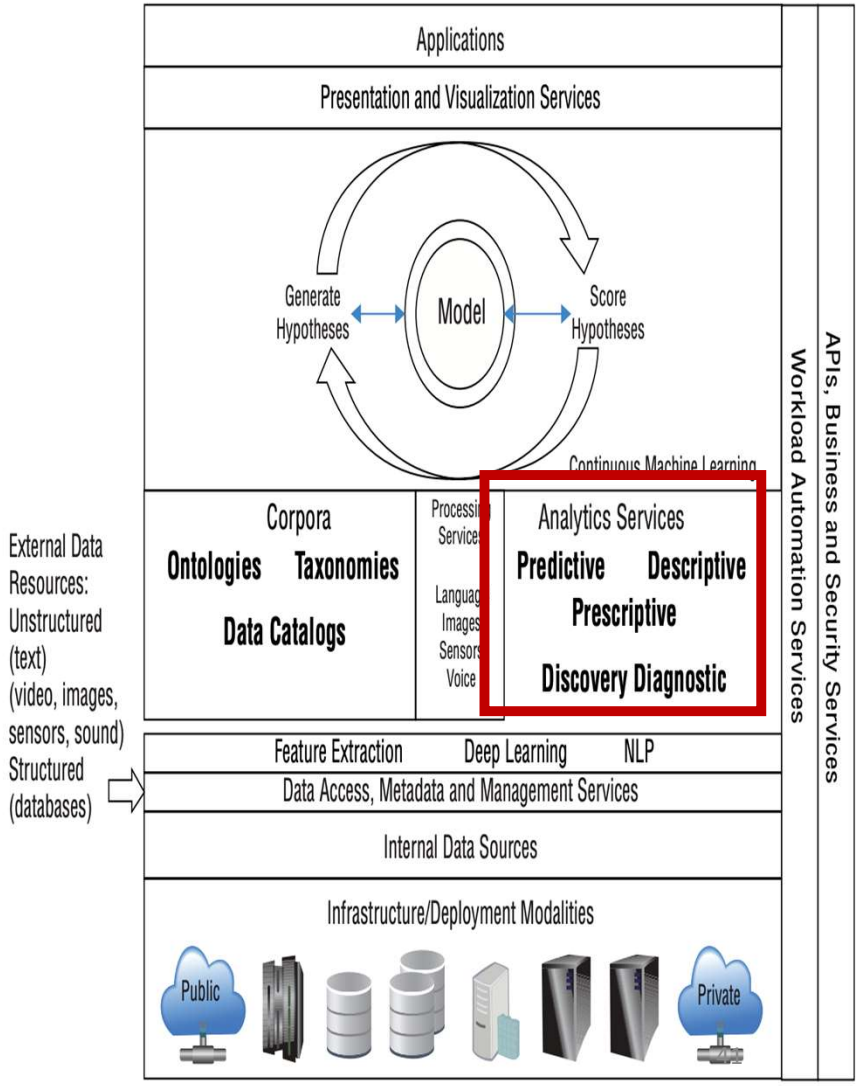
Techniques:

- **Regression analysis:** Predicting outcomes based on relationships between variables.
- **Time series forecasting:** Predicting future values based on past trends.
- **Machine learning models:** Using algorithms to predict future outcomes based on historical data.

Examples:

- Forecasting sales
- Predicting customer churn
- Estimating future demand

Customer churn refers to the percentage of customers who stop using a product or service over a given period.



4. Prescriptive Questions: “What should we do?”

Question: "What should we do to increase website traffic?"

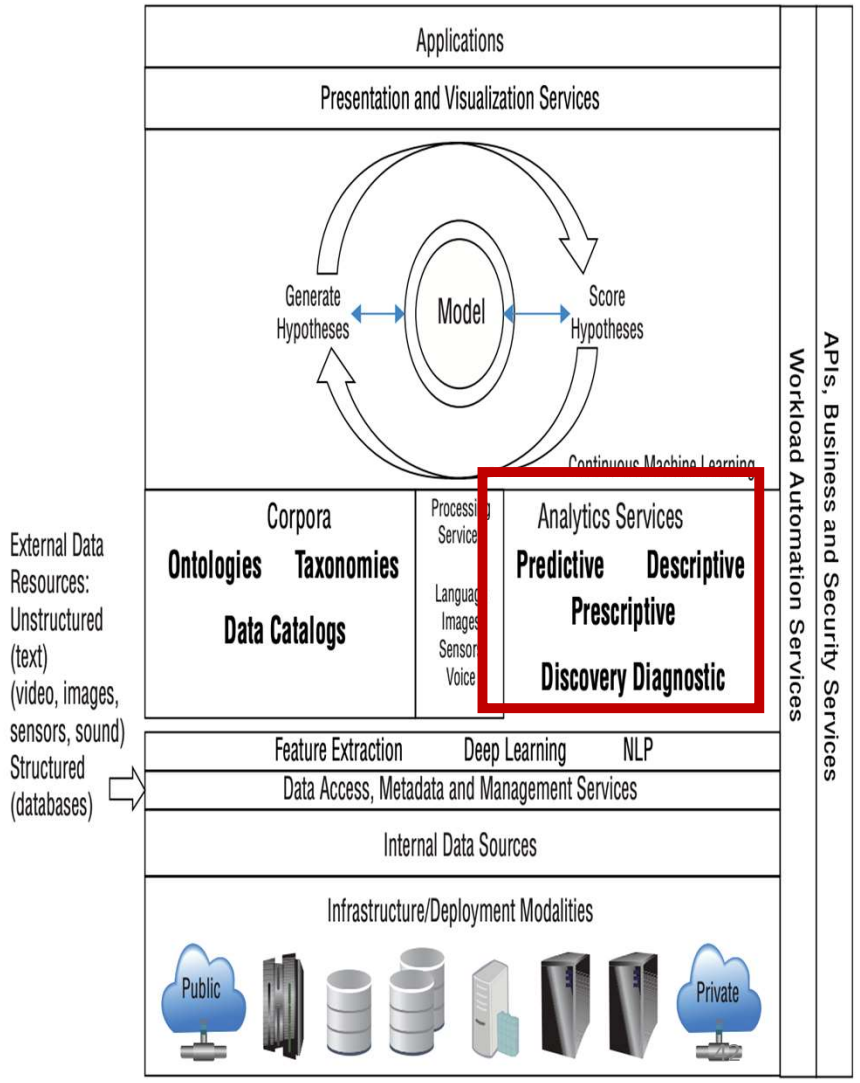
Approach: Prescriptive Analytics

Techniques:

- **Optimization models:** Finding the best solution from a set of alternatives.
- **Simulation:** Modeling scenarios to predict outcomes.
- **Decision analysis:** Evaluating and comparing different decisions.

Examples:

- Recommending inventory levels
- Optimizing marketing campaigns
- Determining pricing strategies



5. Classification Questions: “Which category does this belong to?”

Question: "Which category does this data point belong to**?"**

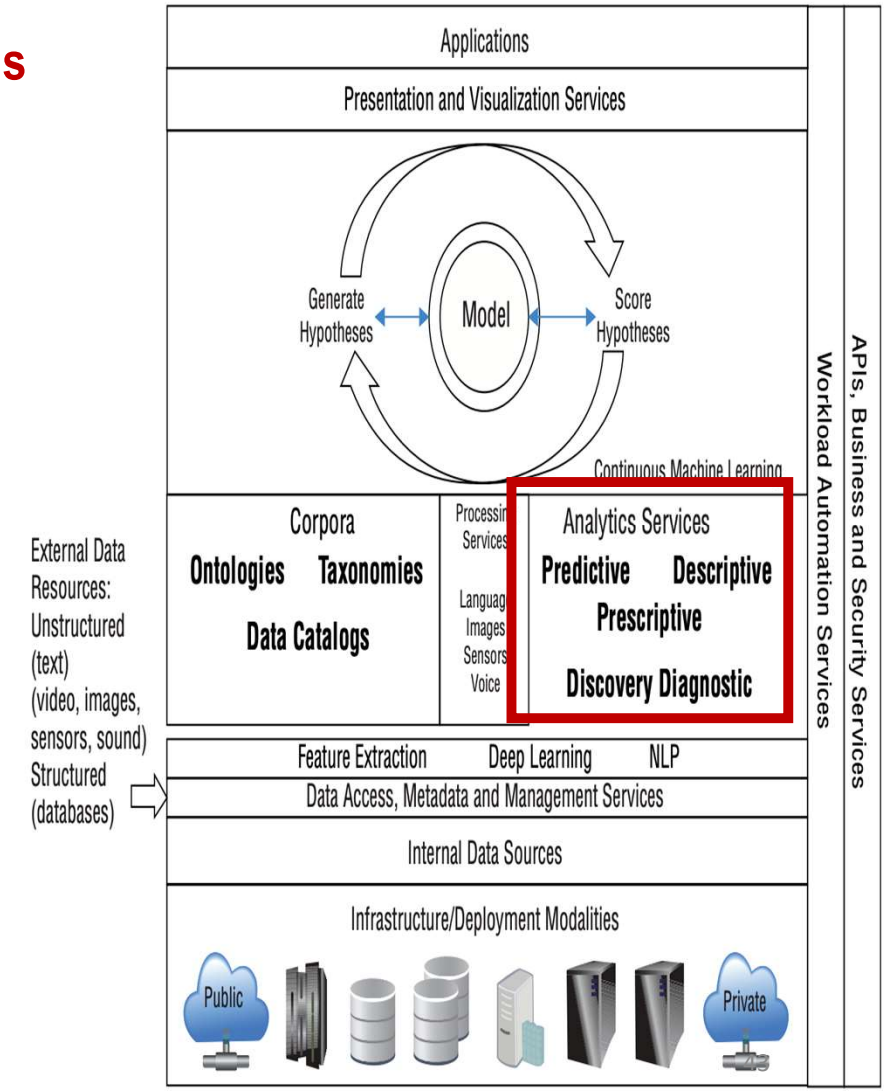
Approach: Classification (Supervised Learning)

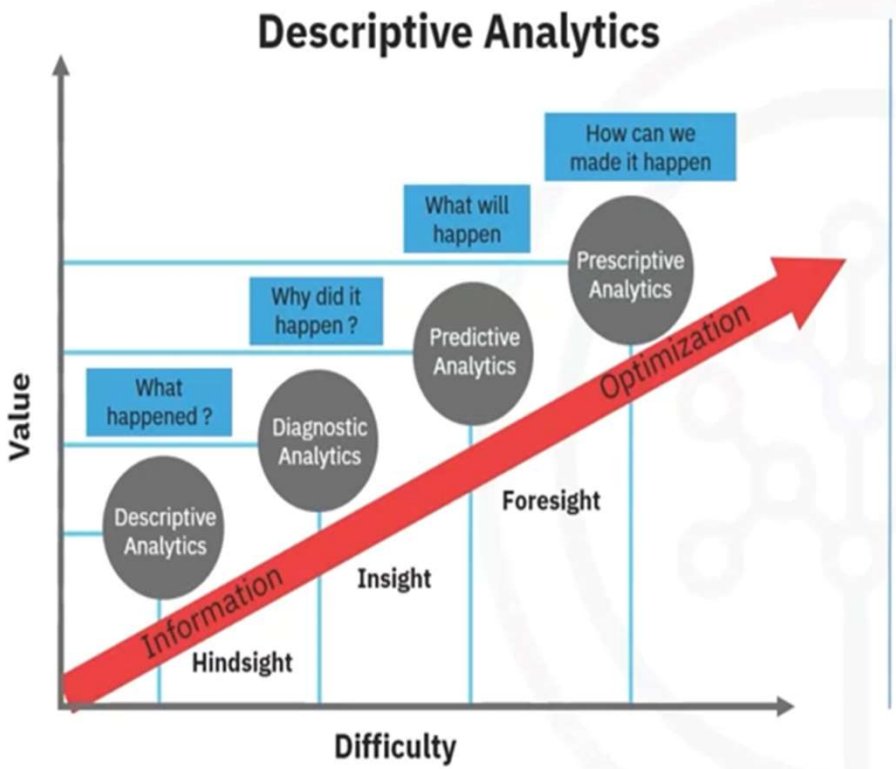
Techniques:

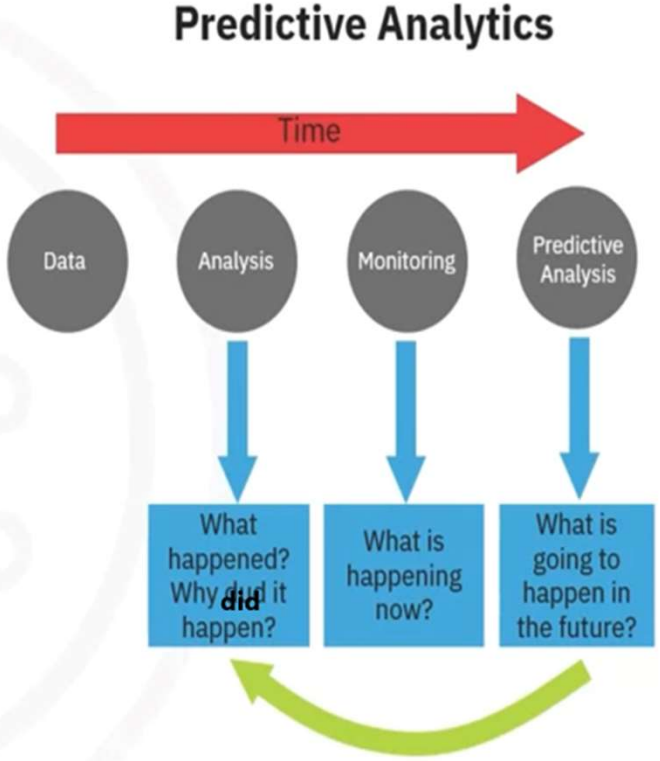
- **Logistic regression:** Predicting the probability of a categorical outcome.
- **Decision trees:** Splitting data into branches to classify it.
- **Support vector machines:** Finding the best boundary to separate categories.
- **Neural networks:** Using interconnected nodes to classify data.

Examples:

- Email spam detection
- Image classification
- Disease diagnosis




Descriptive Analytics
How can we made it happen
What will happen
Why did it happen ?
What happened ?
Prescriptive Analytics
Predictive Analytics
Diagnostic Analytics
Descriptive Analytics
Optimization
Foresight
Insight
Information
Hindsight
Value
Difficulty


Predictive Analytics
Time
Data
Analysis
Monitoring
Predictive Analysis
What happened? Why did it happen?
What is happening now?
What is going to happen in the future?

Pick analytic approach based on the question type

Descriptive

- Current status

Diagnostic (Statistical Analysis)

- What happened?
- Why is this happening?

Predictive (Forecasting)

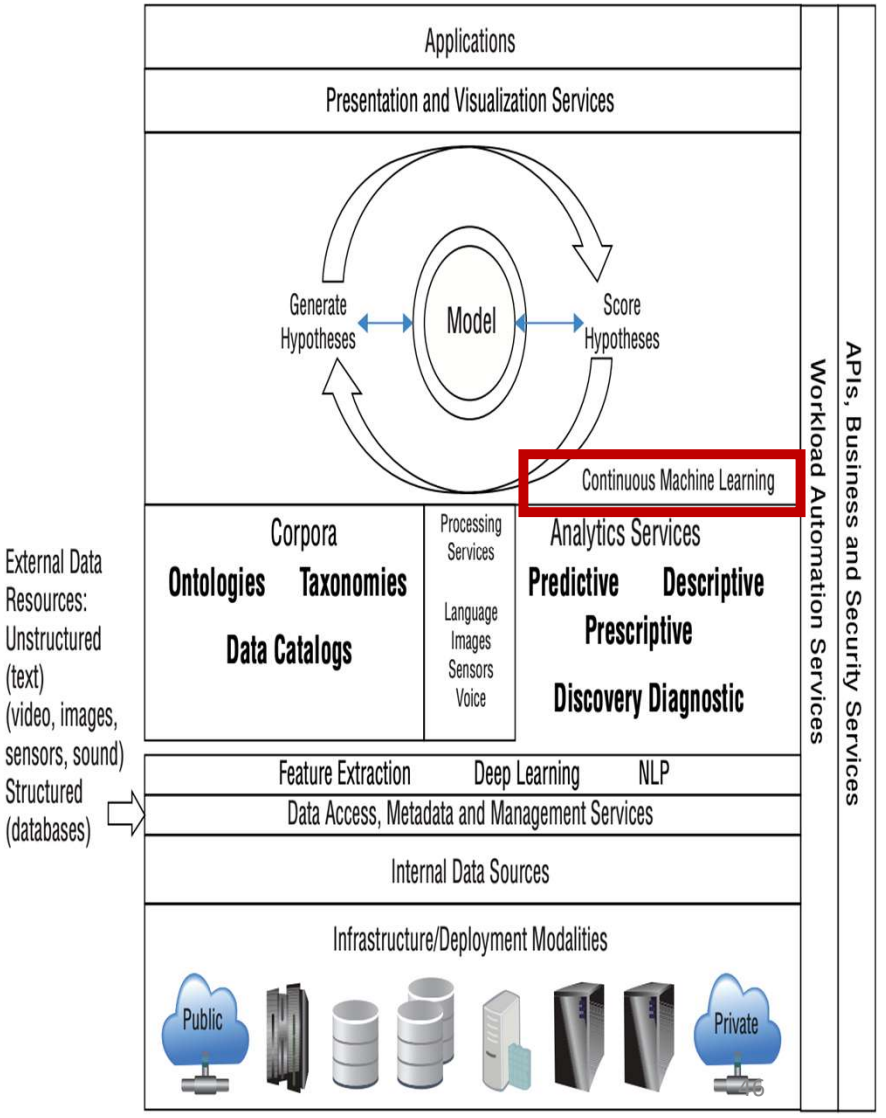
- What if these trends continue?
- What will happen next?

Prescriptive

- How do we solve it?

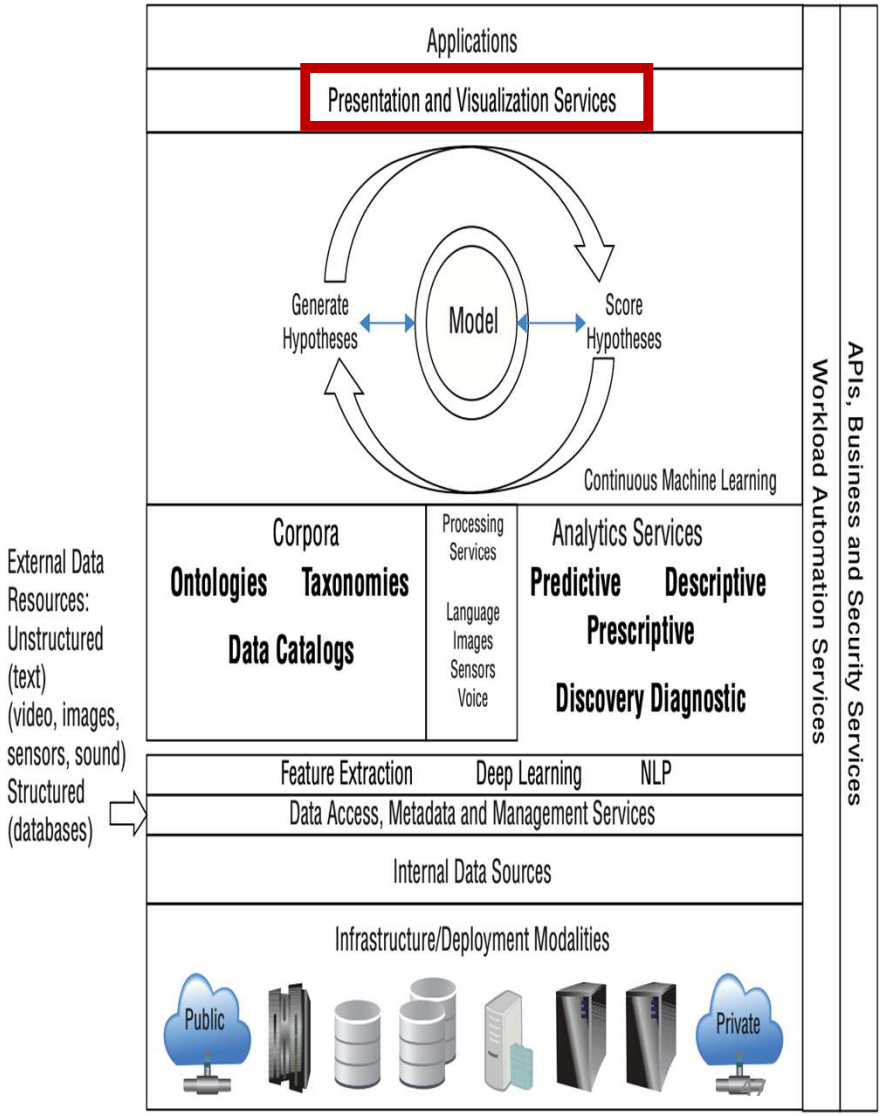
Continuous Machine Learning

- Technique that provides the capability for the data to learn without being explicitly programmed.
- Cognitive systems are not static.
- Rather, models are continuously updated based on new data, analysis, and interactions. A machine learning process has two key elements:
  - Hypothesis generation
  - Hypothesis evaluation



Presentation and Visualization Services

- **Data Visualization**
  - *Visual representation of data as well as the study of data in a visual way*
  - *Types: Static and Dynamic*
- **Presentation Services:** Prepare results for output
- **Visualization Services:** Help to communicate results by providing a way to demonstrate the relationships between data.



Cognitive Applications

- A cognitive system must leverage underlying services to create applications that address problems in a specific domain.
- These applications that are focused on solving specific problems must engage users so that they gain insights and knowledge from the system.