[Overview of the Data Analyst Ecosystem](https://www.coursera.org/learn/introduction-to-data-analytics/lecture/h08bm/overview-of-the-data-analyst-ecosystem?trk_ref=coach_copy)  Jun 4, 2025

The video provides an overview of a **data analyst's ecosystem**, which includes the infrastructure, software, tools, frameworks, and processes used in data analysis. Here’s a detailed summary:

* **Types of Data**:
  + **Structured Data**: Organized in a rigid format, typically found in databases and spreadsheets (e.g., rows and columns).
  + **Semi-Structured Data**: Contains both structured and unstructured elements (e.g., emails with sender/recipient info and email content).
  + **Unstructured Data**: Complex data that cannot be easily organized (e.g., photos, videos, text files, social media content).
* **Data Repositories**:
  + Includes databases, data warehouses, data marts, data lakes, and big data stores. The type and format of data influence the choice of repository.
* **Big Data**:
  + Requires specialized warehouses and frameworks for processing large volumes of data in real-time.
* **Languages**:
  + Query languages (e.g., SQL), programming languages (e.g., Python), and shell/scripting languages are essential for data manipulation and application development.
* **Tools and Frameworks**:
  + Various automated tools and frameworks are used throughout the analytics process, including data gathering, cleaning, mining, analysis, and visualization (e.g., spreadsheets, Jupyter Notebooks, IBM Cognos).

The video sets the stage for deeper exploration of these topics in subsequent sections of the course.

[Types of Data](https://www.coursera.org/learn/introduction-to-data-analytics/lecture/ePgdX/types-of-data?trk_ref=coach_copy)  Jun 4, 2025

I understand you're looking to take notes. Here’s a concise summary of the key points regarding data types from the course material:

**Types of Data**

1. **Structured Data**
   * **Definition**: Well-defined structure, adheres to a specified data model.
   * **Storage**: Stored in relational databases (e.g., SQL).
   * **Examples**:
     + SQL Databases
     + Spreadsheets (Excel, Google Sheets)
     + Online forms
     + Sensors (GPS, RFID)
   * **Analysis**: Easily examined with standard data analysis methods.
2. **Semi-Structured Data**
   * **Definition**: Some organizational properties but lacks a fixed schema.
   * **Storage**: Cannot be stored in traditional rows and columns.
   * **Examples**:
     + E-mails
     + XML and JSON files
     + Binary executables
   * **Characteristics**: Contains metadata for grouping and hierarchy.
3. **Unstructured Data**
   * **Definition**: No easily identifiable structure, cannot be organized in traditional databases.
   * **Storage**: Stored in files/documents or NoSQL databases.
   * **Examples**:
     + Web pages
     + Social media feeds
     + Images, videos, audio files
     + Documents (PDFs, Word files)
   * **Analysis**: Requires specialized tools for analysis.

**Summary**

* **Structured Data**: Organized, easily analyzed.
* **Semi-Structured Data**: Some organization, uses metadata.
* **Unstructured Data**: No conventional organization, diverse sources.