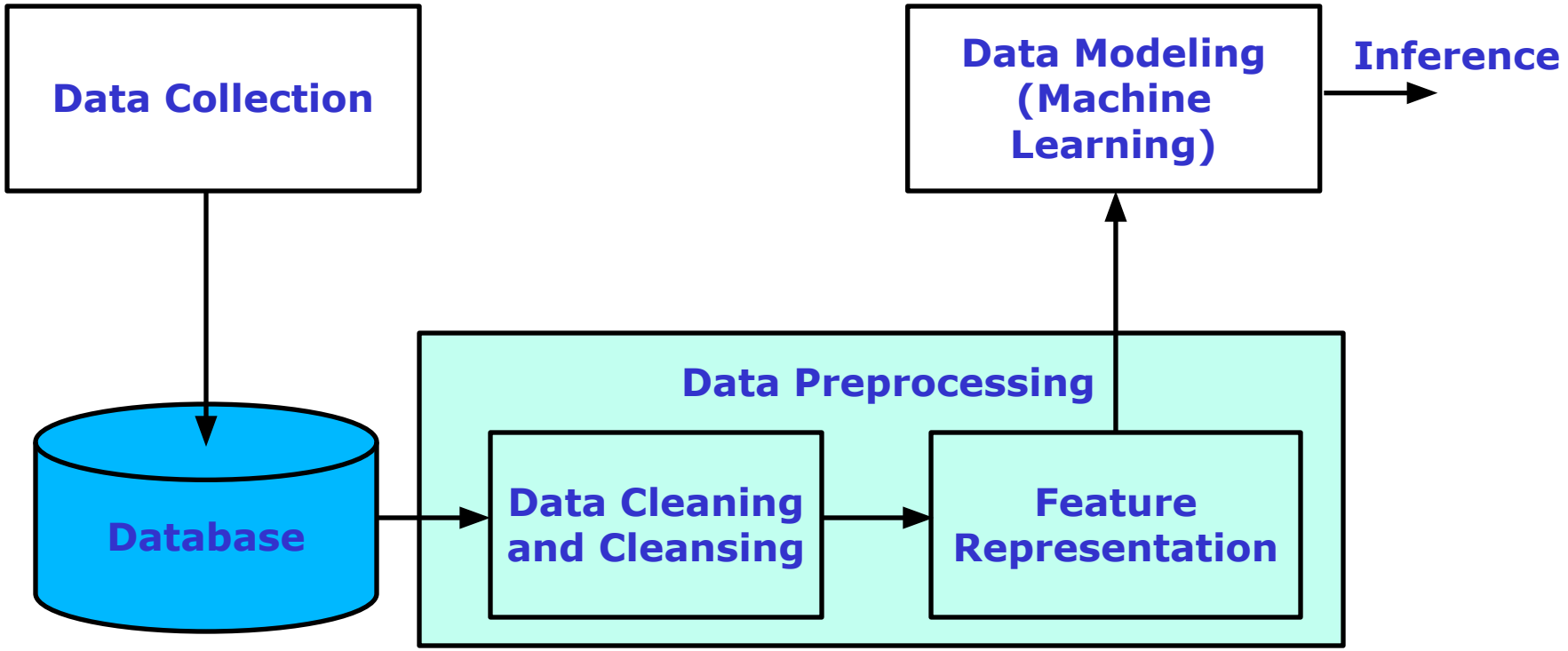
Data, Types of Data and Data Collection using Sensors

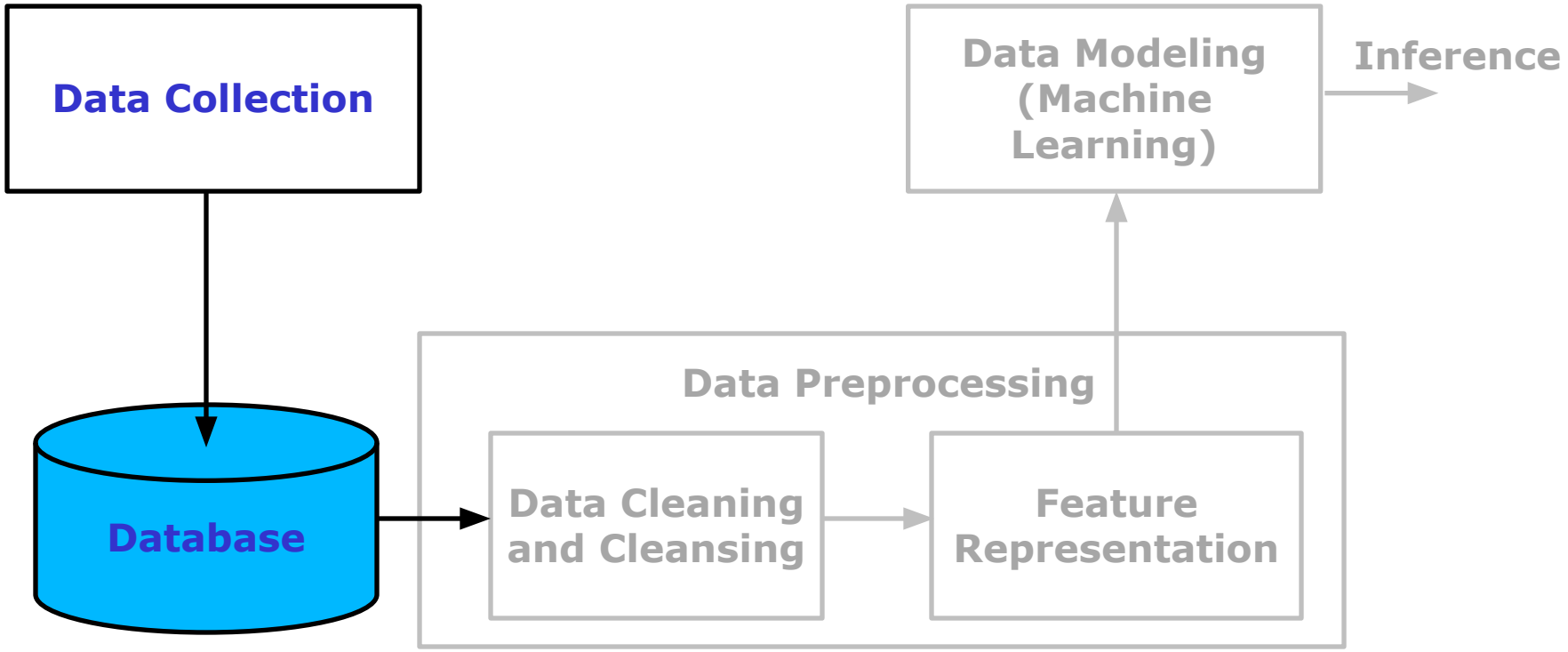
Data Science

- Multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insight from structured and unstructured data
- Central concept is gaining insight from data
- Machine learning uses data to extract knowledge



Data Science

- Multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insight from structured and unstructured data
- Central concept is gaining insight from data
- Machine learning uses data to extract knowledge



Data Collection

- Data manifests itself in many different forms
- Different forms of data require different ways to collect them and different storage solutions
- Collection of data may consists of sending out surveys, polls or doing other experiments
- Data based on the way it is collected:
  - Data that comes from surveys
    - Usually textual form of data or mixed

Sl. No	Timestamp	Email Address	Please mention your name	Please mention your Roll Number	Do you have a Laptop/Tablet which can help you to view videos?	What type of internet connection do you have?	Do you have the Laptop/Desktop for doing programming assignments?
78	9/14/2020 19:14:17	b19042@students.iitmandi.ac.in	Manan Shah	19042	Yes	4G and above	Yes
121	9/14/2020 20:31:32	b19107@students.iitmandi.ac.in	Rishabh Garg	19107	Yes	3G	Yes
25	9/14/2020 18:52:14	b19001@students.iitmandi.ac.in	Aarushi Gajri	B19001	Yes	4G and above	Yes
120	9/14/2020 20:30:22	b19002@students.iitmandi.ac.in	Aditya Narayan khokhar	B19002	Yes	3G	Yes
106	9/14/2020 19:56:31	b19003@students.iitmandi.ac.in	Aditya Sarkar	B19003	Yes	4G and above	Yes
146	9/14/2020 21:58:00	b19004@students.iitmandi.ac.in	Anooshka Bajaj	b19004	Yes	4G and above	Yes
4	9/15/2020 11:12:56	b19005@students.iitmandi.ac.in	Krishna sai	B19005	No	3G	No
83	9/14/2020 19:18:31	b19006@students.iitmandi.ac.in	Chirag	B19006	Yes	Bad internet connection fit only to view/download to lecture pdf files.	Yes
175	9/15/2020 8:30:34	b19008@students.iitmandi.ac.in	Samvivek	B19008	Yes	3G	Yes
215	9/16/2020 3:08:13	b19010@students.iitmandi.ac.in	Kshitij Nair	b19010	Yes	4G and above	Yes
69	9/14/2020 19:07:44	b19011@students.iitmandi.ac.in	Laishram	B19011	Yes	3G	Yes

Data Collection

- Data manifests itself in many different forms
- Different forms of data require different ways to collect them and different storage solutions
- Collection of data may consists of sending out surveys, polls or doing other experiments
- Data based on the way it is collected:
  - Data that comes from surveys
    - Usually textual form of data or mixed
  - Data entered in a database as system entry
    - E.g. Student information entered on academic automation system etc.
  - Data in the form of signals (comes from sensors)
    - Speech/Audio, Images and videos, Temperature readings, Humidity, Seismic data, EEG (all bio-type signals) etc.
- According to the objective of the task, the way the data is collected will change

Types of Data: Based on Organization

1. Unstructured data:
   - Rawest form of data
   - Example: Any type of files like texts, images, sounds or videos etc.
   - This type of data stored in a repository of files
     - Well organised directories on the computer hard drive

Types of Data: Based on Organization

2. Structured data:
   - It is a tabular data (rows and columns), which are very well defined

Date/ Time	Temperature (C)/ Humidity (%)	Pressure (Pa)	Rain (inches)	Light intensity (lux)	Accelerations (g)	Force (N)	Moisture (%)
2017-09-06 18:44:32	23.00,56.00	617.64	0.01	3	0.52,0.31,-0.80,0.00,0.00,0.00,31.36,-159.01	0.02	81.00
2017-09-06 18:33:32	24.00,58.00	619.47	0.01	12	0.52,0.30,-0.79,0.00,0.00,0.00,31.45,-159.12	0.02	82.00
2017-09-06 18:22:39	24.00,58.00	623.37	0.00	71	0.52,0.31,-0.80,0.00,0.00,0.00,31.35,-158.88	0.02	83.00
2017-09-06 18:11:31	25.00,60.00	627.02	0.05	194	0.51,0.31,-0.80,0.00,0.00,0.00,30.80,-159.00	0.02	81.00

   - Stored in databases
     - Spreadsheets [Comma Separated Value (CSV) format]
     - Oracle
     - DB2
     - MySQL etc.

Types of Data: Based on Organization

3. Semi-Structured data:
   - Anywhere between unstructured and structured data
   - A consistent format is defined, however there is no strict structure and parts of data may be incomplete or different type
   - Example: Data in the form of XML and JSON
     - Stored in document oriented databases

Types of Data: Based on Organization

3. Semi-Structured data:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<bookstore>

  <book category="cooking">
    <title lang="en">Everyday Italian</title>
    <author>Giada De Laurentiis</author>
    <year>2005</year>
    <price>30.00</price>
  </book>

  <book category="children">
    <title lang="en">Harry Potter</title>
    <author>J K. Rowling</author>
    <year>2005</year>
    <price>29.99</price>
  </book>

  <book category="web">
    <title lang="en">XQuery Kick Start</title>
    <author>James McGovern</author>
    <author>Per Bothner</author>
    <author>Kurt Cagle</author>
```

- Anywhere between unstructured and structured data
- A consistent format is defined, however there is no strict structure and parts of data may be incomplete or different type
- Example: Data in the form of XML and JSON
  - Stored in document oriented databases

Type of Data: Based on Variables (Value) found in Data

- Mainly in Structured Data:

1. Numerical data:
   - Data represented as numbers
   - Data in which information is measurable
   - This type of data is called quantitative data as its describes a quantity
   - Two types based on the values taken:
     - Continuous valued data:
       - Numbers does not have logical end
       - Range lies in the natural limit of what we are measuring
       - Example: Cost of the books, atmospheric temperature etc.
     - Discrete valued data:
       - Numbers have logical end
       - There is a specific limit on the range of the values
       - Example: number of members of family, number of days in a month, number of colours in flag etc.

Type of Data: Based on Variables (Value) found in Data

2. Categorical data:
   - Data that is not a number. It can be string of text or date
   - It describe an item or event to one of few different categories
   - Example: Ethnicity, gender, eye colour, etc.
   - This type of data is called qualitative data as its describes a quality
   - Three types values they hold:
     - Ordinal values: Values that have a set order to them
       - Example: Severity of a alarm as "Critical", "Medium" and ""Low", Ranking of a running race as " First", Second", Third"
     - Nominal values: Values that have no set order to them
       - Example: Values for the variables "Marital Status", "Country", "Eye Colour" etc.
     - Binary values: Special type of categorical data
       - Have only two values – "Yes" and "No" OR "True" and "False" OR "1" and "0"

Type of Data: Based on Variables (Value) found in Data

3. Time series data:
   - Series of data. It involve time and some kind of value
   - Example: Temperature at every hour
   - It is clearly structured and numeric in nature
   - Special case of numerical data
   - This type of data is important because of IoT and sensors
   - Data from sensors are almost always time-series in nature

Date/ Time	Temperature (C)/ Humidity (%)	Pressure (Pa)	Rain (inches)	Light intensity (lux)	Accelerations (g)	Force (N)	Moisture (%)
2017-09-06 18:44:32	23.00,56.00	617.64	0.01	3	0.52,0.31,-0.80,0.00,0.00,0.00,31.36,-159.01	0.02	81.00
2017-09-06 18:33:32	24.00,58.00	619.47	0.01	12	0.52,0.30,-0.79,0.00,0.00,0.00,31.45,-159.12	0.02	82.00
2017-09-06 18:22:39	24.00,58.00	623.37	0.00	71	0.52,0.31,-0.80,0.00,0.00,0.00,31.35,-158.88	0.02	83.00
2017-09-06 18:11:31	25.00,60.00	627.02	0.05	194	0.51,0.31,-0.80,0.00,0.00,0.00,30.80,-159.00	0.02	81.00