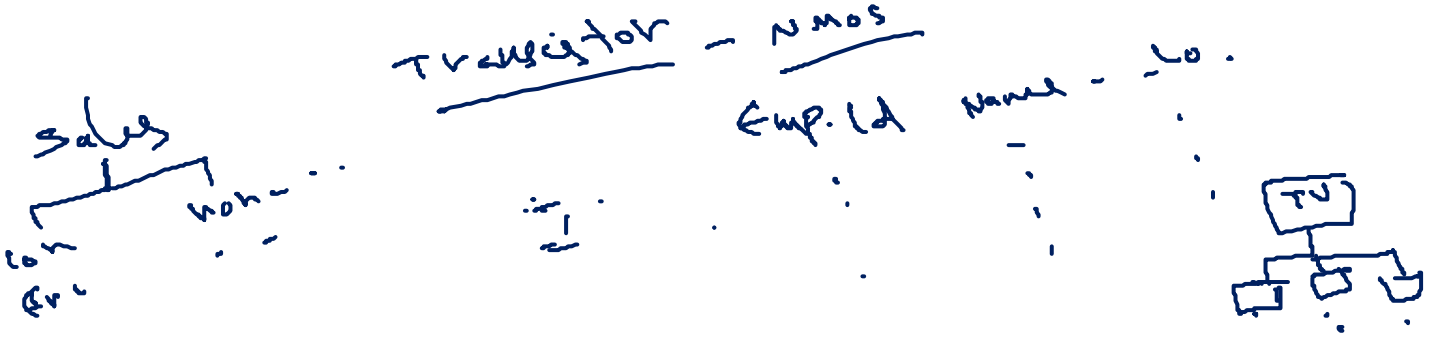
# Module 2

Tags to organize data

# Tags




- Tags are used to organise the data objects. Tagging allows users to categorize the data and organize them functionally.
- Tags are added to data objects like a database, table, column, report, query and files. Each data object can have a single tag or multiple tags. Also, each tag can have multiple data objects.
- **tag** typically refers to a piece of metadata added to a data set for the purpose of improving data organization and findability.
- Organize data coming from a large number of sources.
- Tags are chosen informally and personally by the item's creator or by its viewer, depending on the system

# Tags

- In information systems, a tag is a keyword or term assigned to a piece of information (such as an Internet bookmark, digital image, database record, or computer file).

- This kind of metadata helps describe an item and allows it to be found again by browsing or searching.

- Tags are generally chosen informally and personally by the item's creator or by its viewer, depending on the system, although they may also be chosen from a controlled vocabulary.

- Tags are useful for organizing data at a high level. If you wanted to add more fine-grained detail, like capturing user names or other high-cardinality values, custom attributes or custom events would be a better solution.

# Tag

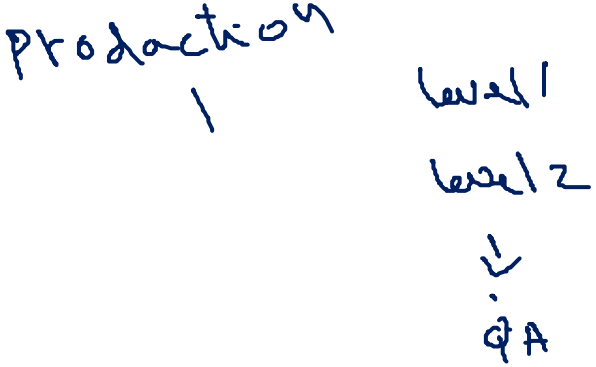
- Labeling and tagging are carried out to perform functions such as aiding in classification and indicating online identity.
- They may take the form of words, images, or other identifying marks.
- Tags are generally chosen informally and personally by the item's creator or by its viewer
- Depending on the system, they may also be chosen from a controlled vocabulary.

# Metadata

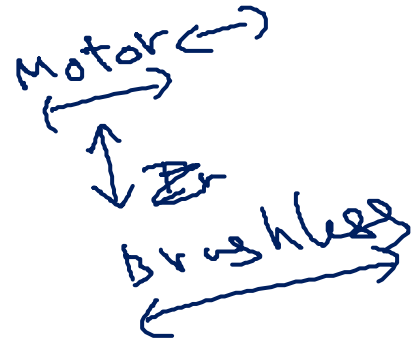
- It is a description and context of the data.
- It helps to organize, find and understand data.
- Metadata is simply **data about data**.
- Typical metadata elements:
  - Title and description,
  - Tags and categories,
  - Who created and when,
  - Who last modified and when,
  - Who can access or update.

# Tag examples


- **1. Team related tags:**
- Creating tags for team names can be helpful to understand which team, group, department, or region was responsible for a change that led to a performance issue.
- Examples: backend, frontend, db
- Roles: Architect, devops, pm
- **2. Environment related tags:**
- Create entities for the environment they belong to
- Examples: production, QA, development, staging
- **3. Critically related tags:**
- Create tags related to criticality levels, easier to track the most critical issues.
- Examples: level1, level2



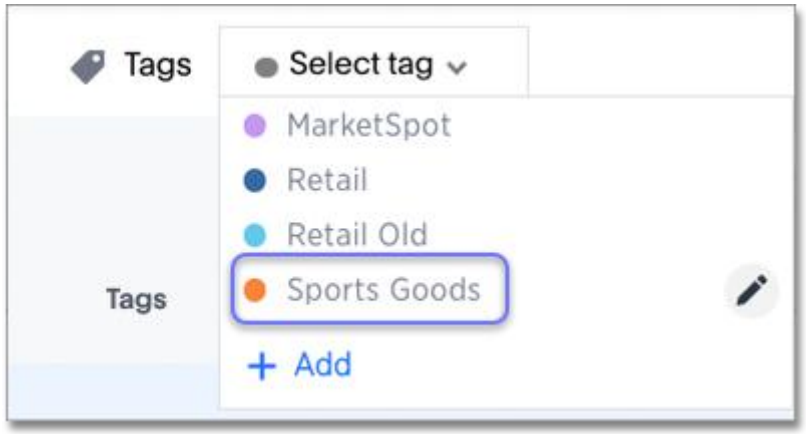
# *Tagging*




- Adding non-hierarchical keywords or terms to
  - Documents
  - Websites
  - Files
  - Lists
  - Social media content

# Tags to organise data

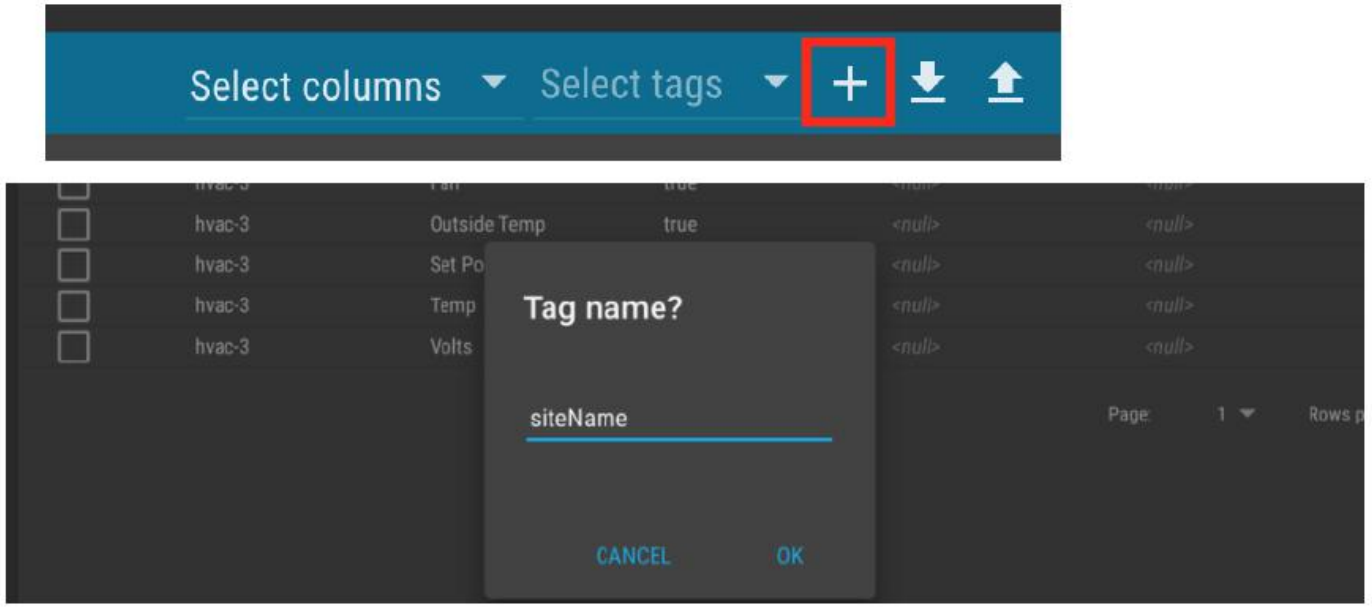
- Tags enable you to create categories for classification of objects, including Pinboards, Answers, data sources, and Worksheets. Tags are often used to designate subject areas, such as sales, HR, and finance.
- Create a tag: Only administrator users can create tags. Anyone can apply the tags you create, or use them as filters when selecting from a list of sources or Pinboards.
- Filter by tag: tags to filter lists of Pinboards or data sources. You can also filter by tag when selecting data sources.



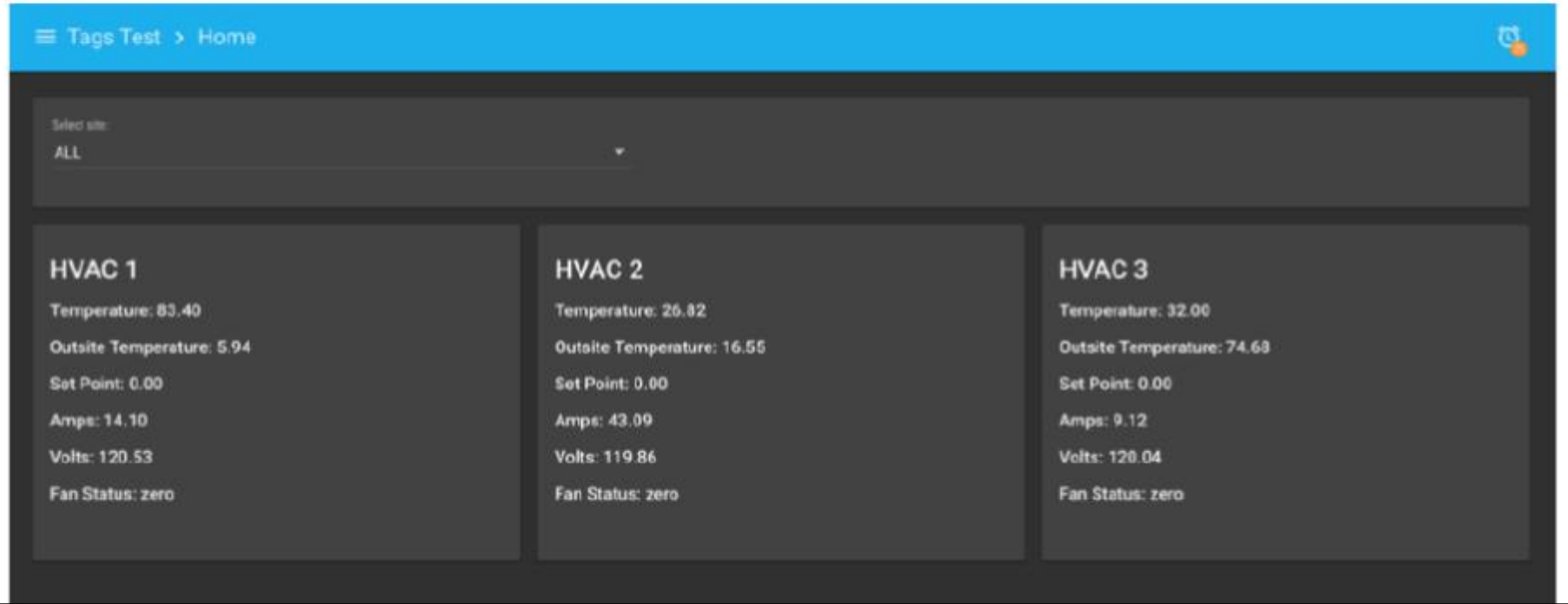
# Tags to organize data

Example: To monitor three HVAC systems (hvac-1, hvac-2, and hvac-3) at different locations, two of them on Site 1 (hvac-1 and hvac-2) and the last one on Site 2. Each HVAC have following variables:

- ❑ Current
- ❑ Voltage
- ❑ Temperature
- ❑ Outside temperature
- ❑ Temperature set point
- ❑ Fan state (ON/OFF)




Tags Test > HVAC Details
Last 1 hours
Select site:
ALL
Select HVAC:
HVAC 2
Temperature Variables
80
70
60
50
40
30
20
10
0
11:10 AM
11:20 AM
11:30 AM
11:40 AM
11:50 AM
12:00 PM
Temperature
Outsite Temperature
Set Point
Electrical Variables
140
120
100
80
60
40
20
0
11:10 AM
11:20 AM
11:30 AM
11:40 AM
11:50 AM
12:00 PM
Amps
Volts


Tags Test > Home
Select site:
ALL
HVAC 1
Temperature: 83.40
Outsite Temperature: 5.94
Set Point: 0.00
Amps: 14.10
Volts: 120.53
Fan Status: zero
HVAC 2
Temperature: 26.82
Outsite Temperature: 16.55
Set Point: 0.00
Amps: 43.09
Volts: 119.86
Fan Status: zero
HVAC 3
Temperature: 32.00
Outsite Temperature: 74.68
Set Point: 0.00
Amps: 9.12
Volts: 120.04
Fan Status: zero

# Tags in web services

- A tag typically refers to a piece of metadata added to a data set for the purpose of improving data organization and data searching.
- Tags are key: value pairs (like team: operations) added to various sets of data, like monitored apps and hosts, agents, dashboards etc.
- Important attributes can be made available as tags (for example, app metadata like app name and language, and host metadata like host name and region).

owner purpose

location.

# Tags in AWS

- Metadata can be assigned to AWS resources in the form of tags.
- Each tag is a label consisting of a user-defined key and value.
- Tags are used to manage, identify, organize, search for, and filter resources.
- Tags can be created to categorize resources by purpose, owner, environment, or other criteria.

Each tag has two parts:

- A tag key (for example, CostCenter, Environment, or Project). Tag keys are case sensitive.
- A tag value (for example, 111122223333 or Production). Like tag keys, tag values are case sensitive.

## Tag naming limits and requirements(AWS)

- Each resource can have a maximum of 50 user created tags.
- For each resource, each tag key must be unique, and each tag key can have only one value.
- The tag key must be a minimum of 1 and a maximum of 128 Unicode characters in UTF-8.
- The tag value must be a minimum of 0 and a maximum of 256 Unicode characters in UTF-8.
- Allowed characters can vary by AWS service. For information about what characters you can use to tag resources in a particular AWS service, see its documentation. In general, allowed characters in tags are letters, numbers, spaces representable in UTF-8, and the following characters: _ . : / = + - @ .
- Tag keys and values are case sensitive. As a best practice, decide on a strategy for capitalizing tags, and consistently implement that strategy across all resource types. For example, decide whether to use Costcenter, costcenter, or CostCenter, and use the same convention for all tags. Avoid using similar tags with inconsistent case treatment.

## Tips on keeping tags simple

- Start by adding only tags you know you'll definitely use. Unused tags create noise and may add confusion.
- Try to use short tags. Shorter tags are easier to parse, and also the UI may sometimes truncate longer tags.
- When possible, use keys and values that are human-readable (for example, region: EMEA is better than Param8323 : 1229072).
- Avoid including several values like regions: EMEA | US | LATAM in a single tag. Instead use three different tags like region: emea, region: us, and region: latam.

## Tips on using tags consistently:

Try to be consistent with tag language across your teams and entities.

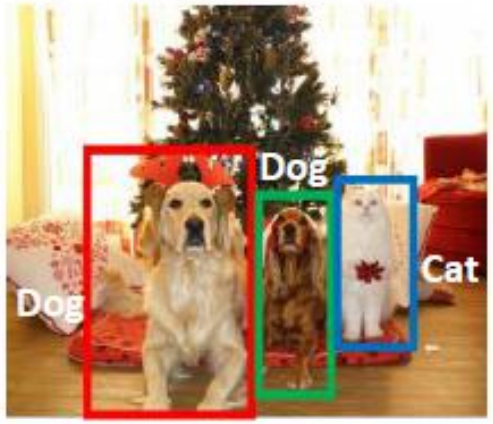
- Be consistent with naming. For example, avoid using both region: emea and reg: emea.
- Be consistent with capitalization. For example, avoid using both env: staging and env: Staging (although tag searching is case-insensitive in the UI and API)

# Tags to process large datasets

- Data labeling is the process of identifying raw data (images, text files, videos, etc.) and adding one or more meaningful and informative labels to provide context so that a machine learning model can learn from it

Examples: labels might indicate whether

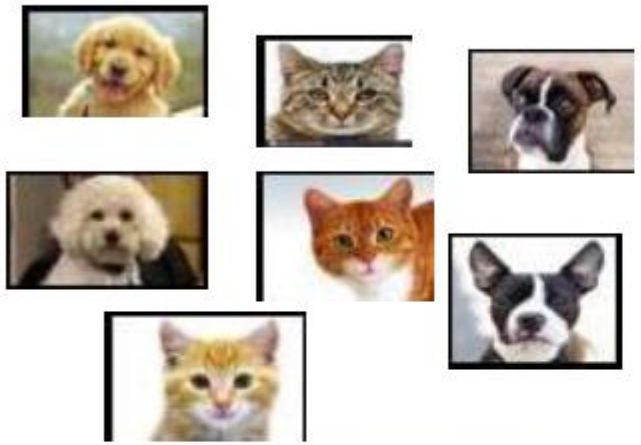
- a photo contains a cat or dog




- which words were uttered in an audio recording
- if an x-ray contains a tumor

Data labeling also called data annotation/tagging is the process of preparing labeled datasets.

## Unlabeled data



| Age | Income | Loan Status |
|---|---|---|
| 21 | 20000 | ? |
| 37 | 55000 | ? |
| 29 | 35000 | ? |
| 23 | 17000 | ? |
| 34 | 70000 | ? |
| 47 | 84000 | ? |
| 25 | 30000 | ? |

## Labeled data

| Input | Image Label |
|---|---|
|  | Dog |
|  | Dog |
|  | Cat |
|  | Dog |
|  | Cat |
|  | Cat |
|  | Dog |

| Age | Income | Loan Status |
|---|---|---|
| 21 | 20000 | Rejected |
| 37 | 55000 | Approved |
| 29 | 35000 | Approved |
| 23 | 17000 | Rejected |
| 34 | 70000 | Approved |
| 47 | 84000 | Rejected |
| 25 | 30000 | Approved |

## Need of Data labeling

- Data labeling is required for a variety of use cases including computer vision, natural language processing, and speech recognition.
- Most practical machine learning models utilize supervised learning, which applies an algorithm to map one input to one output. Supervised learning requires labeled dataset to learn from dataset and to make correct decisions.

## Data labeling Process

- Data labeling typically starts by asking humans to make judgments about a given piece of unlabeled data.
- For example, labelers may be asked to tag all the images in a dataset where “does the photo contain a bird” is true. The tagging can be as rough as a simple yes/no or as granular as identifying the specific pixels in the image associated with the bird.
- The machine learning model uses human-provided labels to learn the underlying patterns in a process called "model training." The result is a trained model that can be used for labelling new set of data.
- In machine learning, a properly labeled dataset which is used as the objective standard to train and assess a given model is often called “ground truth.” The accuracy of trained model depends on the accuracy of ground truth.

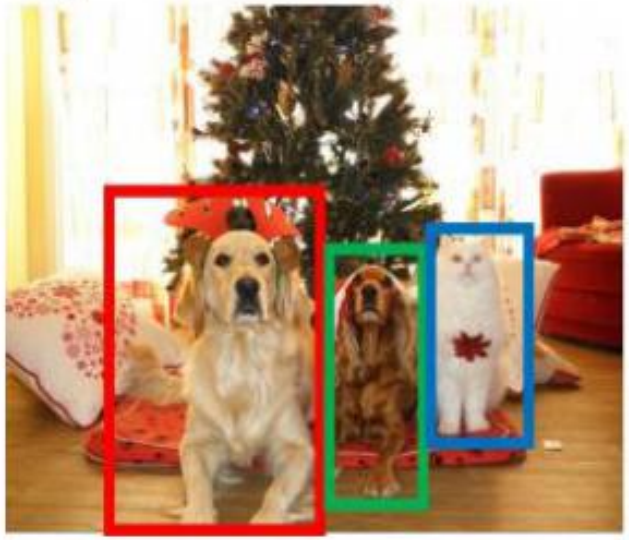
# Machine Learning Model to Label Data

- Successful machine learning models development relies on large volumes of high-quality training data
- Machine learning model can be used to label data automatically
- In this process, a machine learning model for labeling data is first trained on a subset of raw data that has been labeled by humans.
- If the labeling(training) model has high confidence in its results (based on what the labeling model has learned so far), it will automatically apply labels to the raw data(new dataset)
- Whereas if the labeling model has lower confidence in its results, it will pass the data to humans to do the labeling
- The human-generated labels are then provided back to the labeling(training) model for it to learn from and improve its ability to automatically label the next set of raw data.
- Over time, the model can label more and more data automatically and substantially speed up the creation of training datasets.

# Common Types of Data Labeling

- Computer Vision (image classification, Object detection)
- Natural Language Processing
- Audio processing

Computer Vision - Object detection



DOG, DOG, CAT

Computer Vision - Object detection