

AWS MACHINE LEARNING CERTIFICATION



MODULE #1: DATA ENGINEERING (20% EXAM)



AWS ML CERTIFICATION EXAM DOMAINS



Domain	% of Examination
Domain 1: Data Engineering	20%
Domain 2: Exploratory Data Analysis	24%
Domain 3: Modeling	36%
Domain 4: Machine Learning Implementation and Operations	20%
TOTAL	100%

Source: [https://d1.awsstatic.com/training-and-certification/docs-ml/AWS%20Certified%20Machine%20Learning%20-%20Specialty_Exam%20Guide%20\(1\).pdf](https://d1.awsstatic.com/training-and-certification/docs-ml/AWS%20Certified%20Machine%20Learning%20-%20Specialty_Exam%20Guide%20(1).pdf)



DOMAIN #1 OVERVIEW:

SECTION #1: INTRODUCTION, DATA/ML LINGO, AWS DATA STORAGE

- What is Machine Learning and Artificial Intelligence?
- What is Amazon Web Services (AWS)?
- Artificial Intelligence and Machine learning Lingo (data types, Labeled vs. unlabeled, sagemaker groundtruth)
- structured vs. unstructured and database vs. data lake vs. data storage
- AWS Data Storage (Redshift, RDS, S3, DynamoDB)

SECTION #2: AMAZON S3

- Amazon S3 in Depth (partitions, tags)
- Amazon S3 Storage Tiers and Lifecycles
- Amazon S3 Encryption and Security
- Amazon S3 Encryption and Security – Part #2 (ACL, CloudWatch, CloudTrail, VPC)
- Additional Notes (Elasticsearch, ElastiCache, and Database vs. data warehouse)

DOMAIN #1 OVERVIEW:

SECTION #3: AWS DATA MIGRATION, GLUE, PIPELINE, STEP AND BATCH

- AWS Glue (crawlers, features, built-in transformations etc)
- AWS Data pipeline
- AWS Data Migration Service (DMS)
- AWS Batch
- Step Function

SECTION #4: DATA STREAMING & KINESIS

- Kinesis Overview
- Kinesis Video Streams
- Kinesis Data Streams
- Kinesis Firehose
- Kinesis Analytics and Random Cut Forest

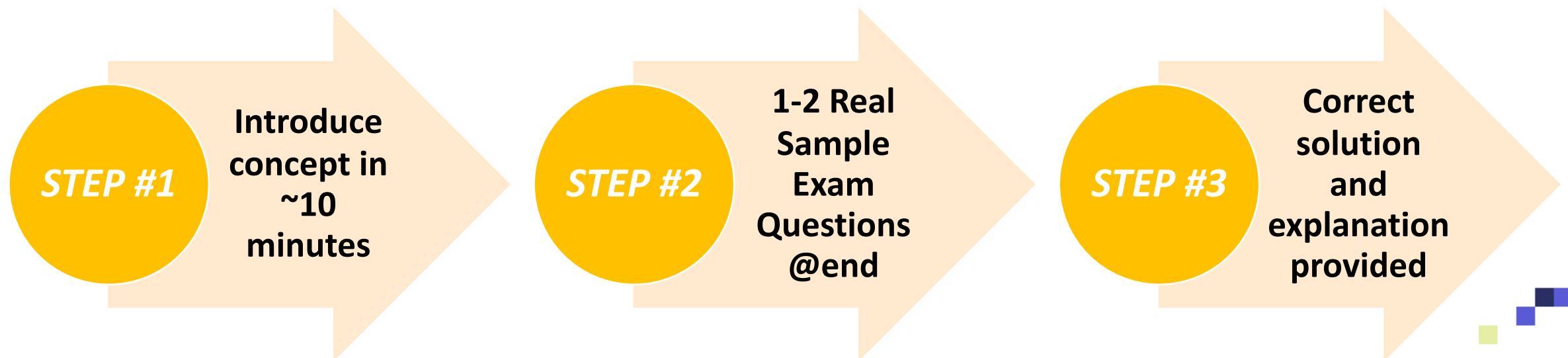
LECTURE DESIGN



- We know how hard it is to study for an exam especially if you have a busy schedule.
- This course is designed to be extremely on point and optimized to pass the exam.

No boring content. Zero unnecessary information.

- Here's the lecture structure that we will follow:



VALUABLE PRIZE!



- For those of you who will successfully complete the entire section and watch the videos till the end, they will receive a valuable prize!

**10 NEW SAMPLE EXAM QUESTIONS + COMPLETE
ANSWER KEY**



GAME AND MINI CHALLENGES!



- Unfortunately, you can't skip the videos.
- You have to collect a code throughout the lectures to unlock the exam.
- Special characters will appear at random moments throughout the video.
- You will need to collect the code and enter it to a website to access the material.
- That's what the final code might look like!

F 2 @ 9 & B



LOOK FOR RED STARS!



- Through the course, important topic will be highlighted with red stars like this!



WHAT IS ARTIFICIAL INTELLIGENCE? AND WHAT IS MACHINE LEARNING? – Part #1



INTRODUCTION



- Artificial Intelligence/Machine learning does not only mean robots or Sci-Fi movies!
- Machine and deep learning applications are everywhere!
- Google search engine, amazon recommender systems, Facebook facial recognition (tagging), Siri.

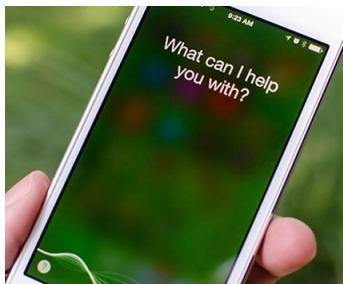
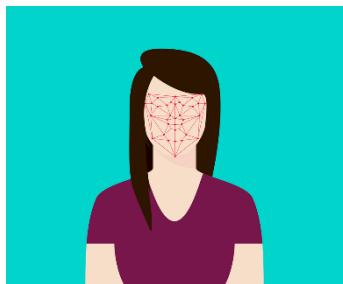


Photo Credit: https://commons.wikimedia.org/wiki/File:Waymo_self-driving_car_front_view.gk.jpg

Photo Credit: <http://blog.etonic.net/index.php?entry=entry110316-081129>

Photo Credit: <https://www.flickr.com/photos/topgold/8325104250>

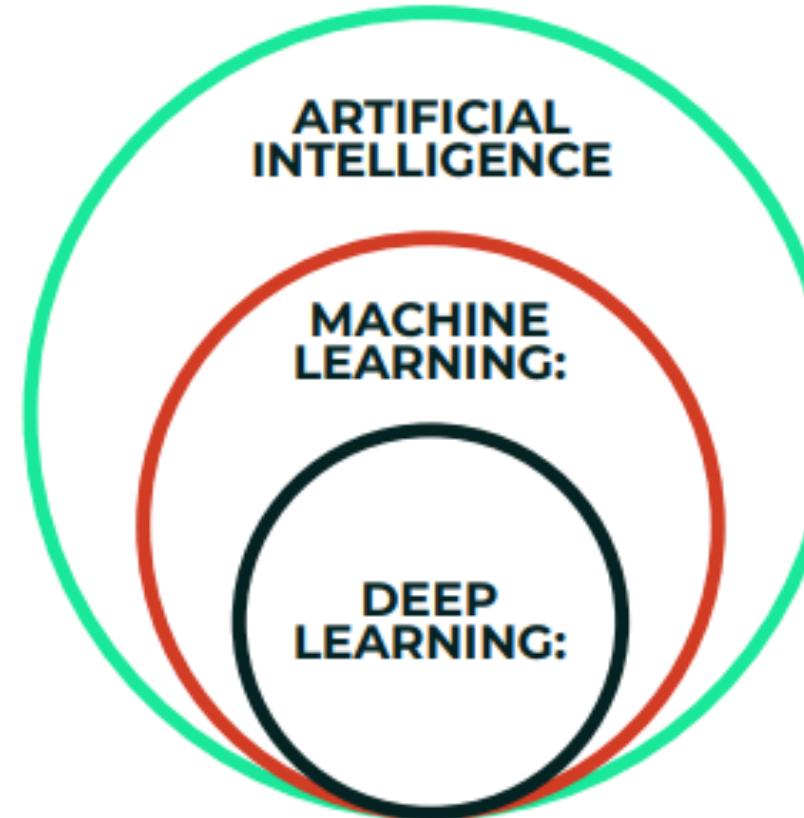
Photo Credit: <https://picryl.com/media/google-search-engine-magnifying-glass-computer-communication-00b825>

Photo Credit: <https://www.flickr.com/photos/iphonedigital/26988770454>

Photo Credit: <https://pixabay.com/illustrations/flat-recognition-facial-face-woman-3252983/>



ARTIFICIAL INTELLIGENCE Vs. MACHINE LEARNING Vs. DEEP LEARNING



1. ARTIFICIAL INTELLIGENCE



- Science that empowers computers to mimic human intelligence such as decision making, text processing, and visual perception.
- AI is a broader field (i.e.: the big umbrella) that contains several subfield such as machine learning, robotics, and computer vision.



Photo Credit: <https://pixabay.com/illustrations/artificial-intelligence-brain-think-4111582/>



2. MACHINE LEARNING



- Machine Learning is a subfield of Artificial Intelligence that enables machines to improve at a given task with experience.
- It is important to note that all machine learning techniques are classified as Artificial Intelligence ones. However, not all Artificial Intelligence could count as Machine Learning since some basic Rule-based engines could be classified as AI but they do not learn from experience therefore they do not belong to the machine learning category.

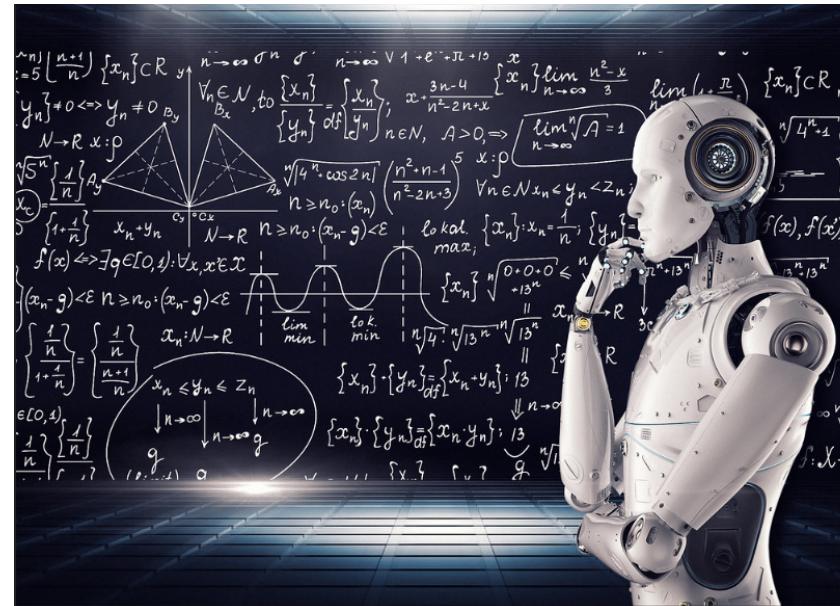


Photo Credit: <https://www.flickr.com/photos/mikemacmarketing/30212411048>



3. DEEP LEARNING

- Deep Learning is a specialized field of Machine Learning that relies on training of Deep Artificial Neural Networks (ANNs) using large dataset such as images.
- ANNs are information processing models inspired by the human brain.
- The human brain consists of billions of neurons that **INPUT DATA** communicate to each other using electrical and chemical signals and enable humans to see, feel, and make decision.

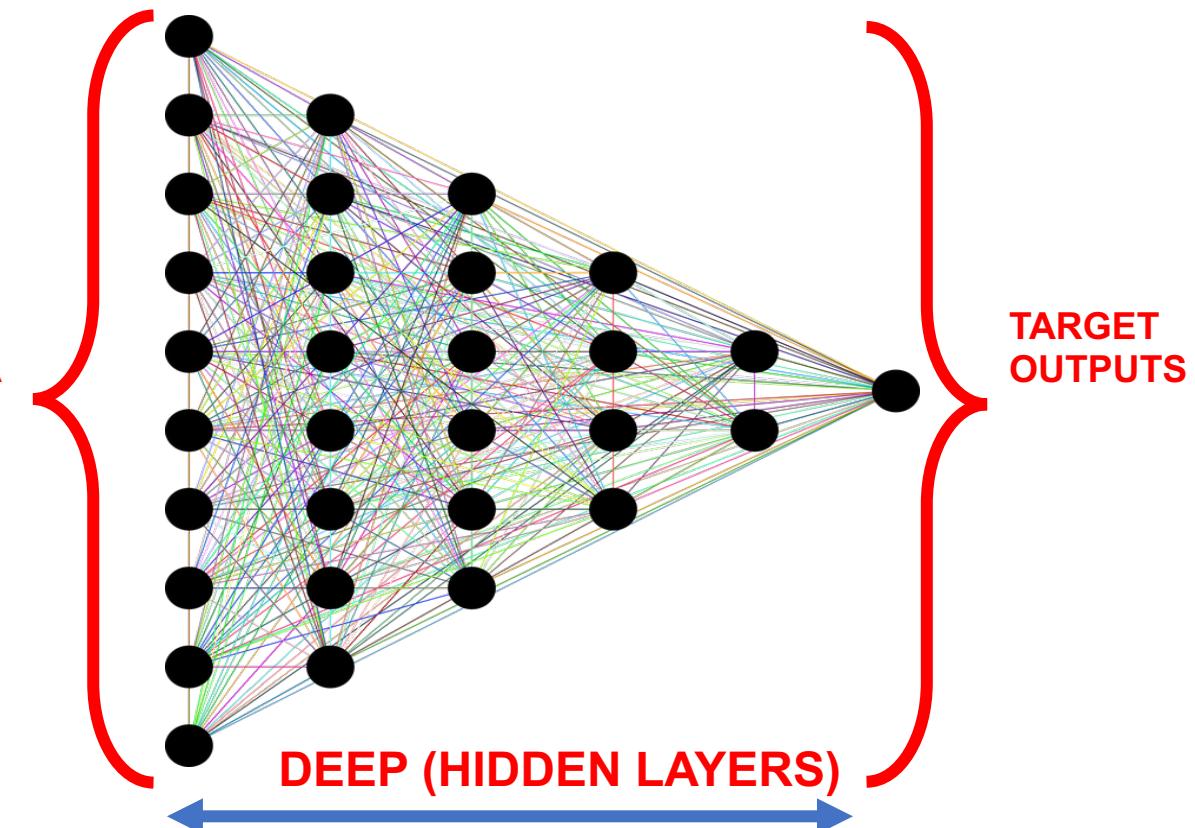
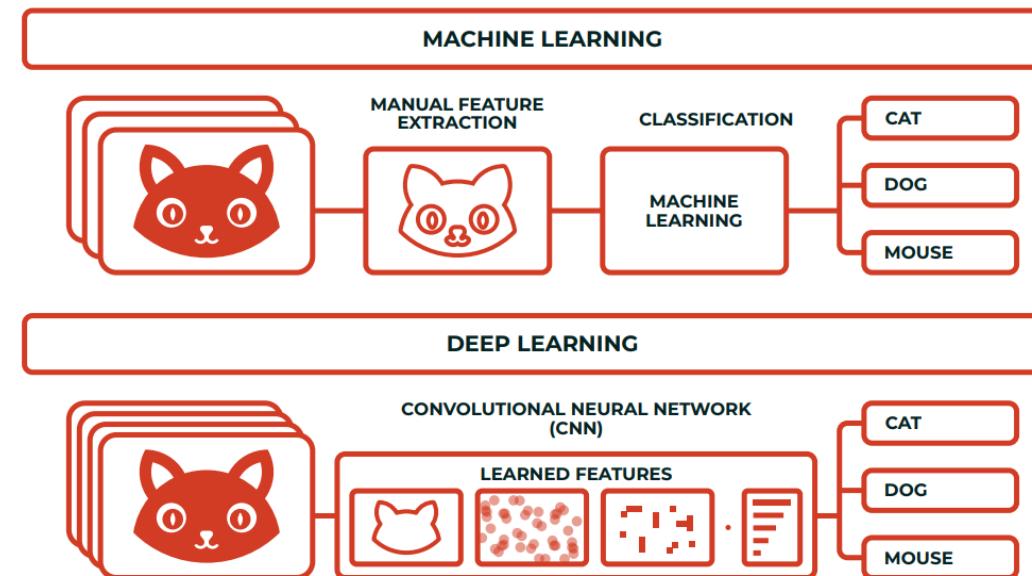


Photo Credit: <https://pixabay.com/en/neural-network-thought-mind-mental-3816319/>

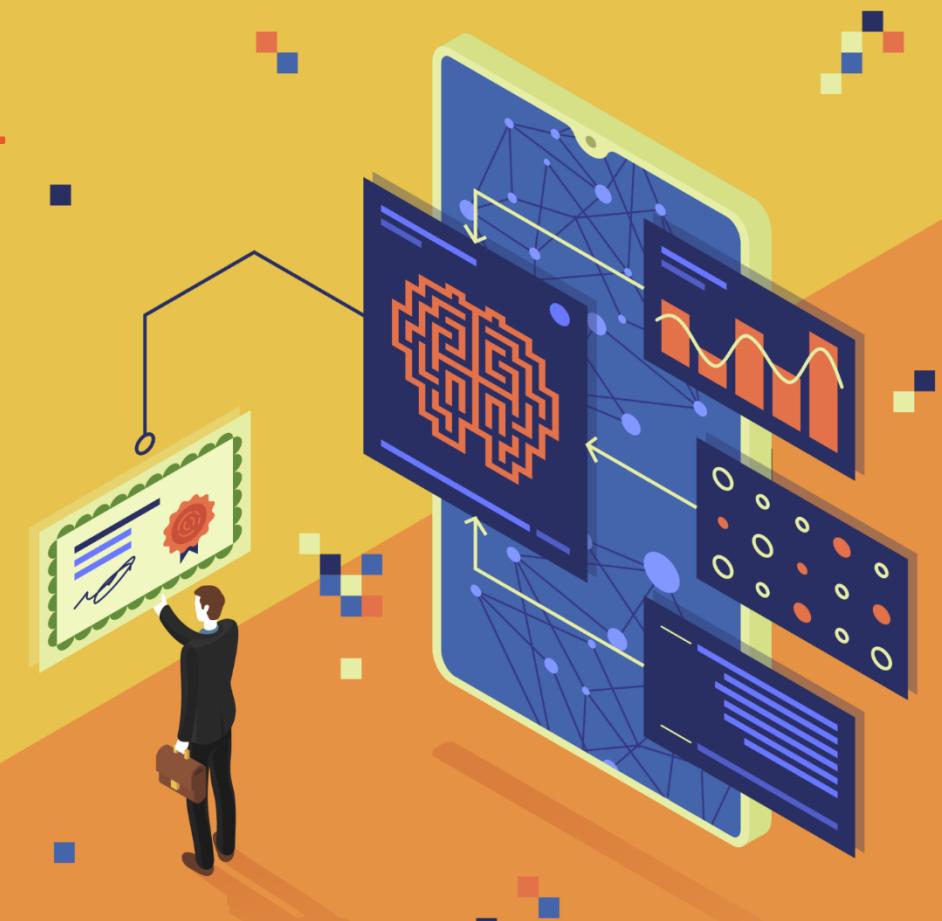
MACHINE VS. DEEP LEARNING



- What differentiates deep learning from machine learning techniques is in their ability to extract features automatically:
 - Machine learning Process: (1) select the model to train, (2) manually perform feature extraction.
 - Deep Learning Process: (1) Select the architecture of the network, (2) features are automatically extracted by feeding in the training data (such as images) along with the target class (label).



WHAT IS ARTIFICIAL INTELLIGENCE? AND WHAT IS MACHINE LEARNING? – Part #2

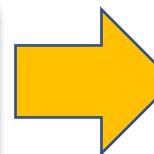


MACHINE LEARNING: BIG PICTURE



ARTIFICIAL INTELLIGENCE
Science that enables computers to mimic human intelligence.
Subfields: Machine Learning, robotics, and computer vision

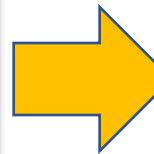
MACHINE LEARNING
Subset of AI that enable machines to improve at tasks with experience



SUPERVISED LEARNING
Training algorithms using labeled input/output data.

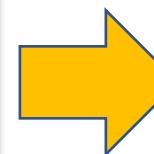


CLASSIFICATION



UNSUPERVISED LEARNING
Training algorithms with no labeled data. It attempts at discovering hidden patterns on its own.

CLUSTERING



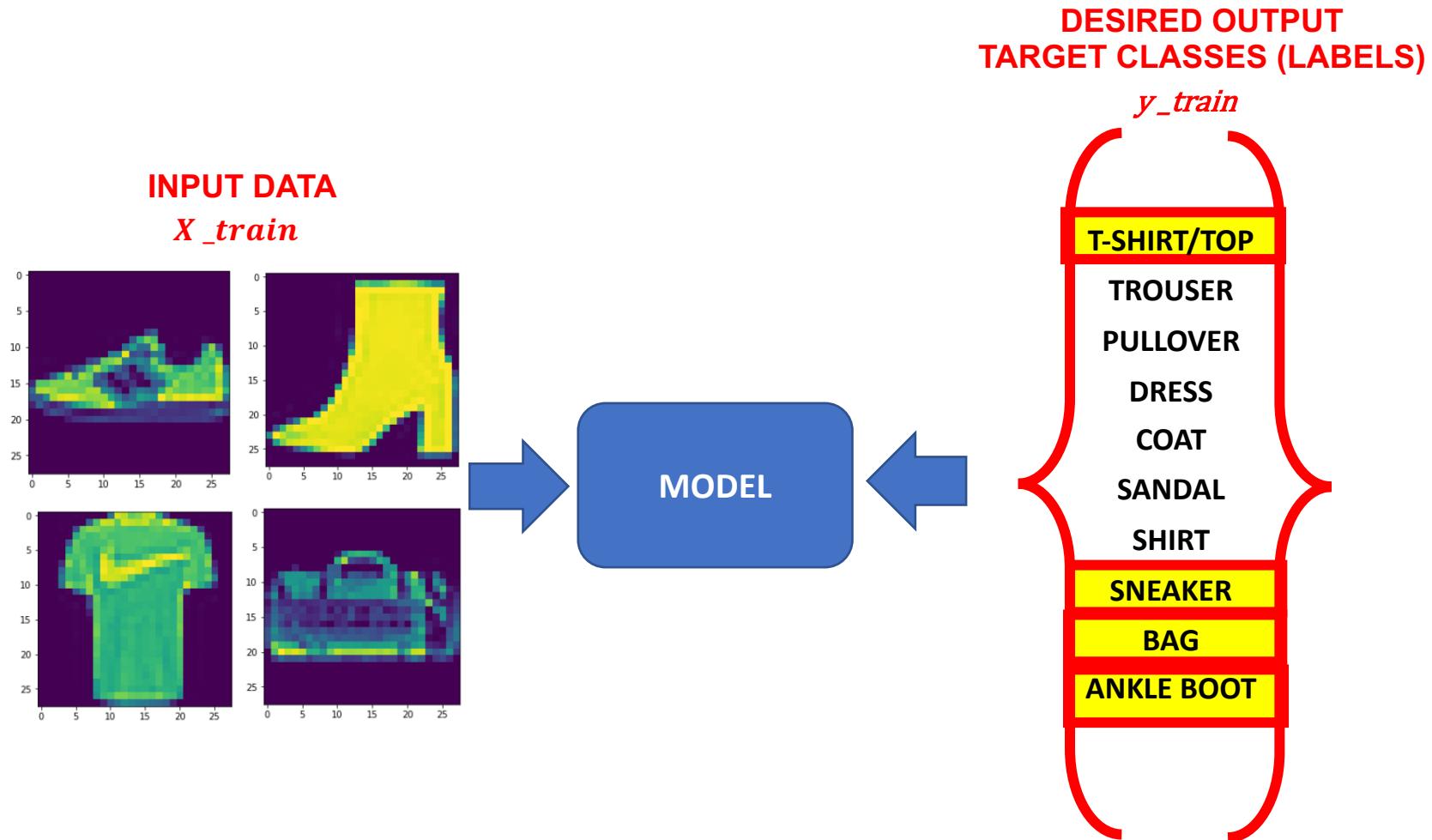
REINFORCEMENT LEARNING
Algorithm take actions to maximize cumulative reward.



MACHINE LEARNING: SUPERVISED LEARNING



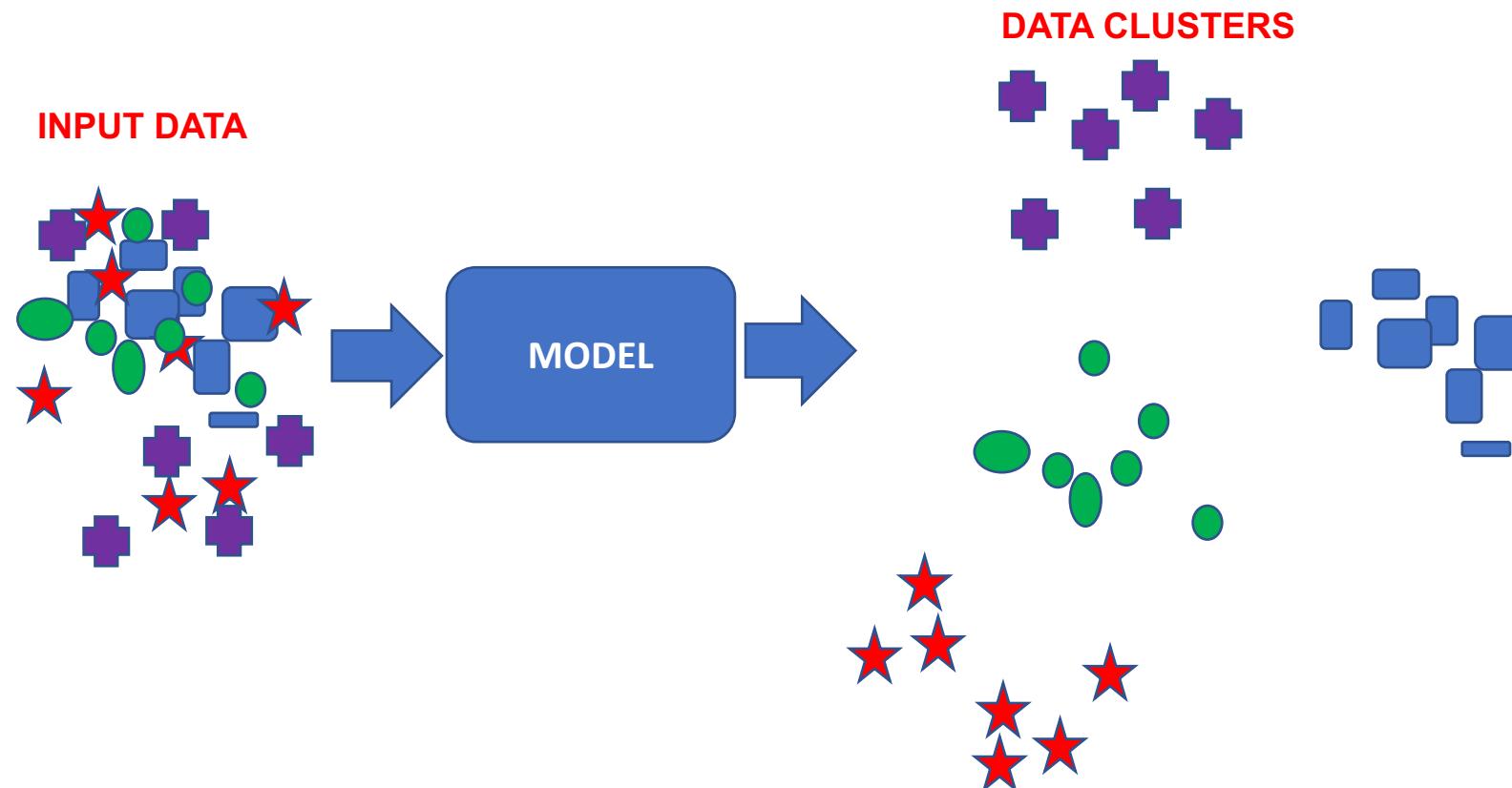
- **Supervised:** used to train algorithms using labeled input and output data.
- Performance is assessed by comparing trained model prediction vs. real output.



MACHINE LEARNING: UNSUPERVISED LEARNING



- **Unsupervised learning:** provides the algorithm with no labeled data.
- The algorithm attempts at discovering hidden patterns within the training data.
- Unsupervised learning methods can analyze complex data that humans might find difficult to interpret.
- No feedback!



MACHINE LEARNING: REINFORCEMENT LEARNING



- Reinforcement learning allows machines take actions to maximize cumulative reward.
- Reinforcement algorithms learn by trial and error through reward and penalty.
- Two elements: environment and learning agent.
- The environment rewards the agent for correct actions.
- Based on the reward or penalty, agent improves its environment knowledge to make better decision.

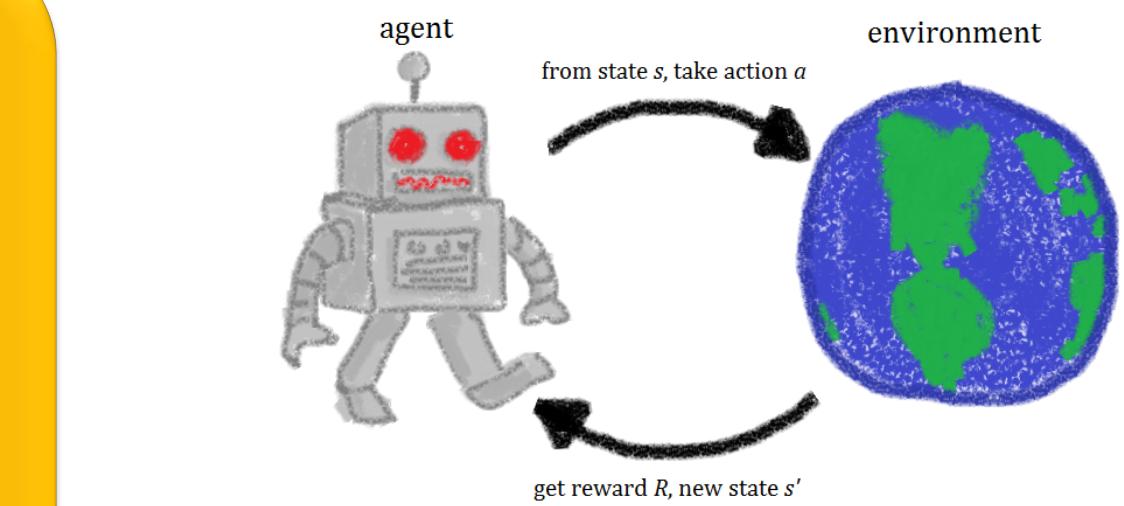
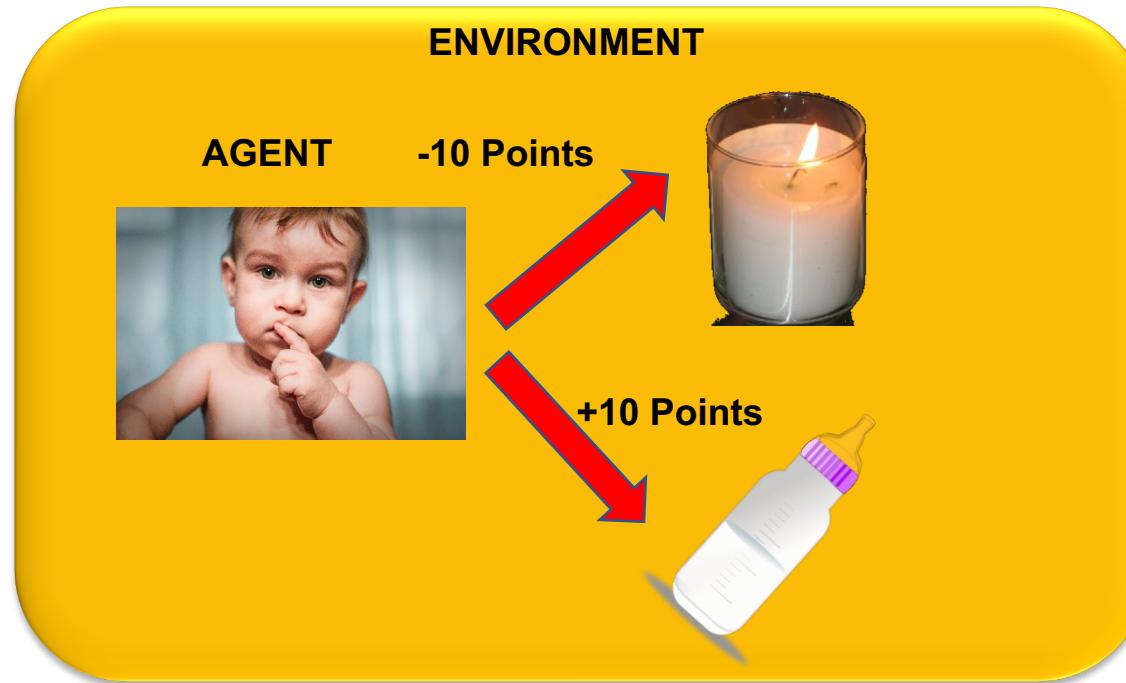


Photo Credit: https://commons.wikimedia.org/wiki/File:RL_agent.png

AMAZON WEB SERVICES (AWS)

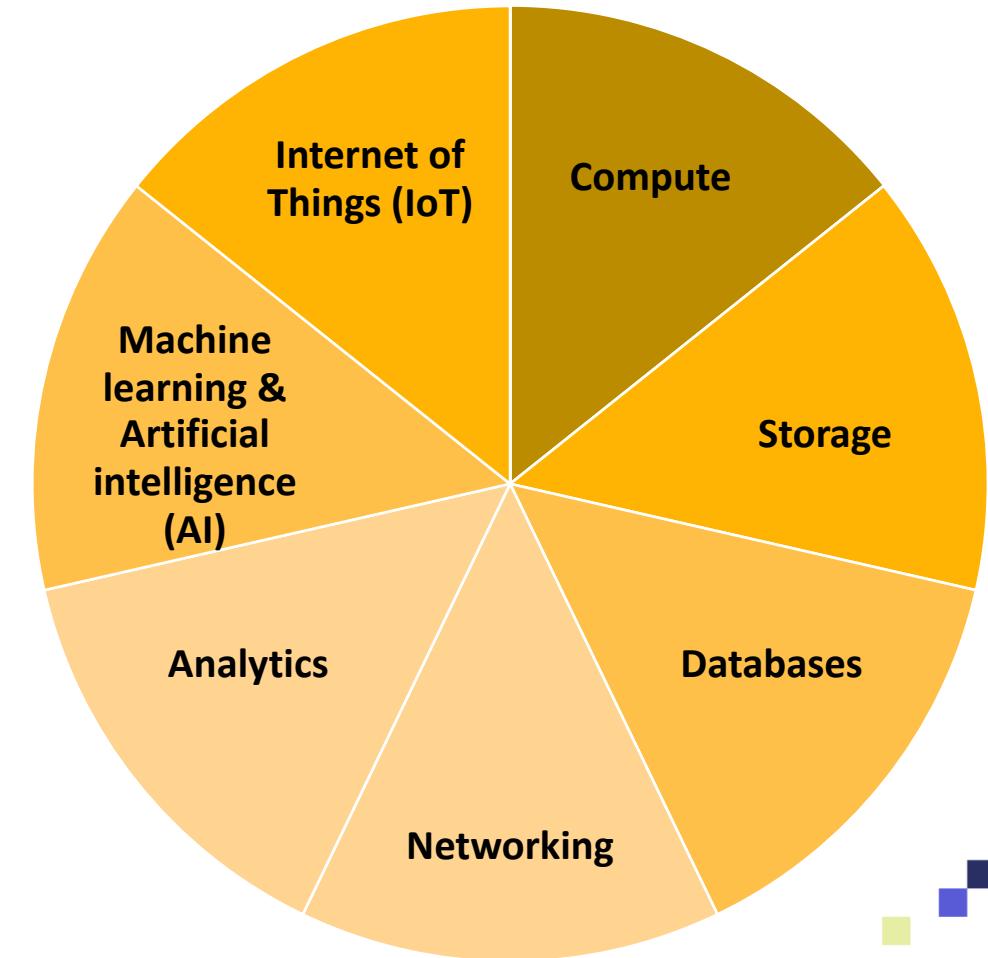


WHAT IS AWS?

- Amazon Web Services (AWS) is the world's top cloud platform.
- AWS offers more than 165 fully featured services (40 of them are not offered anywhere else!).
- AWS is adopted by millions of customers globally including small and large scale enterprises.
- AWS enables companies to be more agile, flexible, secure at a fraction of the cost.
- AWS provides services for broad range of applications such as:



Photo Credit: https://commons.wikimedia.org/wiki/File:AmazonWebservices_Logo.svg



MACHINE LEARNING COMPONENTS?



1. DATA



2. MODEL



3. COMPUTE



AWS SERVICES: STORAGE AND COMPUTE



S RyanAhn

History

Amazon SageMaker

Console Home

Amazon Polly

Amazon Rekognition

Amazon Comprehend

Amazon Translate

Compute

EC2

Lightsail

ECR

ECS

EKS

Lambda

Batch

Elastic Beanstalk

Serverless Application Repository

Storage

S3

EFS

FSx

S3 Glacier

Storage Gateway

AWS Backup

Database

RDS

DynamoDB

ElastiCache

Neptune

Amazon Redshift

Amazon QLDB

Amazon DocumentDB

Analytics

Athena

EMR

CloudSearch

Elasticsearch Service

Kinesis

QuickSight

Data Pipeline

AWS Data Exchange

AWS Glue

AWS Lake Formation

MSK

Customer Enablement

AWS IQ

Support

Managed Services

Blockchain

Amazon Managed Blockchain

Satellite

Ground Station

Management & Governance

AWS Organizations

CloudWatch

AWS Auto Scaling

CloudFormation

CloudTrail

Config

OpsWorks

Service Catalog

Systems Manager

Trusted Advisor

Control Tower

AWS License Manager

AWS Well-Architected Tool

Personal Health Dashboard

AWS Chatbot

Launch Wizard

Business Applications

Alexa for Business

Amazon Chime

WorkMail

End User Computing

WorkSpaces

AppStream 2.0

WorkDocs

WorkLink

Internet Of Things

IoT Core

Amazon FreeRTOS

IoT 1-Click

IoT Analytics

IoT Device Defender

IoT Device Management

IoT Events

IoT Greengrass

IoT SiteWise

IoT Things Graph

Game Development

Amazon GameLift

3

1

This screenshot shows the AWS Services dashboard. A search bar at the top has 's' typed into it. Below the search bar, there's a navigation bar with 'Services' and 'Resource Groups' dropdowns, and a user profile for 'RyanAhn'. The main content area is divided into several sections: 'Compute' (with EC2, Lightsail, ECR, ECS, EKS, Lambda, Batch, Elastic Beanstalk, and Serverless Application Repository), 'Storage' (with S3, EFS, FSx, S3 Glacier, Storage Gateway, and AWS Backup), 'Database' (with RDS, DynamoDB, ElastiCache, Neptune, Amazon Redshift, Amazon QLDB, and Amazon DocumentDB), 'Analytics' (with Athena, EMR, CloudSearch, Elasticsearch Service, Kinesis, QuickSight, Data Pipeline, AWS Data Exchange, AWS Glue, AWS Lake Formation, and MSK), 'Customer Enablement' (with AWS IQ, Support, and Managed Services), 'Blockchain' (with Amazon Managed Blockchain), 'Satellite' (with Ground Station), 'Management & Governance' (with AWS Organizations, CloudWatch, AWS Auto Scaling, CloudFormation, CloudTrail, Config, OpsWorks, Service Catalog, Systems Manager, Trusted Advisor, Control Tower, AWS License Manager, AWS Well-Architected Tool, Personal Health Dashboard, AWS Chatbot, and Launch Wizard), 'Business Applications' (with Alexa for Business, Amazon Chime, and WorkMail), 'End User Computing' (with WorkSpaces, AppStream 2.0, WorkDocs, and WorkLink), 'Internet Of Things' (with IoT Core, Amazon FreeRTOS, IoT 1-Click, IoT Analytics, IoT Device Defender, IoT Device Management, IoT Events, IoT Greengrass, IoT SiteWise, and IoT Things Graph), and 'Game Development' (with Amazon GameLift). Two specific sections are highlighted with red boxes and yellow circles containing numbers: 'Storage' (labeled '1') and 'Compute' (labeled '3').

AWS SERVICES: MACHINE LEARNING



Screenshot of the AWS Services Catalog showing the Machine Learning section highlighted.

The screenshot shows the AWS Services Catalog interface. On the left, there's a sidebar with 'History' and links to various AWS services like S3, Amazon SageMaker, and Amazon Comprehend. The main area has a search bar at the top. Below it, services are categorized into groups:

- Migration & Transfer**: AWS Migration Hub, Application Discovery Service, Database Migration Service, Server Migration Service, AWS Transfer for SFTP, Snowball, DataSync.
- Media Services**: Elastic Transcoder, Kinesis Video Streams, MediaConnect, MediaConvert, MediaLive, MediaPackage, MediaStore, MediaTailor, Elemental Appliances & Software.
- Mobile**: AWS Amplify, Mobile Hub, AWS AppSync, Device Farm.
- AR & VR**: Amazon Sumerian.
- Application Integration**: Step Functions, Amazon EventBridge, Amazon MQ, Simple Notification Service, Simple Queue Service, SWF.
- AWS Cost Management**: AWS Cost Explorer, AWS Budgets, AWS Marketplace Subscriptions.
- Customer Engagement**: Amazon Connect, Pinpoint, Simple Email Service.

A yellow circle with the number '2' is overlaid on the 'Machine Learning' section, which is highlighted with a red rounded rectangle. The 'Machine Learning' section contains the following services:

- Machine Learning**: Amazon SageMaker, Amazon Comprehend, Amazon Forecast, Amazon Lex, Amazon Machine Learning, Amazon Personalize, Amazon Polly, Amazon Rekognition, Amazon Textract, Amazon Transcribe, Amazon Translate, AWS DeepLens, AWS DeepRacer.

- **AI/ML DATA LINGO – LABELED VS.
UNLABELED**



MACHINE LEARNING DATA



- Machine Learning models require data to train.
- There are generally two types of data that we could use to train machine learning models.

UNLABELED DATASET

Unlabeled data consists of data that does not have explanation (class or tag) associated with it.



LABELED DATASET

Labeled data consists of unlabeled data but with a “class” or “tag” associated with it.



LABEL = “CAT”



LABEL = “DOG”



WHERE DOES THIS DATA COME FROM?



- Data can come from so many sources such as images, audio, video, and text.
- Collecting, structuring and analysing this data is critical for companies to gain customers insights and set their marketing and product strategies.

IMAGE/VIDEO



TEXT (CORPUS)



AUDIO/SOUND



TIMESERIES/SIGNALS



Photo Credit: <https://pxhere.com/en/photo/1454351>

Photo Credit: <https://www.flickr.com/photos/29881930@N00/2086641598>

Photo Credit: https://commons.wikimedia.org/wiki/File:Mobile_phone_text_messages.jpg

Photo Credit: https://en.wikipedia.org/wiki/File:Messages_Yosemite.svg

Photo Credit: <https://www.pexels.com/photo/blue-and-yellow-graph-on-stock-market-monitor-159888/>



GOOD Vs. BAD DATA



GOOD DATA

- Many samples (large number of data points)
- Not Biased
- Does not contain missing data points
- Only contains (relevant) important features
- Does not contain duplicate samples

BAD DATA

- Few samples (small number of data points)
- Biased
- Contains missing data points
- Contains many irrelevant (useless) features
- Contains duplicate samples



WHERE DOES THIS DATA COME FROM?



- Data could also come from multiple sources such as Kaggle, UCI, AWS Dataset, and ImageNet.
- ImageNet is an open source repository of images consisting of 21,841 subcategories (classes) and over 14 million images.

The image shows three side-by-side screenshots of data repositories:

- Kaggle Datasets:** A search interface showing results for "UFC Fight historical data from 1993 to 2019".
- UCI Machine Learning Repository:** A search interface showing results for "UFC Fight historical data from 1993 to 2019".
- Registry of Open Data on AWS:** A search interface showing results for "UFC Fight historical data from 1993 to 2019".

Check out website here: <https://archive.ics.uci.edu/ml/datasets.php>
Check out website here: <https://www.kaggle.com/datasets>

HOW TO OBTAIN LABELED DATA USING AWS? SAGEMAKER GROUNDTRUTH



- AWS SageMaker GroundTruth is a service offered by AWS to label data.
- In machine learning terminology, Ground truth means “gold standard”!
- Ground Truth indicates the “true” or “real” class that you would like your model to learn how to predict.



AMAZON
SAGEMAKER
GROUNDTRUTH

The screenshot shows the AWS SageMaker GroundTruth interface. At the top, there's a navigation bar with the AWS logo, 'Services', and 'Resource Groups'. Below it, a dropdown menu is set to 'Image'. A section titled 'Task category' says 'Select the type of data being labeled to view available task templates for it or select 'Custom' to create your own.' Under 'Task selection', there are four options:

- Image classification**: Selected. Description: Get workers to categorize images into specific classes. Info. Options: Basketball (image of two people playing basketball), Soccer (image of a soccer player).
- Bounding box**: Description: Get workers to draw bounding boxes around specified objects in your images. Info. Options:
- Semantic segmentation**: Description: Get workers to draw pixel level labels around specific objects and segments in your images. Info. Options:
- Label verification**: Description: Get workers to verify existing labels in your dataset. Info. Options: Correct label (image of a green car with a blue border), Incorrect label (image of a green car with a blue border).



HOW TO OBTAIN LABELED DATA USING AWS? SAGEMAKER GROUNDTRUTH



The screenshot shows the Amazon SageMaker console with the 'Ground Truth' section highlighted in the sidebar. A red box and arrow point to the 'Labeling jobs' option. The main content area displays the 'How it works' diagram, which illustrates the five-step machine learning pipeline: Label, Build, Train, Tune, and Deploy. Each step is accompanied by a brief description and a corresponding icon.

Amazon SageMaker
Build, train, and deploy machine learning models at scale
The quickest and easiest way to get ML models from idea to production.

How it works

- Label**
Set up and manage labeling jobs for highly accurate training datasets within Amazon SageMaker, using active learning and human labeling
- Build**
Connect to other AWS services and transform data in Amazon SageMaker notebooks
- Train**
Use Amazon SageMaker's algorithms and frameworks, or bring your own, for distributed training
- Tune**
Amazon SageMaker automatically tunes your model by adjusting multiple combinations of algorithm parameters
- Deploy**
Once training is completed, models can be deployed to Amazon SageMaker endpoints, for real-time predictions

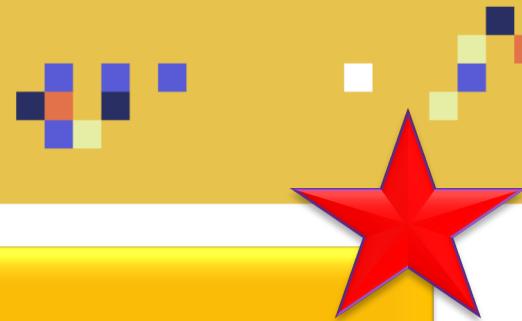
YOU CAN CREATE LABELING JOBS USING AMAZON GROUND TRUTH SERVICE

<https://aws.amazon.com/sagemaker/groundtruth/>

- AI/ML DATA LINGO – DATA TYPES



DATA TYPES



1. QUANTITATIVE (NUMERICAL)

- Numerical data also known as quantitative data represents a measurement or count.
- Examples: weight, blood pressure, and dollars count.
- Either discrete or continuous

2. QUALITATIVE (CATEGORICAL)

- Categorical (Qualitative) data represents data that could be divided into groups.
- Examples: race, sex, age group, and educational level.
- Categorical data can take numerical values but they do not have mathematical meaning so you can't multiply them together for example.

3. ORDINAL

- Ordinal data represents a mix between numerical and categorical data.
- Example: course ratings on Udemy!
- Data consists of categories such as numbers between 1 and 5, in which:
 - 1 star means poor quality course
 - 5 star means great quality course

1. QUANTITATIVE (NUMERICAL) DATA



- Numerical data also known as quantitative data represents a measurement or count.
- Examples: weight, blood pressure, and dollars count.
- Numerical data consists of two types as follows: (1) discrete and (2) continuous.

DISCRETE

- Includes data that are distinct and separable.
- Discrete data could be counted as integers.
- Examples: How many cats do you have? How many products sold?

CONTINUOUS

- Represent measurements that are uncountably infinite
- Continuous data can only be described using intervals on the real number line.
- Examples: How much gas did you put in a car? Values could be anywhere between 0 gallons to 20 gallons. There are infinite possibilities; 8.40 gallons, or 8.41, or 8.414863 gallons.



2. QUALITATIVE (CATEGORICAL) DATA



- Categorical data represents data that could be divided into groups.
- Examples: race, sex, age group, and educational level.
- Categorical data can take numerical values but they do not have mathematical meaning so you can't multiply them together for example (and order does not mean anything).
 - **Example:** 1 = single, 2 = married
- Binary data is a type of qualitative data that consists of only two values such as:
 - **Example:** 1 or 0 ON or OFF True or False
- Binary data is very common in machine learning classification algorithms in which the outcome could be one of two options: healthy or sick, malignant or benign.



SUNNY = 1



CLOUDY = 2



RAINY = 3

Photo Credit: <https://publicdomainvectors.org/en/free-clipart/Vector-graphics-of-weather-forecast-color-symbol-for-brightly-sunny-sky/18972.html>

Photo Credit: <https://pixabay.com/vectors/clouds-weather-rain-drops-308682/>

Photo Credit: <https://publicdomainvectors.org/en/free-clipart/Vector-drawing-of-weather-forecast-color-symbol-for-sunny-to-cloudy-sky/18971.html>



3. ORDINAL DATA

- Ordinal data represents a mix between numerical and categorical data.
- Example: course ratings on Udemy!
- Data consists of categories such as numbers between 1 and 5, in which:
 - 1 star means poor quality course
 - 5 star means great quality course
- The numbers in each category have mathematical meaning.
- This what differentiates ordinal data from categorical data.
- For example, if you take the average of the 1000 reviews on Udemy per course, you will end up with an answer that have a meaning.
- This does not work if you have categorical data, you cannot average single and married and get meaningful results.



THERE IS ONE MORE DATA TYPE: USELESS DATA!



- Useless data is a type of data that is discrete and has no relationship whatsoever with the output.
- We usually drop the useless features from the dataset before training the model
- Example include: random customer IDs at a store or a bank

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	Nan	S
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Heikkinen, Miss. Laina	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1		female	26.0	0	0	STON/O2. 3101282	7.9250	Nan	S
3	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	Nan	S

USELESS
INFORMATION



DATABASE Vs. DATA LAKE Vs. DATA WAREHOUSE



STRUCTURED Vs. UNSTRUCTURED DATA



- In order to understand the difference between database, data warehouse and data lake, we need to cover the difference between structured and unstructured data.

STRUCTURED

Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
Cumings, Mrs. John Bradley (Florence Briggs Th... Heikkinen, Miss. Laina	female	38.0	1	0	PC 17599 STON/O2. 3101282	71.2833 7.9250	C85 NaN	C S
Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	26.0	0	0	313803	53.1000	C123	S
Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

UNSTRUCTURED

- Unstructured data has no schema or structure.



Photo Credit: <https://pixabay.com/vectors/database-schema-data-tables-schema-1895779/>
<https://pixabay.com/vectors/newspaper-article-journal-headlines-154444/>
<https://www.flickr.com/photos/lolololori/2343992441>

DATABASE



- Databases are typically structured with a defined schema.
- Items are organized as a set of tables with columns and rows.
- Columns include attributes and rows indicate an object or entity.
- Database is designed to be transactional and they are not designed to perform data analytics.



AMAZON
AURORA



MySQL®



PostgreSQL



ORACLE®

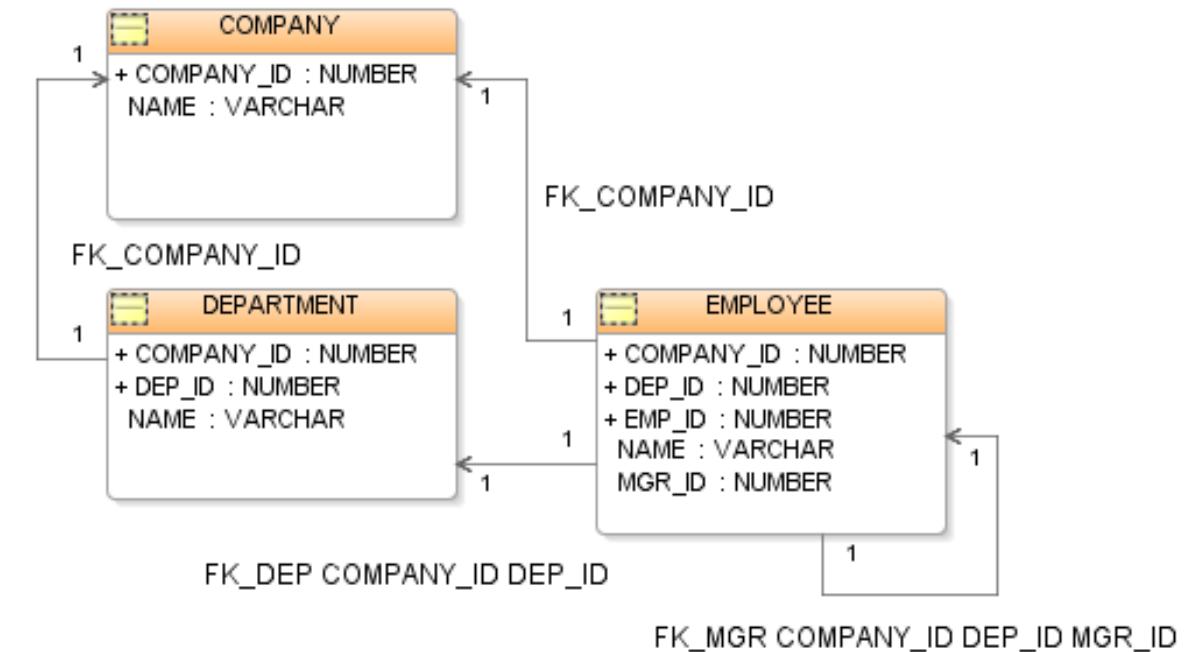
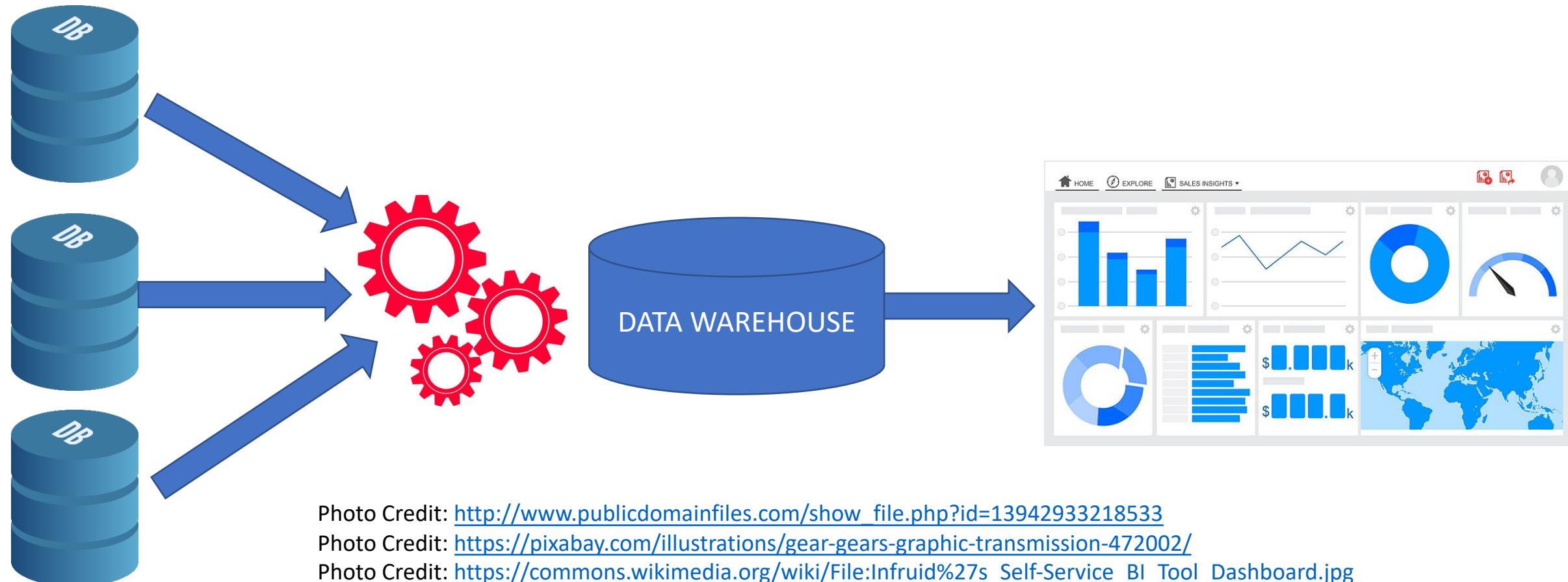


Photo Credit: <https://commons.wikimedia.org/wiki/File:Cascaded-keys.PNG>

DATA WAREHOUSE



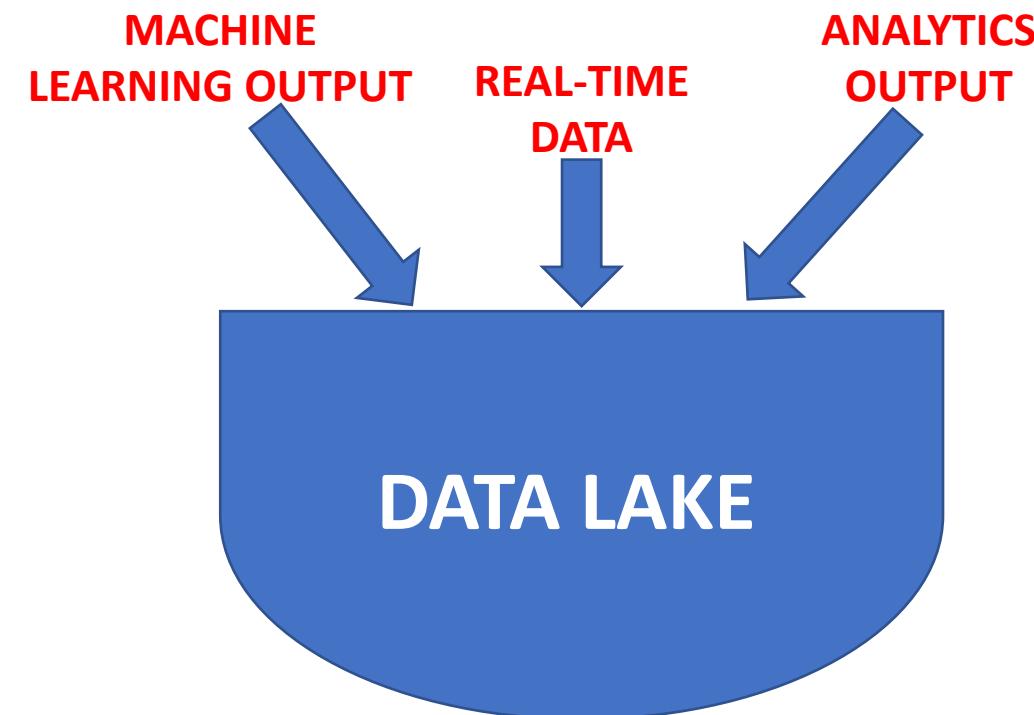
- A data warehouse exists on top of several databases and used for business intelligence.
- Data warehouse consumes data from all these databases and creates a layer optimized to perform data analytics.
- Schema is done on import.



DATA LAKE



- A data lake is a centralized repository for structured and unstructured data storage.
- Data lakes could be used to store raw data as is without any structure (schema).
- There is no need to perform any ETL or transformation jobs on it.
- You can store many types of data such images, text, files, videos.
- You can store machine learning models artifacts, retime data, and analytics outputs in data lakes.
- Processing could be done on export so schema is defined on read.



AWS KEY STORAGE SERVICES



AWS KEY STORAGE TYPES



S

History

Amazon SageMaker

Console Home

Amazon Polly

Amazon Rekognition

Amazon Comprehend

Amazon Translate

Compute

- EC2
- Lightsail
- ECR
- ECS
- EKS
- Lambda
- Batch
- Elastic Beanstalk
- Serverless Application Repository

Analytics

- Athena
- EMR
- CloudSearch
- Elasticsearch Service
- Kinesis
- QuickSight
- Data Pipeline
- AWS Data Exchange
- AWS Glue
- AWS Lake Formation
- MSK

Customer Enablement

- AWS IQ
- Support
- Managed Services

Business Applications

- Alexa for Business
- Amazon Chime
- WorkMail

Blockchain

- Amazon Managed Blockchain

Satellite

- Ground Station

End User Computing

- WorkSpaces
- AppStream 2.0
- WorkDocs
- WorkLink

Internet Of Things

- IoT Core
- Amazon FreeRTOS
- IoT 1-Click
- IoT Analytics
- IoT Device Defender
- IoT Device Management
- IoT Events
- IoT Greengrass
- IoT SiteWise
- IoT Things Graph

Management & Governance

- AWS Organizations
- CloudWatch
- AWS Auto Scaling
- CloudFormation
- CloudTrail
- Config
- OpsWorks
- Service Catalog
- Systems Manager
- Trusted Advisor
- Control Tower
- AWS License Manager
- AWS Well-Architected Tool
- Personal Health Dashboard
- AWS Chatbot
- Launch Wizard

Game Development

- Amazon GameLift

Storage

- S3
- EFS
- FSx
- S3 Glacier
- Storage Gateway
- AWS Backup

Database

- RDS
- DynamoDB
- ElastiCache
- Neptune
- Amazon Redshift
- Amazon QLDB
- Amazon DocumentDB

AWS KEY STORAGE TYPES



1. AMAZON S3

2. AURORA RDS

3. REDSHIFT

4. DYNAMODB



1. AMAZON S3



- Amazon Simple Storage Service (Amazon S3) is a storage service that allows enterprises/individuals to store and protect any amount of data.
- Amazon S3 is extremely easy to use and allows enterprises to organize their data and configure finely-tuned access controls.
- Amazon S3 extremely durable to 99.99999999% (11 9's).



AMAZON S3

Photo Credit: https://commons.wikimedia.org/wiki/File:AWS_Simple_Icons_Storage_Amazon_S3.svg



1. AMAZON S3



**CREATE A
BUCKET AND
SIMPLY UPLOAD
DATA TO IT**

The screenshot shows the AWS S3 console with a red arrow pointing to the '+ Create bucket' button. The main content area displays three steps: 'Create a new bucket', 'Upload your data', and 'Set up your permissions'. Each step has a corresponding icon and a 'Learn more' link.

S3 buckets

Create a new bucket

Buckets are globally unique containers for everything that you store in Amazon S3.

[Learn more](#)

Upload your data

After you create a bucket, you can upload your objects (for example, your photo or video files).

[Learn more](#)

Set up your permissions

By default, the permissions on an object are private, but you can set up access control policies to grant permissions to others.

[Learn more](#)

Get started

2. RDS AURORA



- Amazon Aurora is a fully managed by Amazon Relational Database (RDS) service.
- Transactional style database.
- Amazon Aurora is a MySQL and PostgreSQL-compatible relational database.
- You do not have to deal with administration tasks such as hardware provisioning, creating backups and database setup.
- It features continuous backup to Amazon S3, and replication across three Availability Zones.
- Many engines available to create database



AMAZON RDS AURORA

- [Watch Video: https://aws.amazon.com/rds/aurora/](https://aws.amazon.com/rds/aurora/)

2. RDS AURORA



MANY ENGINE OPTIONS ARE
AVAILABLE TO CREATE
DATABASE

The screenshot shows the 'Create database' page in the AWS RDS console. At the top, there's a message: 'We listened to your feedback! Now, create a database with a single click using our pre-built configurations! Or choose your own configurations. [Switch to your original interface.](#)' Below this, the breadcrumb navigation shows 'RDS > Create database'. The main section is titled 'Create database' and contains a 'Choose a database creation method' section with two options: 'Standard Create' (selected) and 'Easy Create'. The 'Standard Create' option includes a note: 'You set all of the configuration options, including ones for availability, security, backups, and maintenance.' The 'Easy Create' option includes a note: 'Use recommended best-practice configurations. Some configuration options can be changed after the database is created.' A large red arrow points from the text 'MANY ENGINE OPTIONS ARE AVAILABLE TO CREATE DATABASE' to the 'Engine options' section. This section is titled 'Engine type' and shows several database engine options with their icons:

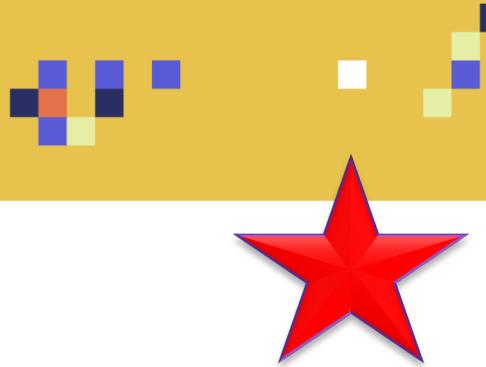
- Amazon Aurora (selected)
- MySQL
- MariaDB
- PostgreSQL
- Oracle
- Microsoft SQL Server

Below the engine type section, there's an 'Edition' section with two options:

- Amazon Aurora with MySQL compatibility (selected)
- Amazon Aurora with PostgreSQL compatibility

3. REDSHIFT

- Amazon Redshift is the fastest cloud **data warehousing** service that could be used to perform business analytics.
- Extremely fast and optimized performance since it relies on columnar storage and data compression.
- Queries are run against data stored in redshift storage or against data stored in S3.
- Redshift uses a unique data warehousing architecture that relies on Massively Parallel Processing (MPP).
- MPP parallelize and distribute SQL operations.
- Redshift uses machine learning to optimize performance.



AMAZON REDSHIFT

3. REDSHIFT



- Redshift is used on top of several databases and used for business intelligence.
- It consumes data from many sources and creates a layer optimized to perform data analytics.

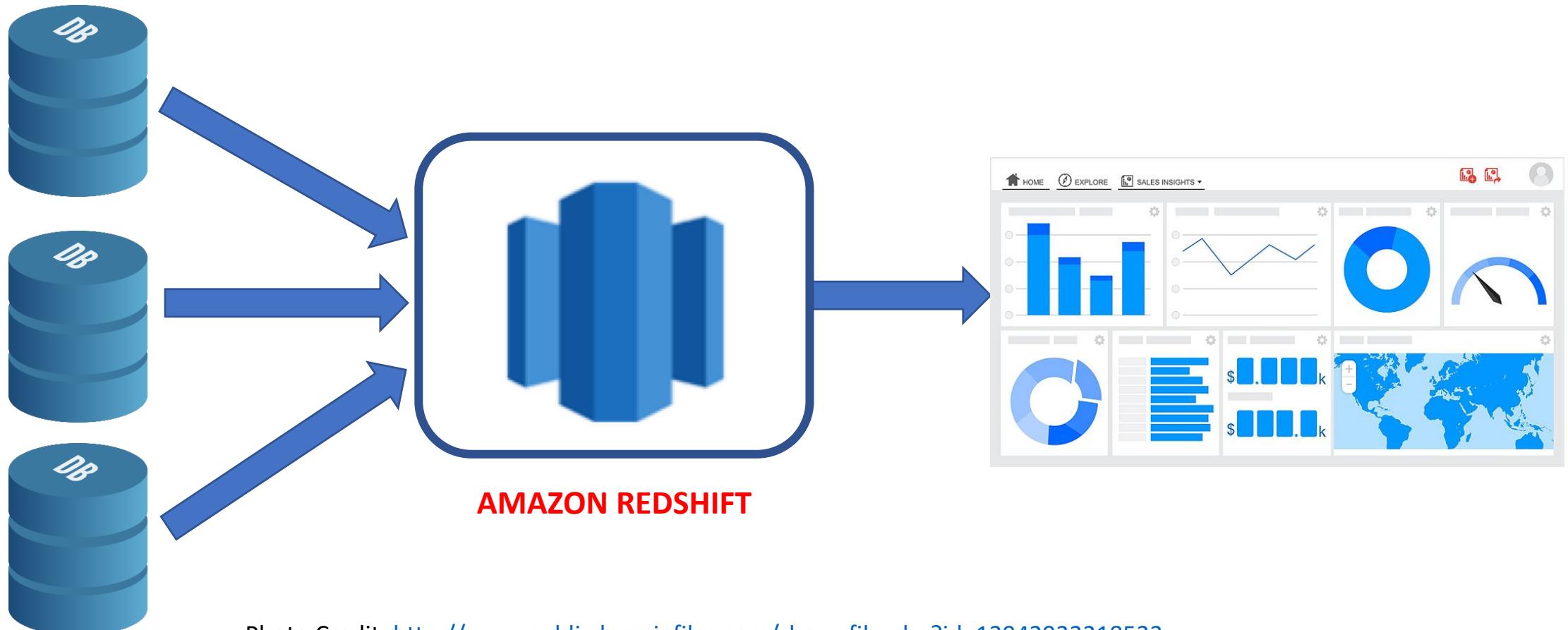


Photo Credit: http://www.publicdomainfiles.com/show_file.php?id=13942933218533

Photo Credit: https://commons.wikimedia.org/wiki/File:Infruid%27s_Self-Service_BI_Tool_Dashboard.jpg

3. REDSHIFT



Switch to the new Amazon Redshift console.
We have been listening to your feedback to make several improvements. Try it out and tell us what you think!

Getting Started

- Getting Started with Amazon Redshift
- Overview and features
- Free Trial
- Evaluation and POC support
- Documentation
- Query your S3 data lake with Redshift Spe
- Pricing and Specs
- Purchase a Reserved Node

AWS Marketplace

- Matillion ETL for Amazon Redshift
 - By Matillion
 - Rating ★★★★☆
 - \$1.37 to \$5.48/hr for software + AWS usage
 - [View all Data Integration](#)
- TIBCO Spotfire® Analytics for AWS (Hourly)
 - By TIBCO Software Inc.
 - Rating ★★★★☆
 - Starting from \$1.20/hr or from \$8,400/yr (24hrs)
 - AWS usage fees
 - [View all Infrastructure Software](#)
- Yellowfin 8.0.3 for AWS (12 Month, 3 User)
 - By Yellowfin
 - Rating ★★★★☆
 - Bring Your Own License + AWS usage fees
 - [View all Infrastructure Software](#)

LAUNCH REDSHIFT CLUSTERS

Clusters Query editor Saved queries Snapshots Security Workload management Reserved nodes Advisor Events Connect count What's new

Launch cluster [Learn more](#)

With a few clicks, you can create your first Amazon Redshift cluster in minutes.

Quick launch cluster

Query Editor [Learn more](#)

Amazon Redshift console now supports writing, running, and saving queries.

Open Query Editor

Find the best cluster configuration for your needs [Learn more](#)

What is your uncompressed data size? [GB](#) [TB](#) [PB](#)

0 250 500 750 1000 20 GB

dc2.large
Best throughput at the lowest cost

On-demand: \$0.50/hour
Reserved (1 yr): \$0.32/hour
Reserved (3 yr): \$0.18/hour

Nodes: 2 Compressed storage: 0.32 TB

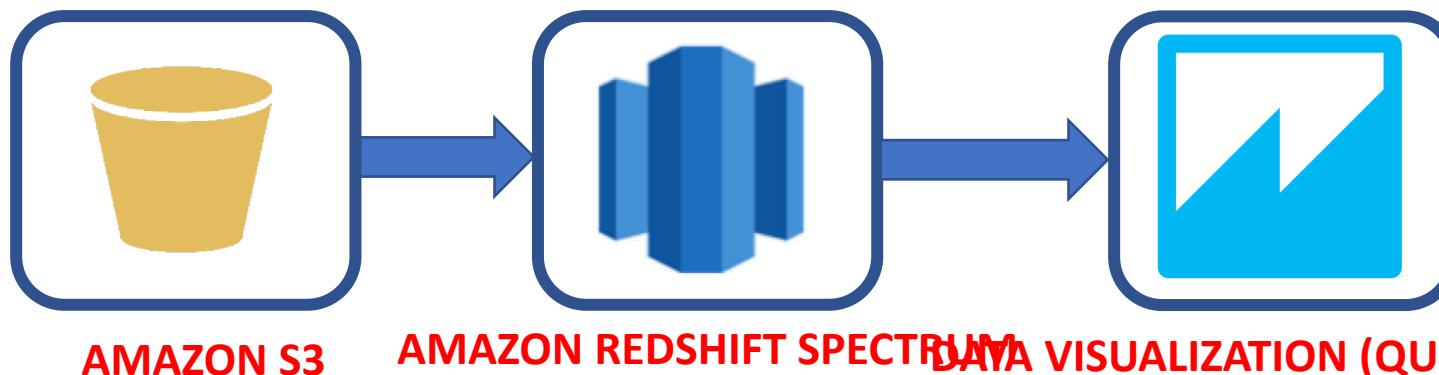
Launch this cluster

If you're doing a Proof-of-Concept on Redshift, [follow this guide](#) or reach the Amazon Redshift team for help.

3. AMAZON REDSHIFT SPECTRUM



- AWS Amazon Redshift Spectrum allows analysts to run SQL queries on data stored in **Amazon S3 buckets directly**.
- Redshift can dramatically save time because it does not require transferring data from S3 to a database.
- Redshift Spectrum can work well with unstructured S3 data lakes.



AMAZON REDSHIFT SPECTRUM



4. AMAZON DYNAMODB

- Amazon DynamoDB is a fully managed NoSQL key-value and document database (not relational database so no schema is required).
- DynamoDB is extremely scalable with minimum latency:
 - 10 trillion requests/day
 - 20 million requests/second
- Create a new table for your application and let DynamoDB handle the rest.
- It works great for storing machine learning models for inference by application
- Watch Video: <https://aws.amazon.com/dynamodb/>



AMAZON DYNAMODB

4. AMAZON DYNAMODB

```
{  
  "Customers": [  
    {  
      "ID": 1,  
      "Name": Ryan,  
      "Age": 28,  
    },  
    {  
      "ID": 2,  
      "Name": Mitch,  
      "Age": 21,  
    }  
  ]  
}  
  
KEY:VALUE
```

