

◆ Member-only story

Day 1 of 30 days of Data Engineering

With examples and projects...



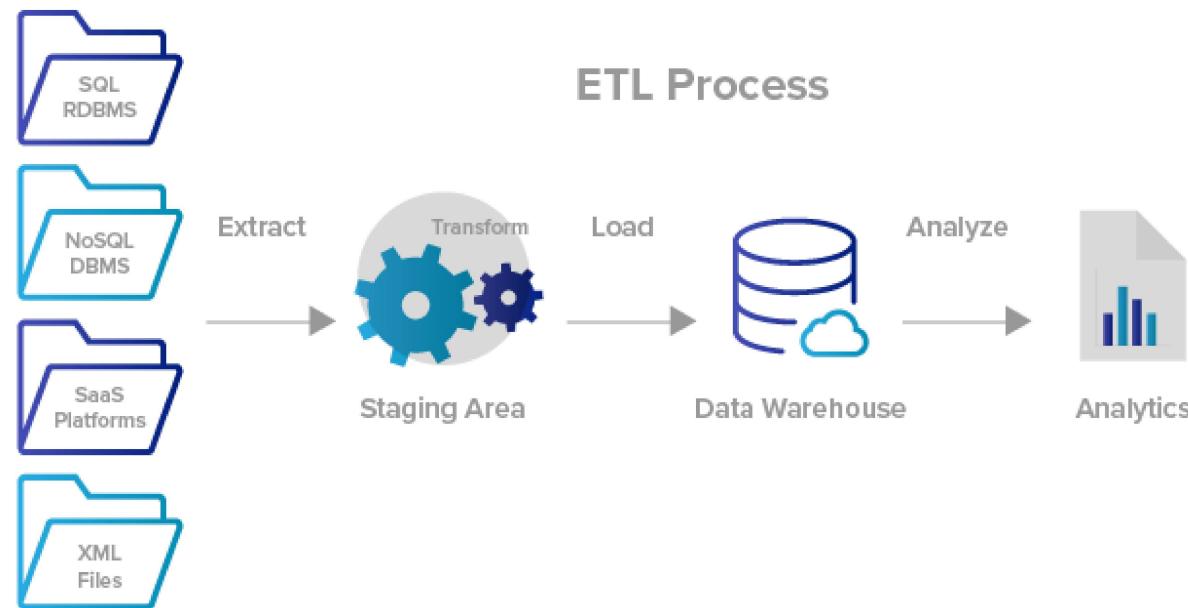
Naina Chaturvedi · [Follow](#)

Published in Coders Mojo · 17 min read · Aug 15, 2022

232



...



Pic credits : infa

Welcome back peeps! Hope all is going well. So, after receiving a great response (and some really good feedback and inputs) for 60 days of Data Science and ML with projects series, I'm excited to share that I'm starting a new Series – **30 days of Data Engineering with (amazing) projects**. PS: I'll be writing as and when I'm free out of my busy work schedule.

What's Covered in 30 days of Data Engineering with Projects Series till now —

[Day 1 : What's Data Engineering, Why Data Engineering, Data Engineers – ML Engineers – Data Scientists, Purpose and Scope](#)

[Day 2 : Complete Python for Data Engineering – Part 1](#)

[Day 3 : Complete Advanced Python for Data Engineering – Part 2](#)

[Day 4: Techniques to write efficient and Optimized Code](#)

[Day 5 : SQL](#)

[Day 6 : Advanced SQL](#)

[Day 7 : BigQuery and SQL vs NOSQL databases](#)

[Day 8 : Advanced Functions](#)

[Day 9 : Query Optimizations](#)

Day 10 : MySQL and PostgreSQL

Day 11: Shell scripting and Linux “touch” command

Day 12 : Map Reduce, Data Warehouse, Data Lakes

Day 13: Pandas, Pandas, Data Cleaning and processing, Outlier Detection, Noisy Data, Missing Data, Pandas Functions, Aggregate Functions, Joins

Day 14 : Numpy

Day 15 : Advanced Pandas Techniques

Day 16 : Data Pre-processing, Handling missing values, Data Cleaning, Mean/mode/median Imputation, Hot Deck Imputation, Rescale Data, Binarize Data, Regression Imputation, Stochastic regression imputation, Feature Scaling

Day 17 : Data Augmentation, Read and Process Large Datasets

Day 18 : Data Visualization basics, Data Visualization Projects, Data Visualization using Plotly and Bokeh, Data Profiling, Summary Functions,

[Indexing](#), [Grouping](#), [Linear Regression](#), [Multi Linear Regression](#),
[Polynomial Regression](#), [Regression](#), [Support Vector Regression](#), [Decision Tree Regression](#), [Random Forest Regression](#), [Feature Engineering](#), [GroupBy Features](#), [Categorical and Numerical Features](#), [Missing Value Analysis](#), [Fill the missing Values](#), [Unique Value Analysis](#), [Univariate Analysis](#), [Bivariate Analysis](#), [Multivariate Analysis](#), [Correlation Analysis](#), [Spearman's \$\rho\$](#) ,
[Pearson's \$r\$](#) , [Kendall's \$\tau\$](#) , [Cramér's V \(\$\varphi_c\$ \)](#), [Phik \(\$\varphi_k\$ \)](#)

[Day 19 : MySQL and PostgreSQL](#)

[Day 20 : ETL \(Extract, Tranform and Load\) basics](#), [Why ETL is important?](#),
[How ETL works](#), [ETL Tools](#)

[Day 21 : Structured Data](#), [Semi Structured Data](#), [Unstructured Data](#), [Data Warehouse](#), [Data Mart](#), [Data Lake](#)

[Day 22 :Big Data](#), [Types of Big Data](#), [Big data tools](#), [SQL and NoSQL Databases](#), [Hadoop](#), [Hadoop HDFS](#), [Hadoop Yarn](#)

[Day 23: Batch Processing](#), [Stream Processing](#), [Apache Spark](#), [Apache Spark Commands](#), [Apache Kafka](#), [How Apache Kafka works](#)

Day 24 : Hive, Zookeper, Pig, Cassandra, Sqoop

Day 25: Docker, Docker vs Virtual Machines, Most important Docker commands, Kubernetes, Snowflake

Day 26 : Data Pipelines, Transformation, Processing, Workflow, Monitoring, Airflow, DAG

Day 27 : Power BI, Which chart to use and When?, Power BI – Data Analysis Expressions, Joins, Data Profiling

Day 28 : REST API, Postman, Data API

Day 29 : Data Engineering on cloud, AWS, AWS Services, Google Cloud Platform, GCP services

Day 30 : Machine Learning Algorithms, Linear Regression, Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, K Nearest Neighbors, K means Clustering, Hierarchical Clustering, Neural Networks

• • •

Complete Data Structures and Algorithm Series

Complexity Analysis

Backtracking

Sliding Window

Greedy Technique

Two pointer Technique

Arrays

Linked List

Strings

Stack

Queues

1- D Dynamic Programming

Divide and Conquer Technique

Recursion

Github —

GitHub — Coder-World04/Complete-Data-Structures-and-Algorithms: This repository contains everything...

This repository contains everything you need to become proficient in Data Structures and Algorithms Start here : Day 1...

[github.com](https://github.com/Coder-World04/Complete-Data-Structures-and-Algorithms)



• • •

Projects Videos —

All the projects, data structures, SQL, algorithms, system design, Data Science and ML, Data Analytics, Data Engineering, , Implemented Data Science and

ML projects, Implemented Data Engineering Projects, Implemented Deep Learning Projects, Implemented Machine Learning Ops Projects, Implemented Time Series Analysis and Forecasting Projects, Implemented Applied Machine Learning Projects, Implemented Tensorflow and Keras Projects, Implemented PyTorch Projects, Implemented Scikit Learn Projects, Implemented Big Data Projects, Implemented Cloud Machine Learning Projects, Implemented Neural Networks Projects, Implemented OpenCV Projects, Complete ML Research Papers Summarized, Implemented Data Analytics projects, Implemented Data Visualization Projects, Implemented Data Mining Projects, Implemented Natural Leaning Processing Projects, MLOps and Deep Learning, Applied Machine Learning with Projects Series, PyTorch with Projects Series, Tensorflow and Keras with Projects Series, Scikit Learn Series with Projects, Time Series Analysis and Forecasting with Projects Series, ML System Design Case Studies Series videos will be published on our youtube channel (just launched).

Subscribe today!

Ignito

Excited to share that we have launched our Youtube channel — Ignito to cover all the projects and coding exercise for ...

www.youtube.com



Ignito

• • •

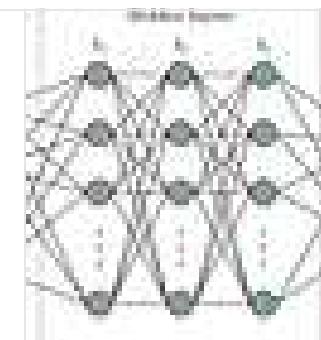
Another series that I'm starting along with Data Engineering is Machine Learning Ops — 30 days of Machine Learning Ops

60 Days of Data Science and Machine Learning with projects Series —

Day 1 — Day 60 : Quick Recap of 60 days of Data Science and ML

Connect the ML dots...

medium.com



The main aim of 30 days of Data Engineering with (amazing) projects series to understand Data Engineering from a practical perspective and get hands on practice by implementing projects (without falling in the rabbit hole of too much theory)

• • •

Solved System Design Case Studies

Design Instagram

Design Netflix

Design Reddit

Design Amazon

Design Messenger App

Design Twitter

Design URL Shortener

Design Dropbox

Design Youtube

Design API Rate Limiter

Design Web Crawler

Design Amazon Prime Video

Design Yelp

Design Uber

Design Tinder

Design Tiktok

Design Whatsapp

Most Popular System Design Questions

Mega Compilation : Solved System Design Case studies

• • •

Let's get started!

I'll be covering only the most important topics in Data Engineering with projects (written below) —

1. Data Engineering

What's Data Engineering

Why Data Engineering

Data Engineers – ML Engineers – Data Scientists

Purpose and Scope

2. Python for Data Engineering

Basic Python with Project

Advanced Python with Project

Techniques to write efficient and optimized code

3. SQL Basics

Structured Query Language

Query Structure

Conditions

Joins

Stored Procedures

4. Aggregations

Wild cards

Grouping Data

Aggregation Functions

Filtering

Sequences

Group By, Order By

Having Clause

Write Sub queries

Grouping Sets

Analytical Functions

5. Window Functions

Row Numbering

Percentile

Advanced windowing techniques

6. BigQuery

BigQuery Basics

SELECT, FROM, WHERE and Date and Extract in BigQuery

Common Expression Table

UNNEST Clause

SQL vs NoSQL Database

7. Advanced Functions

Triggers

Pivot

Cursors

Views

Indexes

Auto Increment

8. Performance Tuning SQL Queries

Query Optimizations in SQL

9. MySQL, PostgreSQL and MongoDB

Introduction to MySQL

Introduction to PostgreSQL

Introduction to MongoDB

Comparison between MySQL and PostgreSQL and MongoDB

Introduction to SQL and NoSQL Databases

MySQL in Depth

10. Scripting and Automation

Shell Scripting

ETL (Extract, Transform and Load) basics

Why ETL is important?

How ETL works

ETL Tools

11. Relational Databases and SQL

[Basic SQL](#)

[Advanced SQL](#)

12. NoSQL Data bases and Map Reduce

[Data Warehouses](#)

[Data Lakes](#)

[Structured Data](#)

[Semi Structured Data](#)

[Unstructured Data](#)

[Data Mart](#)

[Map-Reduce](#)

13.Data Analysis

Pandas

Numpy

Advanced Pandas Techniques

Data Pre-processing

Handling missing values

Data Cleaning

Mean/mode/median Imputation

Hot Deck Imputation

Rescale Data

Binarize Data

Regression Imputation

Stochastic regression imputation

Feature Scaling

Data Augmentation

Read and Process Large Datasets

Data Visualization basics

Data Visualization Projects

Data Visualization using Plotly and Bokeh

Data Profiling

Summary Functions

Indexing

[Grouping](#)

[Linear Regression](#)

[Multi Linear Regression](#)

[Polynomial Regression](#)

[Regression](#)

[Support Vector Regression,](#)

[Decision Tree Regression](#)

[Random Forest Regression](#)

[Feature Engineering](#)

[GroupBy Features](#)

[Categorical and Numerical Features](#)

Missing Value Analysis

Fill the missing Values

Unique Value Analysis

Univariate Analysis

Bivariate Analysis

Multivariate Analysis

Correlation Analysis

Spearman's ρ

Pearson's r

Kendall's τ

Cramér's V (φ_c)

Phik (φk)

14. Data Processing Techniques

Batch Processing

Stream Processing

Apache Spark

Apache Spark Commands

Apache Kafka

How Apache Kafka works

15. Big Data

Big Data

Types of Big Data

Big data tools

SQL and NoSQL Databases

Hadoop

Hadoop HDFS

Hadoop Yarn

Hive

Zookeeper

Pig

Cassandra

Sqoop

16. Data Pipelines and WorkFlows

Data Pipelines

Transformation

Processing

Workflow

Monitoring

Airflow

DAG

17. Infrastructure

Docker

Docker vs Virtual Machines

Most important Docker commands

Kubernetes

Snowflake

18. Power BI

Power BI

Which chart to use and When?

Power BI – Data Analysis Expressions

Joins

Data Profiling

19. Cloud Data Engineering

Data Engineering on cloud

AWS

AWS Services

Google Cloud Platform

Google Cloud Platform services

20. Machine Learning Algorithms

Linear Regression

Logistic Regression

Decision Trees

Random Forest

Support Vector Machines

K Nearest Neighbors

K means Clustering

Hierarchical Clustering

Neural Networks

Let's dive in!

• • •

Data Engineering

Data engineering is the process of preparing data for use in analysis, machine learning, and other applications. This includes tasks such as data ingestion, cleaning, and transformation, as well as building and maintaining the infrastructure and systems needed to store, process, and access the data.

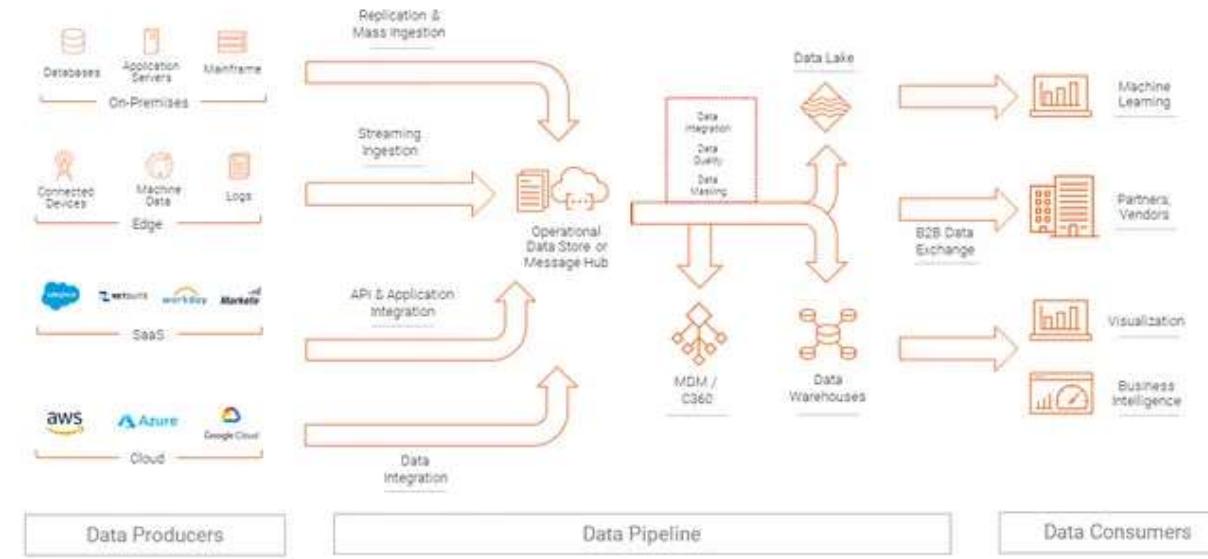
The purpose of data engineering is to make sure that data is in a form that can be easily used and understood by other members of the data team, such as data scientists and machine learning engineers.

In simple words, Data Engineering is the heart of designing, building for collecting, storing, processing, and analyzing large amount of data at scale.

To put it straight, in data engineering we develop and maintain large scale data processing systems to prepare structured and unstructured data to perform analytical modeling and make data driven decisions.

The aim of data engineering is to make quality data available for analysis and efficient data-driven decision making.

Data Pipeline Patterns



Pic credits : statsx

Most importantly, the Data Engineering ecosystem consists of 4 things –

Data – different data types, formats, and sources of data.

Data stores and repositories – Relational and non-relational databases, data warehouses, data marts, data lakes, and big data stores that store and process the data

Data Pipelines – Collect/Gather data from multiple sources, clean, process and transform it into data which can be used for analysis,

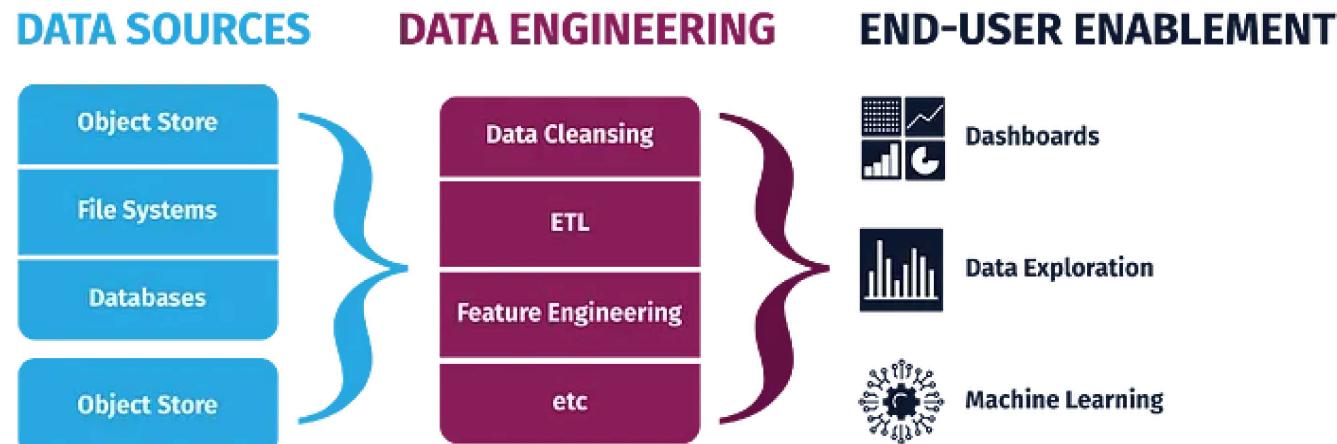
Analytics and Data driven Decision Making – Make the well processed data available for further business analytics, visualization and data driven decision making.



Pic credits : alterx

Why Data Engineering?

Data Engineering lifecycle consists of building/architecting data platforms, designing and implementing data stores and repositories, data lakes and gathering, importing, cleaning, pre-processing, querying, analyzing data, performance monitoring, evaluation, optimization and fine tuning the processes and systems.



Pic credits: techtaregt

It gives a great edge —

1. *To work and process with heterogeneous data formats and in the end get quality data that can be used in production.*
2. *To be able to work with large amount of data at scale and extract optimal value.*
3. *To automate the data pipelines and streams.*
4. *Use meta data efficiently.*
5. *To be able to derive amazing insights from the real time data (quality data).*

• • •

Data engineers play a crucial role in the field of data management and analytics. They are responsible for designing, building, and maintaining the infrastructure required for data acquisition, storage, processing, and delivery. This includes developing robust and scalable data pipelines,

integrating and transforming data from various sources, and ensuring data quality and reliability.

To better understand the role of data engineers, let's compare it with data science:

Data Engineering versus Data Science: Data engineering and data science are two distinct but interconnected fields within the broader realm of data analytics. While data science focuses on extracting insights and knowledge from data, data engineering is concerned with the technical aspects of managing and processing data. Here are some key differences between the two roles:

1. **Data Engineering:** Data engineers primarily work on the infrastructure and data pipelines, ensuring the efficient collection, storage, and processing of data. They focus on building and maintaining the systems that enable data scientists and analysts to work with large volumes of data effectively. Data engineers typically have expertise in data modeling, ETL (Extract, Transform, Load) processes, database systems, and distributed computing.
2. **Data Science:** Data scientists focus on analyzing and interpreting data to uncover patterns, trends, and insights that drive decision-making.

They apply statistical and machine learning techniques to extract actionable knowledge from the data. Data scientists often use programming languages like Python or R, and they possess skills in statistical analysis, machine learning algorithms, data visualization, and domain knowledge.

Popular tools used in data engineering:

Apache Hadoop: Hadoop is an open-source framework that allows distributed processing of large datasets across clusters of computers. It consists of two main components: Hadoop Distributed File System (HDFS) for storing data and MapReduce for processing and analyzing data in parallel. Hadoop is widely used for big data processing and is supported by various tools and libraries in the Hadoop ecosystem.

```
import pydoop.hdfs as hdfs

# Read a file from HDFS
with hdfs.open("/path/to/file.txt") as file:
    data = file.read()
    print(data)
```

Apache Spark: Spark is another open-source framework that provides an in-memory computing engine for big data processing. It supports various programming languages, including Python, and offers high-level APIs for distributed data processing and machine learning. Spark is known for its speed and scalability and is often used for real-time data streaming, batch processing, and iterative algorithms.

```
from pyspark.sql import SparkSession

# Create a SparkSession
spark = SparkSession.builder.appName("WordCount").getOrCreate()

# Read a text file
lines = spark.read.text("file:///path/to/file.txt")

# Count the occurrences of each word
word_counts = lines.rdd.flatMap(lambda line: line.value.split()).countByValue()

# Print the word counts
for word, count in word_counts.items():
    print(f"{word}: {count}")
```

Apache Kafka: Kafka is a distributed streaming platform that provides a publish-subscribe messaging system for real-time data streaming. It allows data engineers to efficiently collect, process, and transmit large volumes of data between different systems or applications. Kafka is commonly used for building data pipelines, event-driven architectures, and real-time analytics.

```
from kafka import KafkaProducer, KafkaConsumer

# Create a Kafka producer
producer = KafkaProducer(bootstrap_servers='localhost:9092')

# Produce a message to a topic
producer.send('my_topic', b'Hello, Kafka!')

# Create a Kafka consumer
consumer = KafkaConsumer('my_topic', bootstrap_servers='localhost:9092')

# Consume messages from a topic
for message in consumer:
    print(message.value.decode('utf-8'))
```

Apache Airflow: Airflow is an open-source platform for orchestrating and scheduling data workflows. It allows data engineers to define, schedule, and

monitor complex data pipelines as directed acyclic graphs (DAGs). Airflow supports various data sources and destinations, and it integrates well with other tools in the data engineering ecosystem.

```
from airflow import DAG
from airflow.operators.bash_operator import BashOperator
from datetime import datetime

# Define the DAG
dag = DAG(
    'my_dag',
    description='A simple DAG',
    start_date=datetime(2023, 1, 1),
    schedule_interval='@daily',
)

# Define the tasks
task1 = BashOperator(
    task_id='task1',
    bash_command='echo "Task 1"',
    dag=dag,
)

task2 = BashOperator(
    task_id='task2',
    bash_command='echo "Task 2"',
    dag=dag,
)
```

```
# Set task dependencies
task1 >> task2
```

Complete Code —

```
import pydoop.hdfs as hdfs
from pyspark.sql import SparkSession
from kafka import KafkaProducer, KafkaConsumer
from airflow import DAG
from airflow.operators.python_operator import PythonOperator
from datetime import datetime

# Apache Hadoop – Reading a file from HDFS
def read_file_from_hdfs():
    with hdfs.open("/path/to/file.txt") as file:
        data = file.read()
        print(data)

# Apache Spark – Word count example
def word_count_with_spark():
    spark = SparkSession.builder.appName("WordCount").getOrCreate()
    lines = spark.read.text("file:///path/to/file.txt")
    word_counts = lines.rdd.flatMap(lambda line: line.value.split()).countByValue
    for word, count in word_counts.items():
        print(f"{word}: {count}")

# Apache Kafka – Producing and consuming messages
def produce_and_consume_messages():
    producer = KafkaProducer(bootstrap_servers='localhost:9092')
    producer.send('my_topic', b'Hello, Kafka!')
```

```
consumer = KafkaConsumer('my_topic', bootstrap_servers='localhost:9092')
for message in consumer:
    print(message.value.decode('utf-8'))

# Apache Airflow - Defining a DAG and tasks
def my_function():
    print("Hello, Airflow!")

dag = DAG(
    'data_engineering_pipeline',
    description='Example data engineering pipeline',
    schedule_interval='0 0 * * *', # Runs daily at midnight
    start_date=datetime(2023, 5, 17),
    catchup=False
)

read_file_task = PythonOperator(
    task_id='read_file_from_hdfs',
    python_callable=read_file_from_hdfs,
    dag=dag
)

word_count_task = PythonOperator(
    task_id='word_count_with_spark',
    python_callable=word_count_with_spark,
    dag=dag
)

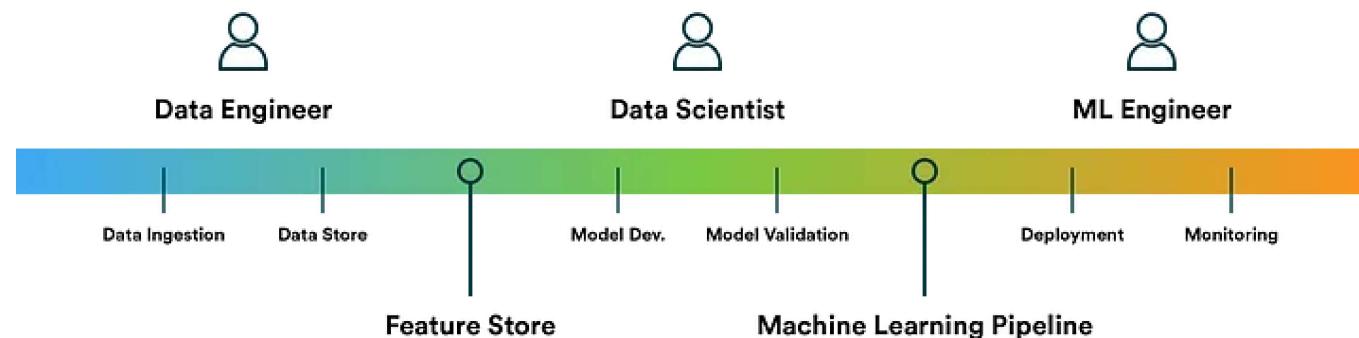
produce_consume_task = PythonOperator(
    task_id='produce_and_consume_messages',
    python_callable=produce_and_consume_messages,
    dag=dag
)

my_function_task = PythonOperator(
```

```
task_id='my_function_task',  
python_callable=my_function,  
dag=dag  
)  
  
read_file_task >> word_count_task >> produce_consume_task >> my_function_task
```

• • •

How Data Engineers are different from ML Engineers and Data Scientists?



Pic credits : valoh

Data Engineers — To put it straight, data engineer is responsible for making quality data available from various resources, maintain databases, build data

pipelines, query data, data preprocessing, Feature Engineering, Apache hadoop and spark, Develop data workflows using Airflow etc

Data Scientists and ML Engineers — On the other hand, ML Engineers and



Search Medium

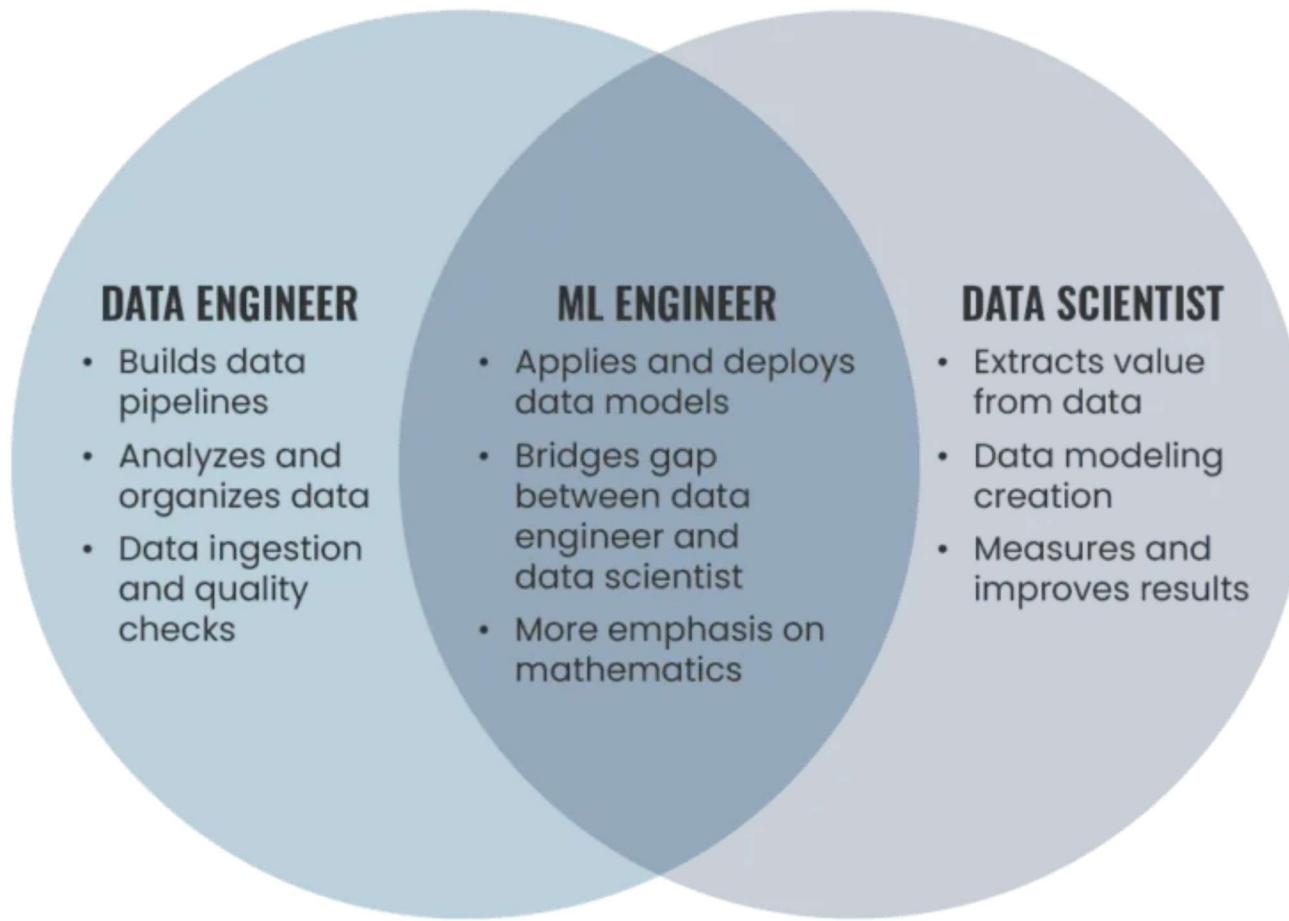
Write



B

ML models and deploy them, have statistical and mathematical knowledge and measure, optimize and improve results.

DATA TEAM AT A GLANCE



Pic credits : phdata

Data engineers and machine learning engineers are two distinct roles, although there is some overlap between the two. Data engineers are responsible for designing

and building the infrastructure and systems that store and process data, while machine learning engineers are responsible for building and deploying machine learning models. Data scientists are responsible for analyzing and interpreting data to gain insights and make decisions.

. . .

Purpose, Scope and Responsibilities

The scope of data engineering includes a wide range of tasks, from data pipeline design and data warehousing to working with big data technologies such as Hadoop and Spark. Data engineers also work closely with data scientists and machine learning engineers to ensure that the data is in a form that can be easily used and understood by these other members of the data team.

Data Engineers are responsible for building the most efficient data infrastructure in order to process large amount of data coming from various sources.

Data Analyst	Data Engineer	Data Scientist
Pre-processing and data gathering	Develop, test & maintain architectures	Responsible for developing Operational Models
Emphasis on representing data via reporting and visualization	Understand programming and its complexity	Carry out data analytics and optimization using machine learning & deep learning
Responsible for statistical analysis & data interpretation	Deploy ML & statistical models	Involved in strategic planning for data analytics
Ensures data acquisition & maintenance	Building pipelines for various ETL operations	Integrate data & perform ad-hoc analysis
Optimize Statistical Efficiency & Quality	Ensures data accuracy and flexibility	Fill in the gap between the stakeholders and customer

Pic credits : datahouse

The purpose and scope of 30 days of Data Engineering has already been discussed above.

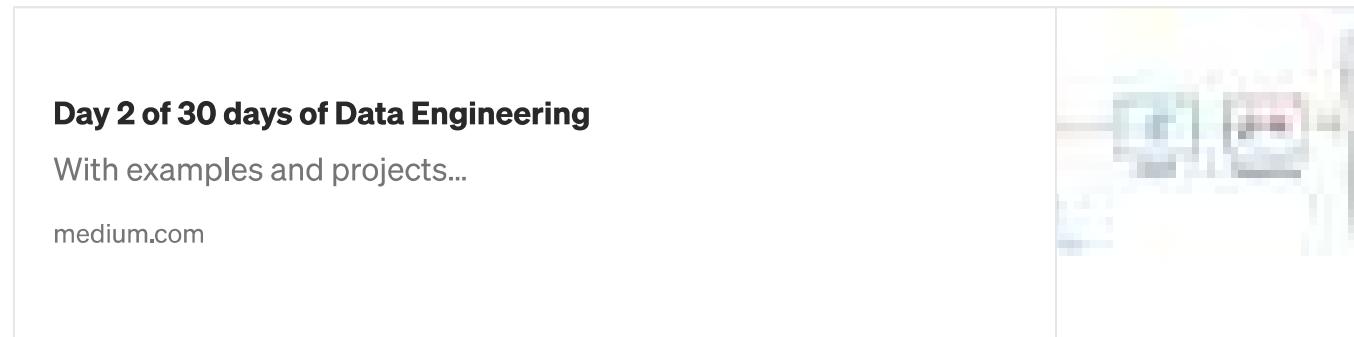
To re-iterate, the goal of this series is to give practical hands-on exposure while covering bits and pieces of important theory concepts.

Join me in this journey!! :)

• • •

That's it for now!

Day 2:



• • •

System Design Case Studies — In Depth

[Design Instagram](#)

[Design Messenger App](#)

[Design Twitter](#)

[Design URL Shortener](#)

[Design Dropbox](#)

[Design Youtube](#)

Design API Rate Limiter

Design Web Crawler

Design Facebook's Newsfeed

Most Popular System Design Questions

Mega Compilation : Solved System Design Case studies

Complete Data Structures and Algorithm Series

Complexity Analysis

Sliding Window

Backtracking

Greedy Technique

Two pointer Technique

1- D Dynamic Programming

Divide and Conquer Technique

Recursion

Github —

GitHub — Coder-World04/Complete-Data-Structures-and-Algorithms: This repository contains everything...

This repository contains everything you need to become proficient in Data Structures and Algorithms Start here : Day 1...

github.com



• • •

Complete System Design Series Parts —

1. System design basics

2. Horizontal and vertical scaling

3. Load balancing and Message queues

4. High level design and low level design, Consistent Hashing, Monolithic and Microservices architecture

5. Caching, Indexing, Proxies

6. Networking, How Browsers work, Content Network Delivery (CDN)

7. Database Sharding, CAP Theorem, Database schema Design

8. Concurrency, API, Components + OOP + Abstraction

9. Estimation and Planning, Performance

10. Map Reduce, Patterns and Microservices

11. SQL vs NoSQL and Cloud

12. Most Popular System Design Questions

Github —

Complete-System-Design/README.md at main · Coder-World04/Complete-System-Design

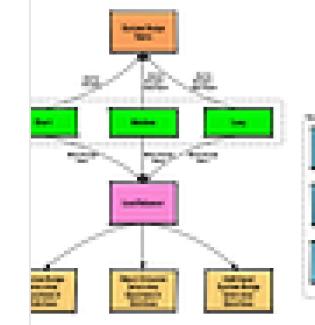
This repository contains everything you need to become proficient in System Design
Topics you should know in System...

github.com

Most Popular System Design Questions — Mega Compilation

Just for your reference...

medium.com



Keep learning and coding :)

Advanced SQL Series

Day 1 : SQL Basics and Kick start of Advanced SQL Series

Day 2 : SQL Basics, Query Structure, Built In functions Conditions

Day 3 : Most Important Commands, Joins and Filters

Day 4 : Set Theory Operations, Stored Procedures and CASE statements in SQL

Day 5 : Wildcards, Aggregation and Sequences in SQL

Day 6 : Subqueries, Group by, order by and Having clauses in SQL and Analytical Functions

Day 7 : Window Functions, Grouping Sets and Constraints in SQL

Day 8 : BigQuery Basics, SELECT, FROM, WHERE and Date and Extract in BigQuery

Day 9 : Common Expression Table, UNNEST Clause, SQL vs NoSQL Databases

Day 10 : Triggers, Pivot and Cursors in SQL

Day 11 : Views, Indexes and Auto Increment in SQL

Day 12 : Query optimizations, Performance tuning in SQL

Day 13 : Introduction to MySQL, PostgreSQL and Mongo DB, Comparison between MySQL and PostgreSQL and Mongo DB, Introduction to SQL and NoSQL Databases

Day 14 : MySQL in Depth

Day 15 : PostgreSQL in Depth

Anyways, For Day 15 of 15 days of Advanced SQL, we will cover —

PostgreSQL inDepth

Github for Advanced SQL that you can follow —

Complete-Advanced-SQL-Series/README.md at main · Coder-World04/Complete-Advanced-SQL-Series

This repository contains everything you need to become proficient in Advanced SQL Structured Query Language Query...

[github.com](https://github.com/Coder-World04/Complete-Advanced-SQL-Series)

Verify or
be Advanced...

• • •

All the projects, data structures, algorithms, system design, Data Science and ML, Data Engineering, MLOps and Deep Learning videos will be published on our youtube channel (just launched).

Subscribe today!

Ignito

Excited to share that we have launched our Youtube channel — Ignito to cover all the projects and coding exercise for ...

www.youtube.com



System Design Case Studies — In Depth

[Design Instagram](#)

[Design Messenger App](#)

[Design Twitter](#)

Design URL Shortener

Design Dropbox

Design Youtube

Design API Rate Limiter

Design Web Crawler

Design Facebook's Newsfeed

Design Yelp

Design Uber

Design Tinder

Design Tiktok

Design Whatsapp

Most Popular System Design Questions

Mega Compilation : Solved System Design Case studies

Complete Data Structures and Algorithm Series

Complexity Analysis

Backtracking

Sliding Window

Greedy Technique

Two pointer Technique

Arrays

Linked List

Strings

Stack

Queues

Hash Table/Hashing

Binary Search

1- D Dynamic Programming

Divide and Conquer Technique

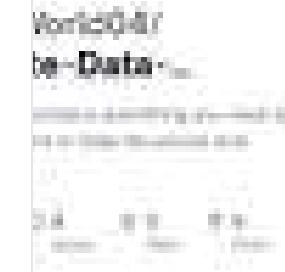
Recursion

Github —

GitHub — Coder-World04/Complete-Data-Structures-and-Algorithms: This repository contains everything...

This repository contains everything you need to become proficient in Data Structures and Algorithms Start here : Day 1...

[github.com](https://github.com/Coder-World04/Complete-Data-Structures-and-Algorithms)



• • •

30 days of Data Analytics Series —

Day 1 : Data Analytics basics and kickstart of Data analytics with projects series

Day 2: Business Understanding – Data Driven Decision Making, Descriptive Analysis, Predictive Analysis, Diagnostic Analysis, Prescriptive Analysis

Day 3 : Data Analytics Ecosystem – Data Life Cycle, Data Analysis complete process (most important things)

Day 4 : Probability, Conditional Probability, Binomial Distribution, Probability Density Function, Sampling Distribution

Day 5 : Statistics

Day 6 : Basic and Advanced SQL

Day 7 : Data Collection, Data Cleaning and Python

Day 8 : Pandas and Numpy

Day 9 : Data Manipulation

Day 10 : Data Visualization – Part 1

Day 11 : Project 1 : Data Visualization – Part 2

Day 12 : Data Visualization – Part 3

Day 13: Tableau – Part 1

Day 14: Tableau – Part 2

Day 15: Tableau – Part 3

Tableau Project

Day 16 : Data Analysis Project 2

Day 17 : Data Analysis Project 3

Day 18: Data Analysis Project 4

Day 19: Data Analysis Project 5

Day 20 : Data Analysis Project 6

Categorical and Numerical Features

Missing Value Analysis

Fill the missing Values

Unique Value Analysis

Univariate Analysis

Bivariate Analysis

Multivariate Analysis

Correlation Analysis

Day 21 : Data Analysis Project 7

Data Profiling

Feature Engineering

GroupBy Features

Categorical and Numerical Features

Missing Value Analysis

Fill the missing Values

Unique Value Analysis

Univariate Analysis

Bivariate Analysis

Multivariate Analysis

Correlation Analysis

Day 22 : Data analysis Project 8

Linear Regression

Data Profiling

Feature Engineering

Sort Values

Categorical and Numerical Features

Missing Value Analysis

Unique Value Analysis

Univariate Analysis

Bivariate Analysis

Multivariate Analysis

Correlation Analysis

Correlation Coefficients

Take Complete Hands On Tableau Course : [Link](#)

Some of the other best Series —

30 Days of Natural Language Processing (NLP) Series

How to solve any System Design Question (approach that you can take)?

Complete System Design Case Studies Series

30 days of Data Structures and Algorithms and System Design Simplified

60 Days of Deep Learning with Projects Series

60 days of Data Science and ML Series with projects

Data Science and Machine Learning Research (.papers) Simplified **

60 Days of Deep Learning with Projects Series

100 days : Your Data Science and Machine Learning Degree Series with projects

23 Data Science Techniques You Should Know

Tech Interview Series – Curated List of coding questions

Complete System Design with most popular Questions Series

Complete Data Visualization and Pre-processing Series with projects

Complete Python Series with Projects

Complete Advanced Python Series with Projects

Kaggle Best Notebooks that will teach you the most

Complete Developers Guide to Git

All the Data Science and Machine Learning Resources

210 Machine Learning Projects

30 days of Machine Learning Ops

• • •

For Python Projects —

Complete Python And Projects — Mega Compilation

Everything that you need to know in Python with Projects...

medium.com



Analyzing Video using Python, OpenCV and NumPy

With Code Implementation...

medium.datadriveninvestor.com





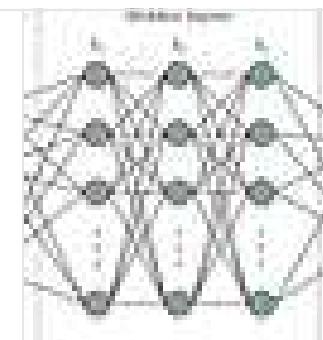
• • •

For complete 60 days of Data Science and ML : Day 1 — Day 60 : Quick Recap of 60 days of Data Science and ML

Day 1 — Day 60 : Quick Recap of 60 days of Data Science and ML

Connect the ML dots...

medium.com



Follow for more updates. Stay tuned and keep coding! Some of the links are affiliates.

• • •

For other projects, tune to —

Build Machine Learning Pipelines(With Code)

Build Machine Learning Pipelines(With Code) — Part 1

Complete implementation...

medium.datadriveninvestor.com

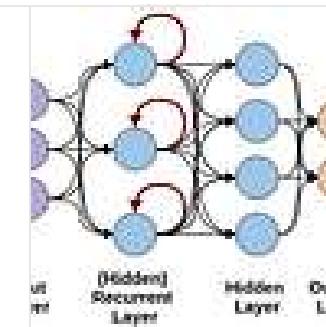


Recurrent Neural Network with Keras

Recurrent Neural Network with Keras

Project Implementation and cheatsheet...

medium.datadriveninvestor.com

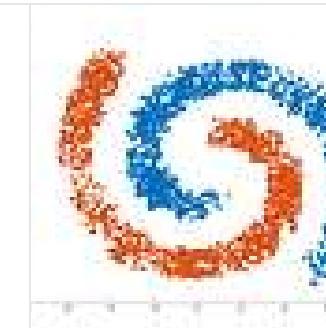


Clustering Geolocation Data in Python using DBSCAN and K-Means

Clustering Geolocation Data in Python using DBSCAN and K-Means

Project Implementation...

medium.datadriveninvestor.com



Facial Expression Recognition using Keras

Facial Expression Recognition using Keras

Project Implementation...

[medium.datadriveninvestor.com](https://medium.datadriveninvestor.com/facial-expression-recognition-using-keras-implementation-894822fcb128)

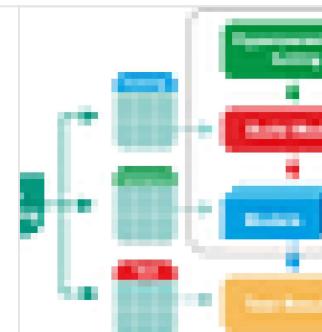


Hyperparameter Tuning with Keras Tuner

Hyperparameter Tuning with Keras Tuner

Project Implementation....

[medium.datadriveninvestor.com](https://medium.datadriveninvestor.com/hyperparameter-tuning-with-keras-tuner-implementation-894822fcb128)

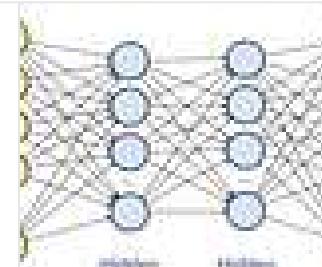


Custom Layers in Keras

Custom Layers in Keras

Code implementation ...

[medium.datadriveninvestor.com](https://medium.datadriveninvestor.com/custom-layers-in-keras-implementation-894822fcb128)





Machine Learning

Data Science

Tech

Artificial Intelligence

Programming



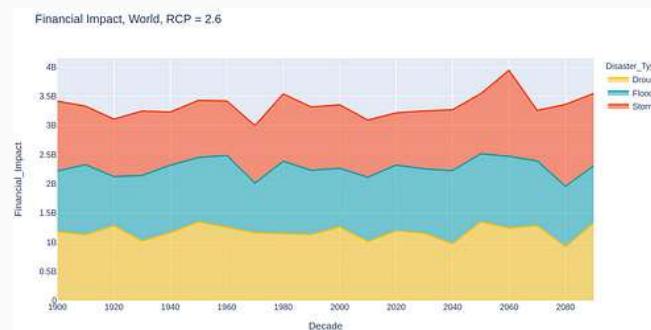
Written by Naina Chaturvedi

8.9K Followers · Editor for Coders Mojo

us,World Traveler, Sr. SDE, Researcher Cornell Uni, Women in Tech, Coursera
Instructor ML & GCP, Trekker, IITB, Reader, I write for fun@AI & Python publications

[Follow](#)

More from Naina Chaturvedi and Coders Mojo



Naina Chaturvedi in Coders Mojo

Day 12 of 30 days of Data Analytics with Projects Series

Welcome back peeps. This is Day 12 of 30 days of data analytics.

⭐ · 8 min read · Nov 21, 2022



117



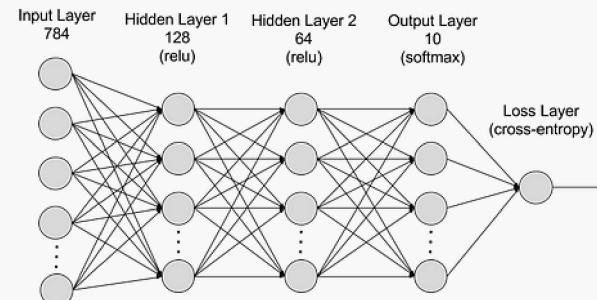
...



207



...

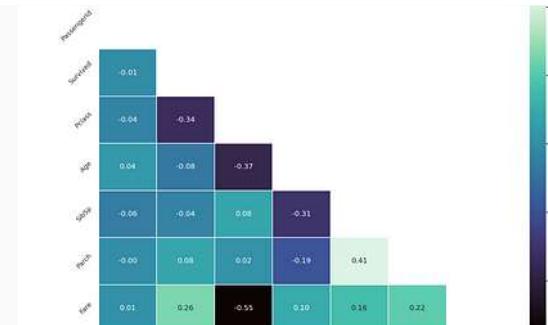
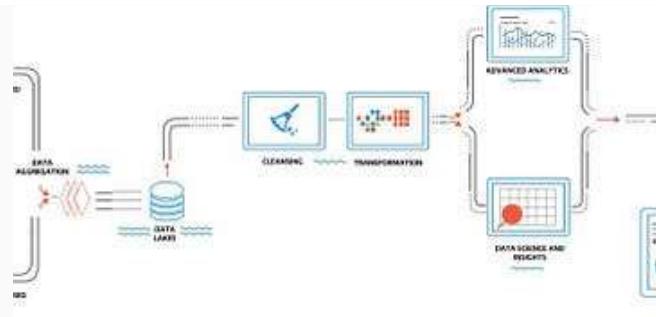


Naina Chaturvedi in Coders Mojo

30 days of PyTorch with Projects Series

Vertical series (One post that will house all the projects as we build/implement them)

⭐ · 5 min read · Dec 30, 2022



Naina Chaturvedi in Coders Mojo

Day 2 of 30 days of Data Engineering

With examples and projects...

★ · 36 min read · Aug 17, 2022

118



...

228



1

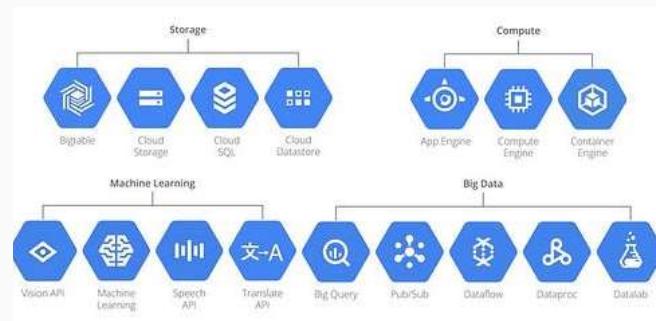


...

[See all from Naina Chaturvedi](#)

[See all from Coders Mojo](#)

Recommended from Medium



Naina Chaturvedi in Coders Mojo

Day 29 of 30 days of Data Engineering Series with Projects

Welcome back peeps to Day 29 of Data Engineering Series with Projects!

★ · 8 min read · Dec 29, 2022



103



...

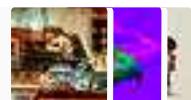


101



...

Lists

**What is ChatGPT?**

9 stories · 95 saves

**Stories to Help You Grow as a Software Developer**

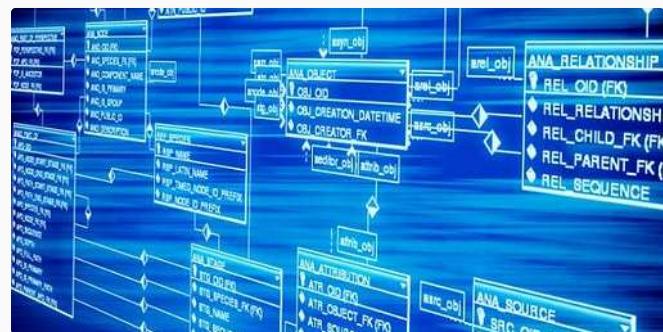
19 stories · 114 saves

**Leadership**

30 stories · 54 saves

**Stories to Help You Level-Up at Work**

19 stories · 92 saves



Naina Chaturvedi in Coders Mojo

Day 6 of 30 days of Data Engineering Series with Projects

Welcome back peeps to Day 6 of Data Engineering Series with Projects!

★ · 10 min read · Dec 23, 2022



...



...



Naina Chaturvedi in Coders Mojo

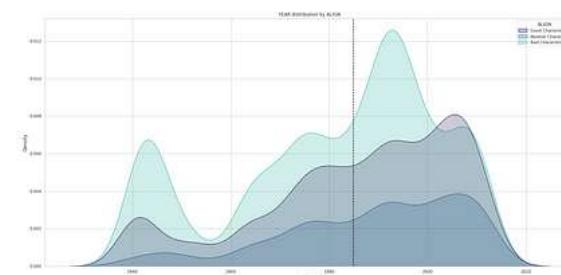
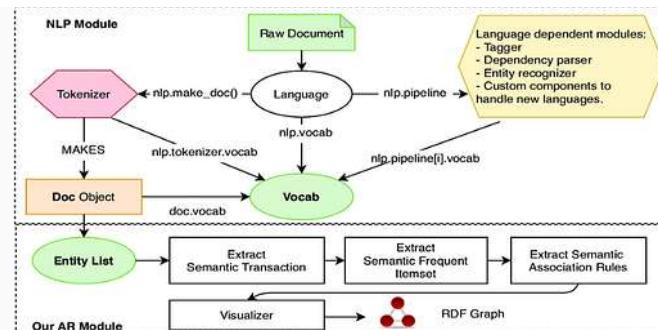
Day 15 of 30 days of Data Engineering Series with Projects

Welcome back peeps to Day 15 of Data Engineering Series with Projects!

★ · 11 min read · Dec 24, 2022



...



Naina Chaturvedi in Coders Mojo

Implemented Natural Leaning Processing Projects

Repo for all the projects (vertical post)...

⭐ · 70 min read · Jan 6



128



...



113



...

[See more recommendations](#)