CILCO

O What is CI/CD?

A CIICO pipeline is a concept central to software. It spins a whole field of processes, taking testing methods, and tooling, all facilitated by the Git code cossioning process.

> CI checks and tests every new piece of code Cor data towns formation logic) you add to your data pipeline.

-> cD ensures that once tested and approved, this code gets added to the live system without manual intervention

oci, CD, in the context of data pipeline deployment, focuses on automating data operations and transformations. This merges develop -ment, testing and operational workflows into a unified, automated process, ensuring the data assests are consistently high quality and the data infrastructure evolves smoothly, even at scale.

Ising CI(CD for data pipelines automation has become more critical in ensuring the development velocity of processes such as toaining machine learning models, supporting a data science team, doing large scale data analysis, business intelligence or data visualization, supporting the growth of unstructed data collection, and other business needs

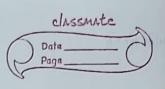
Continuous Integration (CI) in datapipelines.

O Automated testing:

Automated tests check the integrity and quality of data boans formations, ensuring the data is processed as expected and any error is spotted early.

O Version Control:

Pata pipeline code (cg:-SQL surspts) By then transformation is stored in repositeries like bilb, allowing tracking and managing



changes.

3 Consistent Environment:

CI tooks can own test in environments that missos production, ensuring that differences in configuration or dependis don't introduces errors

Data Quality checks:-

These might includes for null values, data range violations, data type mismatches or other custom quality rules.

Continuous data pipelines deployments

O Automated Reployment:

Once code change passall CI checks, CO tools can automate their deployment to production, ensuring seamless data flow.

12 Monitoring and Alest ts:-

Once deployed, monitoring tools keep toack of the data pipelines pertormance, data quality, and any potential issues. Automated alerts can notify on discognacies.

@ Rell backs:

In case on issue is identified post-deployment, co processes allow for quick rollbacks to a previously stable state of the data pipeline

(3) Intros touchure as code CIaC):

Many CO tools support Pac's. For example, cloud resources such as storage or compute can be provisined automatically as part of the deployment process.

Git 's -> Git is a distributed version control system that facilitates collabrative software development by tracking changes across multiple contributions: -2 Git can be paired with data orches toation tooks and integrated into CI/CD workflows, providing the benefits of stream lined deployment and consistency in data engineering tasks. ETL pipelines!-+ ETL (Bx toach, Fransform, Load) pipelines are at the heart of the data engineering. They've the processess that pull data from source Cdatabases, APD's, etc. 2, transform it into a usable format, and then load it into a destination, like databases or a data wavehouse when you deploy an ETL script to Git, you're not just saving the code- you could be toggering a series of events: 1) Testing :-Automated tests are first run to ensure the new code doesn't break anything. @ Deployment: Once tests pass, the BTL processes can be automatically deployed to a staging or production environment. 3 Notifications: If any part of the process fails, or if its successfully completed, notifications can be sent out