

DataOps



by Datics



Speaker



- IT Consultant at Datics Consulting
- DataOps/MLOps Engineer at CARIAD
- Cloud/Security Engineer at Karl Storz
- Cloud/DevOps Engineer at Lufthansa
- Cloud/SRE Engineer at Volkswagen
- System Administrator at Reply

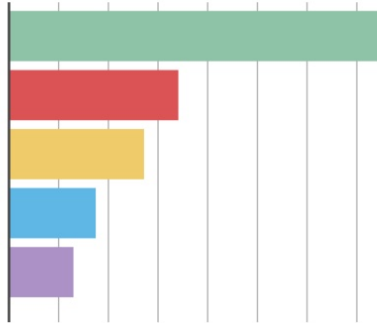
Agenda

- DataOps Introduction with Definitions
- What's improved with DataOps
- Use Case: ETL
- Why Every Data Team Needs a DataOps Framework
- The Roles of DataOps Engineer

DataOps

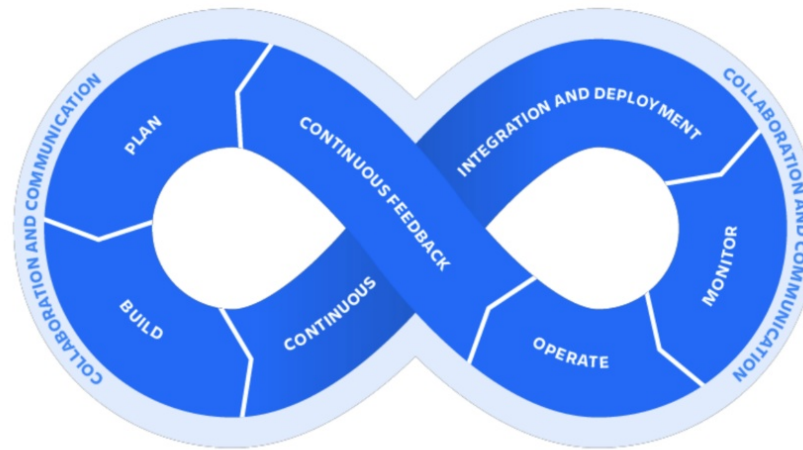


by Datics



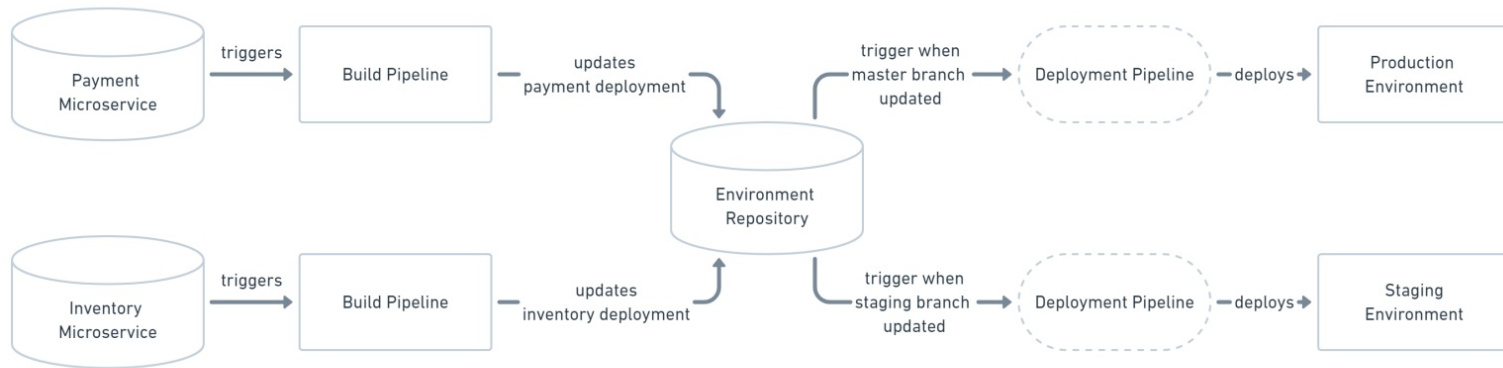
DATAOPS

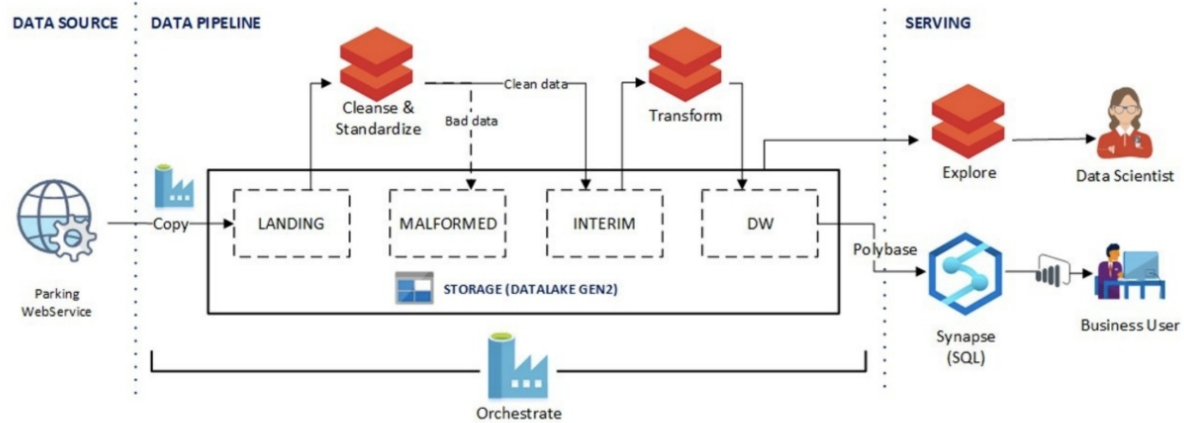
A methodology that associates with DevOps to
promote the automation of data orchestration
improve communication and integration in a data team.



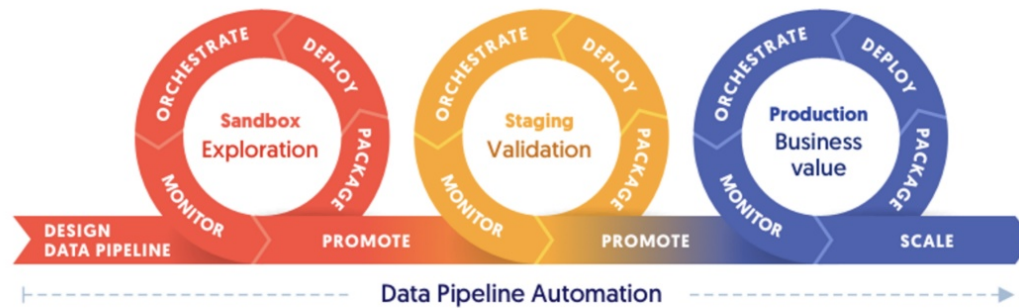
- Shorten the software development life cycle and provide continuous delivery with high software quality.
- Generate (I mean detect) bugs, code fast, deploy faster, promise the application always functioning.
- Improve collaboration between developers and system operation engineers.

Application Deployment for Multiple Environments



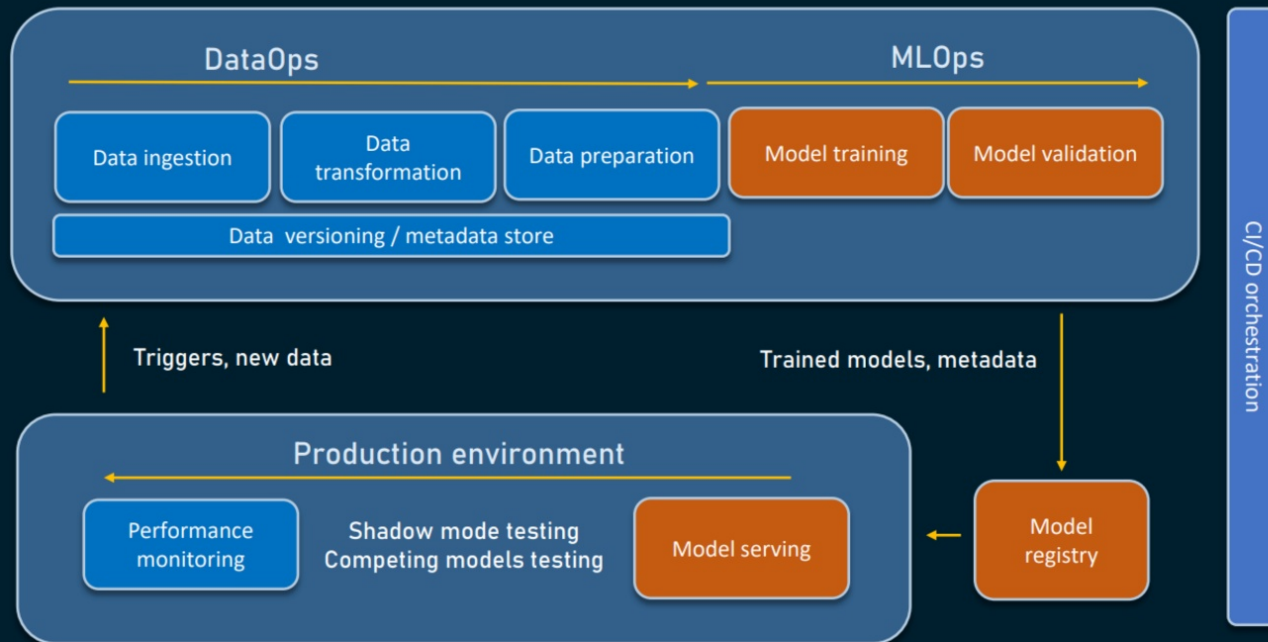


Deliver high quality, on-demand data to organizational customers



- Speed up the development and deployment of automated data workflows (within data pipeline).
- Improve collaboration inside data team:
 - data scientists,
 - data engineer,
 - data/BI analyst,
 - data quality analyst

Components of DataOps in MLOps workflow

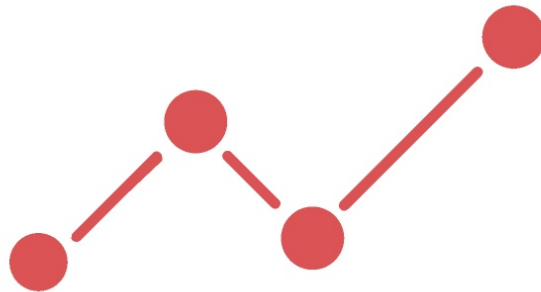


Data●ps

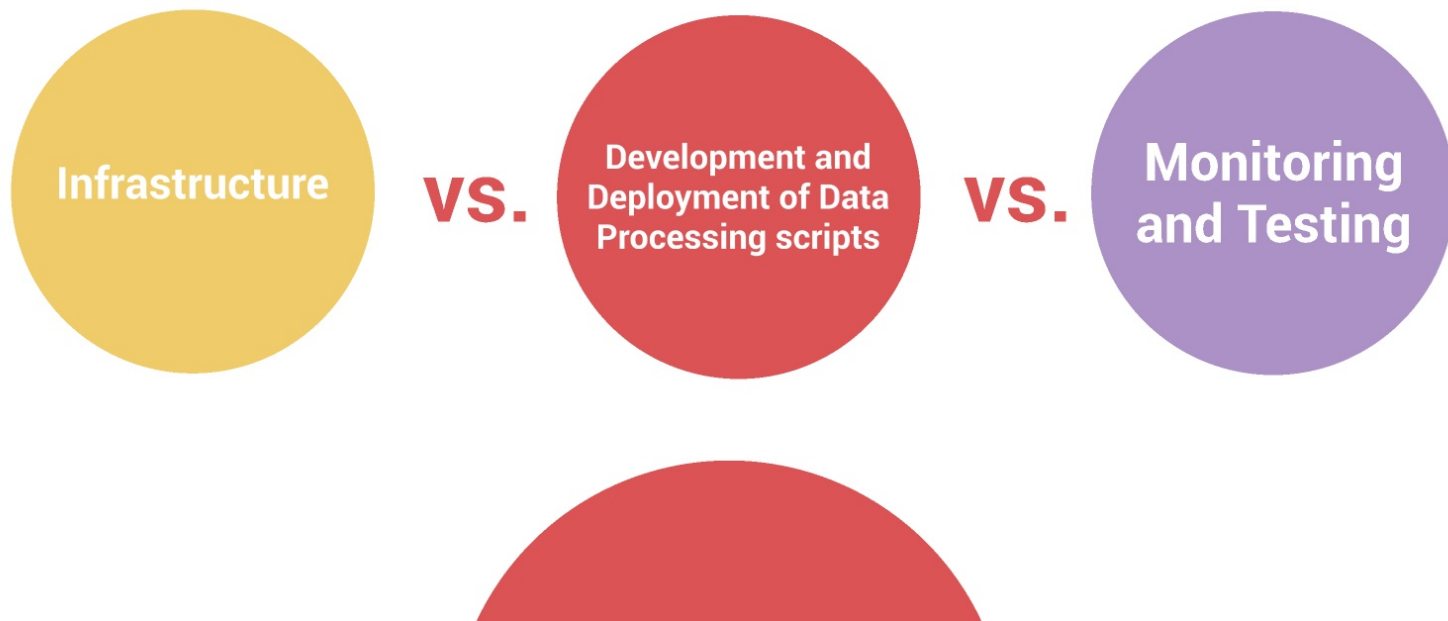


by Datics

What's improved with DataOps



You have data and a story to tell



Infrastructure

Improved:

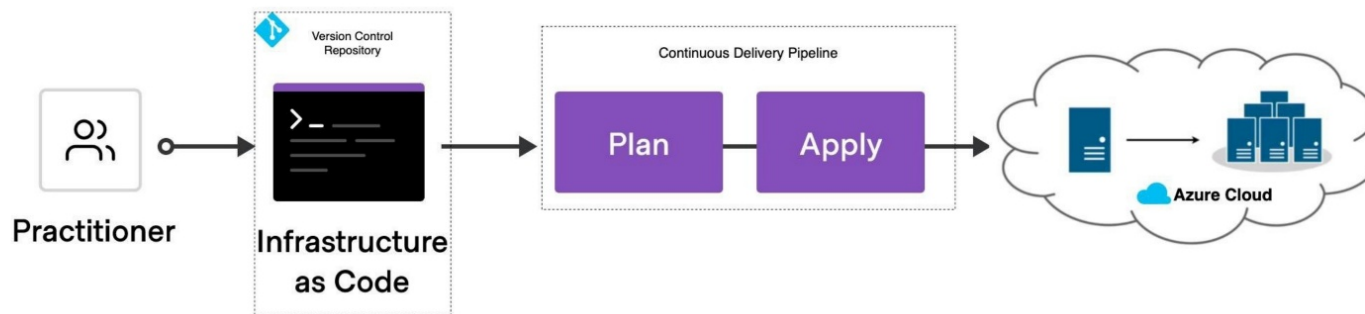
- Provisioning a large-scaled infrastructure is faster with automation jobs.
- Infrastructure states are more manageable with Infrastructure-as-Code.
- Resource availability and scalability can be configured automatically.

Infrastructure

Improved:

- Provisioning a large-scaled infrastructure is faster with automation jobs.
- Infrastructure states are more manageable with Infrastructure-as-Code.
- Resource availability and scalability can be configured automatically.





Infrastructure
Example Concept from DataOps projects
in Automotive and Health-Care Industry

- IaC: Terraform
- CD: GitHub Action
- Cloud: Microsoft Azure

Development and Deployment of Data Processing scripts

Improved:

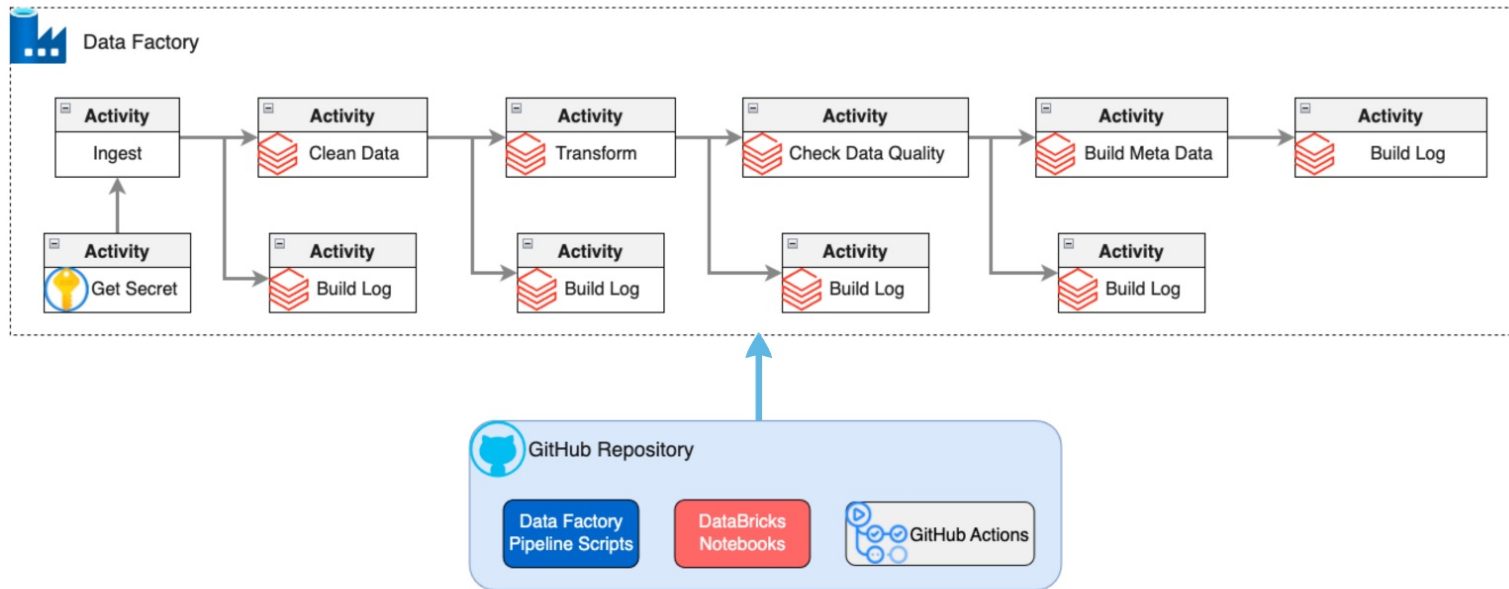
- Differences between local development and production environments are reduced.
- Deployment cycle is shortened and done automatically with the help of CI/CD pipeline.

Development and Deployment of Data Processing scripts

Improved:

- Differences between local development and production environments are reduced.
- Deployment cycle is shortened and done automatically with the help of CI/CD pipeline.

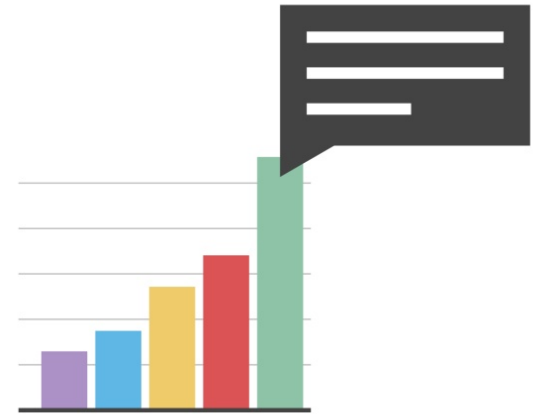




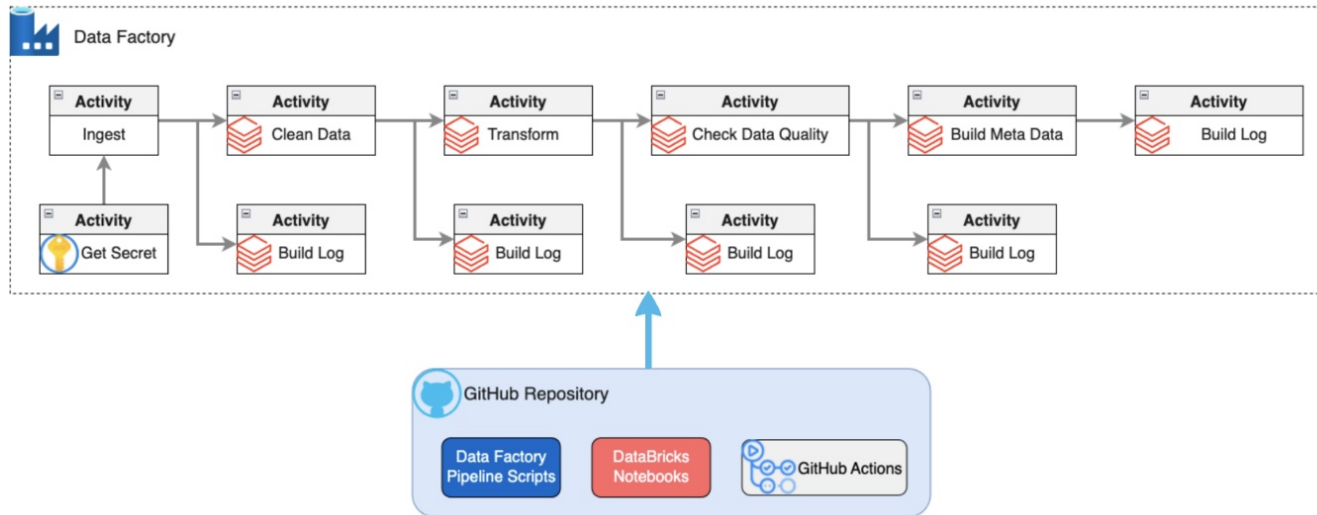
Development and Deployment of Data Processing scripts
Example Concept from DataOps projects in Automotive Project

Monitoring and testing

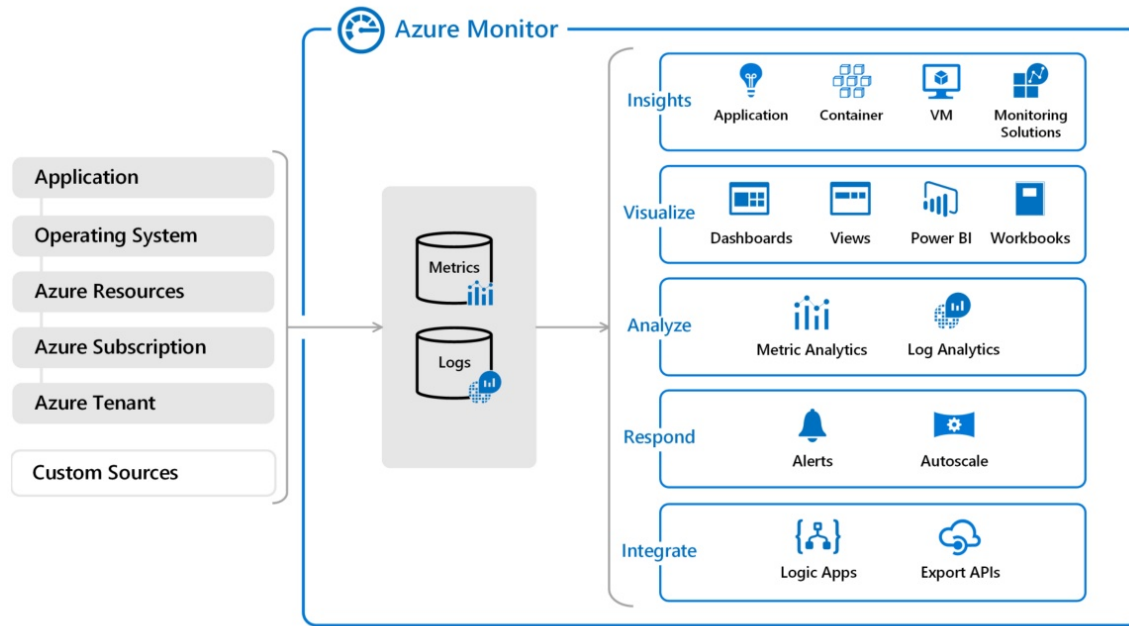
- Monitor for data anomalies. Quality of data model, traceability, computing resource, metrics.



Monitoring and Testing



Monitoring and Testing



External Use © Datics Consulting

21

DataOps

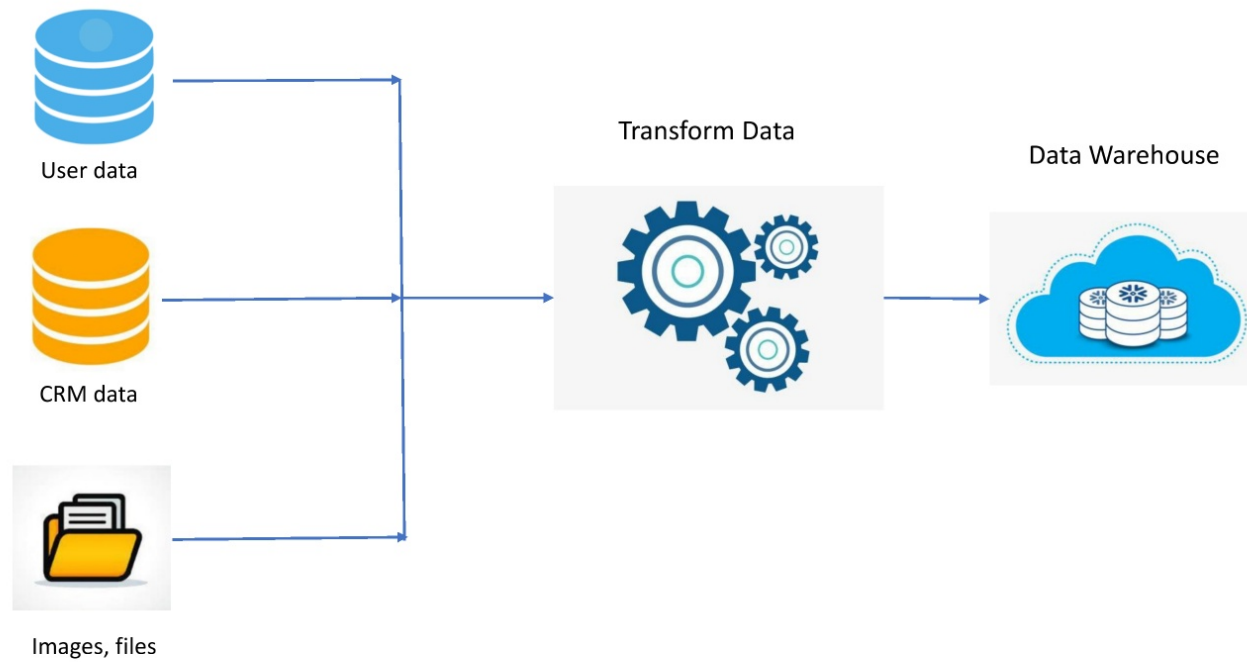


by Datics

The case of ETL

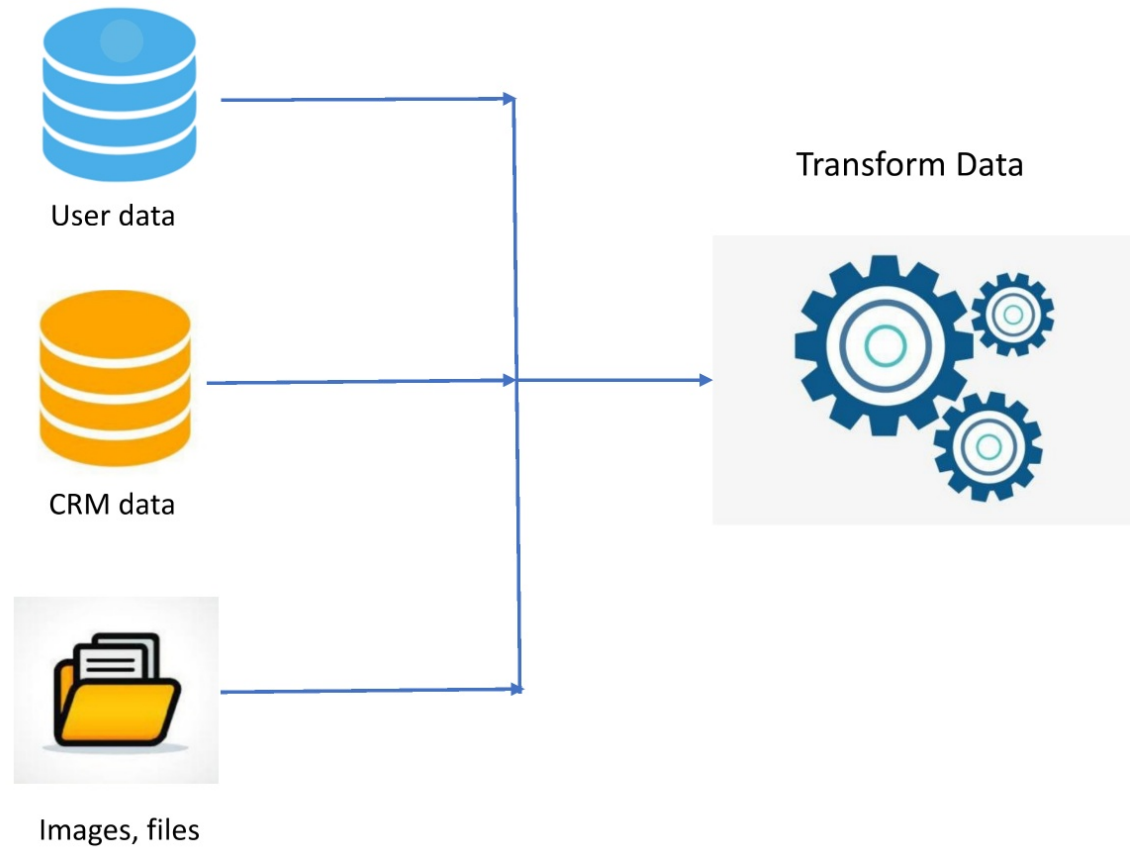


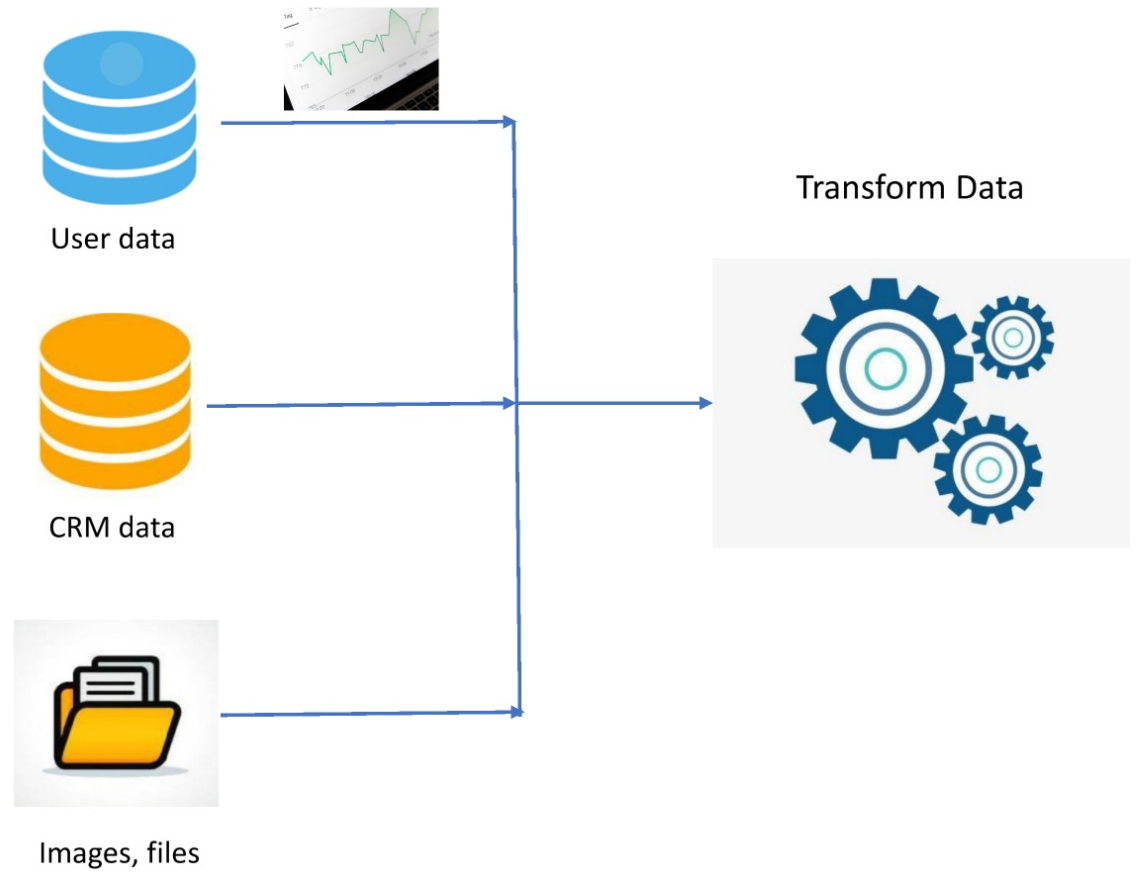
Use case



External Use © Datics Consulting

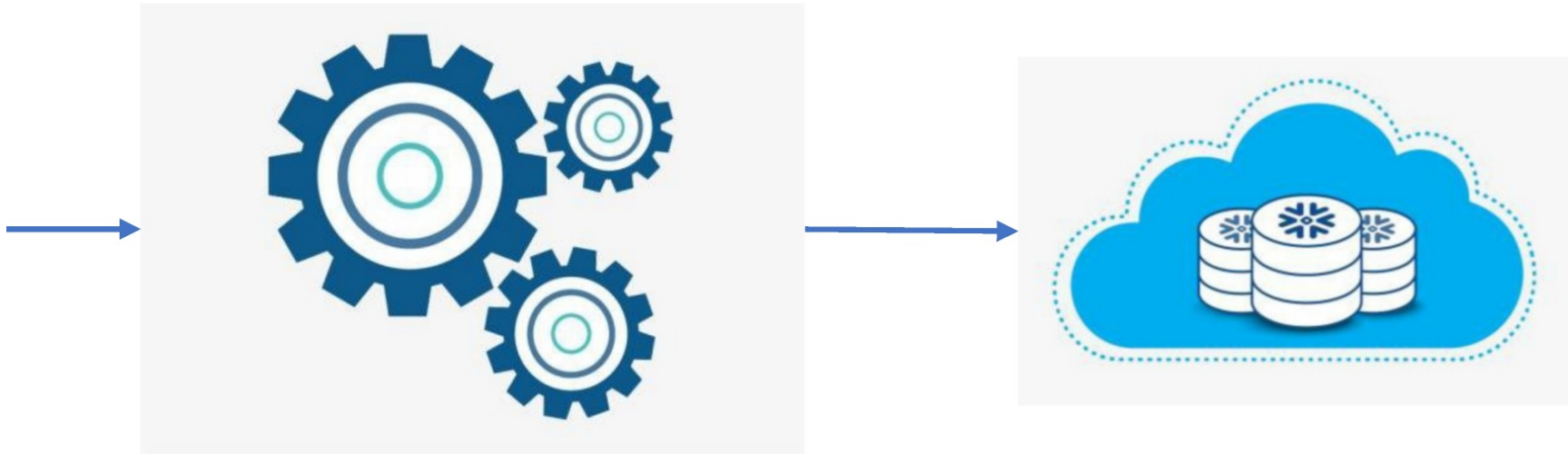
17





Transform Data

Data Warehouse



Traditional Approach for ETL

Data Extraction:

- Data Source – Relational Databases
- Way of extraction- Batch

Data Transformation:

- Location for transformation– On premises
- Flexibility- less
- Scalability- difficult

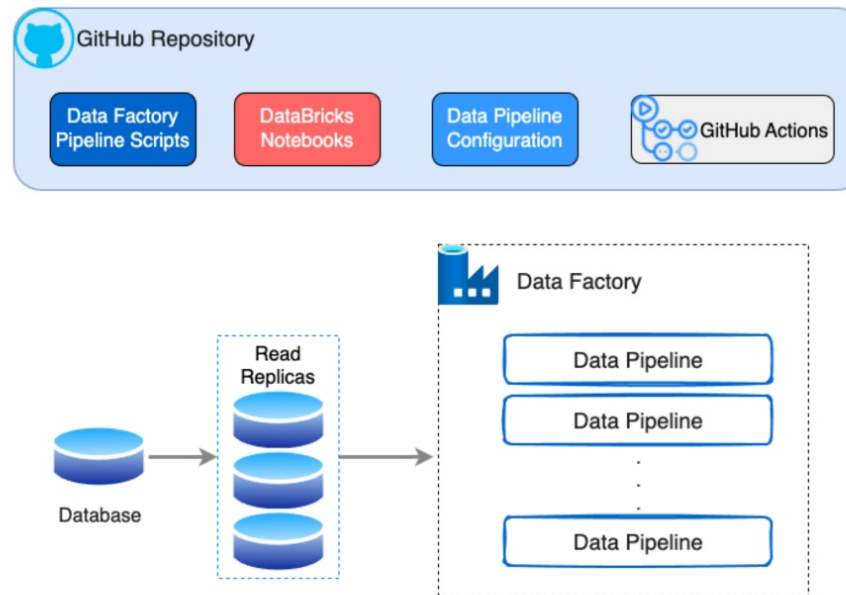
Data Loading:

- Analysis impact- slows down
- Expenses- hardware to suit on-site processing
- Error prone



External Use © Datics Consulting

ETL - DataOps



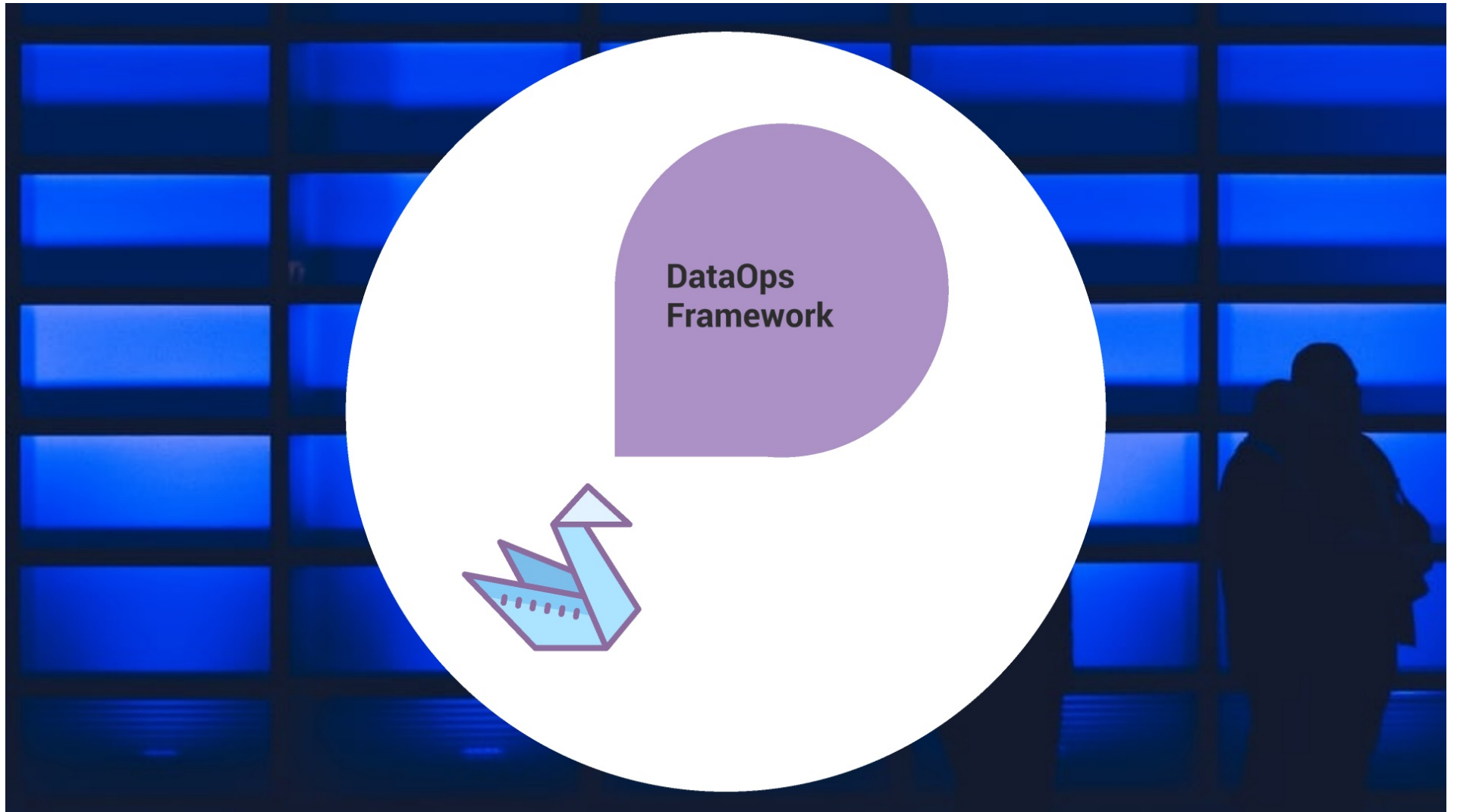
External Use © Datics Consulting

19

DataOps



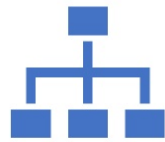
by Datics



Why every data team needs a DataOps Framework



Why every data team needs a DataOps Framework



Data team should focus on dealing with data, not the infrastructure and IT system operation

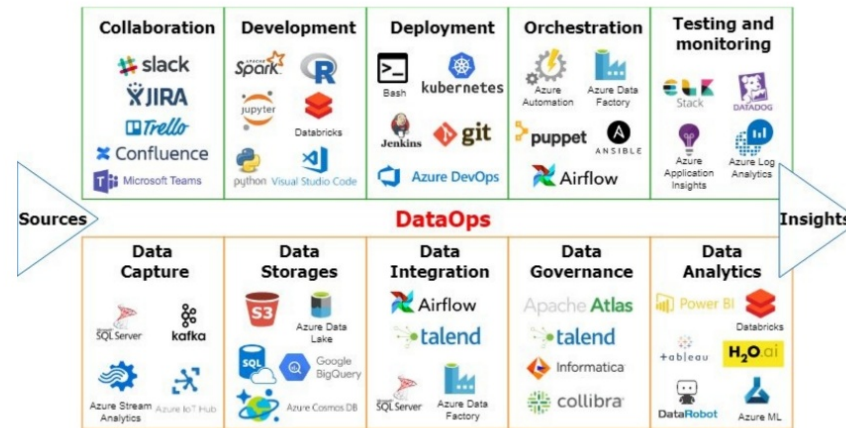


DataOps framework gives a set of tools and technologies >> by using framework, data team can start to work on data faster



Tooling

- DataOps uses DevOps methodology building a DataOps framework, which automizes and shortens the time of data orchestration operation
- So..... Hundred tools to choose



Why every data team needs a DataOps Engineer

- Collaborate with the data team and help to build the framework based on the team's favorite data tools and also suggested tools from DataOps engineers.

Why every data team needs a DataOps Engineer

- Collaborate with the data team and help to build the framework based on the team's favorite data tools and also suggested tools from DataOps engineers.



Why every data team needs a DataOps Engineer

- Collaborate with the data team and help to build the framework based on the team's favorite data tools and also suggested tools from DataOps engineers.



DataOps



by Datics

Conclusion:

Conclusion:



Faster cycle time

Conclusion:



Faster cycle time



Fewer data defects

Conclusion:



Faster cycle time



Fewer data defects



Greater code reuse

Conclusion: More efficiency and agility



Faster cycle time



Fewer data defects



Greater code reuse

Data●ps



by Datics

DataOps



by Datics