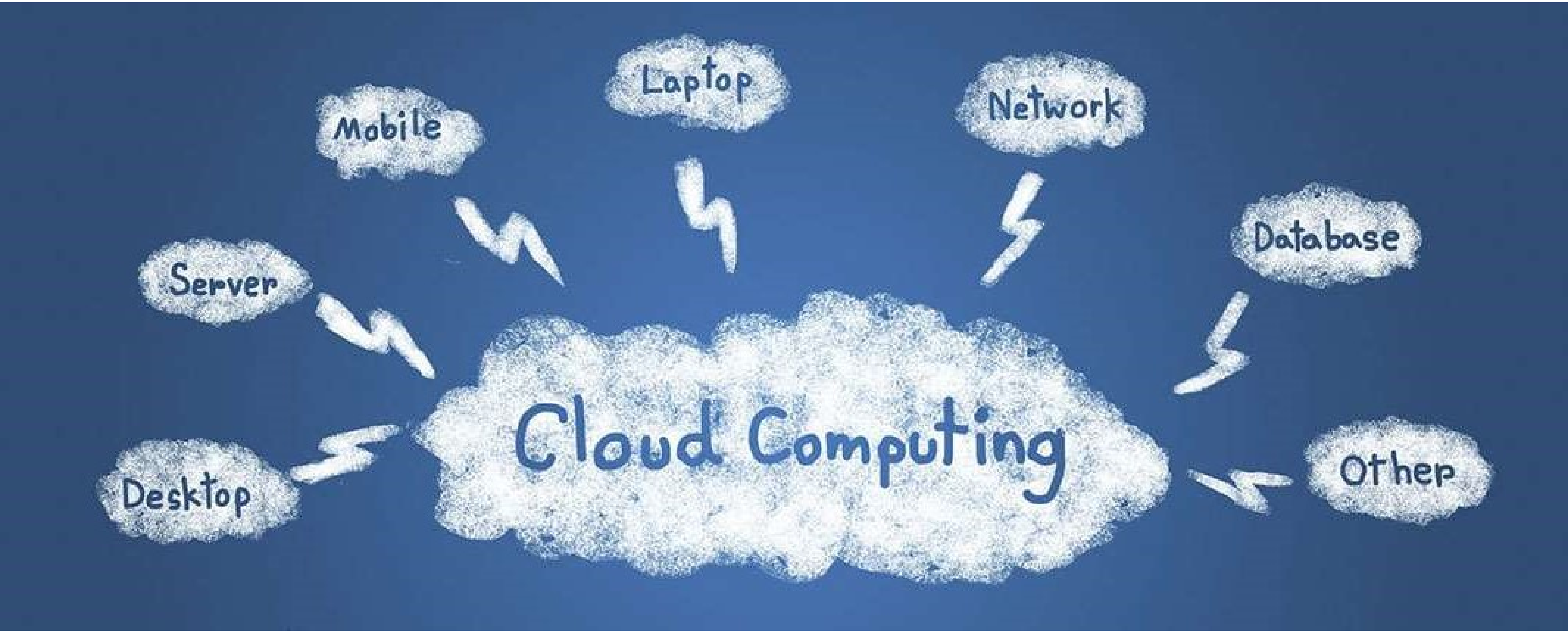
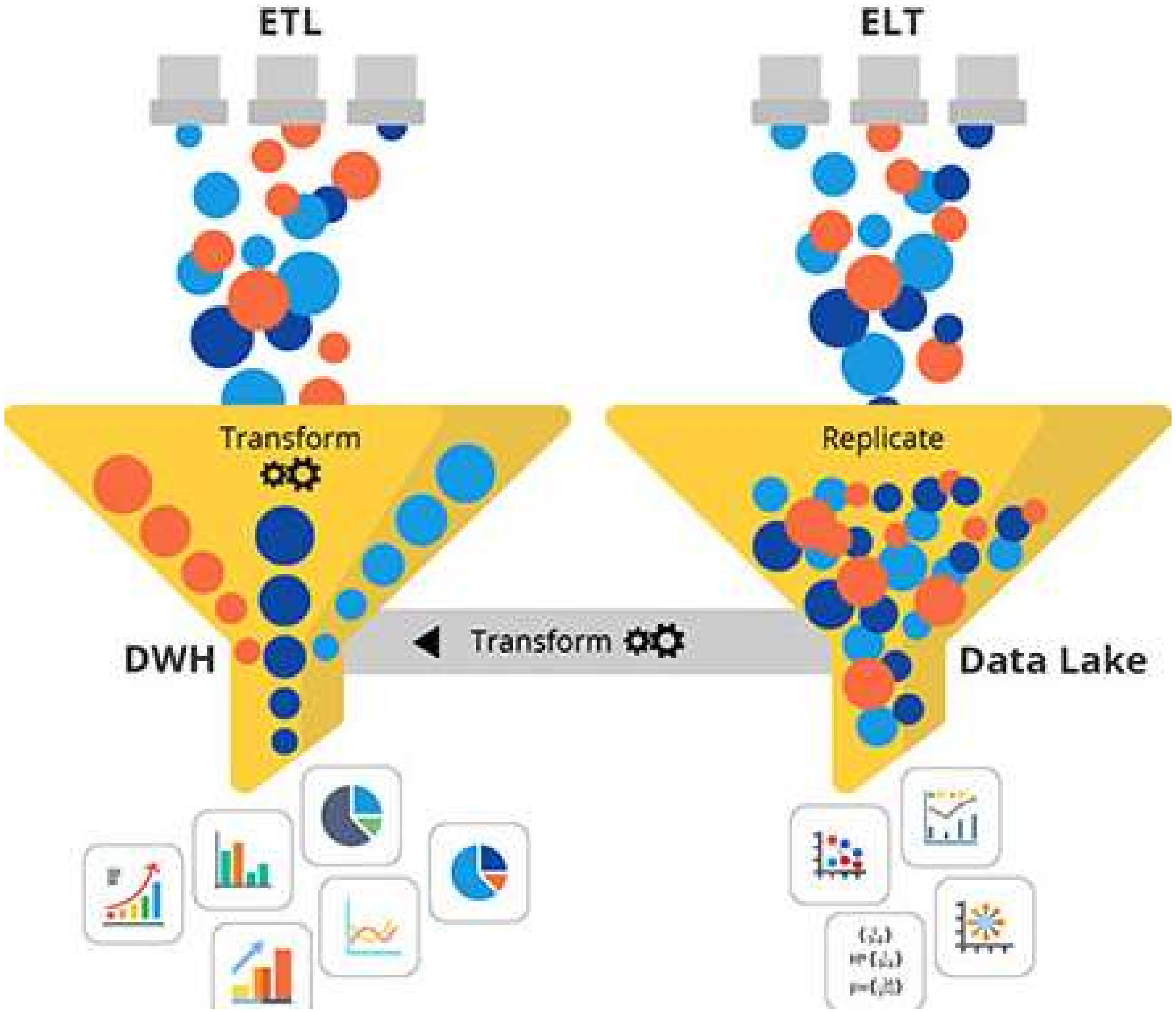
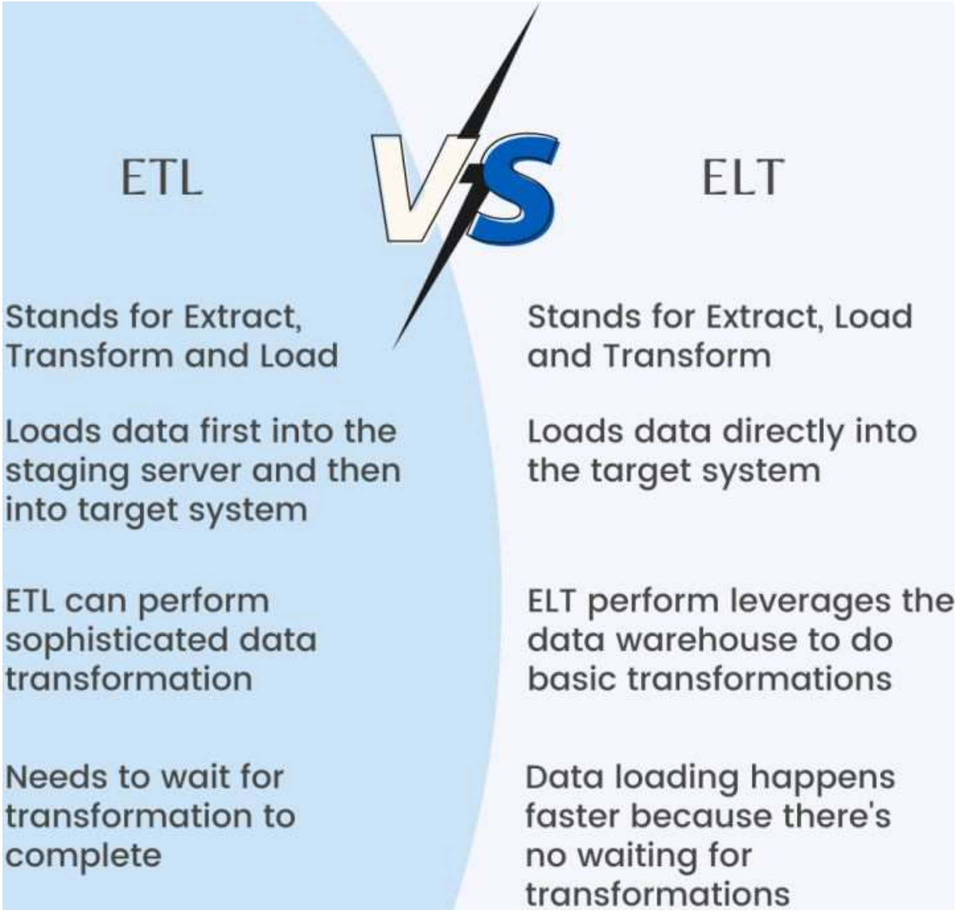
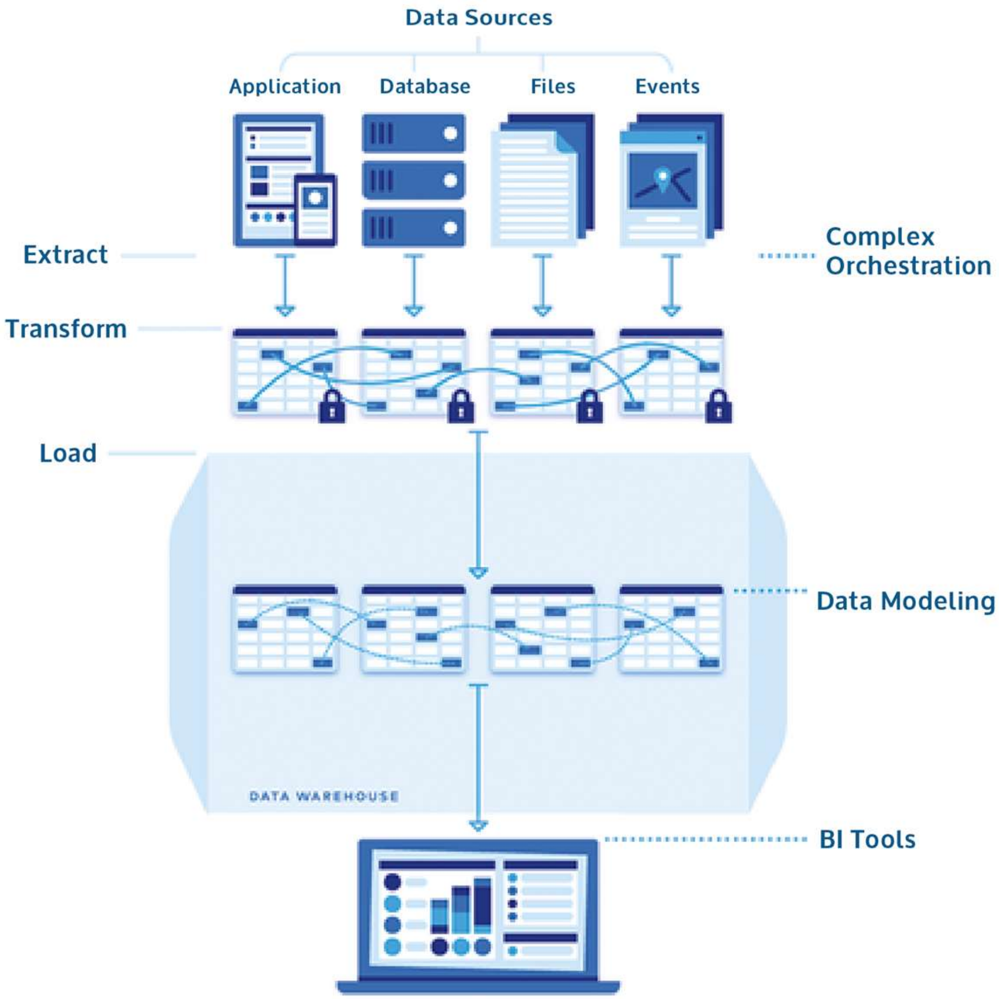

Laptop
Mobile
Network
Database
Server
Cloud Computing
Desktop
Other


ETL
ELT
Transform
Replicate
DWH
Transform
Data Lake

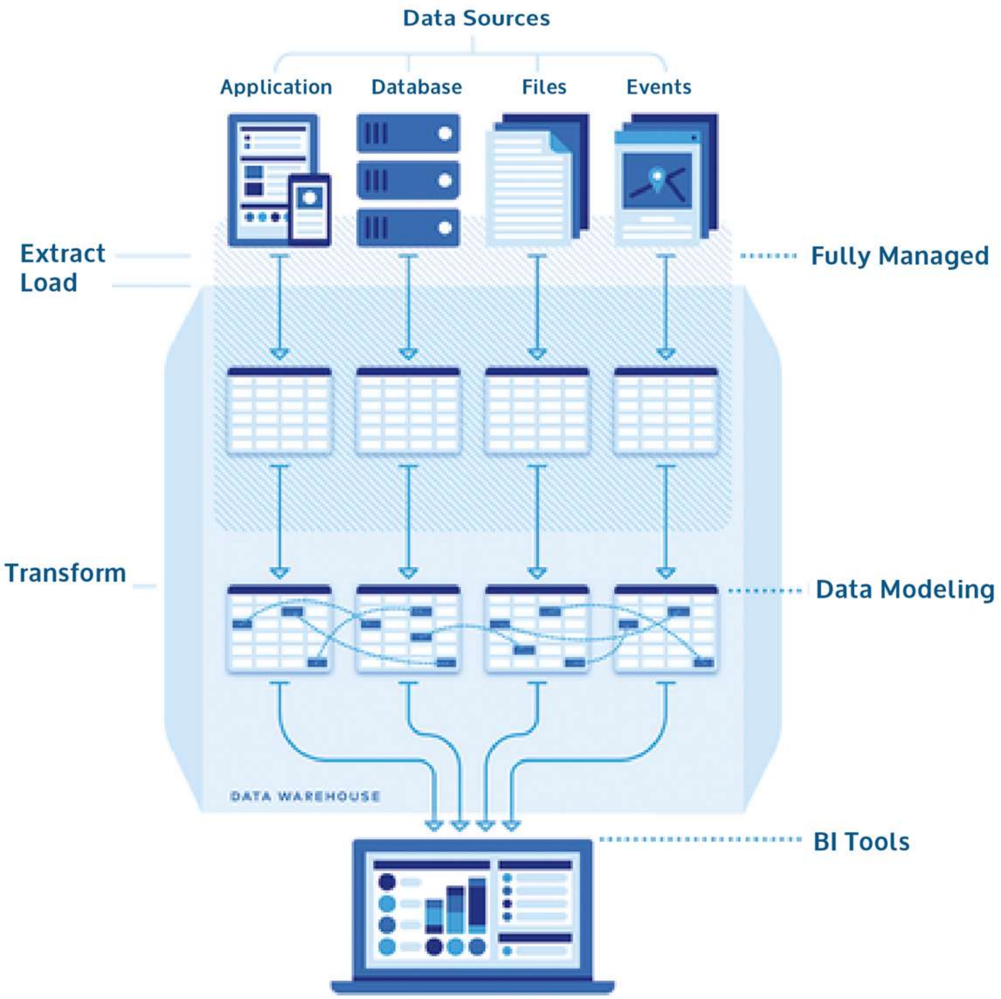


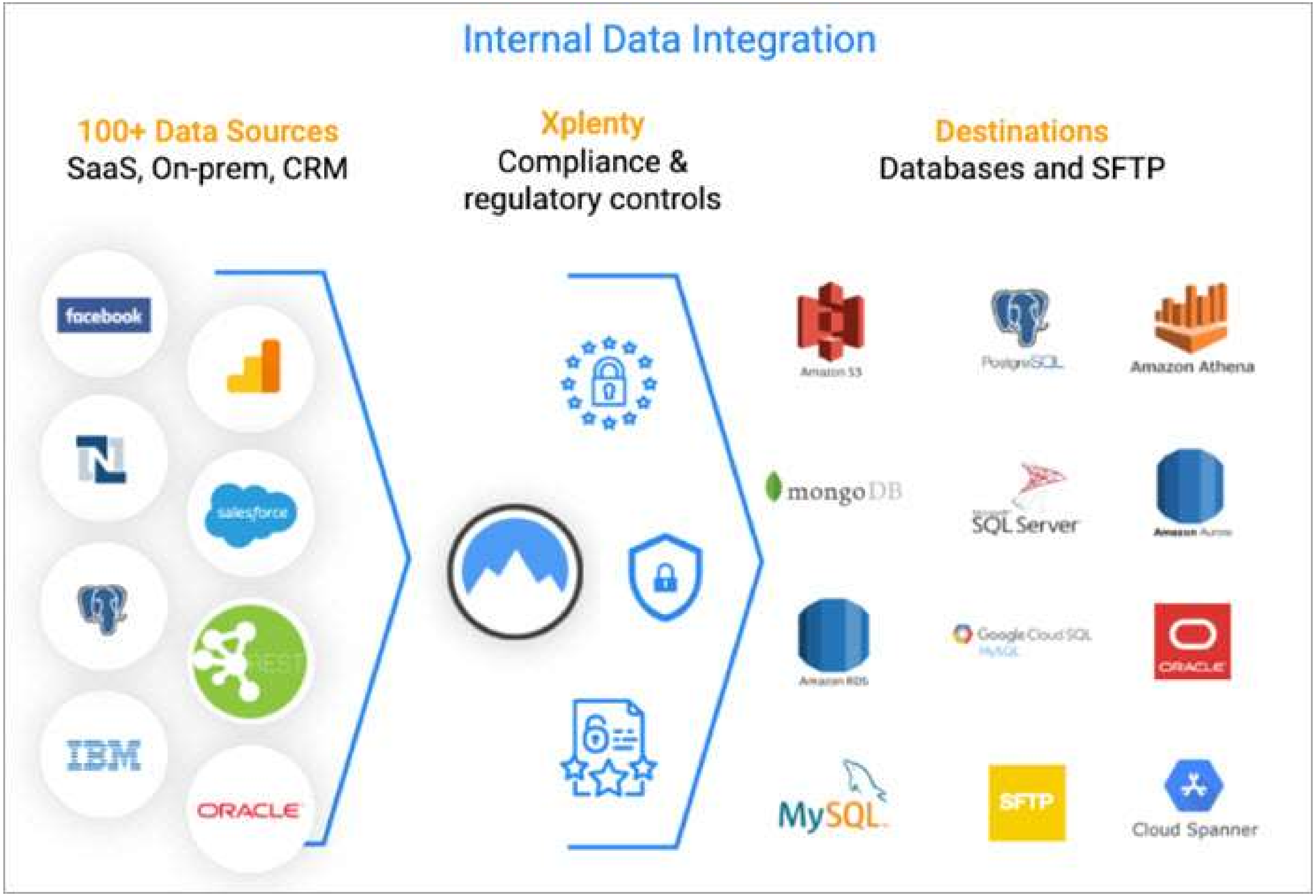
ETL	ELT
Stands for Extract, Transform and Load	Stands for Extract, Load and Transform
Loads data first into the staging server and then into target system	Loads data directly into the target system
ETL can perform sophisticated data transformation	ELT perform leverages the data warehouse to do basic transformations
Needs to wait for transformation to complete	Data loading happens faster because there's no waiting for transformations

Outdated ETL


Data Sources
Application
Database
Files
Events
Extract
Complex
Orchestration
Transform
Load
Data Modeling
DATA WAREHOUSE
BI Tools

Modern ELT


Data Sources
Application
Database
Files
Events
Extract
Load
Fully Managed
Transform
Data Modeling
DATA WAREHOUSE
BI Tools



Internal Data Integration

100+ Data Sources
SaaS, On-prem, CRM

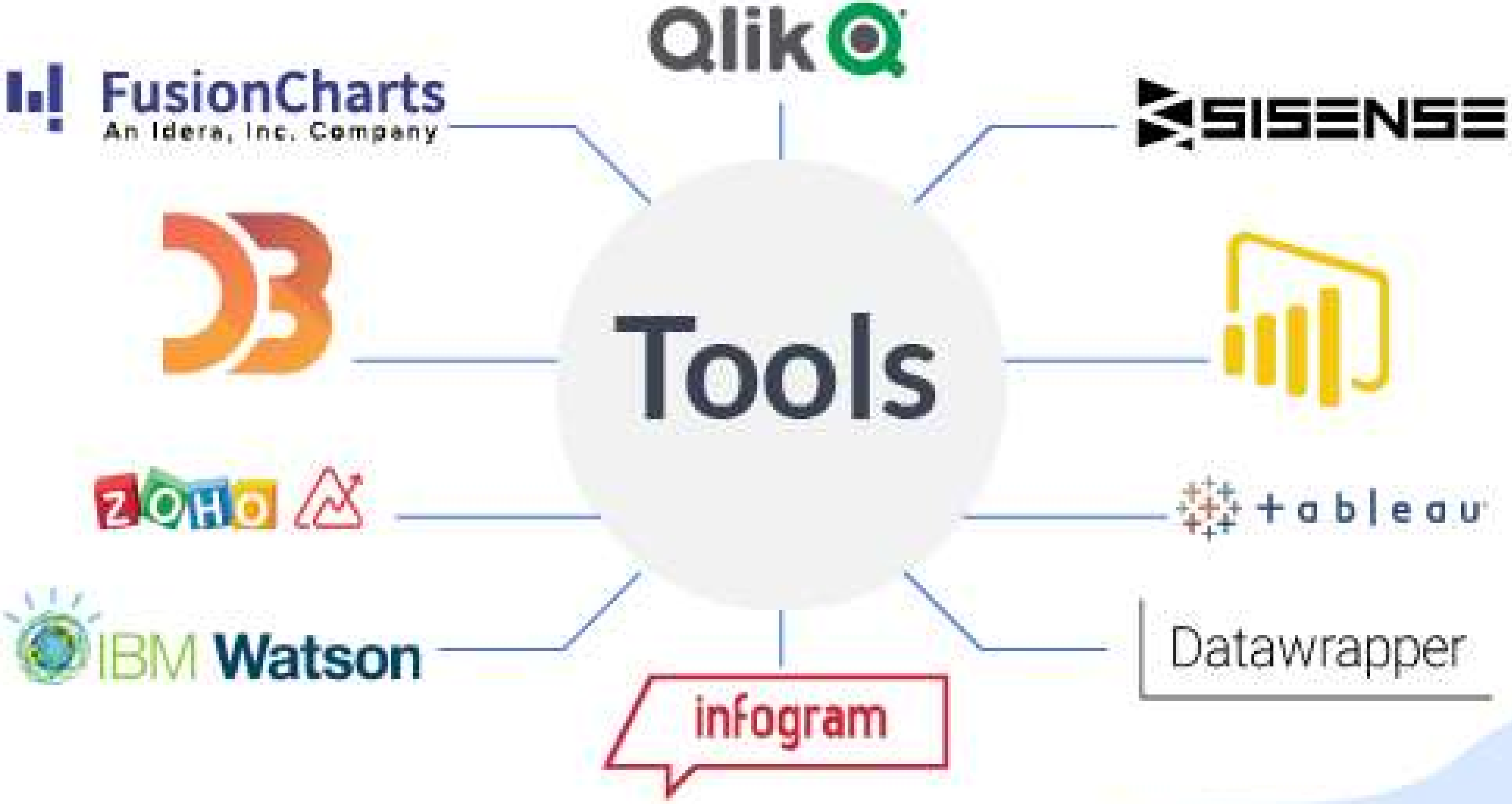
Xplenty
Compliance &
regulatory controls

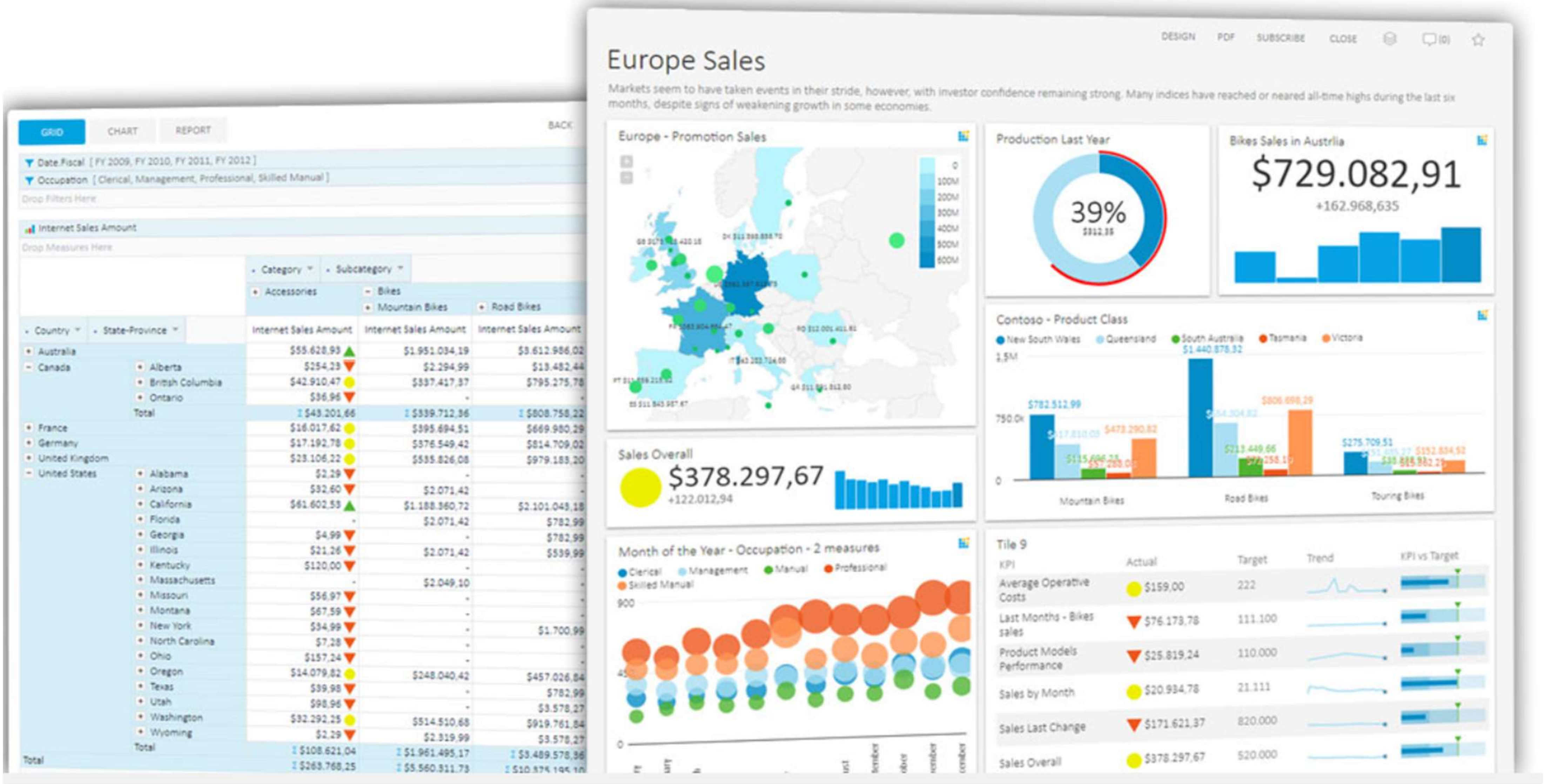
Destinations
Databases and SFTP

facebook
salesforce
REST
IBM
ORACLE
Amazon S3
PostgreSQL
Amazon Athena
mongoDB
SQL Server
Amazon RDS
Google Cloud SQL
MySQL
ORACLE
MySQL
SFTP
Cloud Spanner

Data Visualization Tools

XENONSTACK


Qlik
FusionCharts
An Idera, Inc. Company
SISENSE
Tools
tableau
ZOHO
IBM Watson
Datawrapper
infogram


GRID
CHART
REPORT
BACK
Drop Filters Here
Internet Sales Amount
Drop Measures Here
DESIGN
PDF
SUBSCRIBE
CLOSE
(0)
Europe - Promotion Sales
0
100M
200M
300M
400M
500M
600M
Production Last Year
39%
$312,35
Bikes Sales in Austrlia
$729.082,91
+162.968,635
Contoso - Product Class
New South Wales
Queensland
South Australia
Tasmania
Victoria
1,5M
750,0k
0
$1.440.878,32
$782.512,99
$806.898,29
$473.290,82
$275.709,51
$152.834,52
Mountain Bikes
Road Bikes
Touring Bikes
Sales Overall
$378.297,67
+122.012,94
Month of the Year - Occupation - 2 measures
Clerical
Management
Manual
Professional
Skilled Manual
900
0

Date.Fiscal [FY 2009, FY 2010, FY 2011, FY 2012]
Occupation [Clerical, Management, Professional, Skilled Manual]

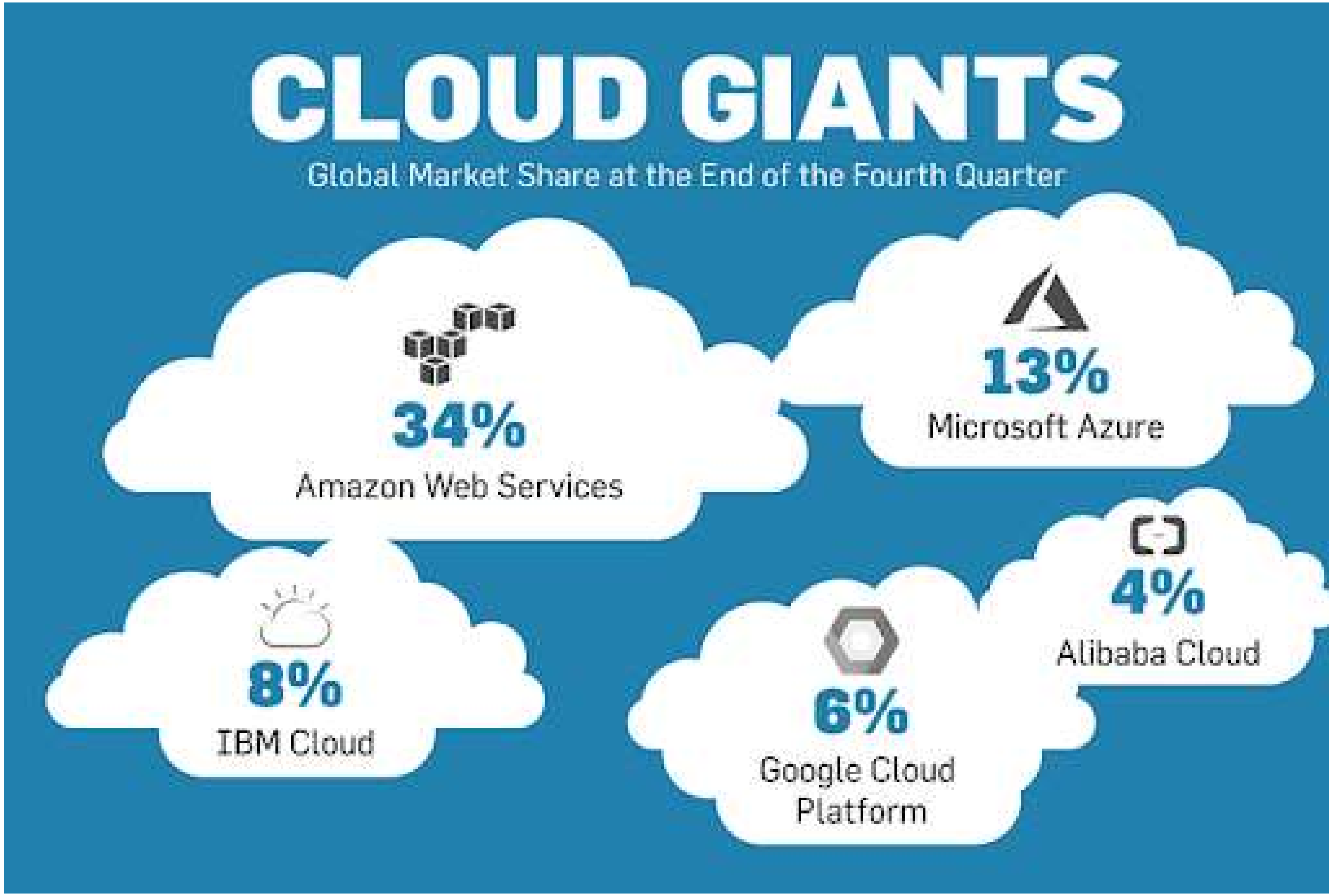
Country	State-Province	Category: Accessories	Bikes: Mountain Bikes	Bikes: Road Bikes
		Internet Sales Amount	Internet Sales Amount	Internet Sales Amount
Australia		$55.628,93 ▲	$1.951.034,19	$3.612.986,02
Canada	Alberta	$254,23 ▼	$2.294,99	$13.482,44
	British Columbia	$42.910,47 ●	$337.417,37	$795.275,78
	Ontario	$36,96 ▼	-	-
	Total	$43.201,66	$339.712,36	$808.758,22
France		$16.017,62 ●	$395.694,51	$669.980,29
Germany		$17.192,78 ●	$376.549,42	$814.709,02
United Kingdom		$23.106,22 ●	$535.826,08	$979.183,20
United States	Alabama	$2,29 ▼	-	-
	Arizona	$32,60 ▼	$2.071,42	-
	California	$61.602,53 ▲	$1.188.360,72	$2.101.043,18
	Florida	-	$2.071,42	$782,99
	Georgia	$4,99 ▼	-	$782,99
	Illinois	$21,26 ▼	$2.071,42	$539,99
	Kentucky	$120,00 ▼	-	-
	Massachusetts	-	$2.049,10	-
	Missouri	$56,97 ▼	-	-
	Montana	$67,59 ▼	-	-
	New York	$34,99 ▼	-	$1.700,99
	North Carolina	$7,28 ▼	-	-
	Ohio	$157,24 ▼	-	-
	Oregon	$14.079,82 ●	$248.040,42	$457.026,84
	Texas	$39,98 ▼	-	$782,99
	Utah	$98,96 ▼	-	$3.578,27
	Washington	$32.292,25 ●	$514.510,68	$919.761,84
	Wyoming	$2,29 ▼	$2.319,99	$3.578,27
	Total	$108.621,04	$1.961.495,17	$3.489.578,36
Total		$263.768,25	$5.560.311,73	$10.375.195,10

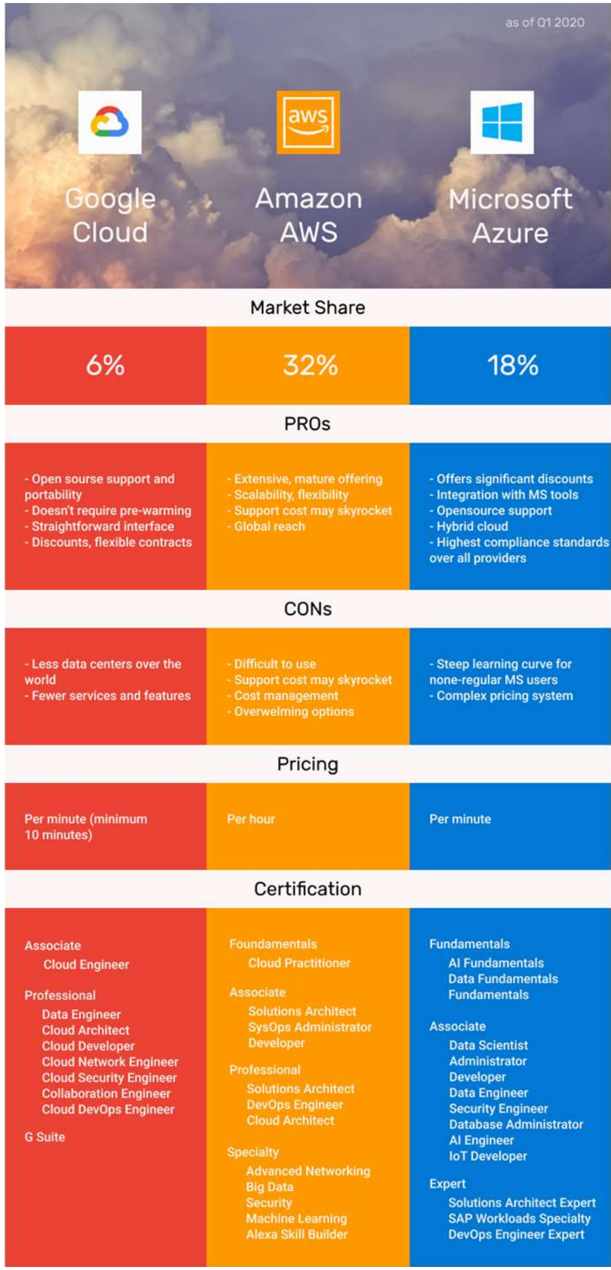
Europe Sales

Markets seem to have taken events in their stride, however, with investor confidence remaining strong. Many indices have reached or neared all-time highs during the last six months, despite signs of weakening growth in some economies.

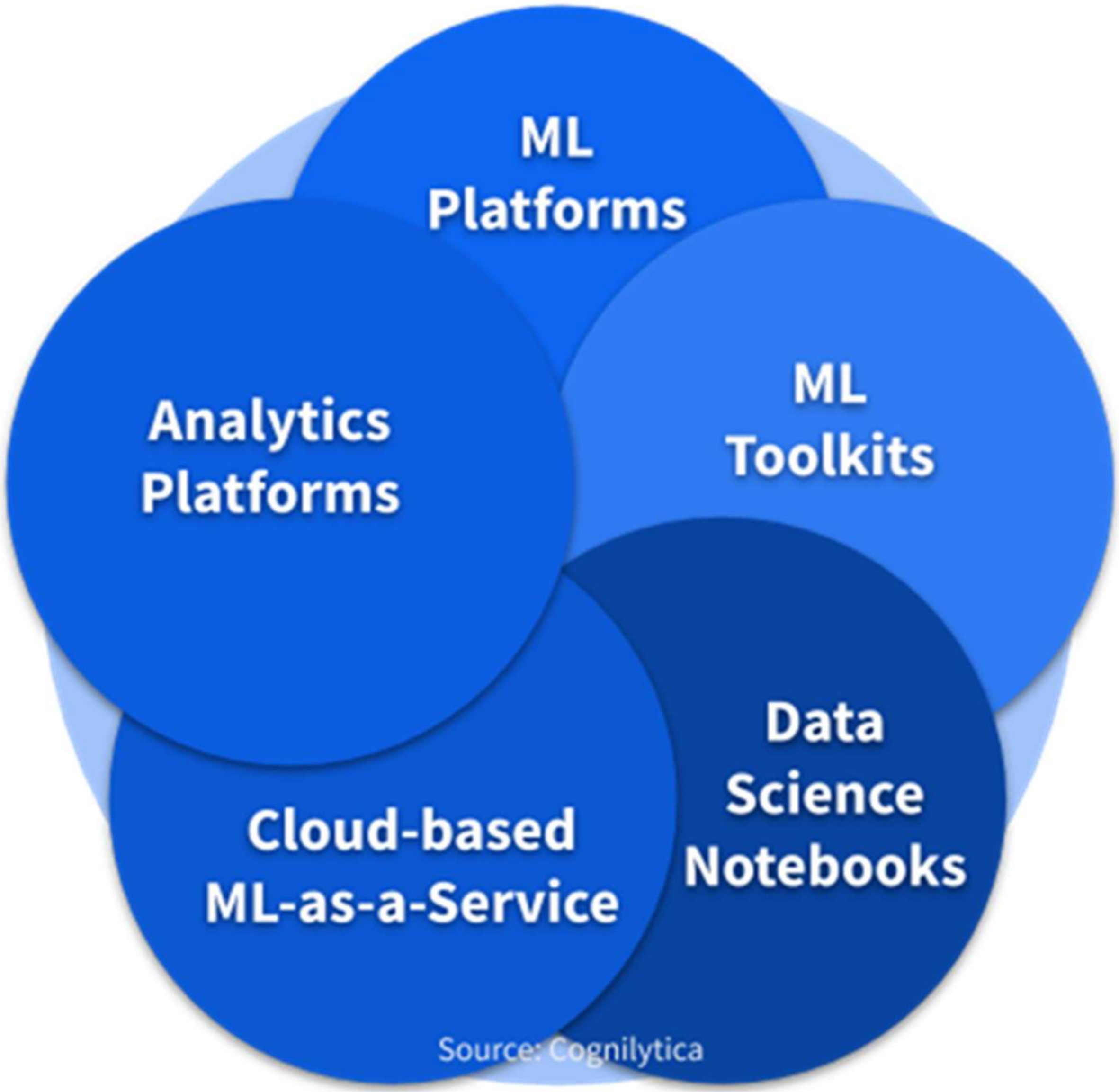
Tile 9

KPI	Actual	Target	Trend	KPI vs Target
Average Operative Costs	● $159,00	222		
Last Months - Bikes sales	▼ $76.173,78	111.100		
Product Models Performance	▼ $25.819,24	110.000		
Sales by Month	● $20.934,78	21.111		
Sales Last Change	▼ $171.621,37	820.000		
Sales Overall	● $378.297,67	520.000		

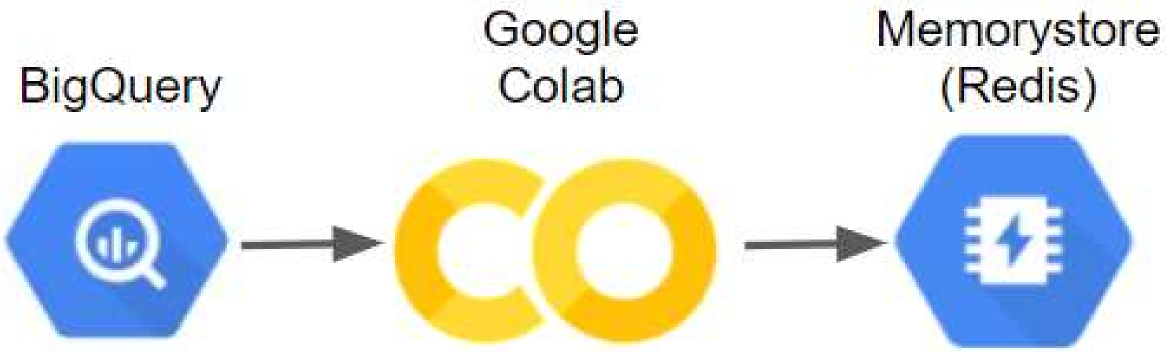


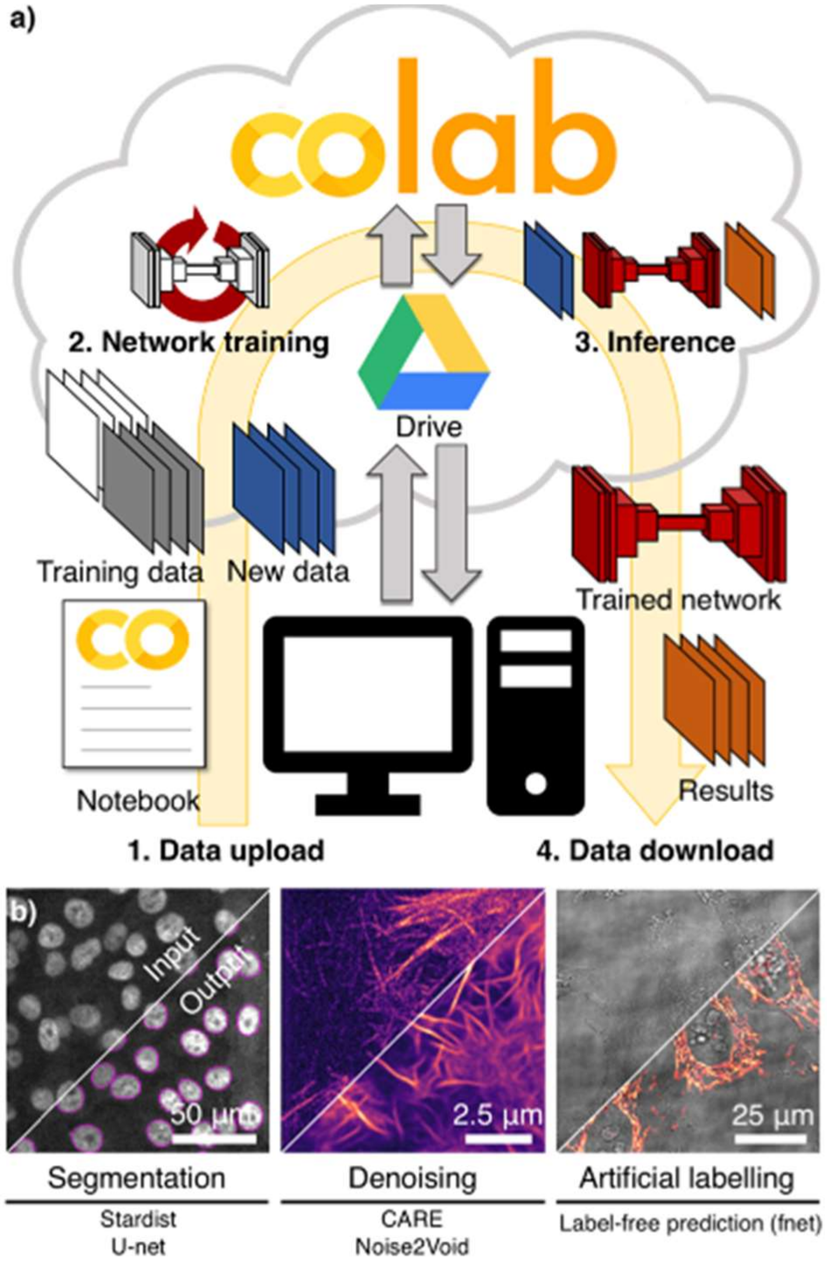

as of Q1 2020

	Google Cloud	Amazon AWS	Microsoft Azure
Market Share	6%	32%	18%
PROs	- Open sourse support and portability - Doesn't require pre-warming - Straightforward interface - Discounts, flexible contracts	- Extensive, mature offering - Scalability, flexibility - Support cost may skyrocket - Global reach	- Offers significant discounts - Integration with MS tools - Opensource support - Hybrid cloud - Highest compliance standards over all providers
CONs	- Less data centers over the world - Fewer services and features	- Difficult to use - Support cost may skyrocket - Cost management - Overwelming options	- Steep learning curve for none-regular MS users - Complex pricing system
Pricing	Per minute (minimum 10 minutes)	Per hour	Per minute
Certification	Associate Cloud Engineer Professional Data Engineer Cloud Architect Cloud Developer Cloud Network Engineer Cloud Security Engineer Collaboration Engineer Cloud DevOps Engineer G Suite	Foundamentals Cloud Practitioner Associate Solutions Architect SysOps Administrator Developer Professional Solutions Architect DevOps Engineer Cloud Architect Specialty Advanced Networking Big Data Security Machine Learning Alexa Skill Builder	Fundamentals AI Fundamentals Data Fundamentals Fundamentals Associate Data Scientist Administrator Developer Data Engineer Security Engineer Database Administrator AI Engineer IoT Developer Expert Solutions Architect Expert SAP Workloads Specialty DevOps Engineer Expert


ML Platforms
Analytics Platforms
ML Toolkits
Cloud-based ML-as-a-Service
Data Science Notebooks

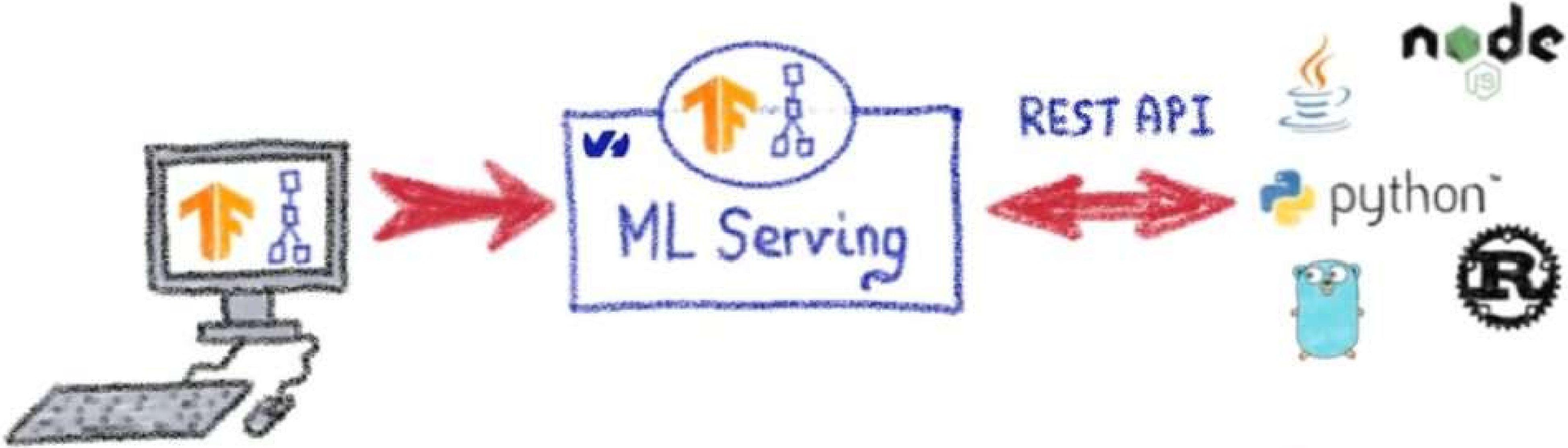
Source: Cognilytica


BigQuery
Google Colab
Memorystore (Redis)

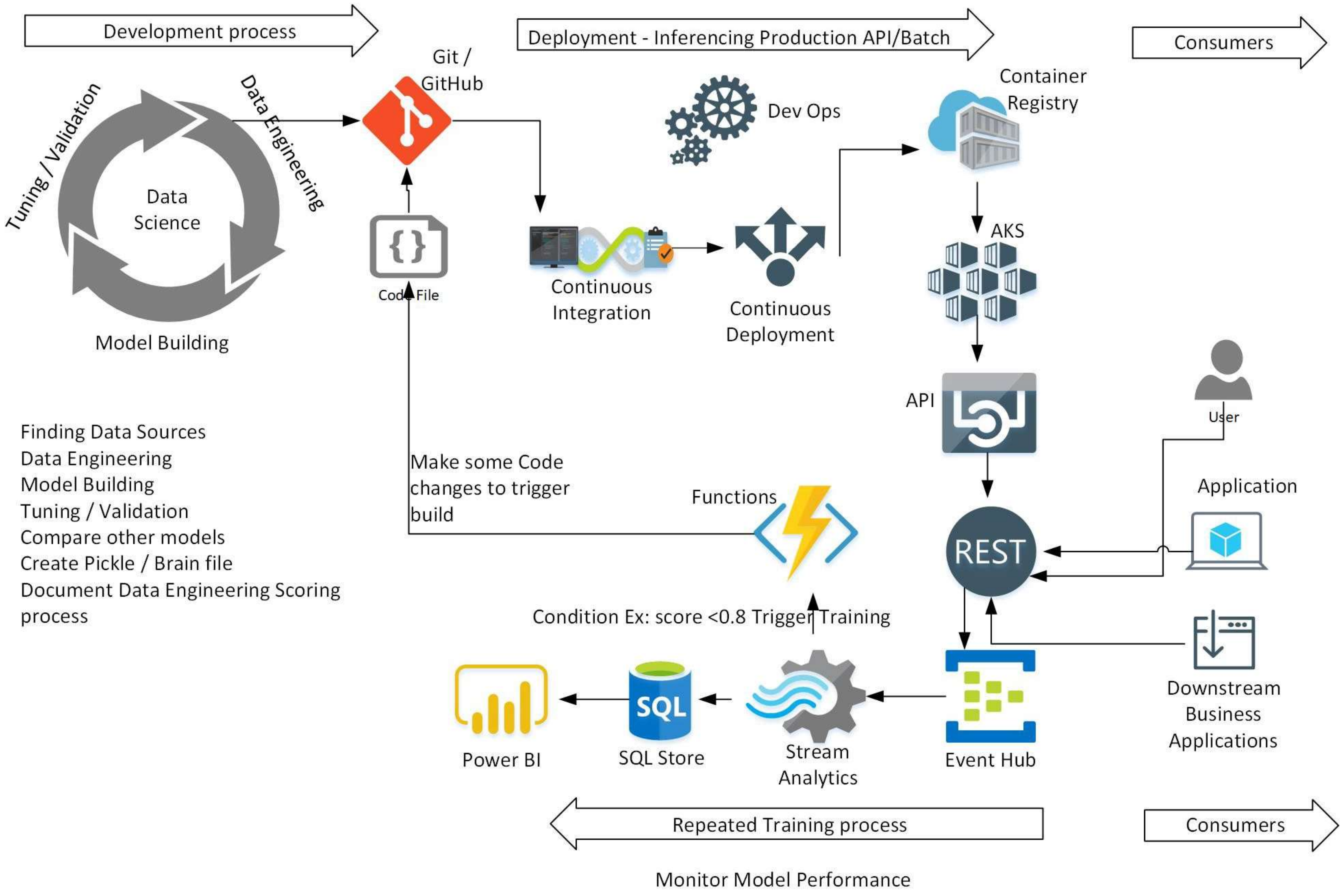

a)
colab
2. Network training
Drive
3. Inference
Training data
New data
Trained network
Notebook
Results
1. Data upload
4. Data download
b)
Input
Output
50 µm
2.5 µm
25 µm
Segmentation
Stardist
U-net
Denoising
CARE
Noise2Void
Artificial labelling
Label-free prediction (fnet)

ML Serving

A cloud based tool
to deploy Machine Learning

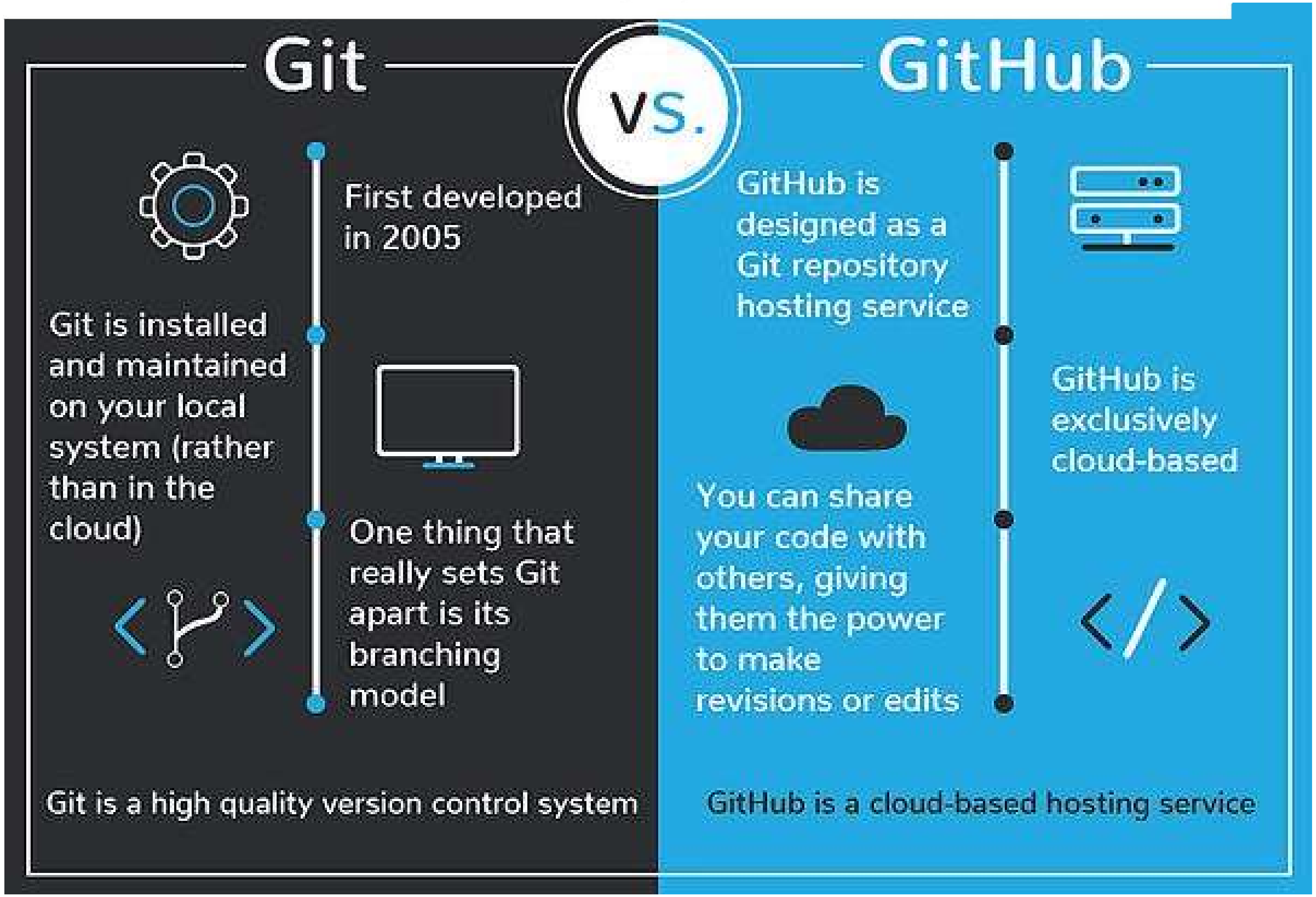

ML Serving
REST API
node
JS
python

Machine Learning deployment and Re-Training deployment with CI/CD using Azure DevOps


Development process
Deployment - Inferencing Production API/Batch
Consumers
Tuning / Validation
Data Engineering
Data Science
Model Building
Git / GitHub
Code File
Dev Ops
Container Registry
Continuous Integration
Continuous Deployment
AKS
API
User
Make some Code changes to trigger build
Functions
Application
REST
Condition Ex: score <0.8 Trigger Training
Downstream Business Applications
Power BI
SQL
SQL Store
Stream Analytics
Event Hub
Repeated Training process
Consumers
Monitor Model Performance

- Finding Data Sources
- Data Engineering
- Model Building
- Tuning / Validation
- Compare other models
- Create Pickle / Brain file
- Document Data Engineering Scoring process

git GitHub



Git VS. GitHub

Git

First developed in 2005

Git is installed and maintained on your local system (rather than in the cloud)

One thing that really sets Git apart is its branching model

Git is a high quality version control system

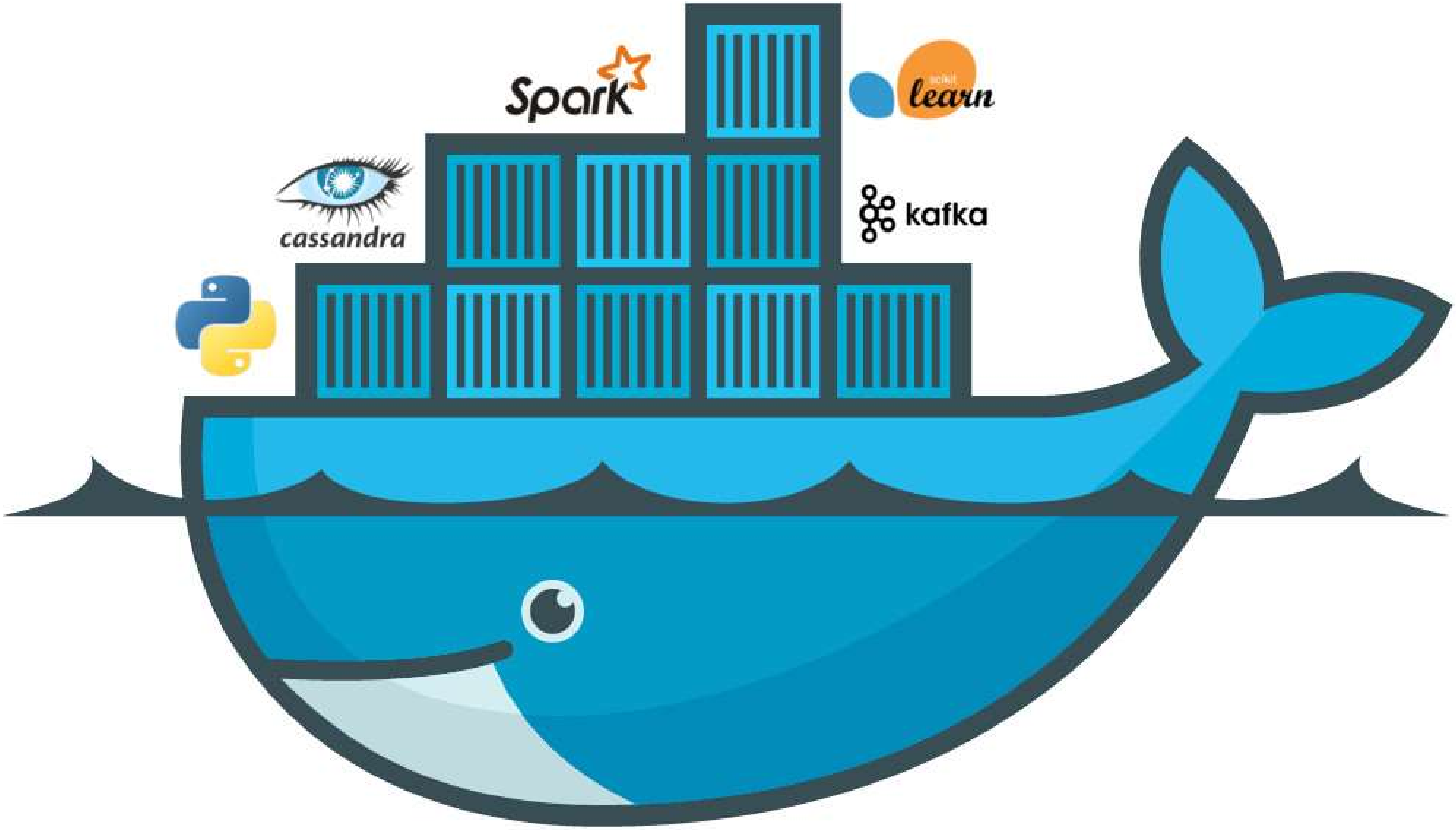
GitHub

GitHub is designed as a Git repository hosting service

GitHub is exclusively cloud-based

You can share your code with others, giving them the power to make revisions or edits

GitHub is a cloud-based hosting service

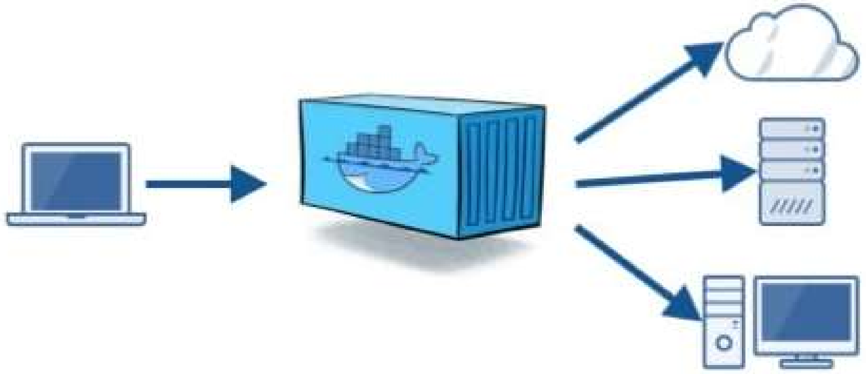

Spark
scikit
learn
cassandra
kafka

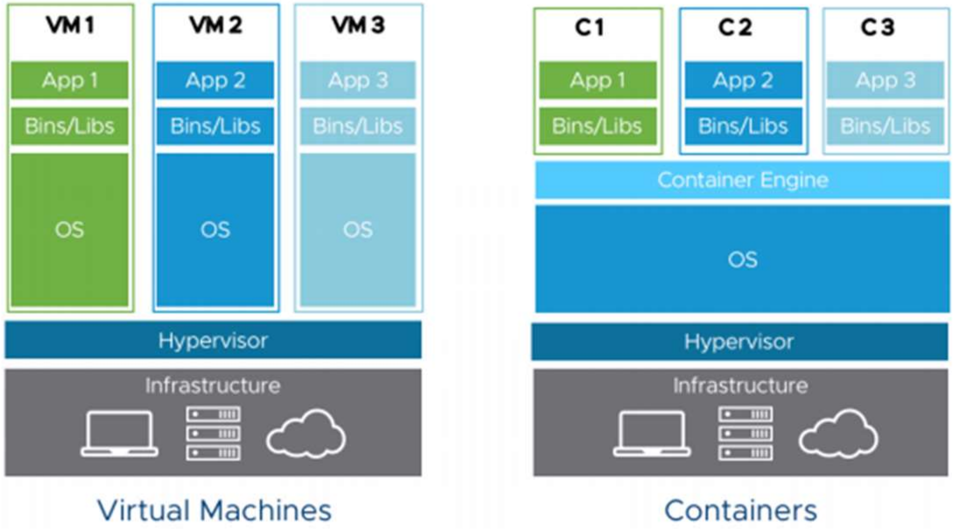
What is it?

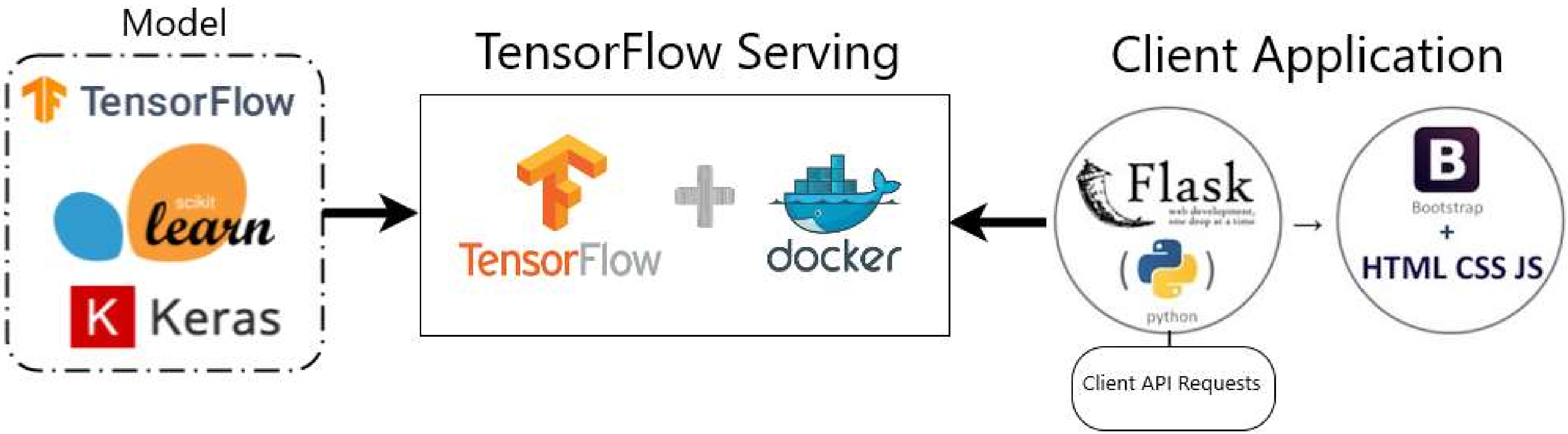
Why use it?

Build - Ship - Run

- Reproducible development environment
- Same container in dev, test and prod




VM 1
VM 2
VM 3
App 1
App 2
App 3
Bins/Libs
Bins/Libs
Bins/Libs
OS
OS
OS
Hypervisor
Infrastructure
Virtual Machines
C 1
C 2
C 3
App 1
App 2
App 3
Bins/Libs
Bins/Libs
Bins/Libs
Container Engine
OS
Hypervisor
Infrastructure
Containers


Model
TensorFlow
scikit learn
Keras
TensorFlow Serving
TensorFlow
docker
Client Application
Flask
web development, one drop at a time
python
Bootstrap
+
HTML CSS JS
Client API Requests