ATTUNITY

ATTUNITY

Emerging DataOps Principles and Technologies for Analytics

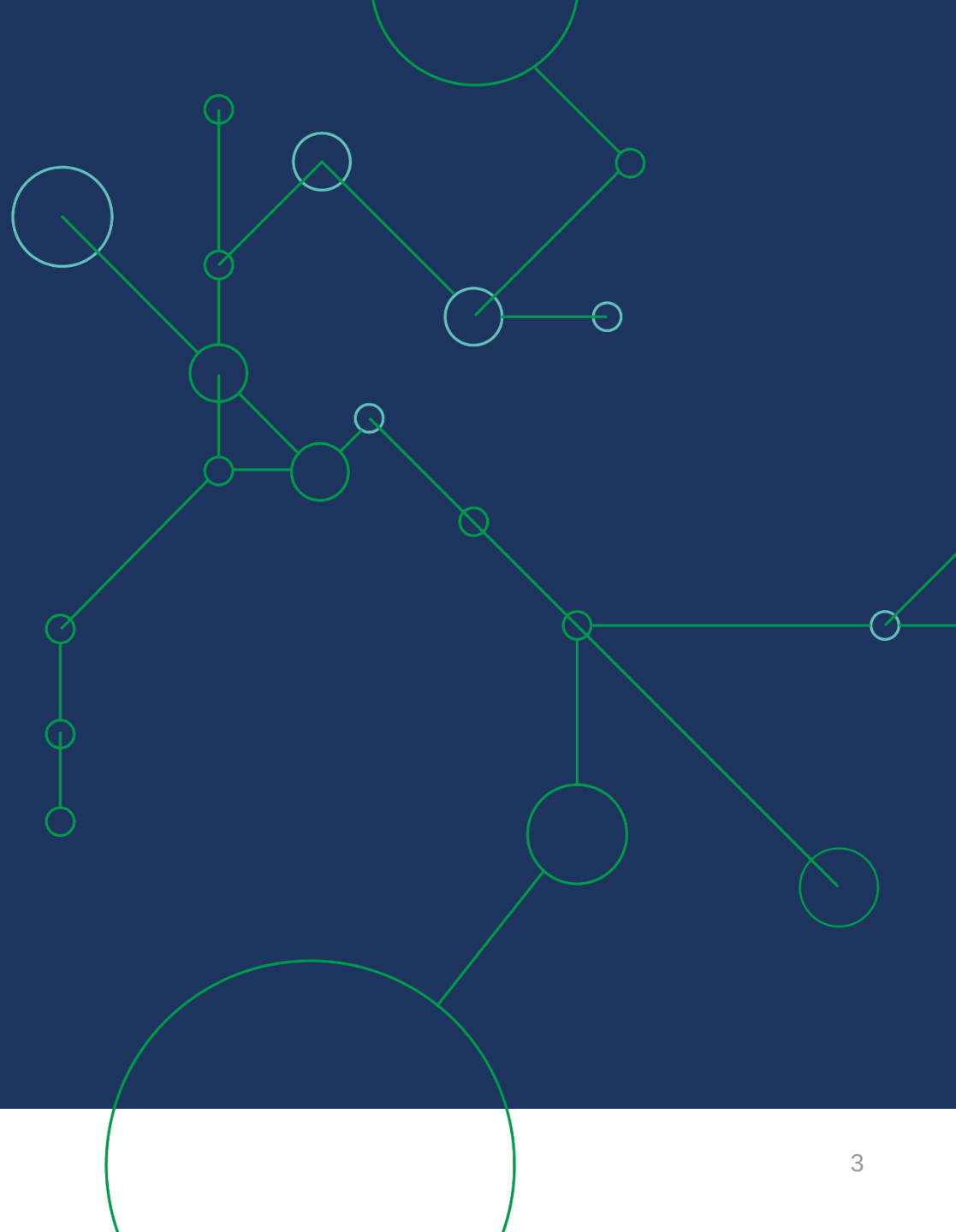
MIDWEST ARCHITECTURE COMMUNITY COLLABORATION

Tim Garrod

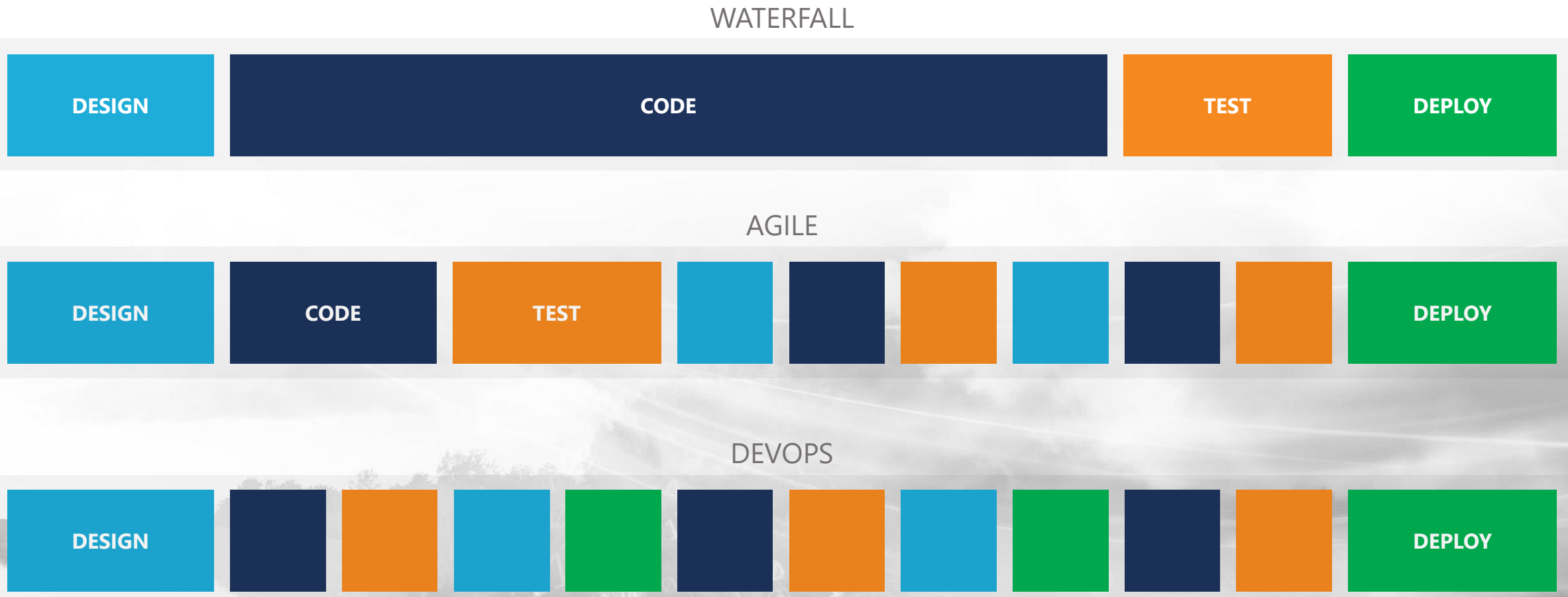
Agenda

- What is DataOps?
- What is driving DataOps?
- Enabling DataOps…
- DataOps technologies and functional requirements

What is DataOps?



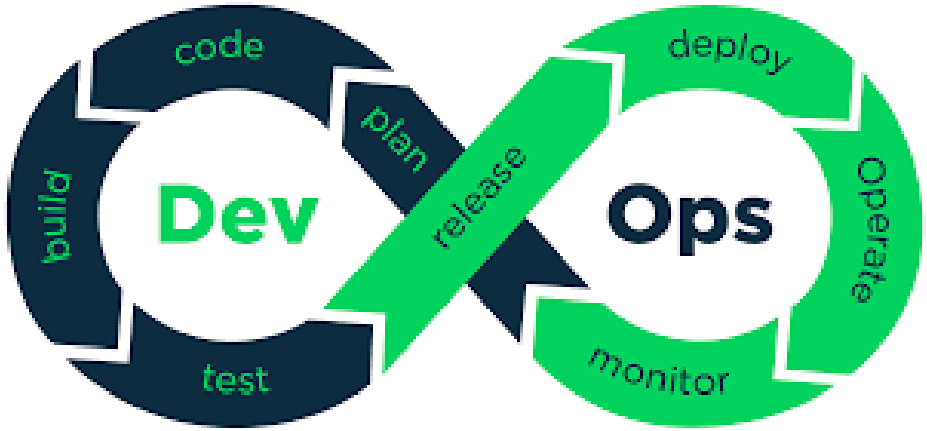
DevOps: Accelerating time to value



DevOps – a grandfather of DataOps

ATTUNITY

- **Built upon the Agile development movement**
- **Collaboration between development and operations**
- **Continuous integration & delivery frameworks**
- **…**



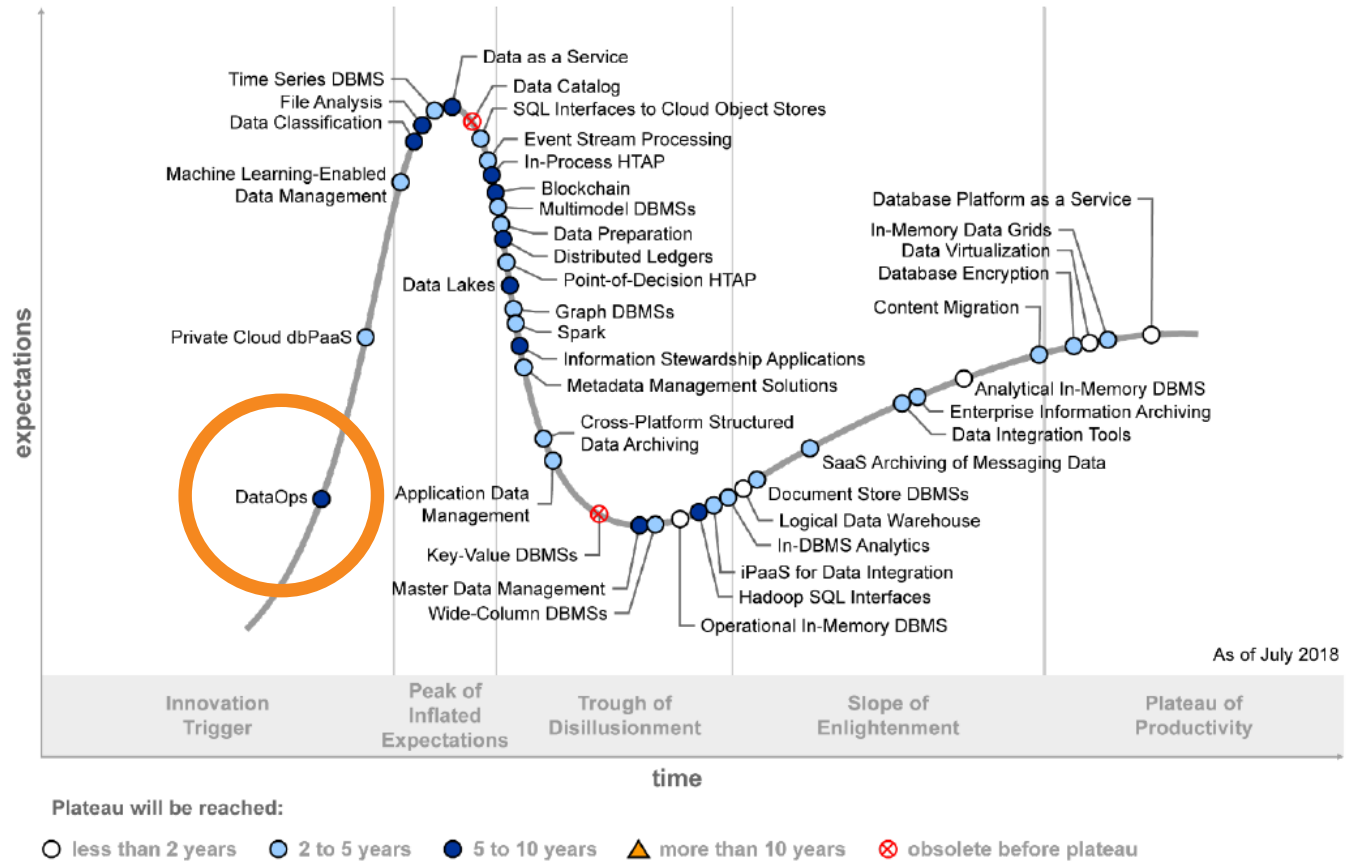
Agile, Collaborative & Continuous

WHAT IS DATAOPS?
WE ARE IN THE EARLY DAYS

ATTUNITY
A Division of Qlik

Figure 1. Hype Cycle for Data Management, 2018

Gartner



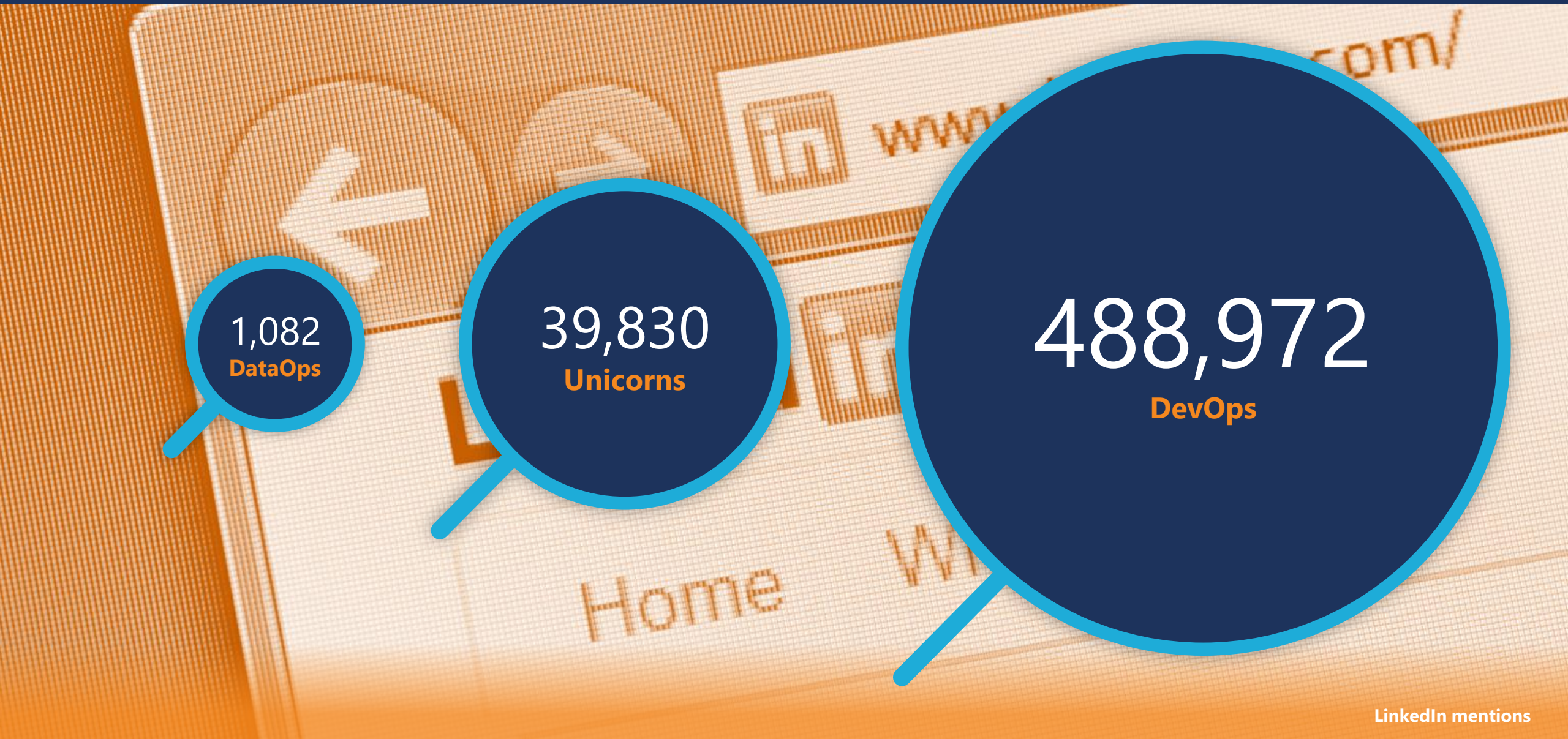
<1%

ADOPTION

Source: Gartner - Hype Cycle for Data Management, July 2018

LINKEDIN RESULTS

ATTUNITY
A Division of Qlik



LinkedIn mentions

WHAT IS DATAOPS

Emerging discipline to build and manage efficient, effective data pipelines

ATTUNITY
A Division of Qlik

PEOPLE PROCESS TECHNOLOGY

CODE

TOOLS

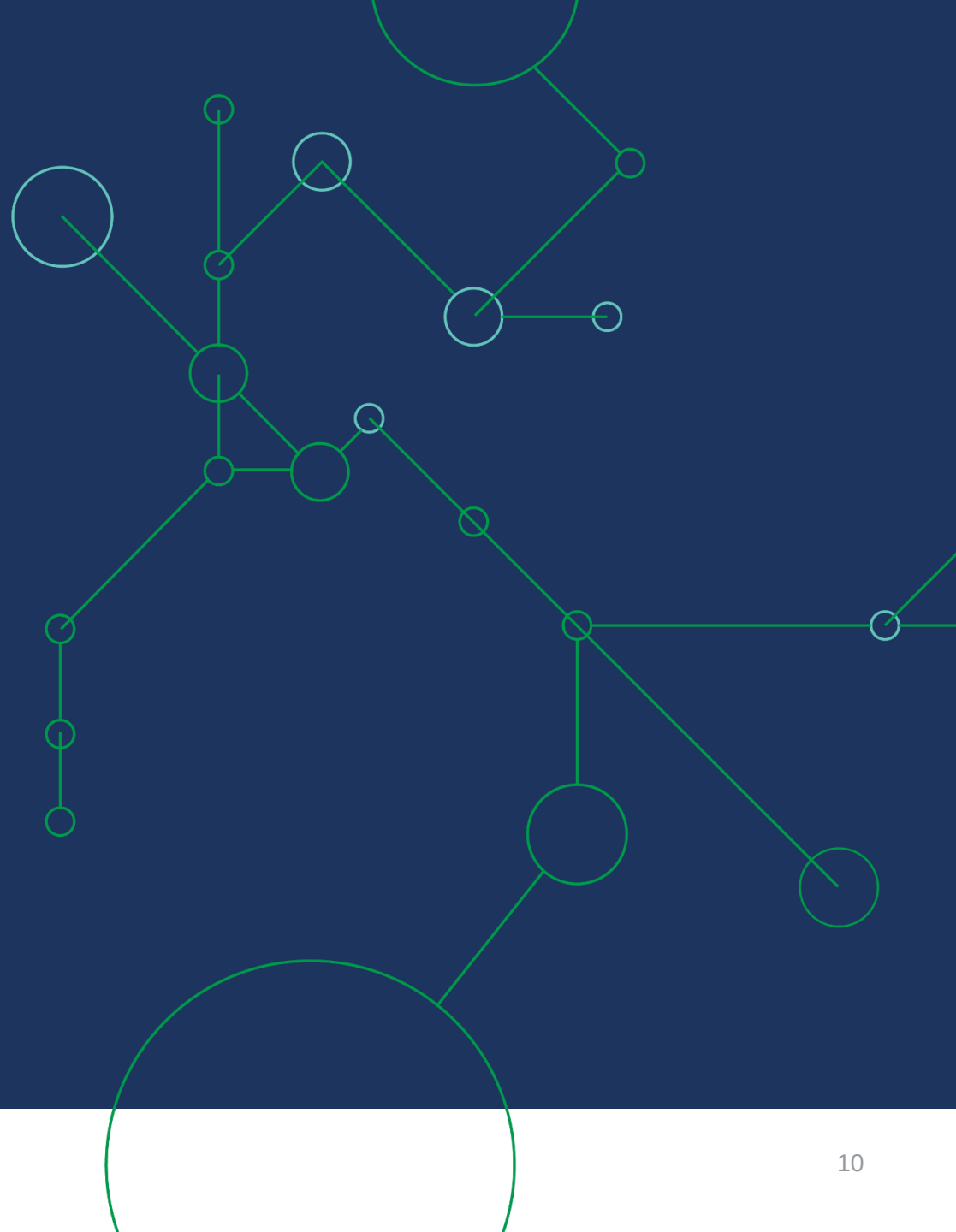
INFRASTRUCTURE

DATA

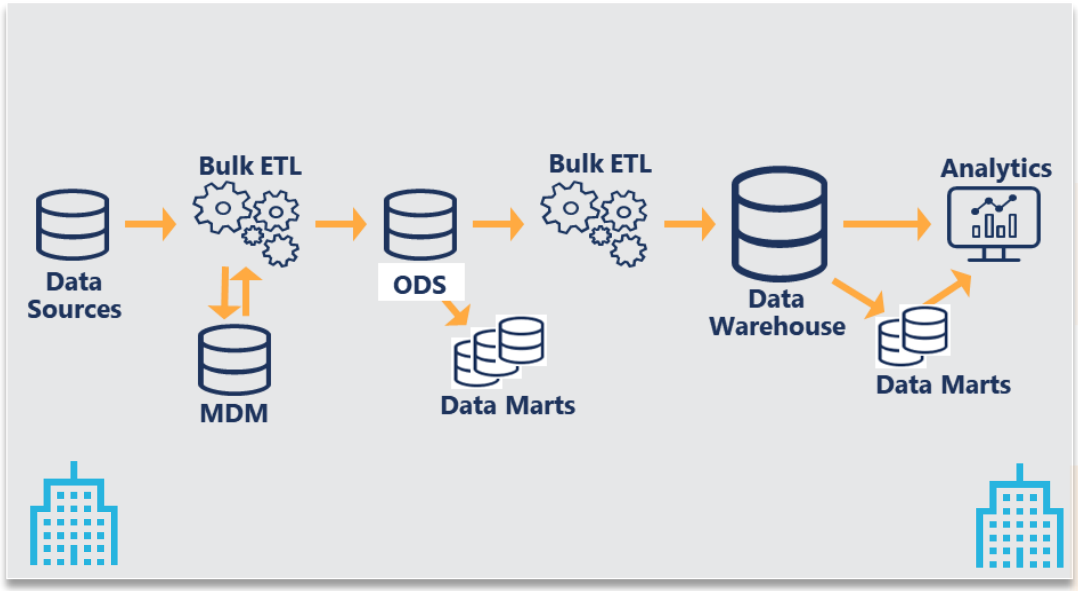
- **Applies DevOps principles of agile development and continuous integration**
- **Seeks to improve collaboration between data managers and consumers**
- **Continuous data delivery in the data pipeline**

Source: Gartner Innovation Insight for DataOps, December 2018

What is driving DataOps?



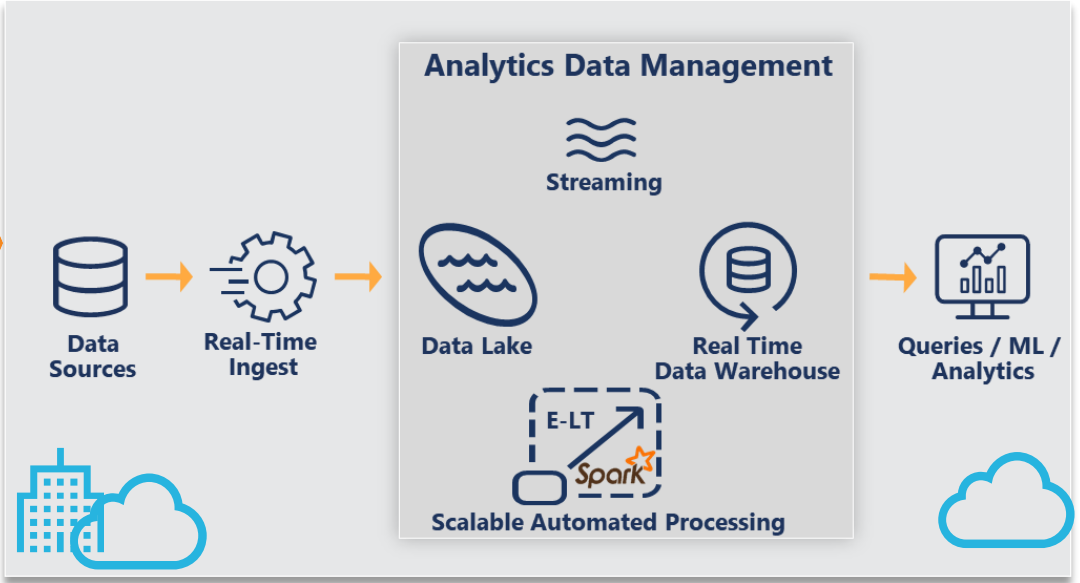
DATA ARCHITECTURE EVOLUTION



Architecture

Platform

Processes



Legacy Data Warehouse Architecture

- ✗ **Bulk** data movement
- ✗ Brittle **hand-coded** ETL
- ✗ **Monolithic** / appliance driven architectures (one size fits all)
- ✗ ***Slow time to market / react to change***

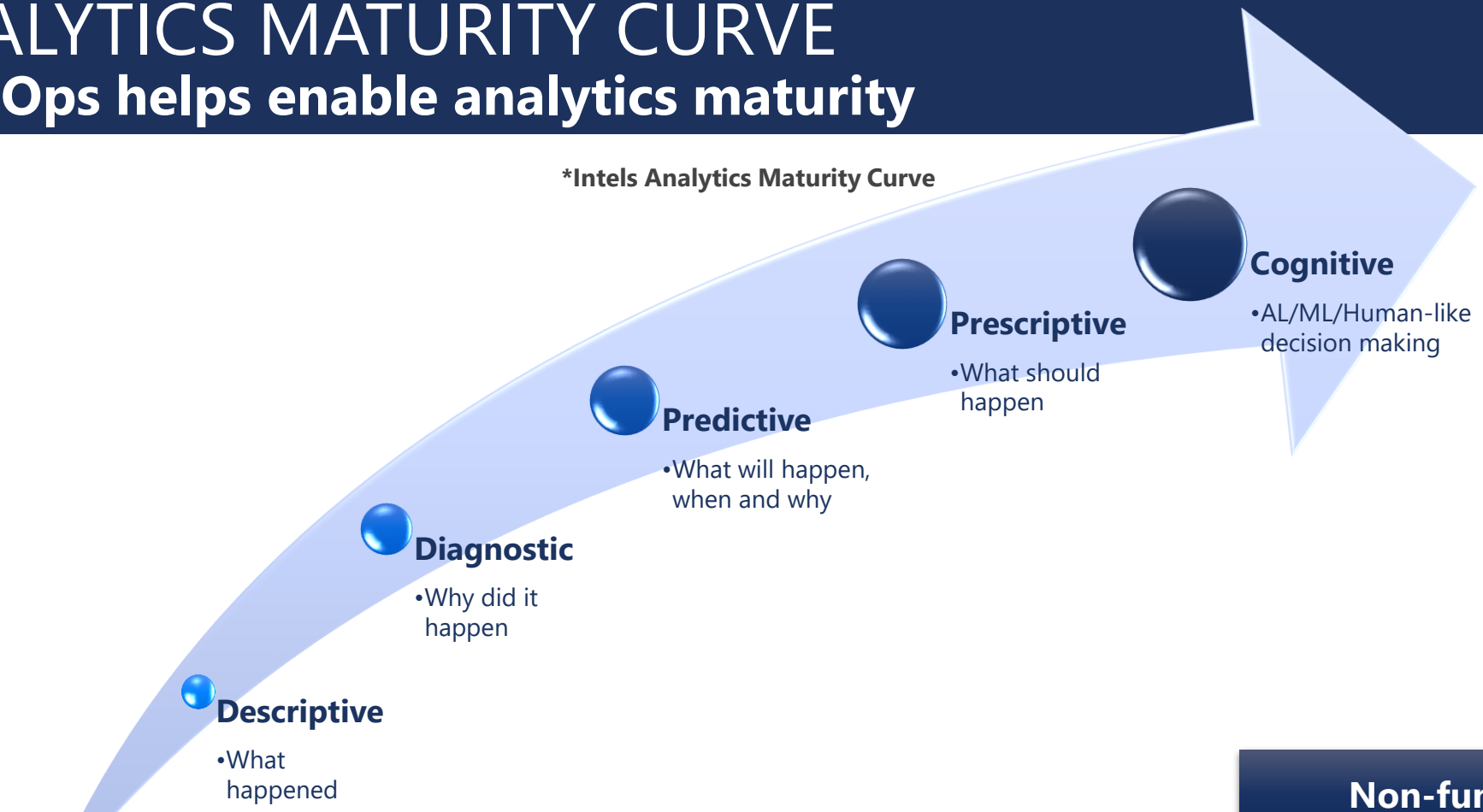
Modern Analytics Data Management

- ✓ **Real-time** data movement
- ✓ **Automated** design and code generation
- ✓ Use-case driven **scalable** technologies
- ✓ ***Faster Time to Market / Value***

ANALYTICS MATURITY CURVE

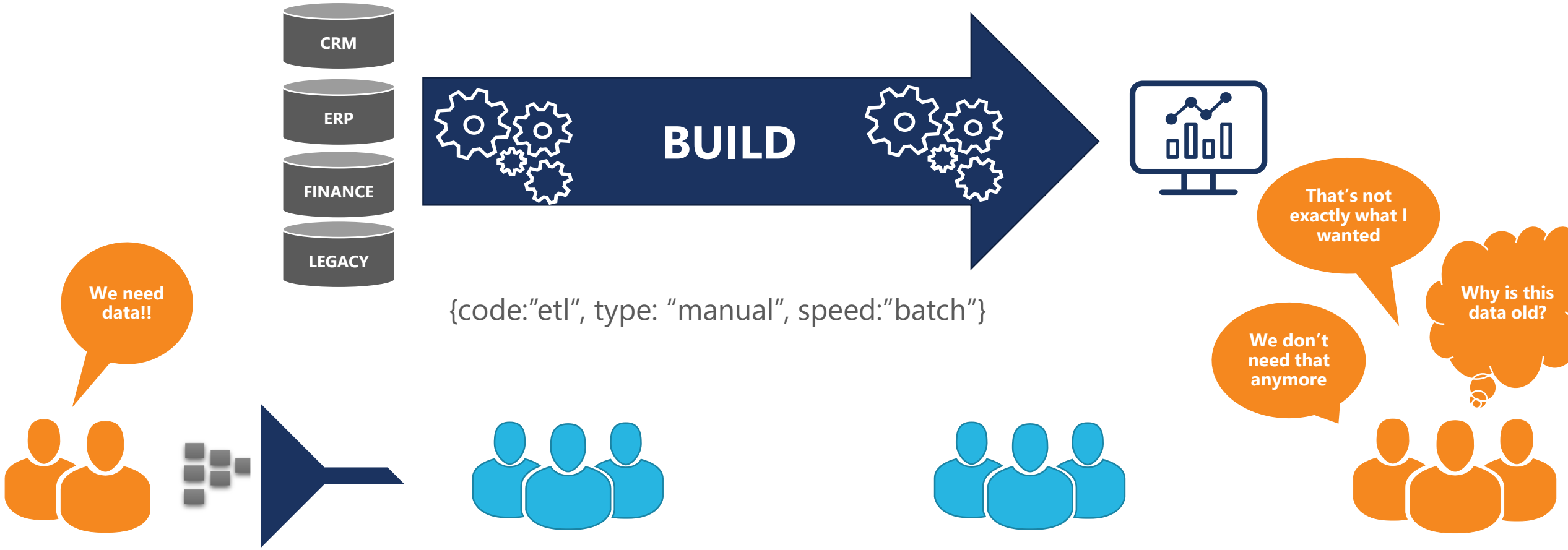
DataOps helps enable analytics maturity

*Intels Analytics Maturity Curve



Non-functional Enablers	
HISTORICAL → REAL-TIME	DATA LATENCY
YEARS & MONTHS → WEEKS & DAYS	TIME TO VALUE
SILOED SPECIALISTS → EMBEDDED TEAMS	COLLABORATION

HOW WE USED TO DO THINGS!



WHY DATAOPS?

ATTUNITY
A Division of Qlik

- **Increasing analytics requirements create complexity and data flow bottlenecks**
- **Data consumers drive demands that IT cannot meet with existing processes and technologies**
- **Projects are failing due to this friction**

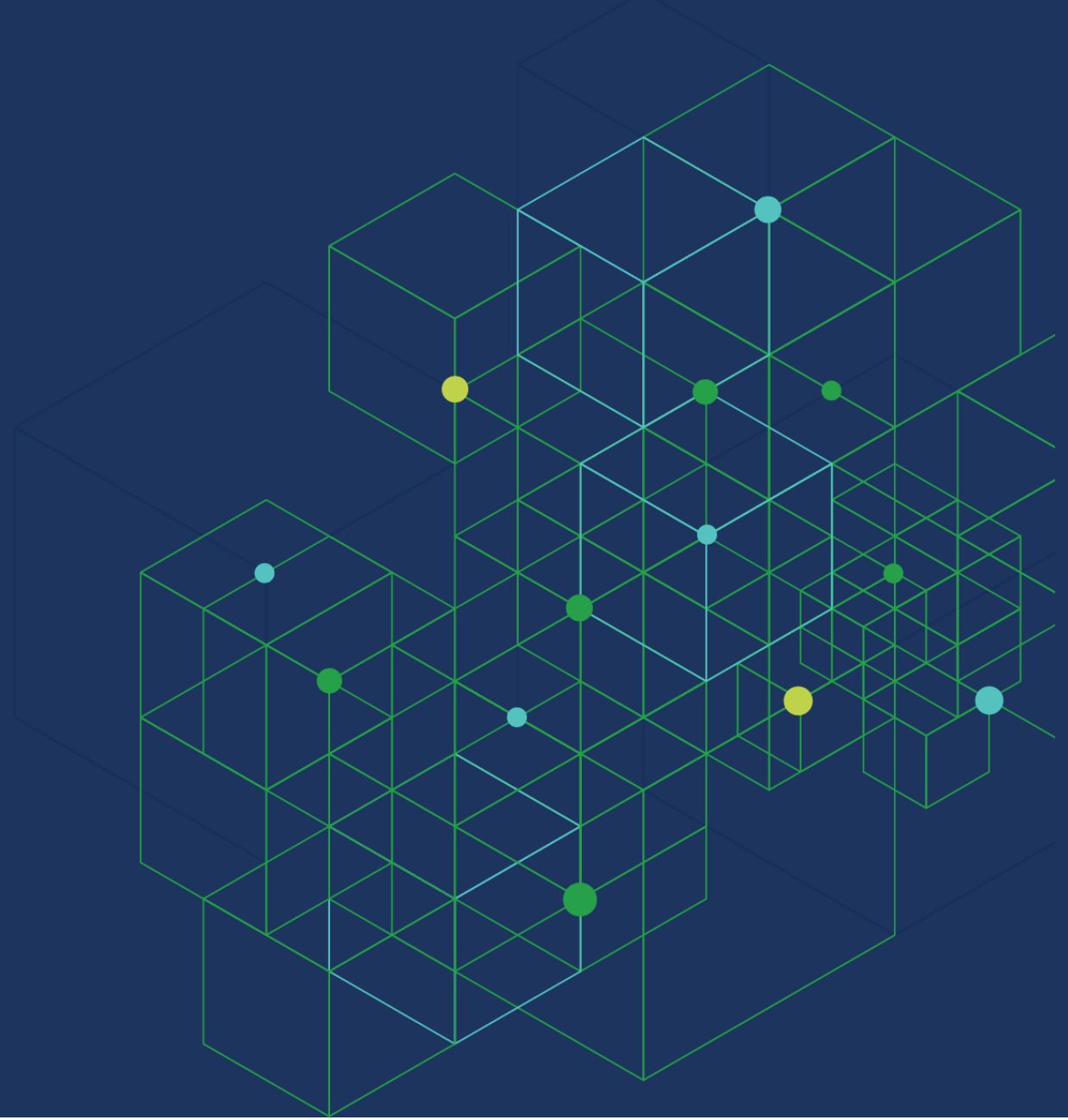
"In every pipeline, data must be identified, captured, formatted, tagged, validated, profiled, cleaned, transformed, combined, aggregated, secured, cataloged, governed, moved, queried, visualized, analyzed, and acted upon. Phew!"

Wayne Eckerson PRESIDENT, ECKERSON GROUP

RISING CHALLENGES

DATA VOLUME, VARIETY, VELOCITY	**NEW PLATFORMS**
SPEED OF CHANGE	**CODING COMPLEXITY**

Enabling DataOps



YOU CAN'T BUY DATAOPS

ATTUNITY
A Division of Qlik

PEOPLE PROCESS TECHNOLOGY

CODE

TOOLS

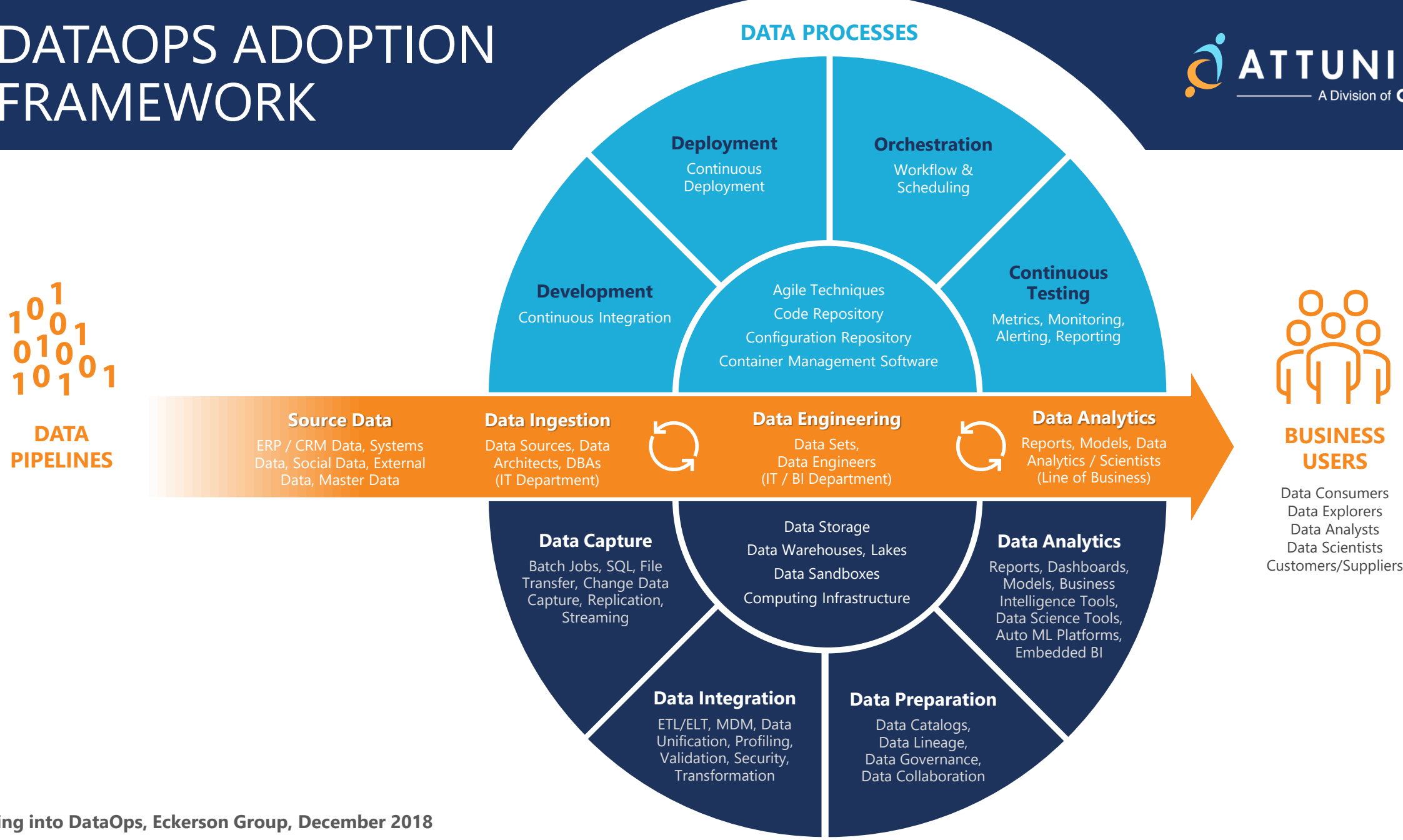
INFRASTRUCTURE

DATA

You should invest in modern technology which supports key DataOps principles and significantly accelerates your time to insight

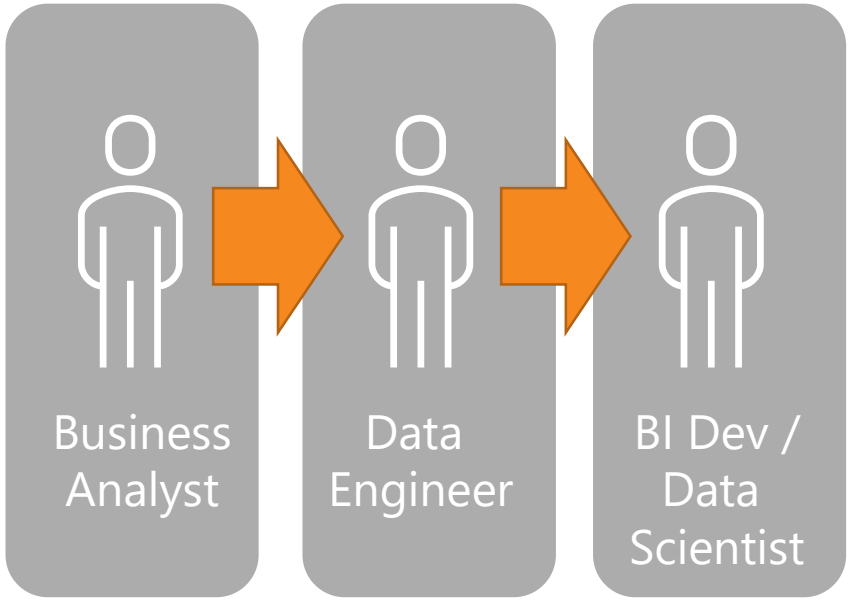
DATAOPS ADOPTION FRAMEWORK

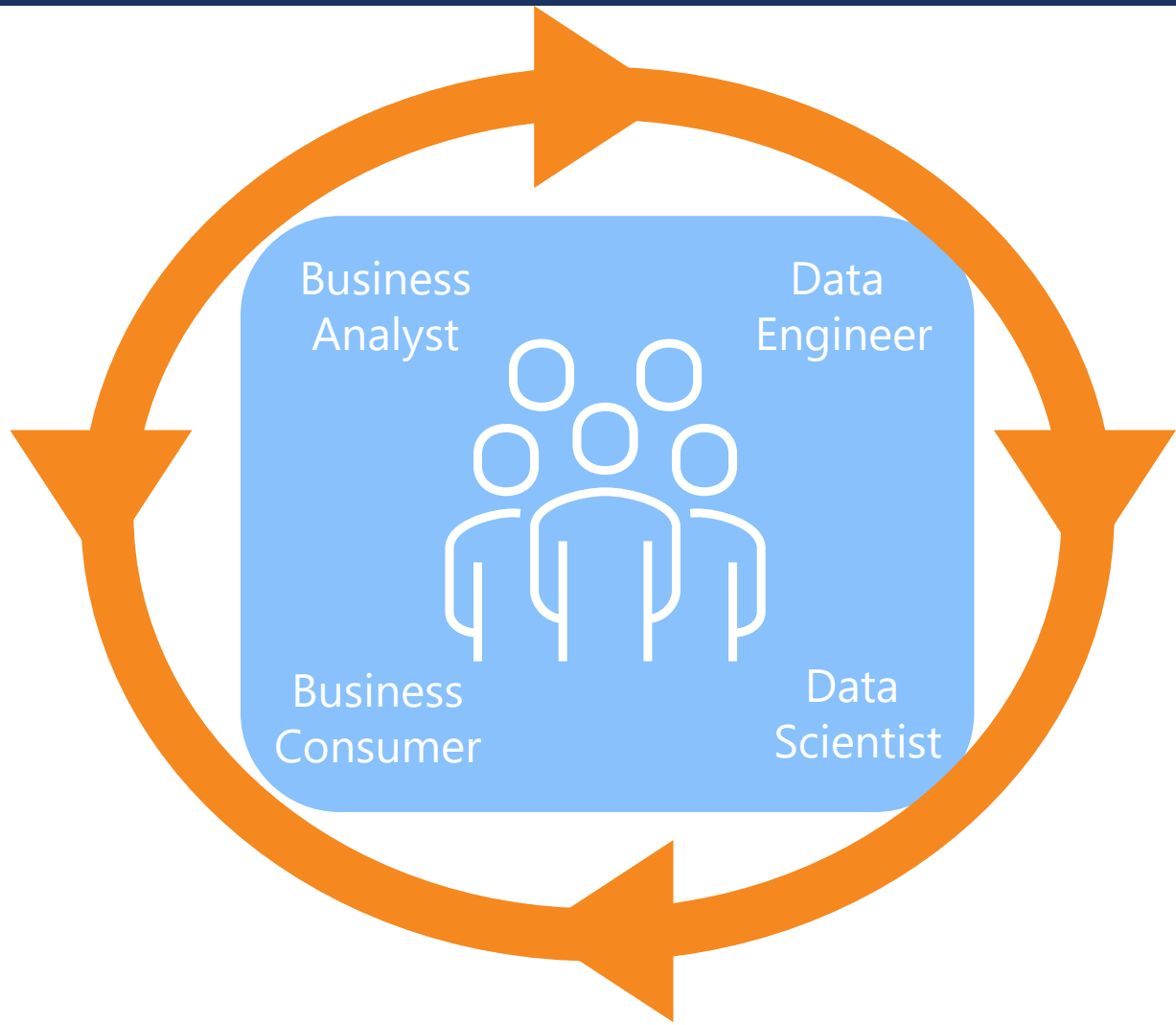
ATTUNITY
A Division of Qlik



Source: Diving into DataOps, Eckerson Group, December 2018

DATAOPS team structure

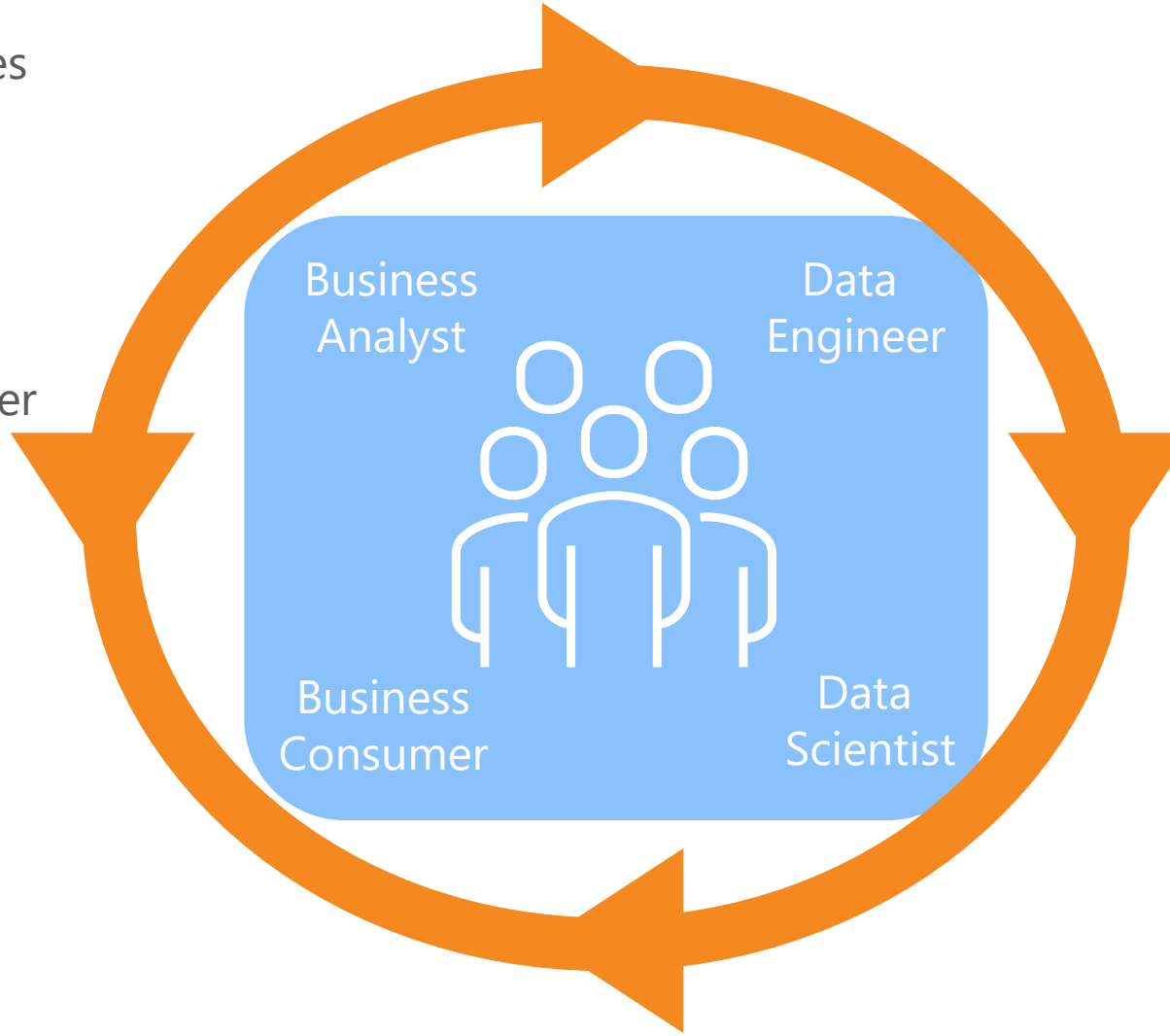




From siloed to centralized – DataOps teams need to be enabled to collaborate

DATAOPS team structure

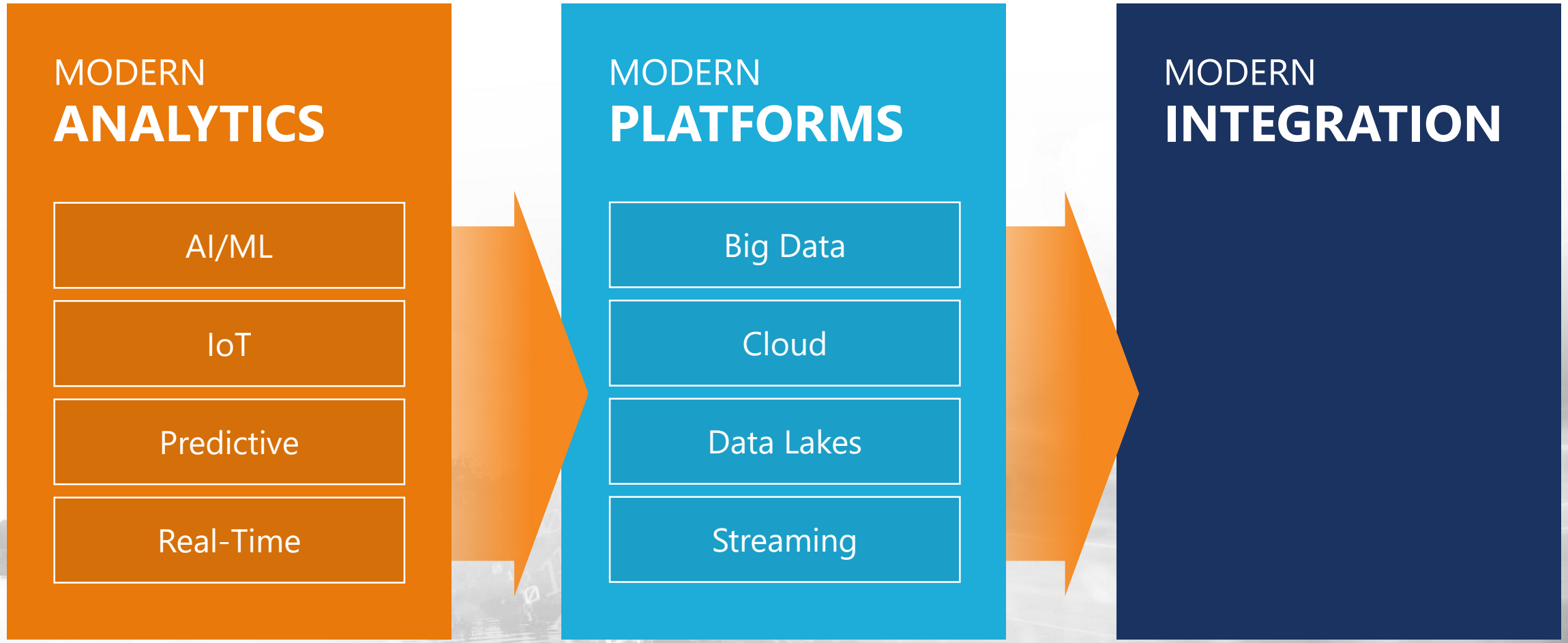
- **Data Engineer**
  - Responsible for building Data Lakes, Data Warehouses
  - DBA, ETL Engineer
  - ETL, Spark programming
- **Data Analyst**
  - Responsible for visualizations, charts, dashboards
  - BI Engineer, Report analyst, Data visualization engineer
  - Data storytelling
- **Data Scientist**
  - Responsible for algorythms, models
  - ML Engineer, AI programmers w/ domain expertise
  - Spark, R, Python, Notebooks
- **Business Consumer***
  - Virtual team member
  - Responsible for business feedback loop



Source: Building a DataOps Team - DataKitchen

DATAOPS BENEFITS FROM A MODERN DATA INFRASTRUCTURE

ATTUNITY
A Division of Qlik



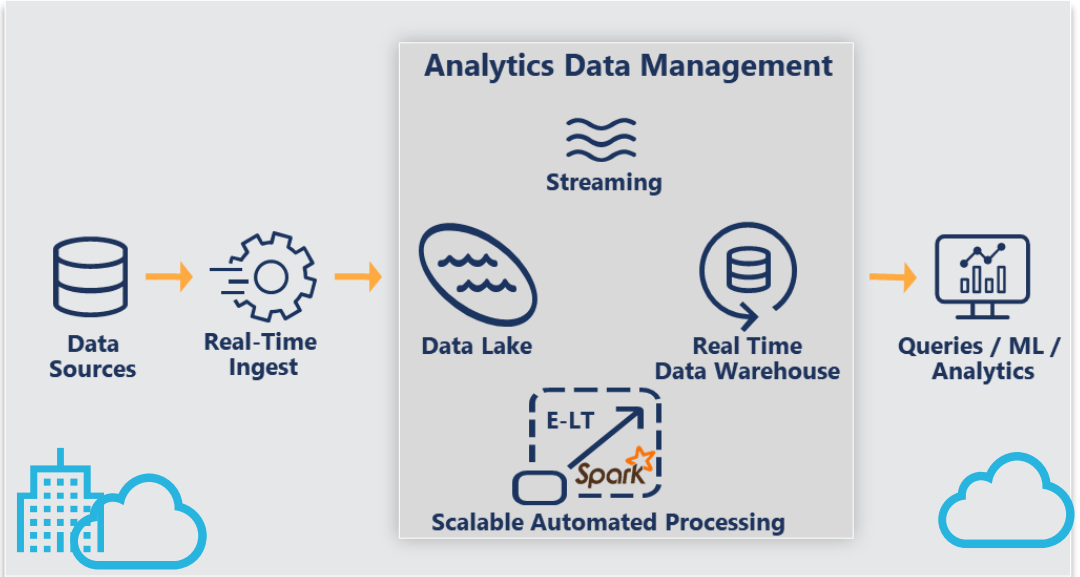
DATAOPS BENEFITS FROM A MODERN DATA INFRASTRUCTURE

ATTUNITY
A Division of Qlik

MODERN
INTEGRATION

A MODERN DATA INFRASTRUCTURE

- Cloud platforms – AWS, Azure, Google, Snowflake
- Data democratization



Modern Analytics Data Management

- ✓ **Real-time** data movement
- ✓ **Automated** design and code generation
- ✓ Use-case driven **scalable** technologies
- ✓ ***Faster Time to Market / Value***

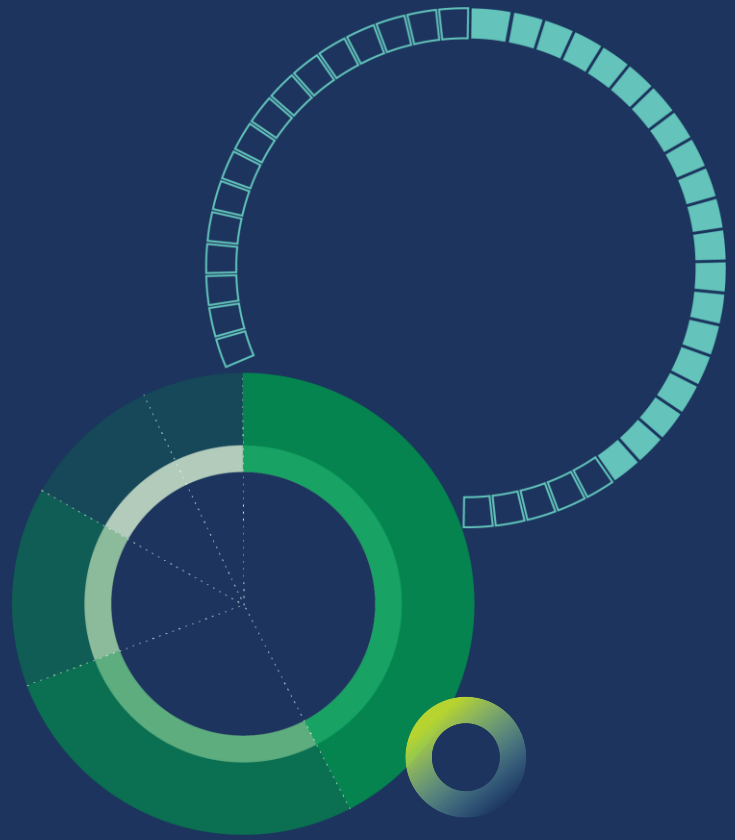
5 KEY DATAOPS TECHNOLOGY REQUIREMENTS

ATTUNITY
A Division of Qlik

MODERN
INTEGRATION

1. Continuous
2. Universal
3. Automation
4. Agility
5. Trust

DataOps technologies



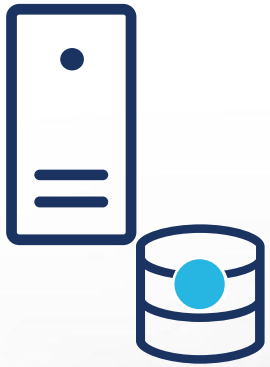
DATAOPS REQUIREMENT #1

CONTINUOUS INTEGRATION

ATTUNITY
A Division of Qlik

MODERN
INTEGRATION

1. **Continuous**
2. Universal
3. Automation
4. Agility
5. Trust



AGENTLESS CDC

REAL-TIME STREAMS

STREAM DATA AND METADATA CHANGES

MINIMAL IMPACT USING TRANSACTION LOGS WITH NO AGENTS

OPTIMIZED FOR EVERY SOURCE AND TARGET

DATAOPS REQUIREMENT #2
UNIVERSAL

ATTUNITY
A Division of Qlik

MODERN INTEGRATION

1. Continuous
2. **Universal**
3. Automation
4. Agility
5. Trust

30 SOURCES

DATABASE
- Oracle
- SQL Server
- DB2 iSeries
- DB2 z/OS
- DB2 LUW
- MySQL
- PostgeSQL
- Sybase ASE
- Informix
- ODBC

EDW
- Exadata
- Teradata
- Netezza
- Vertica
- Pivotal

HADOOP
- Hortonworks
- Cloudera
- MapR

SAP
- ECC on Oracle
- ECC on SQL
- ECC on DB2
- ECC on HANA
- S4 HANA

CLOUD
- AWS RDS
- Amazon Aurora
- Amazon Redshift
- Salesforce

OTHER LEGACY
- SQL/MP
- Enscribe
- RMS

MAINFRAME
- DB2 for z/OS
- IMS/DB
- VSAM

FLAT FILES
- Delimited (e.g., CSV, TSV)

40 TARGETS

DATABASE
- Oracle
- SQL Server
- DB2 LUW
- MySQL
- PostgreSQL
- Sybase ASE
- Informix
- MemSQL

EDW
- Microsoft PDW
- Exadata
- Teradata
- Netezza
- Vertica
- Sybase IQ
- Amazon Redshift
- Actian Vector
- SAP HANA

CLOUD
- Amazon RDS
- Amazon Redshift
- Amazon S3
- Amazon Aurora
- Google Cloud SQL
- Azure SQL DW
- Azure SQL DB
- Azure MySQL
- Azure PostgreSQL
- Snowflake

STREAMING
- Kafka
- Azure Event Hubs
- MapR-ES
- AWS Kinesis

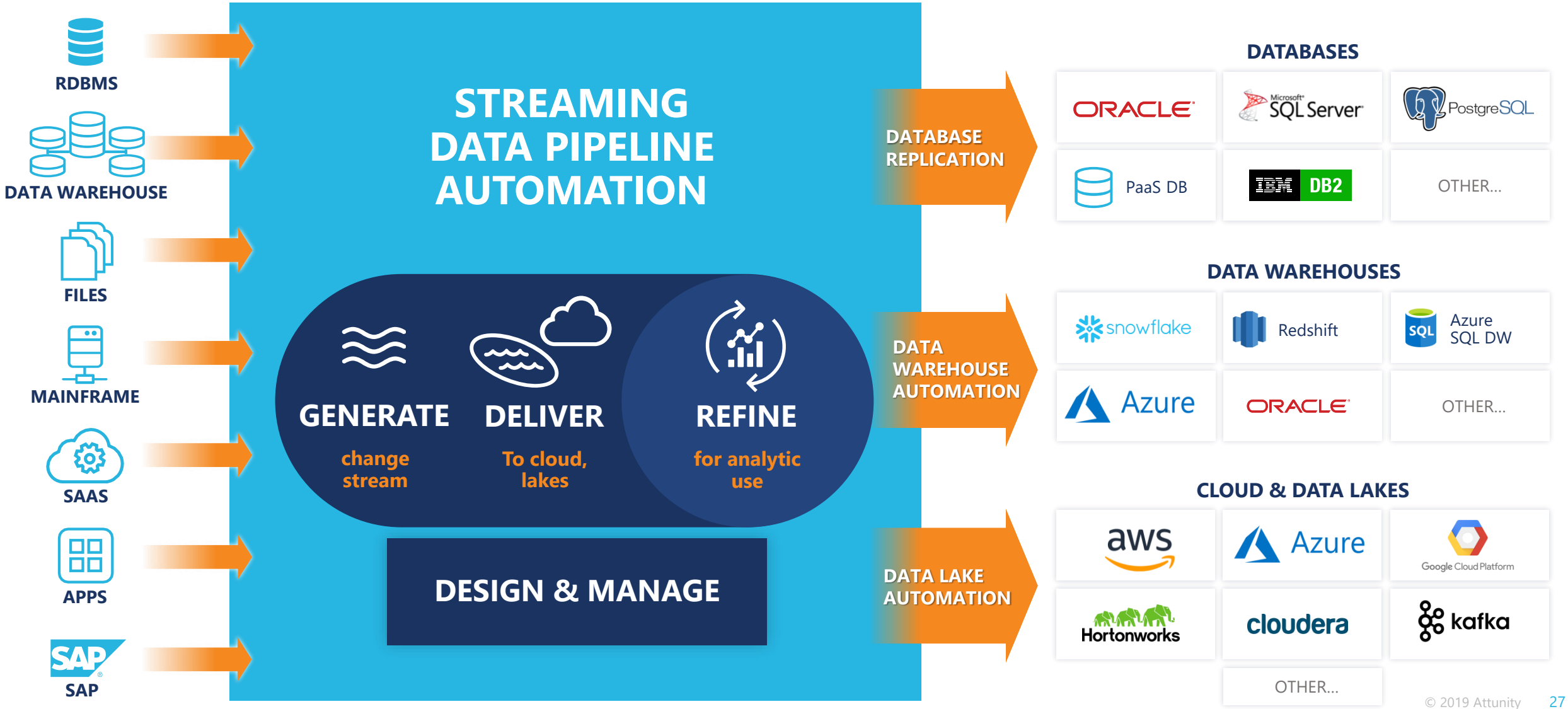
SAP
- HANA

FLAT FILES
- Delimited (e.g., CSV, TSV)

HADOOP
- Hortonworks
- Cloudera
- MapR
- Amazon EMR
- HDInsight

DATAOPS REQUIREMENT #2
UNIVERSAL

ATTUNITY
A Division of Qlik



DATAOPS REQUIREMENT #3
AUTOMATION

ATTUNITY
A Division of Qlik

MODERN
INTEGRATION

1. Continuous
2. Universal
3. **Automation**
4. Agility
5. Trust

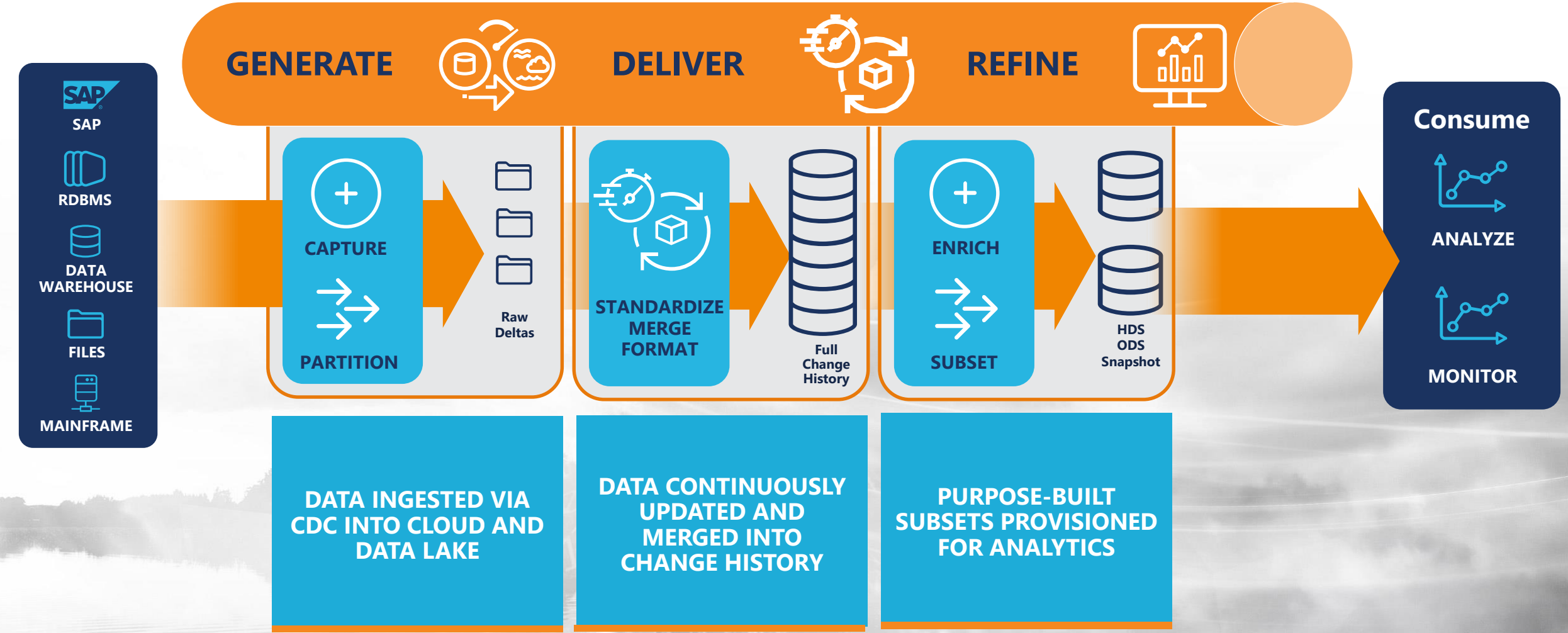
AUTOMATE THE DATA PIPELINE

GENERATE  DELIVER  REFINE  MANAGE

AUTOMATE MULTI-STAGE PROCESSING WITHOUT CODING	ADAPTS TO SOURCE AND TARGET CHANGES	HETEROGENEOUS & DISTRIBUTED WORKLOADS

DATAOPS REQUIREMENT #3
AUTOMATION

ATTUNITY
A Division of Qlik



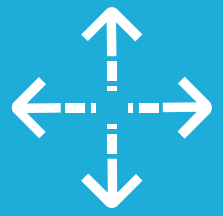
DATAOPS REQUIREMENT #4
AGILITY

ATTUNITY
A Division of Qlik

MODERN
INTEGRATION

1. Continuous
2. Universal
3. Automation
4. **Agility**
5. Trust

RUNS IN THE CLOUD(S), ON-PREM, OR BOTH

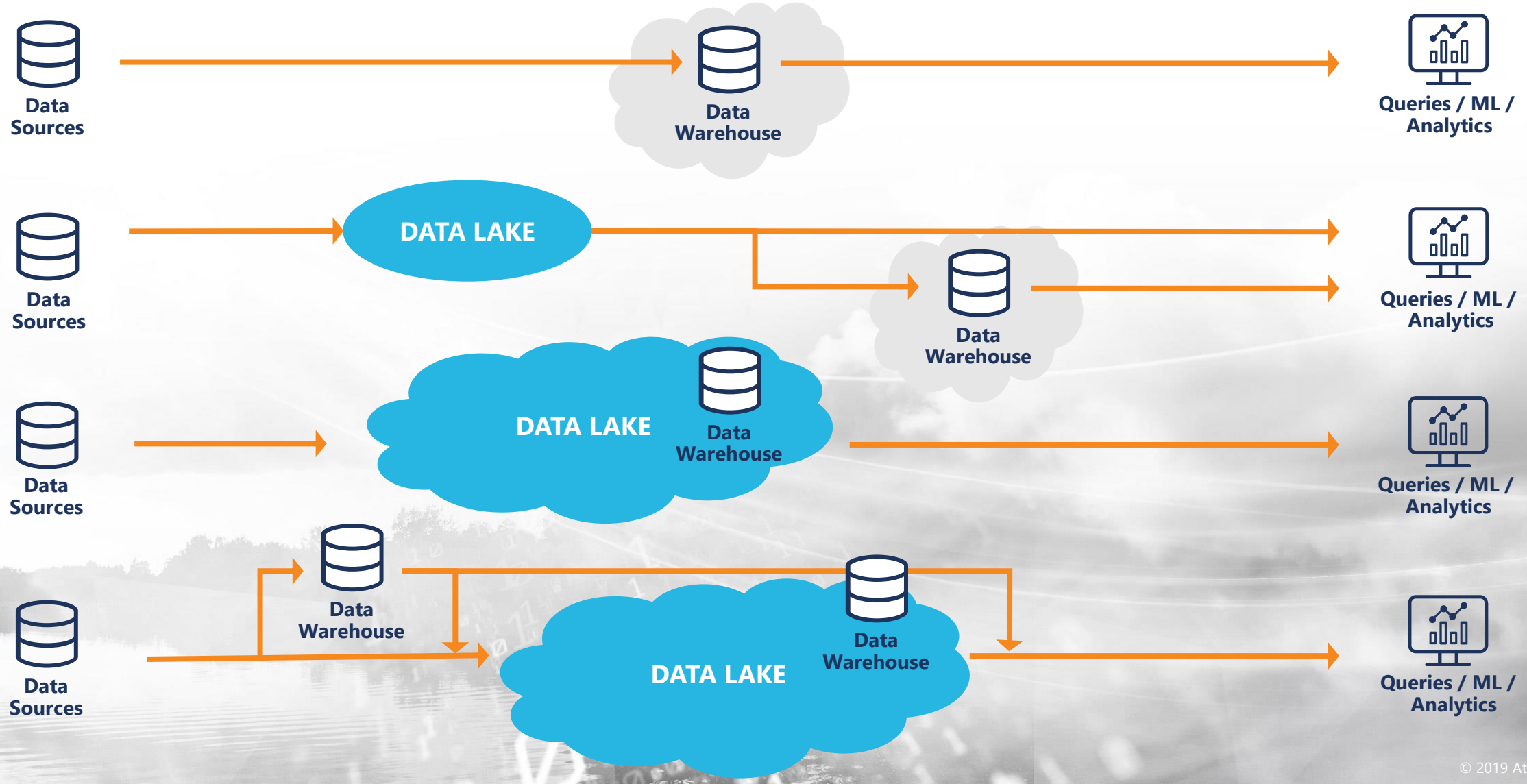


DATA LAKE AND WAREHOUSE AUTOMATION

FUTURE-PROOF FOR 'ARCHITECTURES IN MOTION'

DATAOPS REQUIREMENT #3
AUTOMATION

ATTUNITY
A Division of Qlik



DATAOPS REQUIREMENT #5
TRUST

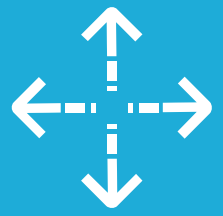
ATTUNITY
A Division of Qlik

MODERN
INTEGRATION

1. Continuous
2. Universal
3. Automation
4. Agility
5. **Trust**

Metadata creates greater trust and confidence from data consumers

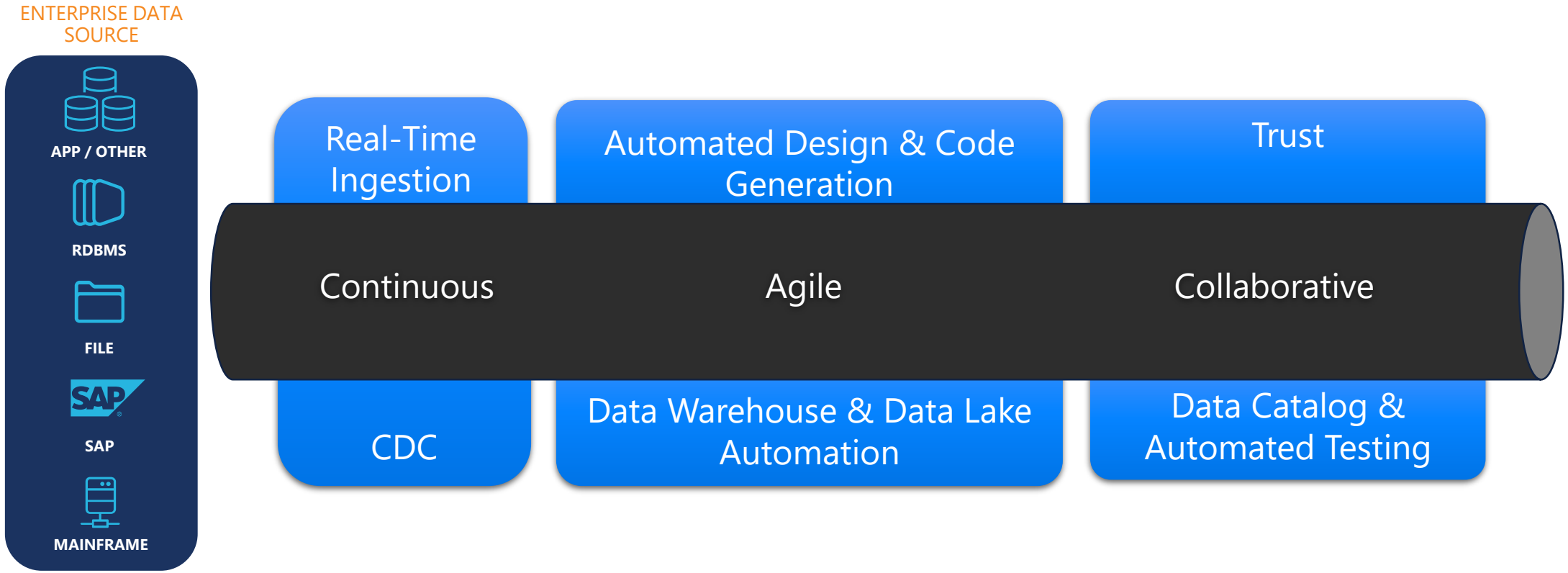
DATA CATALOG TO ACCESS IT-MANAGED AND USER-CURATED DATA SETS



DATA LINEAGE TO UNDERSTAND WHERE THE DATA CAME FROM AND HOW IT WAS TRANSFORMED

DATA VALIDATION TO ENSURE THAT ALL OF THE SOURCE DATA WAS REPLICATED

Technologies to support DataOps Data Pipelines



Demonstration

DataOps pipeline imagined

DATAOPS: ACCELERATING TIME TO INSIGHT

ATTUNITY
A Division of Qlik

TRADITIONAL

DESIGN	CODE	TEST	DEPLOY

Data Manager

- Long cycles
- Batch processing
- Brittle ETL scripts

Data Consumer

NEEDS

- Fast cycles
- Real-time data
- Resilience

DATAOPS: ACCELERATING TIME TO INSIGHT

ATTUNITY
A Division of Qlik

Data Manager

- ✓Code Automation
- ✓Real-time delivery
- ✓Run-time evolution

Data Consumer

NEEDS

- Fast cycles
- Real-time data
- Resilience

DATAOPS

Questions?

Thank you…