Preparing and delivering data for big data analytics

Preparing and delivering data for big data analytics

- **Big data needs to be:**

  - High-quality and up-to-date → Data teams can **trust** it for analysis

  - Well-structured → Data teams can **easily** and **quickly** analyze the data

- **Traditional methods (ETL and Python/Perl/Bash scripts) fall short because they are**

  - Tedious, time-consuming, and error-prone

  - Rigid, without the ability for analysts to define which data they need, when, and how it should be prepared

  - Inefficient, requiring data analysts to wait for hours while their data is prepared

Entity-based approach

Preparing and delivering data for big data analytics

- **Collect, process, and serve data from source systems to data lakes and DWHs by *business entities***
  - Customer, party, product, location, order, etc.
  - Data integrity by design
  - Data is normalized into a common business language
  - Data is structured for quick and easy analysis

- **Real-time, scheduled or on-demand data movement**

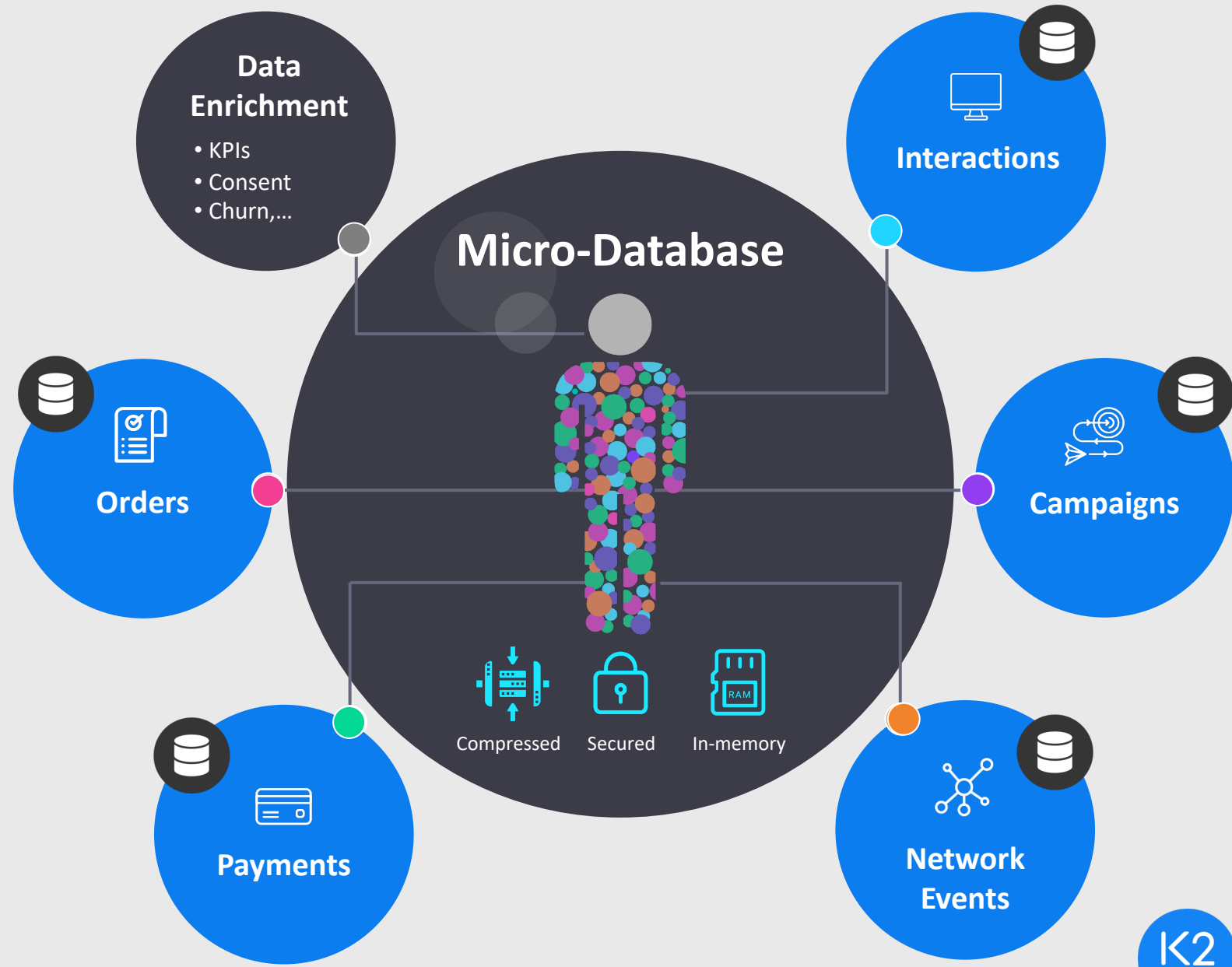- **Reliable, scalable, and cost-efficient**
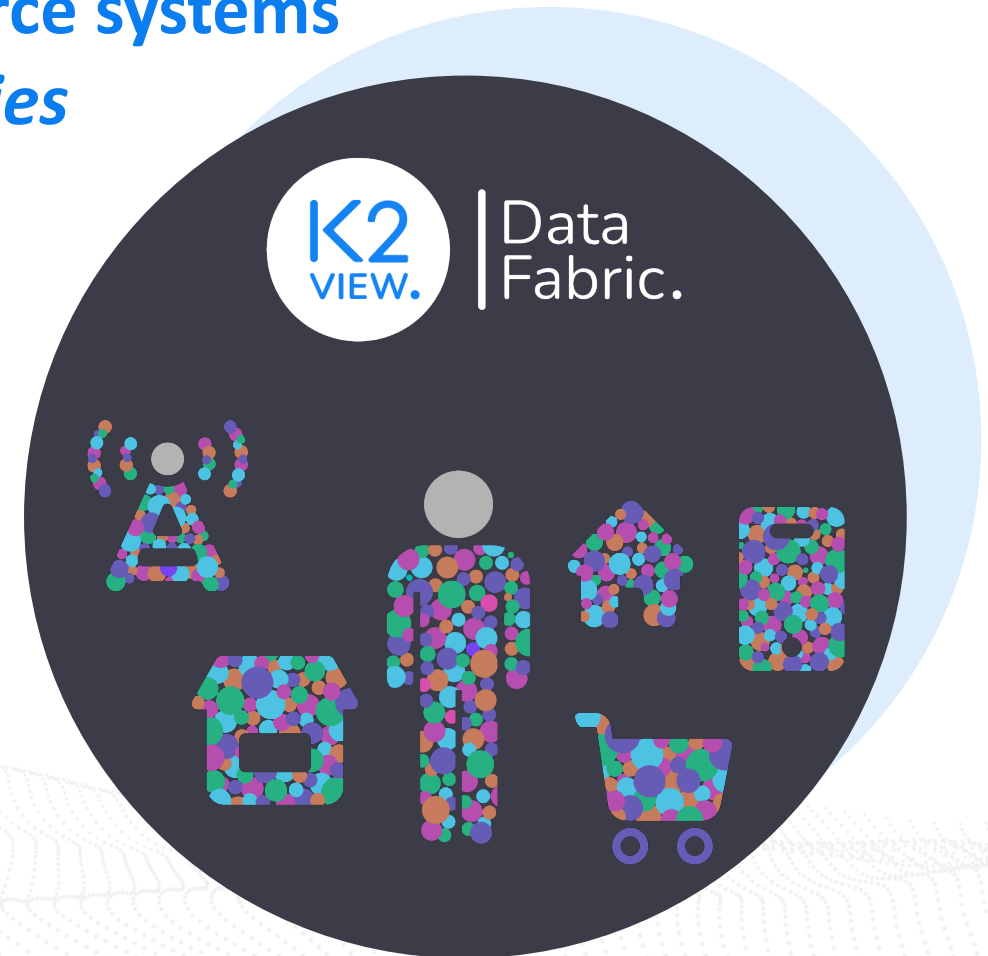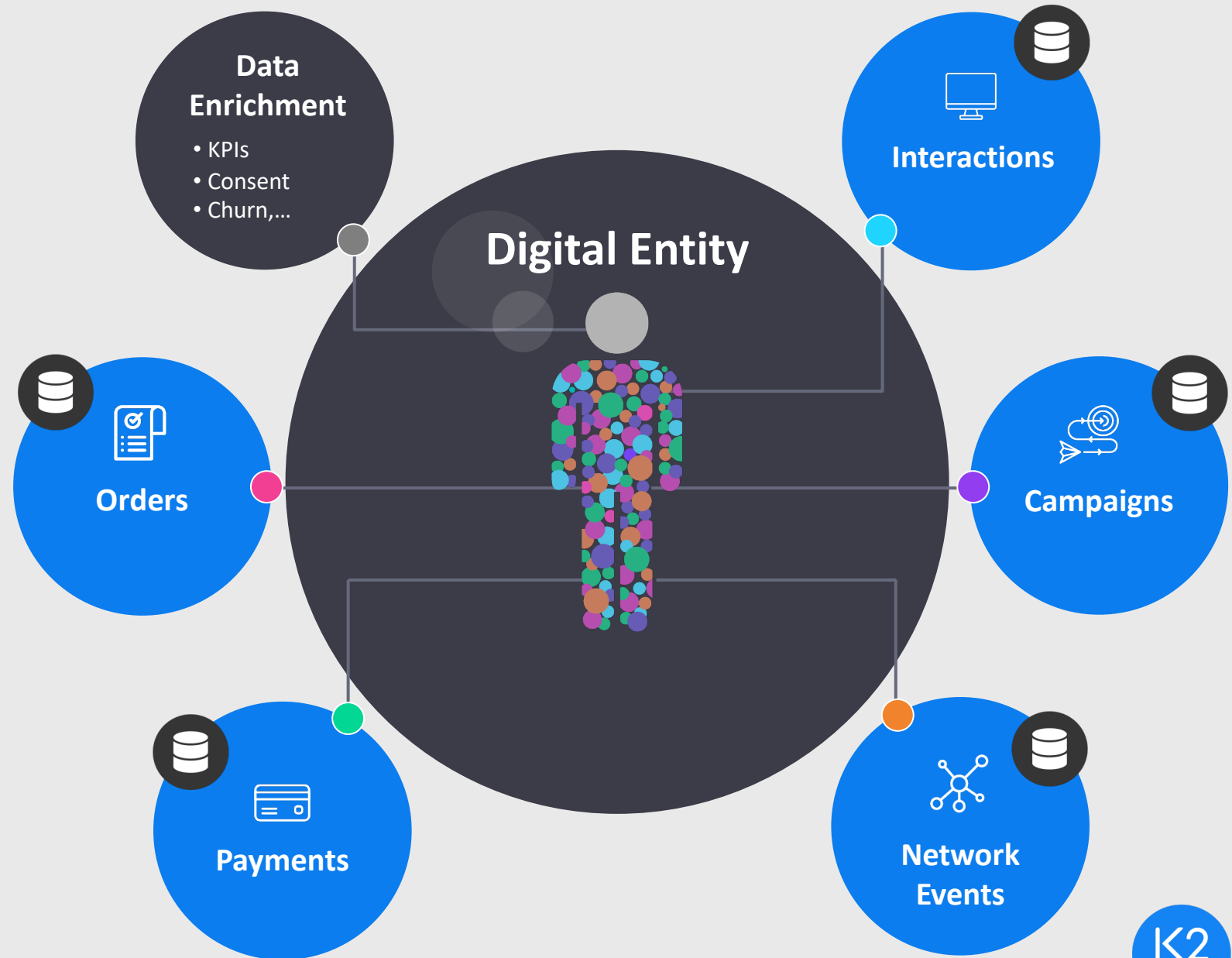
Micro-Database

- Data is ingested in real time, enriched, transformed, and stored into micro-DBs

- One micro-DB per instance of an entity, compressed by 90%

- Each micro-DB is secured with its own key

- Data sync rules for optimal performance

- Distributed, high-scale, high-performance

- Patented

Data Enrichment
- KPIs
- Consent
- Churn,…

Interactions

Micro-Database

Orders

Campaigns

Compressed   Secured   In-memory

Payments

Network Events

K2 VIEW.

Subscribe

Message Bus

Data Lakes

MICROSERVICES

🌐 Data Delivery Layer

👤 Authentication Engine

🔒 Data Masking Layer

⚙️ Data Transformation and Orchestration

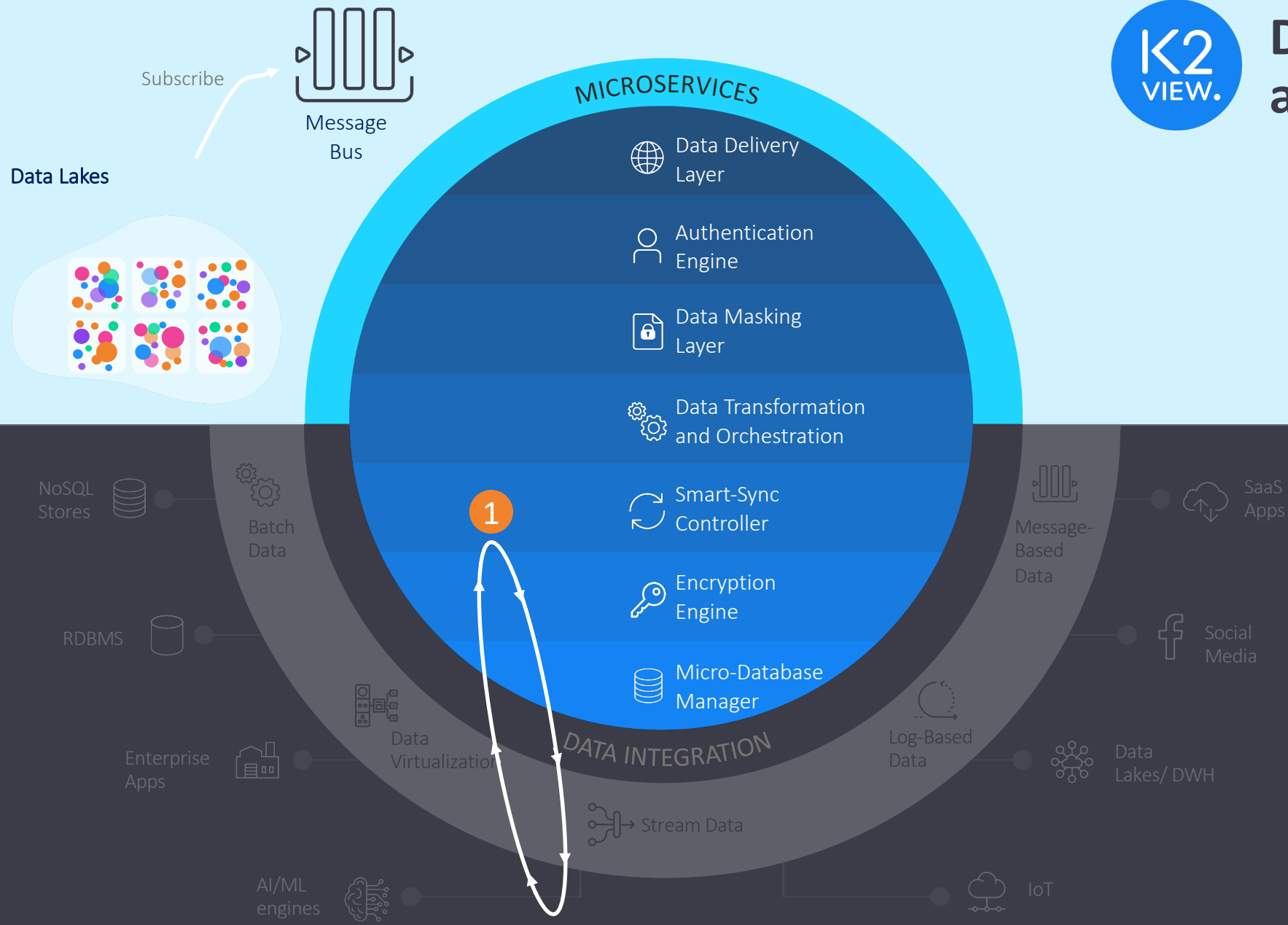🔄 Smart-Sync Controller

🔑 Encryption Engine

🗄️ Micro-Database Manager

DATA INTEGRATION

NoSQL Stores

RDBMS

Enterprise Apps

Batch Data

Data Virtualization

AI/ML engines

Stream Data

Log-Based Data

Message-Based Data

SaaS Apps

Social Media

Data Lakes/ DWH

IoT

K2 VIEW.

Data preparation and delivery

1
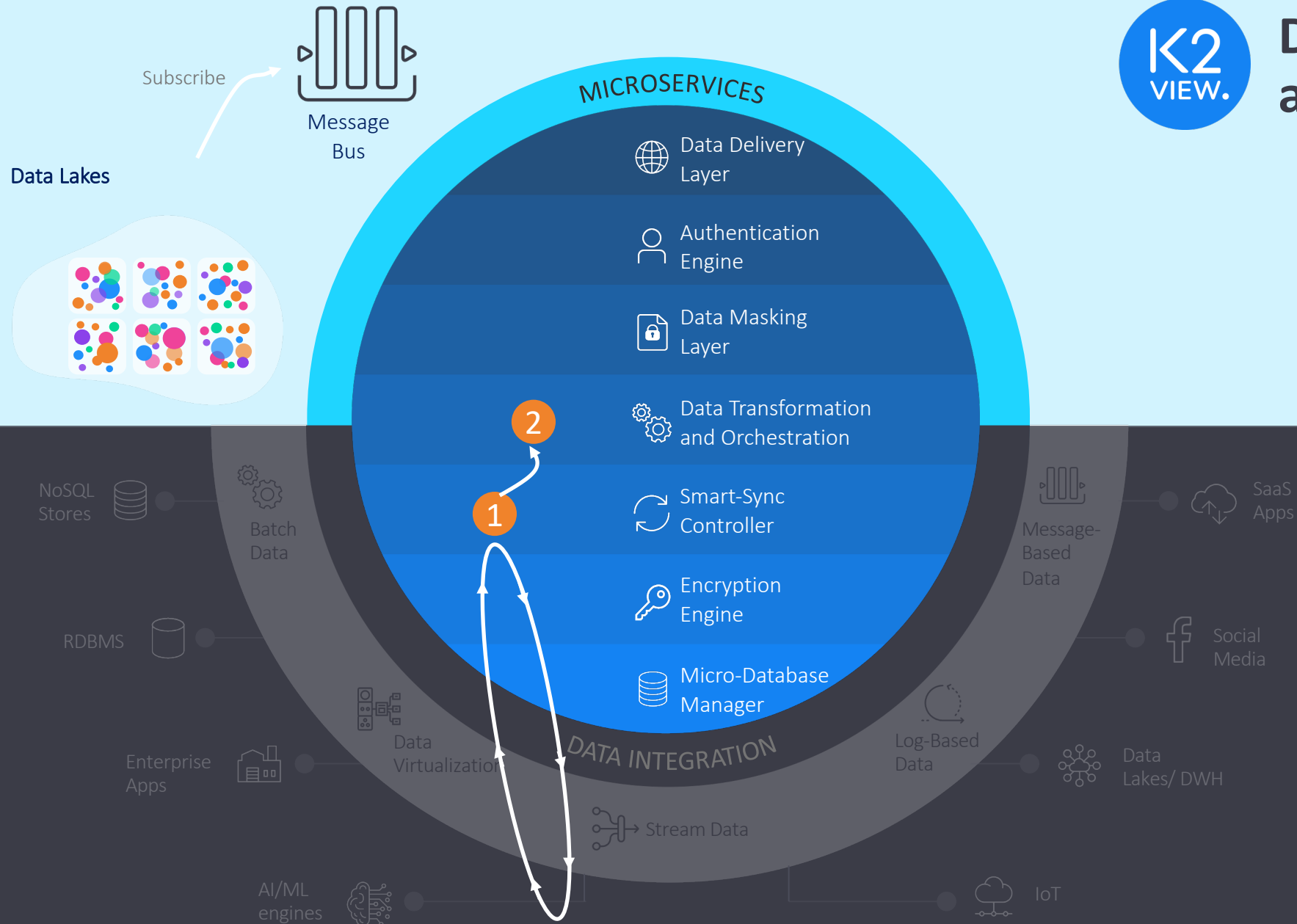
Ingests data messages in real time via CDC.

Applies identity resolution algorithms to identify relevant micro-DB.

Encrypts, compresses and stores updates in the micro-DB.

Places data into memory for processing.

Data preparation and delivery

Subscribe

Message Bus

Data Lakes

MICROSERVICES

Data Delivery Layer

Authentication Engine

Data Masking Layer

Data Transformation and Orchestration

Smart-Sync Controller

Encryption Engine

Micro-Database Manager

DATA INTEGRATION

NoSQL Stores

Batch Data

RDBMS

Data Virtualization

Enterprise Apps

AI/ML engines

Stream Data

Log-Based Data

Message-Based Data

SaaS Apps

Social Media

Data Lakes/ DWH

IoT

2 Performs relevant data transformation, orchestration, and enrichments as defined, and on the fly.

Data preparation and delivery

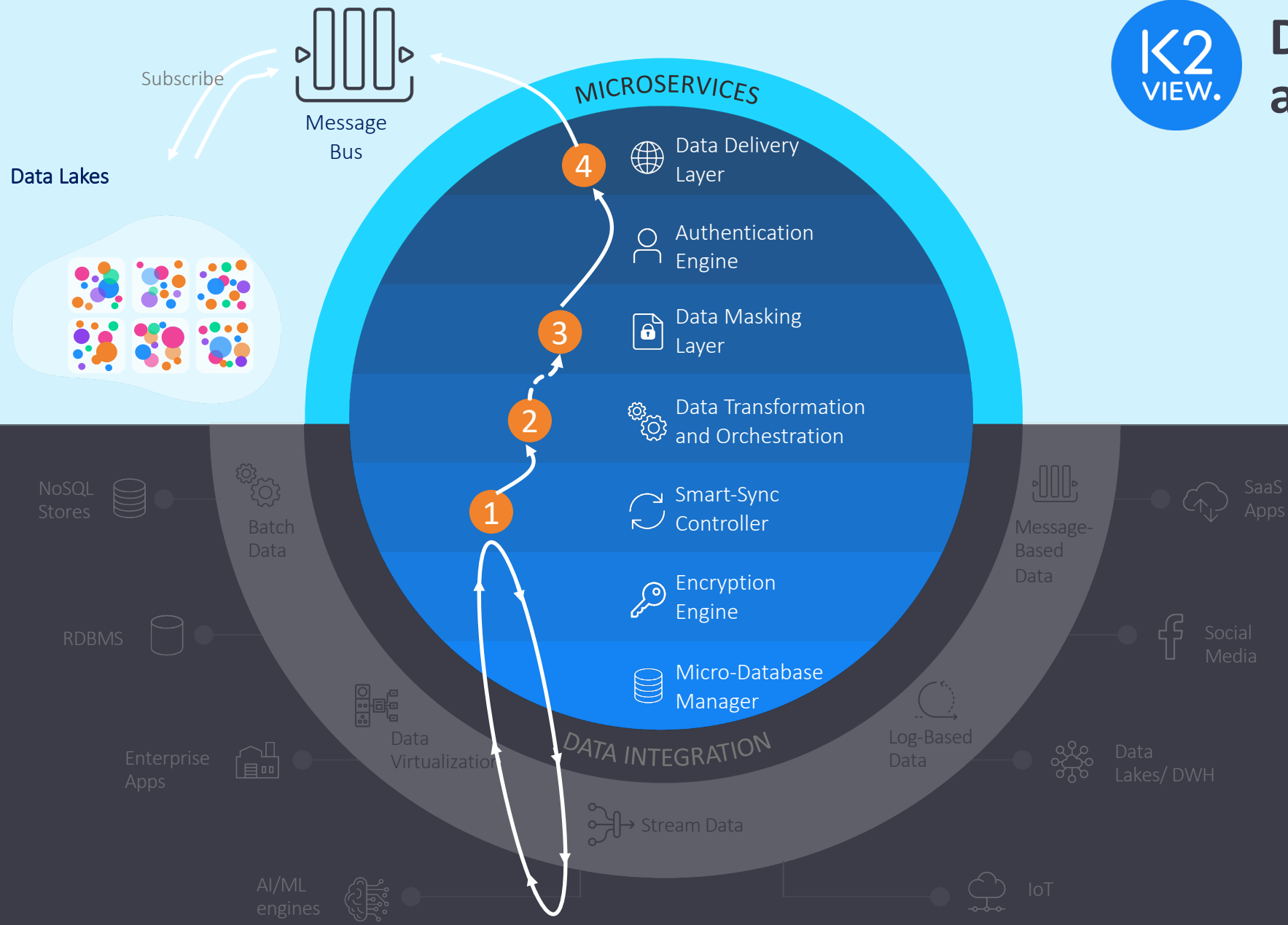# Data preparation and delivery

Subscribe

Message Bus

Data Lakes

MICROSERVICES

4 — Data Delivery Layer

Authentication Engine

3 — Data Masking Layer

2 — Data Transformation and Orchestration

1 — Smart-Sync Controller

Encryption Engine

Micro-Database Manager

DATA INTEGRATION

NoSQL Stores
Batch Data
RDBMS
Enterprise Apps
Data Virtualization
Stream Data
AI/ML engines
Log-Based Data
Message-Based Data
SaaS Apps
Social Media
Data Lakes/ DWH
IoT

4

Publishes data to message bus for consumption by the data lake

Your data is always complete
Insights are always trusted

- The micro-DB always holds a complete view of a business entity:
  - Business rules define when data is synced with the source systems
  - Business rules define how to handle data that cannot be synced - e.g., source/connectivity failure
  - Built-in retry mechanism to overcome failure

- Data is always associated with a Digital Entity → there is never "orphan" data, which isn't associated with an entity → you can always trust derived insights

- Data is pushed/published to data lakes by business entity
  - No risk of missing data at the entity level

Example: Consider a marketing analyst's query to find all Customers that: (i) transferred > $10K funds, (ii) in the last 48-hours, and (iii) have a remaining balance < $20K.
The query could be trusted only if the balance data and transfer data for all customers were transferred to the data lake at the same time

Your data is always up to date
Insights are always current

- K2View continually ingests data from all source systems in real time, and publishes it to consuming datastores
  - Any data lake that subscribes to updates from the data fabric will be updated in real time by the data fabric
- All data transport methods are supported, in and out of the data fabric
  - Bulk (ETL), streaming, log-based (CDC), and message bus

Your data is easily understood and quickly accessible
Analysis by non-data pros

- The K2View Digital Entity normalizes and standardizes data from all sources into a common "business language" that is easily understood by data teams

- By moving complete entity data to the data lake, it is possible to easily correlate data by entity within the data lake
  - Data analysts and data scientists don't have to correlate data at query time, making data queries **much easier to write, and much faster to run**

Your data is always governed and safe
You are always compliant

- K2View dynamically masks sensitive data at the entity level
  - Data from different systems is masked at once at the entity level – preserving data integrity
- Data is encrypted from the time it is ingested from the source system to the moment it is served to the data lakes
  - Data stored in the data fabric is encrypted by micro-DB (entity instance), each micro-DB secured with its own 256-bit encryption key

You always know what data changed and when
Gain time-based insights

- Data teams often need to analyze time between events - e.g., average time elapsed between certain order statuses – but the timestamp of the events (order status change date in the example) are not captured in the source systems
  - This requires data teams to frequently upload entire datasets → costly, lengthy
- K2View can serve data lakes just with the changes to a business entity, when the data changes, together with the timestamp of the change

6

Your source systems are never impacted
No system performance degradation

- K2View syncs its micro-DBs with the source systems in real time ("push" from source) and/or according to user-defined business rules ("pull" from source)

- Push/pull is defined by field in the micro-DB

- Push from source:
  - Data changes (change logs) are served to the data fabric in real time using CDC → no impact to the source systems
  - Real-time identity resolution algorithm updates the right business entity with arriving data fragments

- Pull from source:
  - Business rules "protect" your source systems from being unnecessarily stressed by redundant queries, while ensuring that data freshness is not compromised.
    For e.g., if customer invoices are generated once a month, invoice data will be synced once a month, a day after sync is performed; conversely, payments will be synced 1-2 times a day

- K2View queries data by business entity, and thus doesn't stress source systems

Your data is published for all
Become a data-driven enterprise

- K2View enables a data-driven enterprise:
  - Data fabric publishes business entity data changes to the enterprise via a message queue
  - Data consumers (marketing, R&D, customer service,...) can subscribe to the data (topics) that they need for business analysis to ensure that they have up-to-date, complete, and trusted data

Your data is encrypted and compressed
Safe, fast, and cost-effective data movement to the cloud

K2 VIEW.

- K2View Data Fabric can be deployed on-premise and on-cloud, as a single distributed cluster
  - The data fabric nodes are deployed close to the source (on-premise) AND target (cloud)
  - Data is encrypted and compressed before it is moved to the cloud, to minimize network costs and increase data transfer speed

Use Cases

K2 VIEW. + **AT&T**

Customer 360 architecture

130M Customers

625 Backend systems

K2 VIEW. | Data Fabric.

Fresh data

No backend load

True real-time performance

One technology, one team
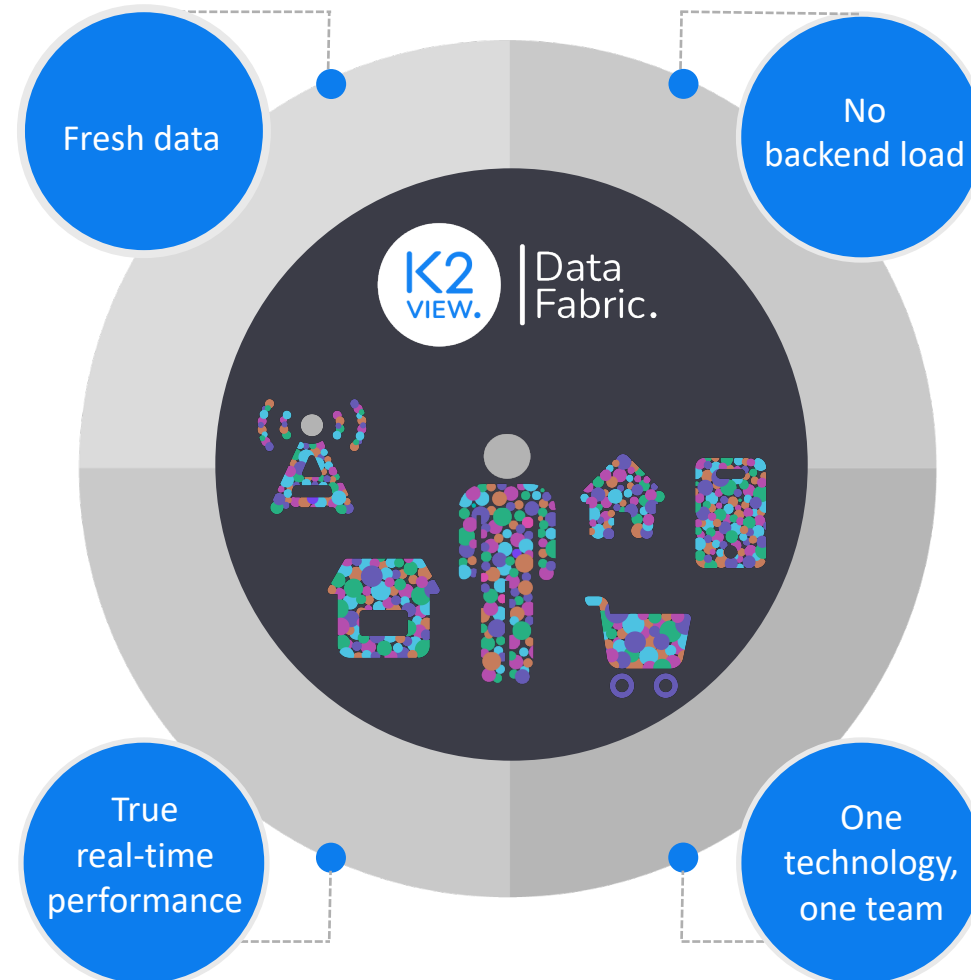
Impact

☑ **$425M / year** saved

☑ **30x** Speed to delivery

☑ Atomic security

Case Study

Fabric powers Digital Transformation with micro-service based access to customer data

AT&T

Industry: **Telco**

Location: **USA**

Product: **K2View Data Fabric for Digital Transformation**

+500
Oracle SP transformed into Web-Services

80% Reduction
in hardware infrastructure size and cost

Faster TTM
Services delivered in Days vs Months

Challenge

- Massive centralized customer DB was costly and slow to maintain and couldn't keep up with the demands of the business.

- DB was the single gateway to data for 130M customers and had become a bottleneck.

- Big data and MDM solutions wouldn't work because data had to be current.

Solution

- K2View Data Fabric

- Access and organize data from 609 systems into 130M customer micro-dbs.

- Replace 500 Oracle stored procedures with simplified web-services.

- Enterprise access via micro-services.

Results

- **Performance:** K2view data Fabric performance is orders of magnitude faster than the existing solution.

- **Strategic Value:** Client adopting Fabric as the foundation for their transformation to a micro-services based architecture.

- **Cost Savings**: K2view data Fabric will reduce the hardware infrastructure size and cost by 80% and dramatically reduce the time and cost for new projects.

- **Speed to Market**: New services delivered in days vs 6 months average TTM.

Case Study

AT&T accelerates speed to market with self-service test data provisioning

AT&T

Industry: **Telco**

Location: **USA**

Product: **K2View Test Data Management**

1,000 Testers
Across AT&T

+79%
Test automation coverage

-70%
Cost of automated regression testing

609
Applications being tested

Challenge

- Process to create, secure and provision data for testing was cumbersome, lengthy & manual
- Speed-to-market for development Cycles - typically 30-45 days just to make test data available
- Budgets - Costs to support TDM were rising & budget cuts were required to bring them in-line.

Solution

- K2View Test Data Management
- Easily integrated with existing tools and apps
- Self-service portal for testers to easily define required test data
- Automates manual processes around collecting, securing & provisioning test data

Results

- Improved speed-to-market by 80%
- Self-service test data creation in minutes
- Reduction in resources needed for manual processes
- Simple, one-time configuration

Case Study

Verizon saves millions by improving network alarming with event correlation

Industry: Telco

Location: USA

Product: K2View Data Fabric

2,000 Dispatches
Avoided/month

250M
Alarms/month

75,000
Records/day

Challenge

- Increasing Average Handling Time (AHT) to research for outages
- Unnecessary customer dispatch
- Delay in handling outage information to mitigate customer support calls
- Data is available, but cannot be processed in real time for optimal utilization

Solution

- Real-time correlation of upstream network outages to customers
- Real-time aggregation of customer-level alarms to identify fiber cuts
- Real-time calculations on Customer Premise Equipment connection and performance data

Results

- Alarm correlation & aggregation generates 2,000 decrease in redundant dispatches per month
- Significant improvement in Average Handling Time (AHT), and customer satisfaction
- Device metrics provides insight to allow for improved AI around Customer Premise Equipment troubleshooting and dispatch management
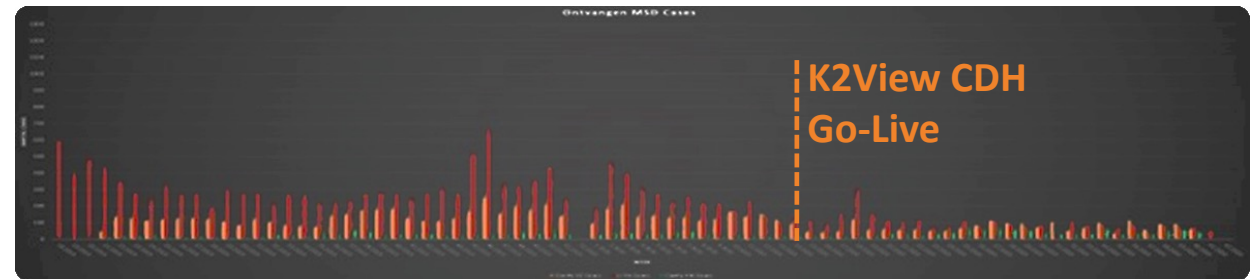
Case Study

K2 VIEW.

Vodafone Ziggo cuts time for eligibility checks & service provisioning from 24hr to minutes

vodafone ● Ziggo

Industry: **Telco**

Location: **Netherlands**

Product: **K2View Customer Data Hub (CDH)**

K2View CDH Go-Live

Solution

- Increase in customer churn and dissatisfaction
- Information is not up-to-date and lack of real-time synchronization
- Integration of data spread across multiple systems from both companies

Results

- Integration of real-time customer data, directly from the source, into a holistic view
- Provision of data at any time, in any format, for any application, in any place (e.g., web portal, IVR, field technician tools)
- Flexibly support additional business needs of all kinds (B2B,B2C)

20M
Subscribers

70% Decrease
Inbound service tickets

Increase
Customer satisfaction

Case Study

DIRECTV wins big with customers while cutting costs

DIRECTV

Industry: **Telco**

Location: **Latin America**

Product: **K2View Customer Data Hub (CDH)**

3 days
To initial deployment

13M
Subscribers

$5+ M
Annual savings

Challenge

- Customer data scattered across multiple countries and systems
- Customers had **poor experience with self-service portal** because of significant data latency issues
- Legacy architecture, costly licensing fees, and long development cycles caused a high TCO

Solution

- K2View CDH integrates and unifies data from **8 countries**
- K2View CDH hosted on **Oracle Cloud**, source systems remain **on-premise**
- Customer-facing web applications were integrated directly
- **Data masking** applied to ensure security

Results

- Cost-Savings: Immediate and on-going cost reduction of more than **$5M/ year**
- Performance: Data latency was cut **from minutes to milliseconds**
- Speed: Quick time to market, improved customer experience
- Know your customer: Unified 360-degree master view for 1M customers was created **in 3 days**
- Efficiencies: Customer Care cost reduction; efficient development; licensing fee savings

Making **every** customer experience personalized and profitable

K2 VIEW.
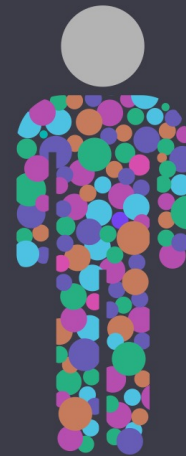
Our secret sauce:

The Digital Entity
Connects Everything

- Automate customer data processing: Access, Collect, Rectify, Erase
- Expedite, scale, and future-proof regulatory compliance
- Build customer trust with fast and complete response
- Increase operational efficiency saving time and resources