



Unify Analytics

Making Production Data Accessible for Data Science at Scale

Paige Roberts

Vertica Open Source Relations Manager

Analytics Use Cases are Everywhere

Companies that enable our data-driven world use both BI and data science.

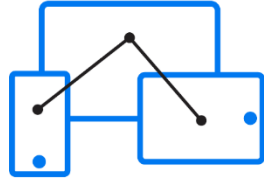
IoT Analytics

PHILIPS



AdTech Analytics

theTradeDesk



COVID-19 Tracking

DOMO



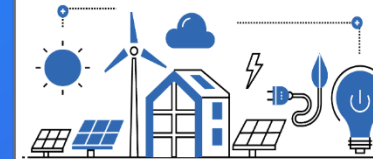
Customer Analytics



Energy Trading



ccee



Security Analytics

CYBERBIT
PROTECTING A NEW DIMENSION



Network Optimization

Hewlett Packard Enterprise



Predictive Maintenance

NATIONAL INSTRUMENTS



Energy Optimization

JAGUAR RACING
FORMULA-E TEAM



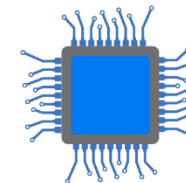
Smart Agriculture

THE CLIMATE CORPORATION



Software Optimization

intuit

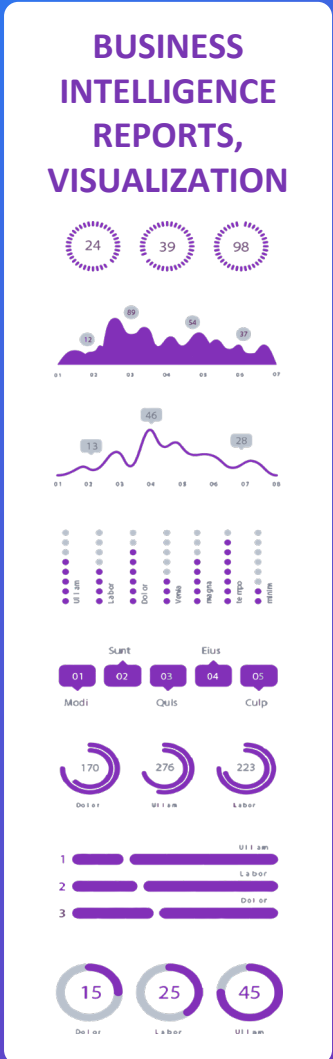
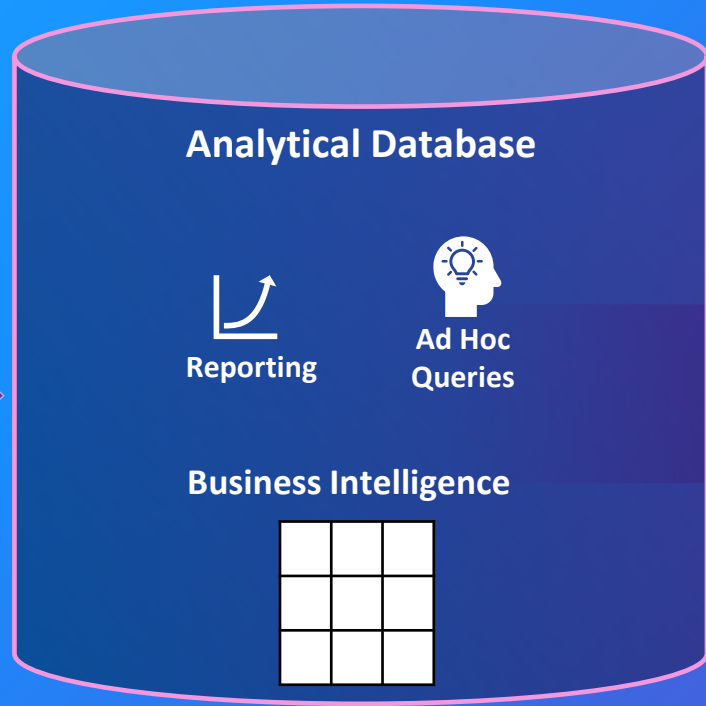
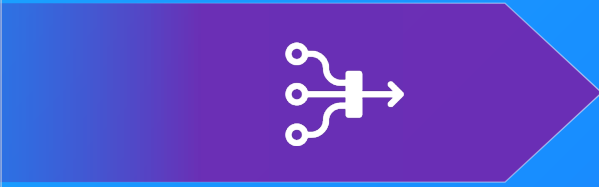
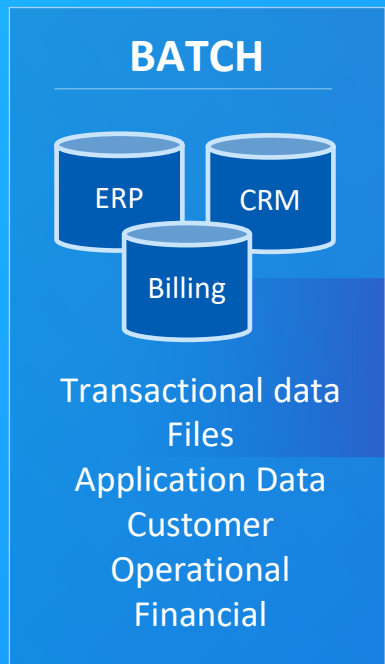


Clickstream Analytics

Taboola



Data Warehouse Architecture



Data Warehouse Strengths



DATA WAREHOUSE

Reporting/
Business Intelligence

High Performance

High Concurrency

Reliability

Security

Governance

SQL



Data Warehouse Weaknesses



DATA WAREHOUSE

Reporting/
Business Intelligence

High Performance

High Concurrency

Reliability

Security

Governance

SQL



WEAKNESSES

Expensive to scale

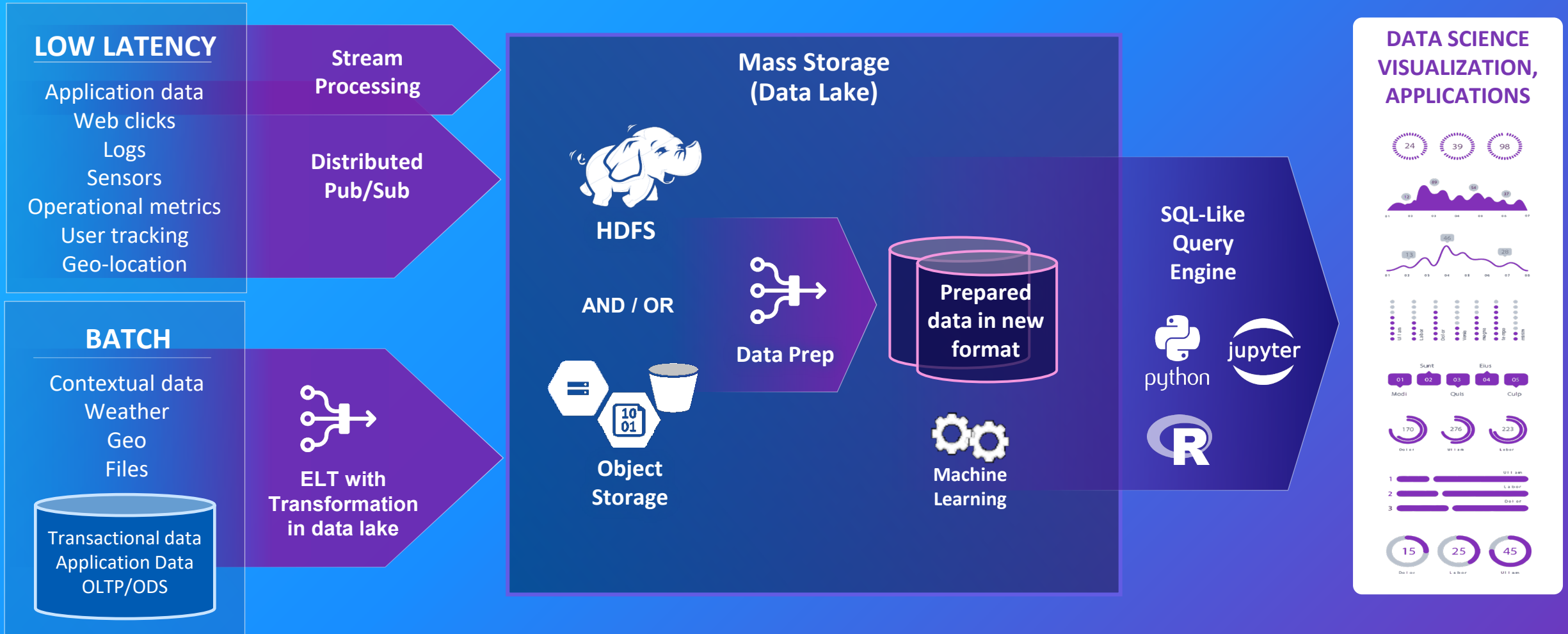
Structured data only

ETL caused data to be stale

Business intelligence only

Can't handle streaming data

Data Lake Architecture



Data Lake Strengths



DATA LAKE

Machine Learning/
Data Science

Unlimited Scale

Streaming Data

Semi-Structured Data
(JSON, AVRO, ...)

Complex Data Types
(Maps, Structs, Arrays)

Schema on Read

Python, R, Jupyter

Data Lake Weaknesses



WEAKNESSES

Slow performance

Poor concurrency

Complex to build / maintain

Immature reliability, security, governance

Difficulty operationalizing ML



DATA LAKE

Machine Learning/
Data Science

Unlimited Scale

Streaming Data

Semi-Structured Data
(JSON, AVRO, ...)

Complex Data Types
(Maps, Structs, Arrays)

Schema on Read

Python, R, Jupyter

Cooperative Architecture Strengths



DATA WAREHOUSE

Reporting/
Business Intelligence

High Performance

High Concurrency

Reliability

Security

Governance

SQL



DATA LAKE

Machine Learning/
Data Science

Unlimited Scale

Streaming Data

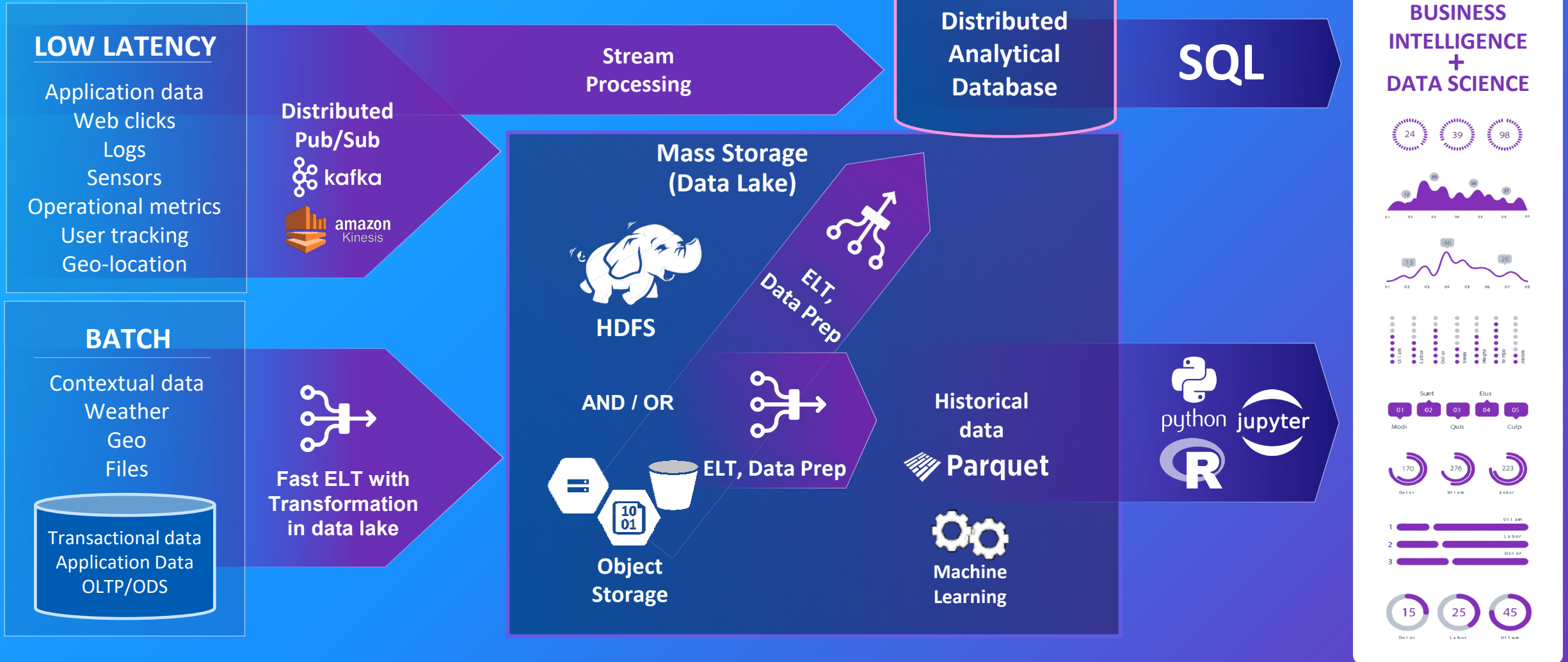
Semi-Structured Data
(JSON, AVRO, ...)

Complex Data Types
(Maps, Structs, Arrays)

Schema on Read

Python, R, Jupyter

Cooperative Architecture



LOW LATENCY

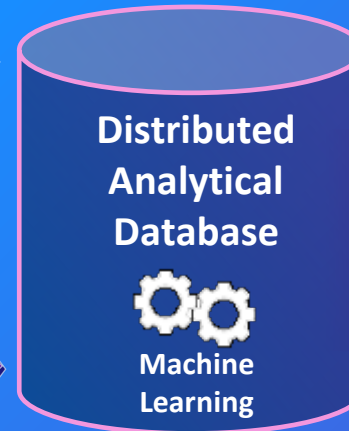
Planting and harvest equipment
Weather stations, probes, satellite imagery
Application data - clickstreams



Mass Storage (Data Lake)



Separate cluster for data Ingest, ETL



SQL

DATA SCIENCE + BUSINESS INTELLIGENCE



BATCH

Bayer research trials
Climate research farms (CRF)
Climate research partners (CRP)
FieldView data
Environmental data
Platform partner data
Sales data
Marketing campaigns



Cooperative architecture weaknesses



DATA WAREHOUSE

Reporting/
Business Intelligence

High Performance

High Concurrency

Reliability

Security

Governance

SQL

BI Visualization tools



WEAKNESSES

Complexity

Duplicated effort

Division of BI and Data Science

DATA LAKE

Machine Learning/
Data Science

Unlimited Scale

Streaming Data

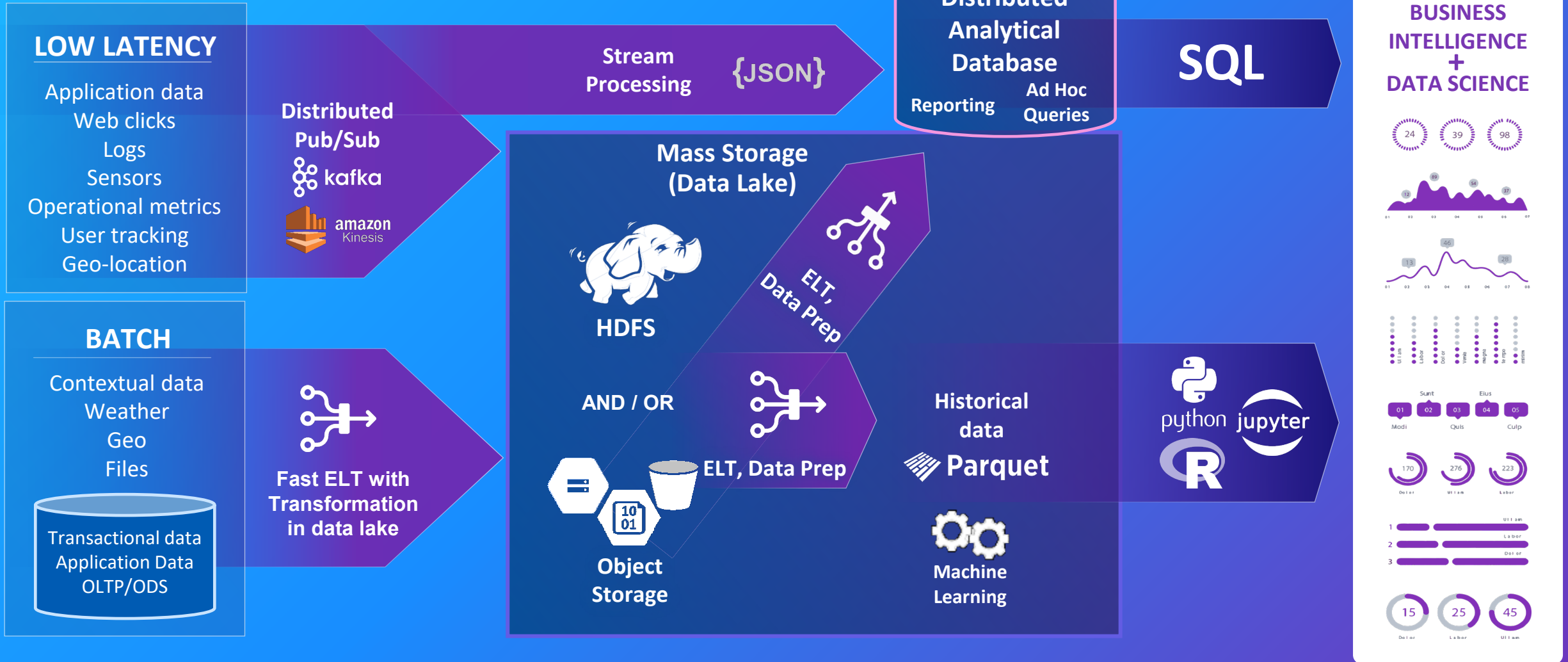
Semi-Structured Data
(JSON, AVRO, ...)

Complex Data Types
(Maps, Structs, Arrays)

Schema on Read

Python, R, Jupyter

Cooperative Architecture



Unify Storage Location



LOW LATENCY

- Application data
- Web clicks
- Logs
- Sensors
- Operational metrics
- User tracking
- Geo-location

Stream Processing

amazon Kinesis

BATCH

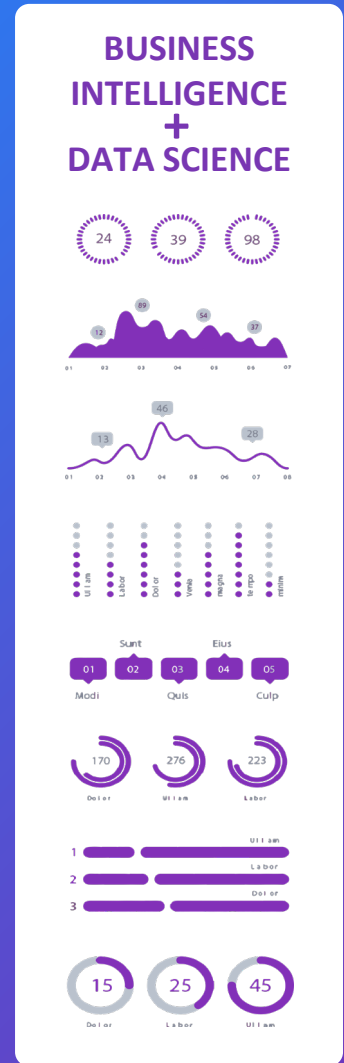
- Contextual data
- Weather
- Geo
- Files

Transactional data
Application Data
OLTP/ODS

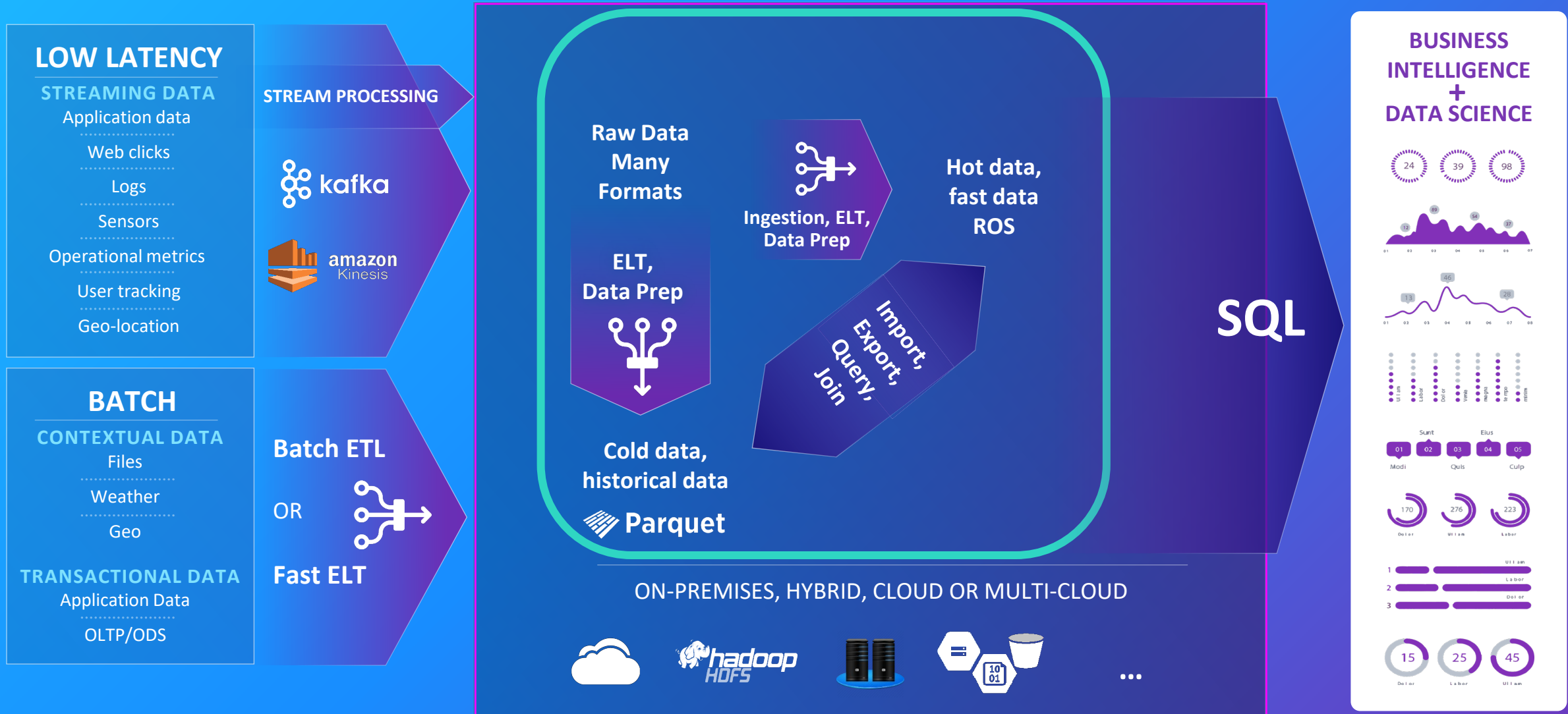
ELT



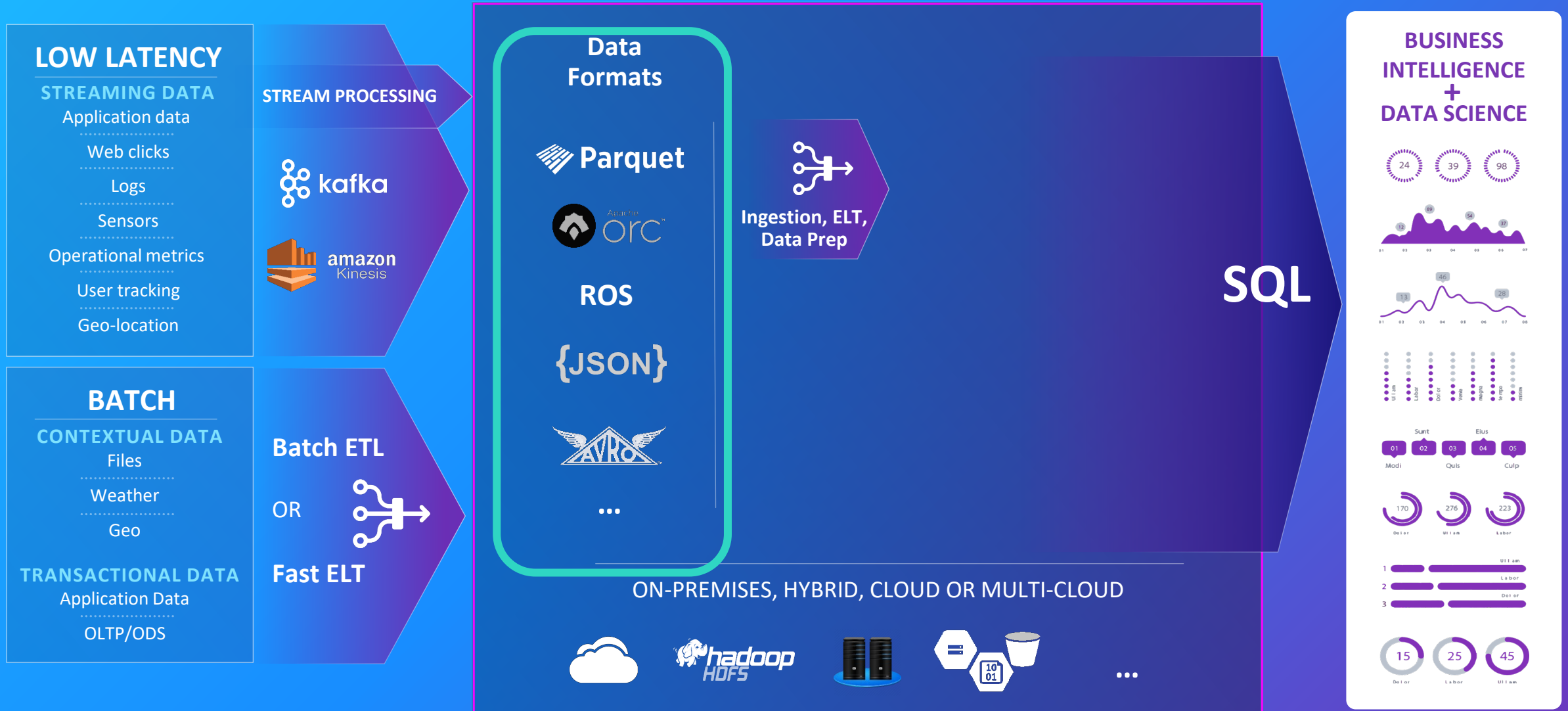
SQL



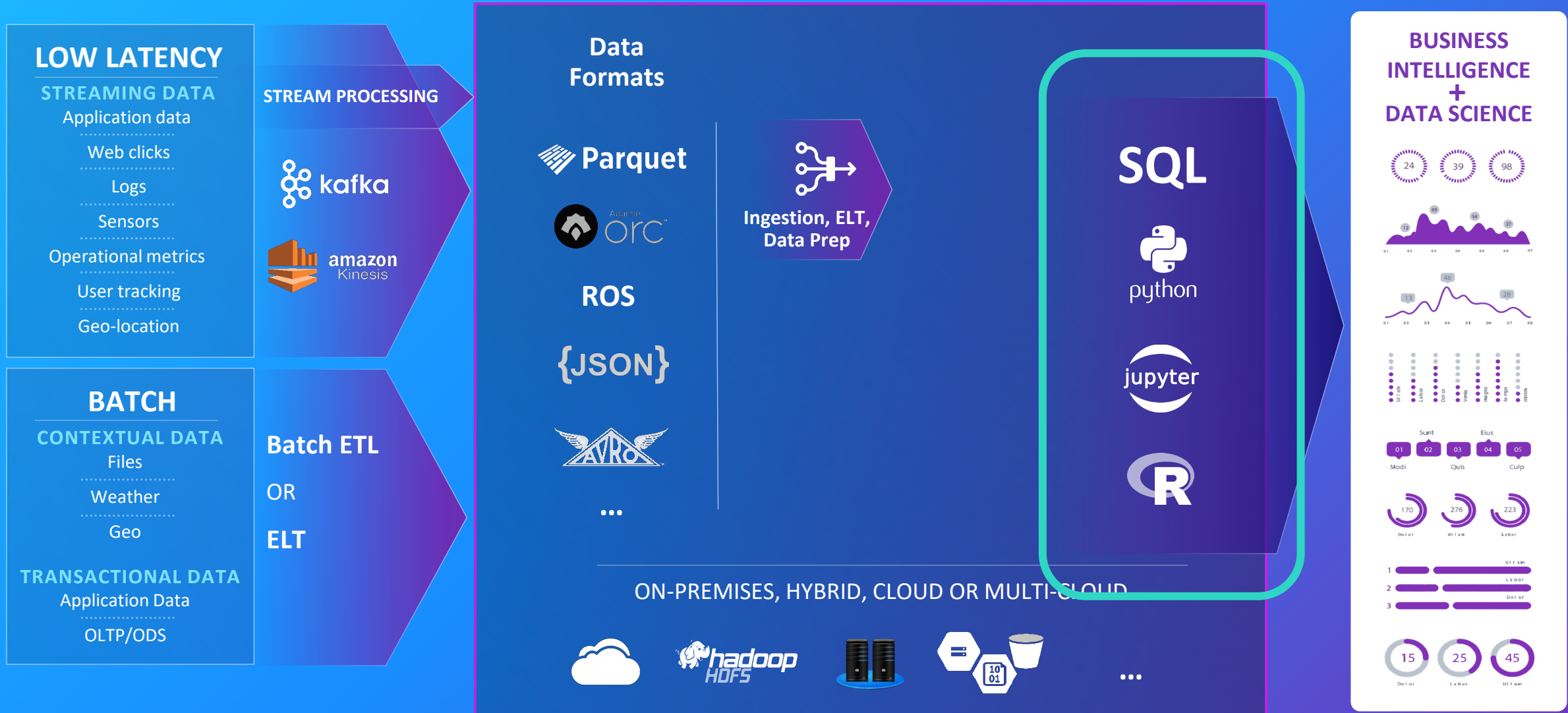
Manage Data Life Cycle with Data Formats



Separate Compute from Storage



Unify Analytics



Isolate Workloads



LOW LATENCY
STREAMING DATA
 Application data
 Web clicks
 Logs
 Sensors
 Operational metrics
 User tracking
 Geo-location

STREAM PROCESSING

BATCH
CONTEXTUAL DATA
 Files
 Weather
 Geo

TRANSACTIONAL DATA
 Application Data
 OLTP/ODS

Batch ETL

OR

Fast ELT

Data Formats

ROS

{JSON}

...

Isolated Compute Workloads

SQL

ON-PREMISES, HYBRID, CLOUD OR MULTI-CLOUD

...

BUSINESS INTELLIGENCE + DATA SCIENCE

24 39 98

Start Elus

01 02 03 04 05

Modi Qeis Culp

170 276 223

De Lor Ullam Labor

1 Ullam Labor

2 Labor De Lor

3 De Lor

15 25 45

De Lor Labor Ullam

Import/Export/Score/Manage models



LOW LATENCY

STREAMING DATA

- Application data
- Web clicks
- Logs
- Sensors
- Operational metrics
- User tracking
- Geo-location

STREAM PROCESSING

BATCH

CONTEXTUAL DATA

- Files
- Weather
- Geo

TRANSACTIONAL DATA

- Application Data
- OLTP/ODS

Batch ETL

OR

Fast ELT

Data Formats

- Parquet
- Apache ORC
- ROS
- {JSON}
- AVRO
- ...

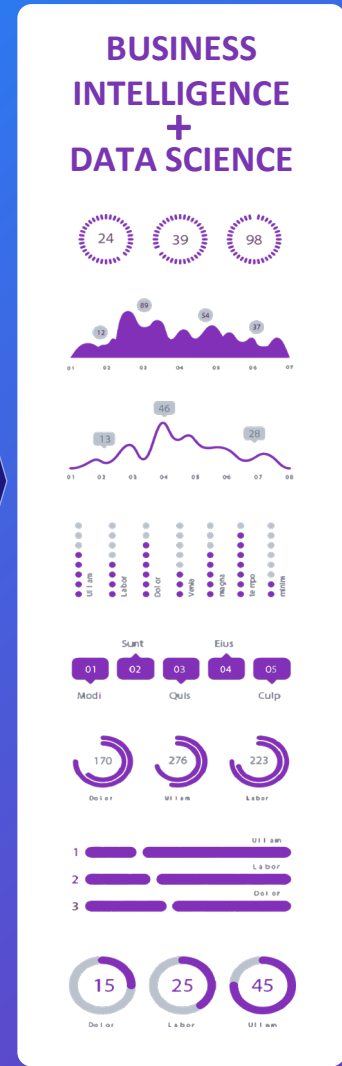
Isolated Compute Workloads

- Ingestion, ELT, Data Prep
- Machine Learning
- Reporting
- Ad Hoc Queries
- Model Evaluation, Deployment, Management

SQL

- python
- jupyter
- R

ON-PREMISES, HYBRID, CLOUD OR MULTI-CLOUD



Unified Analytics Platform



LOW LATENCY

STREAMING DATA

Application data

Web clicks

Logs

Sensors

Operational metrics

User tracking

Geo-location

STREAM PROCESSING



BATCH

CONTEXTUAL DATA

Files

Weather

Geo

Batch ETL

OR



Fast ELT

TRANSACTIONAL DATA

Application Data

OLTP/ODS

Data Formats



ROS



...

Isolated Compute Workloads



Machine Learning



Ad Hoc Queries



Reporting



Model Evaluation, Deployment, Management

SQL



python



ON-PREMISES, HYBRID, CLOUD OR MULTI-CLOUD



...

BUSINESS INTELLIGENCE + DATA SCIENCE



Unified Analytics Warehouse Strengths

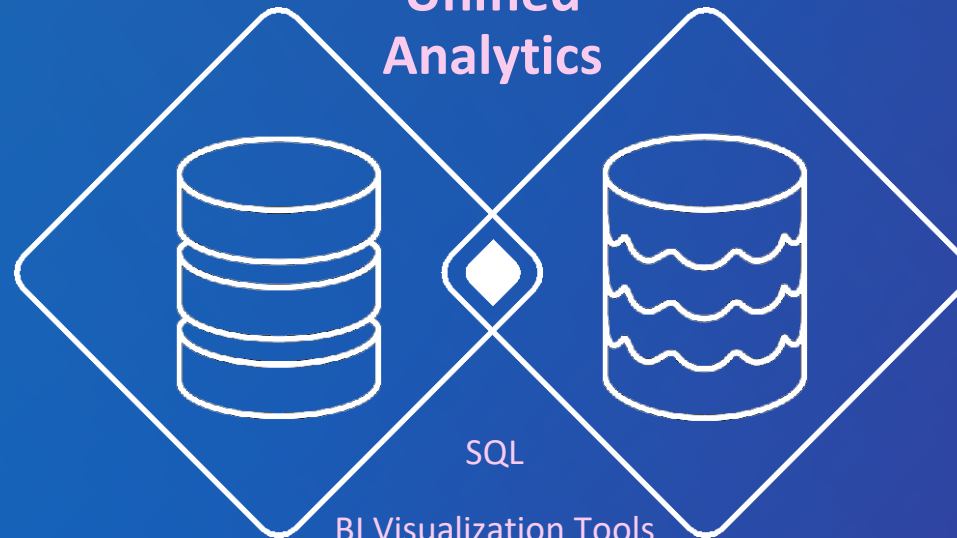


VERTICA

DATA WAREHOUSE

- Reporting/
Business Intelligence
- High Performance
- High Concurrency
- Reliability
- Security
- Governance

Unified Analytics



SQL

BI Visualization Tools

Python

Jupyter

R

DATA LAKE

- Machine Learning/
Data Science
- Unlimited Scale
- Streaming Data
- Semi-Structured Data
(JSON, AVRO, ...)
- Complex Data Types
(Maps, Structs, Arrays)
- Schema on Read



LOW LATENCY



Philips Remote Service Network

STREAM PROCESSING

Mass Storage (Data Lake)



IBM Spectrum Scale



Distributed Analytical Database

SQL



BUSINESS INTELLIGENCE + DATA SCIENCE



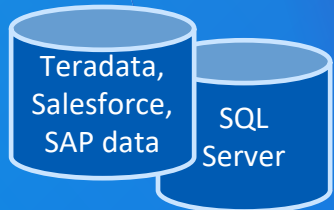
R & D Access

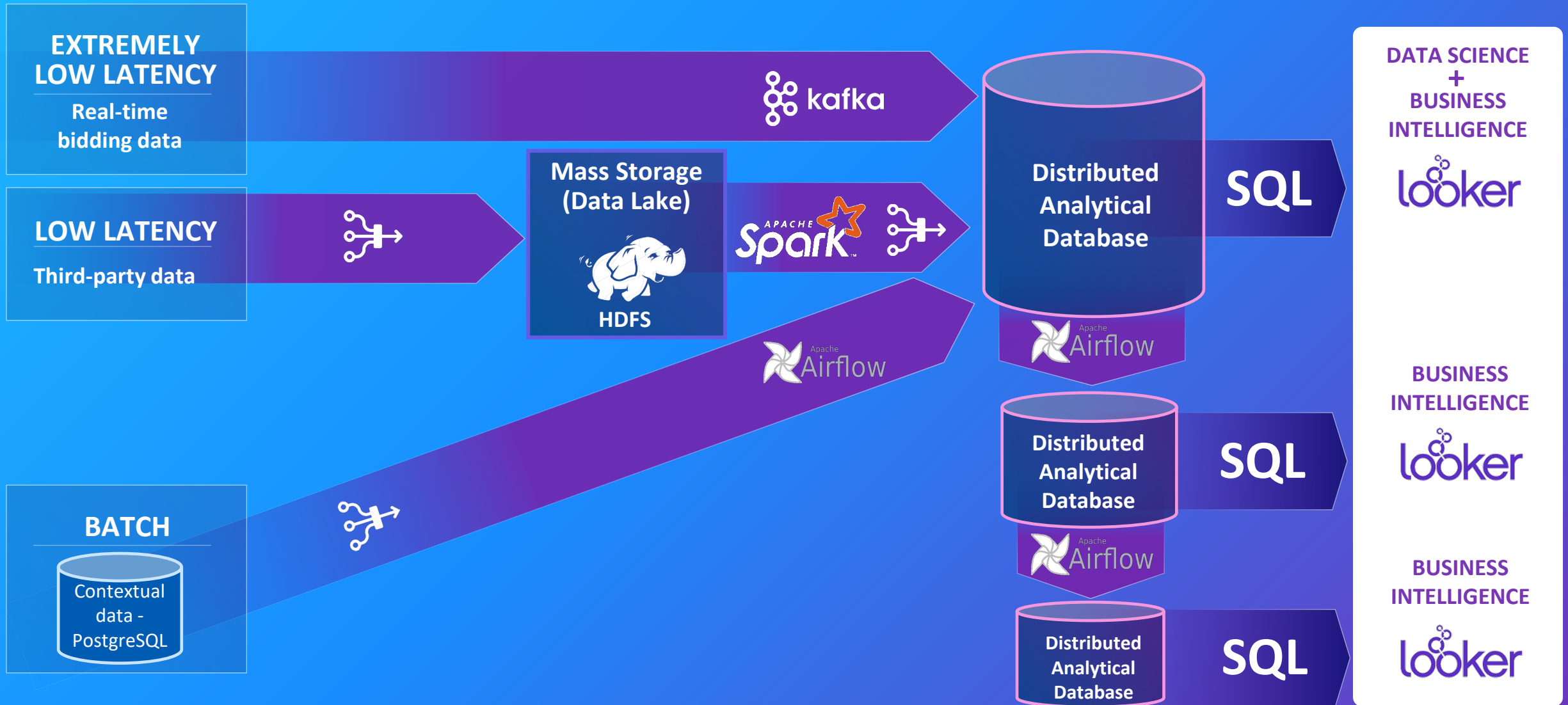
Remote Monitoring

Remote Service

BATCH

CRM data
Repair shop data
Factory data







UNITED GROUP



VERTICA

LOW LATENCY

EON TV Service
On/Off, Channel
Change Data



BATCH



Contact Center
Geo/Mapping
Customer
CX
Operational
Financial
Service Quality
TV Schedule



Informatica



Extract, Load

Transformation
pushed down



Ingest,
Transform,
Data Prep

Unified Analytics Platform



Machine
Learning

Real-time ad targeting

Customer profile analysis
for real-time apps



Ad Hoc
Queries



Reporting

SQL

DATA SCIENCE
+
BUSINESS
INTELLIGENCE



Content Analytics



Reporting



Financial Reporting



Data-driven Apps



EXTREMELY LOW LATENCY
Real-time bidding data

LOW LATENCY
Third-party data

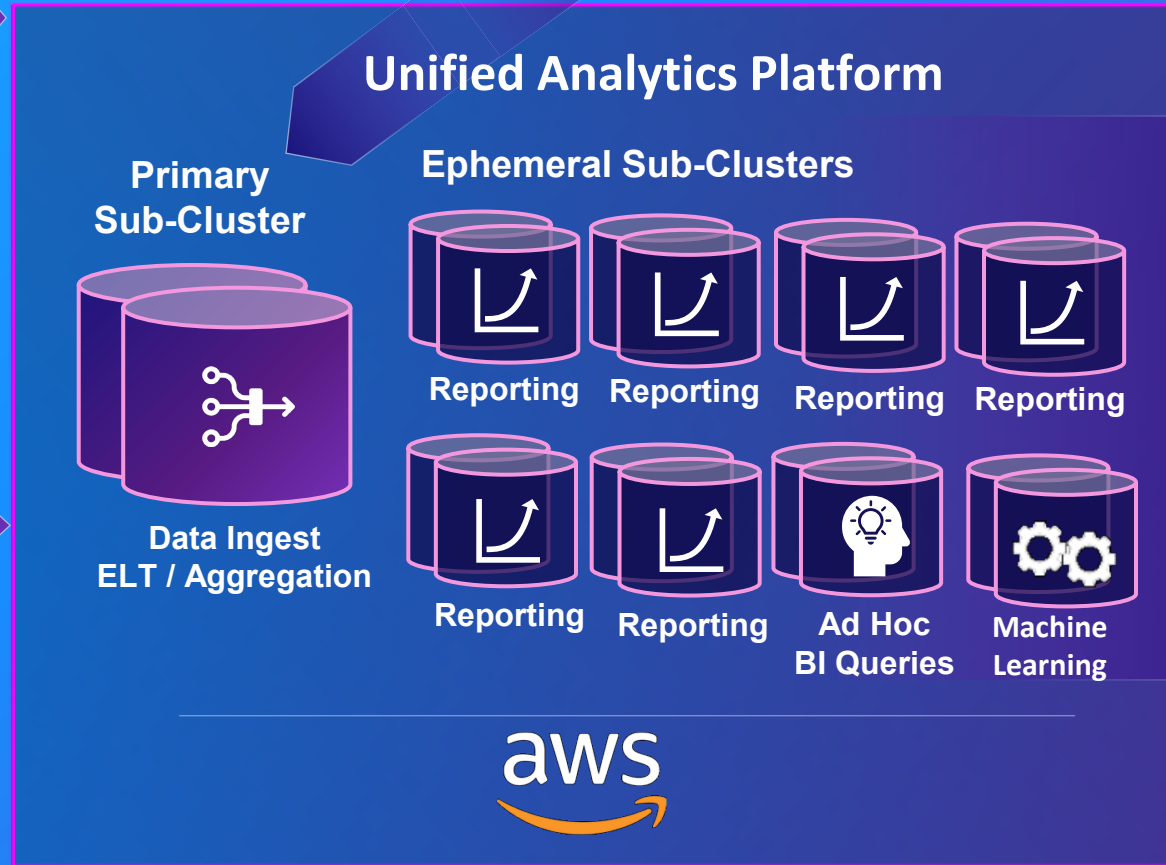
BATCH
Contextual data



SQL
python

APPLICATION
ML APPLIED

BUSINESS INTELLIGENCE
ML TRAINING
REPORTING



SQL
python

EMA Radar Report: Unified Analytics Warehouse A Guide for Investing in Unified Analytics



Request your copy today:
<https://www.vertica.com/success-ema-radar-report/>

VERTICA



Q&A

Learn More: academy.vertica.com

Try it Free: vertica.com/try

Paige Roberts
Open Source Relations Manager
E: Paige.Roberts@vertica.com

VERTICA



Thank you

Paige Roberts
Open Source Relations Manager
E: Paige.Roberts@vertica.com