Streaming History

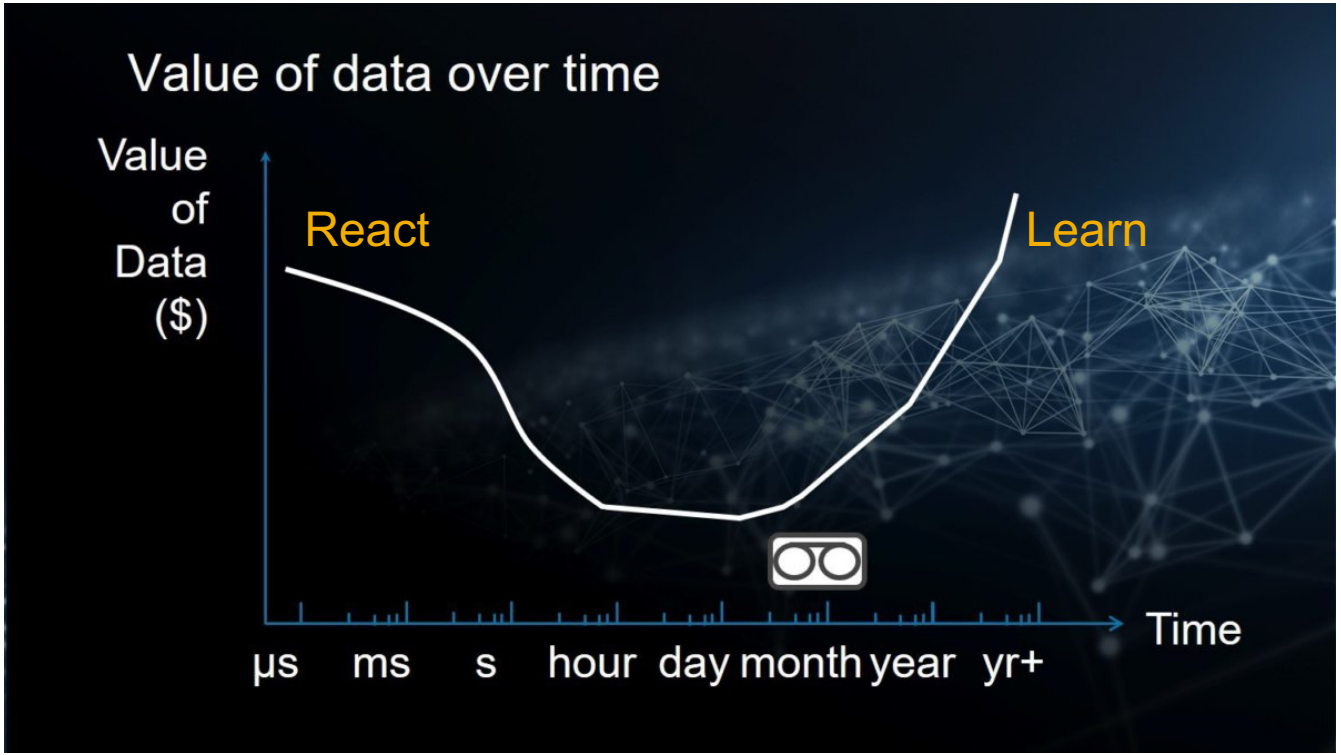
Minimizing Complexity to Maximize Value

Russ Caldwell

Dell EMC ▪ Product Management

DELL Technologies

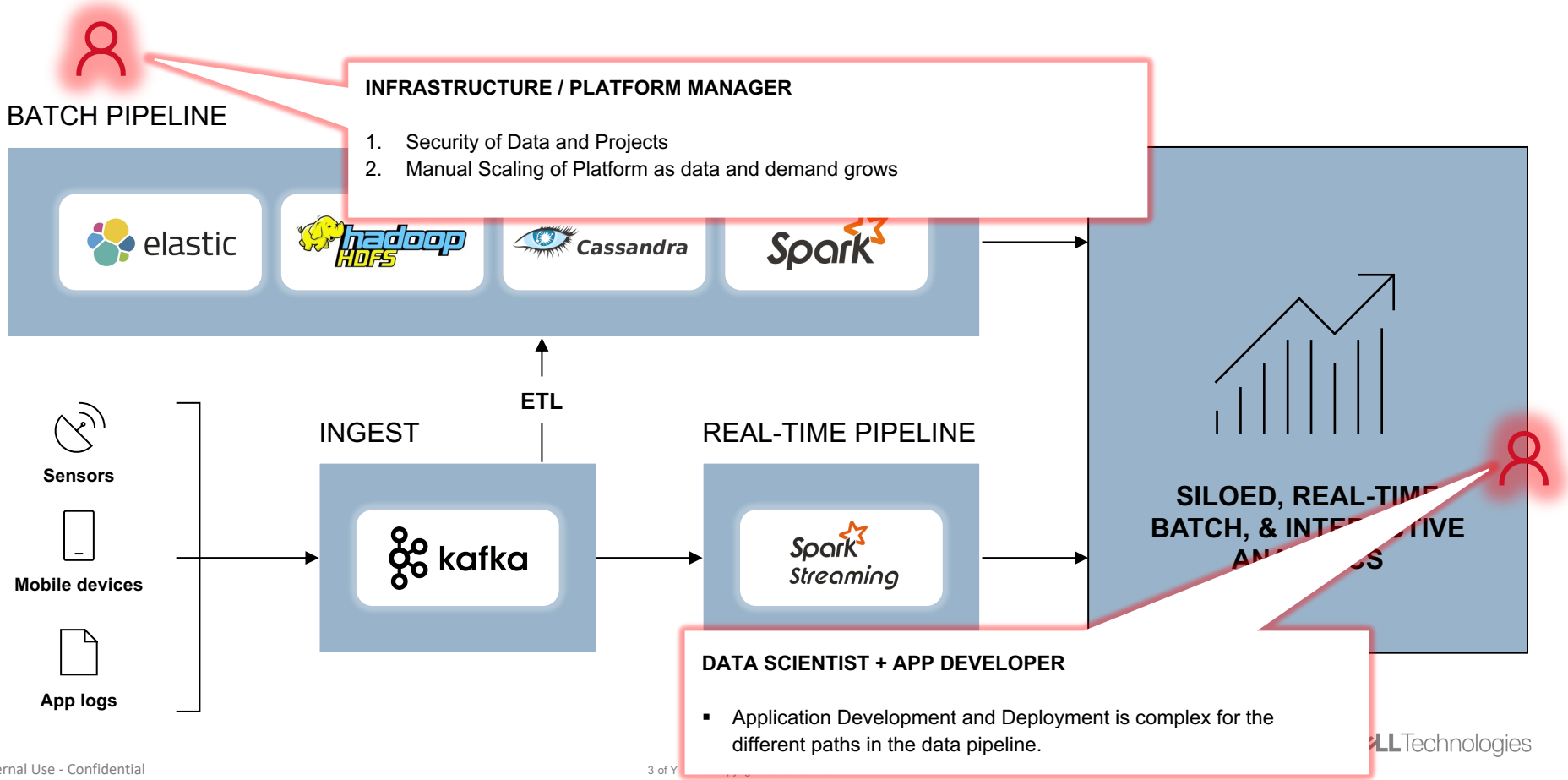
Value of data over time



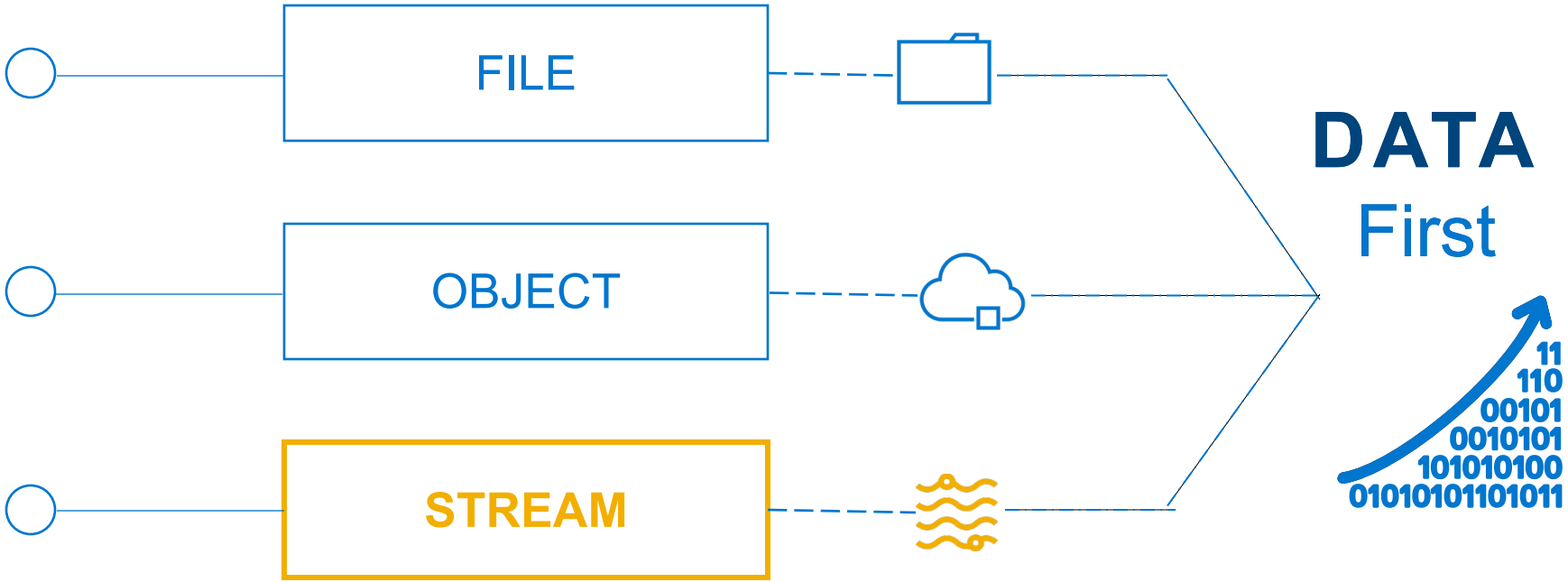
https://www.cio.com/article/3314742/trends-in-big-data-and-artificial-intelligence-data.html

© Copyright 2020 Dell Inc.

Today's DIY stream solutions are complex



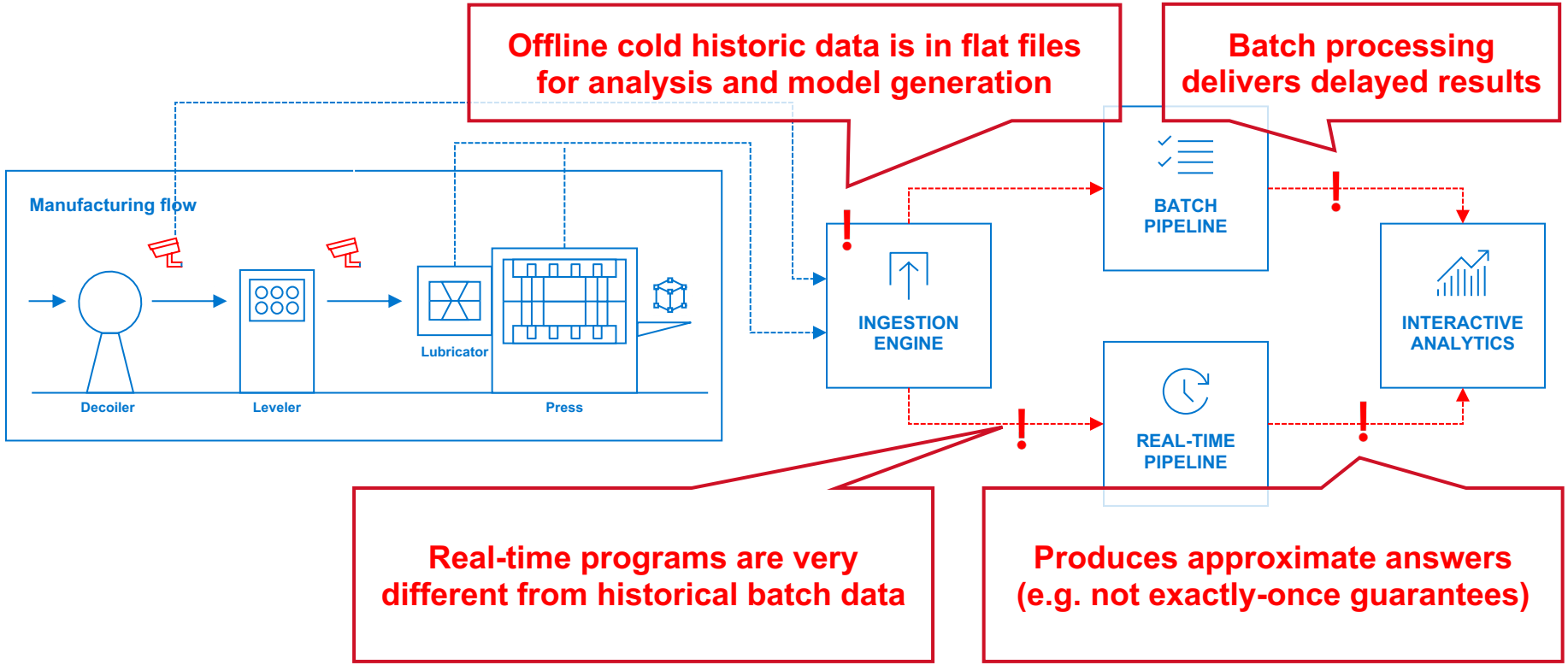
Stream: A First-Class Data Primitive



© Copyright 2020 Dell Inc.

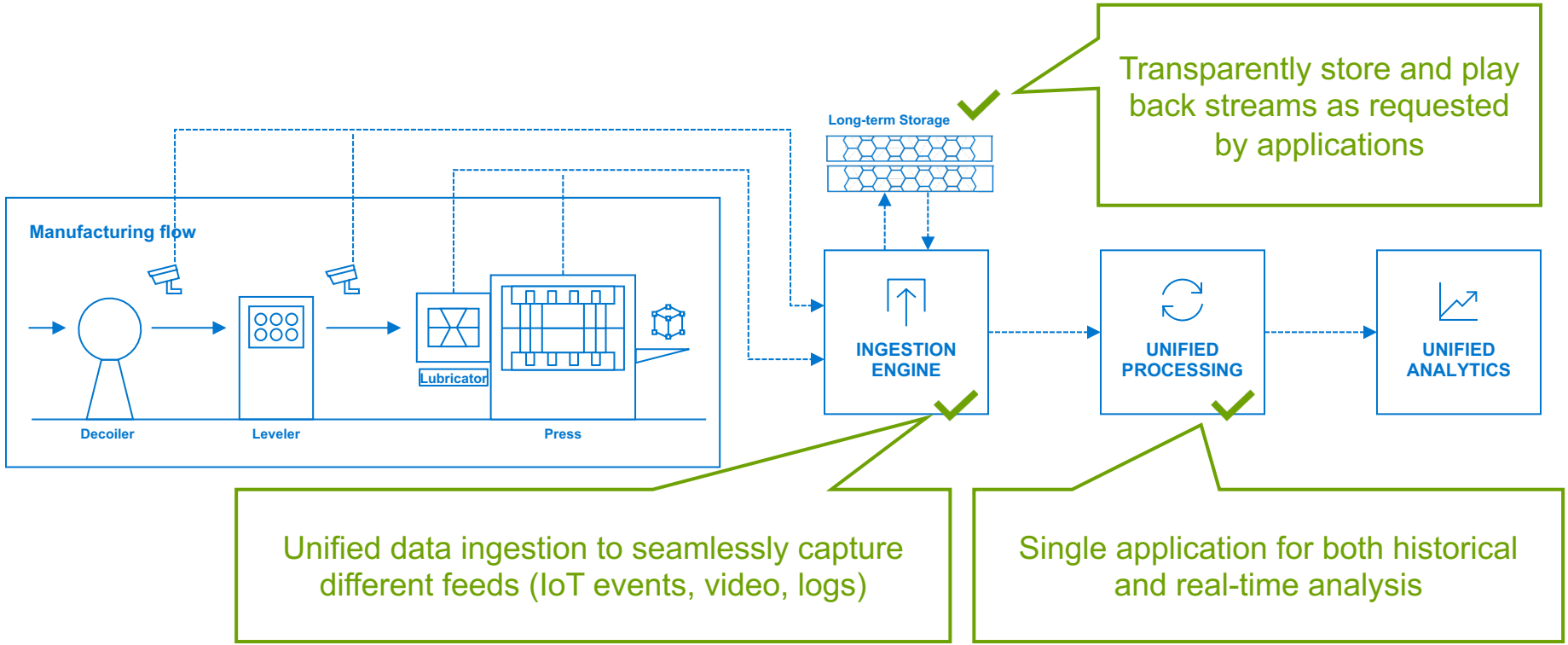
Industrial Sensor Data Anomaly Detection

DIY Lambda Architecture



© Copyright 2020 Dell Inc.

Industrial Sensor Data Anomaly Detection



© Copyright 2020 Dell Inc.

Streaming Storage

© Copyright 2020 Dell Inc.

DELL Technologies

Architectural Streaming Goals

A. Durable Unbounded Data Streams

- Data is **replicated and persisted to disk** before being acknowledged

B. Low-Latency Response AND Long-term Storage

- **Tier 1 Cache** for fast real-time reads, < 10ms writes
- **Long-term (Tier 2) Storage** should provide transparent access to unlimited historical data

C. Strict ordering guarantees, Exactly-once semantics & Transactions

- **In-Order and exactly-once** of events guaranteed
- SQL-like **Rollback and Commit** transactions within data streams

D. Enterprise Scale, Performance and Support

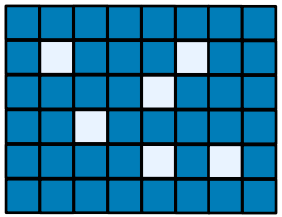
- Support **Tens of Millions** of streams
- **Dynamic partitioning** of streams based on load and throughput SLO
- Size is **not bounded** by the capacity of a single node
- Can easily develop and deploy production solutions

© Copyright 2020 Dell Inc.

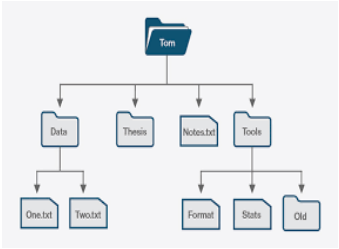
Storage Types

Traditional Apps/Middleware ⟷ Streaming Apps/Middleware

BLOCKS
- Structured Data
- Relational DBs

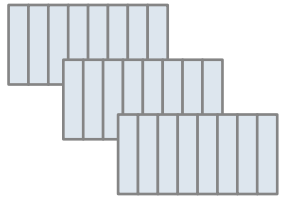


FILES
- Unstructured Data
- Pub/Sub
- NoSQL DBs



OBJECTS
- Unstructured Data
- Internet Friendly (REST)
- Scale over Semantics
- Geo

LOGS
- **Append-only**
- **Low-latency**
- **Tail Read/Write**

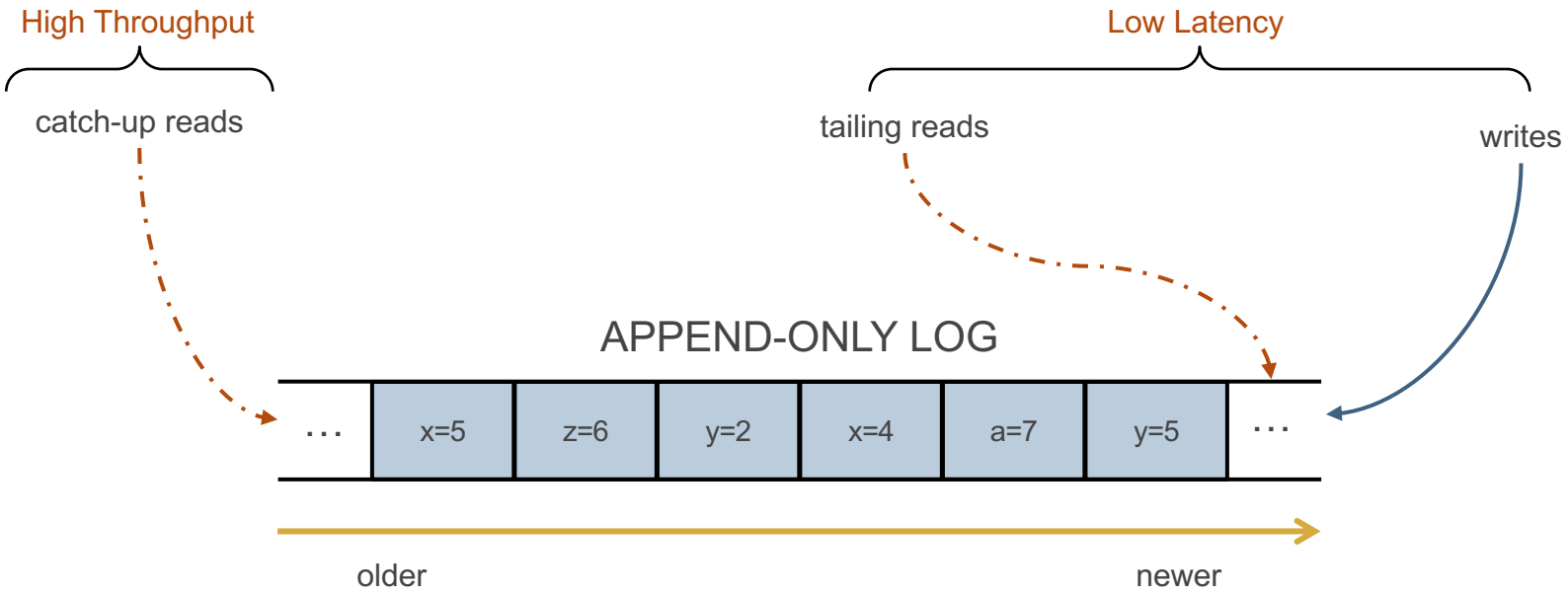


Dell - Internal Use - Confidential

Internal Use - Confidential

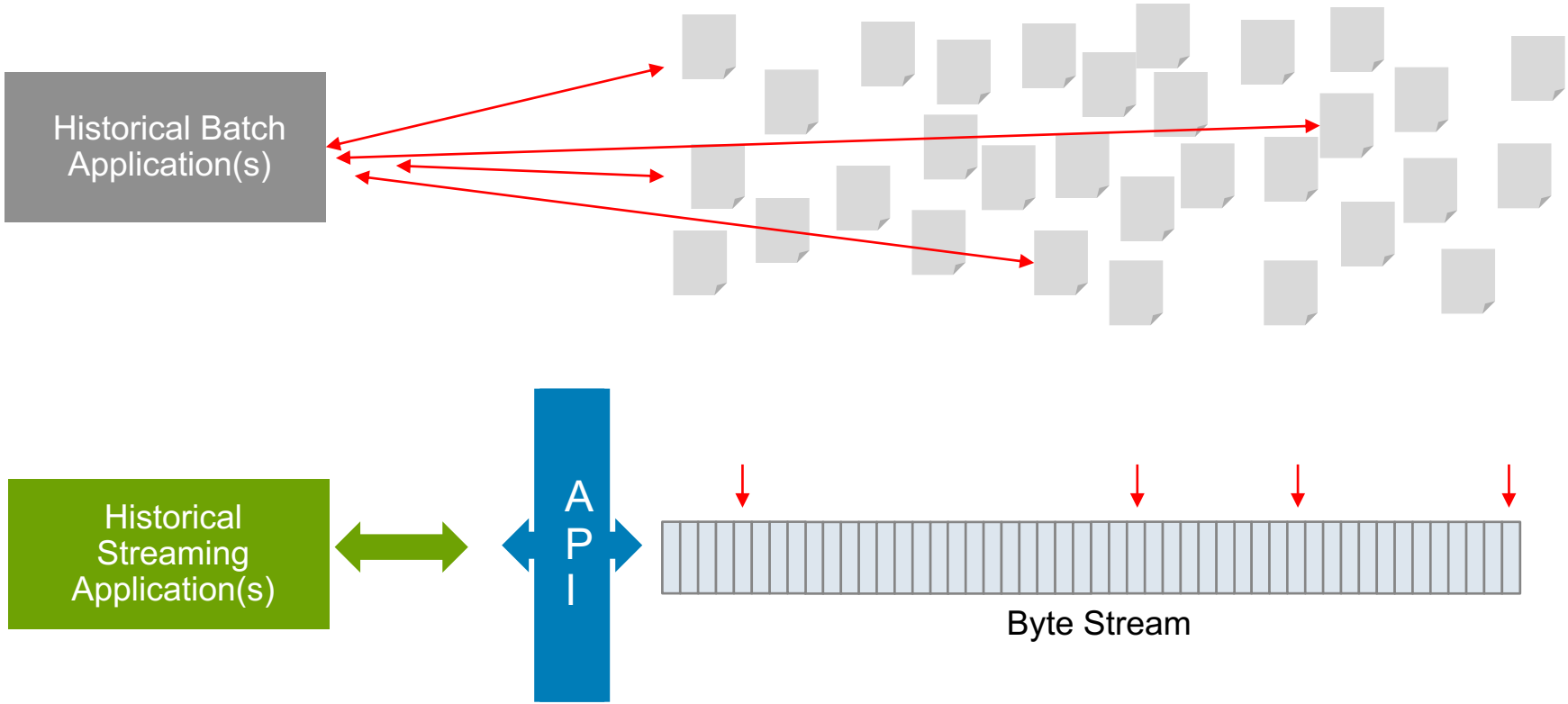
The Importance of Log Storage

The Fundamental Data Structure for Scale-out Distributed Systems



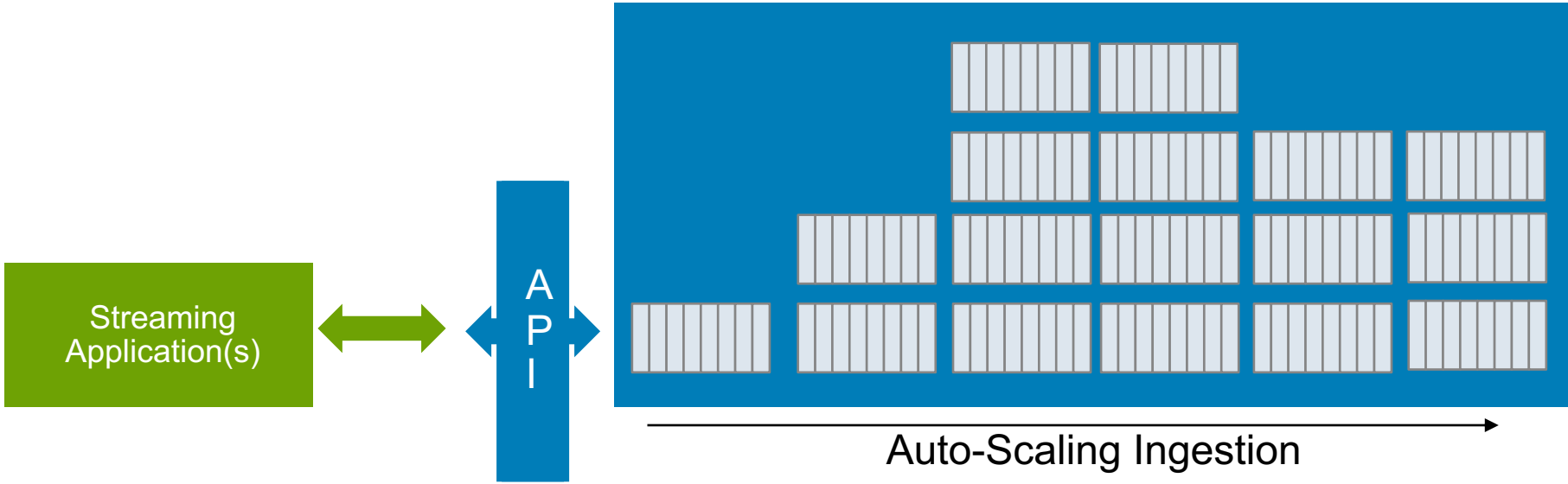
Internal Use - Confidential

Historical File/Object vs Stream Data Primitive



Internal Use - Confidential

Transparent Auto-Scaling



- Still guarantees in-order events
- Scales up & down based on ingestion rates

Dell - Internal Use - Confidential
Internal Use - Confidential
DELL EMC

Developing with Streams

© Copyright 2020 Dell Inc.

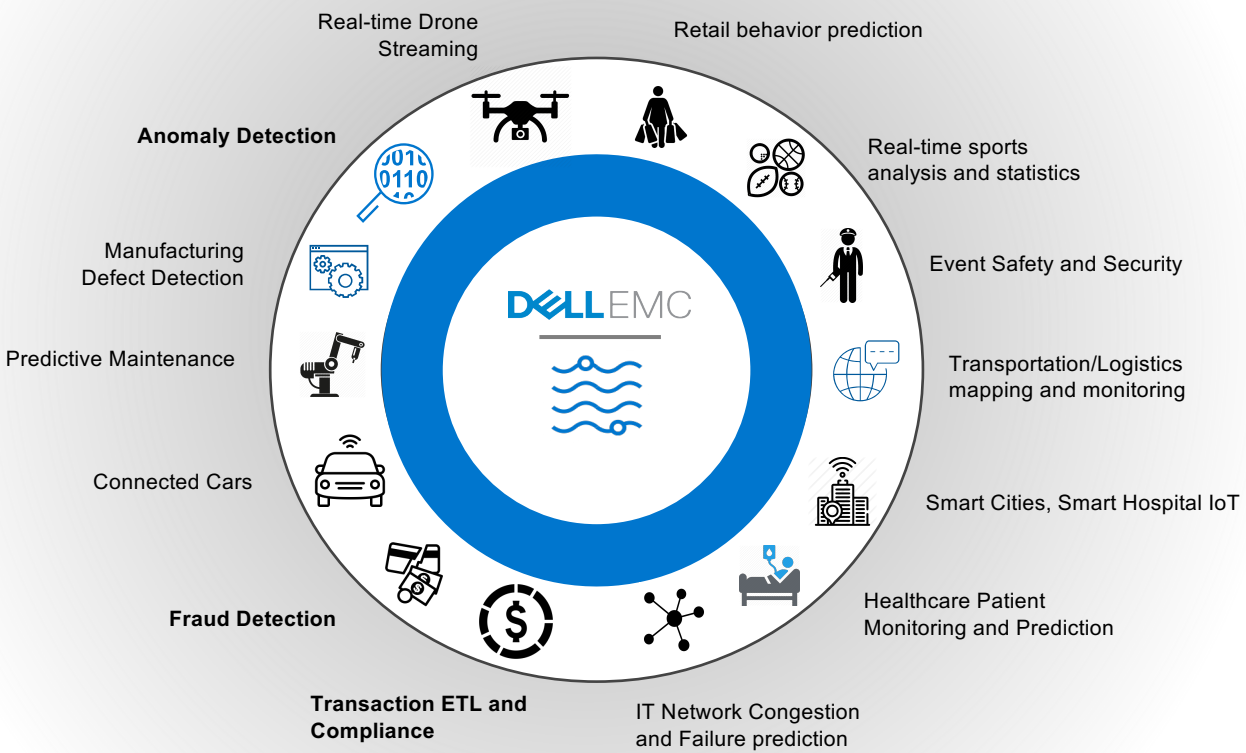
DELL Technologies

Use-Cases

© Copyright 2020 Dell Inc.

Streaming Data Platform

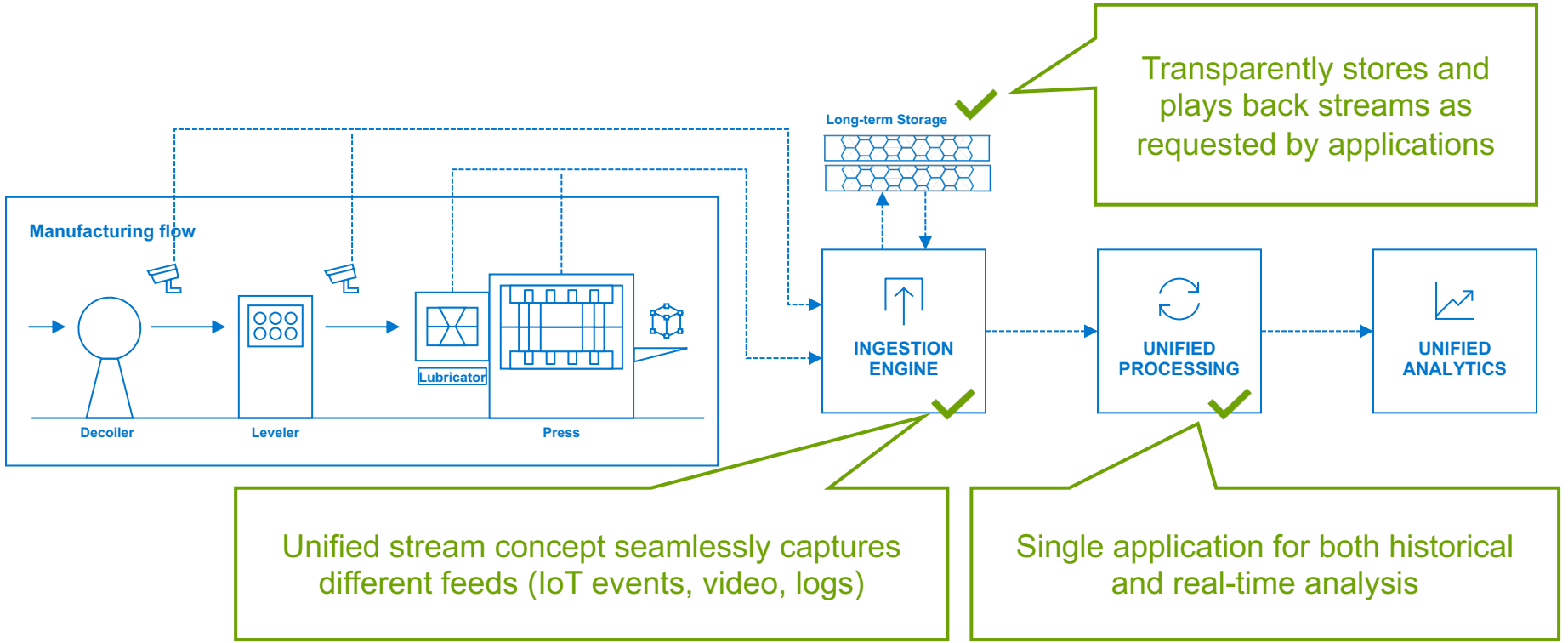
Uses Cases



© Copyright 2018 Dell Inc.
Internal Use - Confidential

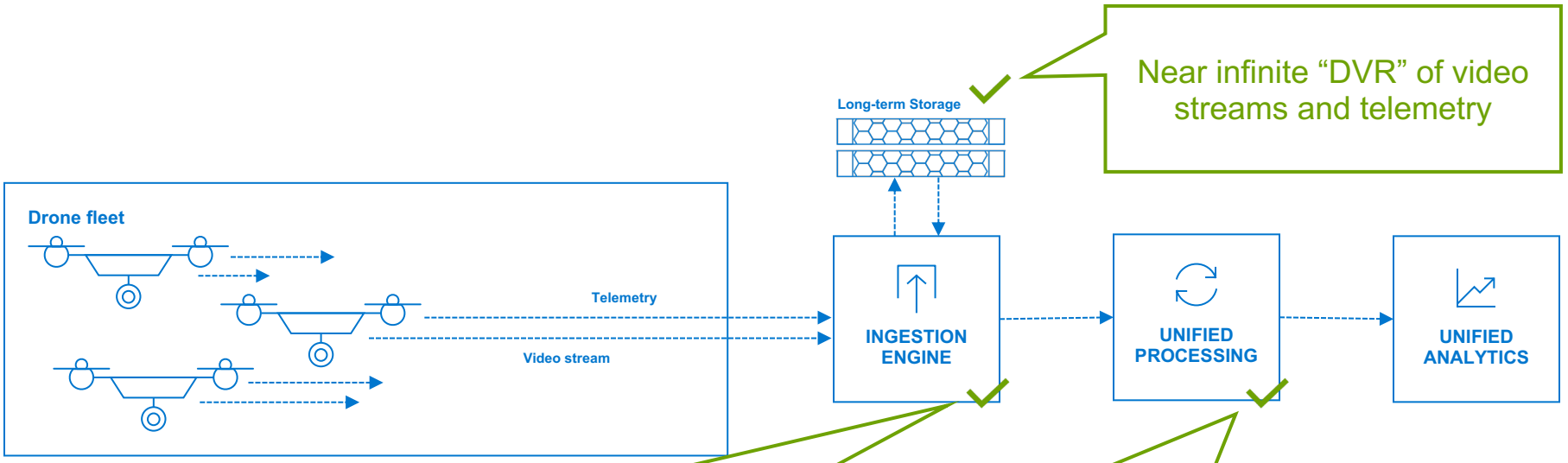
© Copyright 2020 Dell Inc.

Industrial Sensor Data Anomaly Detection



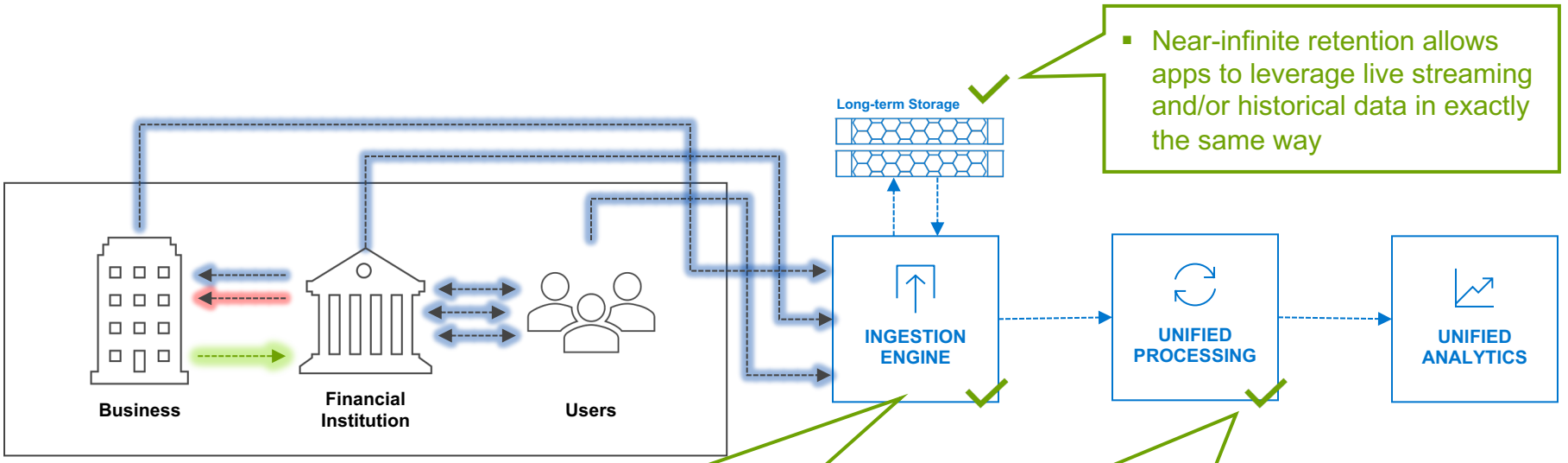
© Copyright 2020 Dell Inc.

Unbounded Data Streams



© Copyright 2020 Dell Inc.

Exactly once, high consistency transactions



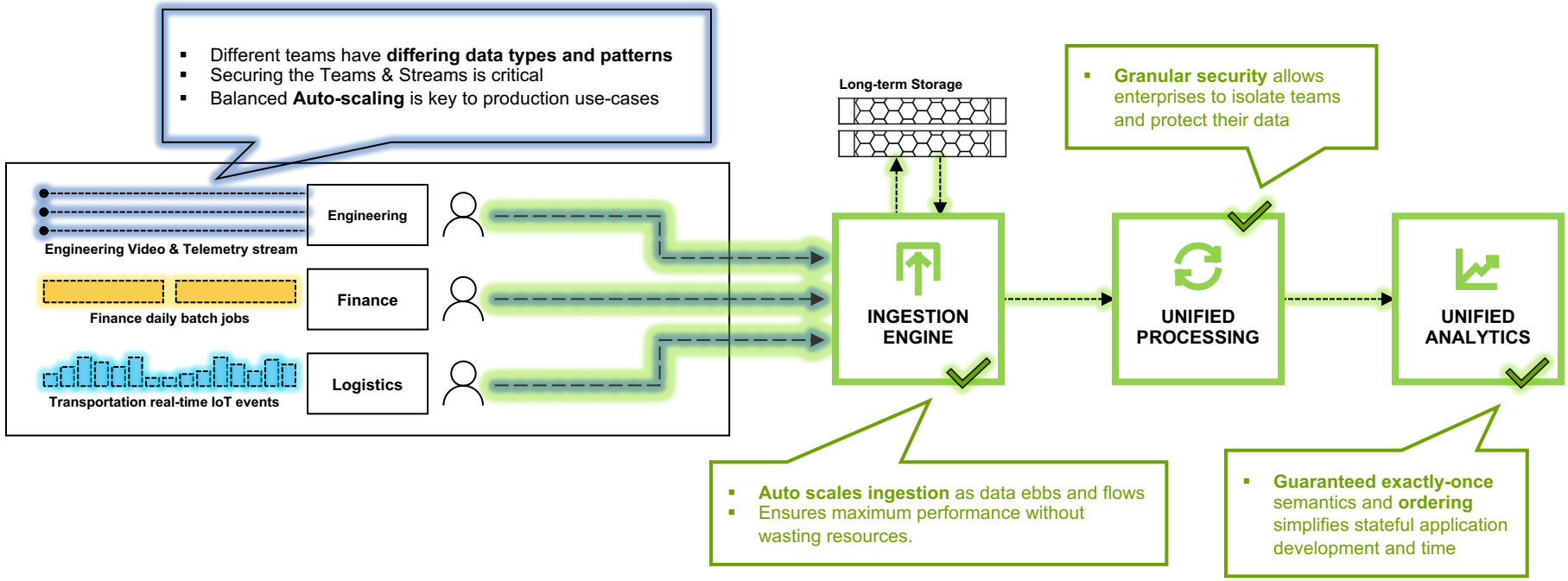
- Near-infinite retention allows apps to leverage live streaming and/or historical data in exactly the same way

- Portfolio Risk Management
- Fraud Detection and Anomaly Detection

- Unified batch and real-time applications easier to develop and maintain

© Copyright 2020 Dell Inc.

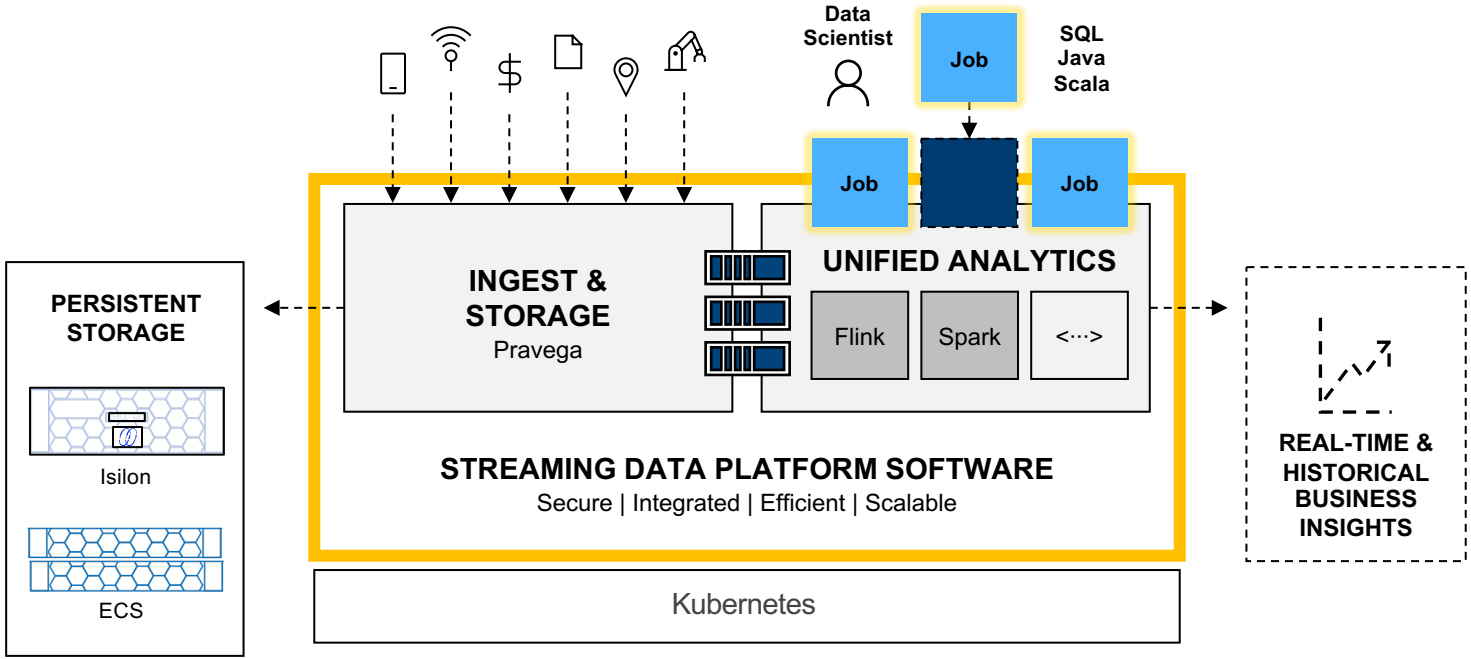
Multiple Teams & Use-Cases



© Copyright 2020 Dell Inc.

Streaming Data Platform

Architectural deep dive



© Copyright 2020 Dell Inc.

DELL Technologies

DELL Technologies