StoreEdge RippleStream

Versatile Infrastructure for IoT Data Transfer

Presenter: Peter Corbett
Authors: Madhumita Bharde, Annmarie Justine K, Suparna Bhattacharya, Dileep Deepa Shree
Affiliation: HPE HCI Engineering

USENIX HotEdge, July 10, 2018

Motivation: Versatile Edge Infrastructure

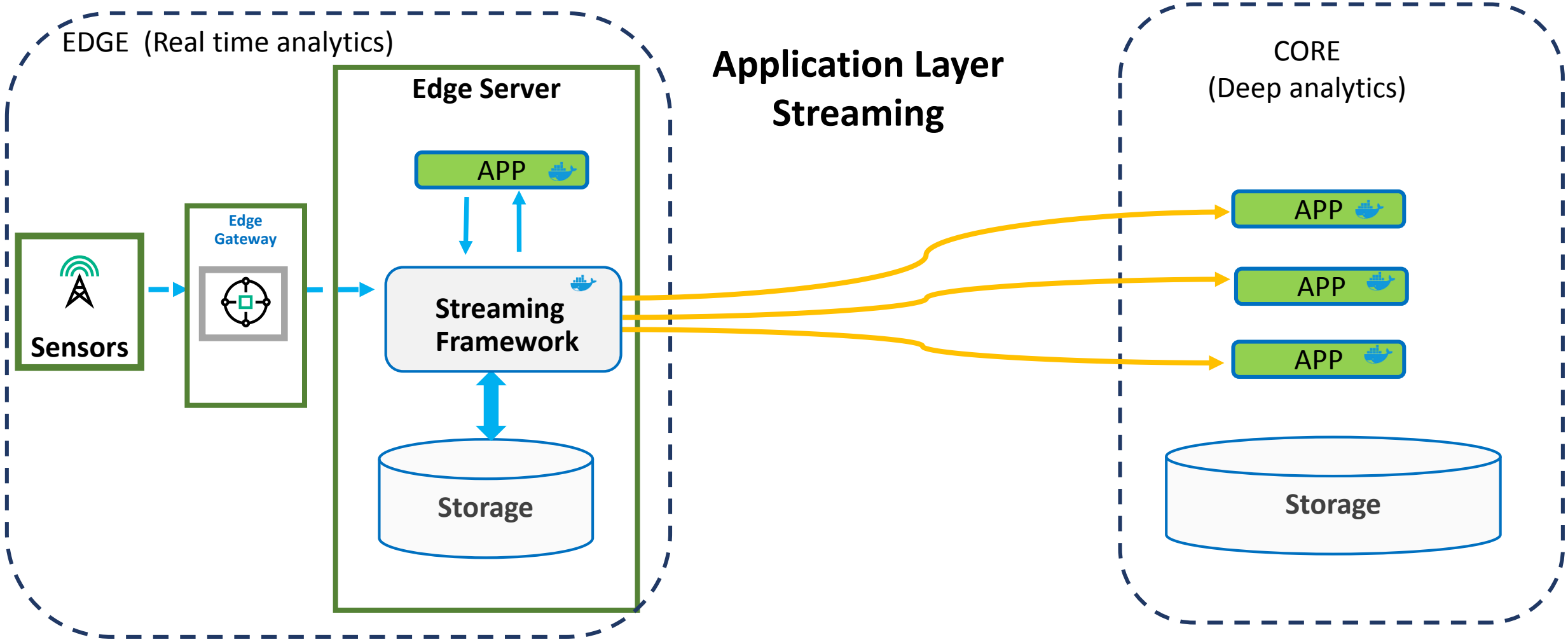
- Benefits of traditional software defined infrastructure
  - End to end management across a federation of edges + core
  - In-built data protection with asynchronous WAN optimized replication
  - High efficiencies for data-at-rest
    - compression (short range redundancies) + de-duplication (long range redundancies)
    - efficiencies improve over time when data has redundancies over days, weeks, months, ..
- Benefits of an edge computing optimized stack
  - Optimized for streaming data flows
  - Low latency action
  - Application defined filtering

What would it take to leverage the the best of both worlds?

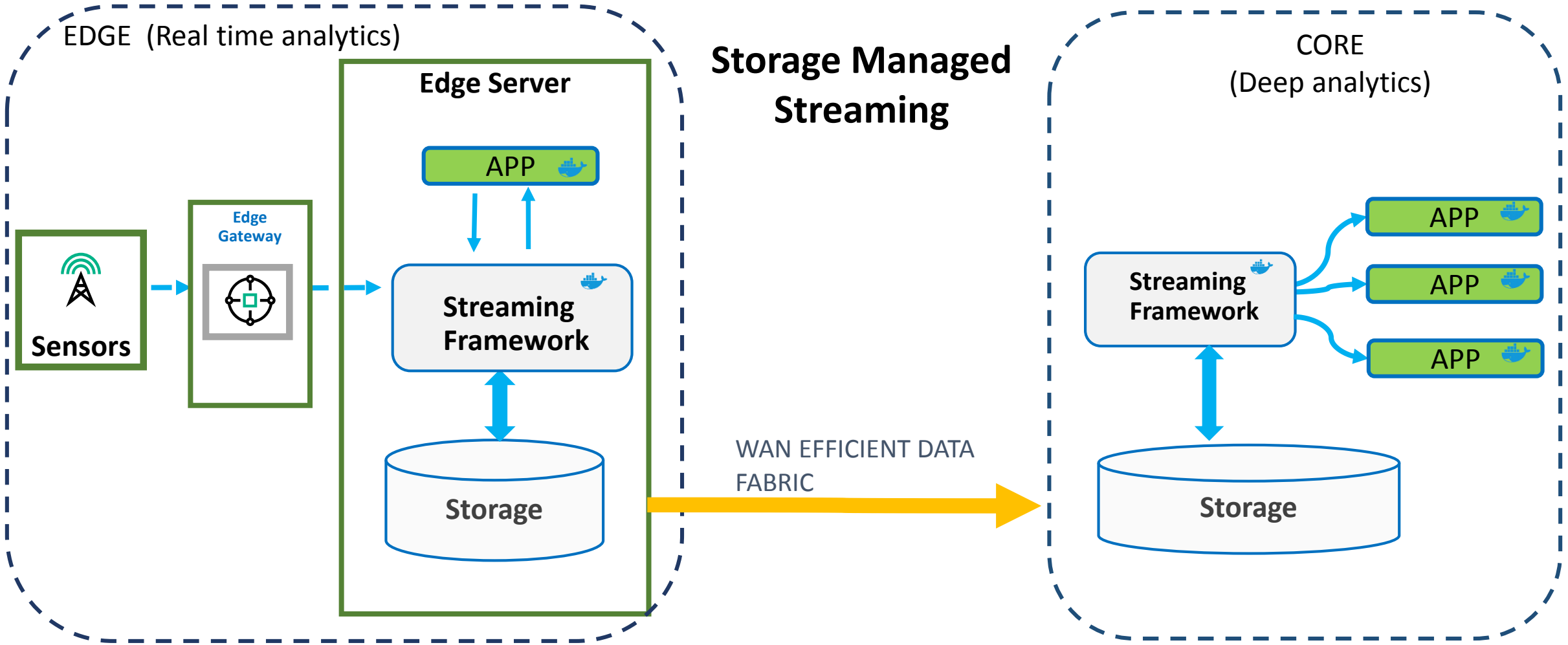
Motivation: Efficient IoT Data Transfer

- Most data generated at edge will never transmit to the core
- It is necessary to transmit significant information to the core
  - High value data with low redundancy
- Storage replication techniques provide a useful mechanism for high-value information transfer
  - Already well-established in commercial storage and HCI products
  - Eliminate redundancy via compression & deduplication of transmitted data
- Study considers combining storage replication with other (semantic) data reduction techniques to increase the information density of transmitted data
- Technique is suitable when significant latency is tolerable in transmitting edge data to the core

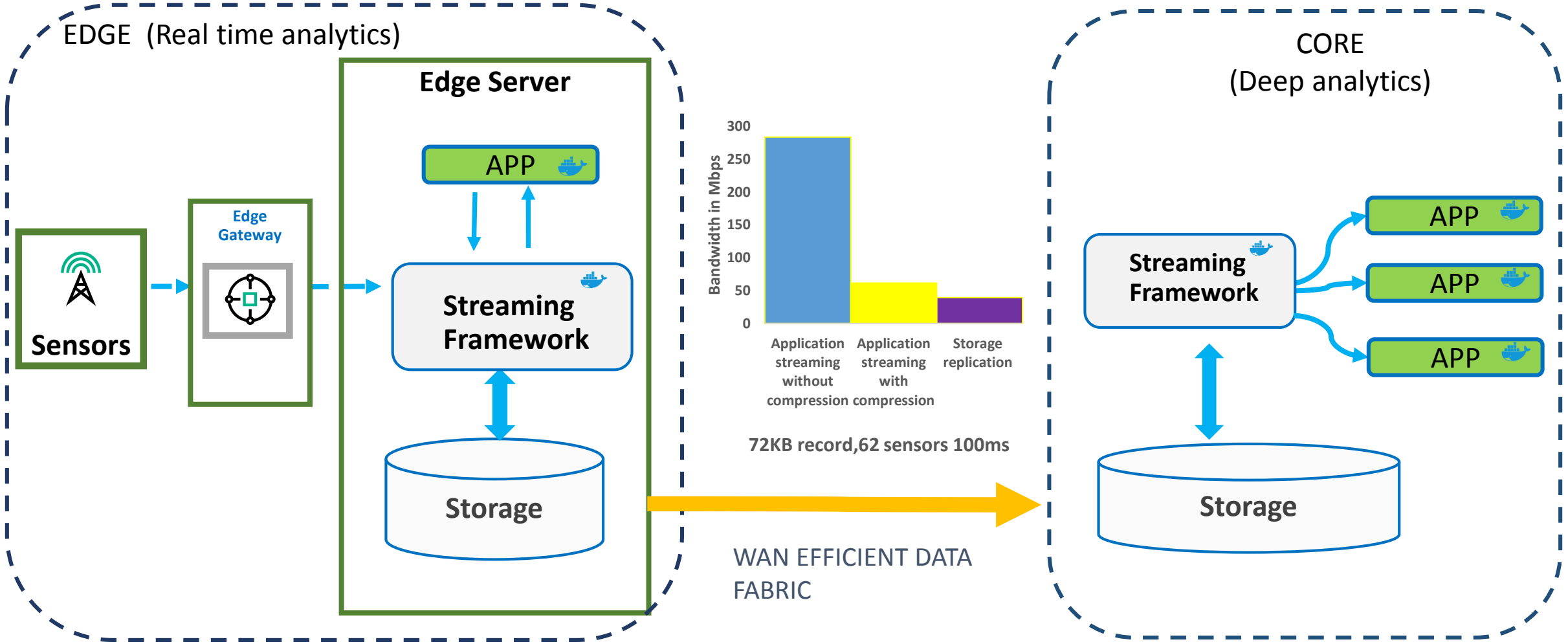
Storage Replication for Streaming



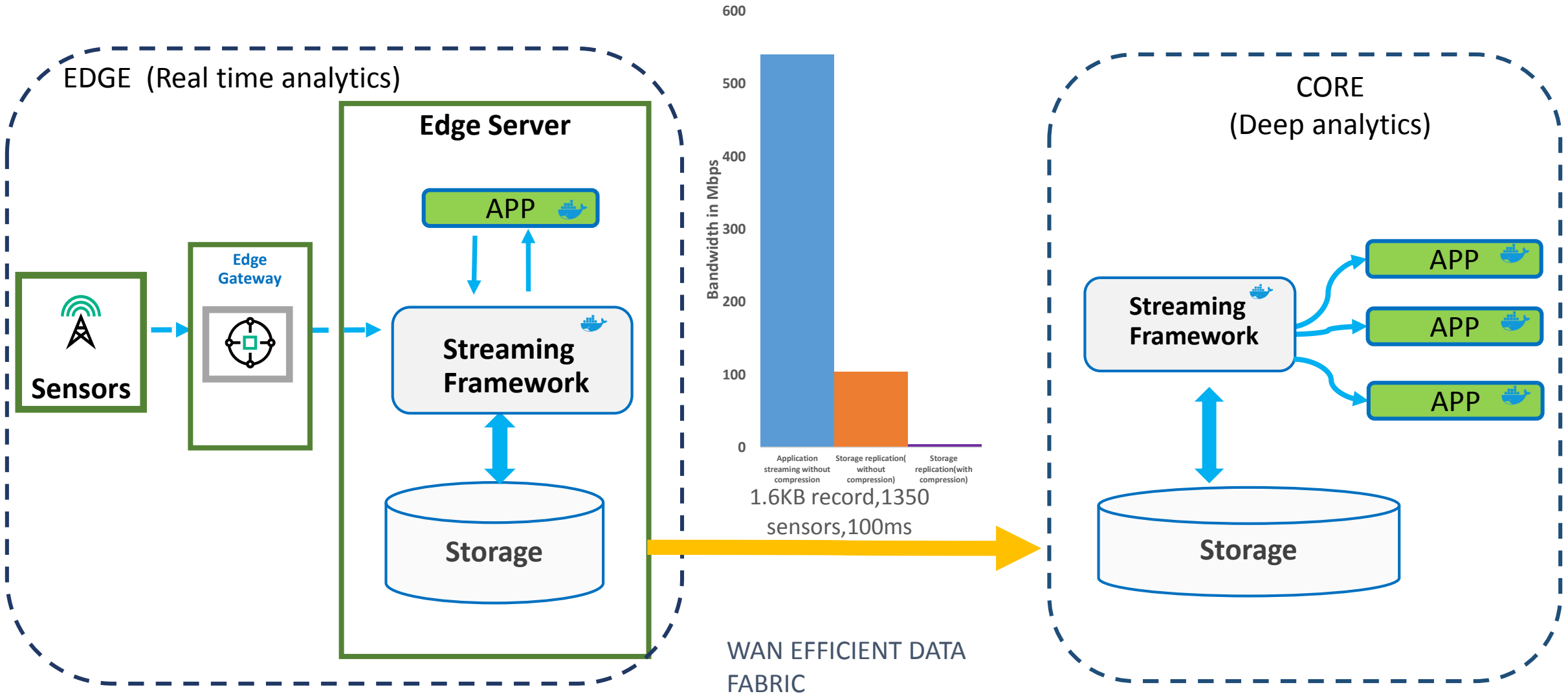
Storage Replication for Streaming



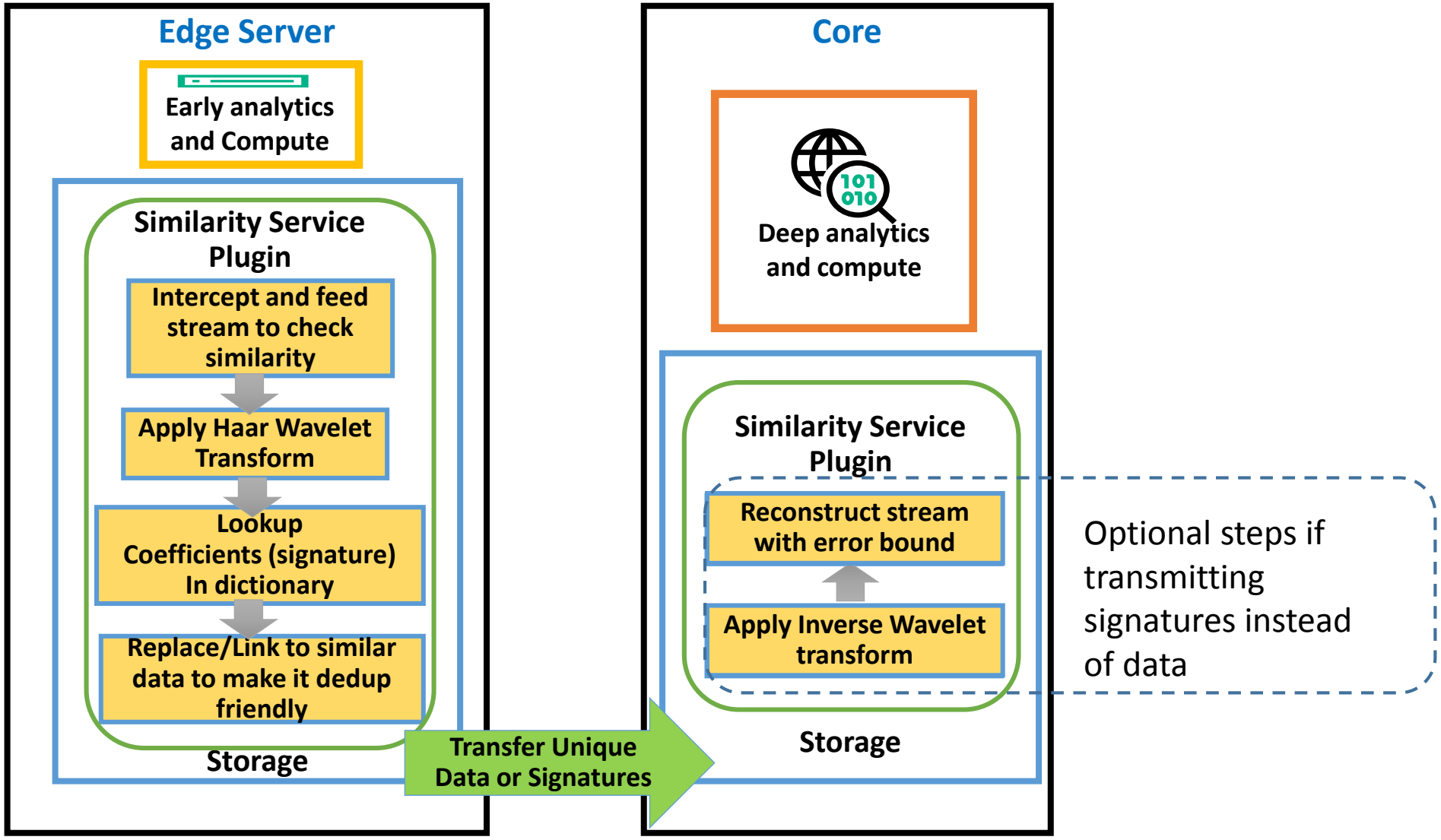
Storage Replication for Streaming



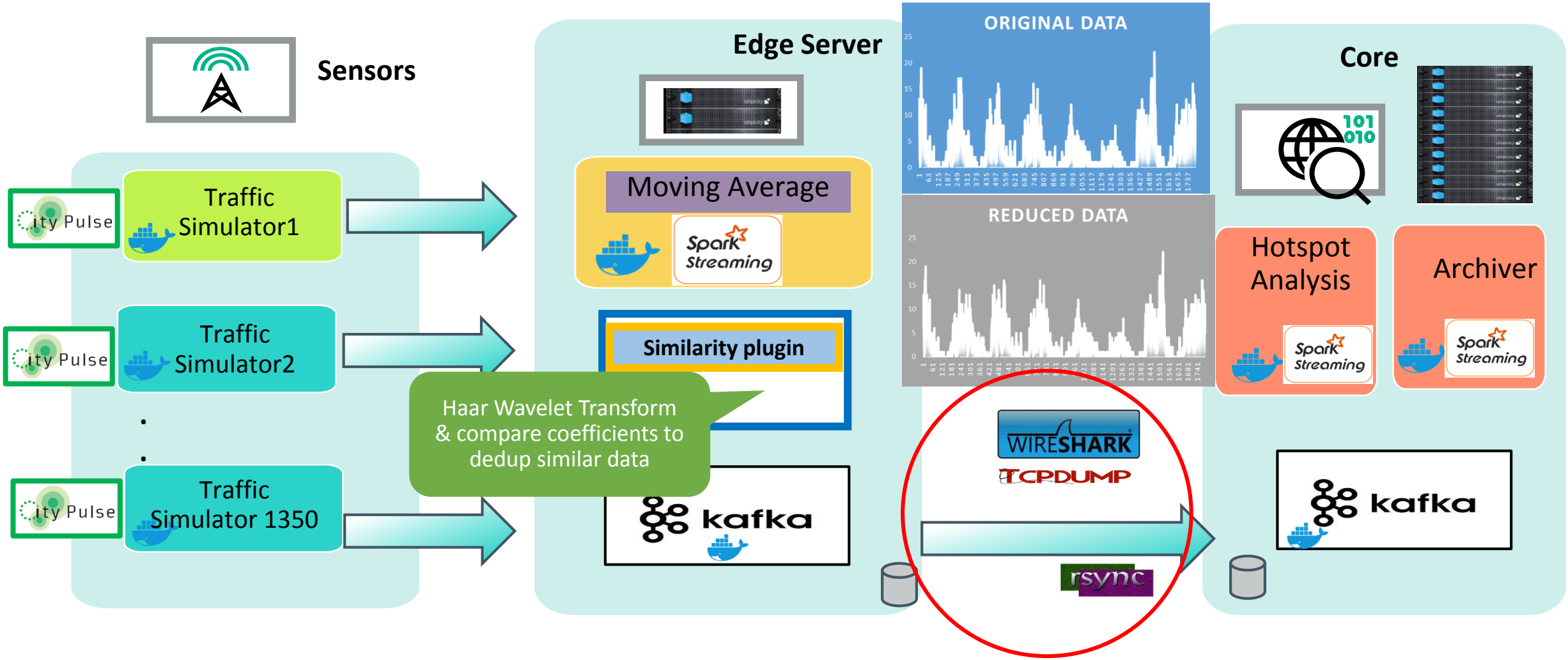
Storage Replication for Streaming



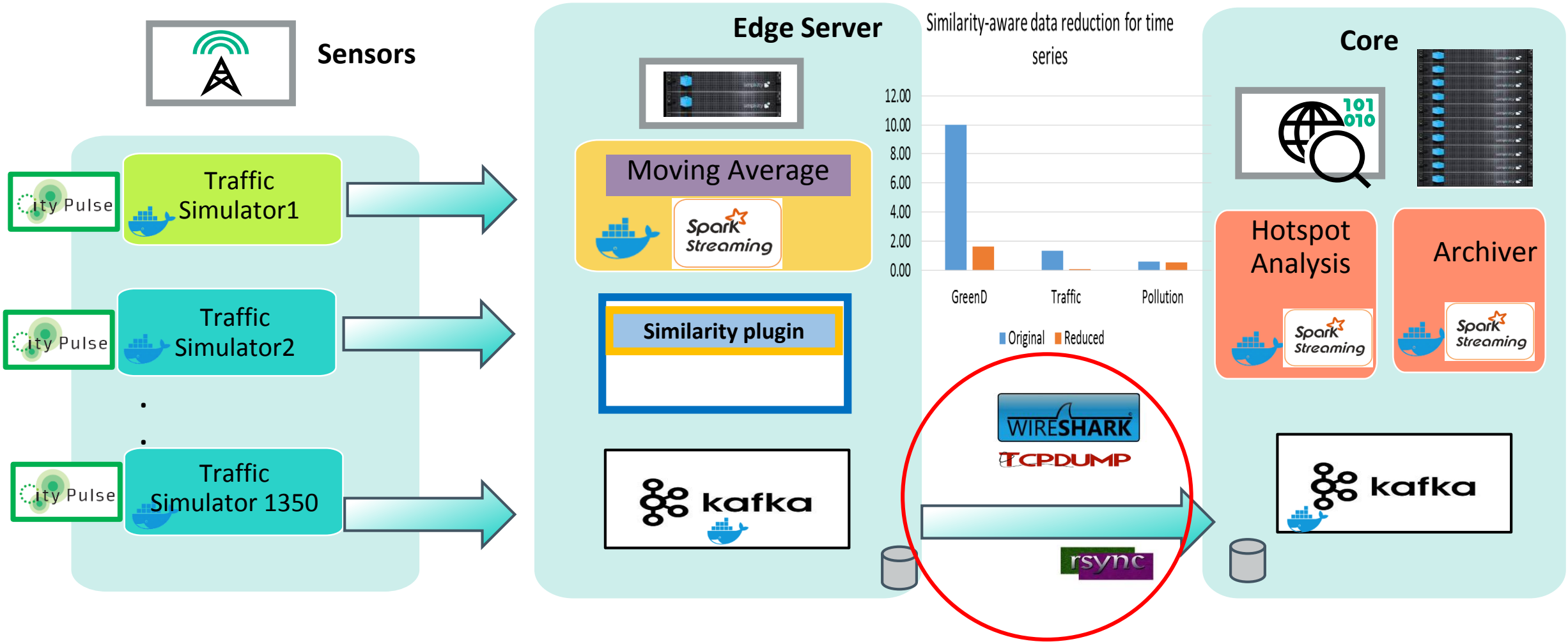
Similarity Awareness: Time Series



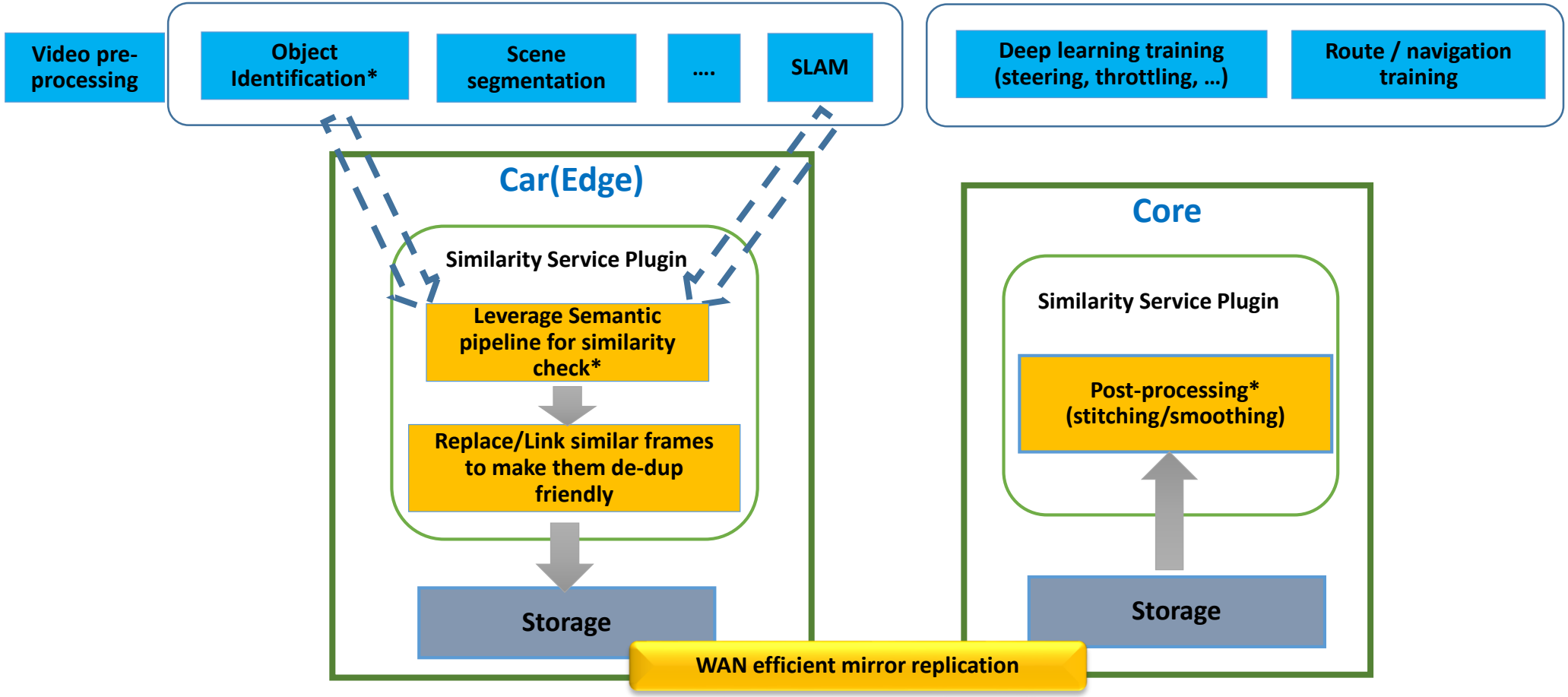
Similarity Awareness: Time Series



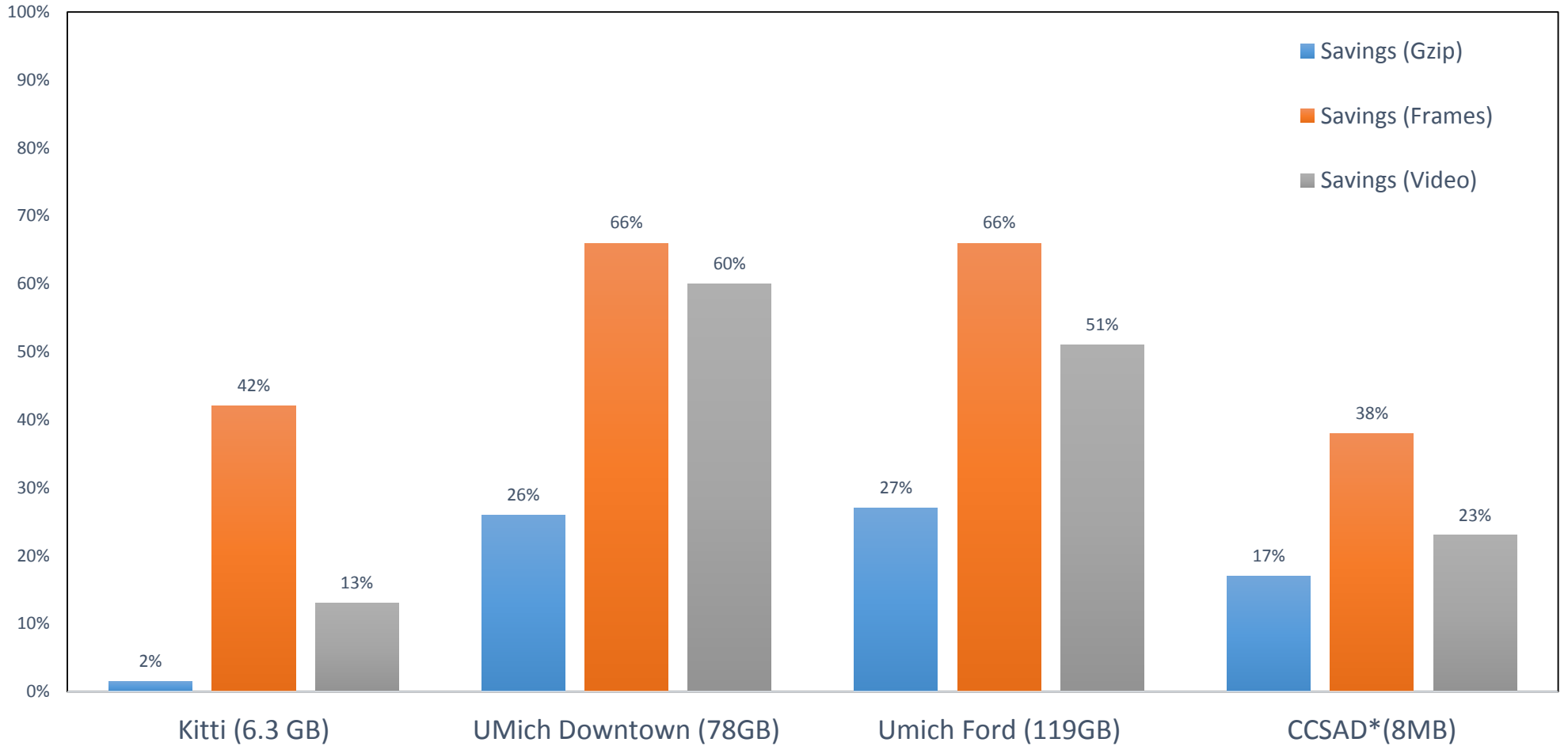
Similarity Awareness: Time Series



Similarity Awareness: Autonomous Car Video



Data Reduction for Autonomous Cars



Discussion and Research Issues

- Co-design framework and infrastructure layers for versatility
  - Apply storage replication to facilitate high-value data transmission from edge to core
  - In conjunction with other data reduction techniques that work on different scopes of data, therefore different timescale ranges of input values
  - Needs separable streams, Application consistent triggers, Semantic similarity plugin
  - Balance Edge-Core compute distribution to enable detection of semantic redundancies
- Managing accuracy vs data reduction
  - Adjusting similarity metrics (application guided or automatic learning)
  - Exact vs semantic views
  - Shifting exact data sensitive computation
  - Post-process correction
- Achieves efficiency at expense of latency and/or accuracy
  - Under what conditions is this sufficient? When would this category of approach fail?