

Please evaluate
my talk via the
mobile app!



Data Movement at Very Large Scale

(With Real-World Examples in Capital Markets,
Customer Service and Sensor Networks)

Aaron Lee

QCon London

Big Data Landscape

Vertical Apps



Ad/Media Apps



Business Intelligence



Analytics and Visualization



Log Data Apps



Data As A Service



Analytics Infrastructure



Operational Infrastructure



Infrastructure As A Service



Structured Databases



Technologies



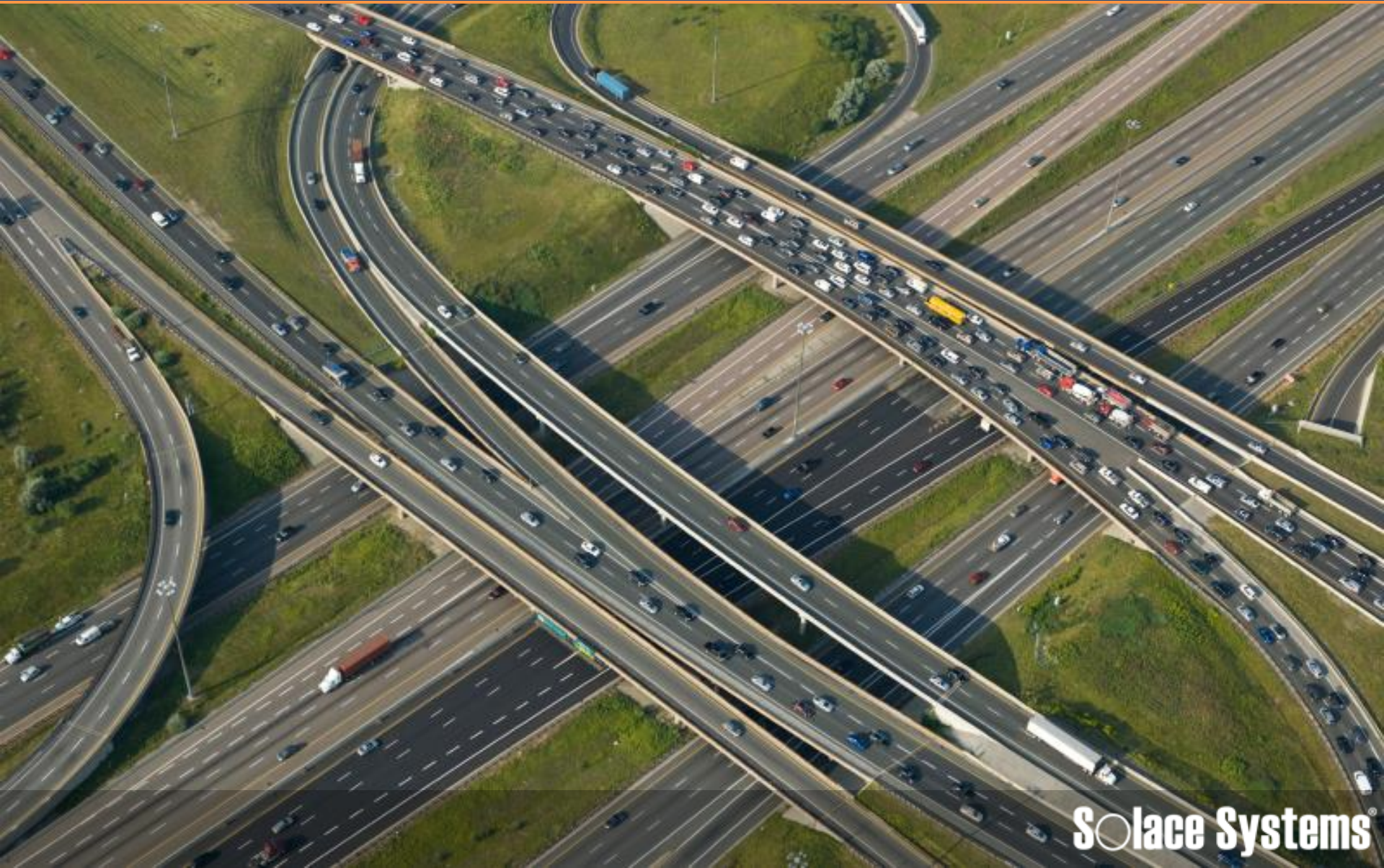
Data Movement for the Past 30 Years



Remember When...

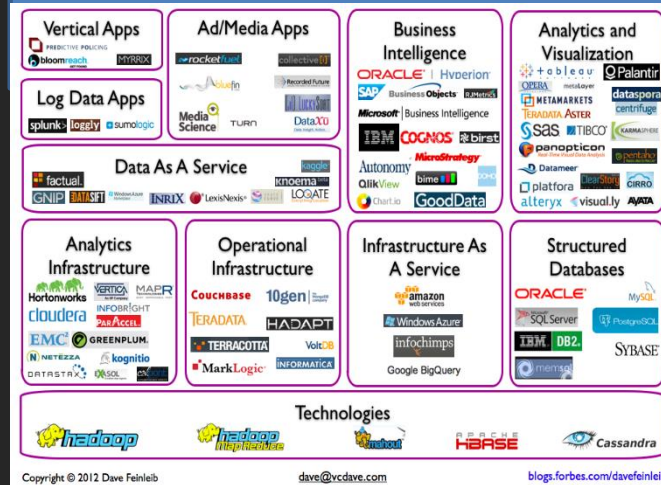


Modern Big Data Movement

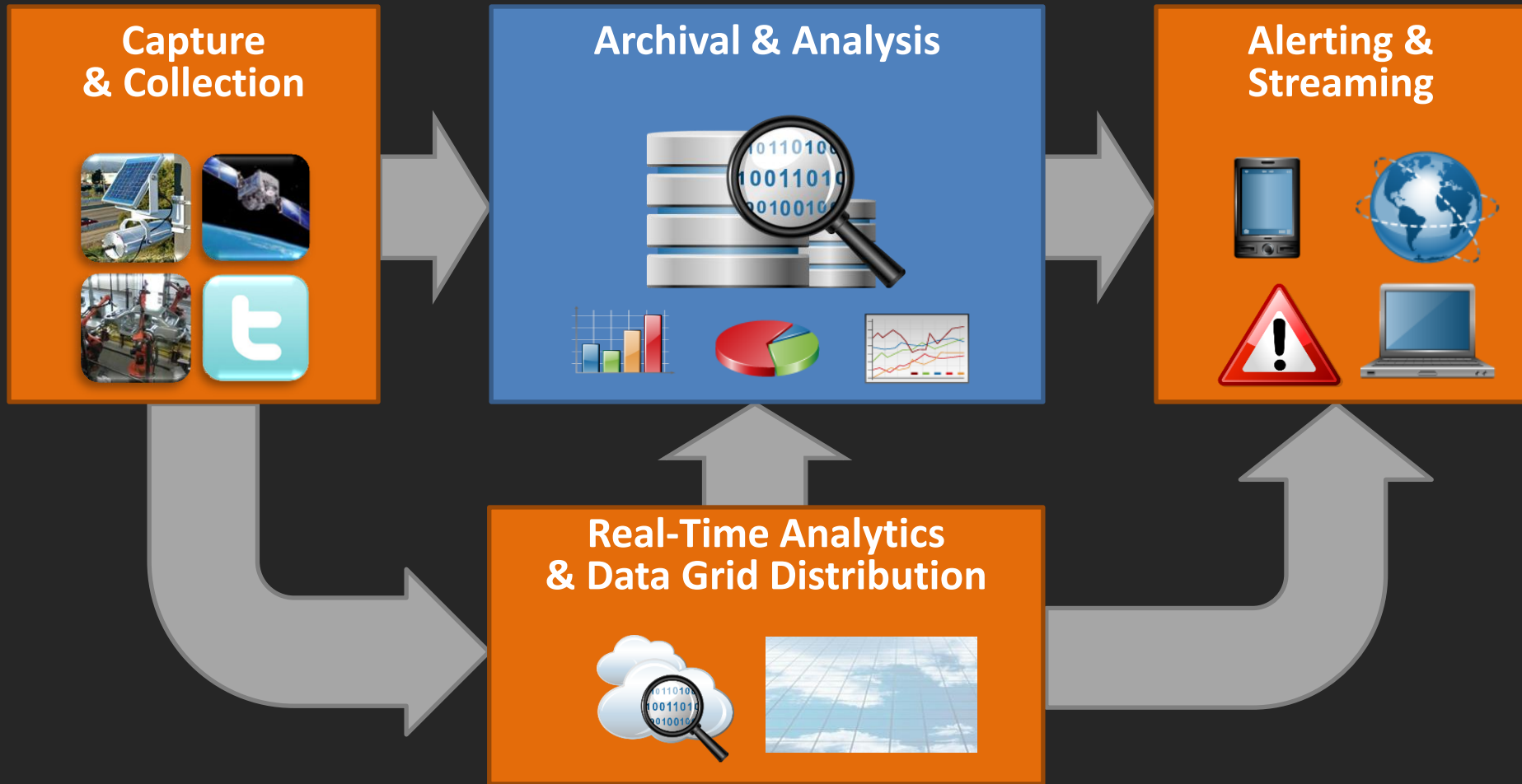


Big Data Landscape

Archival & Analysis



Putting Big Data in Motion



Capture & Collection



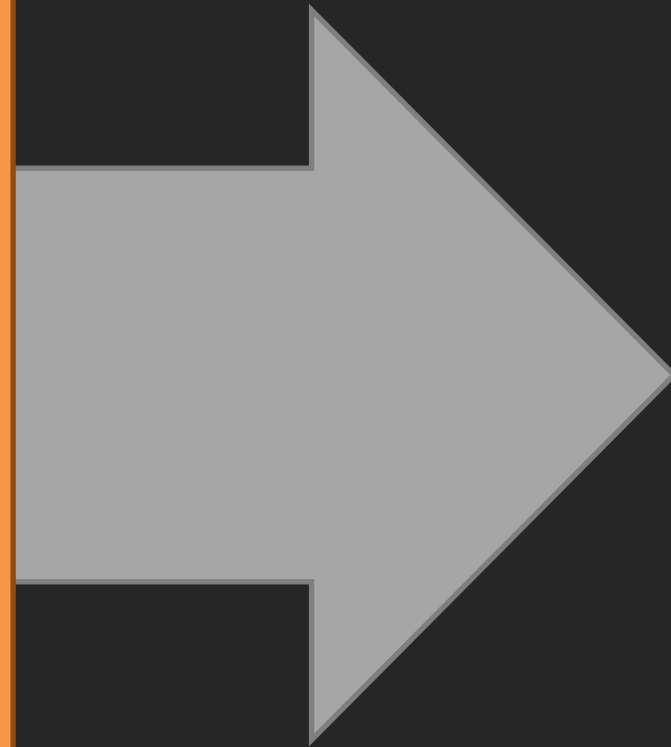
Capture & Collection Challenges



Managing *Millions*
of Data Sources
(M2M, Mobile, Social)

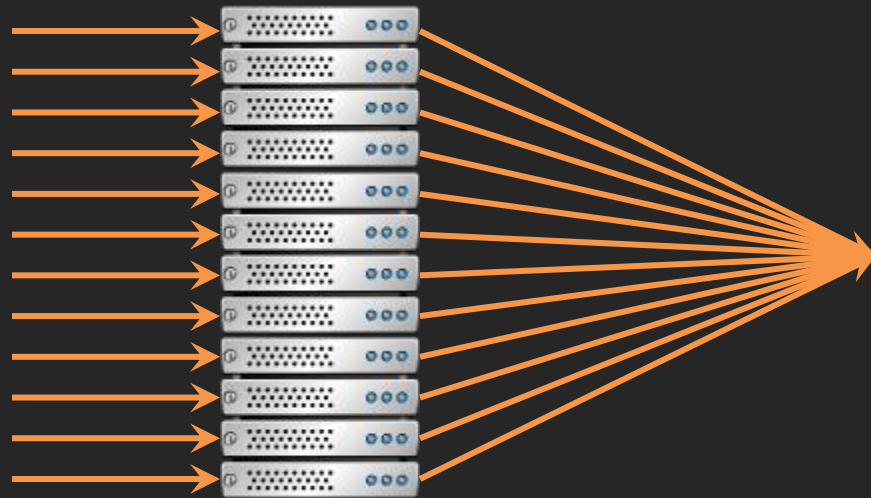
Buffering Spikes
and Overflow

Optimizing
WAN Traffic



Managing Millions of Sources

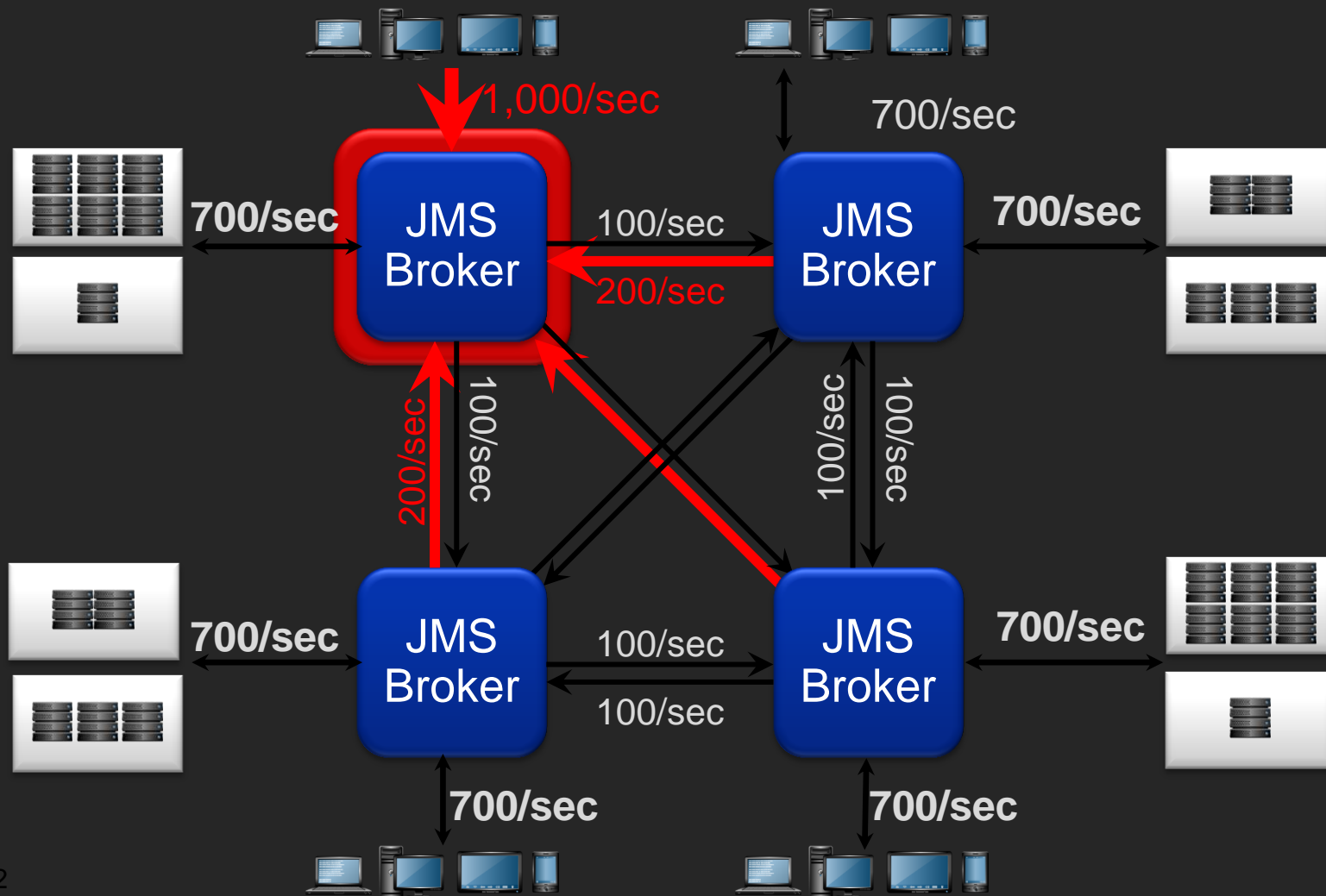
Horizontal



Vertical



“Leakage” of Horizontal Scaling



Buffering Spikes and Overflow



**Inconsistent
Aggregate
Input
Stream**

**Big Data
Shock Absorber**

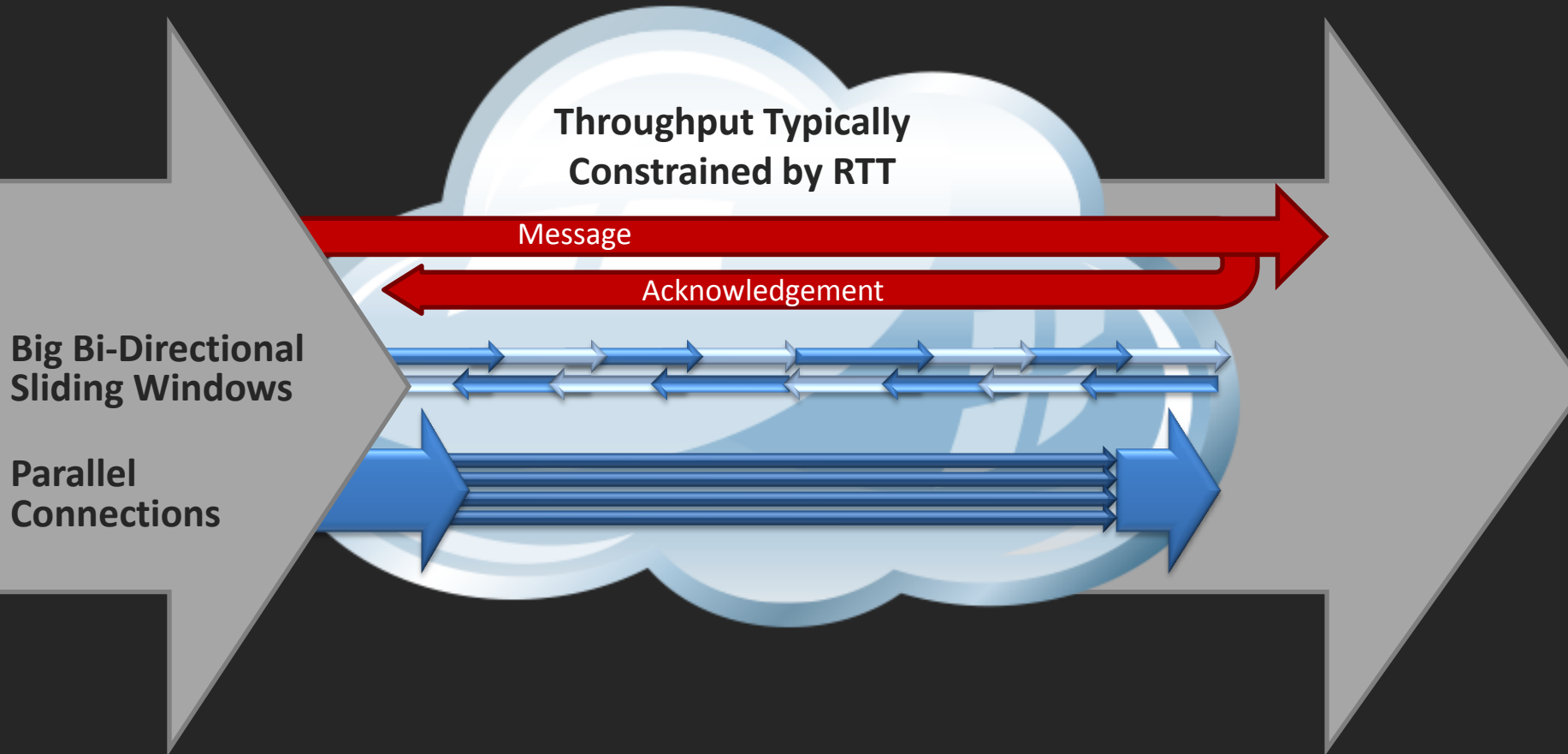
**Network
Processing
Storage**

**Capacity &
Availability Limits**

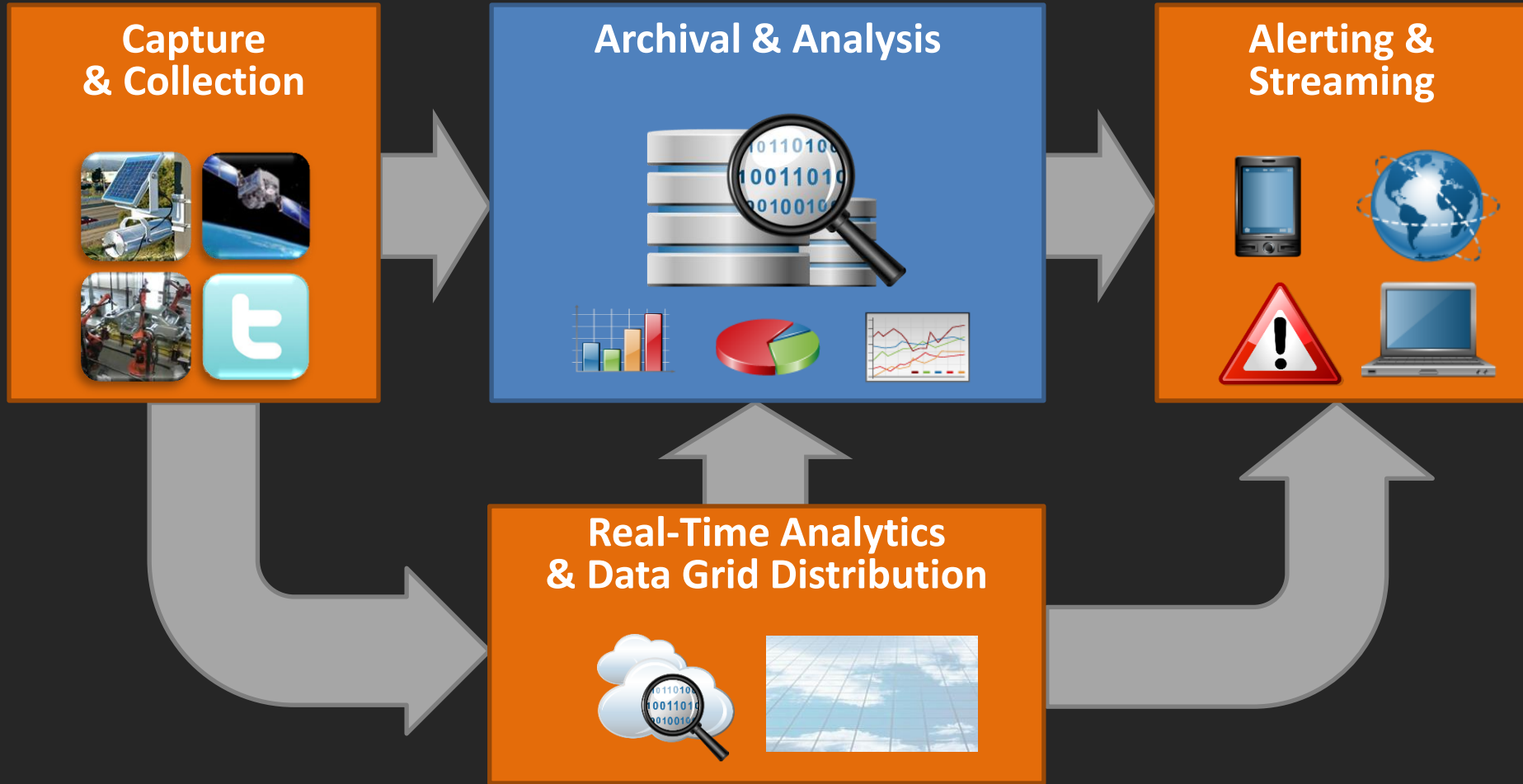
**Outages
Upgrades**

Optimizing WAN Traffic

Addressing The “Long Fat Pipe” Problem



Putting Big Data in Motion



Alerting & Streaming



Alerting & Streaming Challenges



Reaching Many Recipients

Dealing with Slow Subscribers

Conflating Data Stream



Reaching Many Recipients

Fanout to Many,
Maybe Millions,
of Subscribers

M2M

Mobile

Social



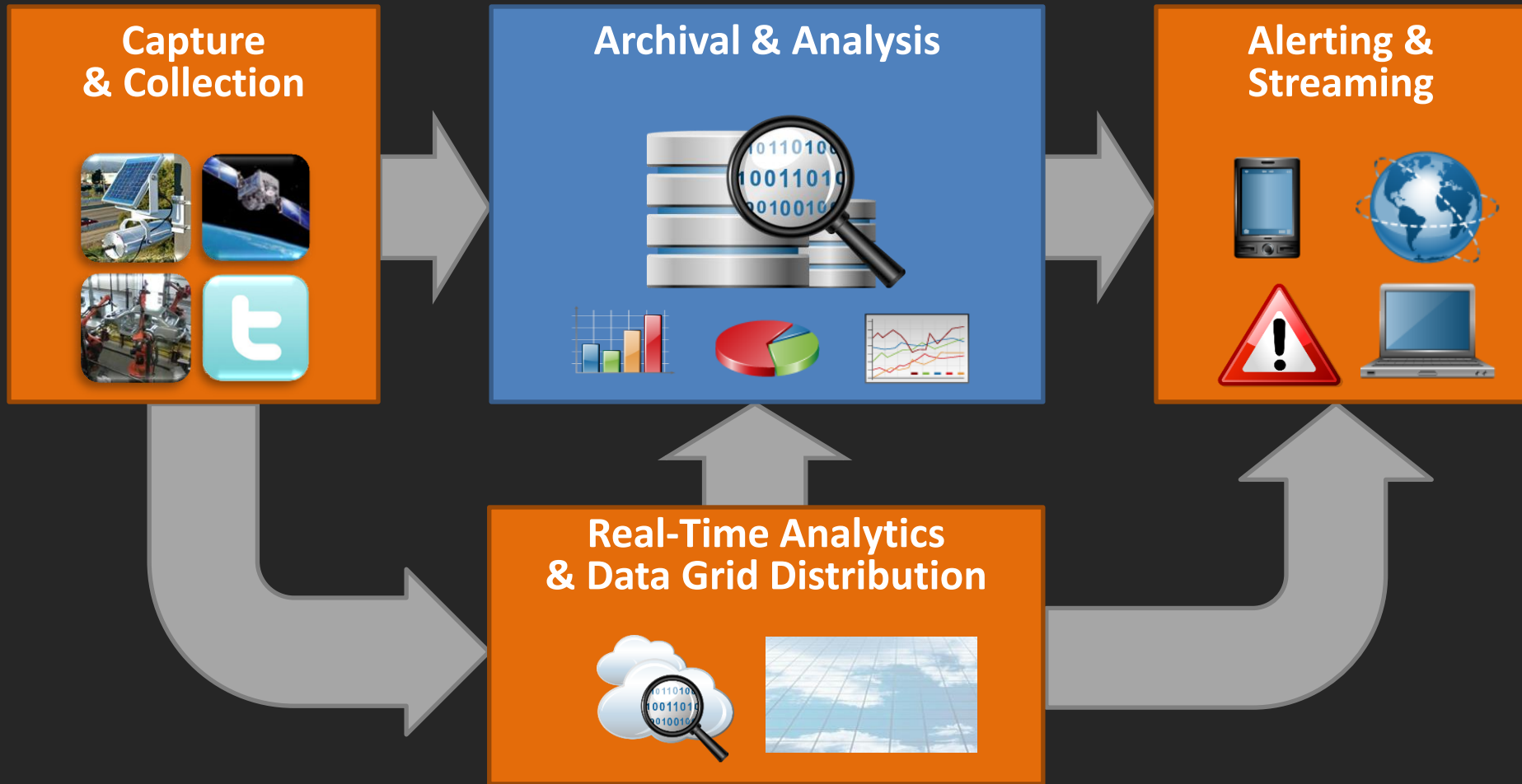
Dealing with Slow Subscribers

Accommodate
Recipients of
Varying Speed
and Availability



Guaranteed Scenarios: Persistence, Buffering
Continuous Streams: Conflation, Eliding

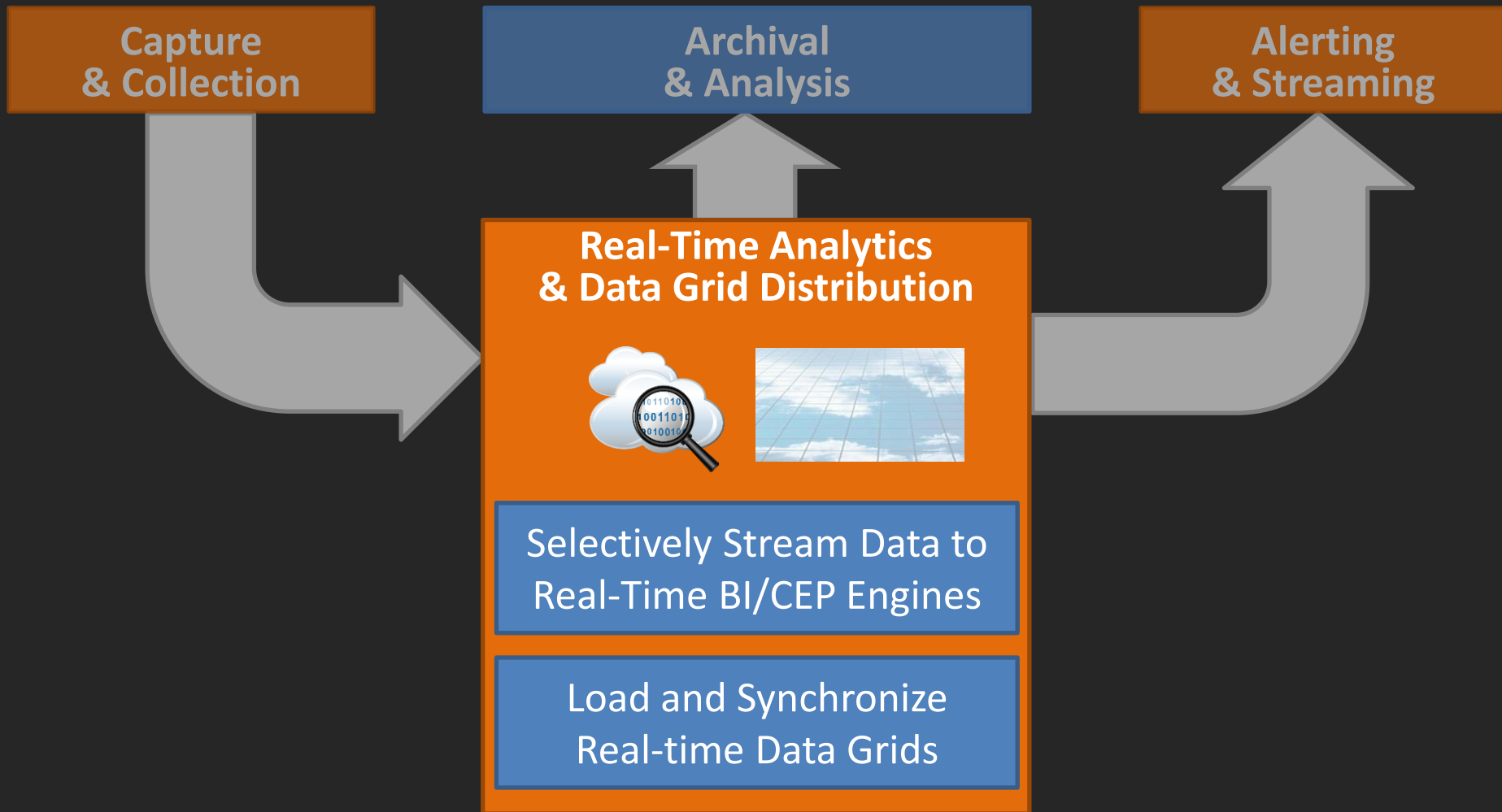
Putting Big Data in Motion



Real-Time Analytics & Distribution



Real-Time Analytics & Distribution



Stream Data to Real-time BI and CEP

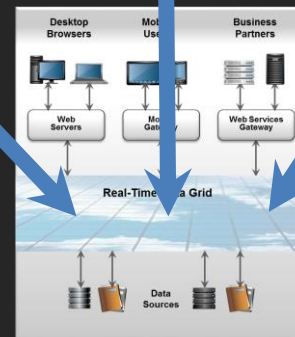
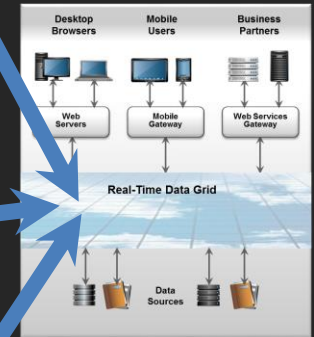
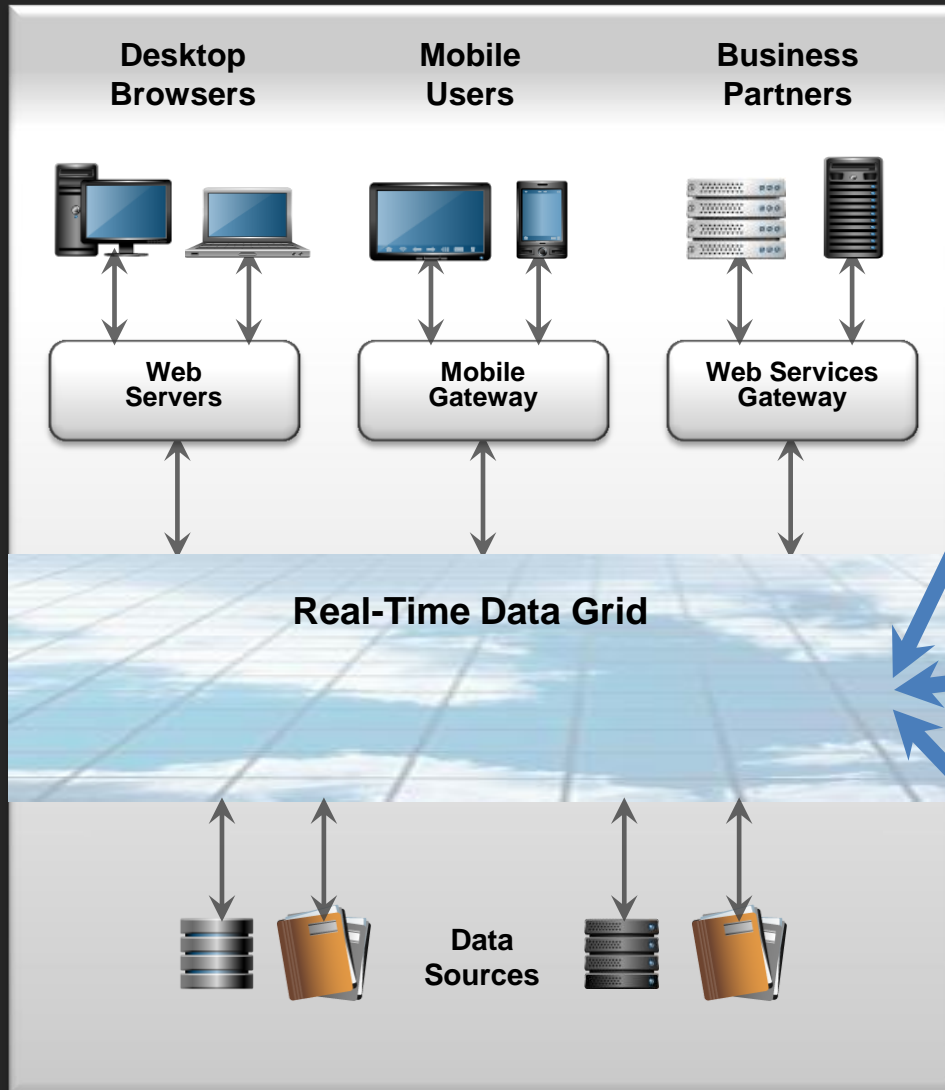
Storage and
After-the-Fact
Analysis

Subsets of data shared with real-time
analytics engines for in-flight analysis...

BI

CEP

Load & Sync Real-Time Data Grids



Use Cases

Major Railroad



- Track sensors and wayside detectors
- Real-time Analytics, BI and CEP
- “Positive Train Control”

Asset Utilization

- Tracks, Yards, Depots
- Locomotives and Cars
- Personnel

Safety and Security

Scheduling and Unscheduled Activity

Very Large Sensor Network



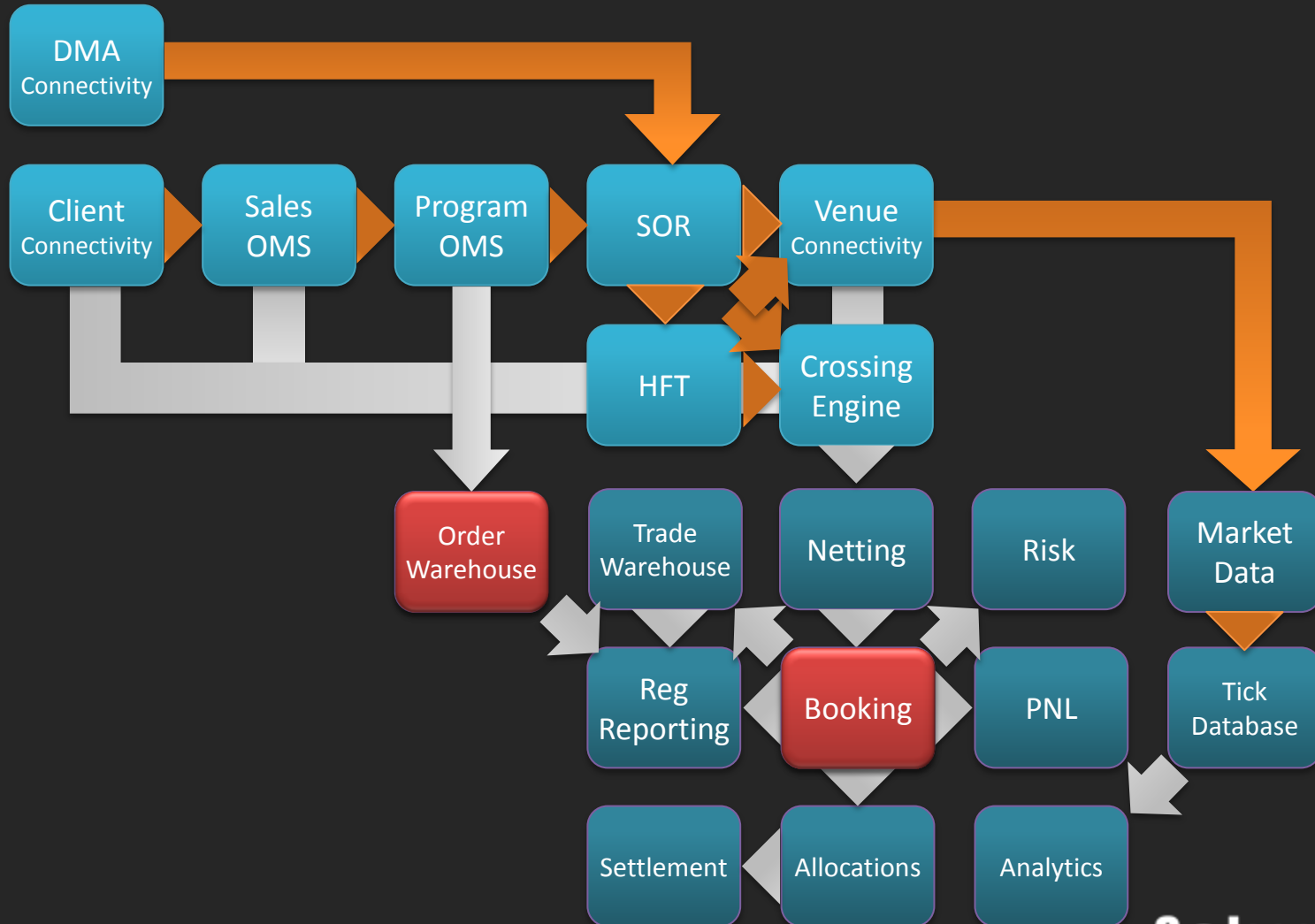
**Business
Intelligence**

**Complex Event
Processing**

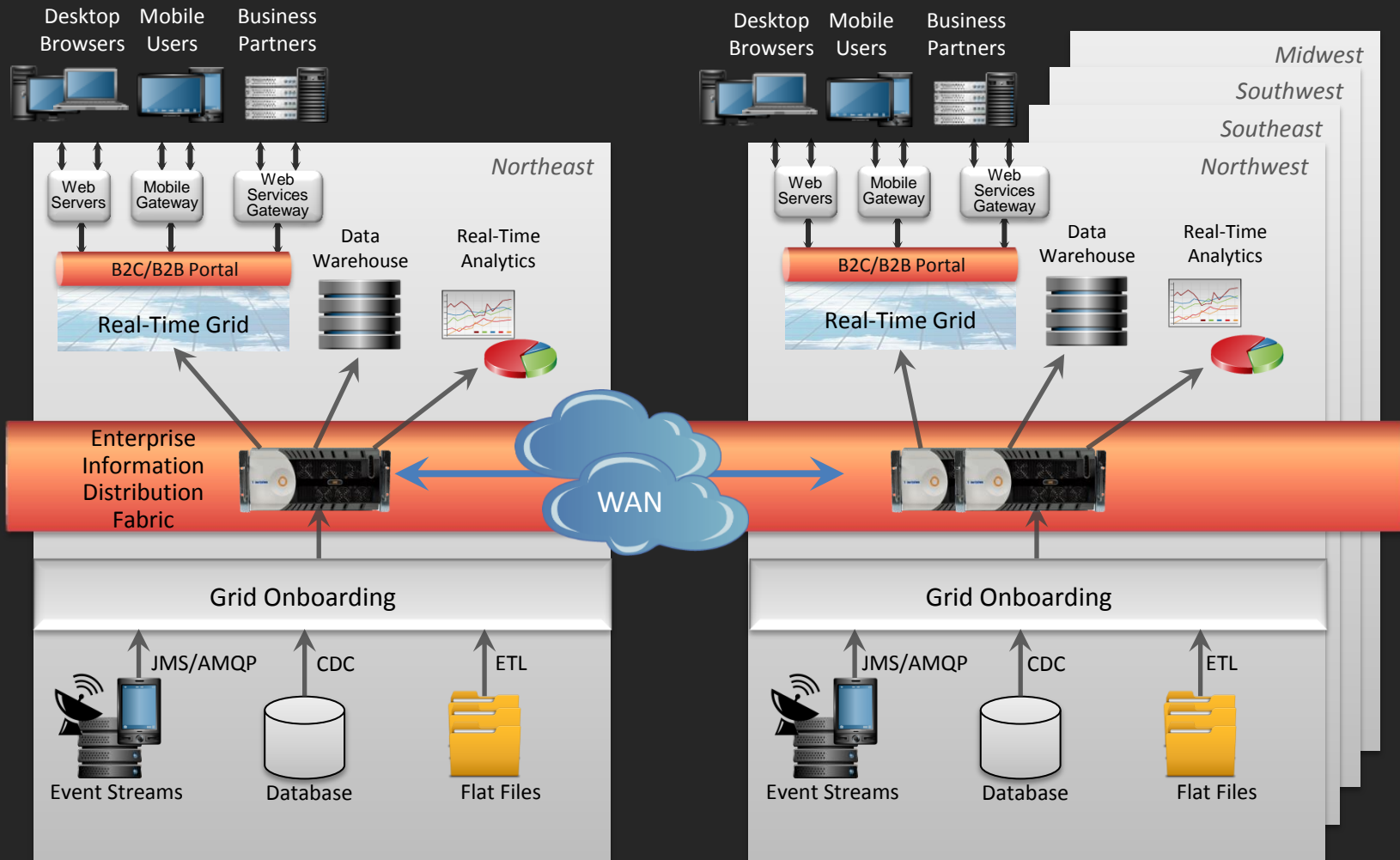
**Location
Services**

Archival

Major Investment Bank



Major Mobile Carrier



Questions?



Please evaluate
my talk via the
mobile app!

