

Design Patterns for Large Scale Data Movement

Aaron Lee
aaron.lee@solacesystems.com

Data Movement Patterns

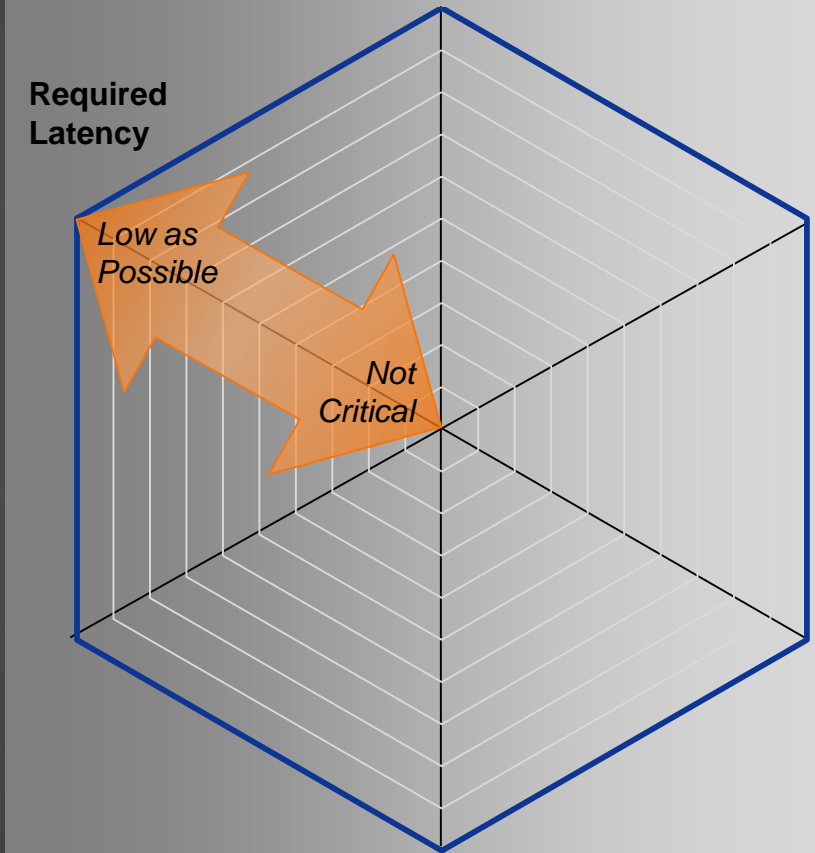
- **The right solution depends on the problem you're solving**

- Real-time or intermittent?
- Update rates?
- Any weird networks?
- Fan-in or fan-out?
- Acceptable latency?
- Payload size?
- Humans or machines?
- Guarantee required?

Latency Required



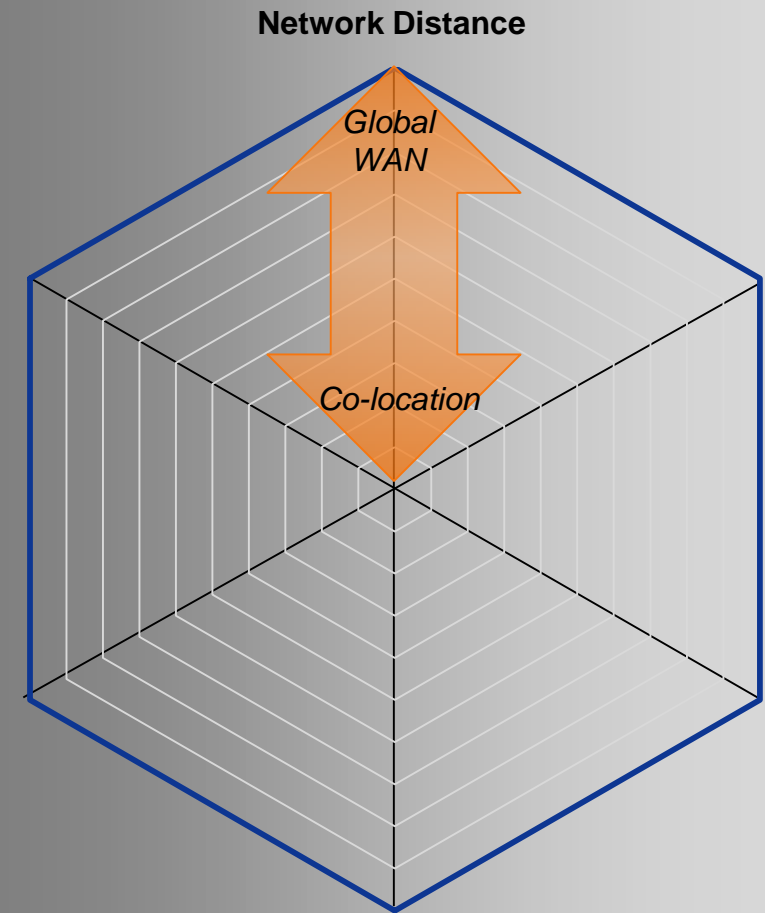
- Some not sensitive at all
 - Batch updates
- Seconds often good enough
 - Database sync
 - User interfaces
- Others measure in milli- or micro-seconds
 - Algo trading
 - Industrial controls



Network Distance



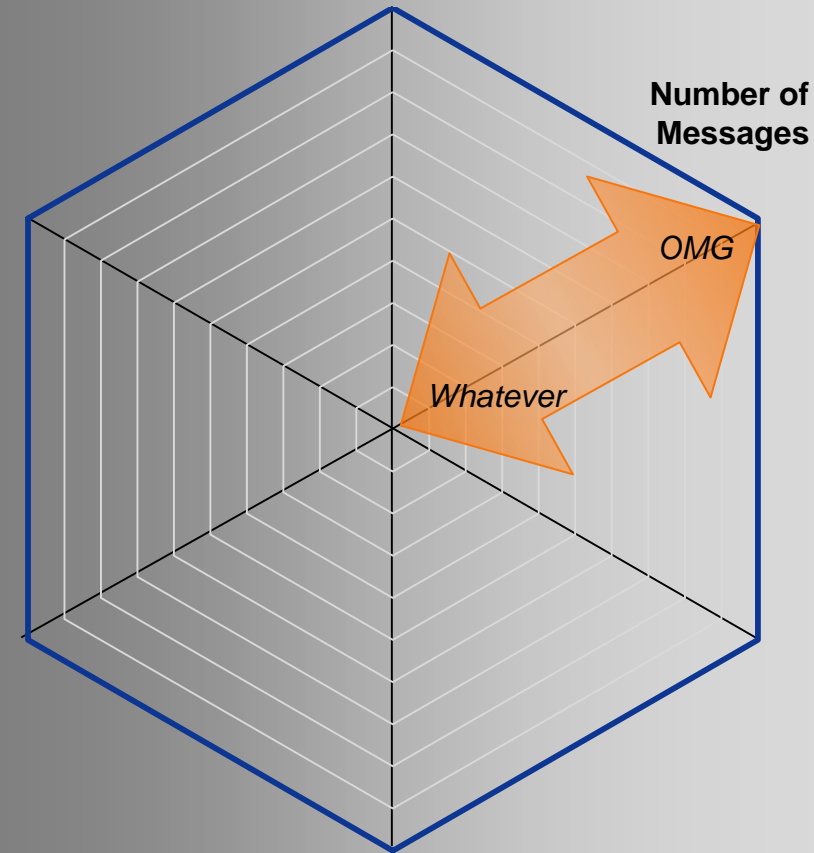
- Co-location for max speed
 - Minimize speed of light
- LAN for many apps
 - 10GigE networks
- Long distance WAN
 - Expensive, limited pipes
 - Creates mismatches with other networks



Number of Messages



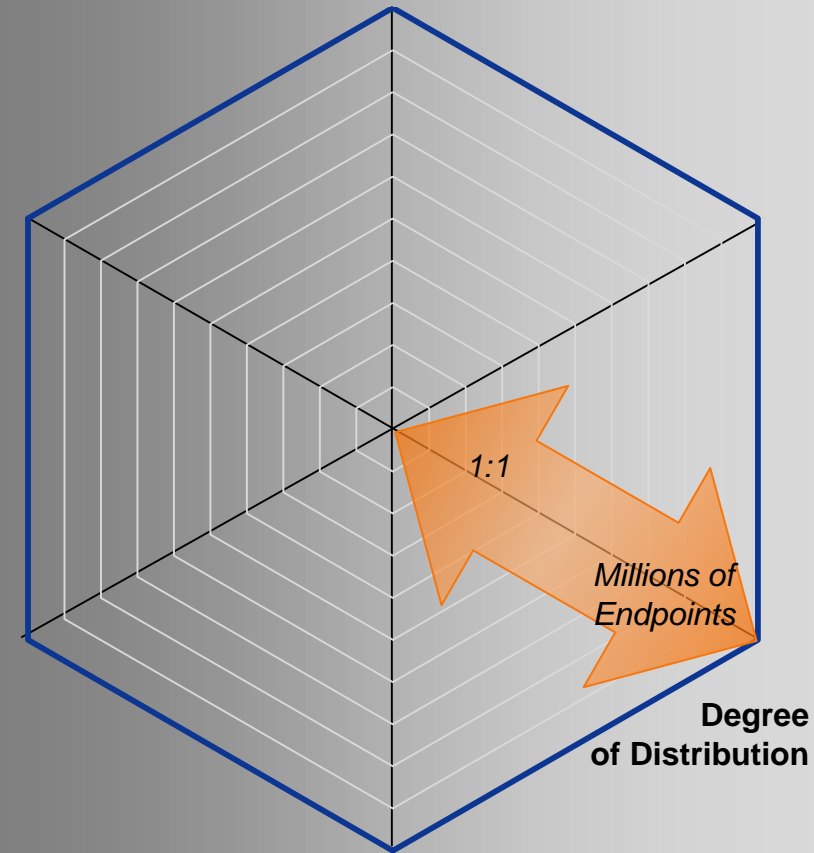
- Few
 - Batch updates
 - Simple applications
- Moderate
 - Risk management
 - Order routing
- Insane
 - Market data
 - Click stream analysis



Degree of Distribution



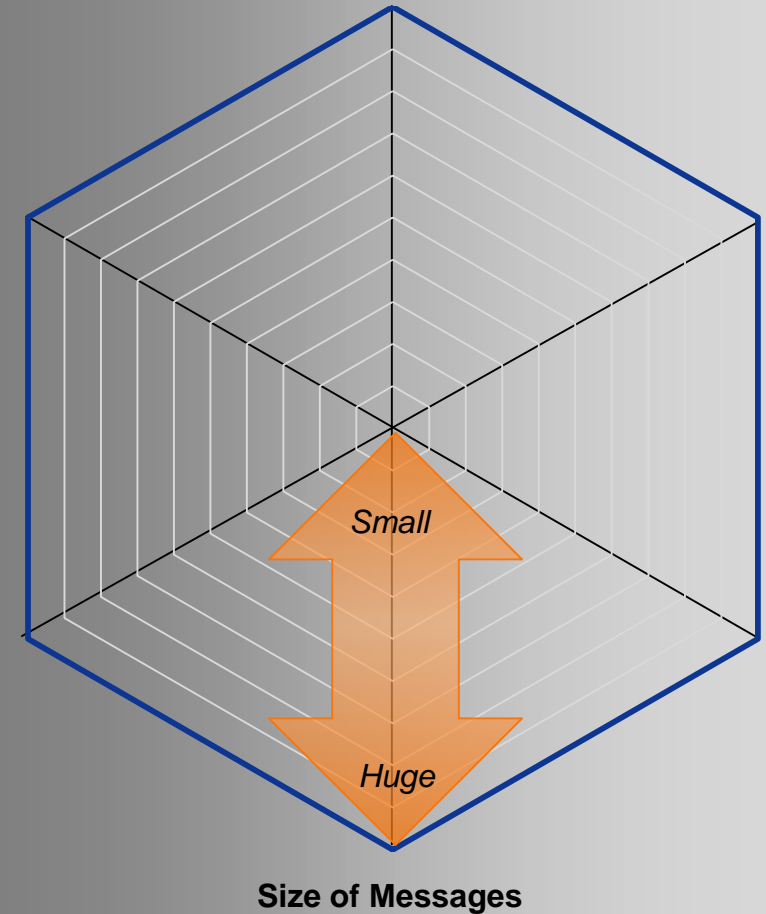
- Point-to-point
- Fan-out (many subs)
- Fan-in (many pubs)
- Mesh
 - Syncing data between many endpoints



Message Size



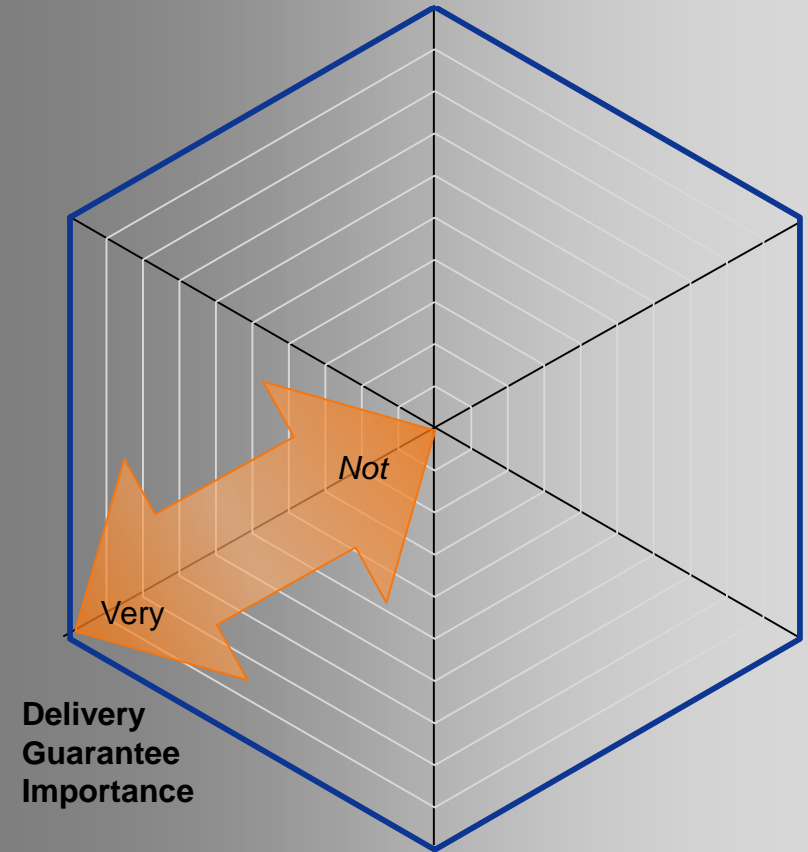
- Small
 - Status updates, activity logging events
- Medium
 - Orders, product BOMs
- Large
 - Batch updates, media files, product catalogs
- Very different stresses on system based on message size and frequency.



Importance of Delivery Guarantee



- “Best effort” fine for some scenarios
- Others require “once and only once”
- Sequence matters for some
- Some demand failsafe even in DR scenarios



Other Considerations



○ Message

- Format
- Protocol
- Structured/Unstructured

○ Network

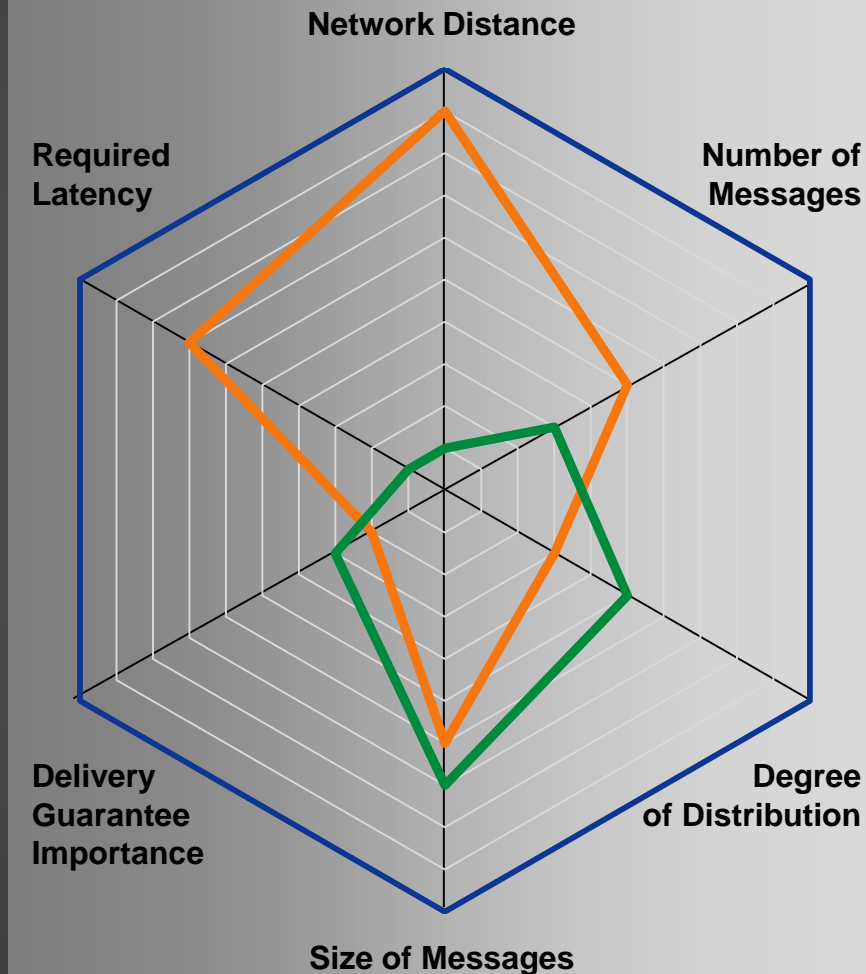
- Availability
- RTT
- Bandwidth cost

○ Robustness

- Archival
- Caching
- Acceptable MTBF
- HA switchover times
- DR requirements

Combination of Factors Yields Design Patterns

- Some attributes tend to correlate
 - # of messages and degree of distribution
- Others usually contradict
 - Network distance and latency
 - Guarantee and latency
- Tradeoffs and creative solutions



Identifying Patterns in Real-World Use Cases

Use cases unique,
but patterns emerge

Examples in this section:

Trade Order Flow

Manufacturing Data Sync

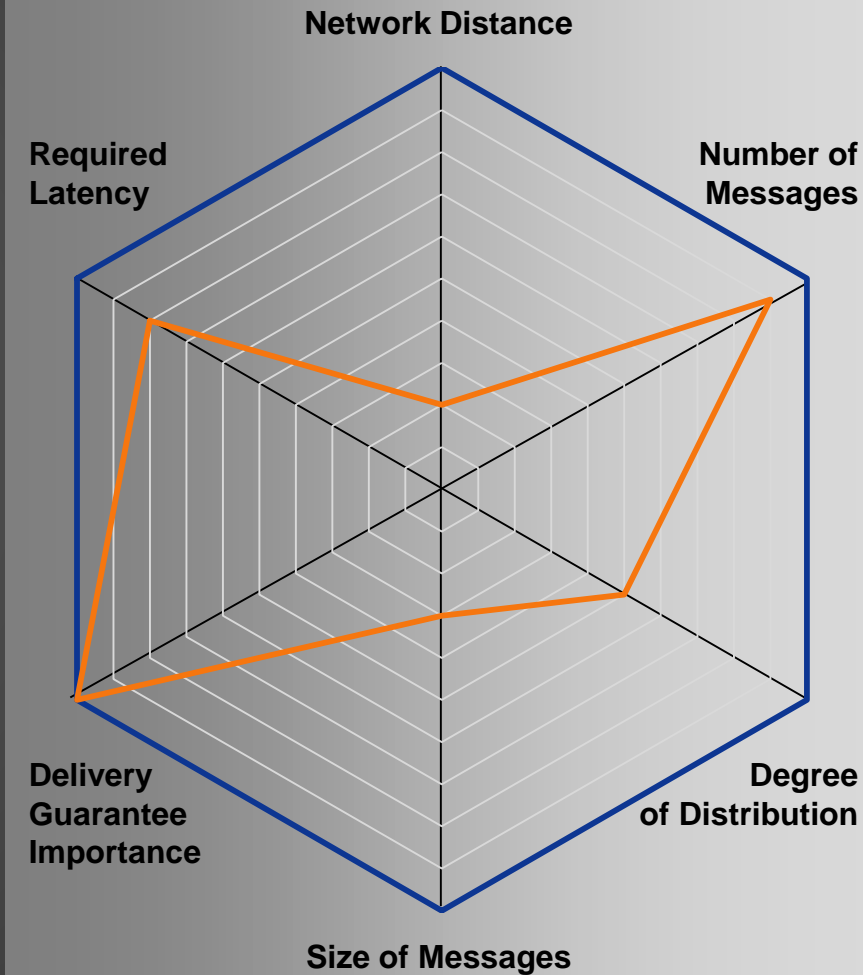
Oil and Gas Monitoring

Real Time Sports Betting

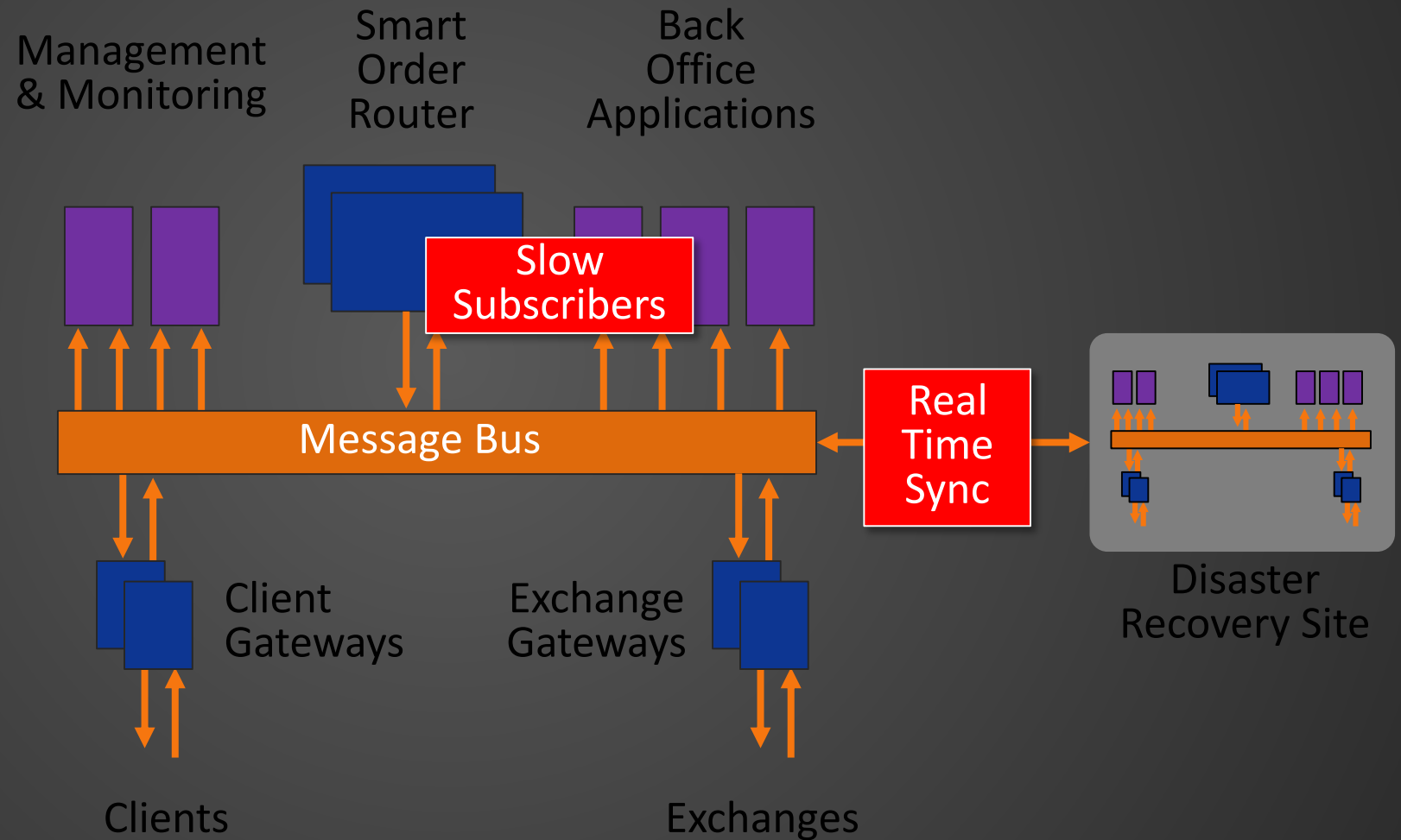
Order Flow



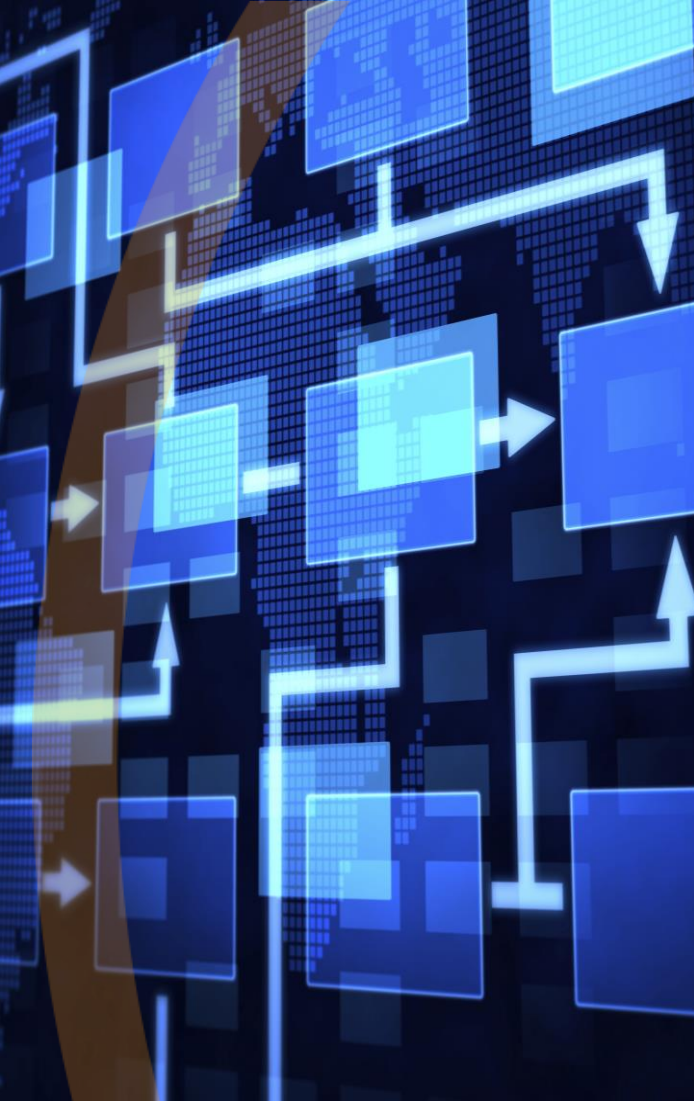
- Latency matters, but not every microsecond
- Usually localized
- Continuous, high-rate message flow
- Mid-sized messages (1-2Kb)
- Messages absolutely must be guaranteed



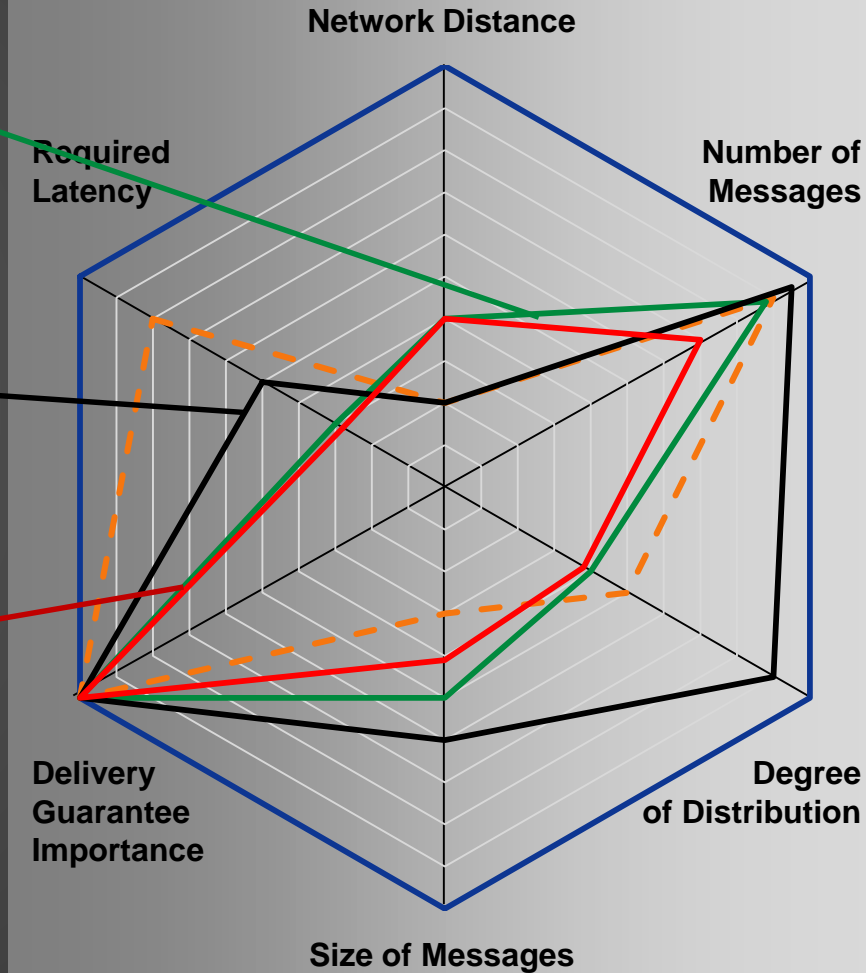
Order Flow; Architecture



Order Flow; Similar Use Cases



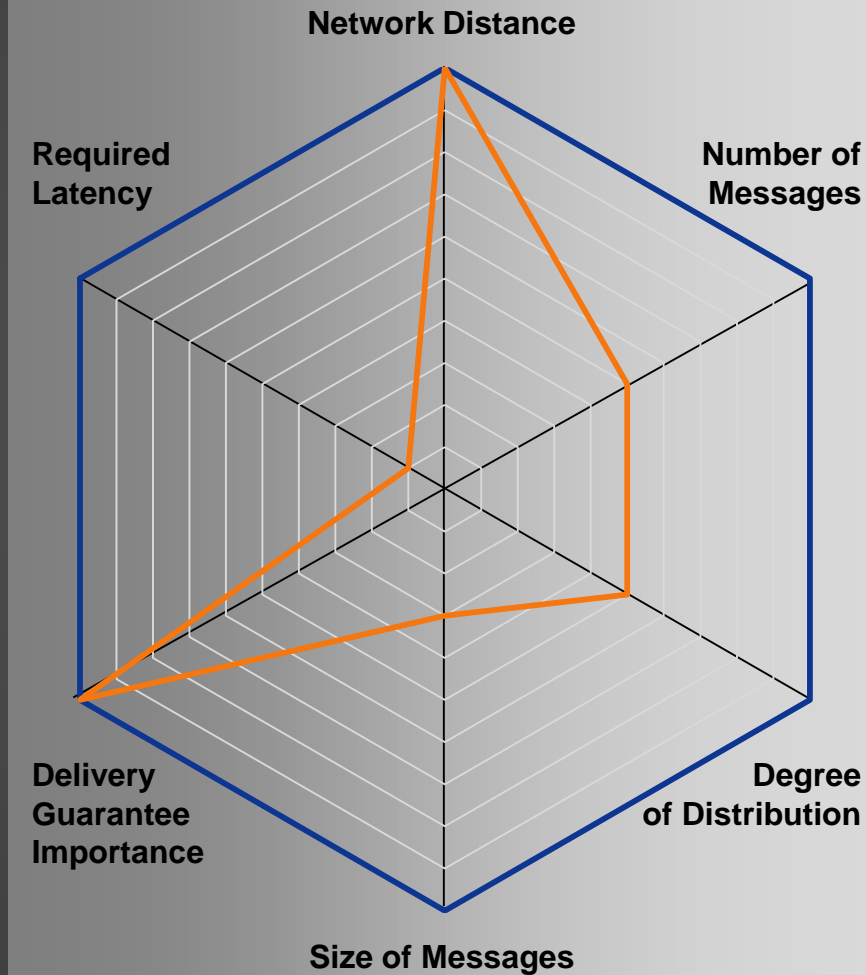
- Credit card processing
 - Long-distance WANs
 - latency in hundreds of milliseconds
- E-commerce
 - Higher volumes
 - Higher guarantee required
- Logistics scheduling
 - Less latency sensitive
 - More likely to include WANs



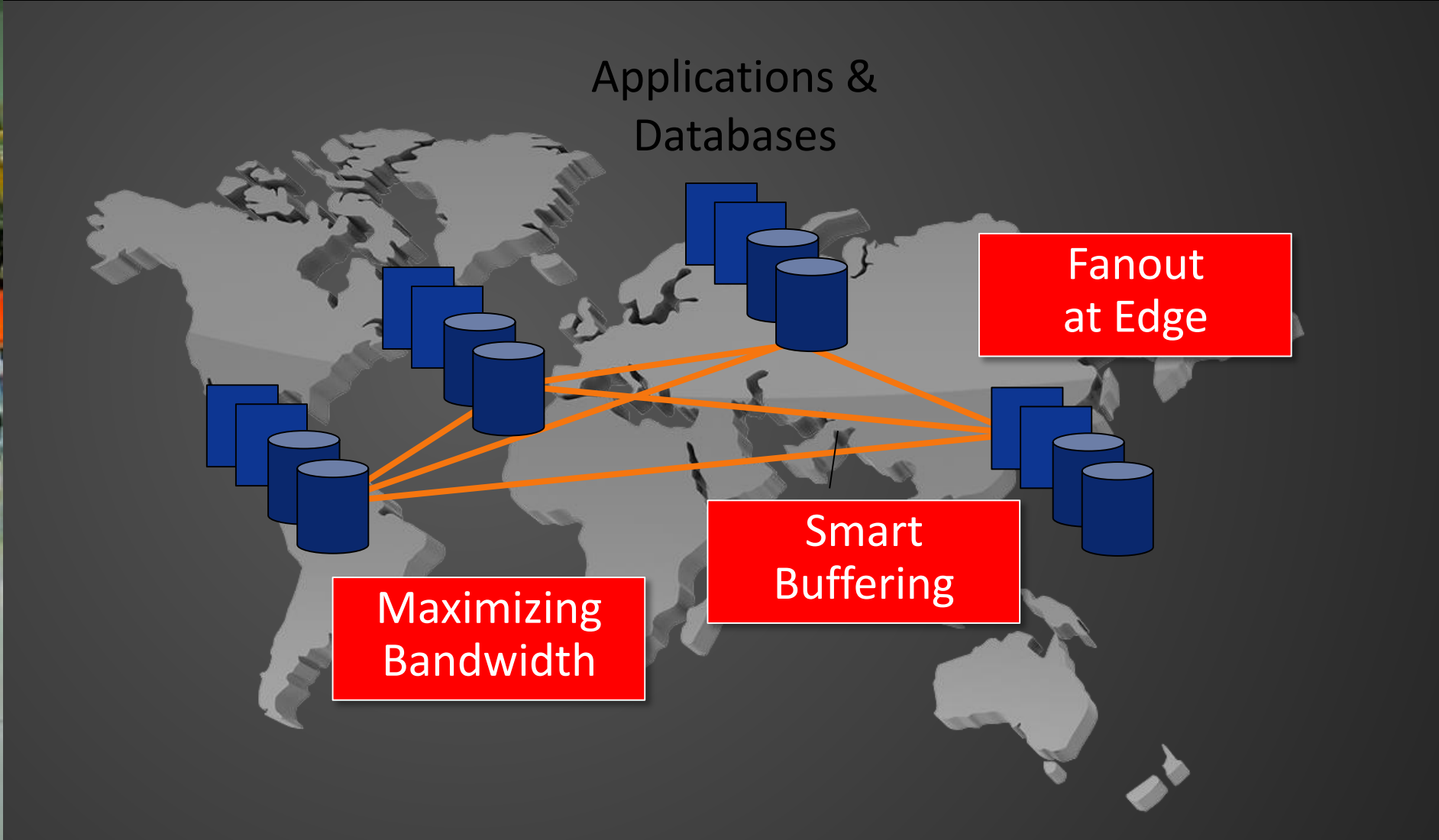
Manufacturing Data Sync



- Geographically distributed
- 100% delivery guarantee required
- Data rate is use case specific – will assume lots of medium (< 5K) messages.
- Number of endpoints use case specific, assume 10 manufacturing locations



Manufacturing Data Sync; Architecture



Manufacturing Data Sync; Similar Use Cases



- Real Time Risk Management

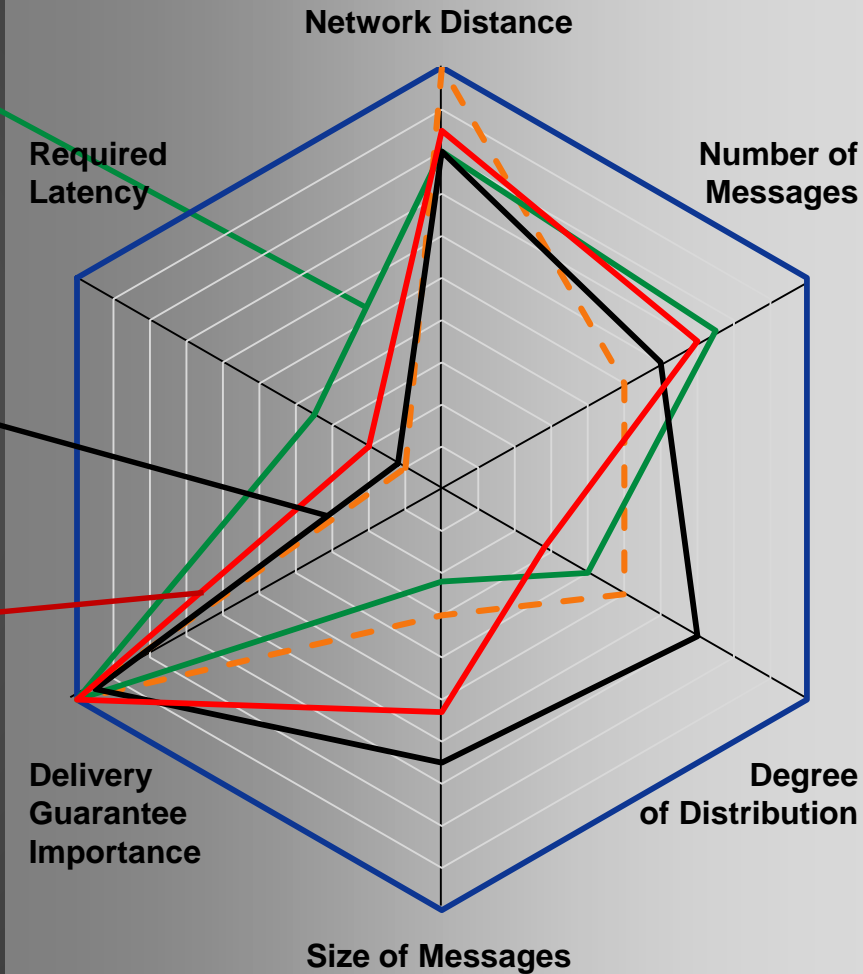
- Smaller messages
- Latency more important

- Retail Global Inventory

- Messages can be larger
- Distribution can be more

- Real Time Financials

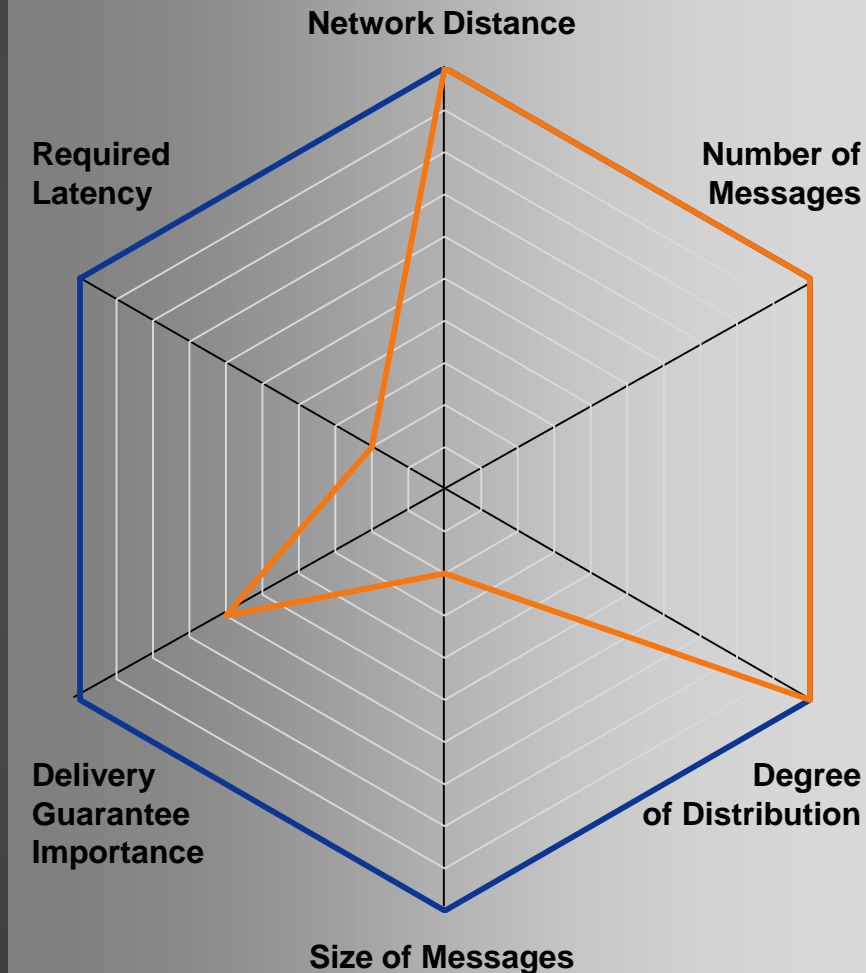
- Messages larger
- Distribution less (collecting to 1 location)



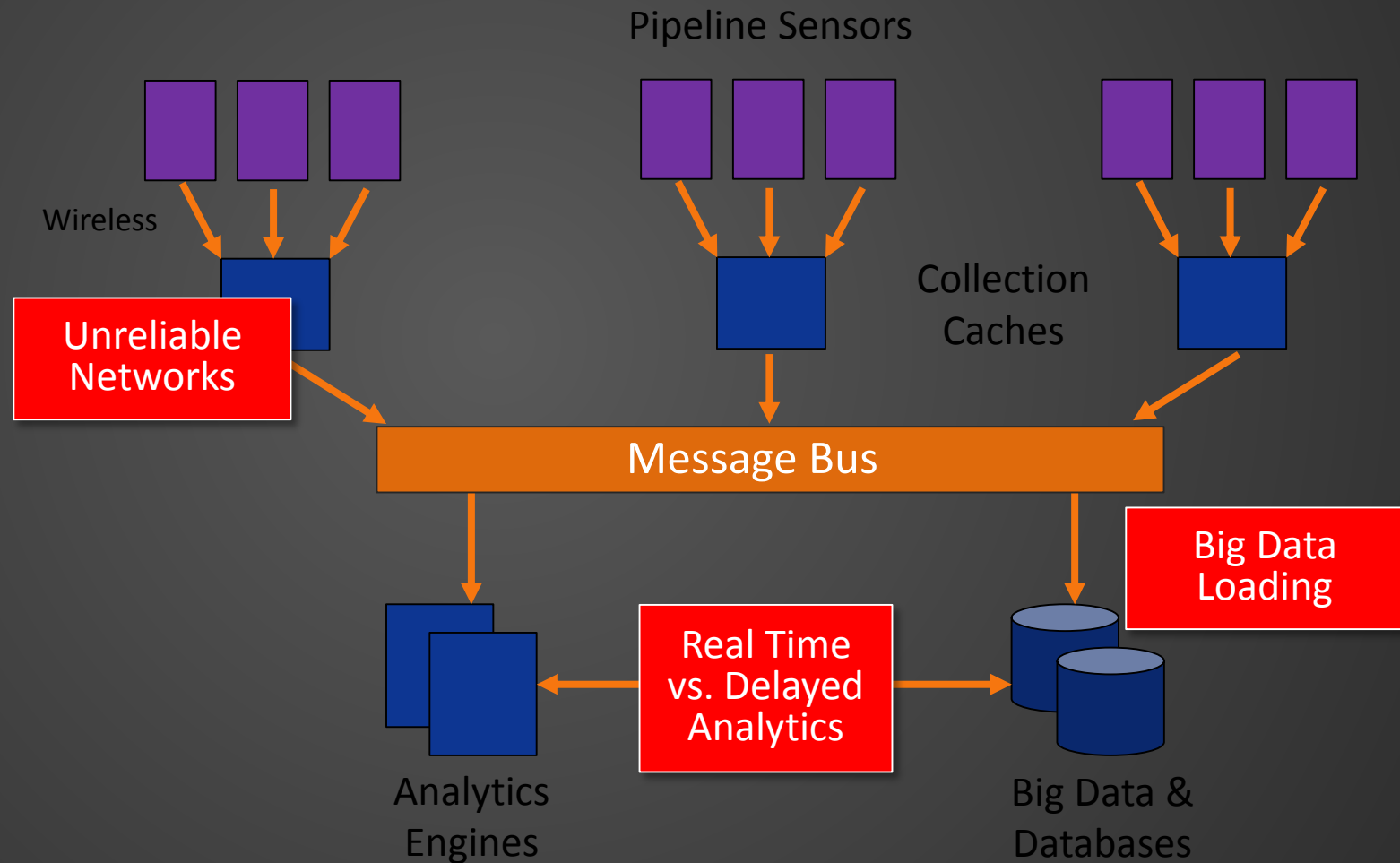
Oil & Gas Pipeline Monitoring



- Wifi, Satellite, proprietary and other unreliable networks
- Degree of distribution off the charts. In this case, fan-in.
- Messages usually pretty small, unless batch
- Latency unimportant
- Level of guarantee use case specific, assume status messages (ie. guarantee not essential)



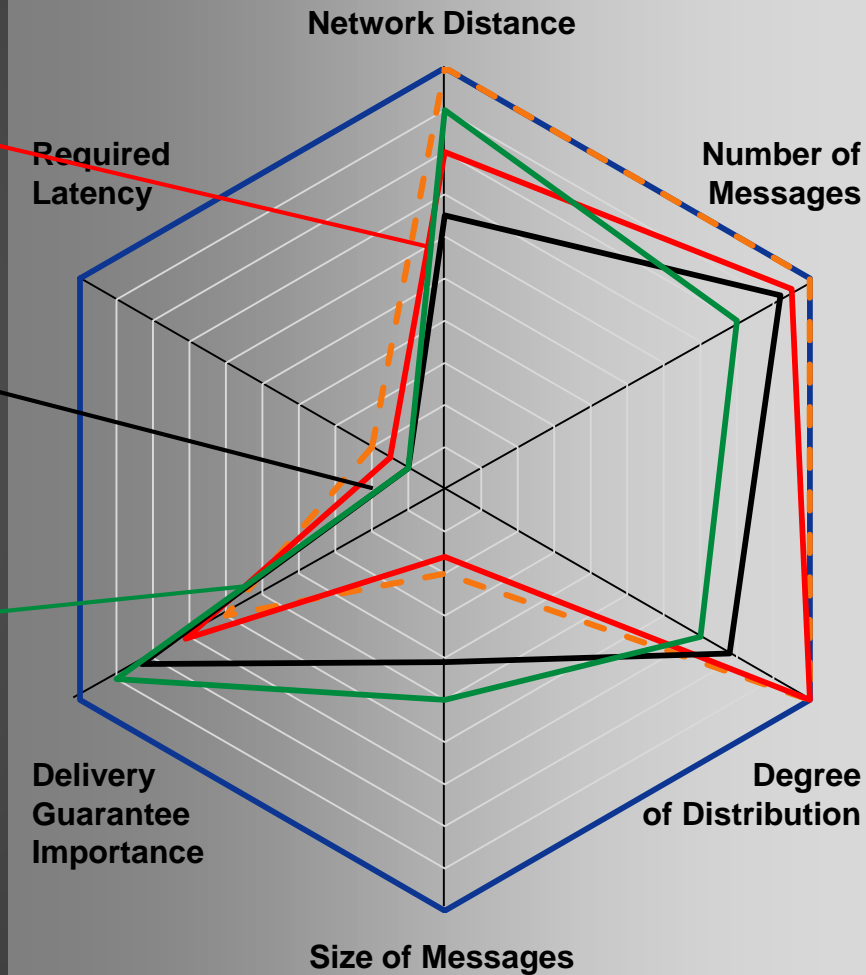
Oil & Gas Pipeline Monitoring; Architecture



Oil & Gas Pipeline Monitoring; Similar Use Cases



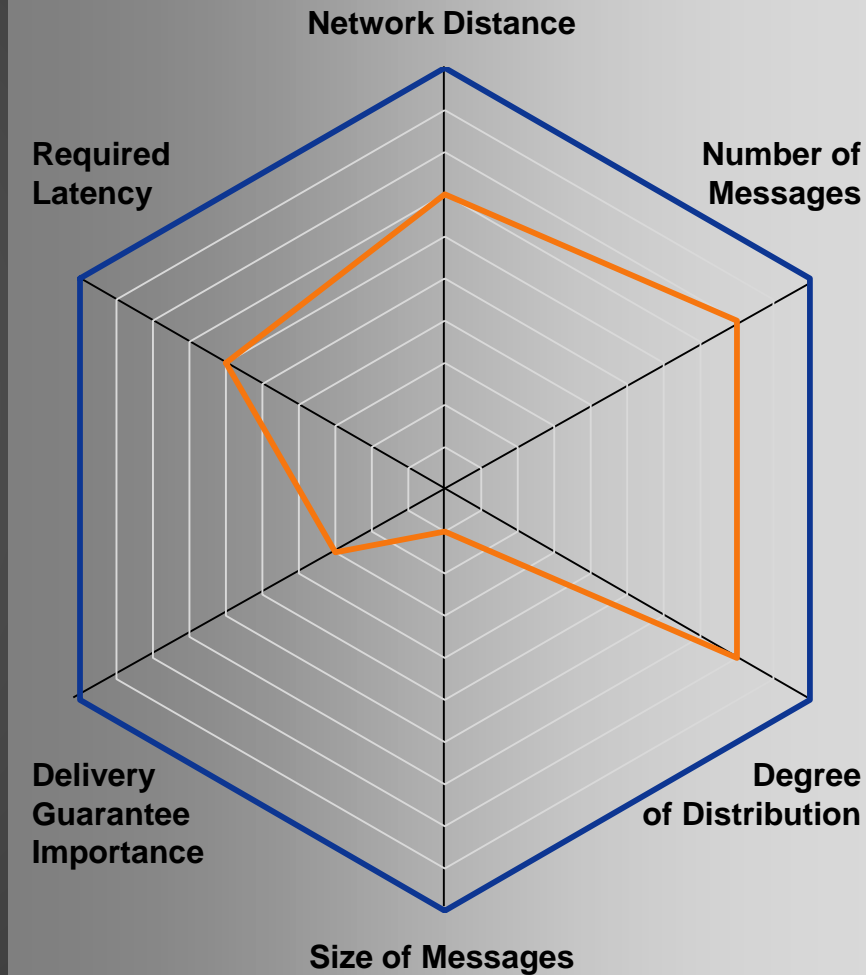
- Smart Grid
 - Small messages
 - Massive distribution
- Transportation Monitoring
 - Fewer endpoints
 - Bigger messages
- Retail Point of Sale
 - More predictable networks
 - Guarantee more important



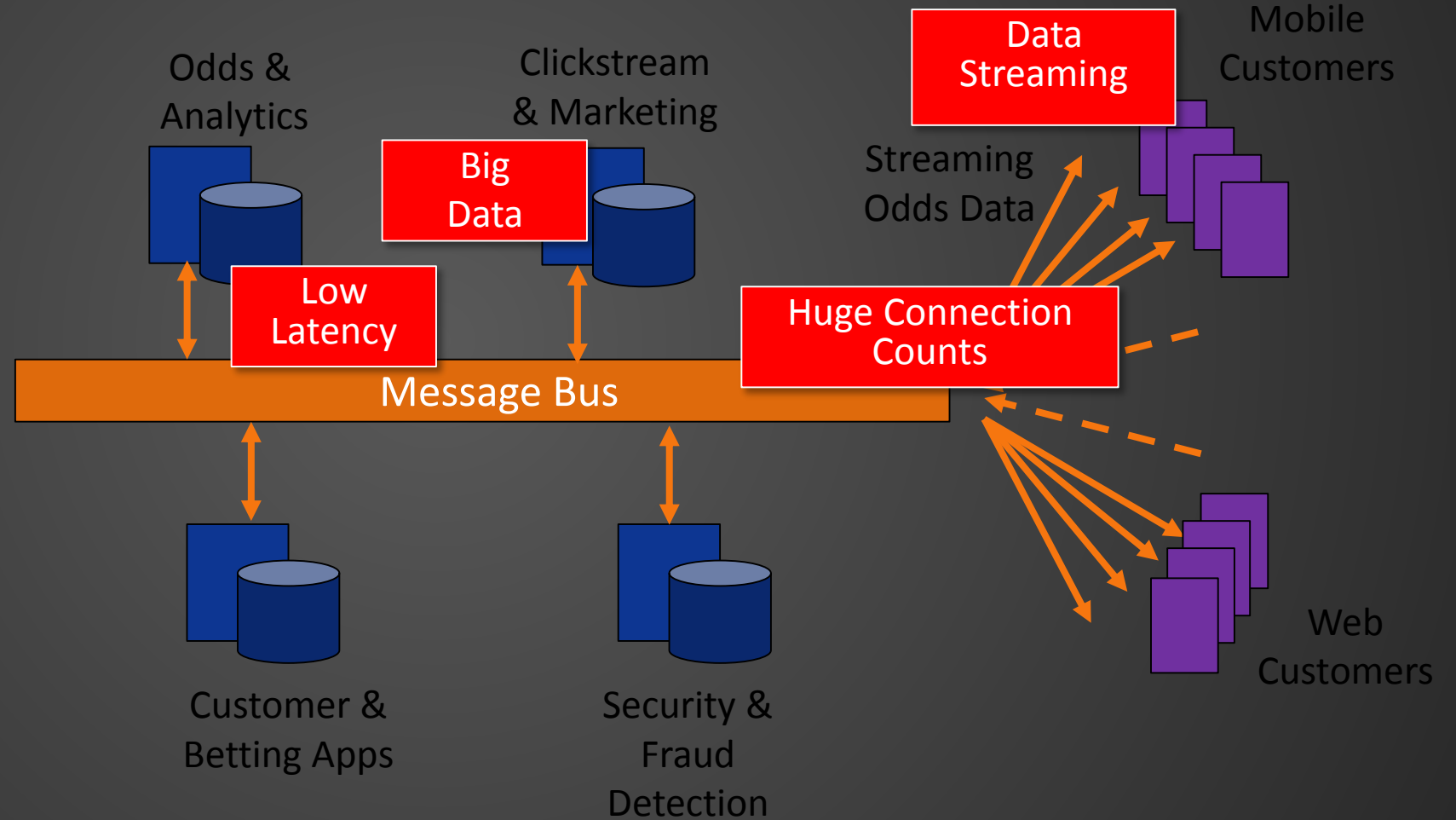
Real-Time Sports Betting



- Huge message volumes (in this case fan-out)
- Low level of guarantee for any one outbound message
- High level of guarantee for inbound messages
- Tiny messages
- Network is the internet + mobile carriers
- Latency (beyond network latency) is important



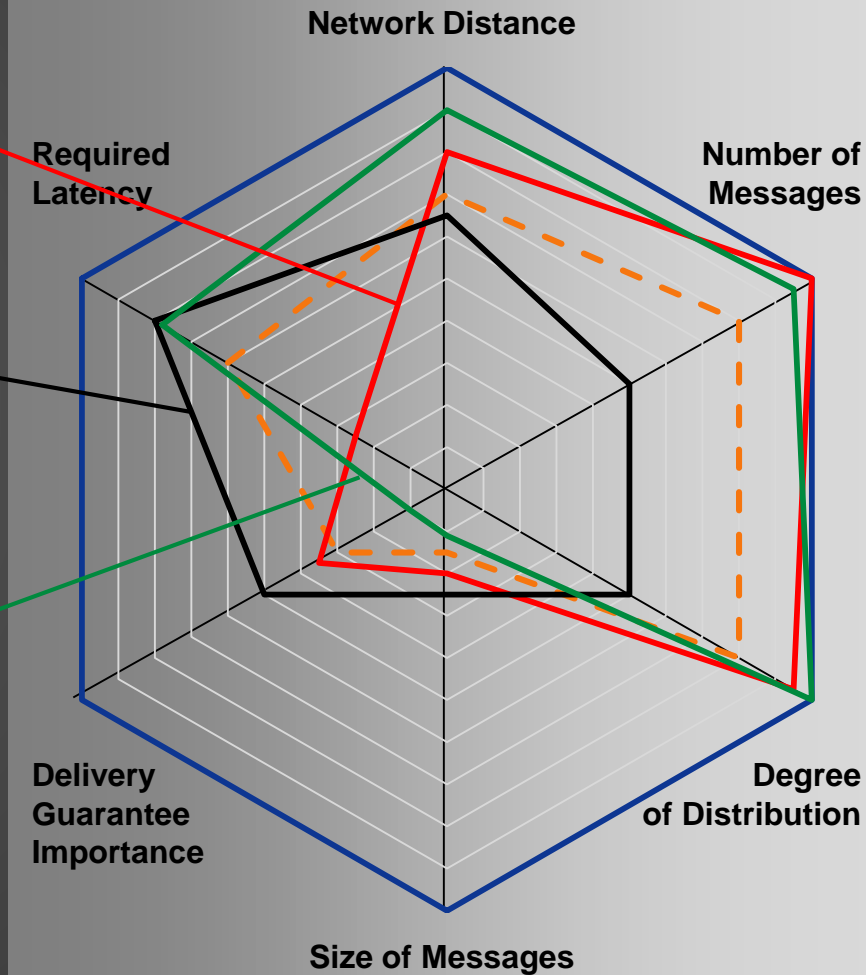
Real-Time Sports Betting; Architecture

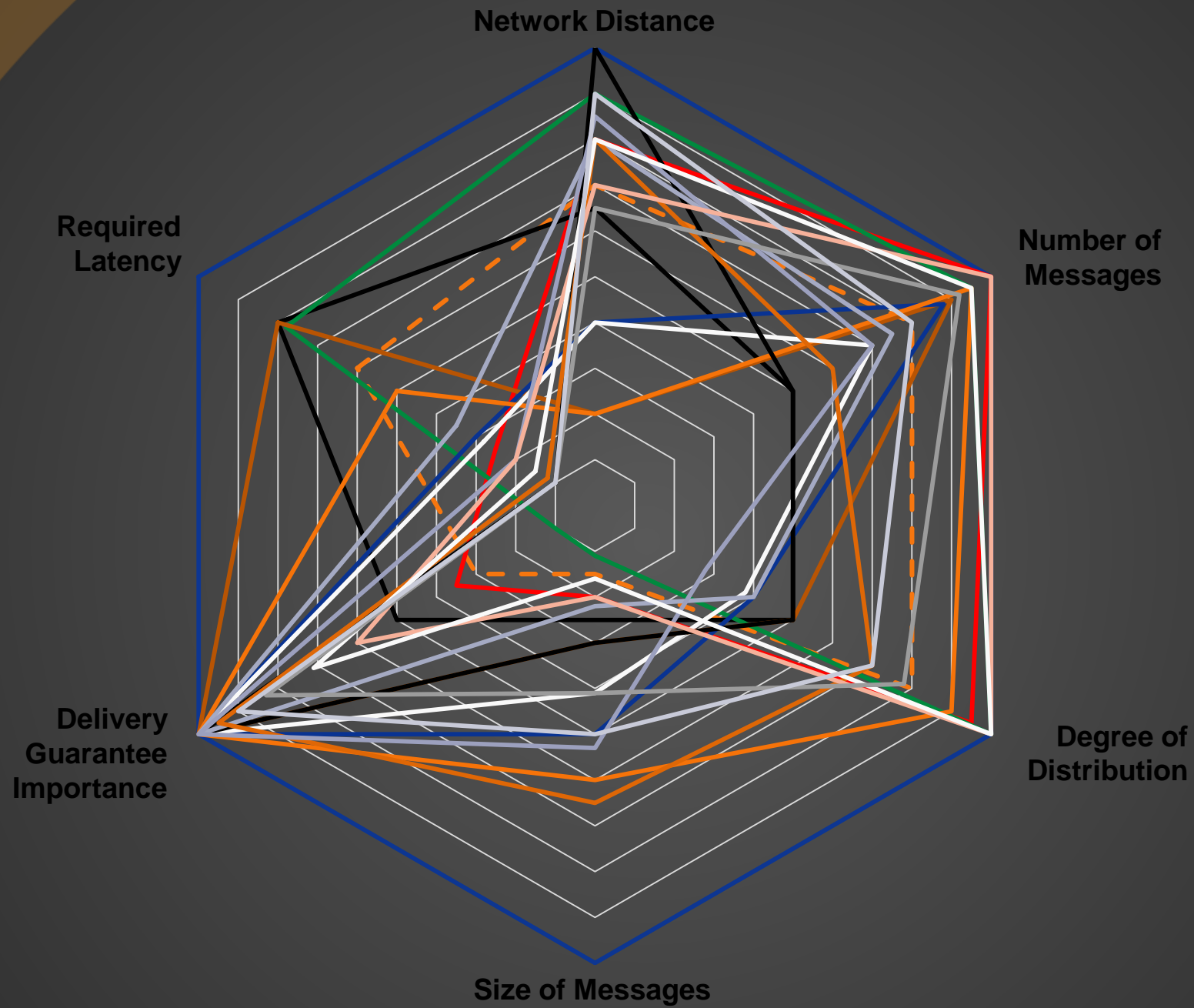


Real-Time Sports Betting; Similar Use Cases



- Mobile Social Updates
 - Latency less important
 - Distribution far greater
- Real Time Travel Alerting
 - Each message more important
 - Volumes much lower
- Market Data Distribution
 - Latency even more important
 - Volumes often much higher
 - Loss often tolerable







Questions?