informatica

Cyber Integration, Message Fabric and Streaming Analytics

SCRE Workshop

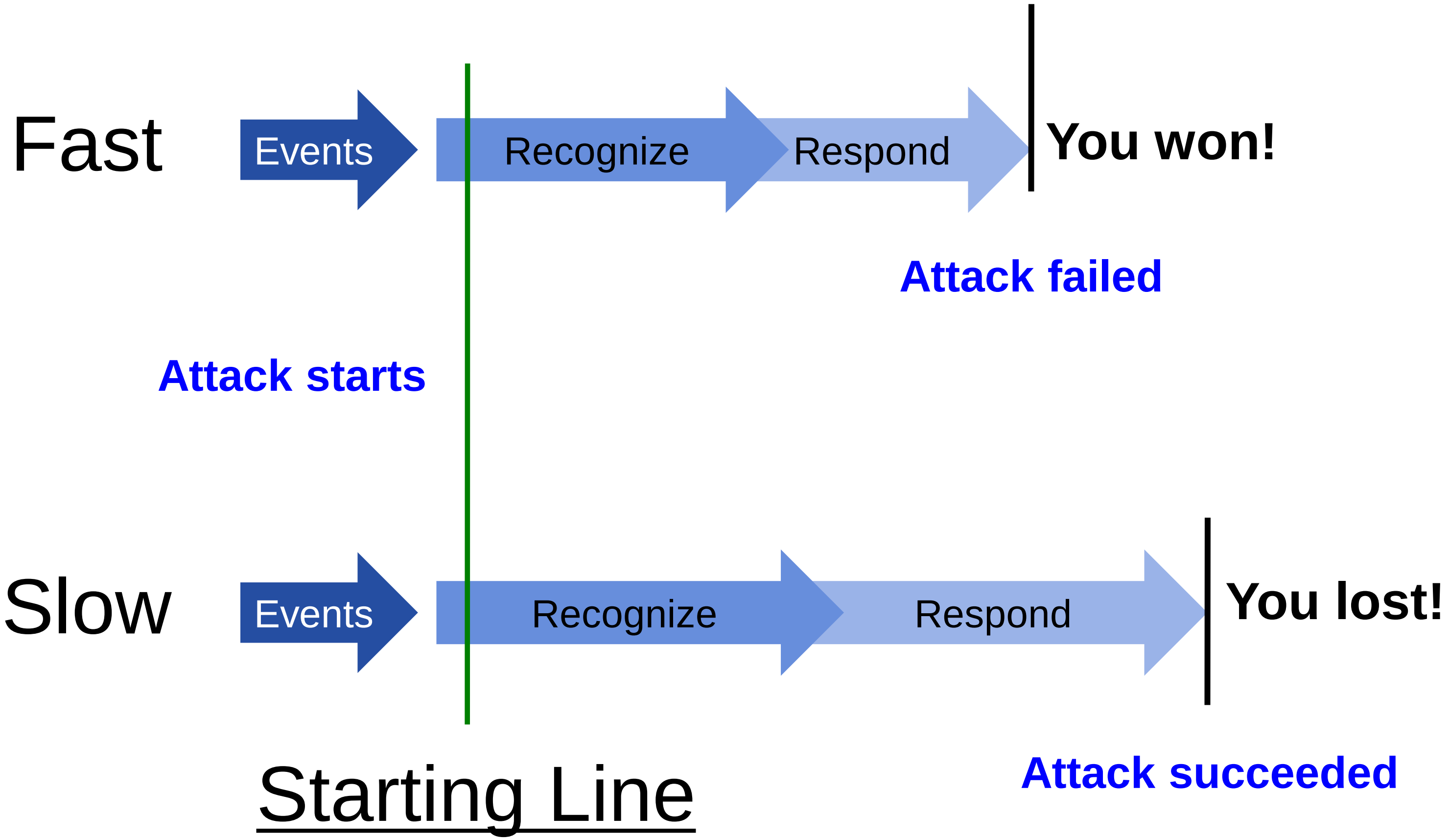
November 17, 2015

Why a Message Fabric for Cyber Integration?

- Abstraction (Pub-Sub, Request / Response, Queuing)
  - **Separate physical systems from communication**; use any infrastructure without changing system behavior
  - **Single point of web-based management**
- Modularity
  - **Quickly add new technologies/algorithms to stay ahead**
- Efficiency
  - Instant response needed? Maybe not, but **latency matters**!
- Functionality
  - Discovery, connectivity, **reliable exchange of data**
  - Guaranteed delivery, fault tolerance, load balancing
  - **Commodity hardware = lower entry and O&M costs**

The Race to Respond

why speed is critical for Cyber Defense



Fast

Events

Recognize

Respond

You won!

Attack failed

Attack starts

Slow

Events

Recognize

Respond

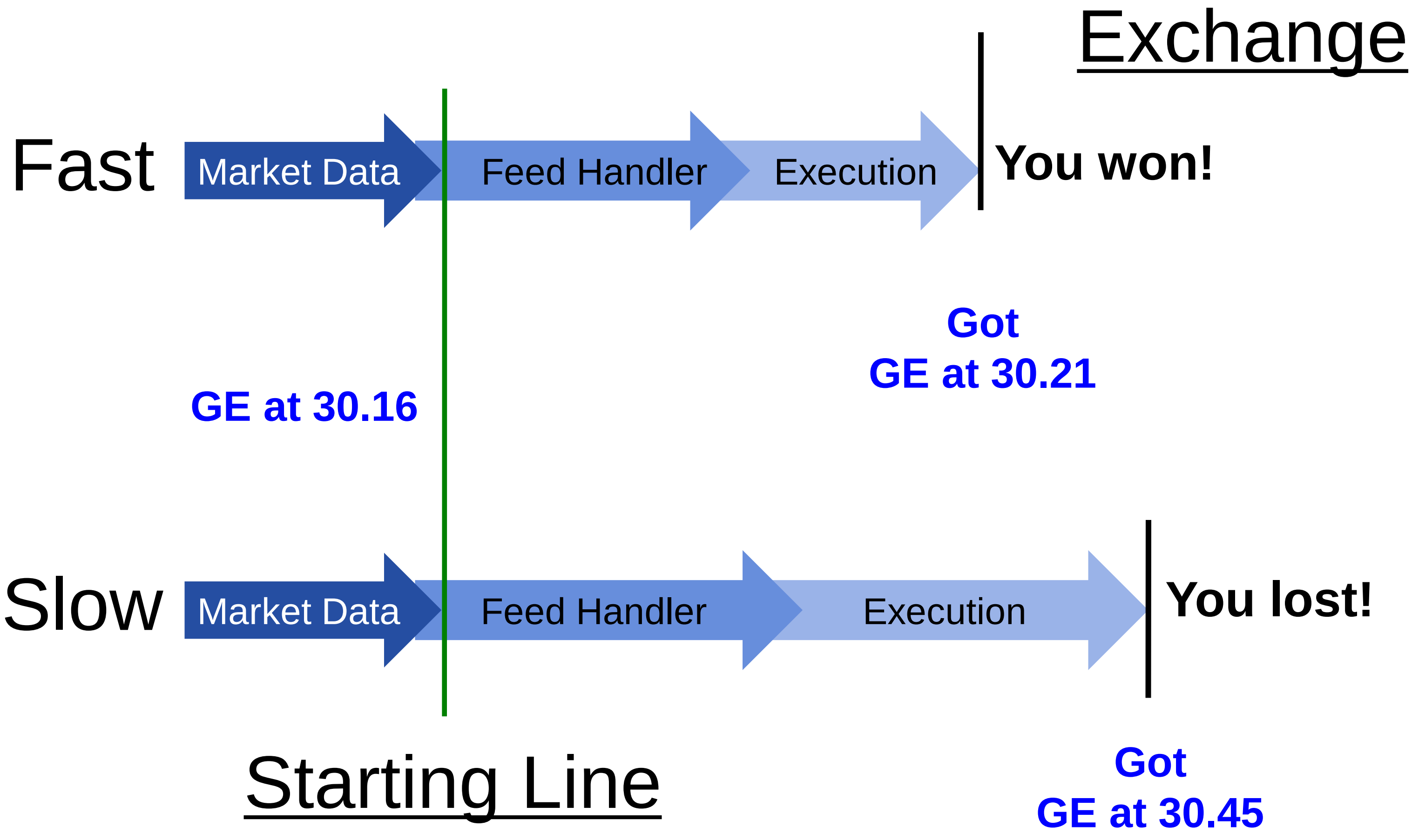
You lost!

Starting Line

Attack succeeded

The Race to the Exchange

why speed is critical for Capital Markets



Exchange

Fast

Market Data → Feed Handler → Execution

You won!

Got
GE at 30.21

GE at 30.16

Slow

Market Data → Feed Handler → Execution

You lost!

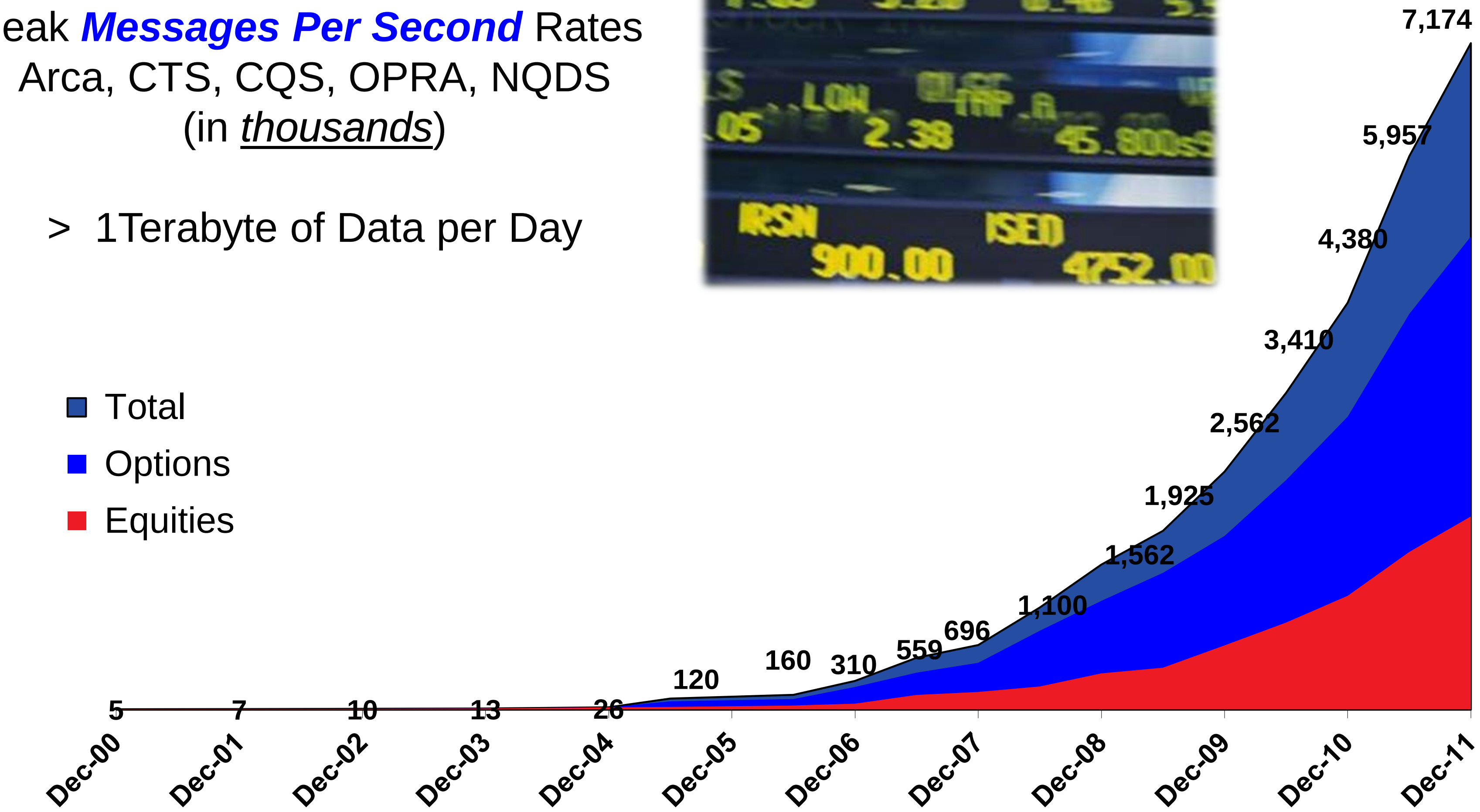
Starting Line

Got
GE at 30.45

Market Data Growth = Data Deluge

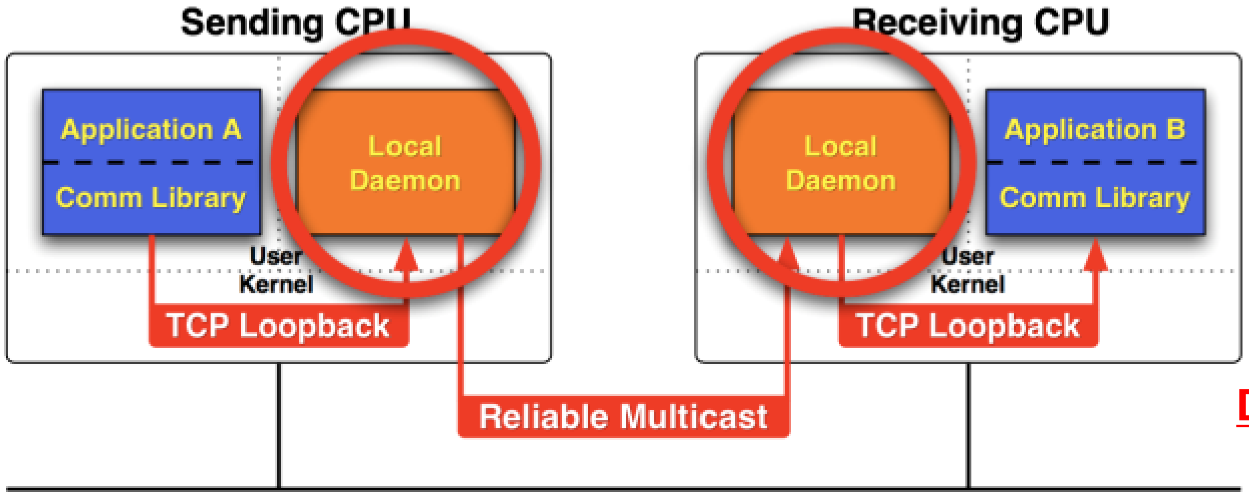
Aggregated One Minute
Peak ***Messages Per Second*** Rates
Arca, CTS, CQS, OPRA, NQDS
(in *thousands*)

> 1Terabyte of Data per Day

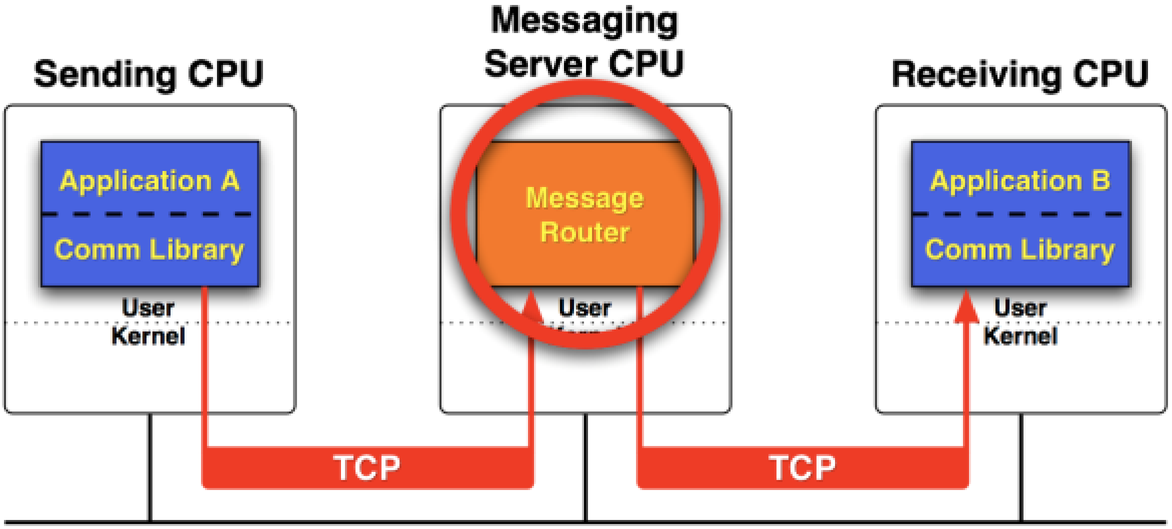


informatica

Legacy Messaging Architectures



Daemon Based Design
6 Data Hops



Broker Based Design
4 Data Hops

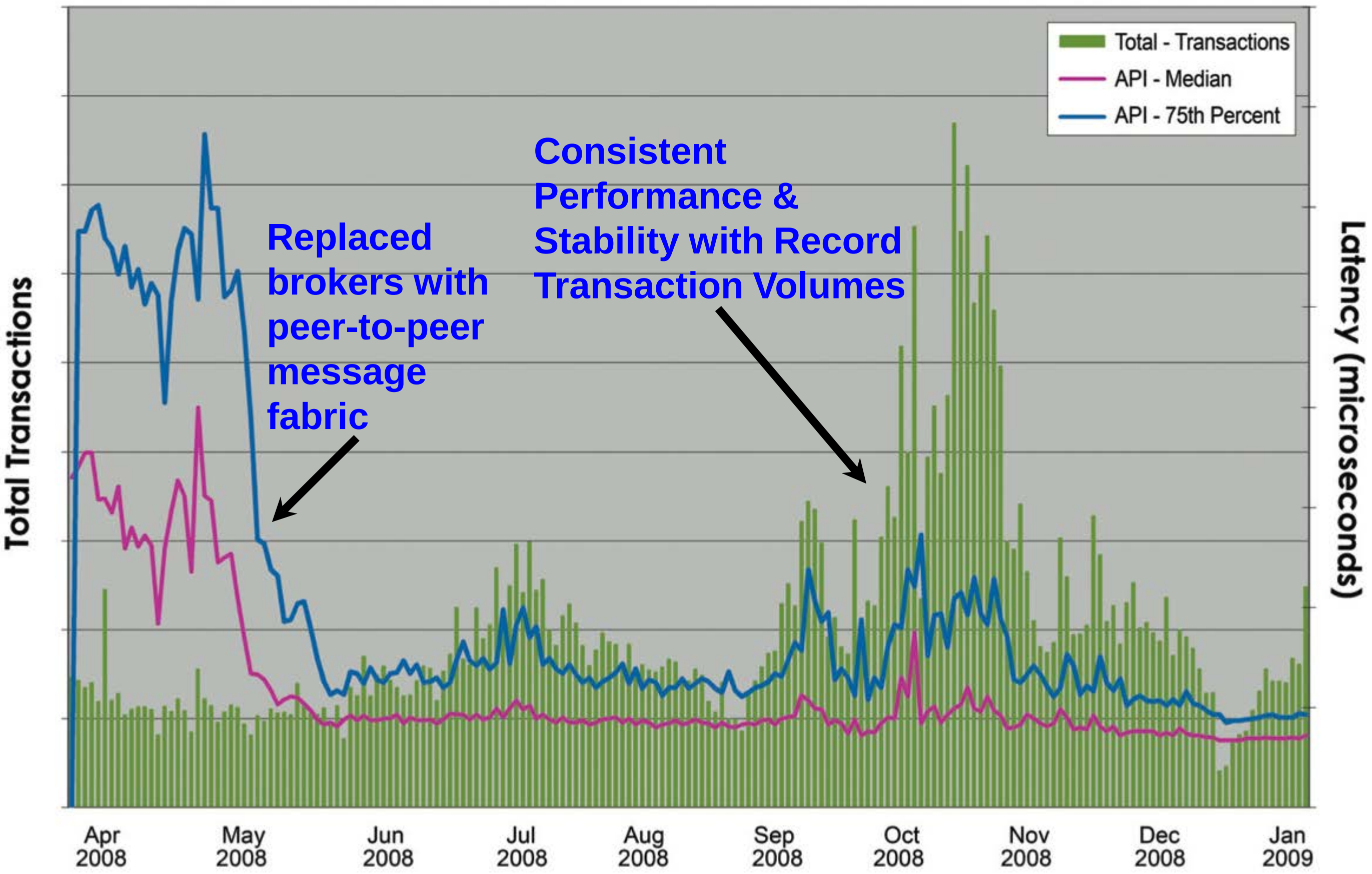
2004 – Need for a State Change

- Motivations / Challenges
  - **Not scaling** to today’s needs (yet alone tomorrow’s!)
  - Availability at risk due to **single points of failure**
- Brokers are a bottleneck
  - Broker is a **source of contention** that limits scaling
  - Broker **failure disastrous** to latency and stability

Remove the Broker from the Message Path!

Case Study: Direct Edge

3rd Largest US Stock Exchange in 2008 (after NYSE and NASDAQ)



Replaced brokers with peer-to-peer message fabric

Consistent Performance & Stability with Record Transaction Volumes

Source: Direct Edge 2008-2009

✓75% lower latency

✓Increased resiliency

✓50% reduction in hardware cost

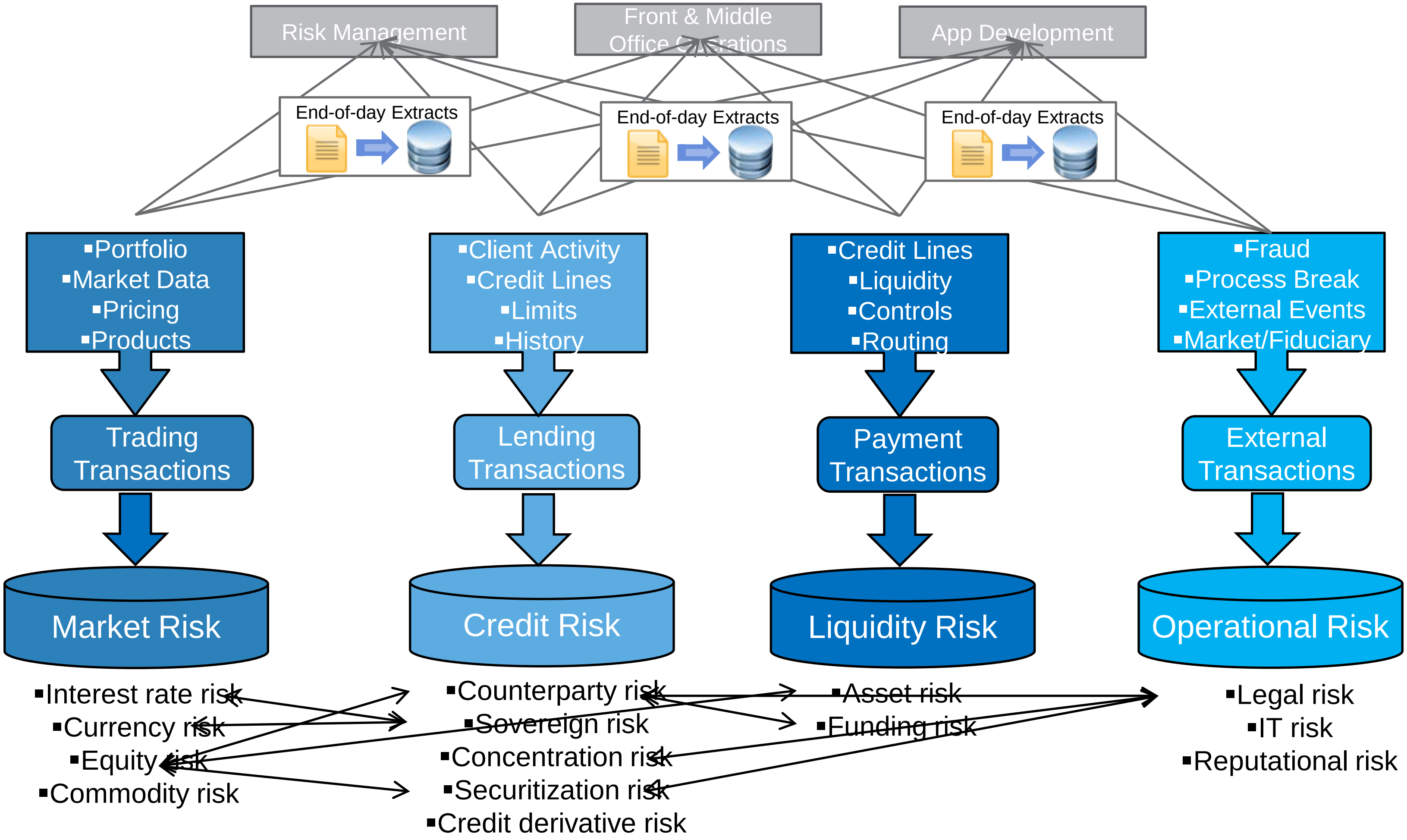
✓Predictable performance

Near Real-Time Financial Data Analytics Framework

Counterparty Risk Assessment

Current State – Disparate Data Siloes

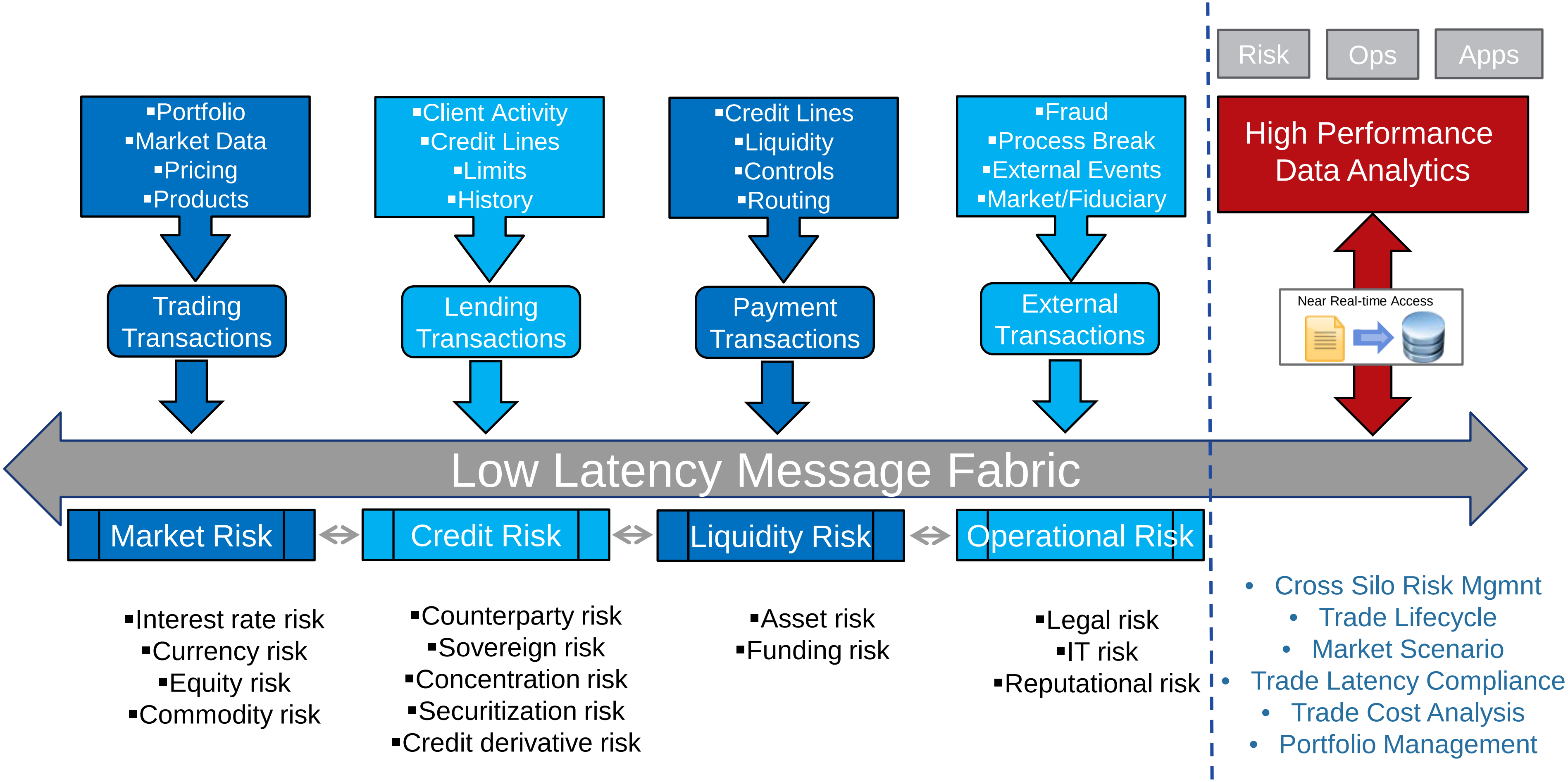
End-of-day extracts and long load/processing times



informatica

Desired State – Correlation across all Data

An Open “Single Source of Truth” for Financial Data



- ▪Portfolio
- ▪Market Data
- ▪Pricing
- ▪Products

Trading Transactions

- ▪Client Activity
- ▪Credit Lines
- ▪Limits
- ▪History

Lending Transactions

- ▪Credit Lines
- ▪Liquidity
- ▪Controls
- ▪Routing

Payment Transactions

- ▪Fraud
- ▪Process Break
- ▪External Events
- ▪Market/Fiduciary

External Transactions

High Performance Data Analytics

Low Latency Message Fabric

Market Risk ⇔ Credit Risk ⇔ Liquidity Risk ⇔ Operational Risk

Market Risk:
- ▪Interest rate risk
- ▪Currency risk
- ▪Equity risk
- ▪Commodity risk

Credit Risk:
- ▪Counterparty risk
- ▪Sovereign risk
- ▪Concentration risk
- ▪Securitization risk
- ▪Credit derivative risk

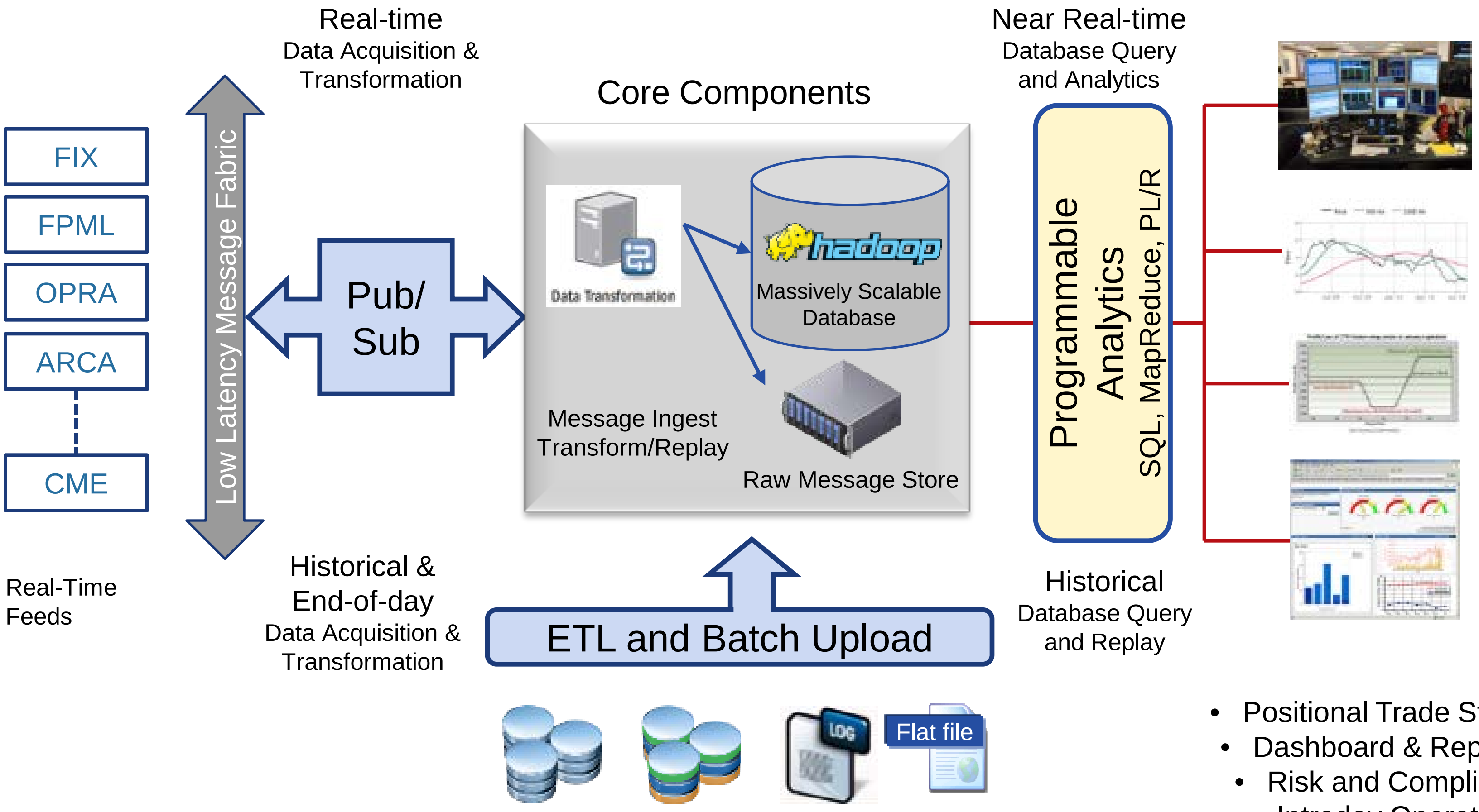
Liquidity Risk:
- ▪Asset risk
- ▪Funding risk

Operational Risk:
- ▪Legal risk
- ▪IT risk
- ▪Reputational risk

- Cross Silo Risk Mgmnt
- Trade Lifecycle
- Market Scenario
- Trade Latency Compliance
- Trade Cost Analysis
- Portfolio Management

Near Real-Time Data Analytics

Real-time & Historical Stock Data with Near Real-time Query



FIX

FPML

OPRA

ARCA

CME

Real-Time Feeds

Low Latency Message Fabric

Real-time
Data Acquisition & Transformation

Pub/ Sub

Historical & End-of-day
Data Acquisition & Transformation

Core Components

Data Transformation

Massively Scalable Database

hadoop

Message Ingest
Transform/Replay

Raw Message Store

ETL and Batch Upload

LOG

Flat file

Near Real-time
Database Query and Analytics

Programmable Analytics
SQL, MapReduce, PL/R

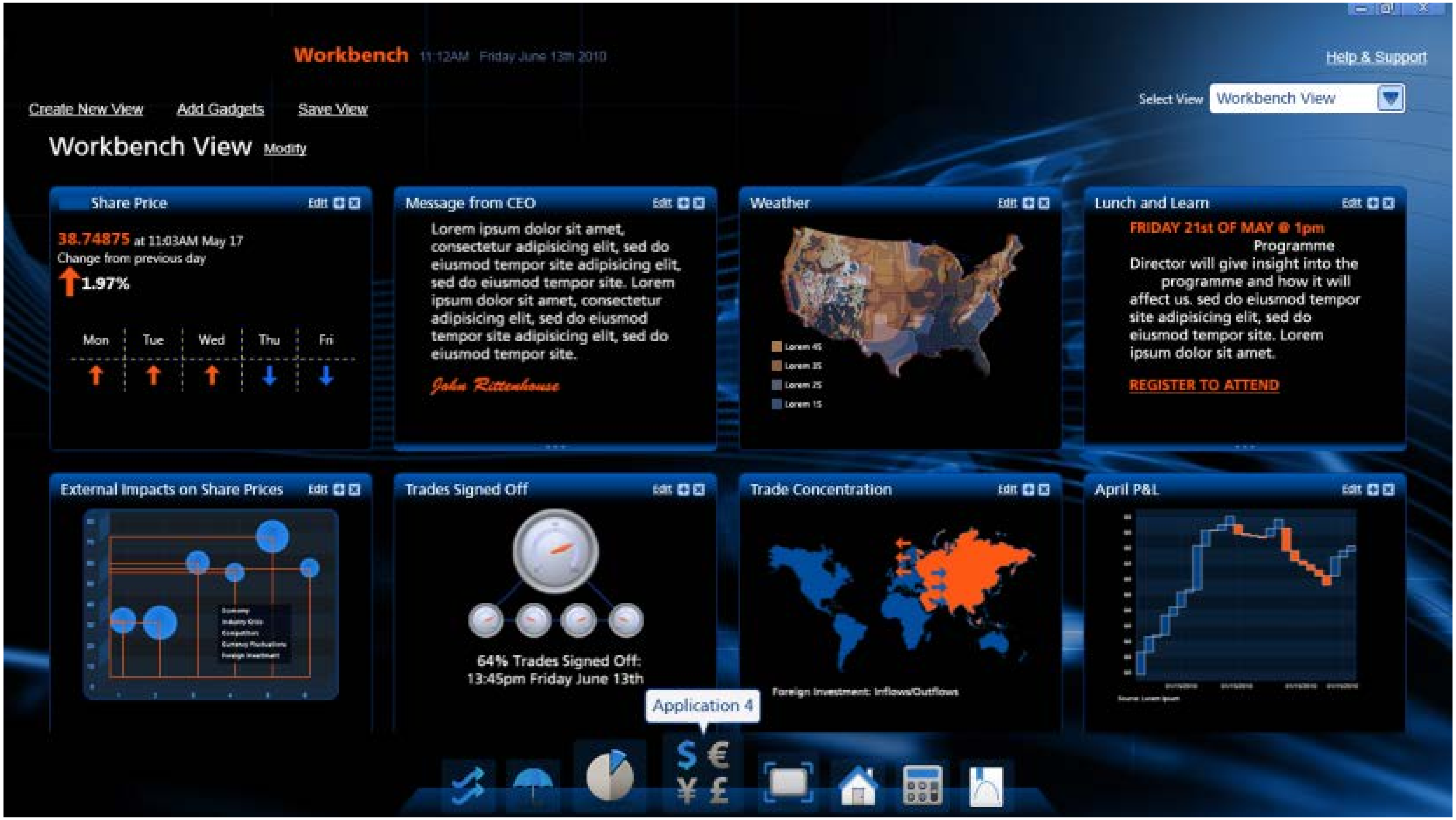
Historical
Database Query and Replay

- Positional Trade Strategy
- Dashboard & Reporting
- Risk and Compliance
- Intraday Operations

informatica

Sample Trade Workbench Real-Time Dashboard

Not a Production View



38.74875 at 11:03AM May 17

Change from previous day

↑1.97%

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor site adipisicing elit, sed do eiusmod tempor site. Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor site adipisicing elit, sed do eiusmod tempor site.

John Rittenhouse

FRIDAY 21st OF MAY @ 1pm

Programme

Director will give insight into the programme and how it will affect us. sed do eiusmod tempor site adipisicing elit, sed do eiusmod tempor site. Lorem ipsum dolor sit amet.

REGISTER TO ATTEND

64% Trades Signed Off: 13:45pm Friday June 13th

What about real-time?

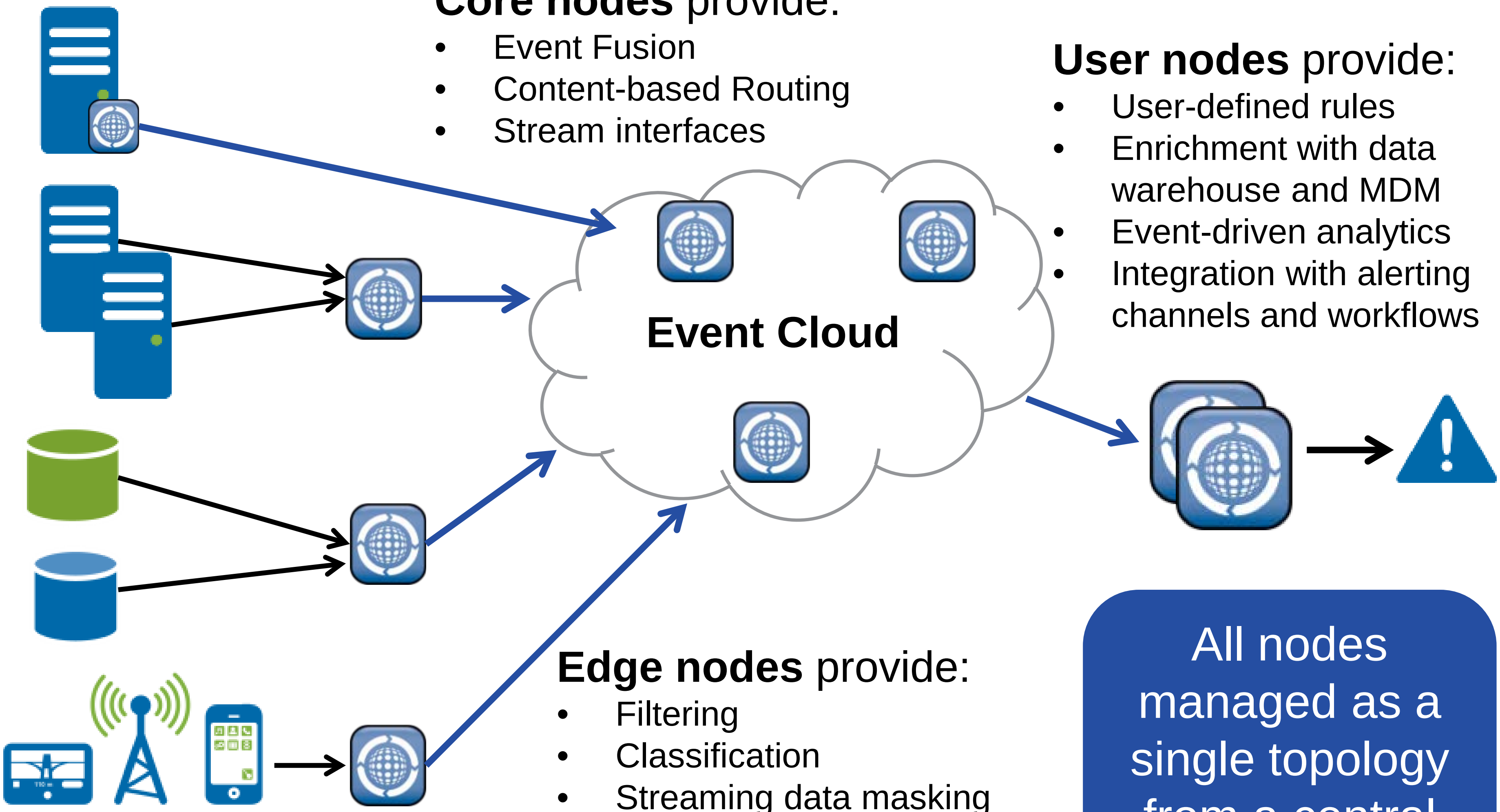
Streaming Analytics and Processing at the Edge

Processing “at the Edge” (and elsewhere)

- Considerations
  - **Aggregation** and **correlation** necessary for “**big picture**”
  - More **distributed** processing power than centralized
  - Raw data is necessary for some types of analysis
    - Is it more efficient to send raw + processed or process later?
- Strategies
  - Derive as much as possible as early as possible
    - **Continuous computation** – counters, distribution statistics
    - **Enrich** (tag/classify unstructured events, add provenance details – origin, identity, versioning, chain of custody)
    - **Exception monitoring** – deviations from norm, trending up/down to exceed thresholds
  - Filter, summarize, compress, transform, mask, encrypt
    - Focus on **state changes** (1111$\underline{0}$000$\underline{1}$11$\underline{0}$0$\underline{1}$11$\underline{0}$0)
    - No-change is data too, but **heartbeats** may be enough

Scalable Deployment w/ Distributed Nodes

AFOC, DCGS-A, DCGS-AF, NATO, IC



Core nodes provide:

- Event Fusion
- Content-based Routing
- Stream interfaces

User nodes provide:

- User-defined rules
- Enrichment with data warehouse and MDM
- Event-driven analytics
- Integration with alerting channels and workflows

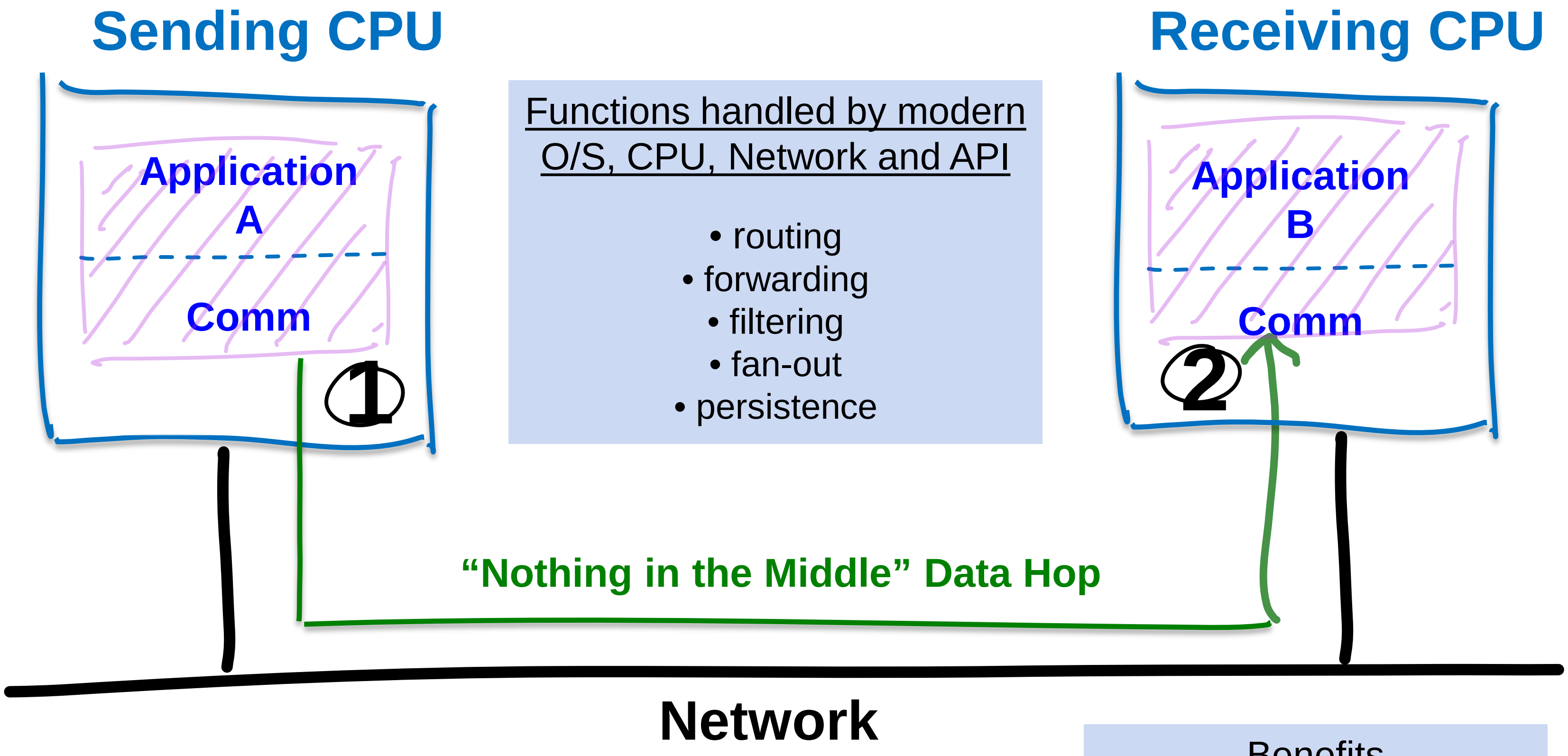
Edge nodes provide:

- Filtering
- Classification
- Streaming data masking
- Real-time aggregates
- In-line enrichment

All nodes managed as a single topology from a central console.

What we want in a Message Fabric

Peer-to-Peer Message Fabric



Functions handled by modern O/S, CPU, Network and API

- routing
- forwarding
- filtering
- fan-out
- persistence

Just 2 steps to move from A to B!!!
Less is more!!

Benefits

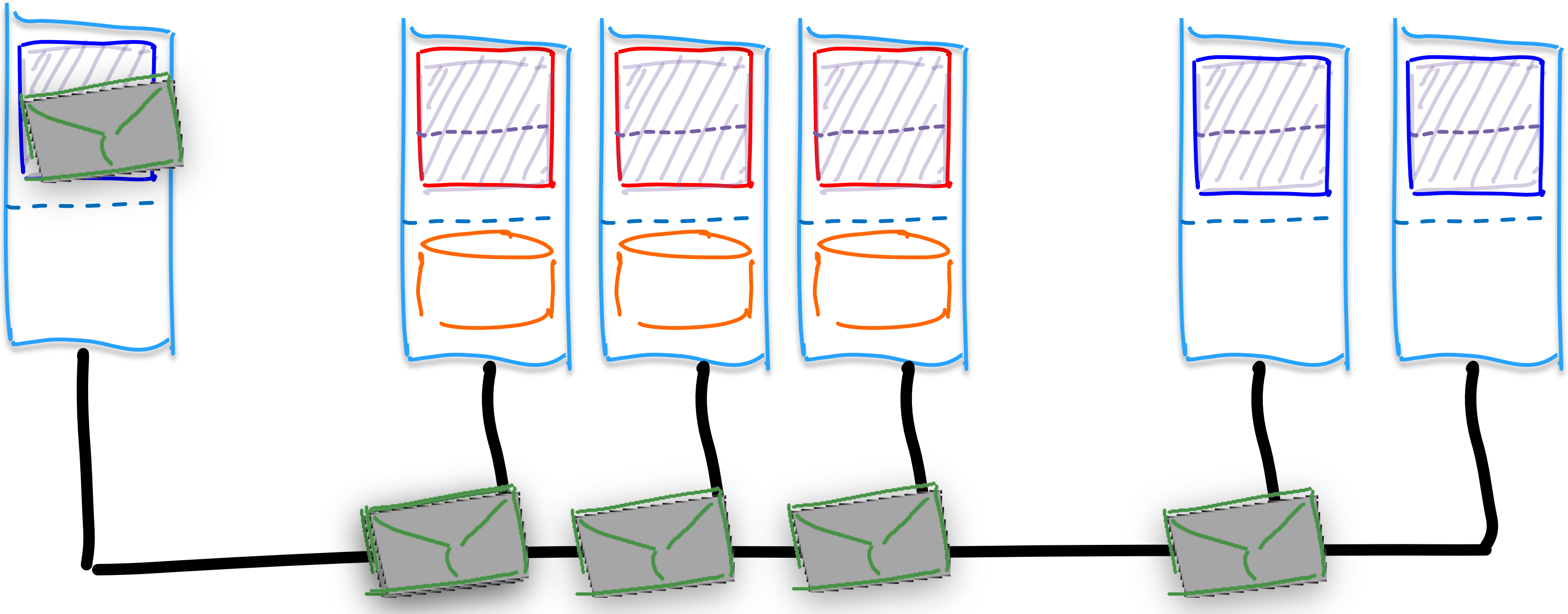
- efficient (single data hop)
- maximizes performance
- no single points of failure
- scalable and flexible
- easier to administer

Parallel Persistence®

Sending Application

Persistent Data Stores

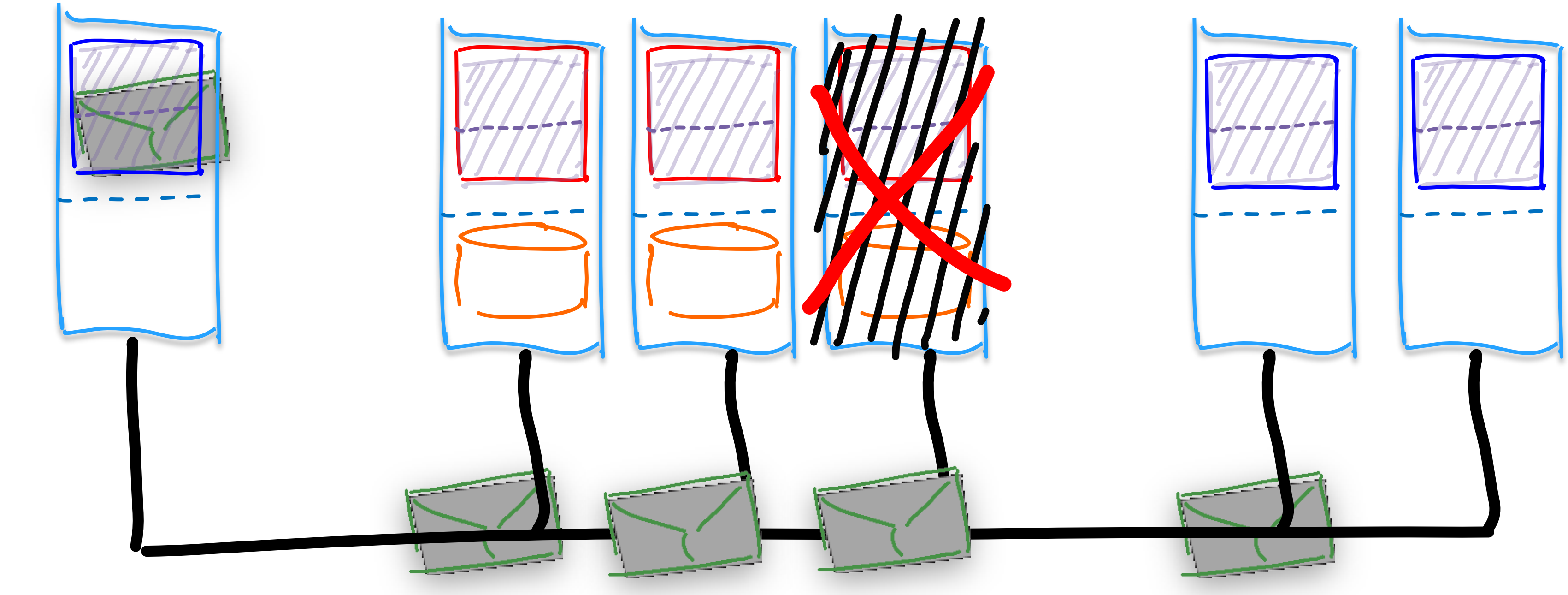
Receiving Applications



Parallel Persistence®

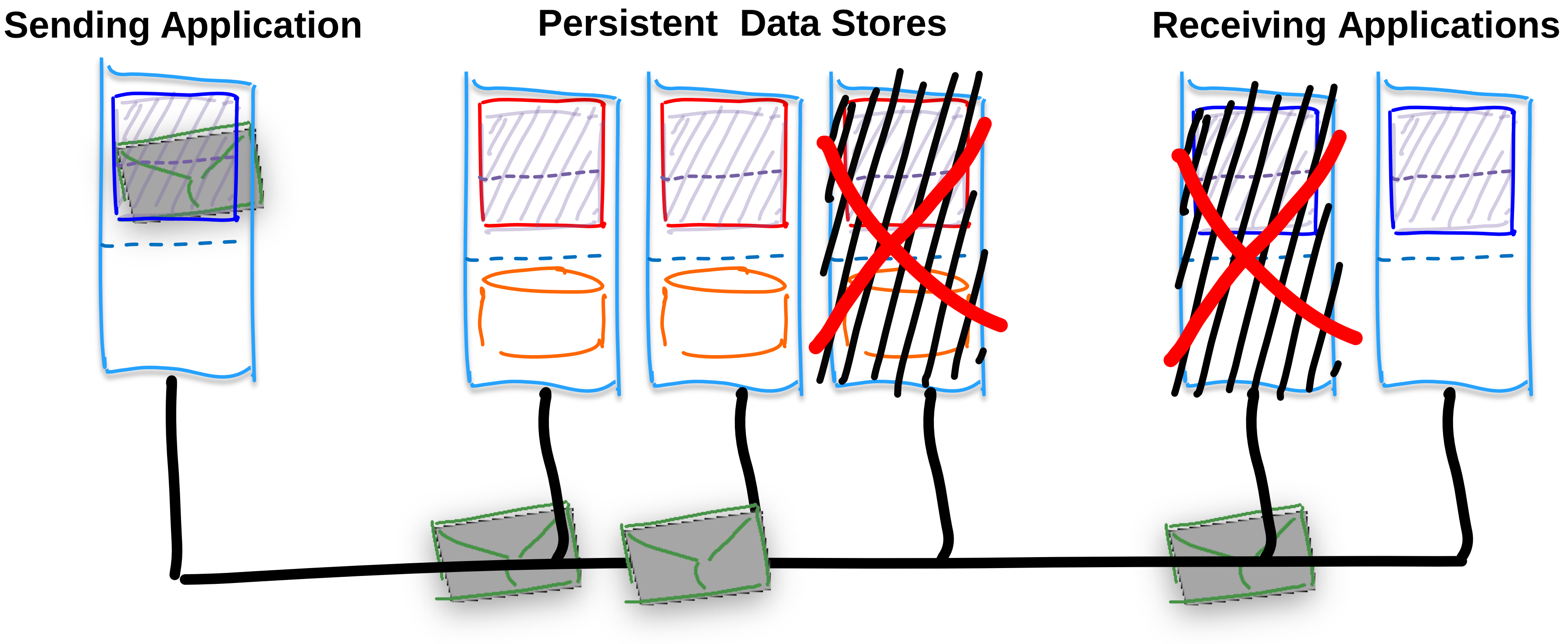
Zero System Downtime!
Zero Latency Failover!

Sending Application **Persistent Data Stores** **Receiving Applications**



informatica

Parallel Persistence®



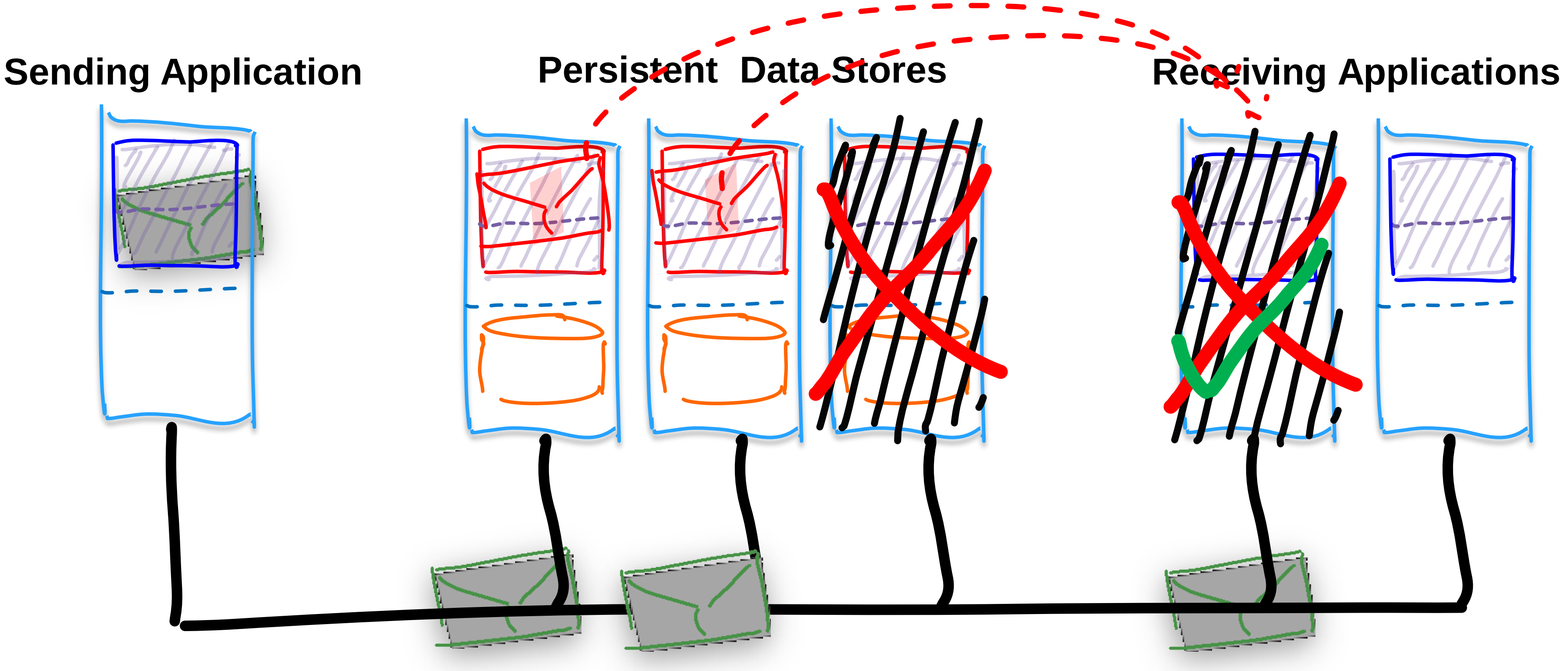
Sending Application

Persistent Data Stores

Receiving Applications

Parallel Persistence®

Receiver recovers with no impact to live message stream, then rejoins the live stream!



Sending Application

Persistent Data Stores

Receiving Applications

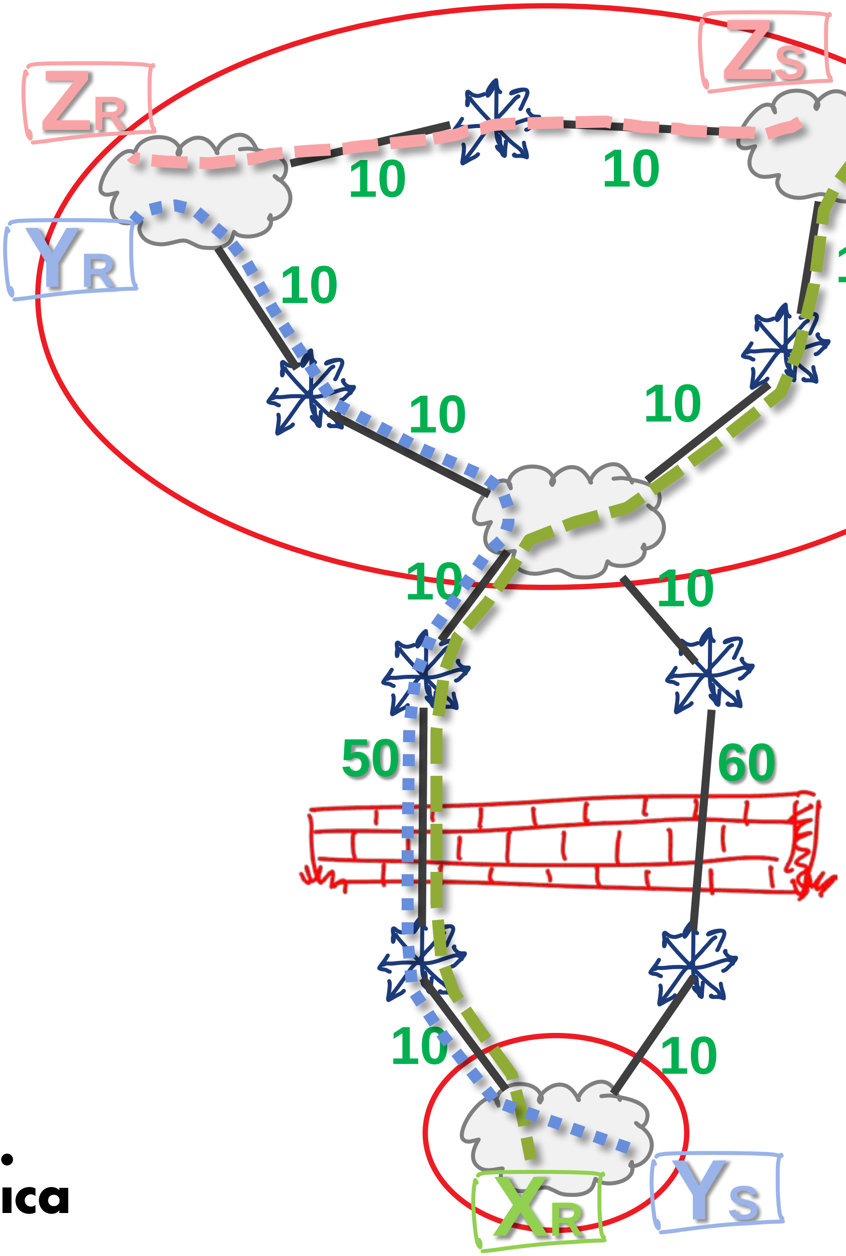
informatica

Extended Enterprise

Considerations:

- Availability
- Authentication
- Authorization
- Bandwidth
- Encryption
- Filtering
- Firewalls
- Protocols
- Routing

Dynamic Routing - Least Cost Path



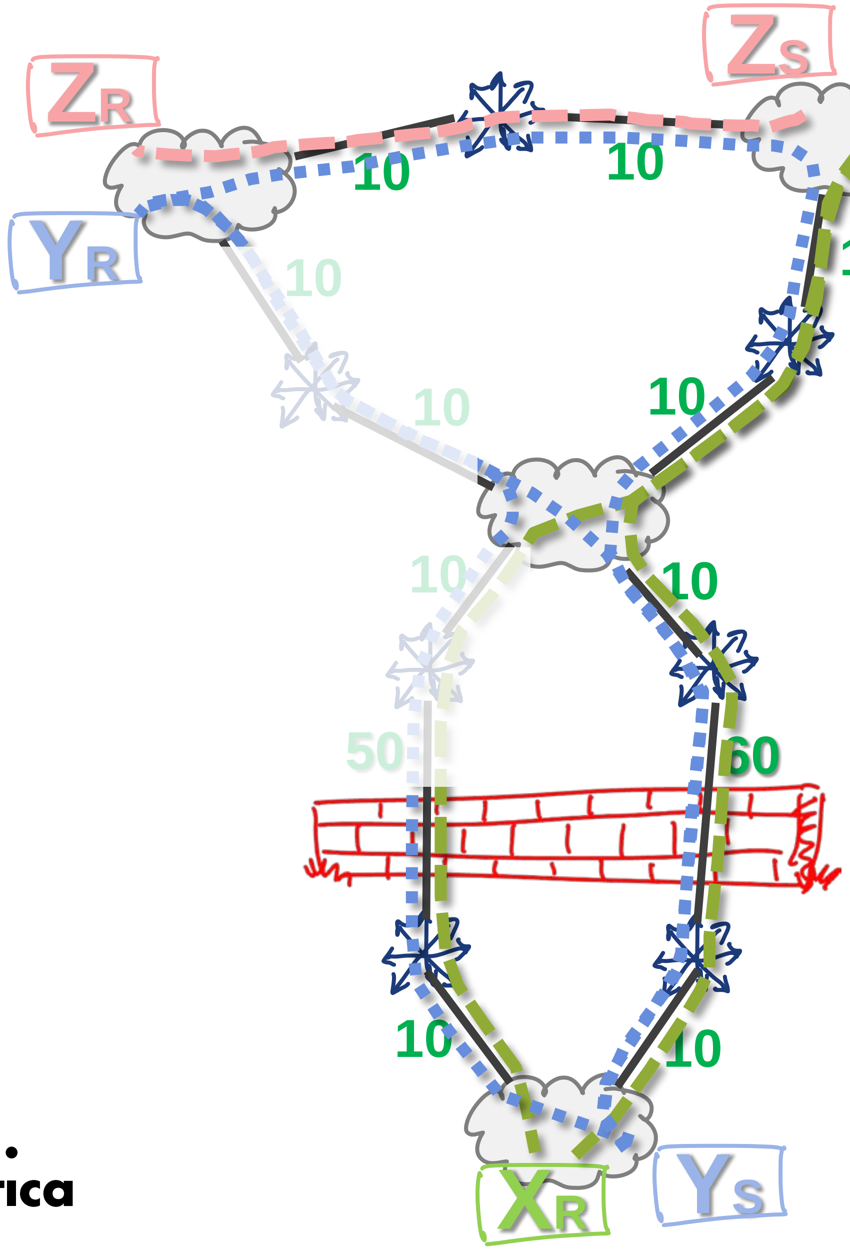
= App Domain

= WAN / firewall

= Message Router

X_S X_R = Sender / Receiver

Dynamic Routing - Least Cost Path



= App Domain

= WAN / firewall

= Message Router

XS
XR
= Sender / Receiver

How can you combine a peer to peer message fabric with standardized interfaces and centralized management?

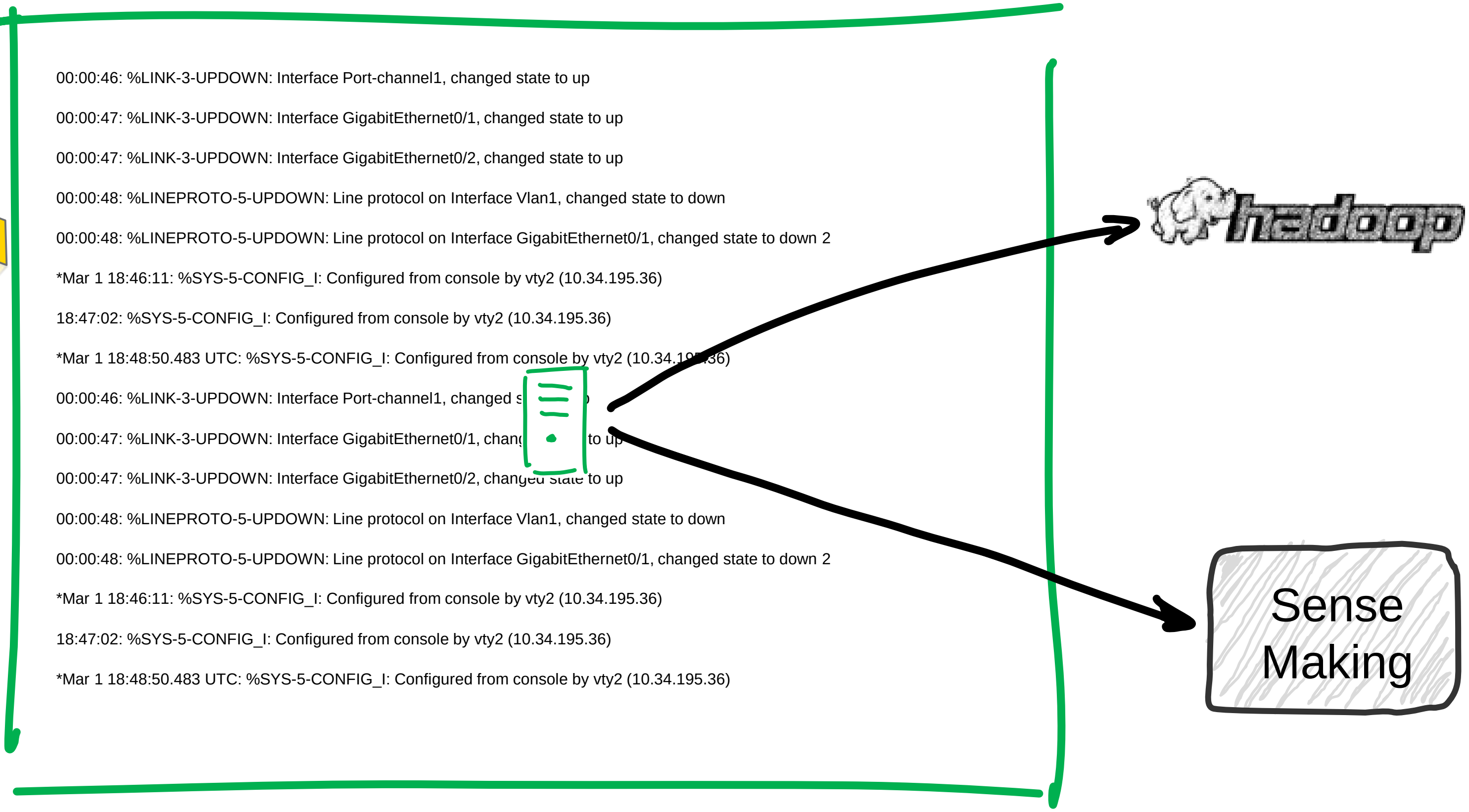
Streaming data collection…

WEB LOG DATA

SERVER LOG DATA

SENSOR DATA

EVENT DATA

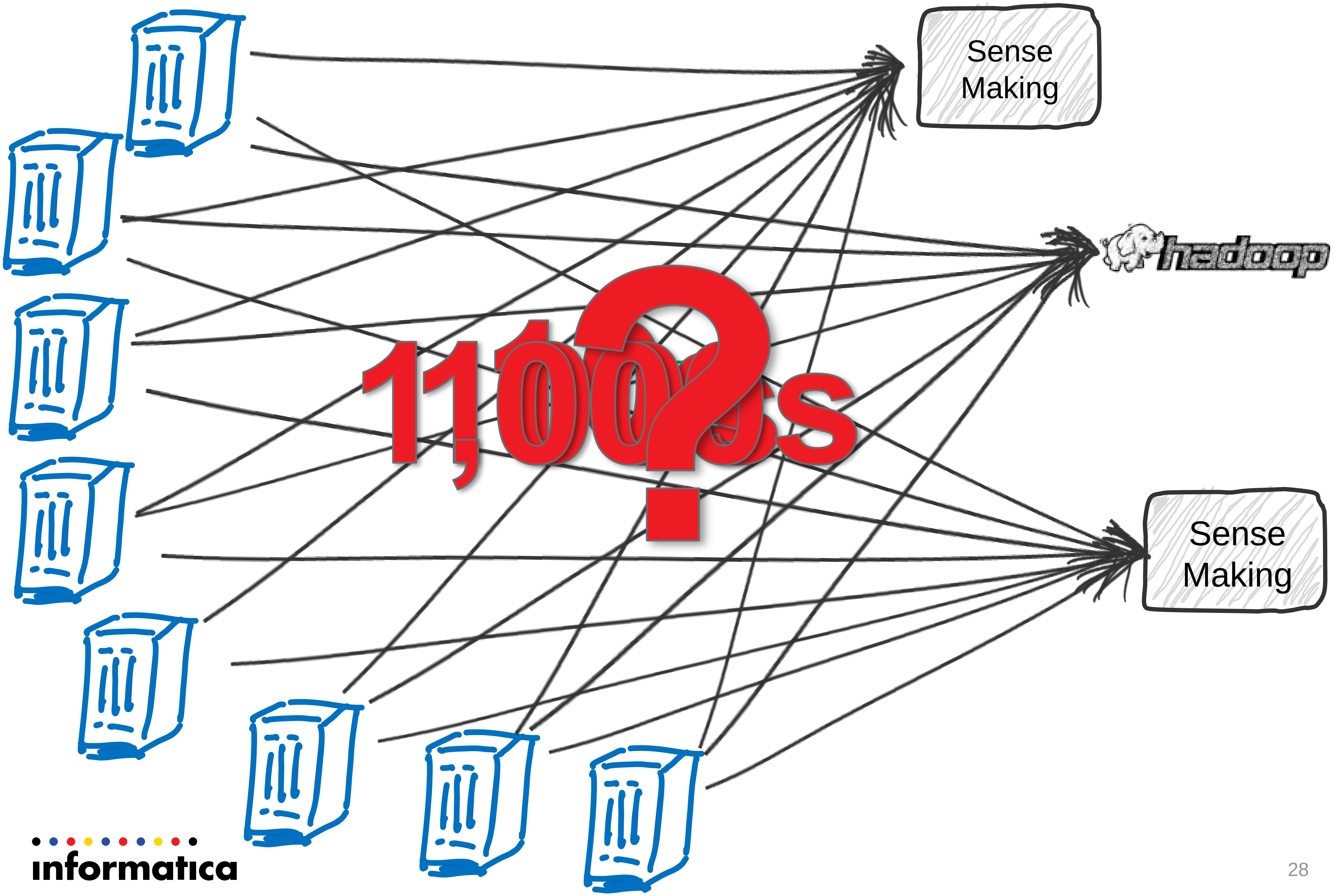


```
00:00:46: %LINK-3-UPDOWN: Interface Port-channel1, changed state to up
00:00:47: %LINK-3-UPDOWN: Interface GigabitEthernet0/1, changed state to up
00:00:47: %LINK-3-UPDOWN: Interface GigabitEthernet0/2, changed state to up
00:00:48: %LINEPROTO-5-UPDOWN: Line protocol on Interface Vlan1, changed state to down
00:00:48: %LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet0/1, changed state to down 2
*Mar 1 18:46:11: %SYS-5-CONFIG_I: Configured from console by vty2 (10.34.195.36)
18:47:02: %SYS-5-CONFIG_I: Configured from console by vty2 (10.34.195.36)
*Mar 1 18:48:50.483 UTC: %SYS-5-CONFIG_I: Configured from console by vty2 (10.34.19 .36)
00:00:46: %LINK-3-UPDOWN: Interface Port-channel1, changed s      )
00:00:47: %LINK-3-UPDOWN: Interface GigabitEthernet0/1, chan      to up
00:00:47: %LINK-3-UPDOWN: Interface GigabitEthernet0/2, changed state to up
00:00:48: %LINEPROTO-5-UPDOWN: Line protocol on Interface Vlan1, changed state to down
00:00:48: %LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet0/1, changed state to down 2
*Mar 1 18:46:11: %SYS-5-CONFIG_I: Configured from console by vty2 (10.34.195.36)
18:47:02: %SYS-5-CONFIG_I: Configured from console by vty2 (10.34.195.36)
*Mar 1 18:48:50.483 UTC: %SYS-5-CONFIG_I: Configured from console by vty2 (10.34.195.36)
```

LOCATION DATA

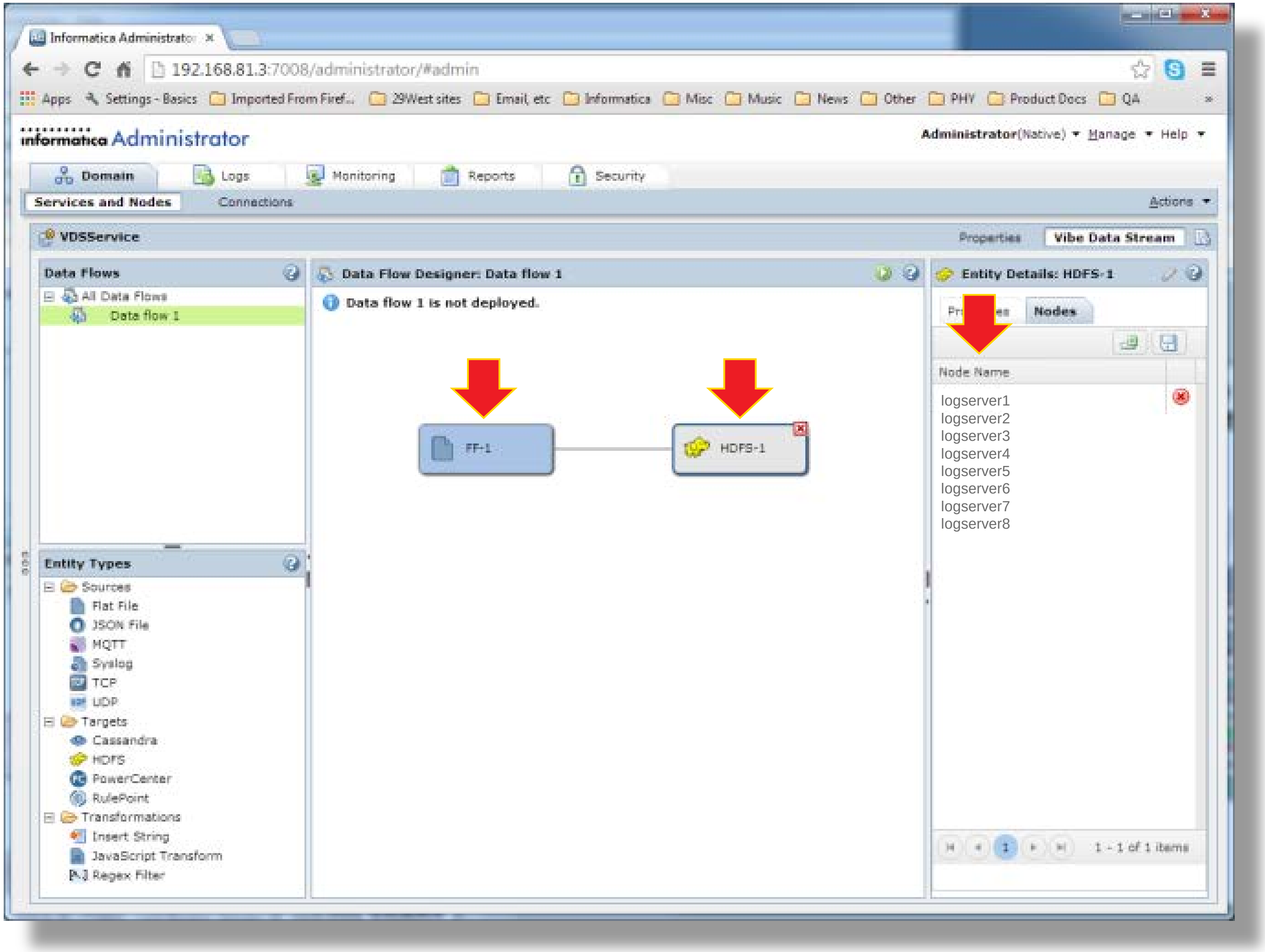
DEVICE DATA

How do you manage this?



informatica

Centralized management, peer to peer data flow



Data flow 1 is not deployed.

logserver1
logserver2
logserver3
logserver4
logserver5
logserver6
logserver7
logserver8

Summary: Essential Characteristics

- No daemons or servers in delivery path
  - Maximize speed and scalability
  - No single points of failure
- Choice of protocols (data “payload” agnostic)
  - TCP, UDP, AMQP, unicast, multicast, shared memory, etc.
- Secure transports, handshakes and storage
  - Integrity, with or without confidentiality
- Secure message routing for extended enterprise
  - Intelligently bridge segmented networks and applications
- Centralized monitoring (with API)
  - Integrated insight from every endpoint (other layers too!)

Summary: Essential Characteristics (cont'd)

- Dynamic service and peer discovery
  - Move applications without changing configuration or code
  - Establish data flows out-of-band to minimize overhead
- Full range of qualities of service
  - From reliable (best-effort) to durable (guaranteed)
- Standards-based interfaces
  - Easily plug in third-party products and services
- Centralized management (with API)
  - Configure top-down; implement locally
- No custom hardware
  - Pure software to always run on best infrastructure

Thank You!

Gay Adams

gadams@informatica.com

cell: 301-980-9148