JUDCon

JBoss Users & Developers Conference

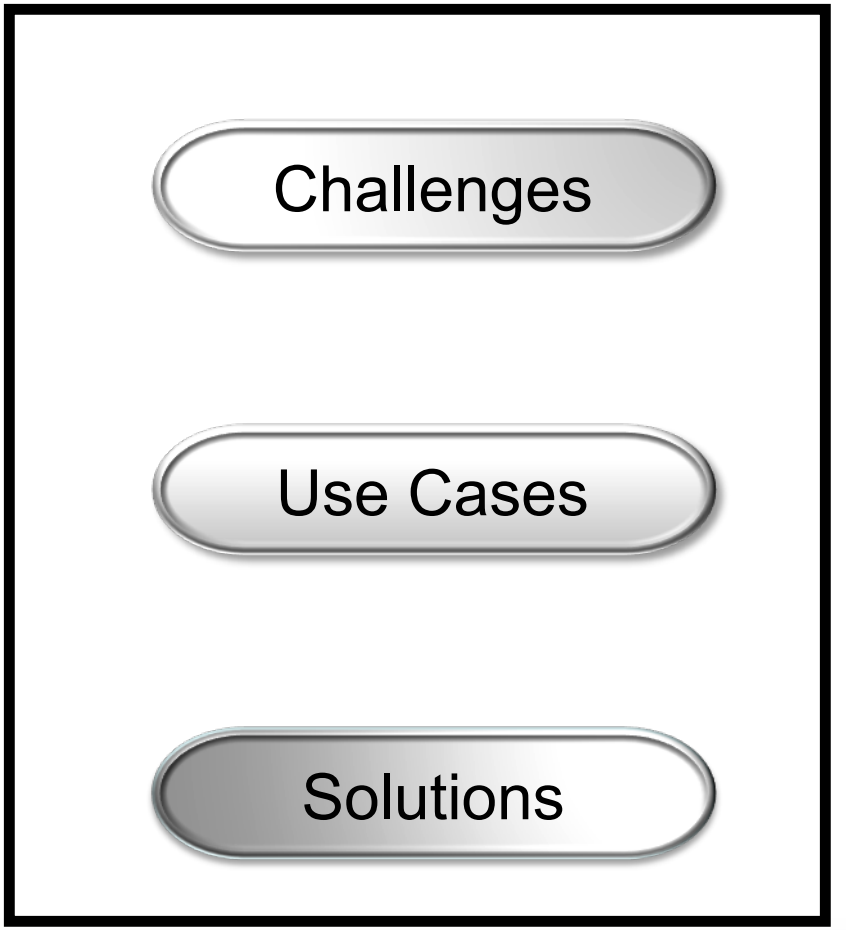
2012:India

Using MRG and Infinispan for Large Scale Integration

Prajod Vettiyattil

What this session is about

- Challenges in *Large Scale integration*
- Use cases for *Large Scale integration*
- How to solve the challenges
  - Solution
  - Products to implement the solution
- The *Open Source* difference



Key Phrases

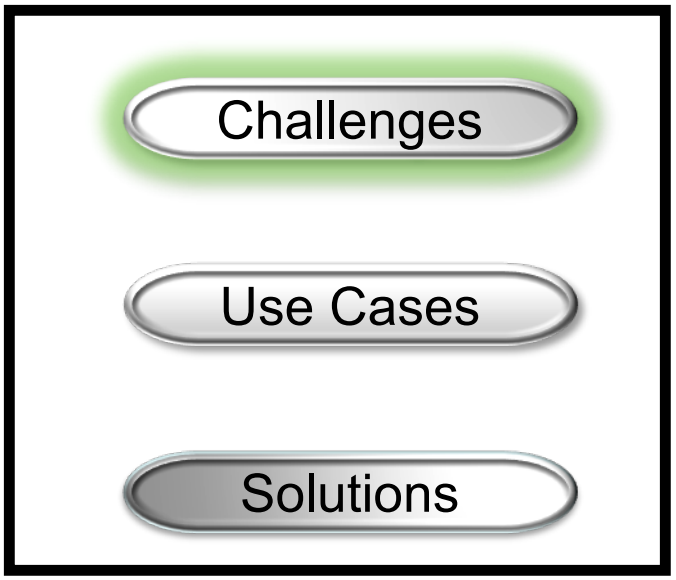
Phrases: 1

- Large Scale Integration
  - Integration of 10s or 100s of systems, and exchange GBs of messages in a day
- Big Data
  - A changing threshold
  - Data in the Terabytes, Petabytes, Exabytes…
- Asynchronous Messaging
  - Message oriented middleware
- Real time systems
  - Systems that are built to respond to requests in real time, with *predicable, consistent* response times

Phrases: 2

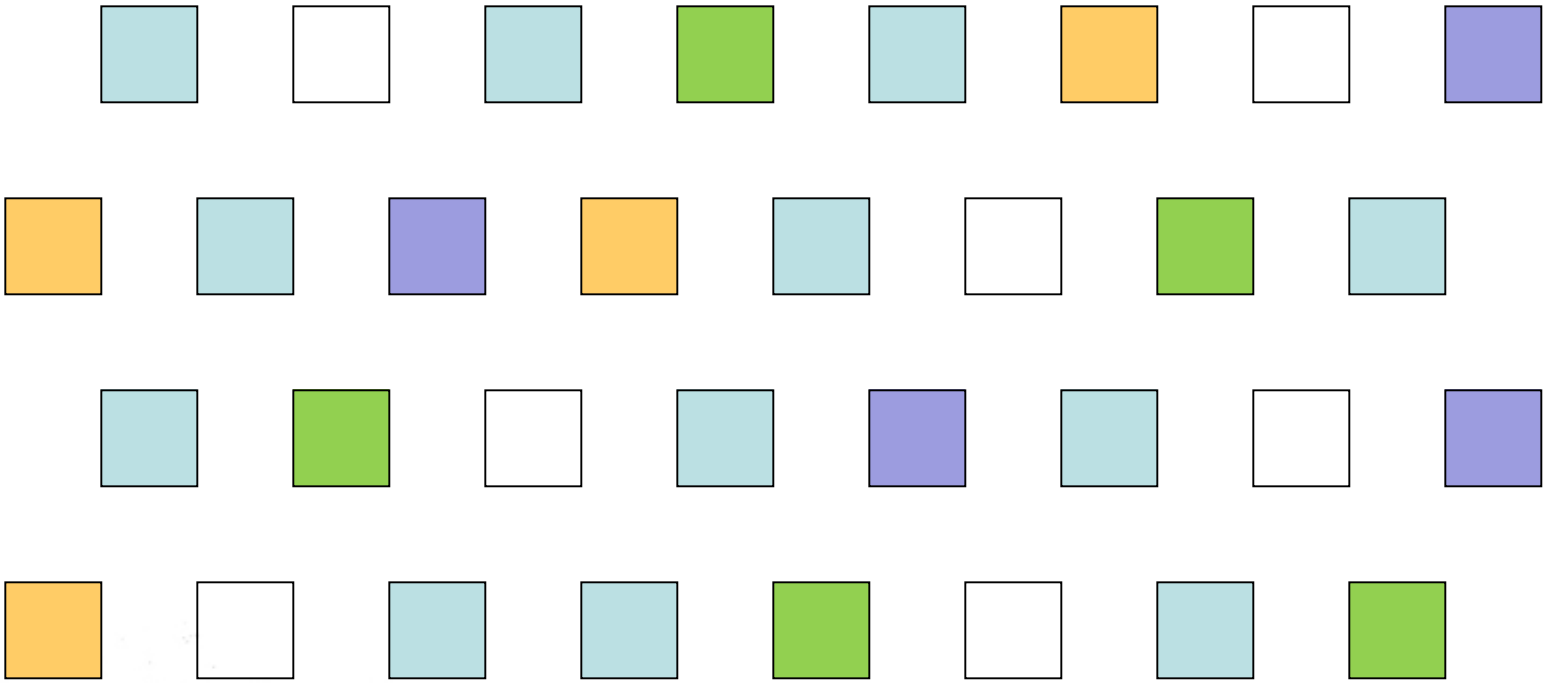
- Grid
  - A set of interconnected computers that work in parallel to solve a computing problem
- Cloud Computing
  - Computing as a service
  - Client of the cloud is isolated from the details of the implementation of the service

Challenges



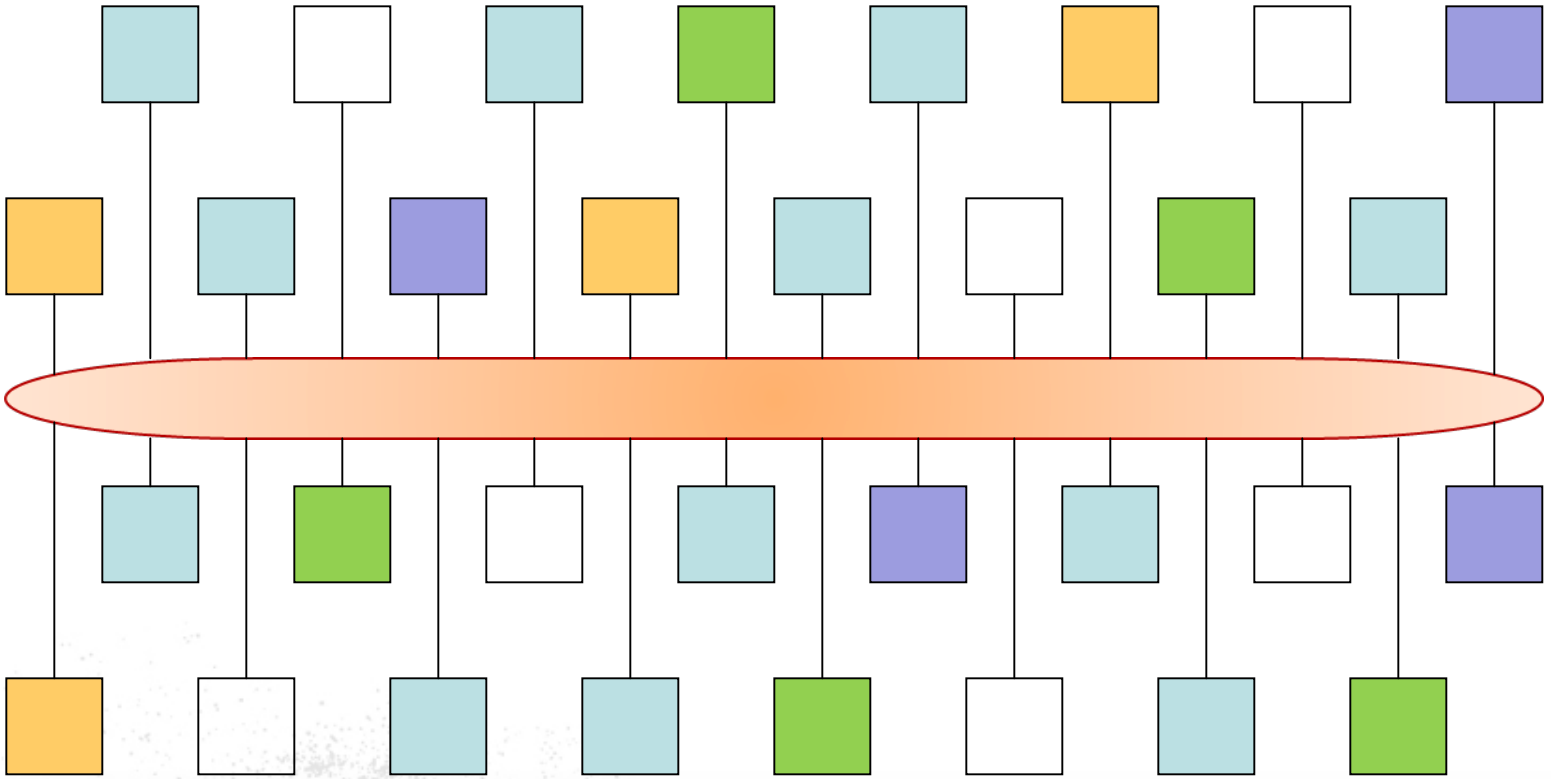
The Key Challenges

Large number of systems



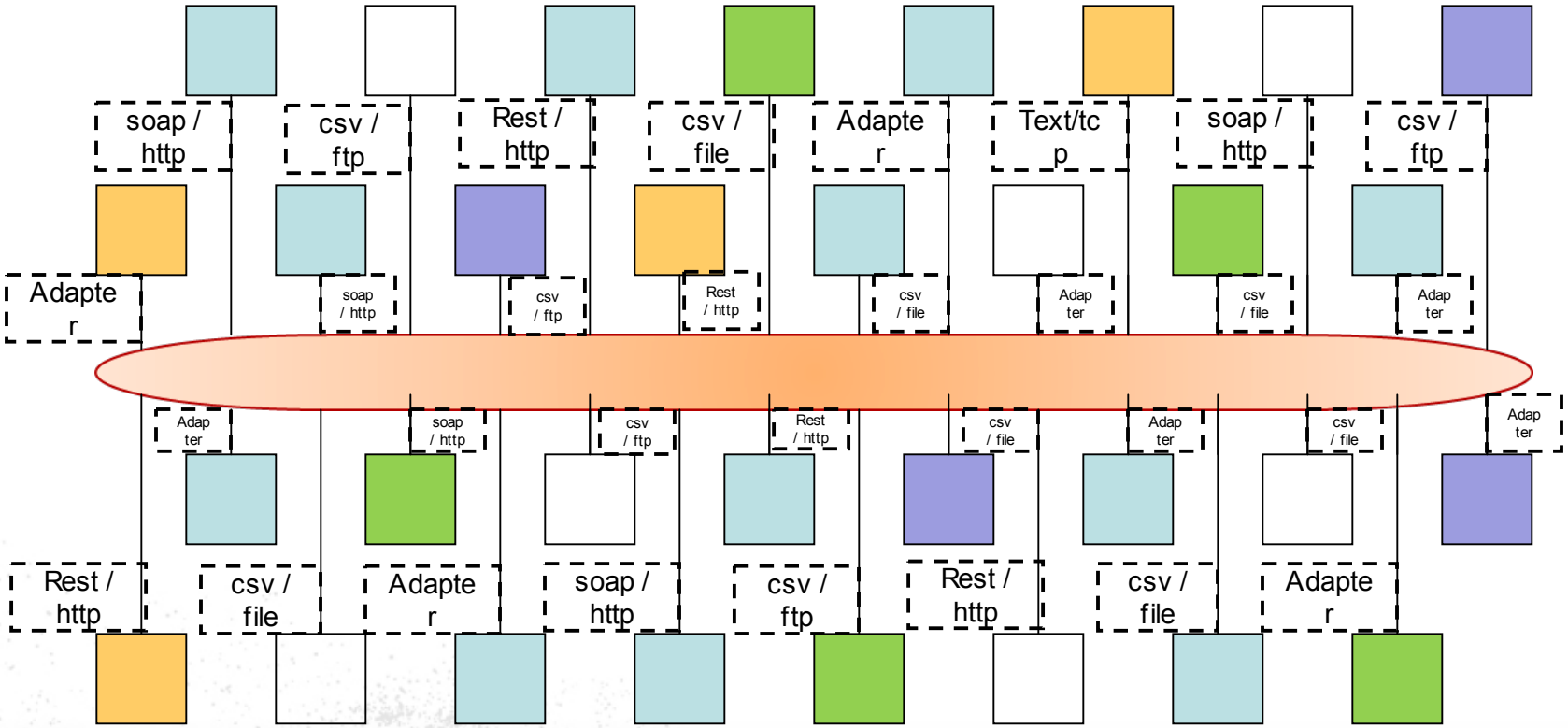
The Key Challenges

Complexity of connection between these systems



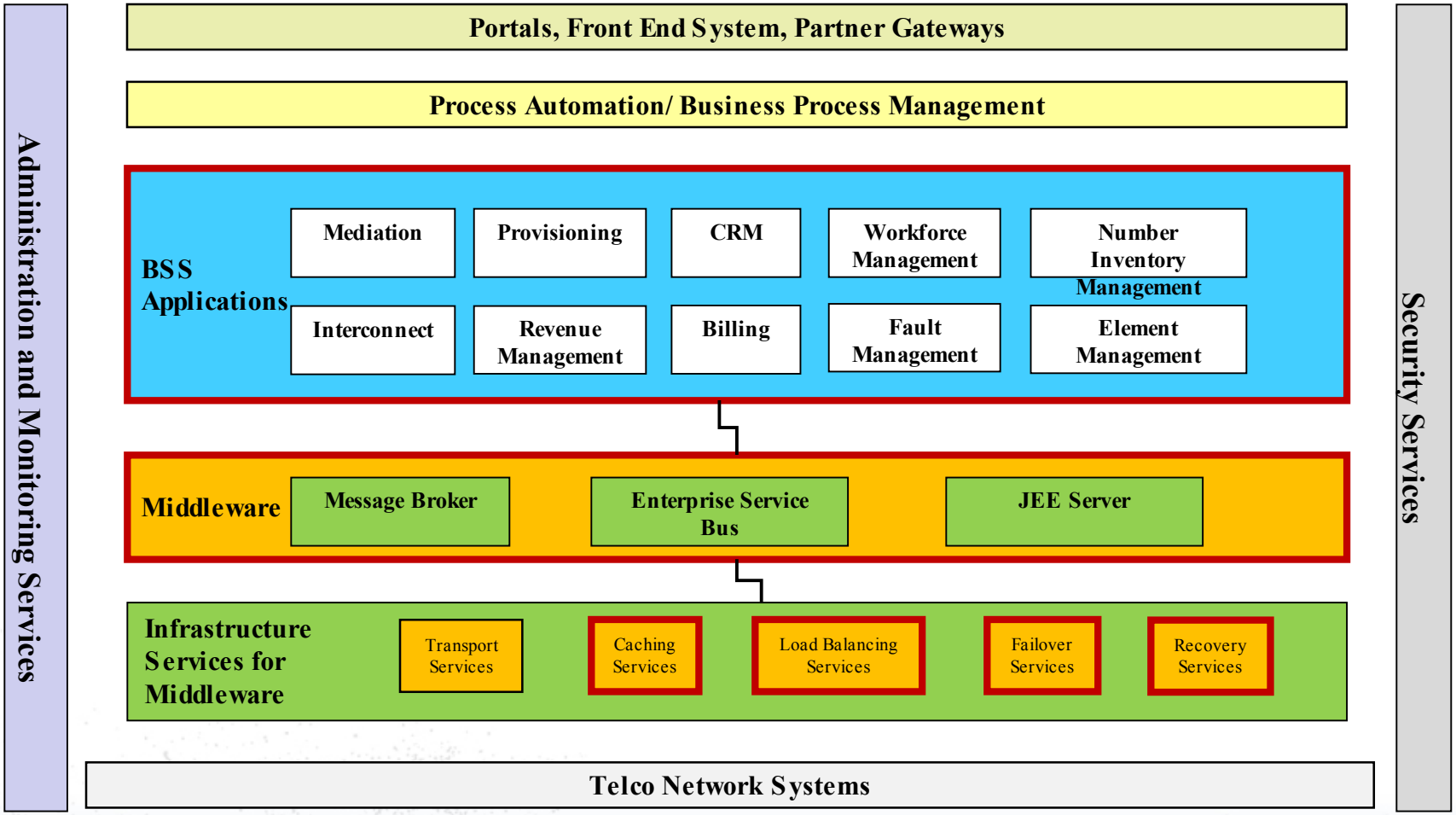
The Key Challenges

Constraints on the systems and on the connections



Use Cases and Solutions

Architecture
Wireless Telco BSS Integration



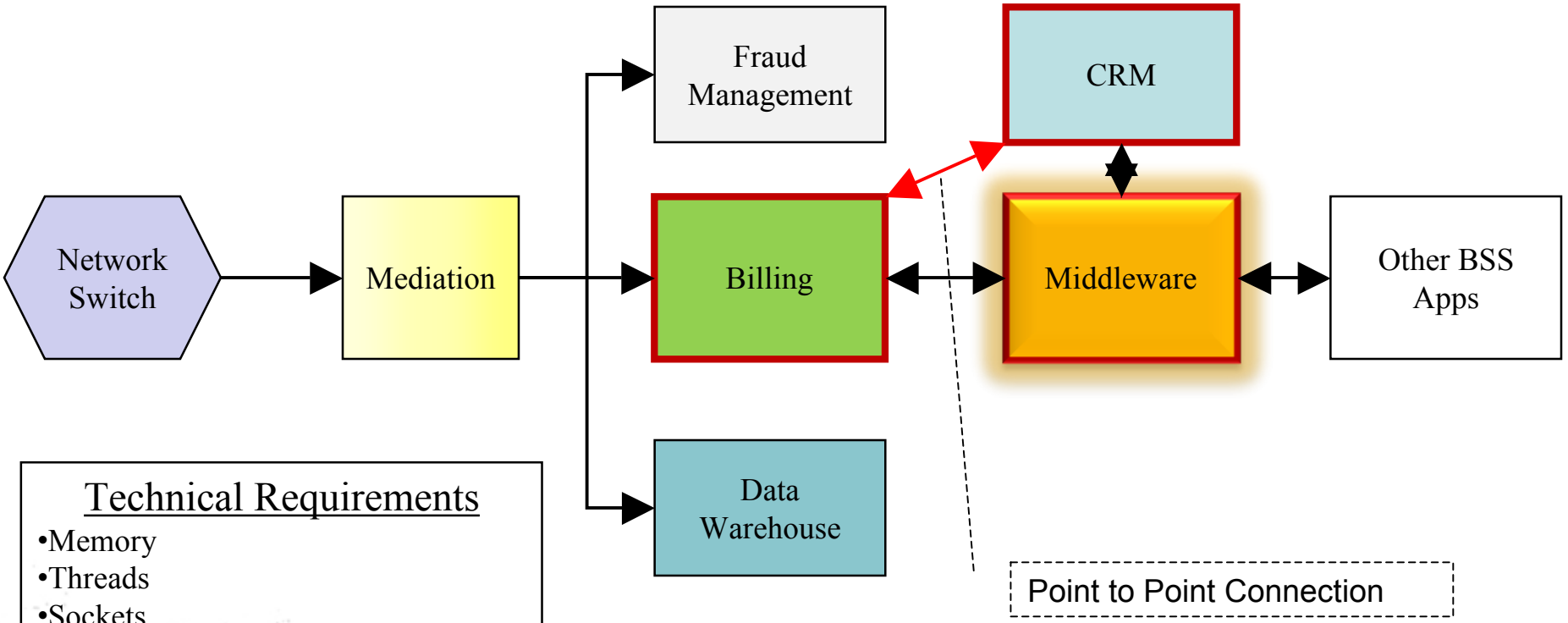
Telco scalability

Some requirements

- 75+ million customers
- Plan for Terabytes of CDRs and other messages per day
- Performance is critical to customer experience and retention
- CRM, Billing, Mediation, Middleware, Data warehouse

High volume use case 1

CRM to Billing Integration



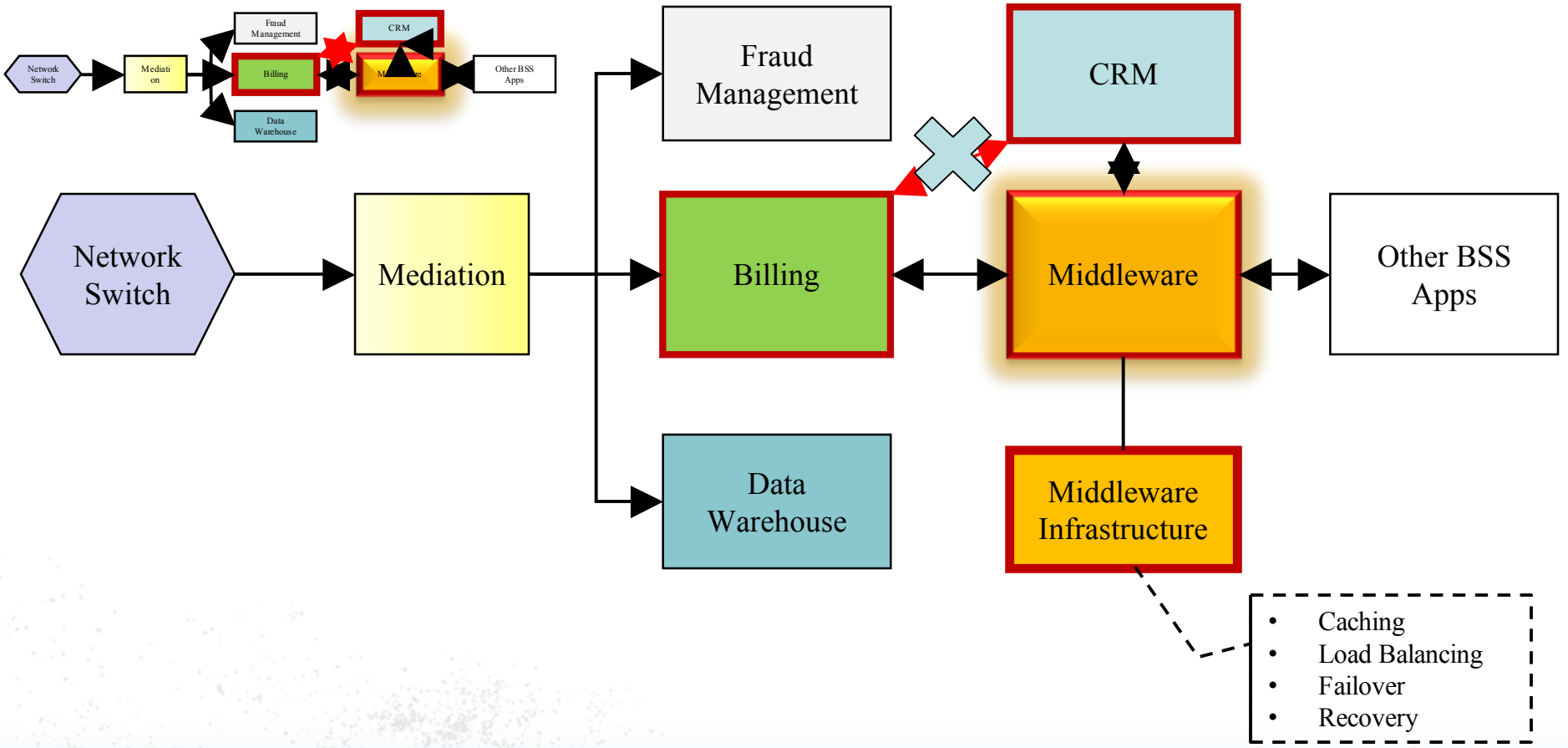
Technical Requirements

- Memory
- Threads
- Sockets
- Sender performance
- Receiver performance

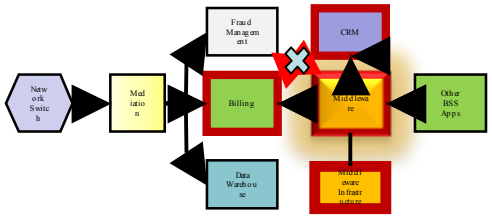
High volume use case 1

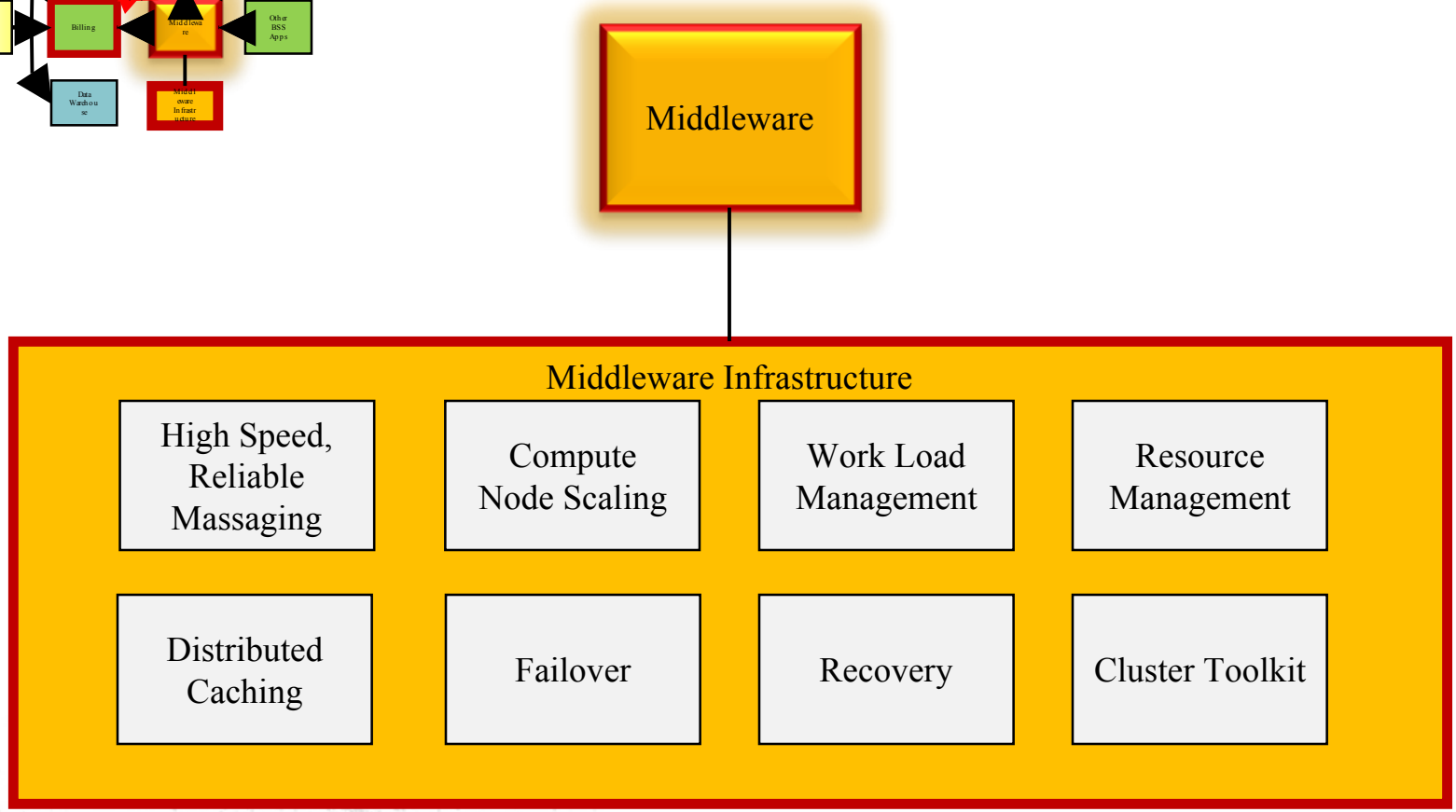
CRM to Billing Integration:

with Middleware Infrastructure



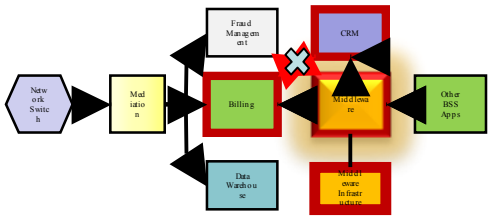
Middleware Infrastructure expanded





Middleware Infrastructure

Products



Middleware

- Any ESB
- Any JEE Server
- Any Message Broker

MRG Grid

Middleware Infrastructure

High Speed, Reliable Massaging

Compute Node Scaling

Work Load Management

Resource Management

Distributed Caching

Failover

Recovery

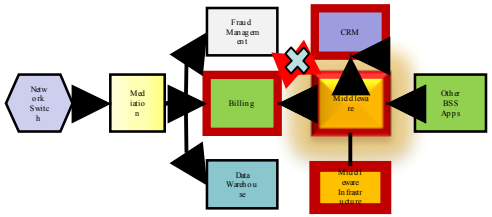
Cluster Toolkit

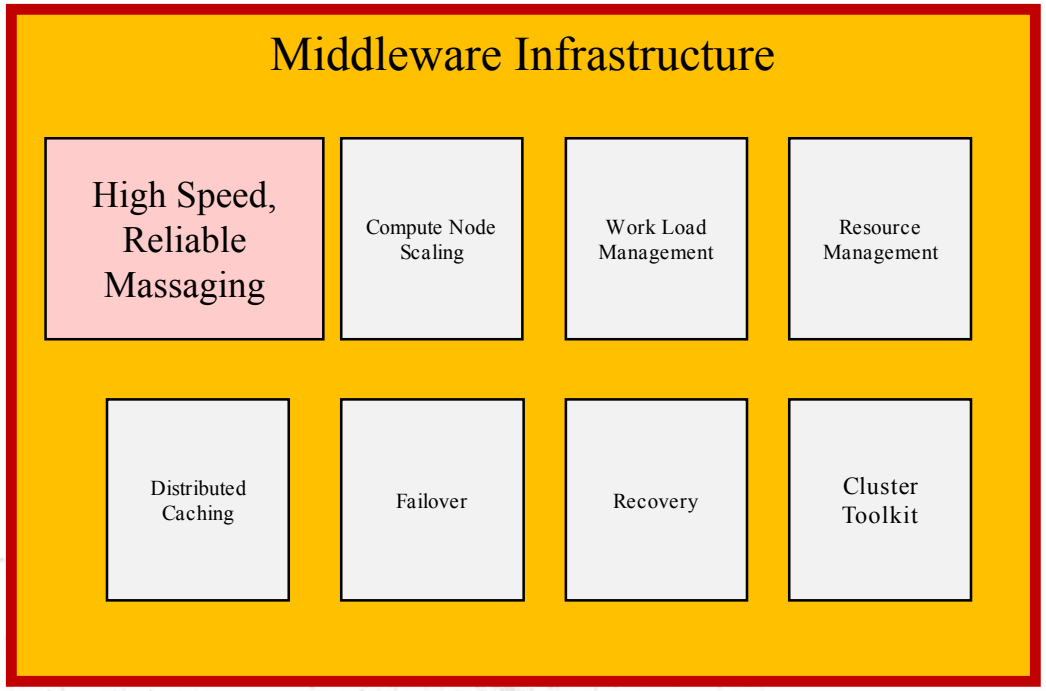
MRG Messaging

Infinispan

1 2 3 4 5 6 7 8

Middleware Infrastructure: Products
MRG Messaging

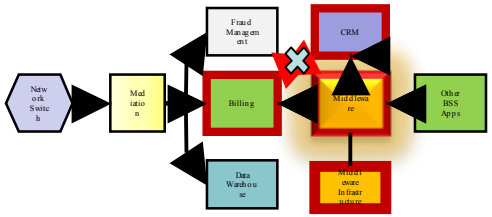


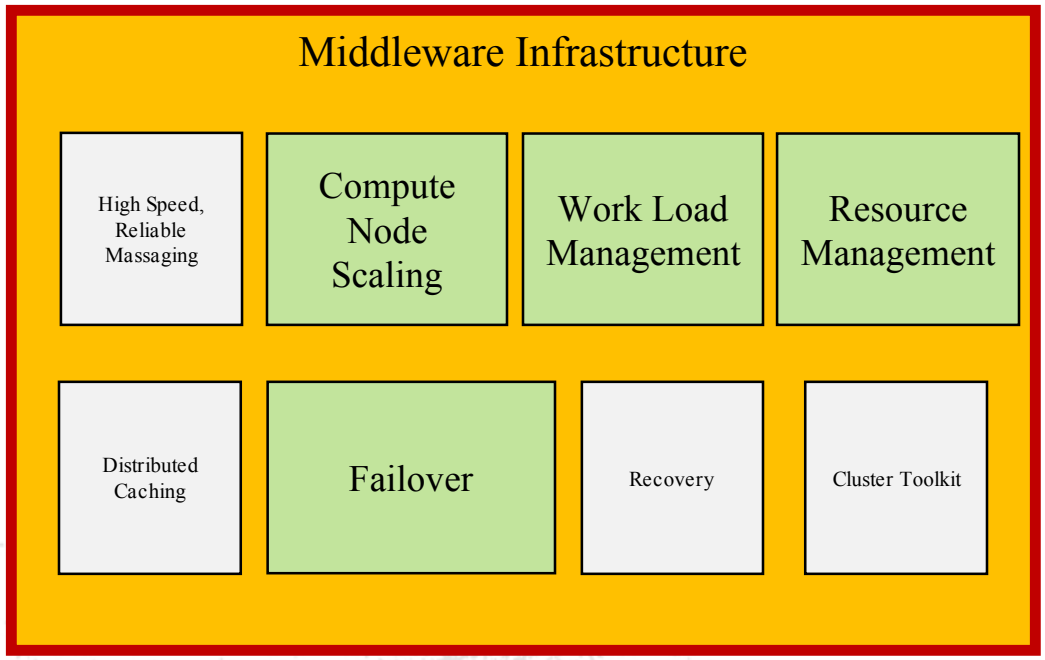


MRG Messaging

- AMQP support
- Native RDMA, Infiniband
- Can use MRG Realtime
- Large message support(> GB)
- Clustering and Failover
- High speed, journal based persistence
- Java and C++ brokers
- Based on Apache Qpid

Middleware Infrastructure: Products
MRG Grid

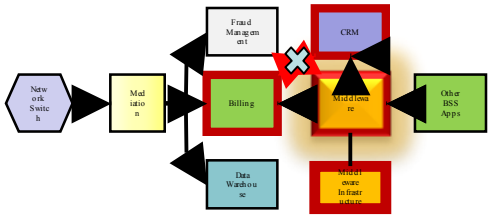


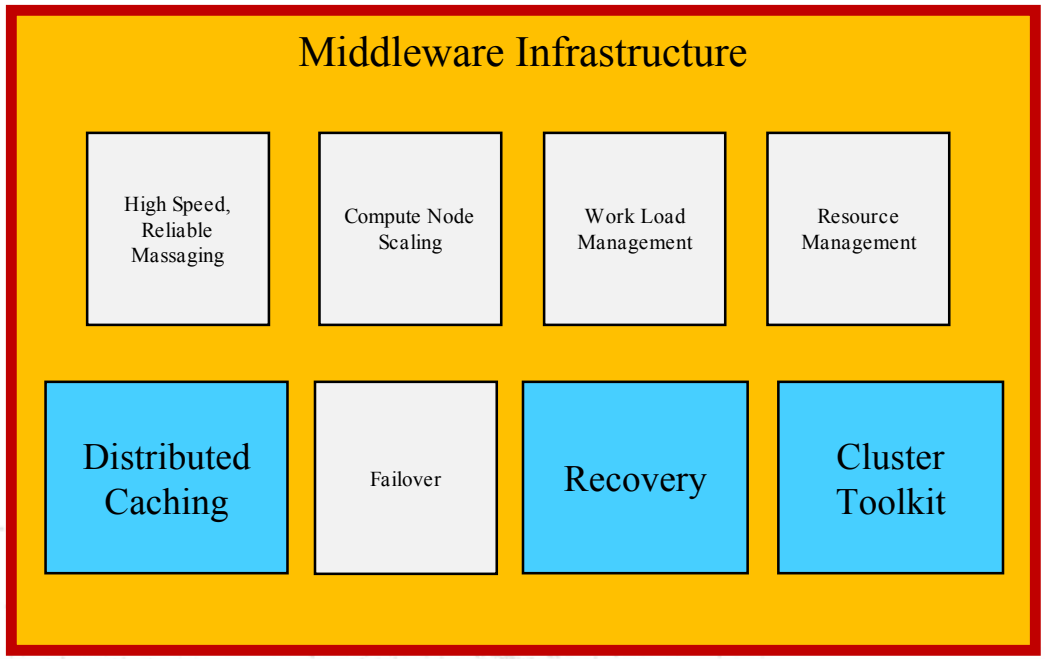


MRG Grid

- Scalable Grid Scheduler
- Resource variety: Desktop to Cloud schedulers
- Low latency results: Using MRG Messaging
- Dynamic provisioning
- High Availability
- Grid Federation
- Is based on the Condor Grid project

Middleware Infrastructure: Products
Infinispan



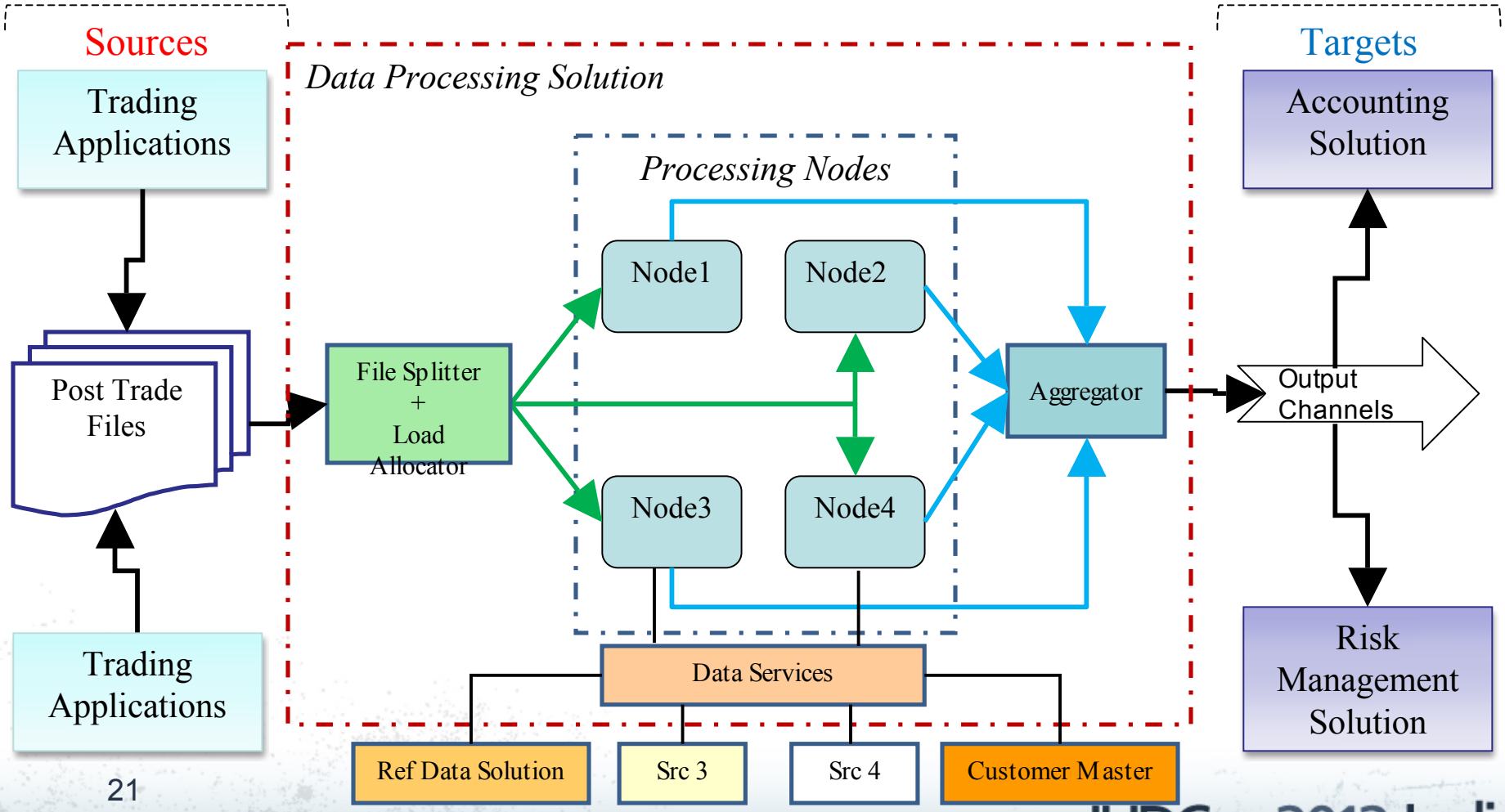


Infinispan

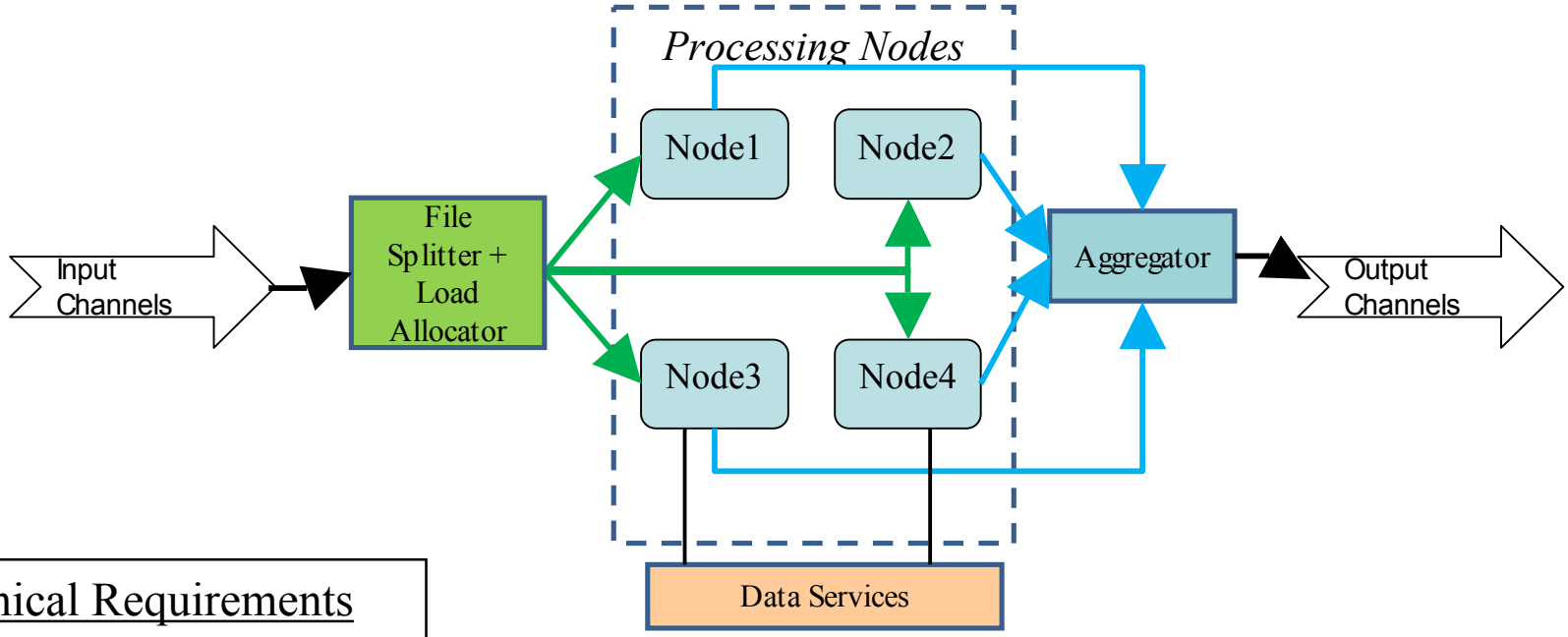
- In memory Data Grid
- Distributed cache
- Peer to peer communication between nodes
- Flexible persistence: JDBC, File, Amazon S3
- Map reduce: node local computing
- Implementation for performance

High volume use case 2

Post Trade Securities Processing



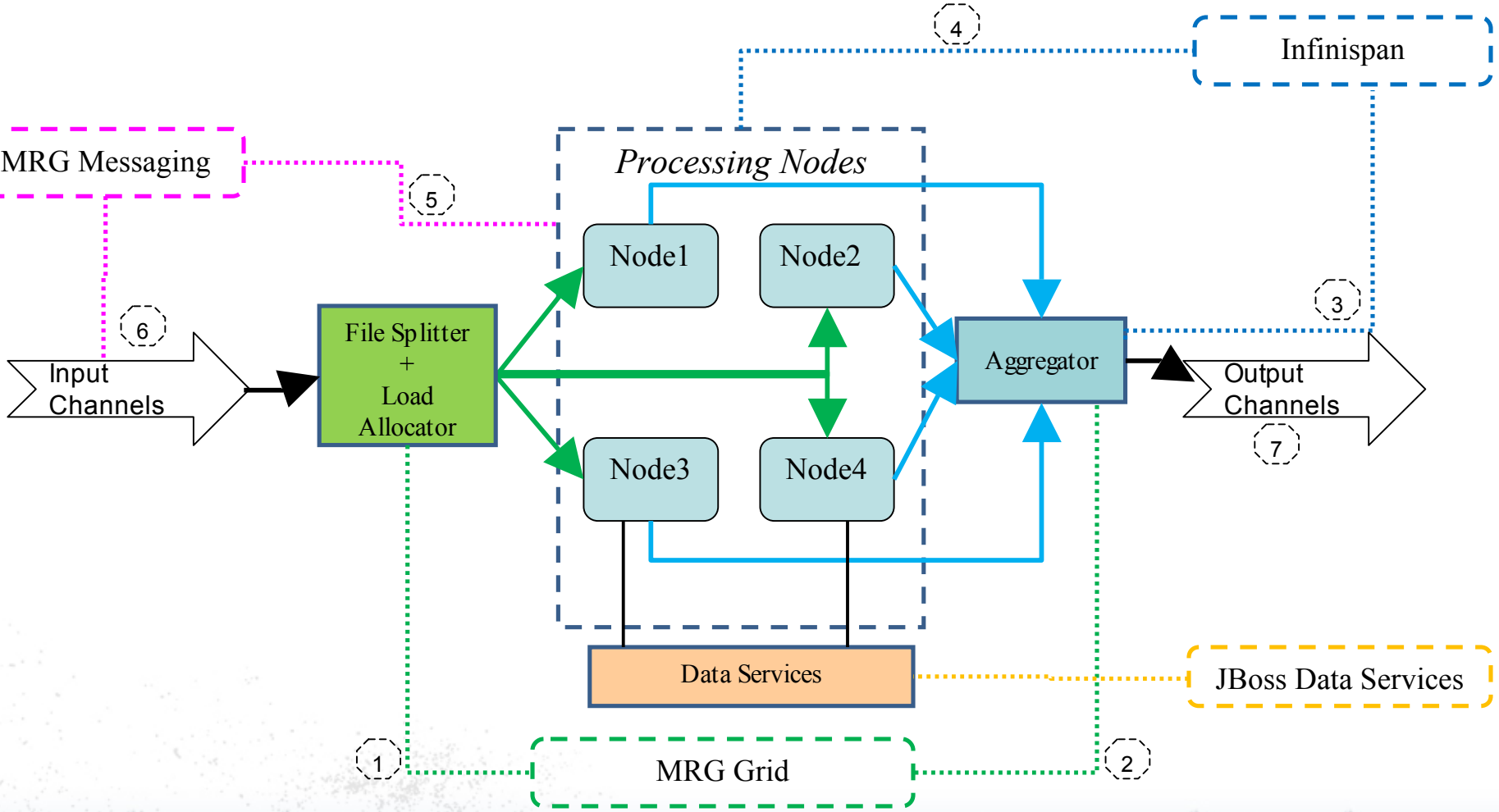
Post Trade Securities Processing
Technical Requirements



Technical Requirements

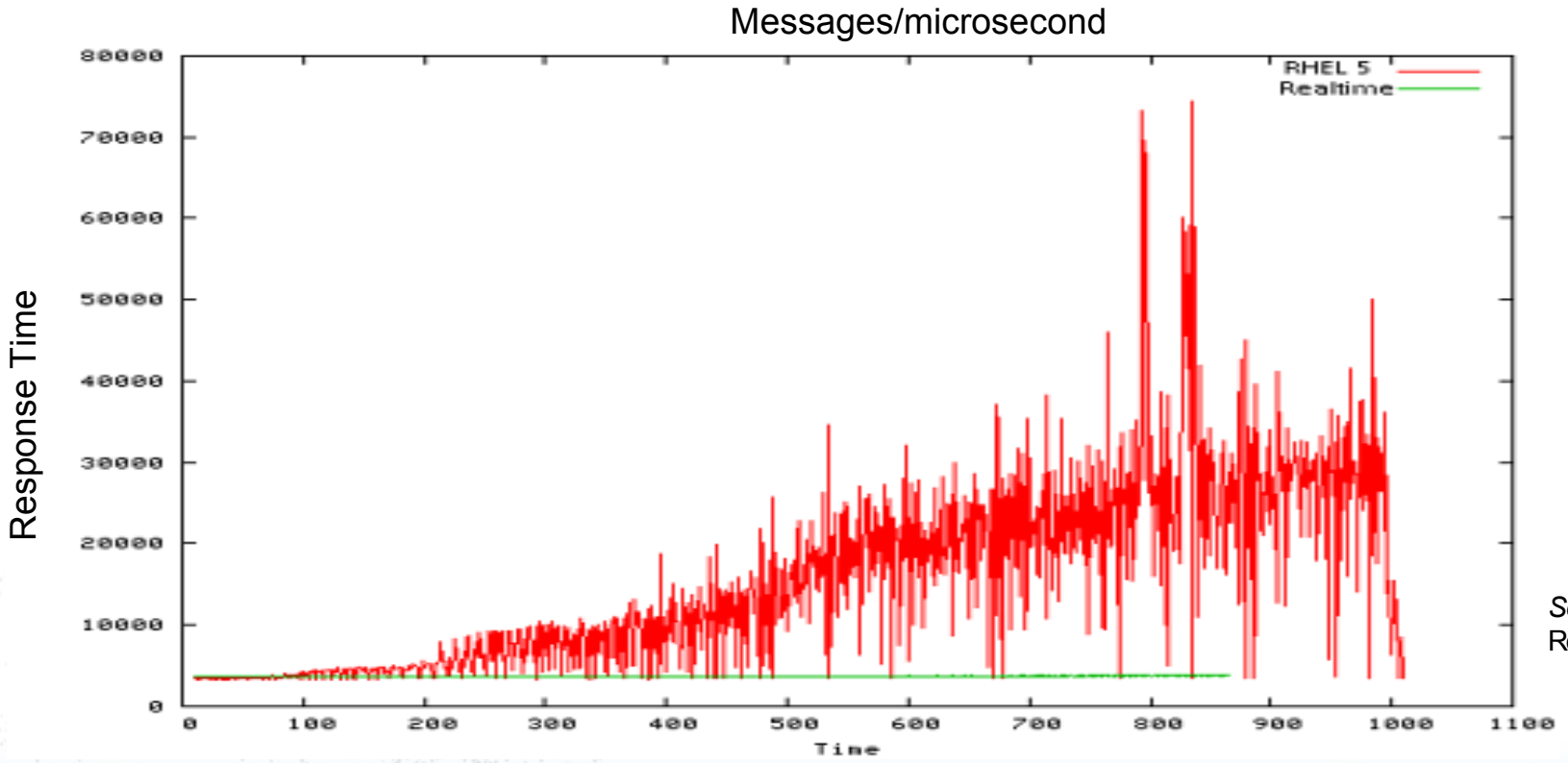
- File Streaming
- Multiple Data views, Data sources
- Data Aggregation
- Reliable Messaging

Post Trade Securities Processing
Products



MRG Realtime

- Consistent, predictable response
- Websphere Realtime: RTSJ



Source:
Red Hat

A Recap of the Solution

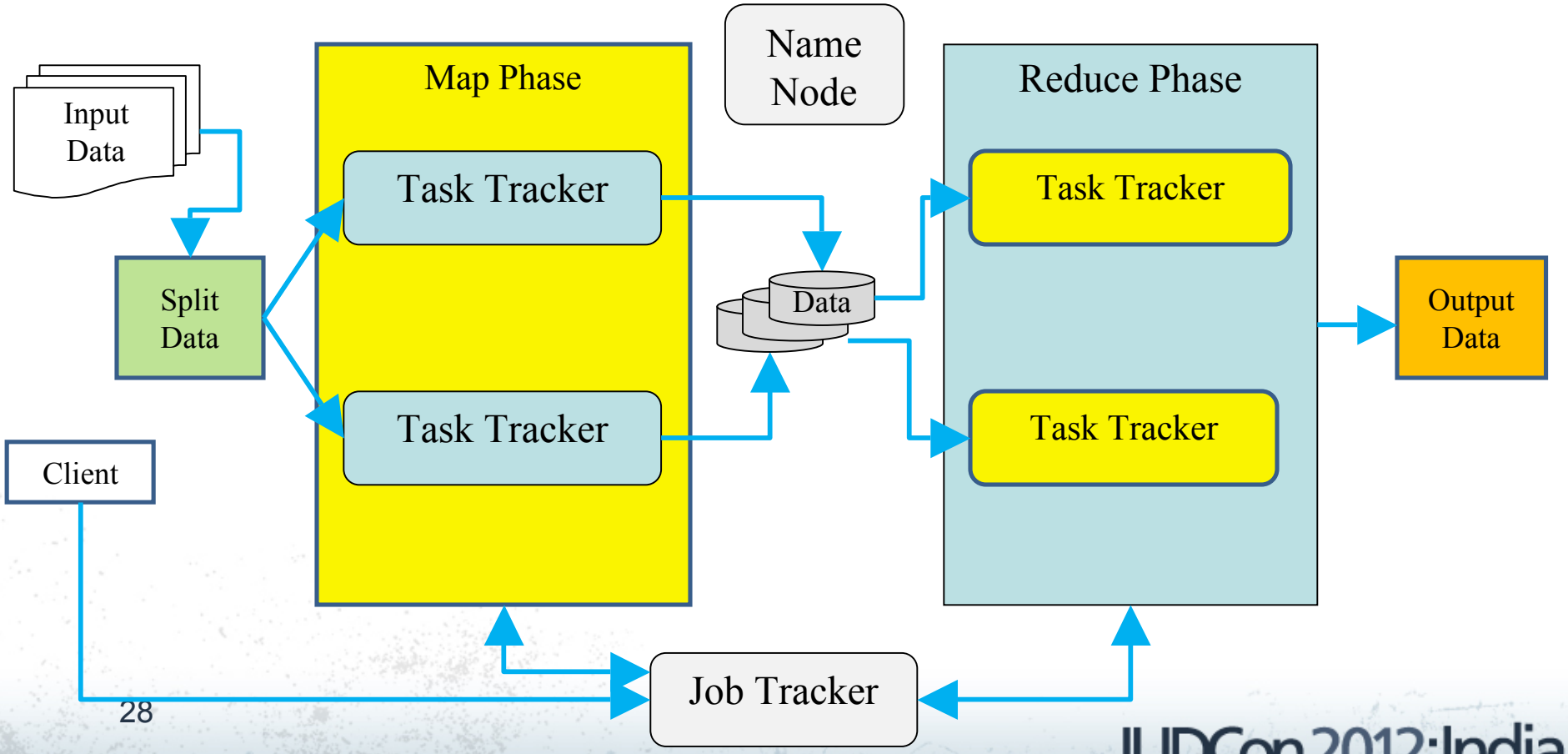
Challenges, Solutions and Products

	Challenge	Solution	Products
1	Small data elements, high volume	Distribution, load balancing and partitioning	MRG Messaging, MRG Grid
2	Large data elements	File splitting, distribution, in place processing	MRG Messaging, MRG Grid
3	Data views, Many data sources	Data Services	JBoss Data Services
3	Predictability	Real time kernels, real time JVMs	MRG Realtime, RTSJ
4	Availability	Load balancing, clustering, failover	MRG Grid, Infinispan
5	Reliability	File based caches, DB persistence	MRG Messaging, Grid, Infinispan
6	Scalability	Compute grids, Data grids, Asynchronous messaging	MRG Messaging, Grid, Infinispan

Solution Alternatives

The Map Reduce method

- Split data, process in parallel, aggregate results



GT3 and Condor

- Globus Toolkit
  - Open source toolkit for Compute Grids
  - Architecture, Service Model and Implementation
  - Job Tracking, Management, Monitoring, Resource Management
  - Data Management: Movement, Location Registry
- Condor
  - Grid Framework from University of Wisconsin
  - Compute Node Scaling
  - Job Scheduling
  - Idle time utilization

Commercial Tools

- IBM
  - IBM Cloudburst
  - Websphere Virtual Enterprise
  - Websphere Realtime
- Oracle
  - Oracle Exalogic
  - Oracle Coherence
  - Oracle Grid Engine(Sun Grid Engine)
- Terracotta
  - Quartz
  - Big Memory

The Open Source difference

The Advantages

- Smaller adoption steps to reduce risk
- Flexible Cost Model
  - Subscription based pricing
  - Incident based pricing
- Cloud alignment
  - Elastic pricing model
  - Cloud Software platforms use open source
- Innovation from wider community
- Custom enhancements

Conclusion

Key Points Discussed

- Large Scale Integration
- Impact of Big Data on Integration
- Use cases
  - Telecom
  - Securities
- Solutions
  - MRG
  - Infinispan
  - Data Services
- Open source differentiators

Questions