

High-Throughput Processing

With Streaming-OODT

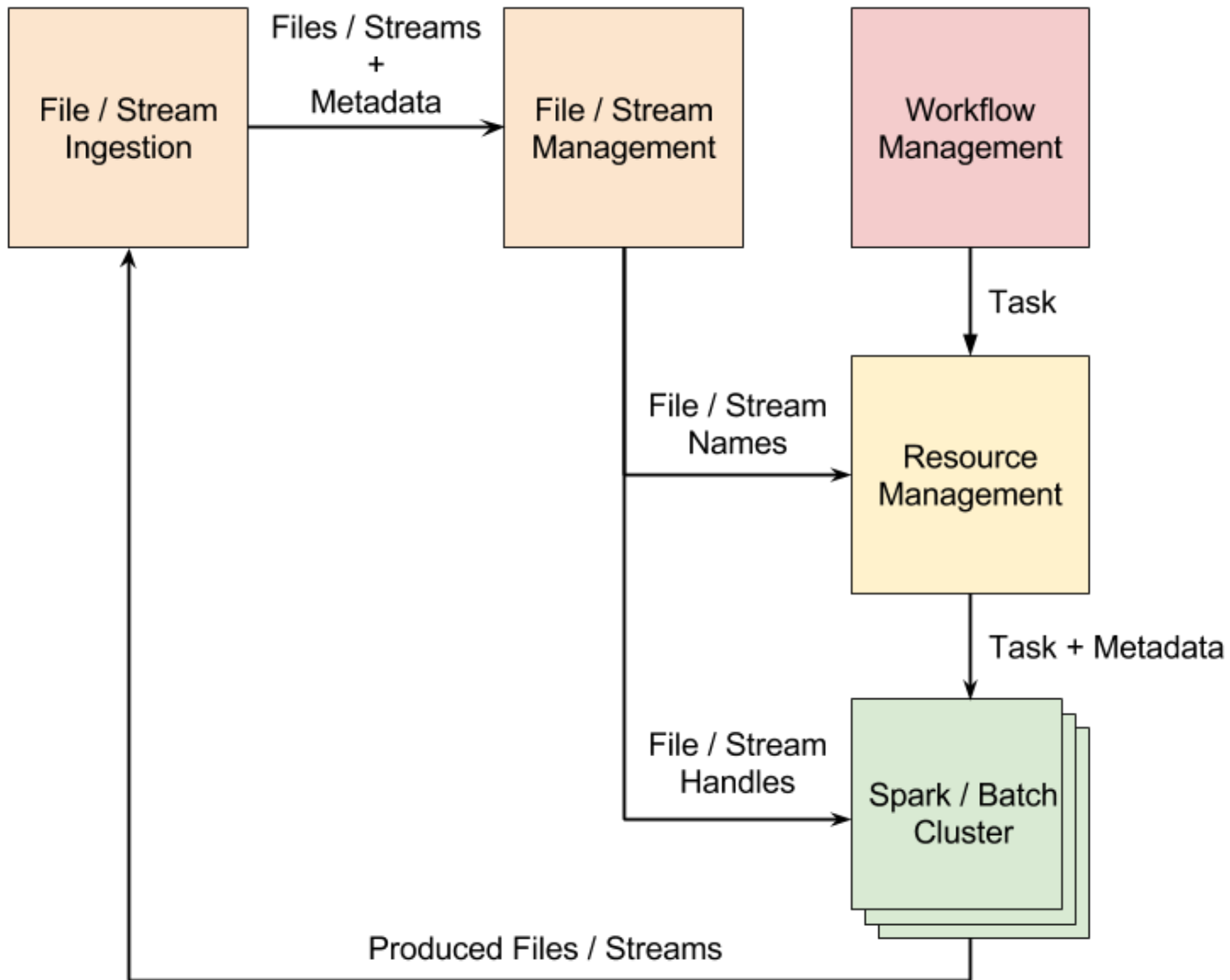
Michael Starch – NASA Jet Propulsion Laboratory

Agenda

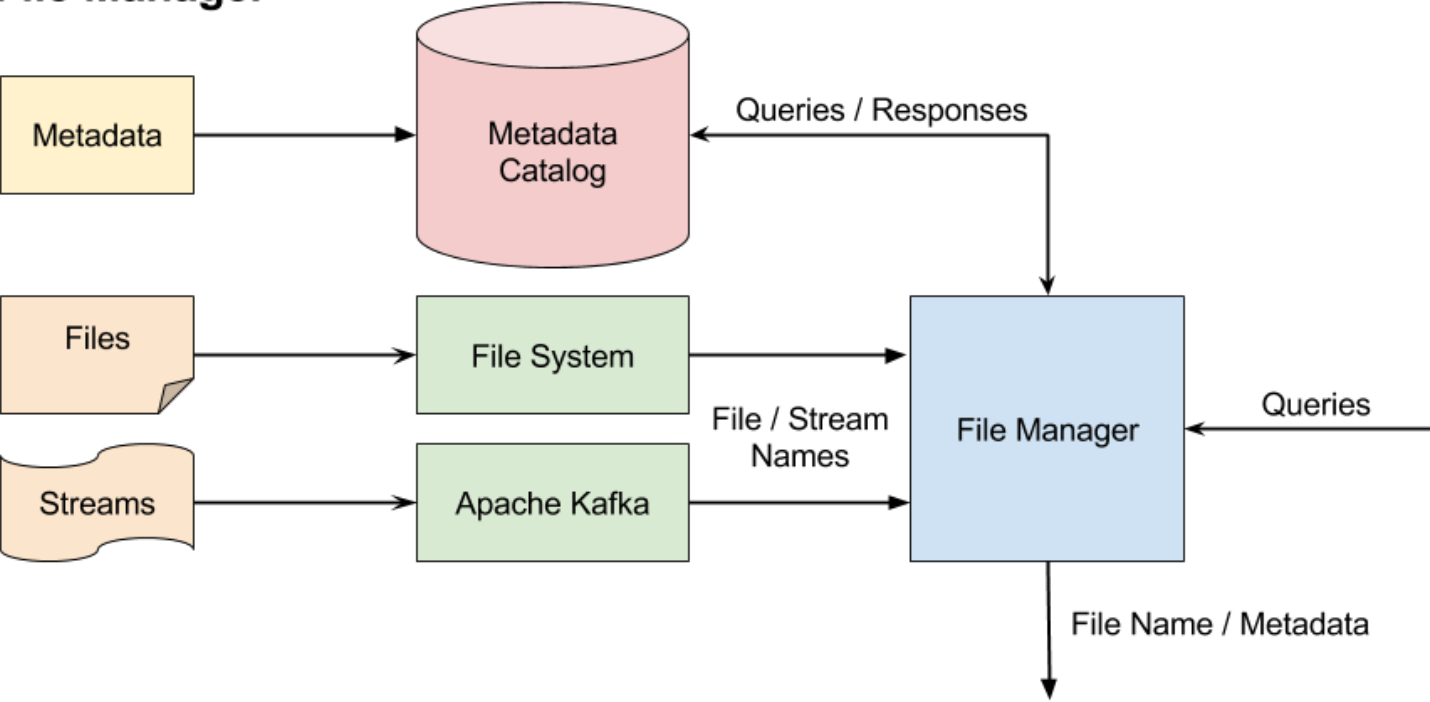
- OODT and Streaming-OODT
- Supporting Technologies
- Setup
- Performance
- Recommendations
- Questions

OODT and Streaming-OODT

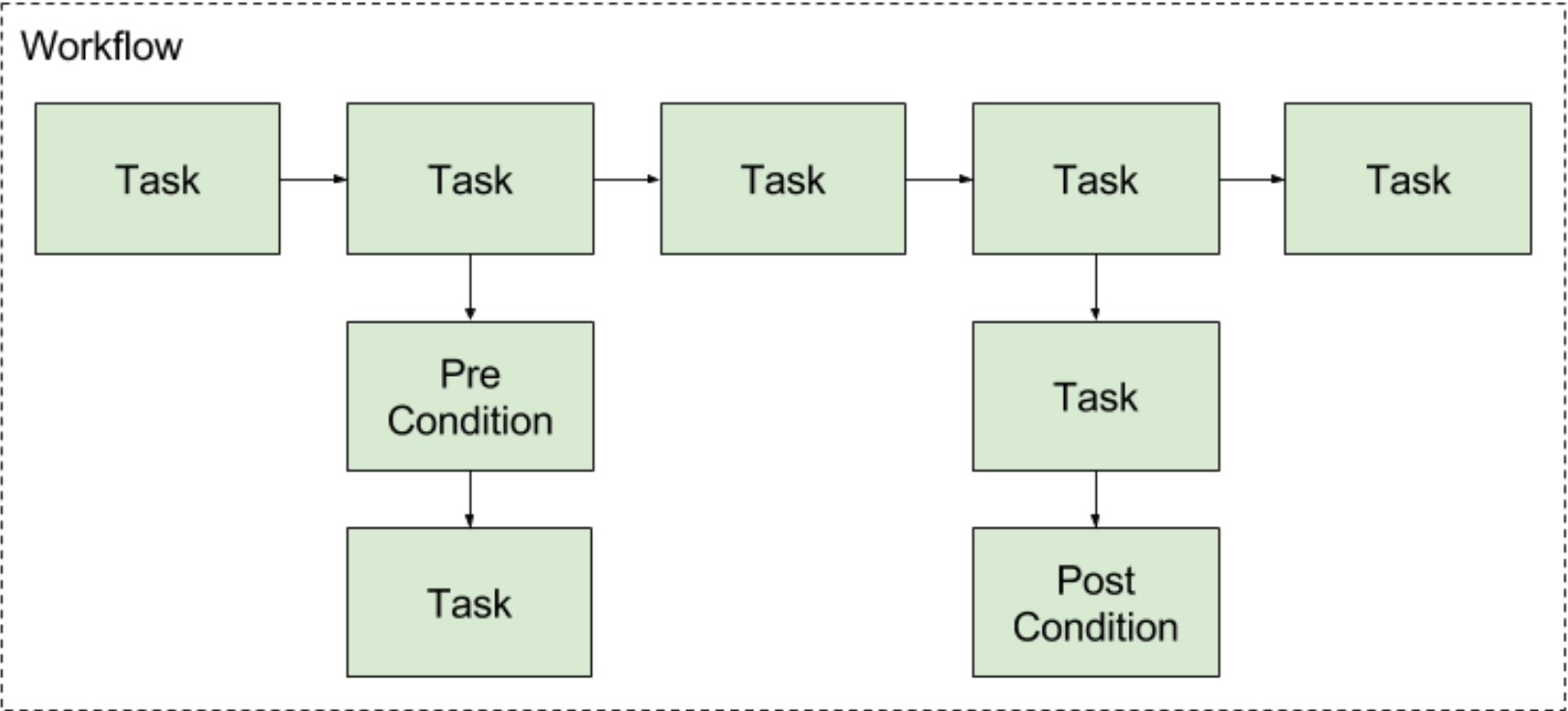
OODT and Streaming OODT



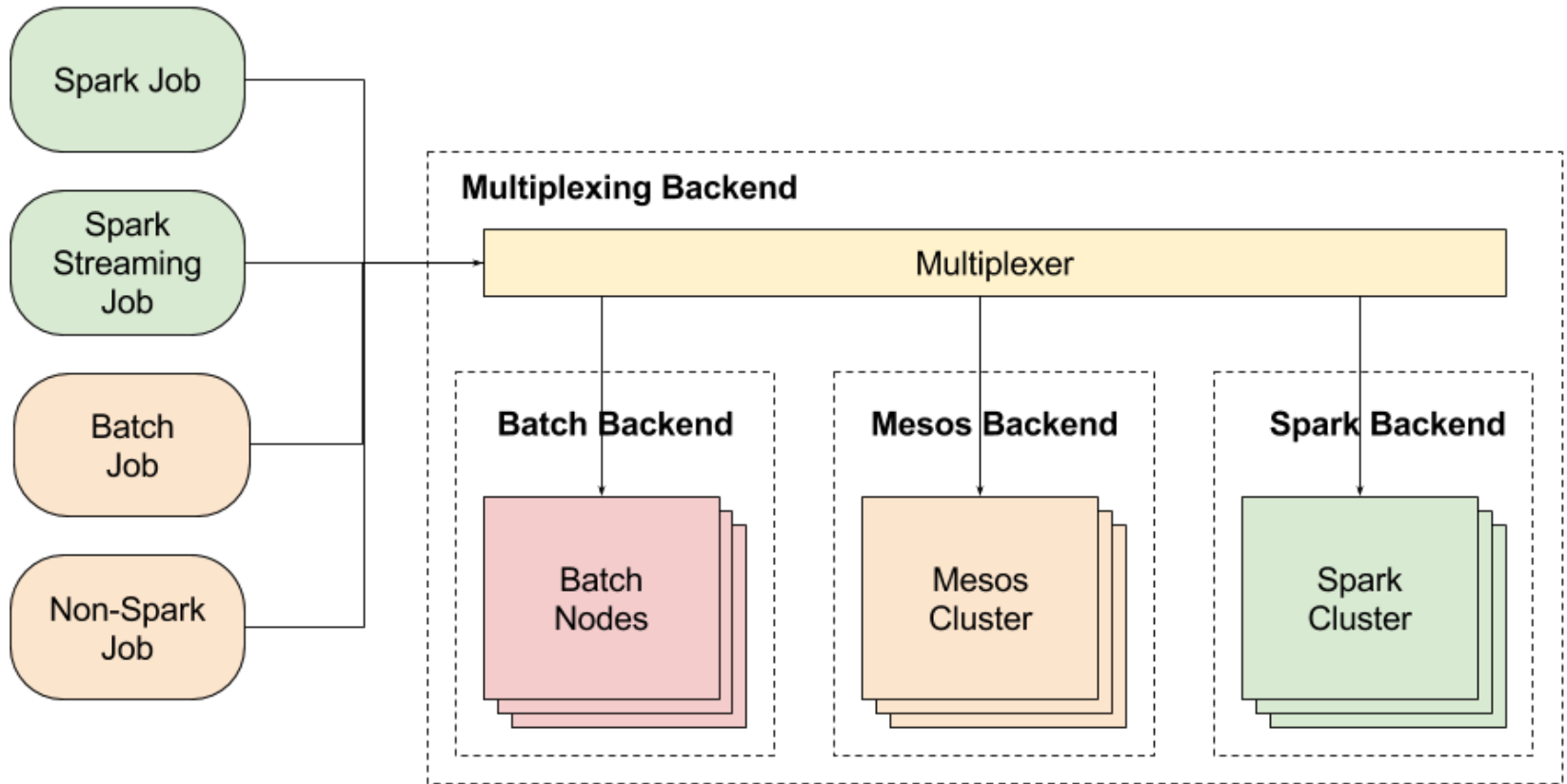
File Manager



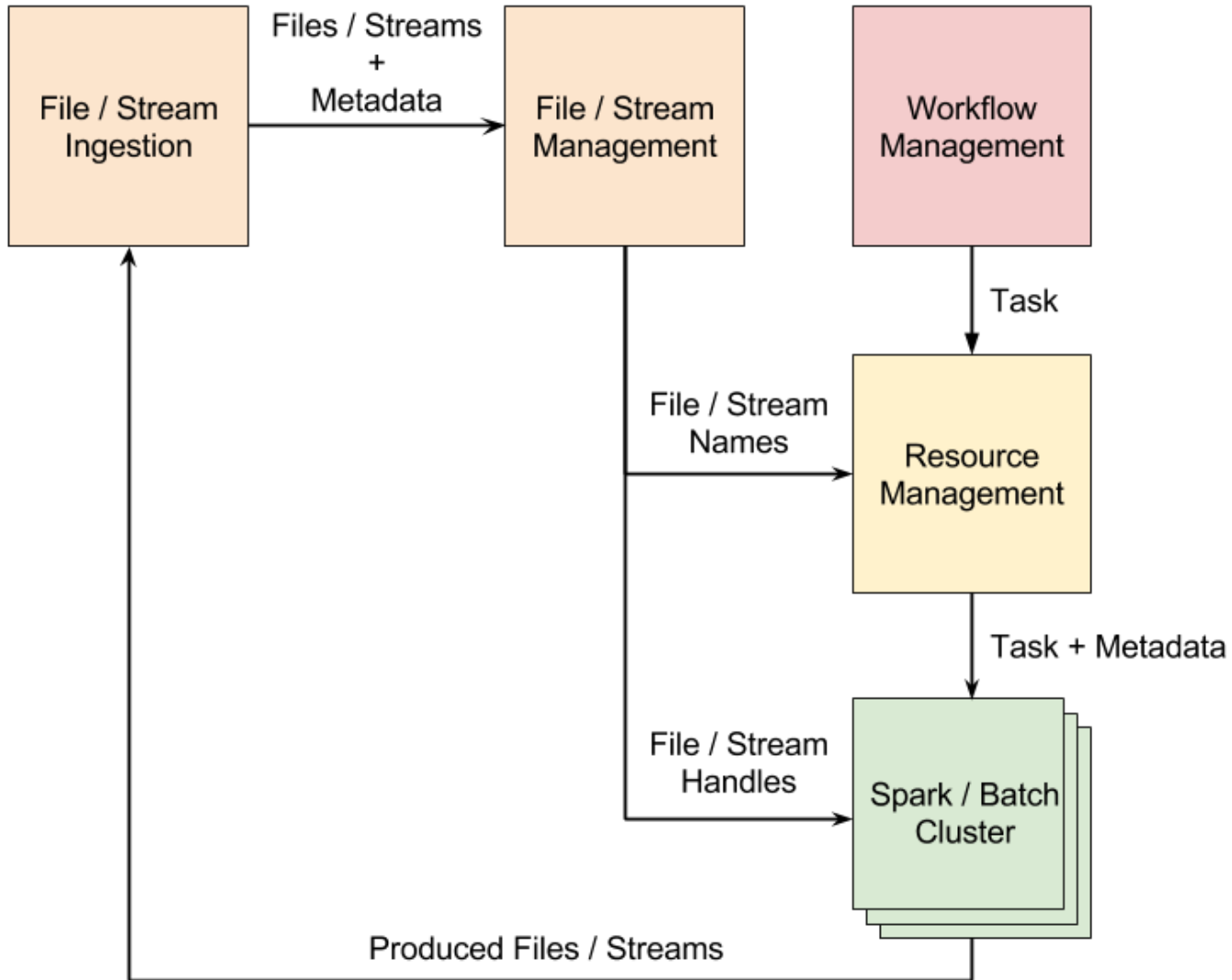
Workflow Manager



Resource Management

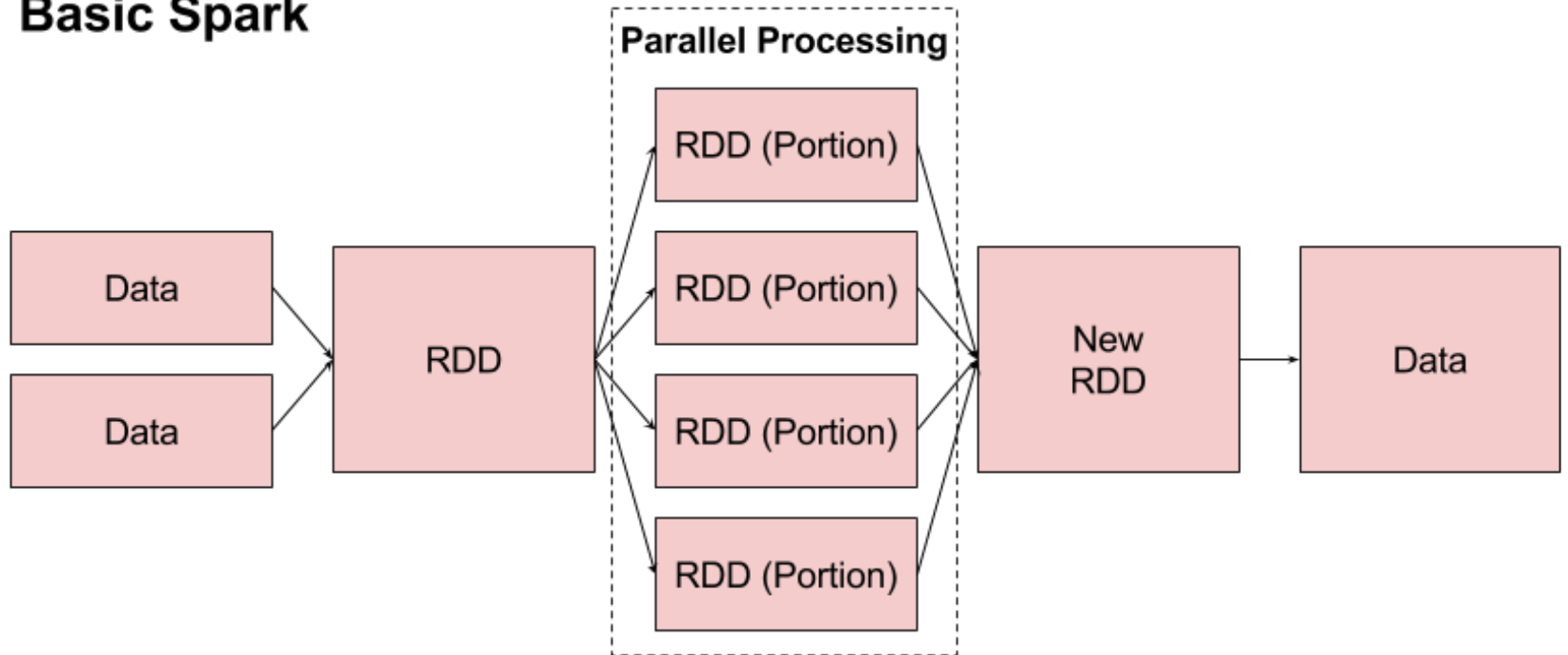


OODT and Streaming OODT

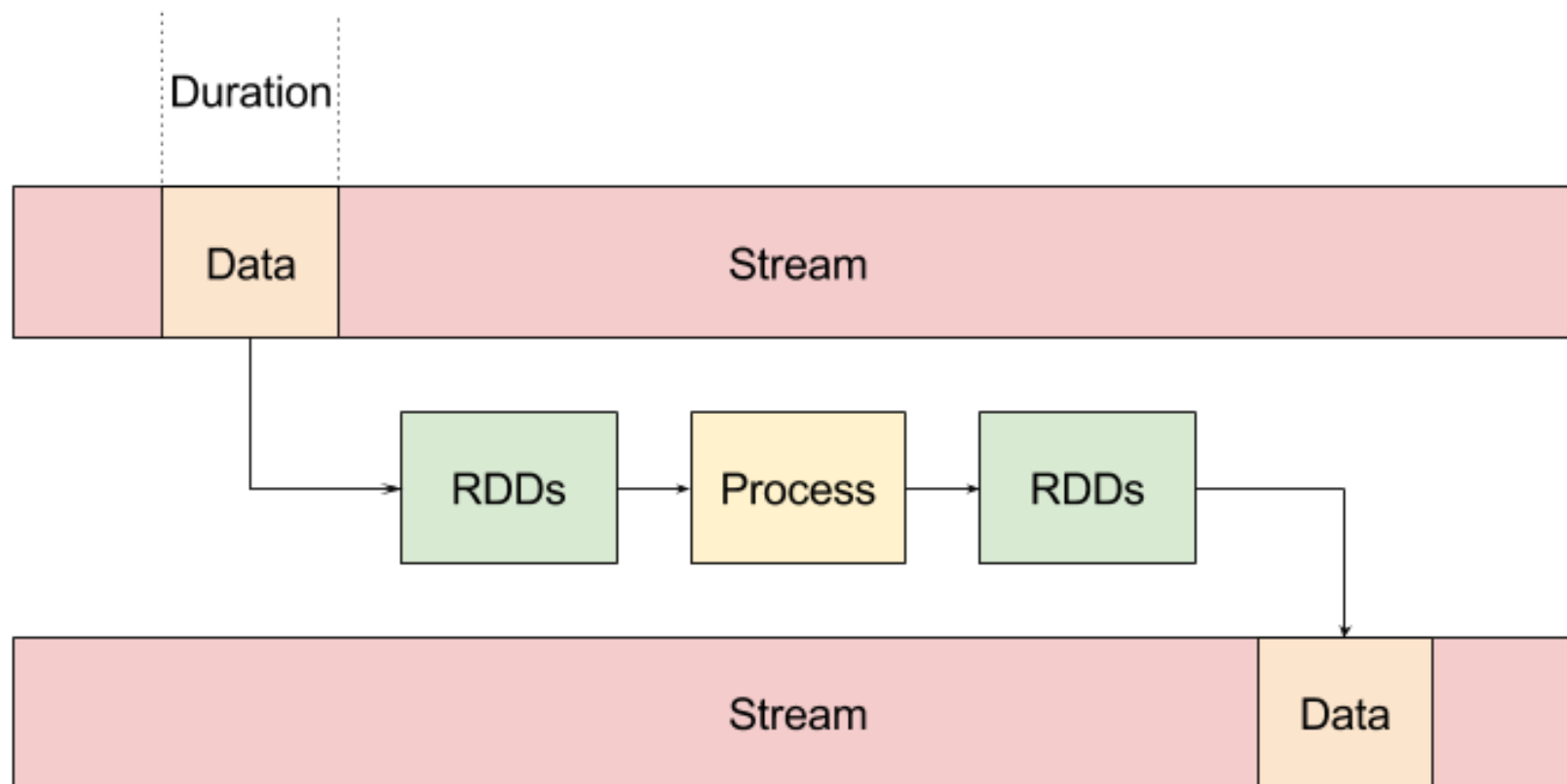


Supporting Technologies

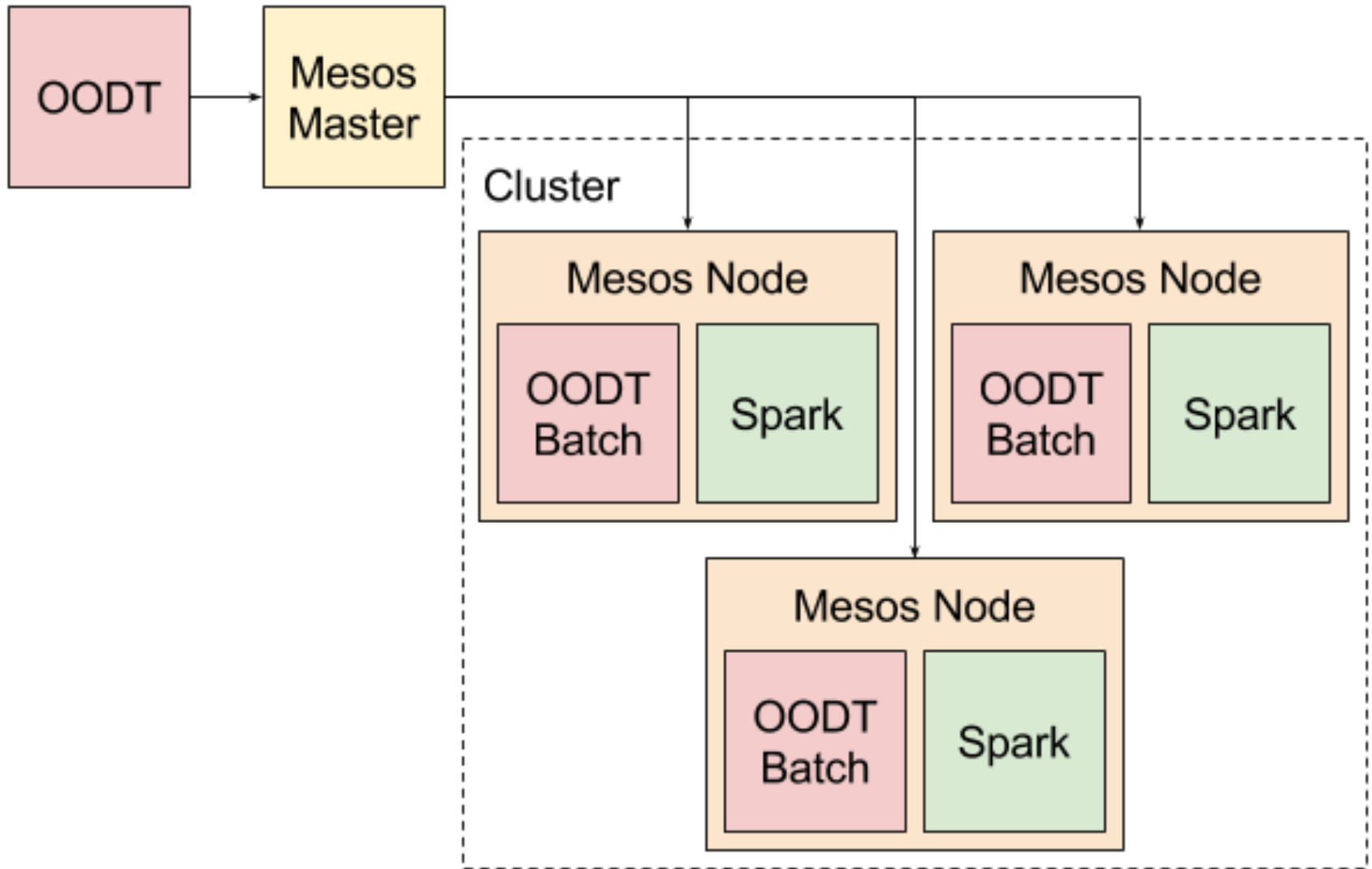
Basic Spark



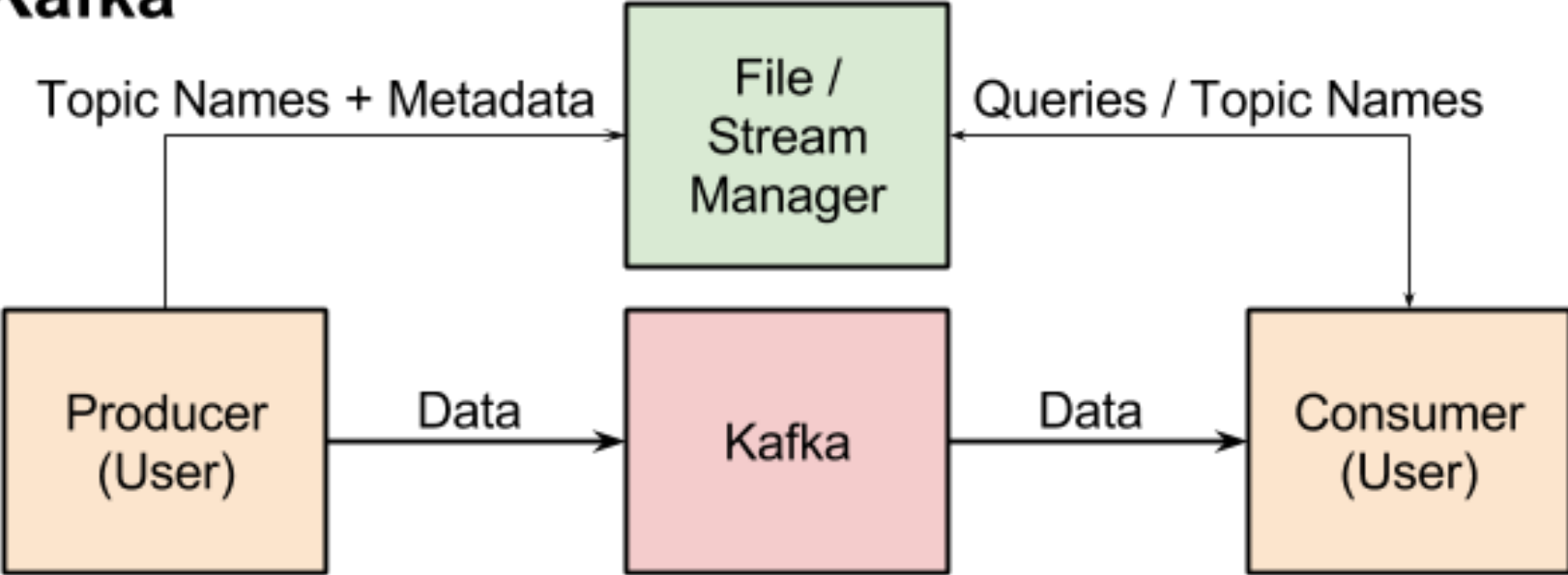
Spark Streaming



Mesos



Kafka

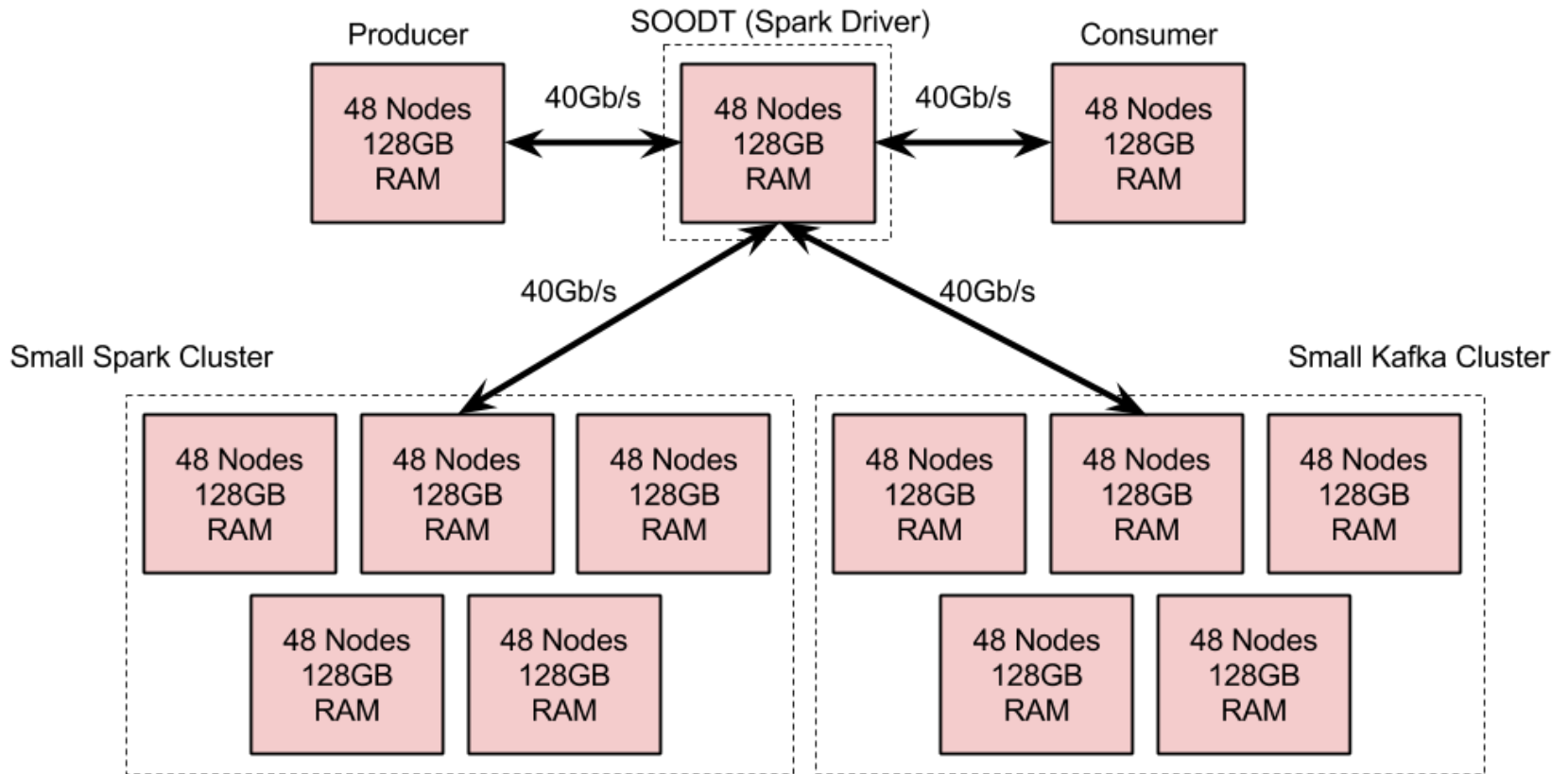


Setup

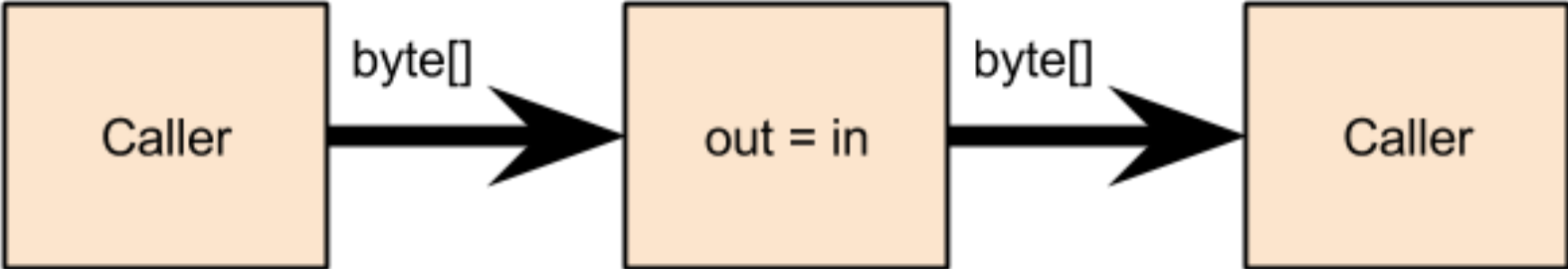
Task Goal



Environment - HPC Cluster



Computation



Performance

System Performance



Recommendations

- Understand the underlying technology
 - OODT handles data-system functions
 - Does not optimize underlying technology
- Understand your problem
 - Different approaches yield better results
 - No all problems are solved with one hammer

Where can I get the code?

It's Open Source! Jump on in!

<https://oodt.apache.org/>

<https://github.com/apache/oodt>

Acknowledgments

NASA Jet Propulsion Laboratory

Research & Technology Development

“Archiving, Processing and Dissemination for the Big Data Era”

Apache Software Foundation

Apache Spark, Apache OODT

Texas Advanced Computing Center

Wrangler

Avez-vous des questions?

Haben Sie Fragen?

Questions?

你
有
沒
有
問
題
?

¿Tienen preguntas?