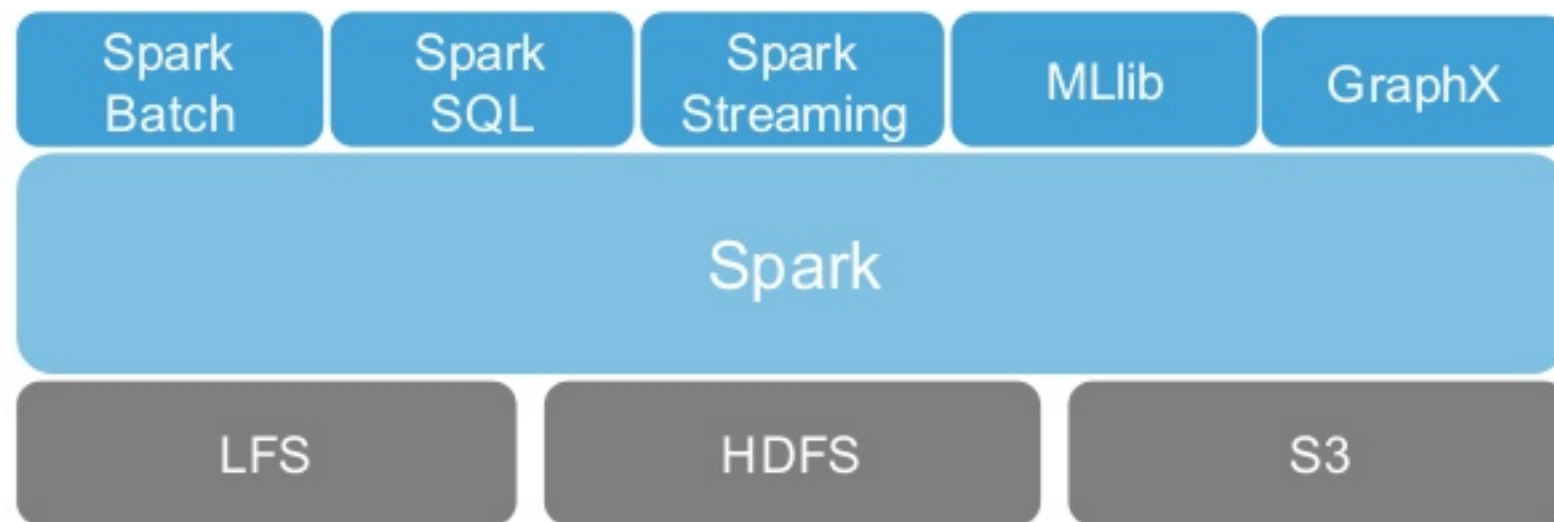




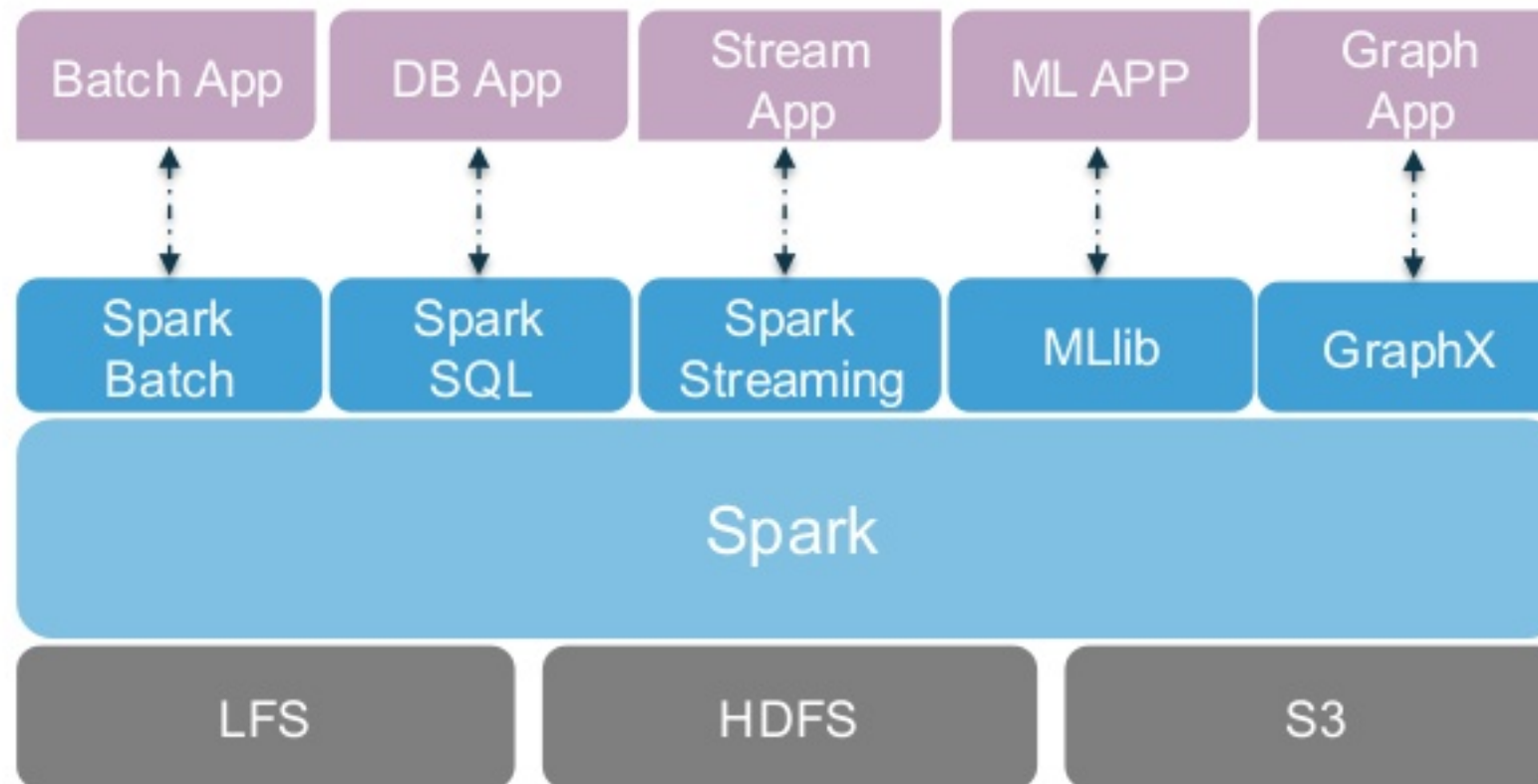
## INTEROPERATING A ZOO OF DATA PROCESSING PLATFORMS

YASSER IDRIS, SEBASTIAN KRUSE

# Spark Stack

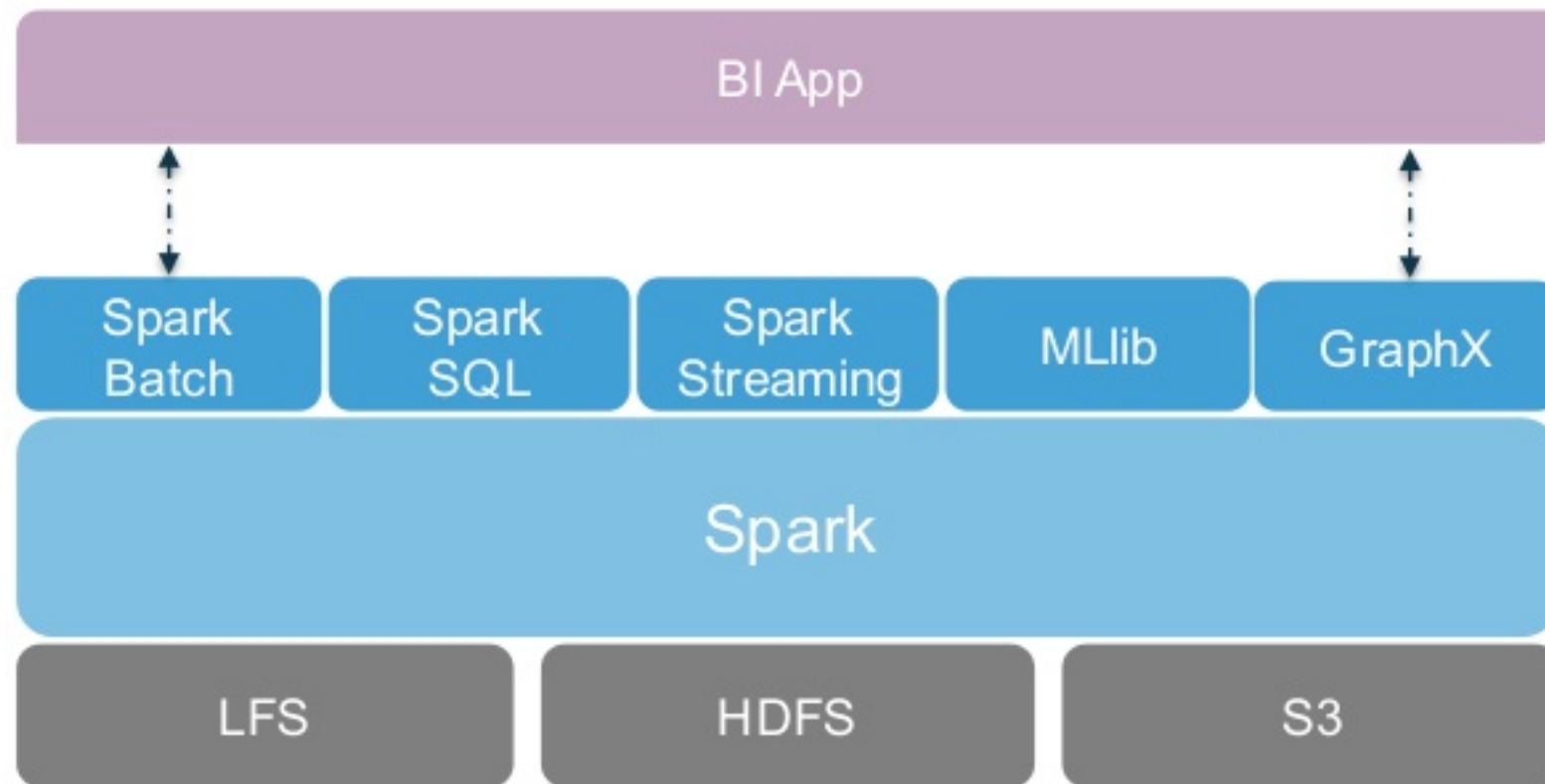


# Apps On Spark



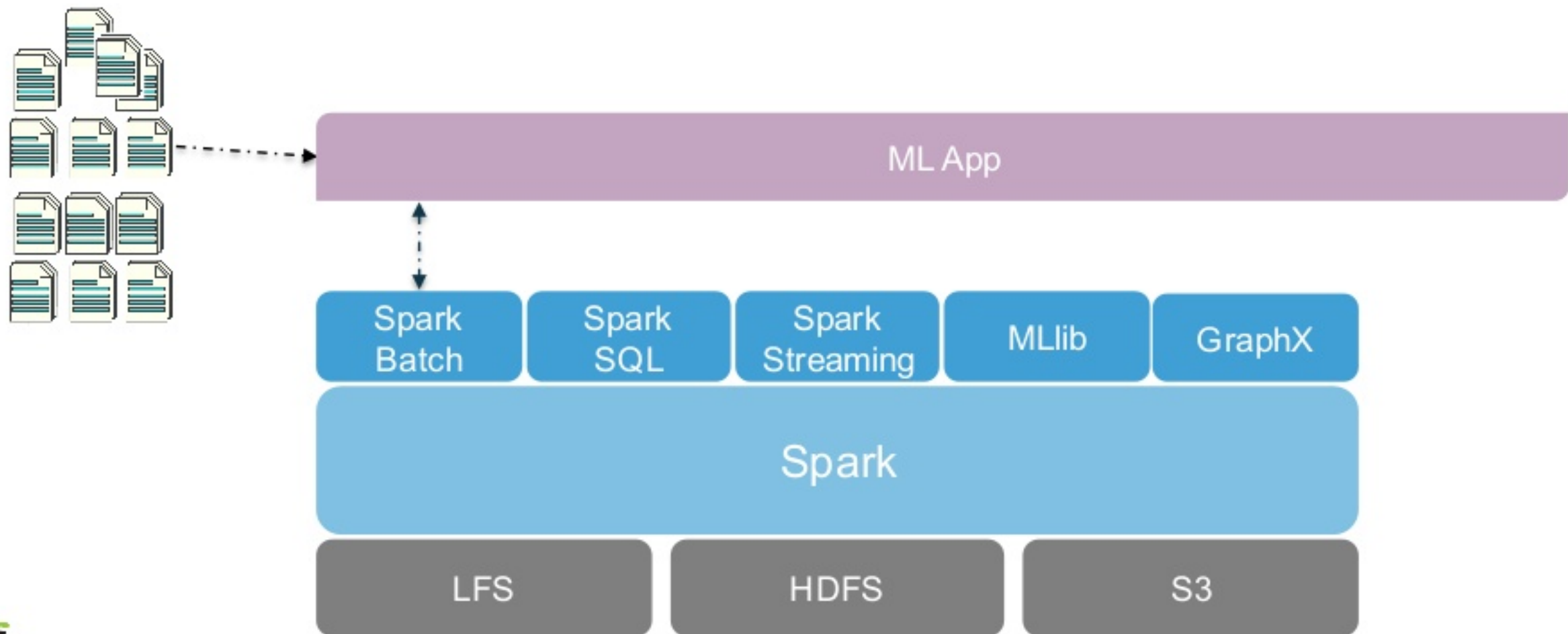
# Beyond a Single Platform

## Complex Apps



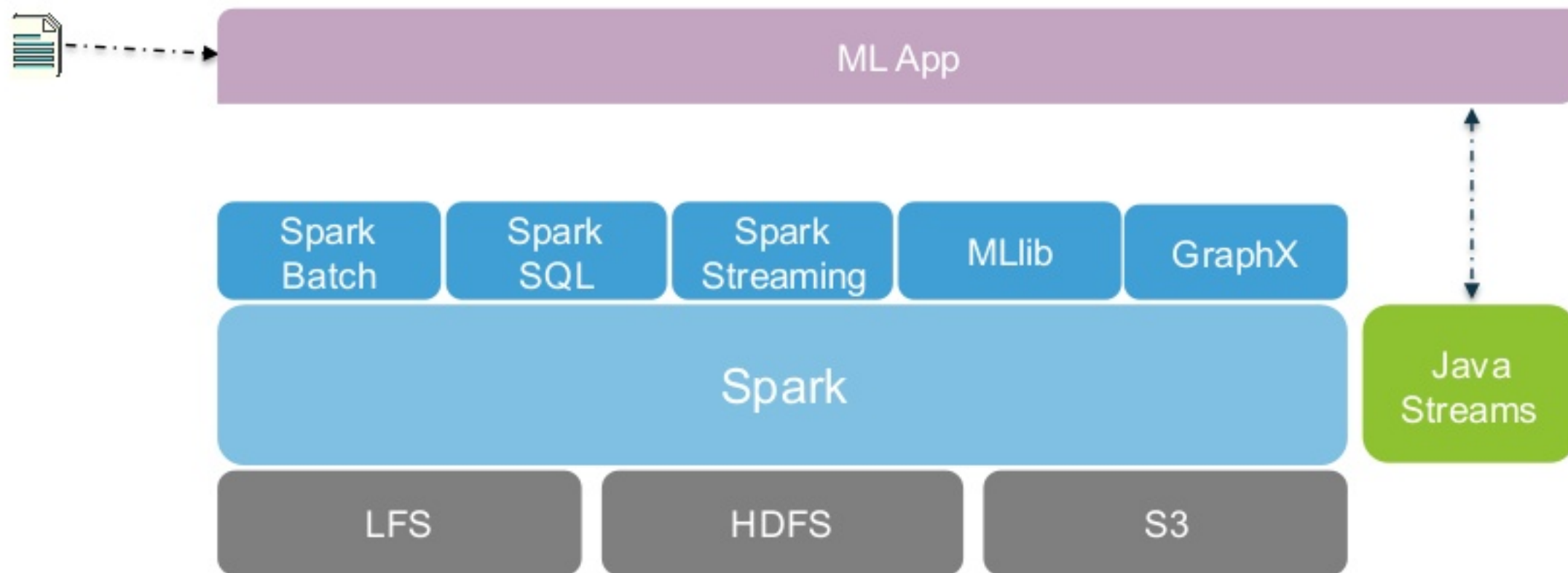
# Beyond a Single Platform

## Big vs. Small Data



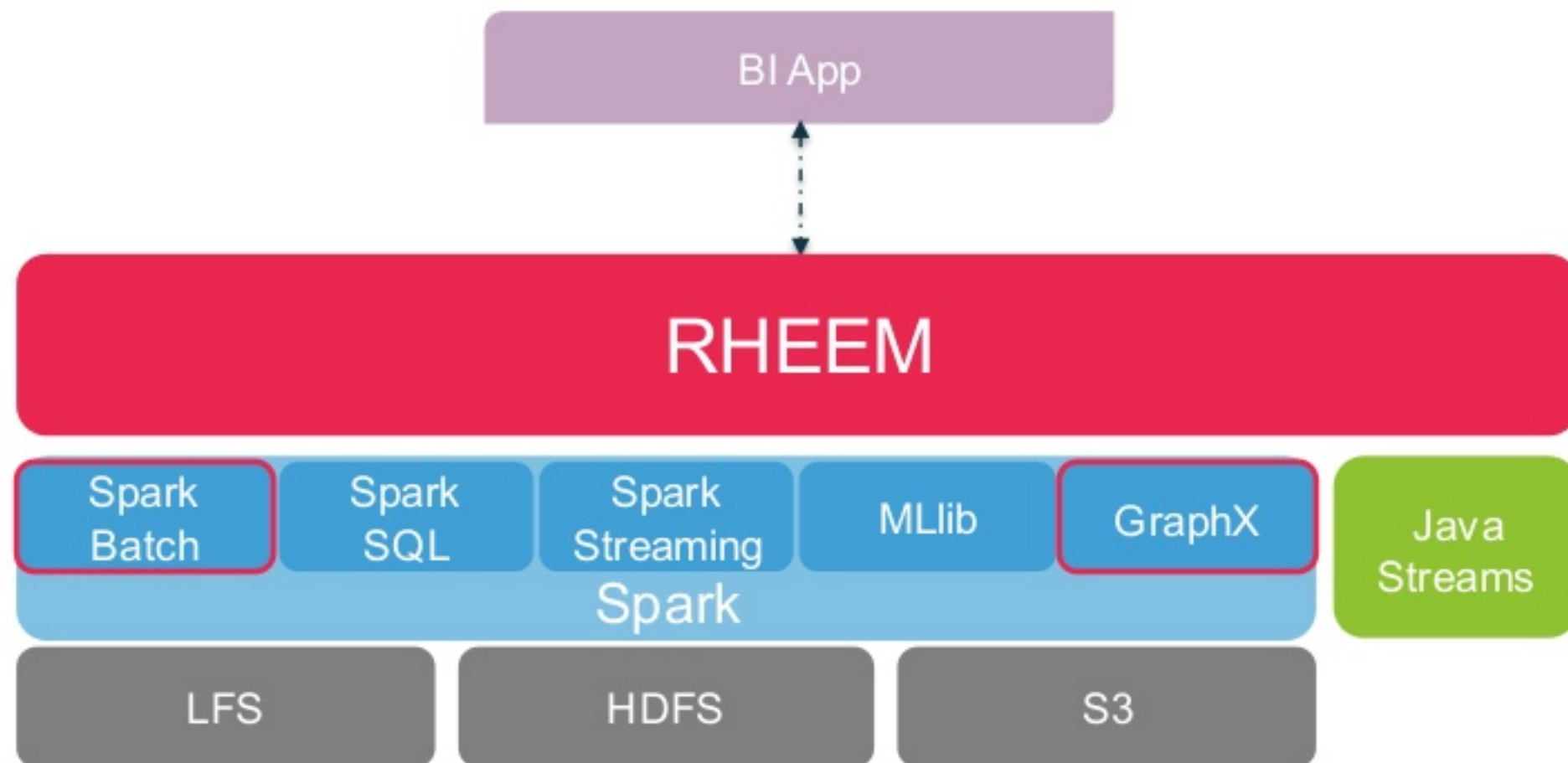
# Beyond a Single Platform

## Big vs. Small Data



# RHEEM

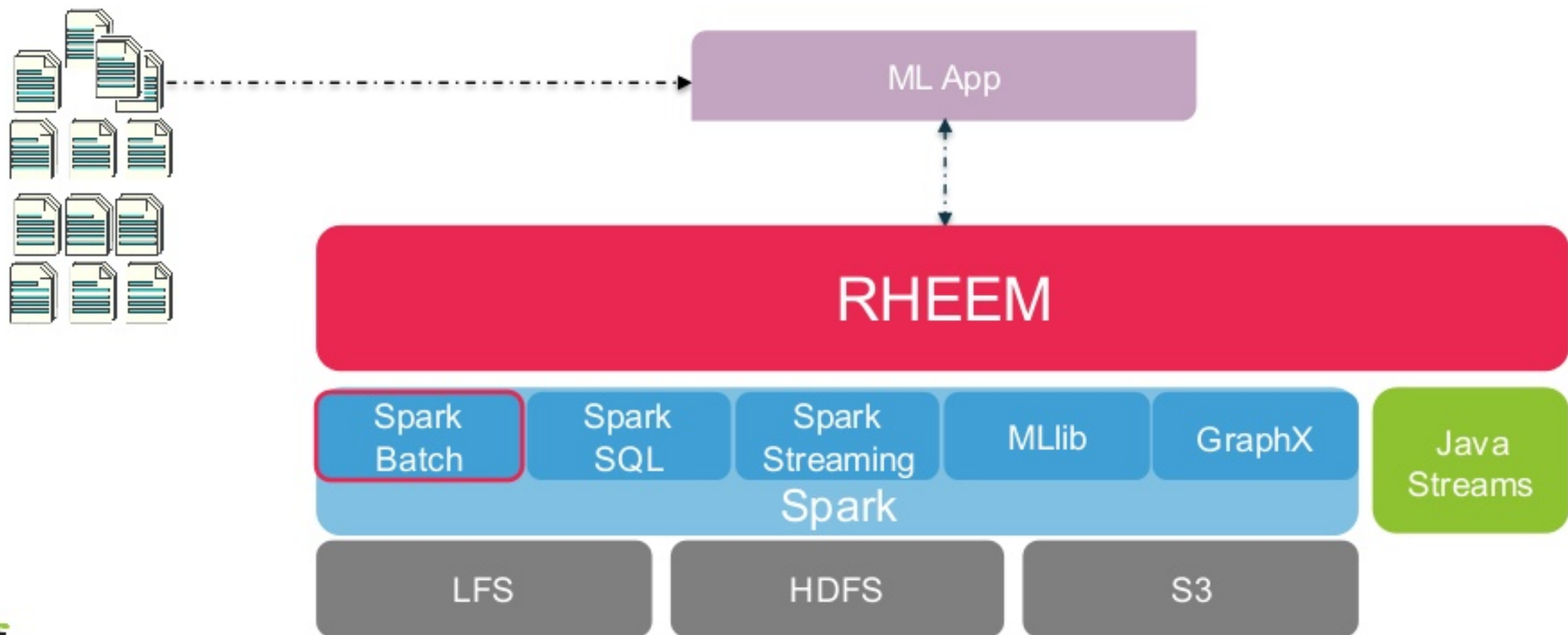
## Complex Apps





# RHEEM

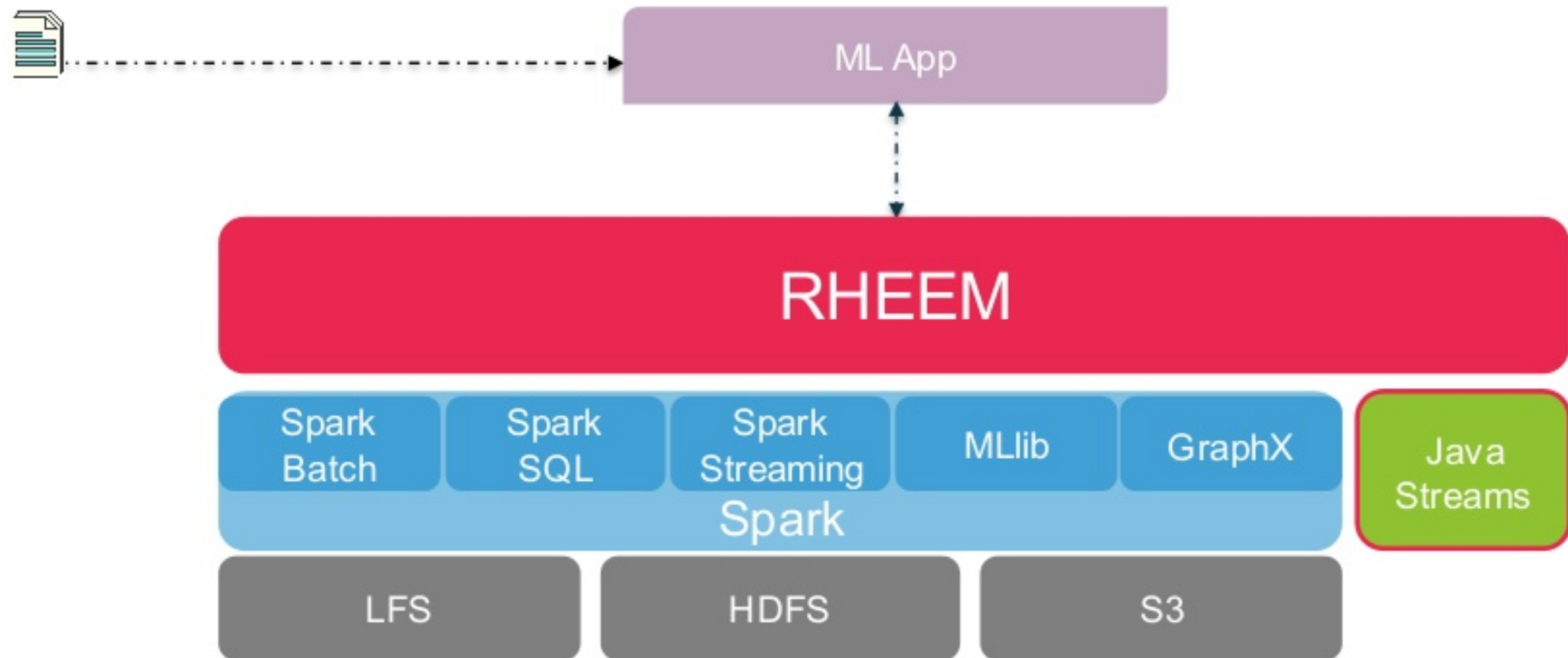
## Big vs. Small Data





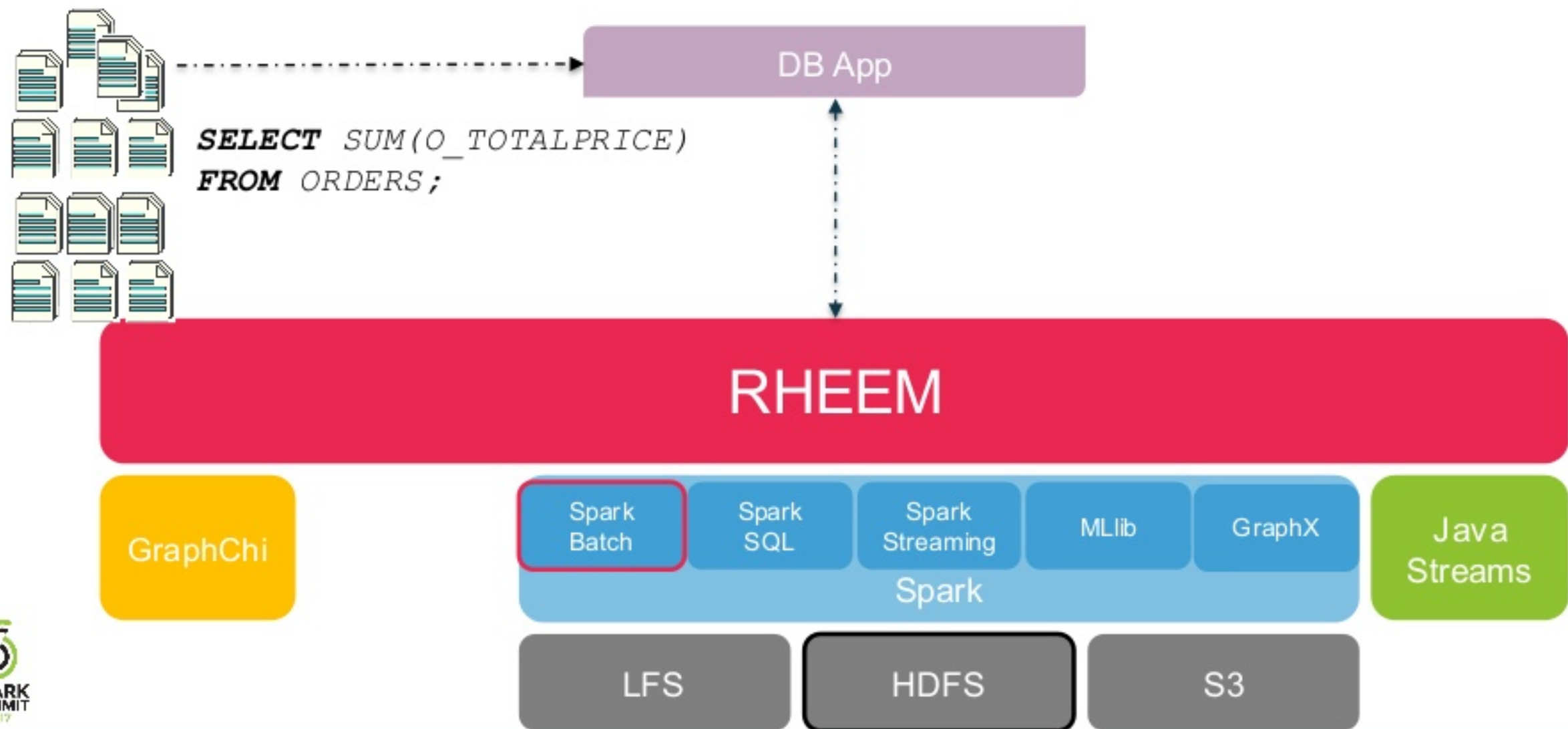
# RHEEM

## Big vs. Small Data

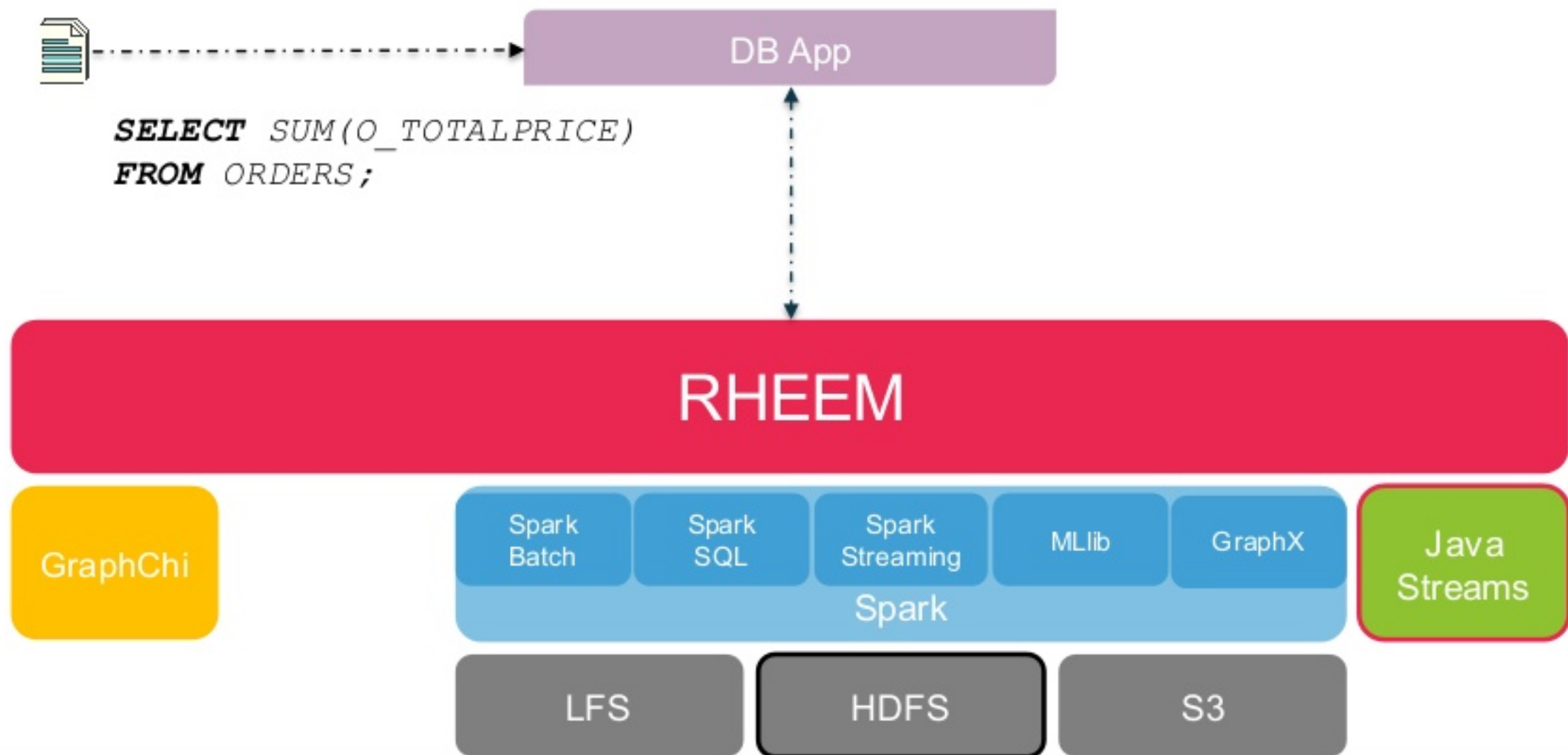


# CROSS-PLATFORM USE CASES

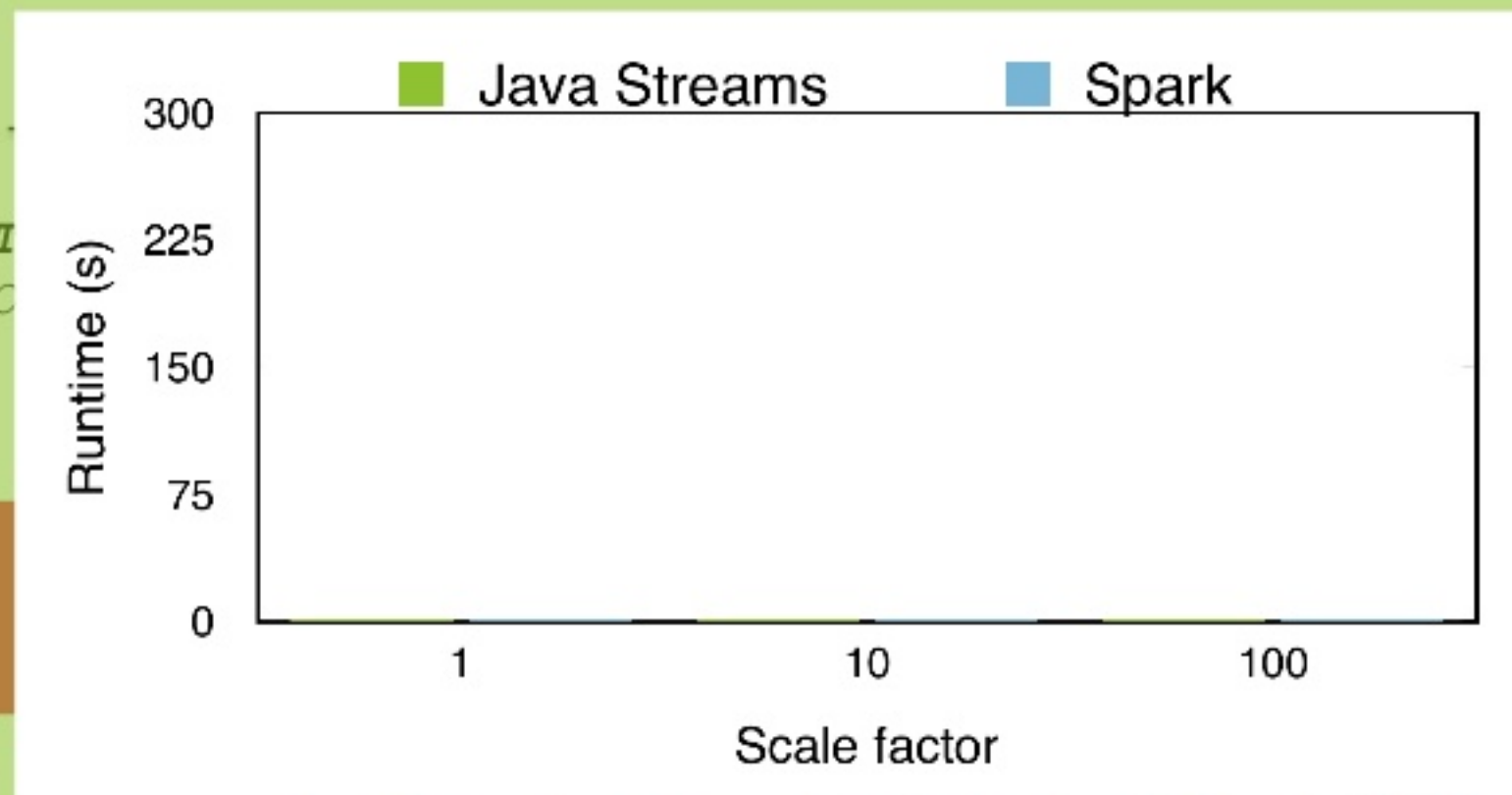
# Platform Independence



# Platform Independence



# Platform Independence



*SELECT*  
*FROM C*

GraphChi

Batch

SQL

Streaming

MLlib

GraphX

Spark

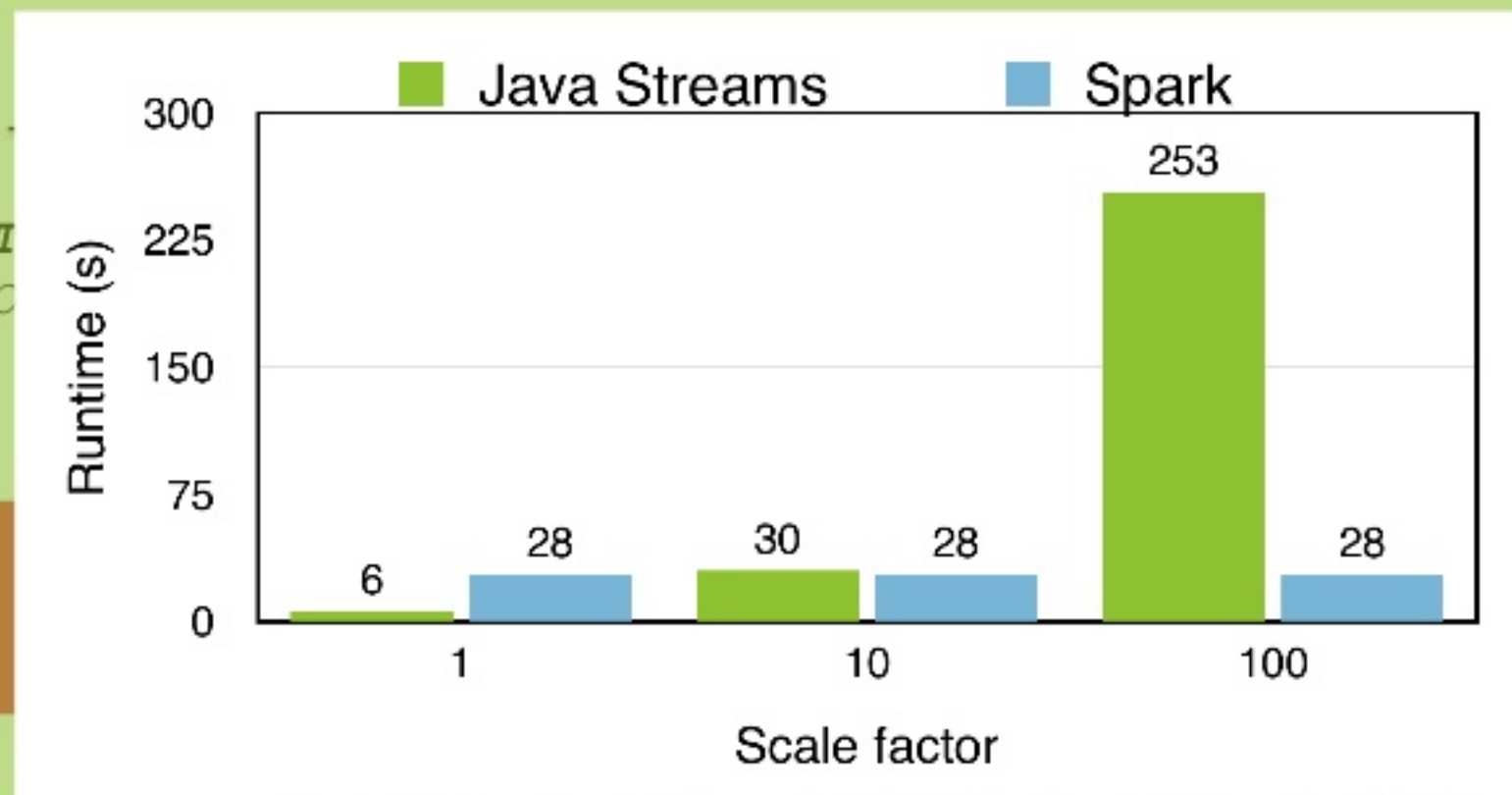
Java  
Streams

LFS

HDFS

S3

# Platform Independence



*SELECT*  
*FROM C*

GraphChi

Batch

SQL

Streaming

MLlib

GraphX

Spark

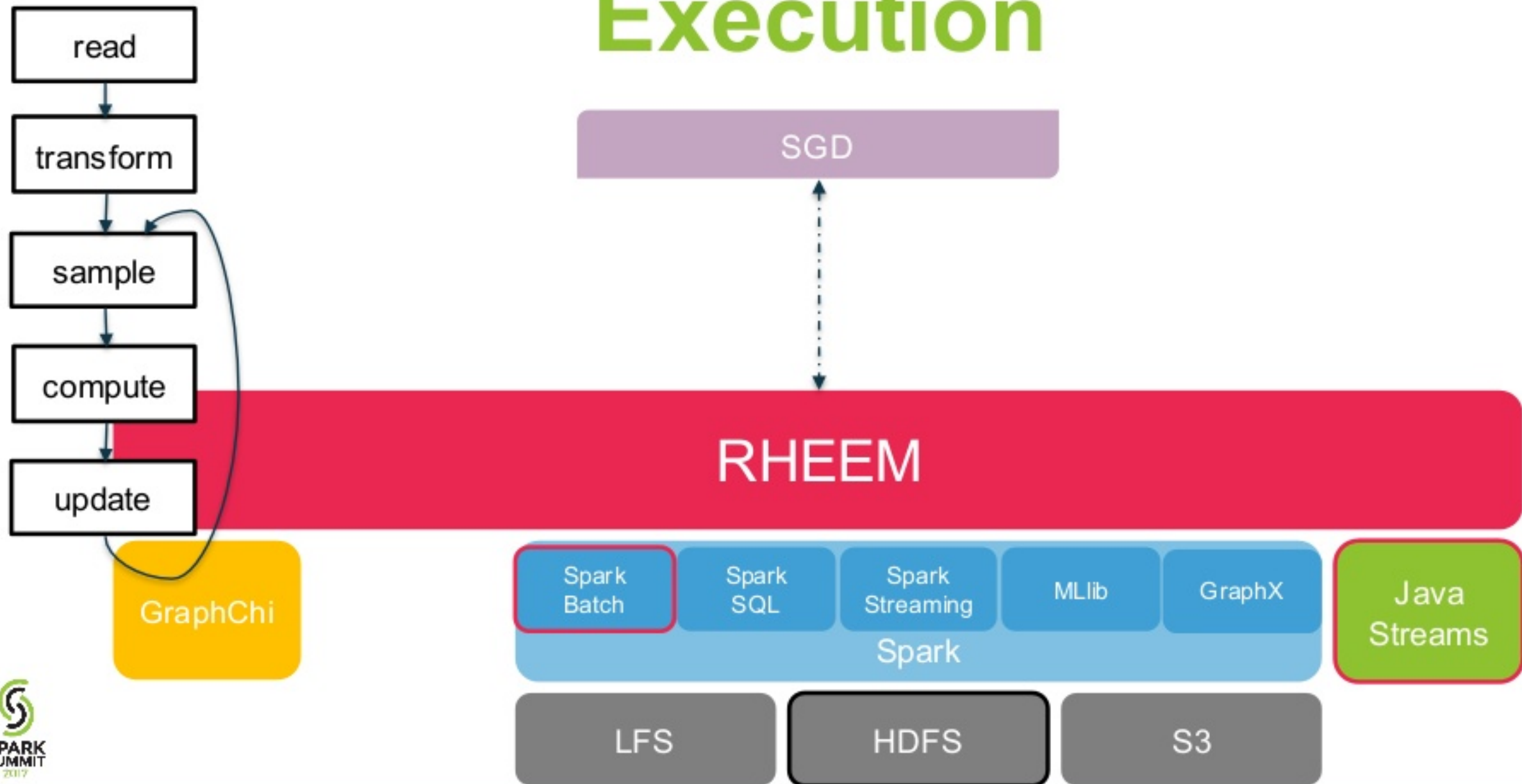
Java  
Streams

LFS

HDFS

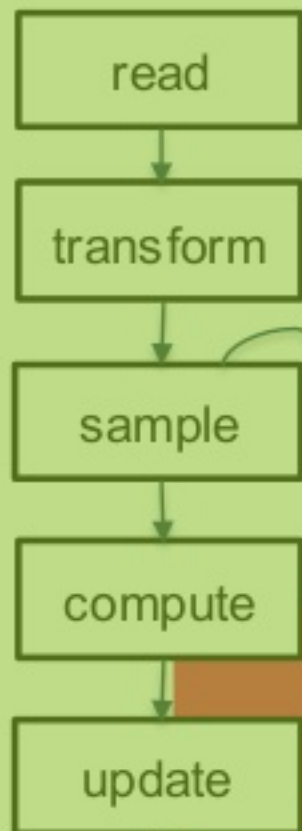
S3

# Opportune Multi-Platform Execution

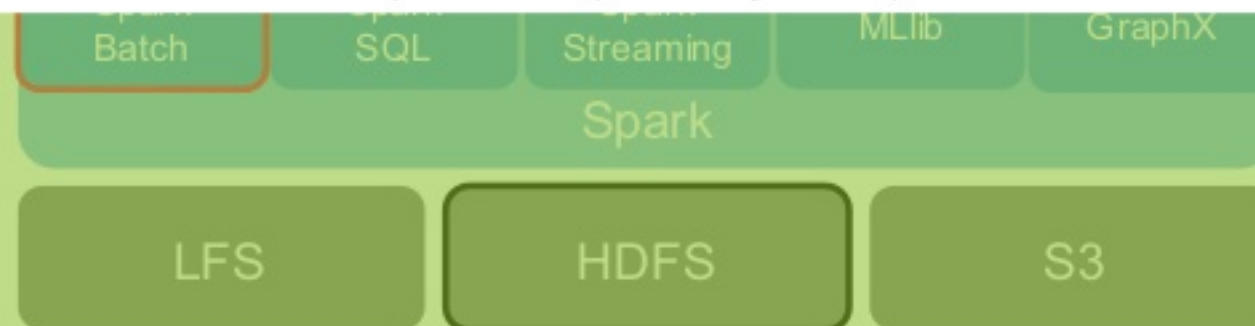
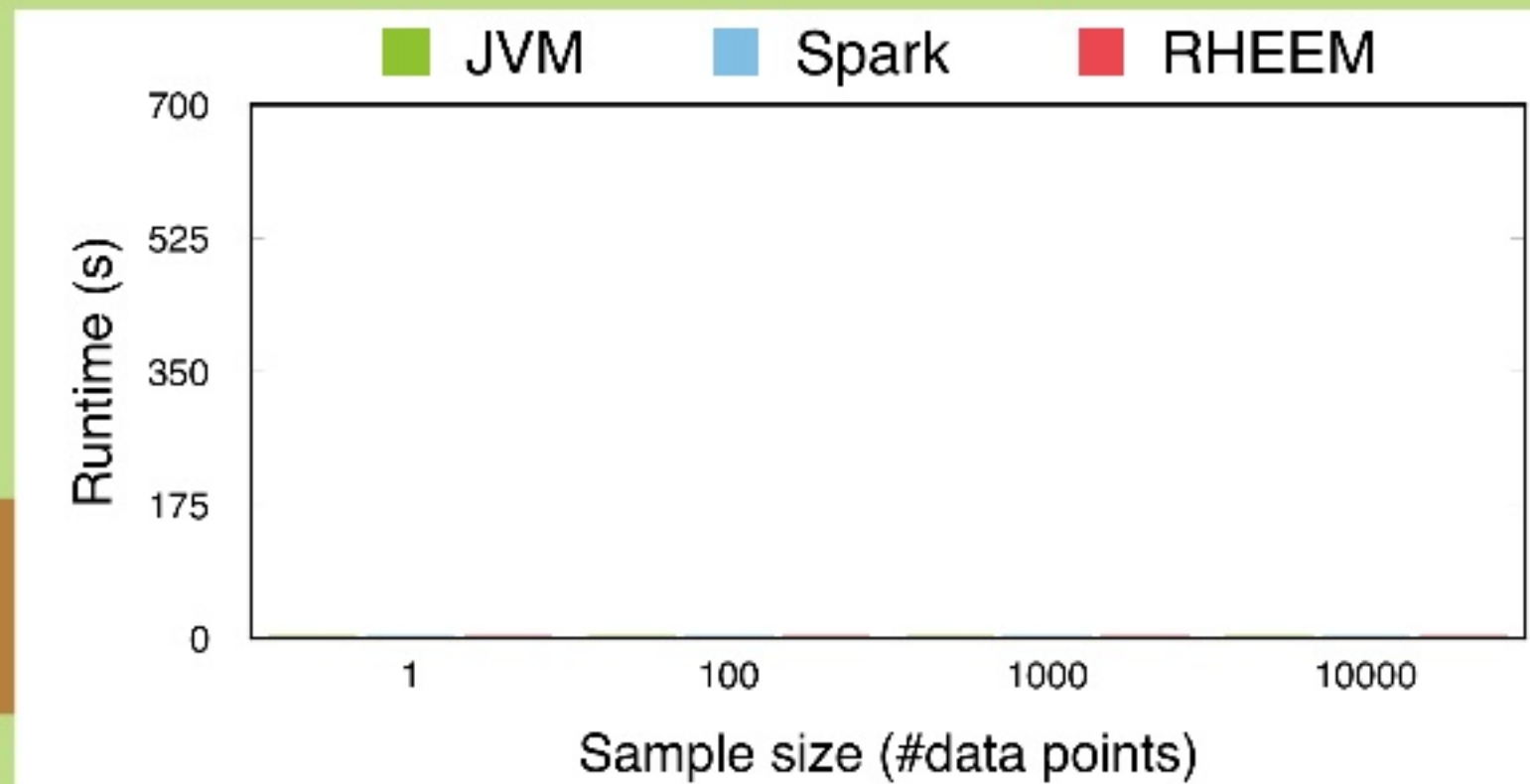




# Opportune Multi-Platform Execution

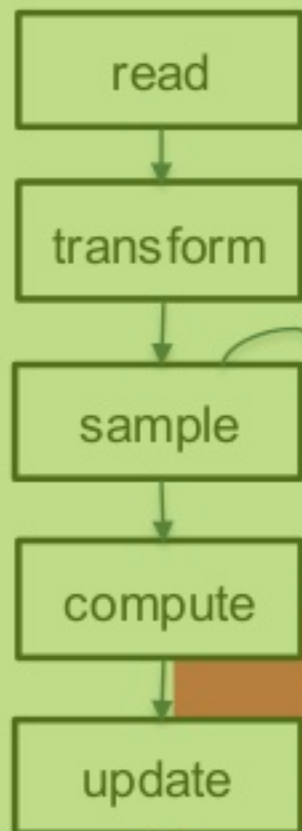


GraphChi

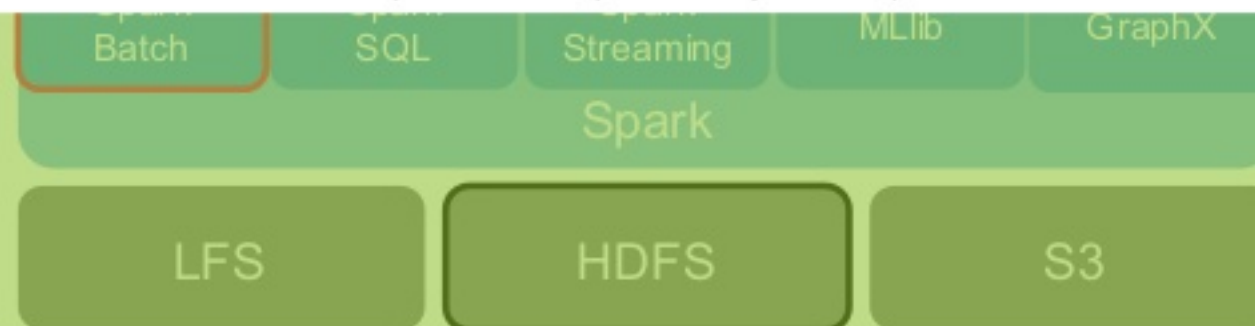
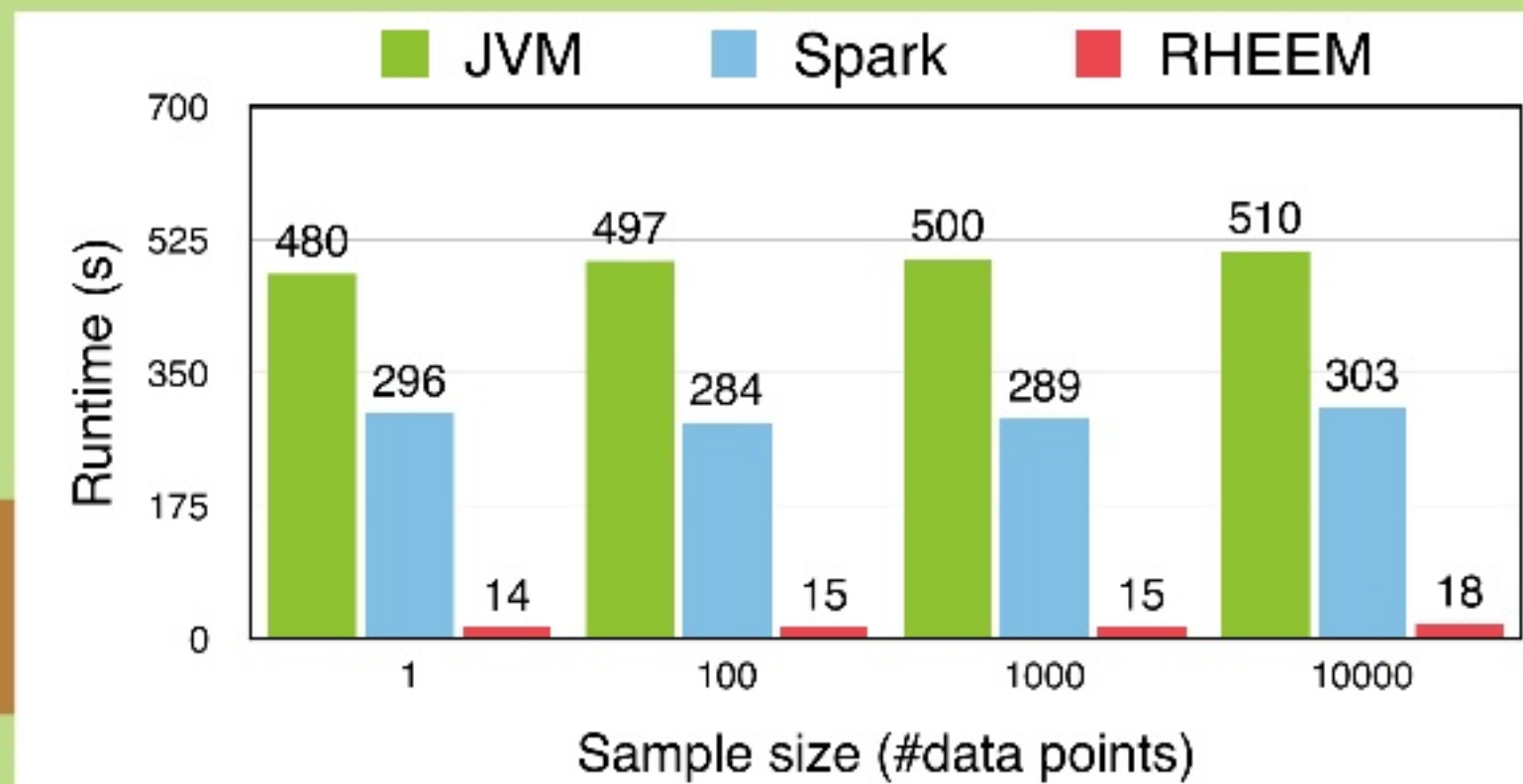


Java Streams

# Opportune Multi-Platform Execution



GraphChi



REGION  
SUPPLIER  
CUSTOMER

NATION

ORDERS  
LINEITEM



# Polystore

TPC-H Q5

RHEEM

GraphChi

Spark  
Batch

Spark  
SQL

Spark  
Streaming

MLlib

GraphX

Java  
Streams

Spark

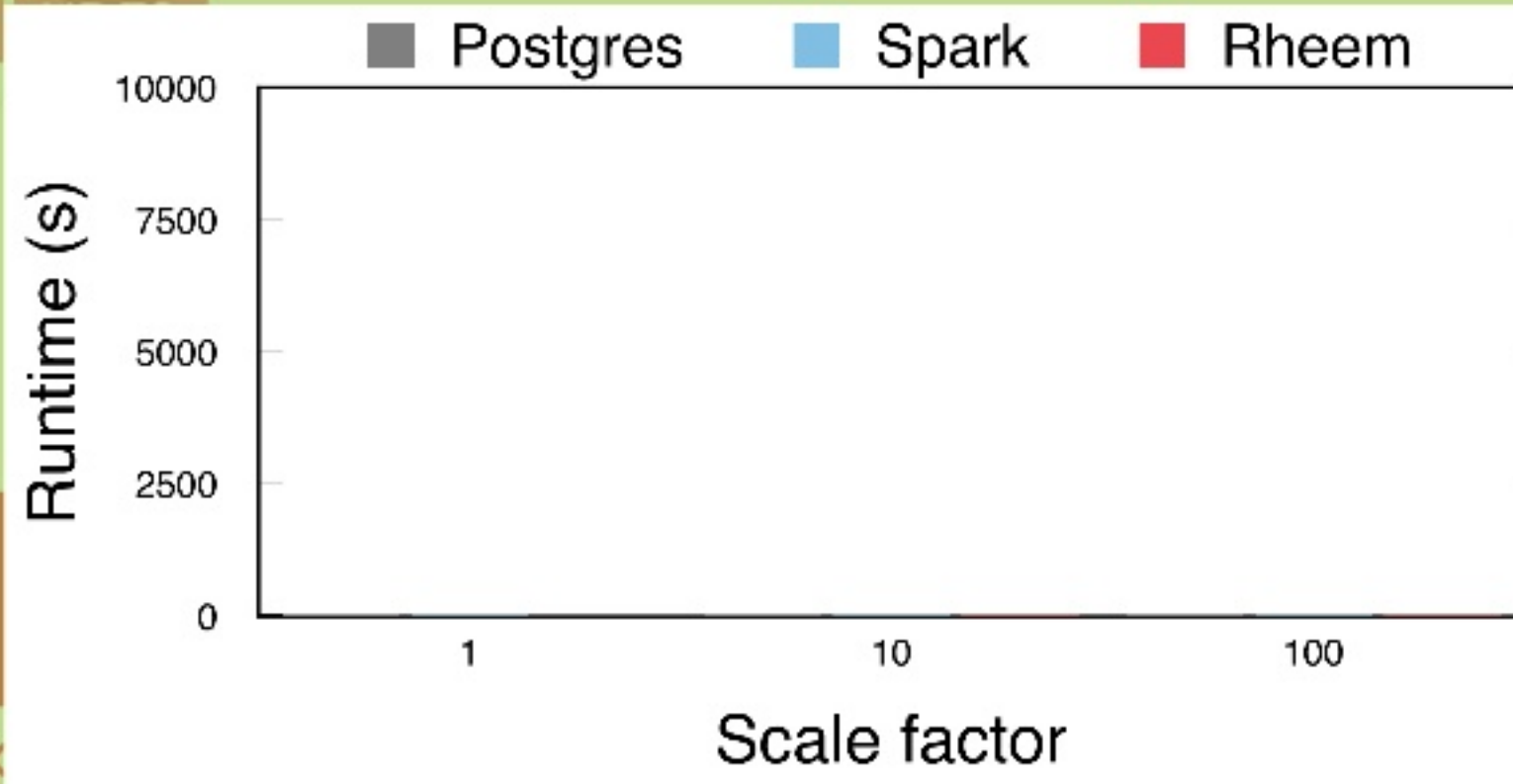
LFS

HDFS

S3

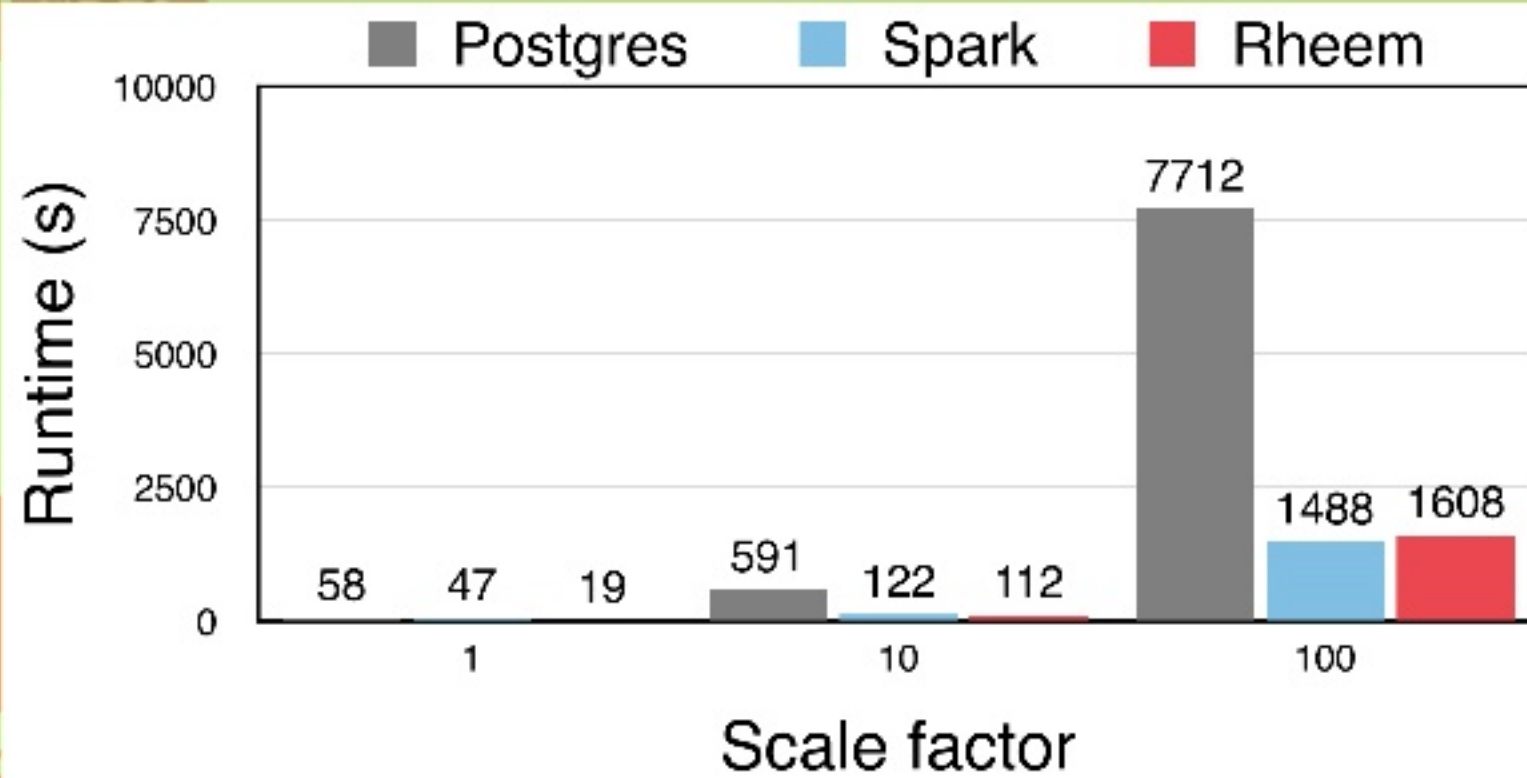
REGION  
SUPPLIER  
CUSTOMER  
NATION  
ORDERS  
LINEITEM

# Polystore



REGION  
SUPPLIER  
CUSTOMER  
NATION  
ORDERS  
LINEITEM

# Polystore



GraphChi

Batch

SQL

Streaming

MLlib

GraphX

Java  
Streams

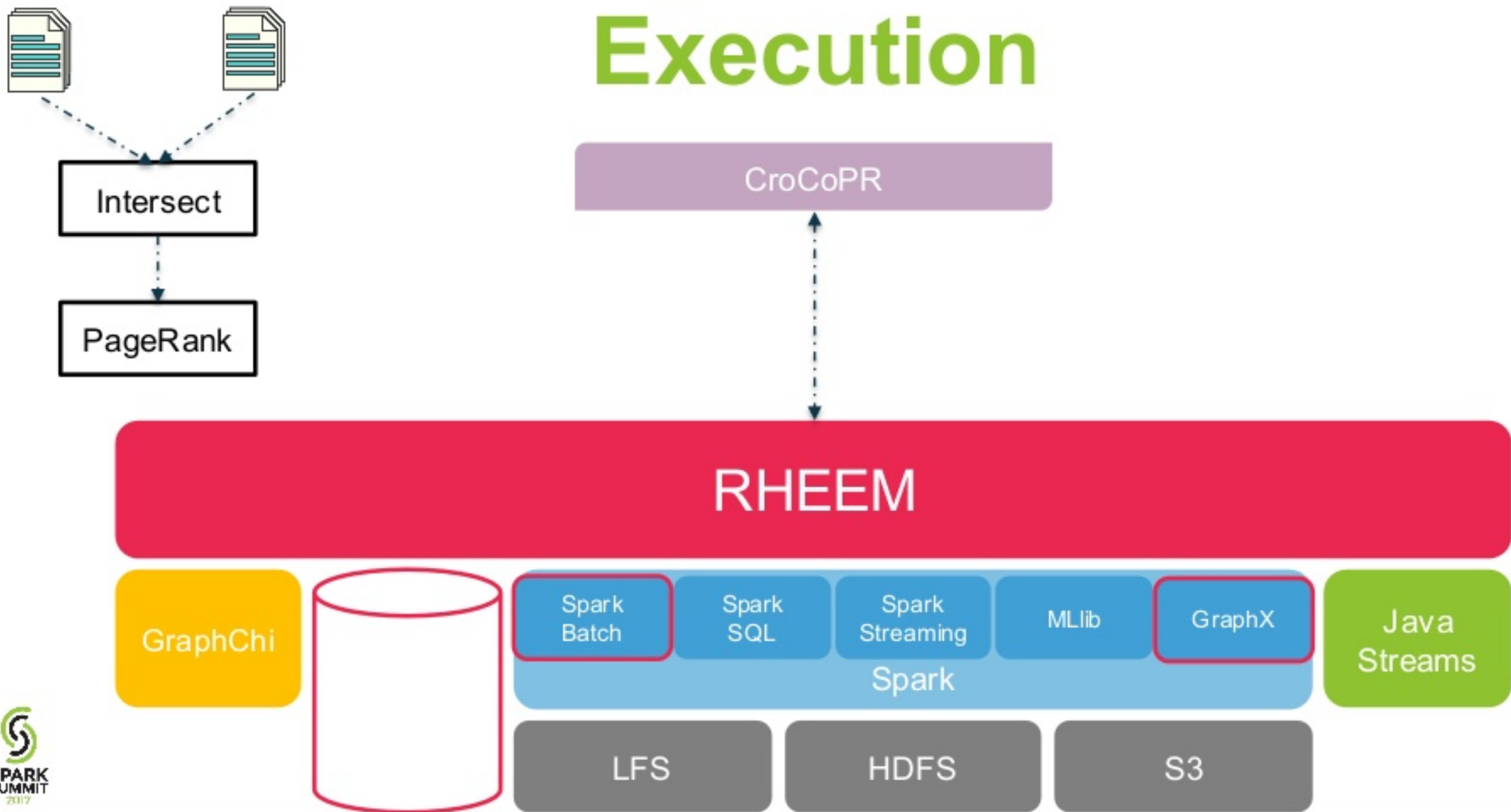
Spark

LFS

HDFS

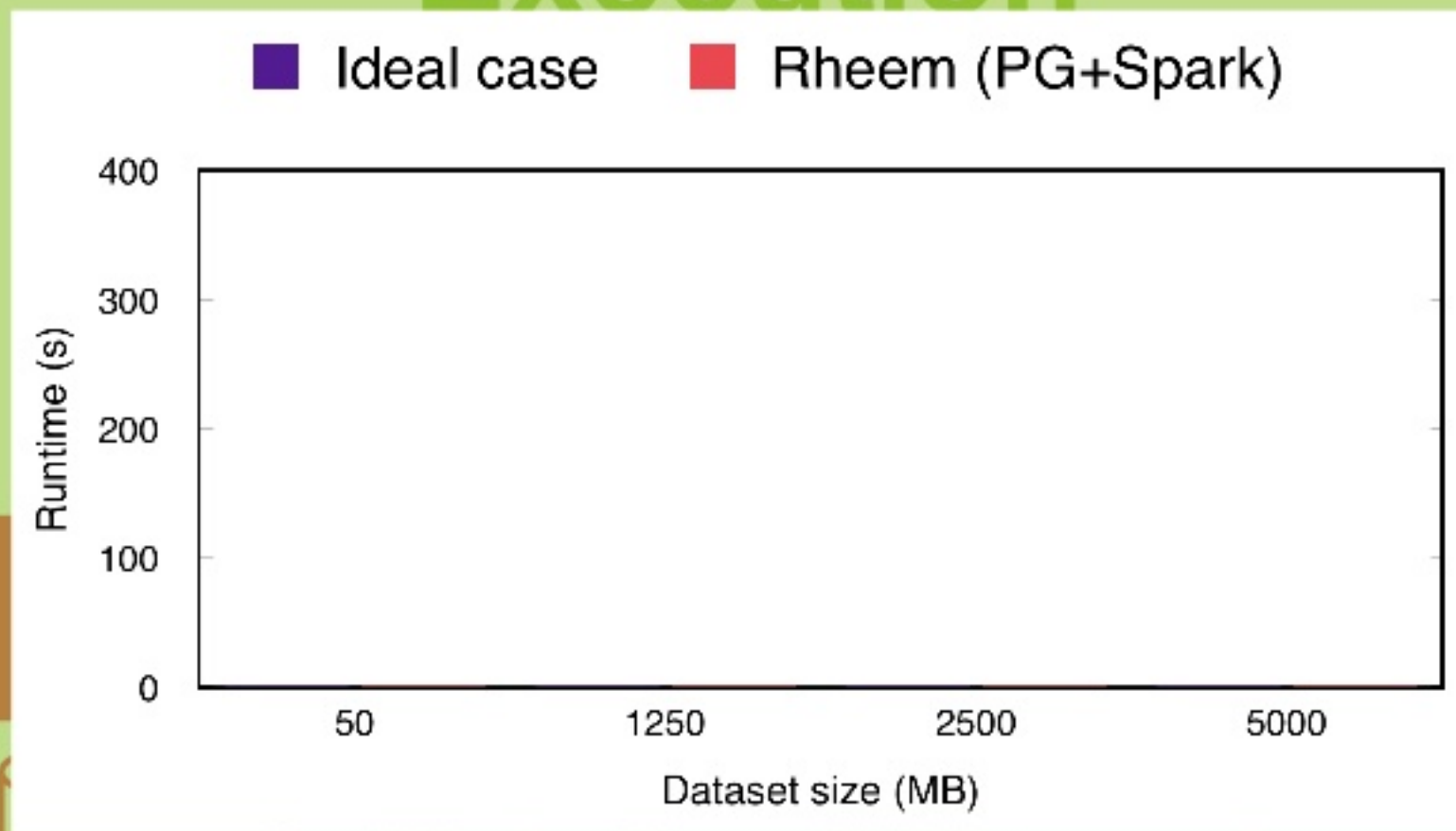
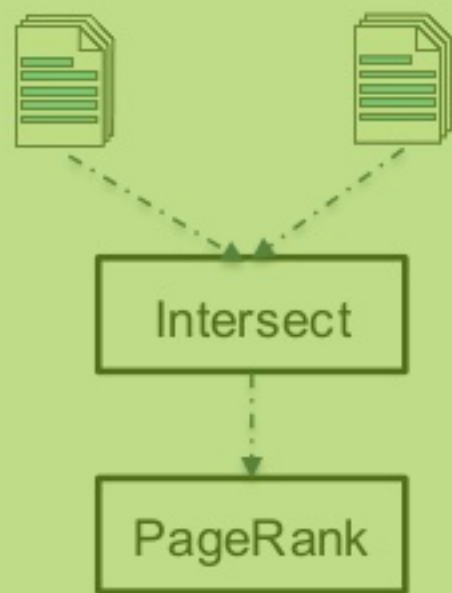
S3

# Mandatory Multi-Platform Execution



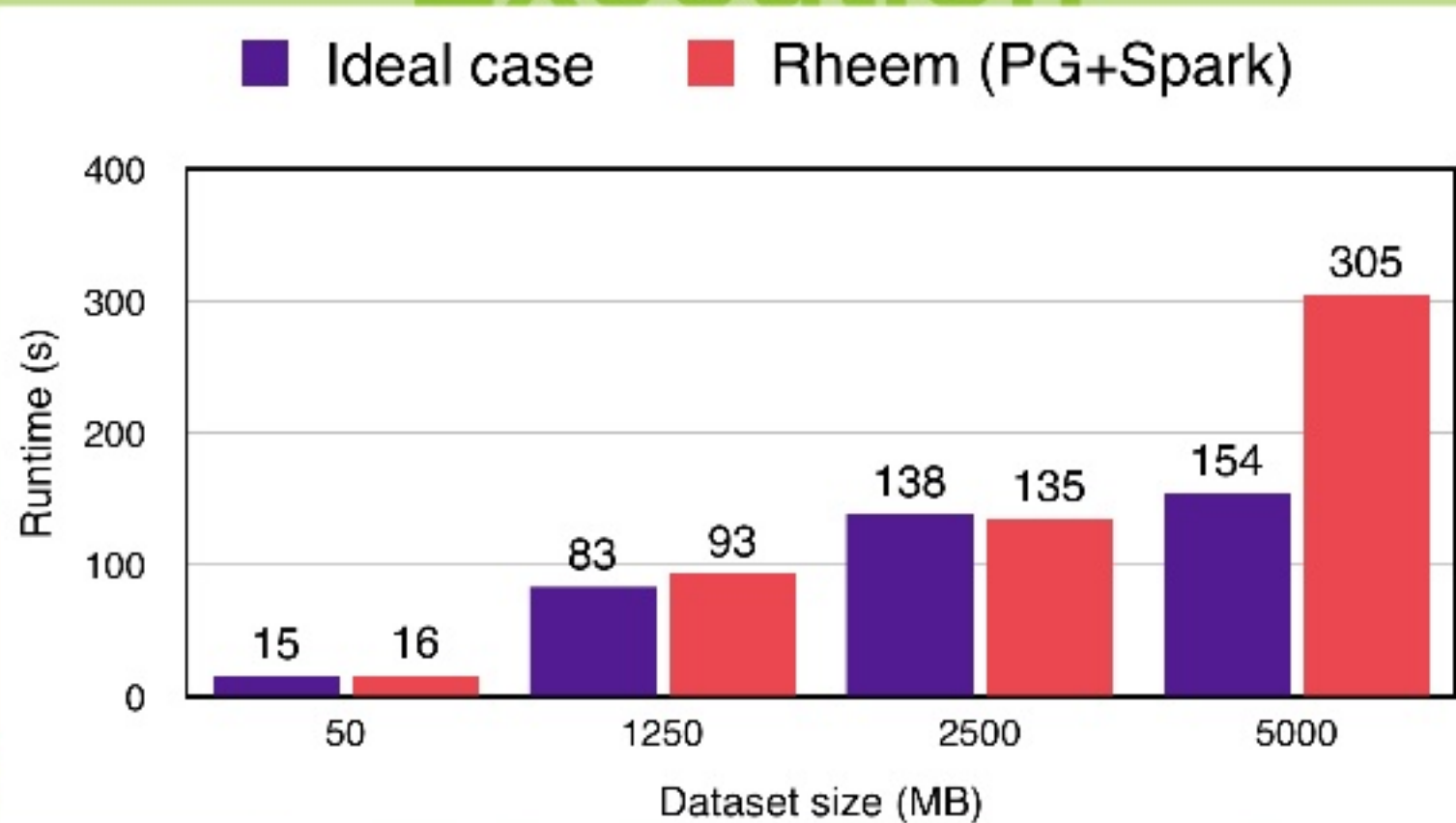
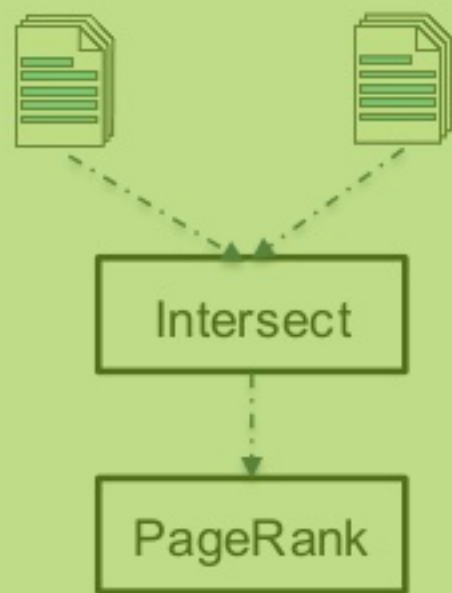


# Mandatory Multi-Platform Execution





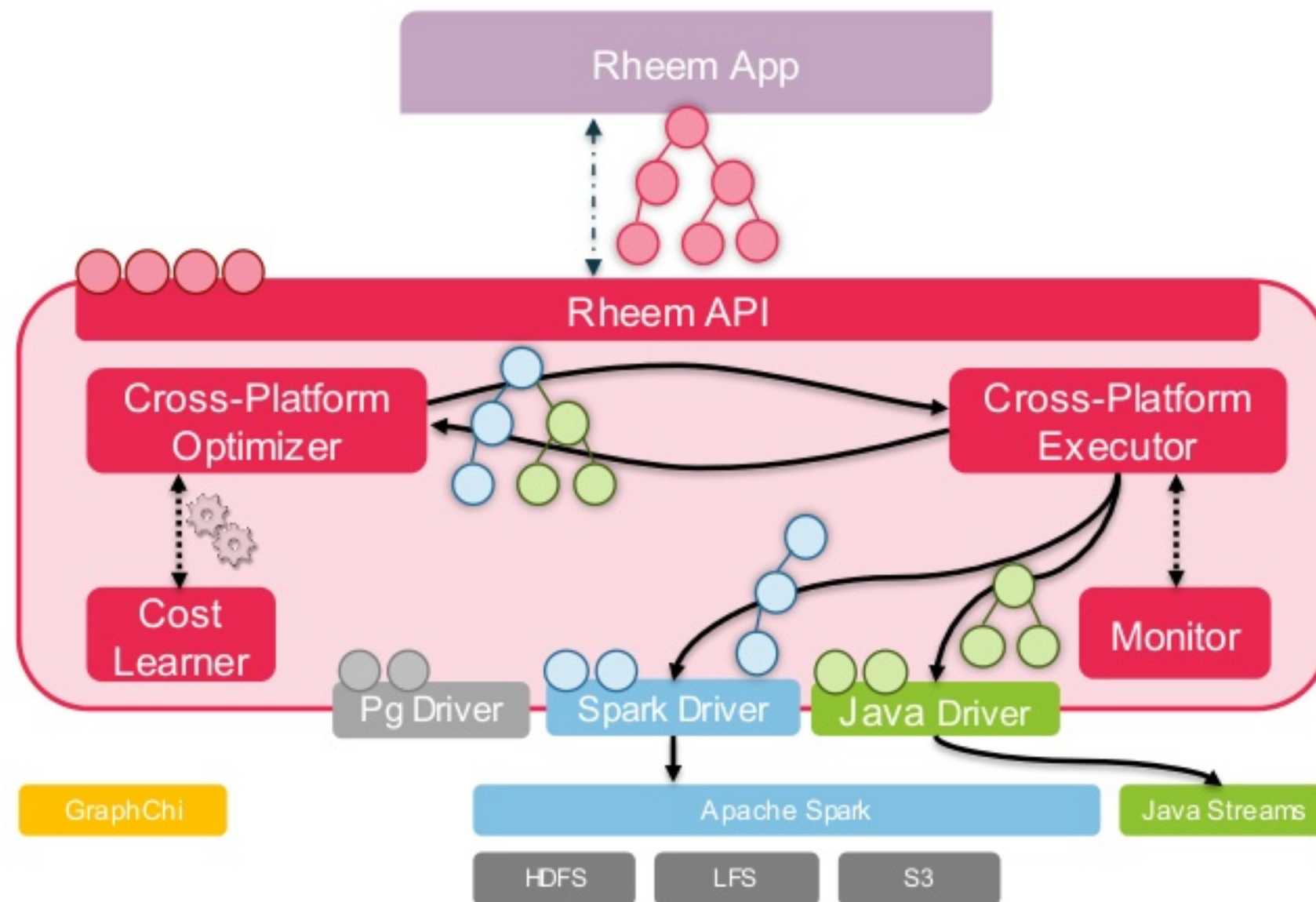
# Mandatory Multi-Platform Execution



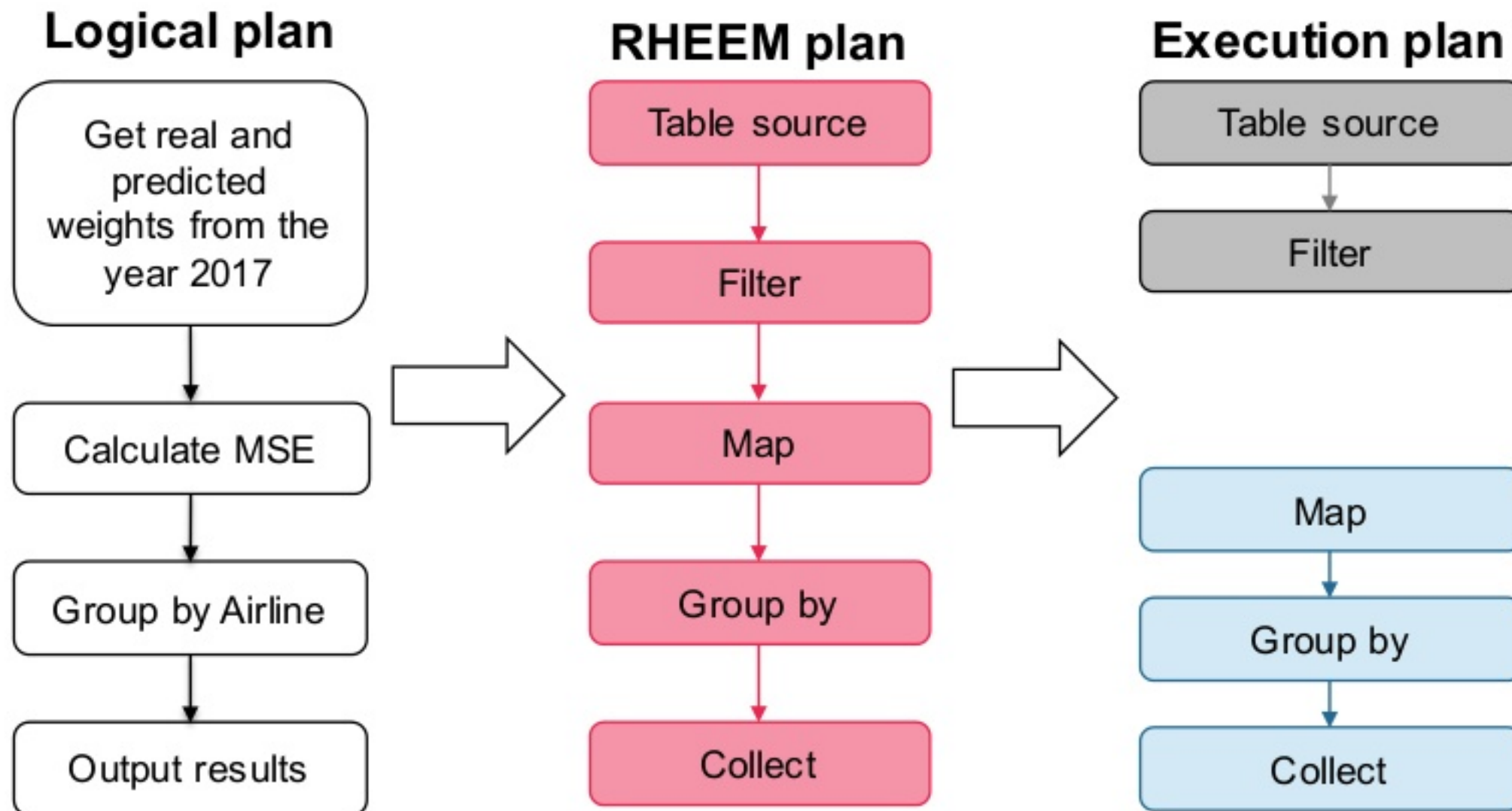
# RHEEM INTERNALS

*Let's have a look under the hood...*

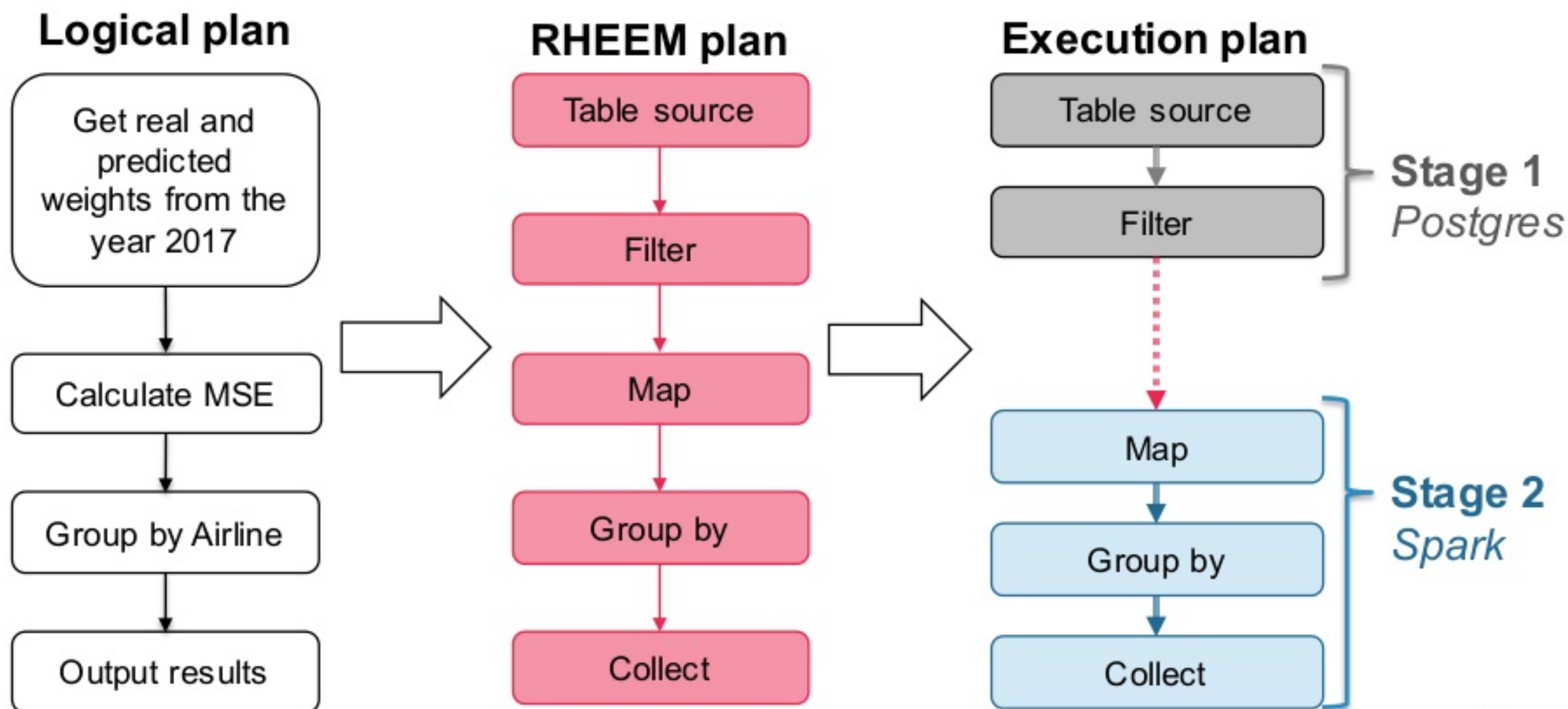
# RHEEM in a Nutshell



# Fine-grained Platform Selection

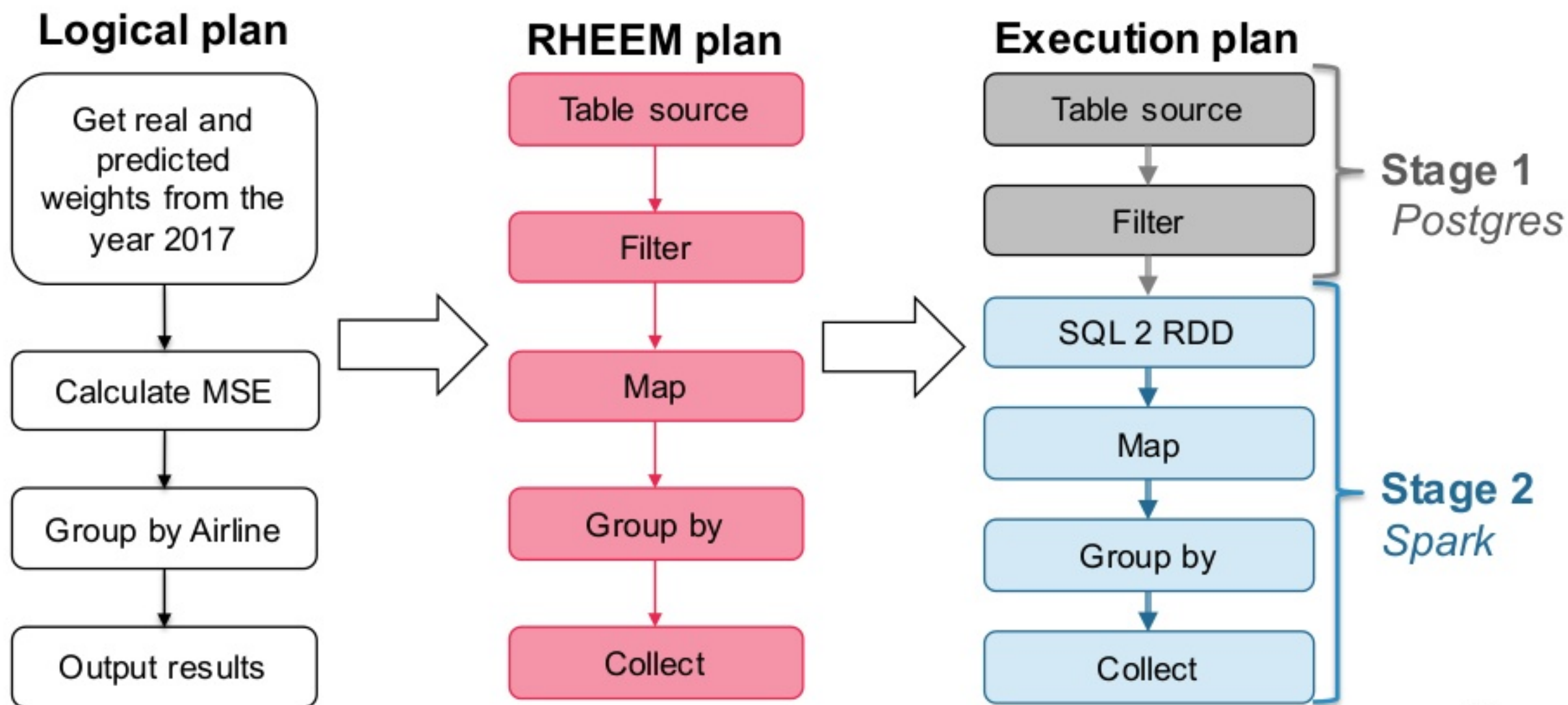


# Automatic Data Movement

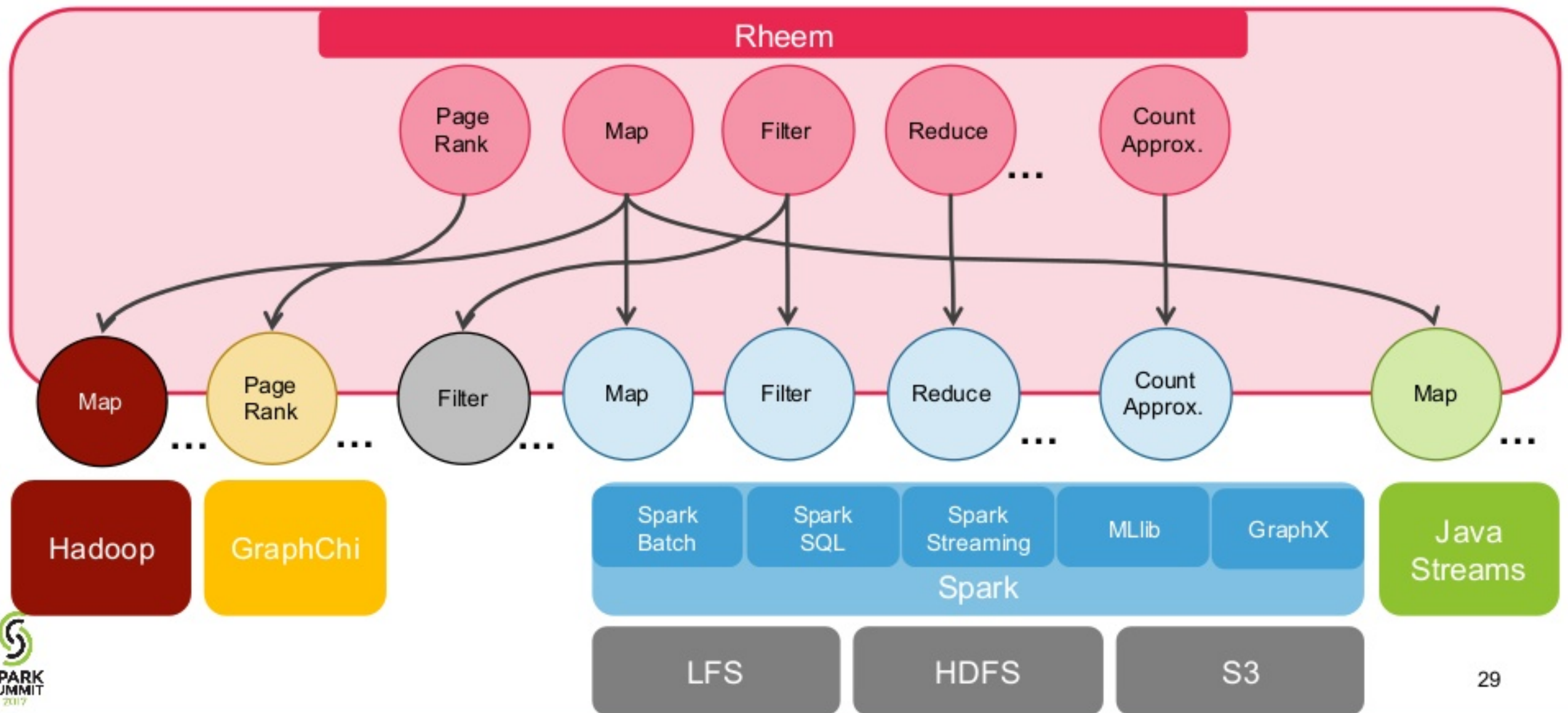




# Automatic Data Movement

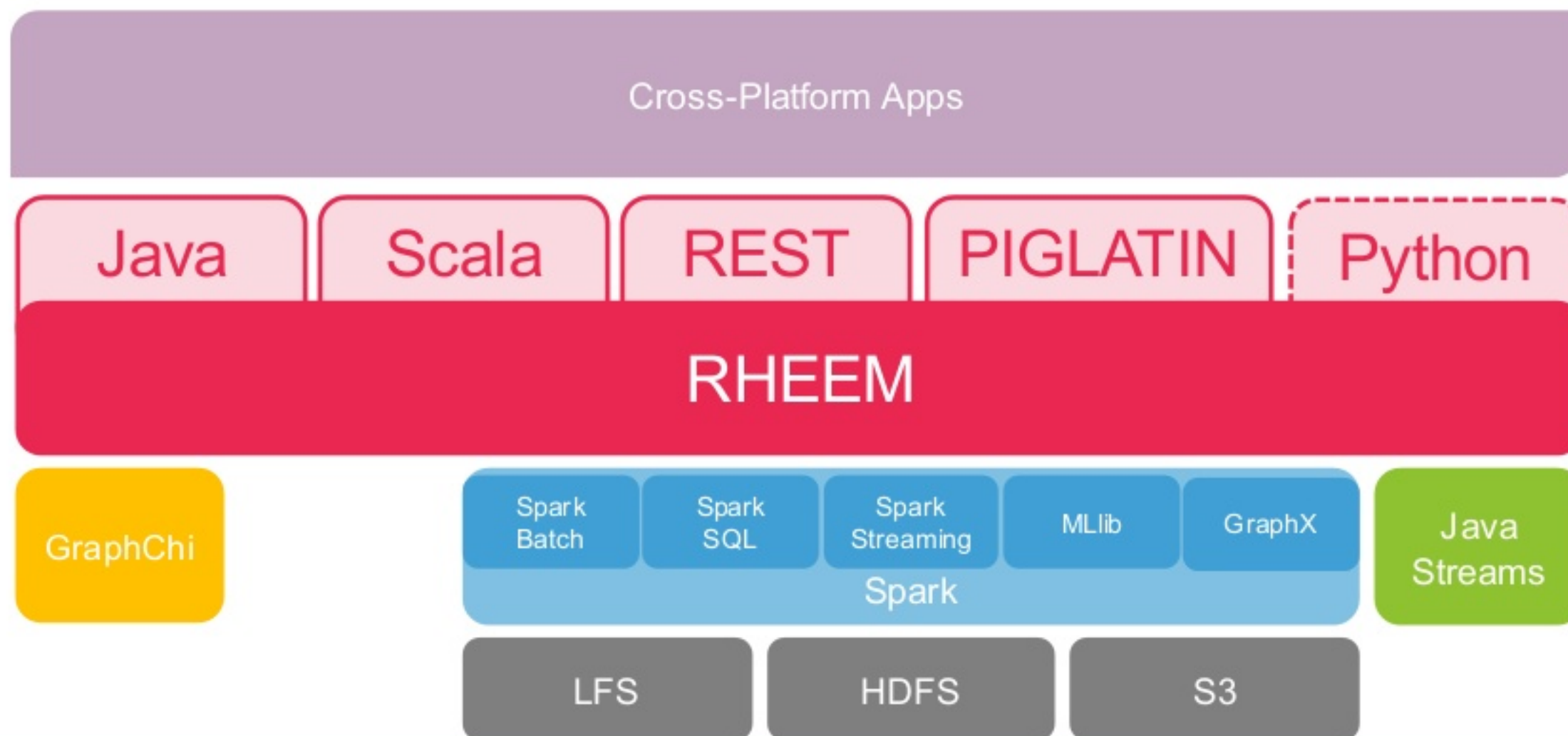


# Extensibility

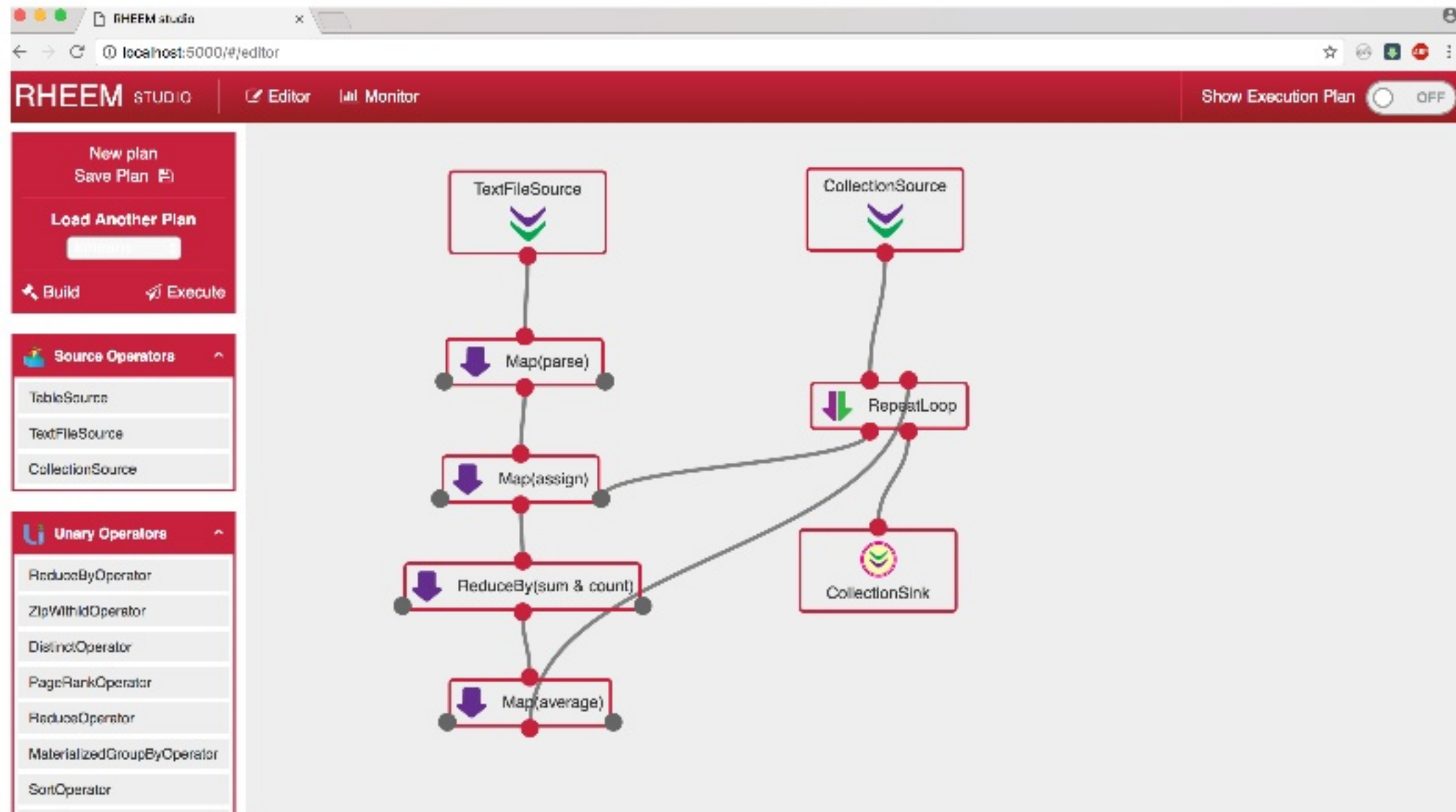




# Multiple APIs



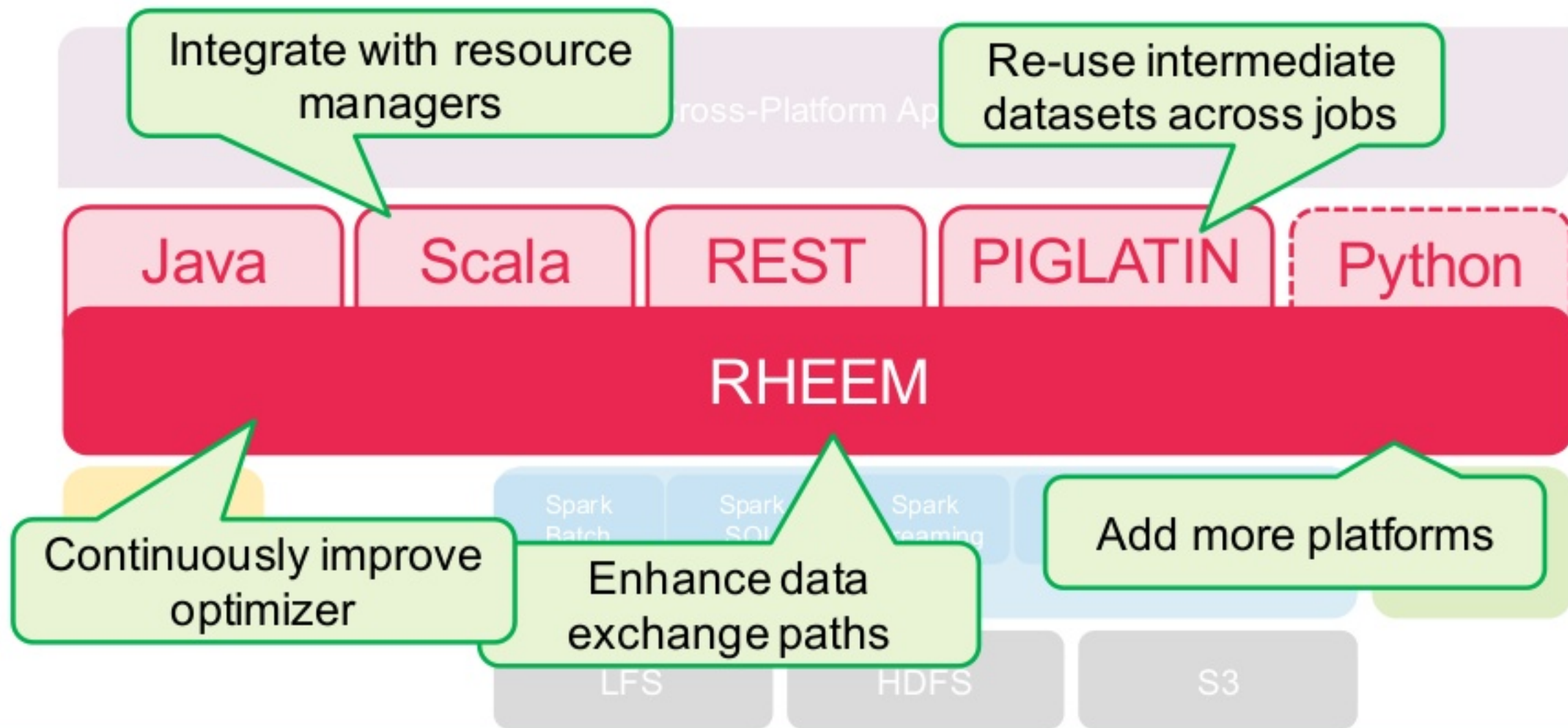
# Rheem Studio



# DEMO

*Let's play...*

# The dRHEEM goes on...





Website: <http://da.qcri.org/rheem/>

GitHub: <https://github.com/rheem-ecosystem>

Gitter: <https://gitter.im/rheem-ecosystem/Lobby>

# We want you!

Apache Incubator Very Soon!

Interested? Then, become a  
**dRHEEMer!**

