

# Quiltdata

Consulting Project

## Glowing Quilt: A Fabric for Spark Data Feb. 2017

**Rich H. Inman**

PhD Candidate, UCSD

Insight Data Engineering Fellow

# What is Quiltdata?

Science is hard enough.  
Why wrestle with code  
and outdated tools?



**Data ready to roll**



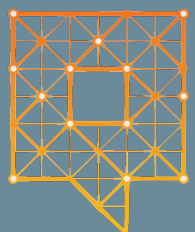
**Analyze without coding**



**Built for big data**



**Cite and leverage**



**Quiltdata**

# Consulting

Understand

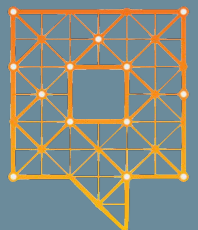


## Understand

Where are you now?

Where would you like to go?

What are the main issues?



Quiltdata

# Consulting

Understand

Implement

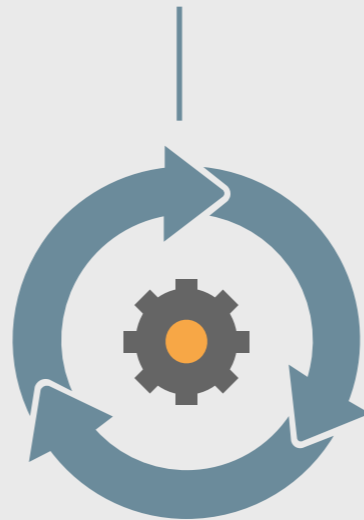


## Understand

Where are you now?

Where would you like to go?

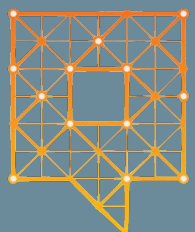
What are the main issues?



## Implement

Work with existing code base

Extend code base to address issues



Quiltdata

# Consulting

Understand

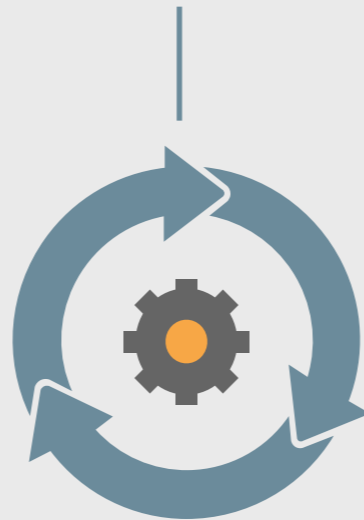
Implement

Test & Disseminate



## Understand

Where are you now?  
Where would you like to go?  
What are the main issues?



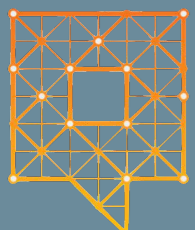
## Implement

Work with existing code base  
Extend code base to address issues



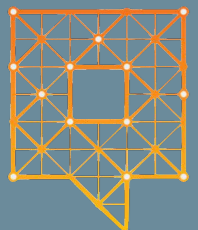
## Test & Disseminate

Does the solution work as expected?  
Does it work for various use cases?  
Tell others about it!



Quiltdata

# Understand

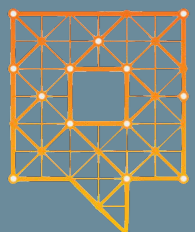


Quiltdata

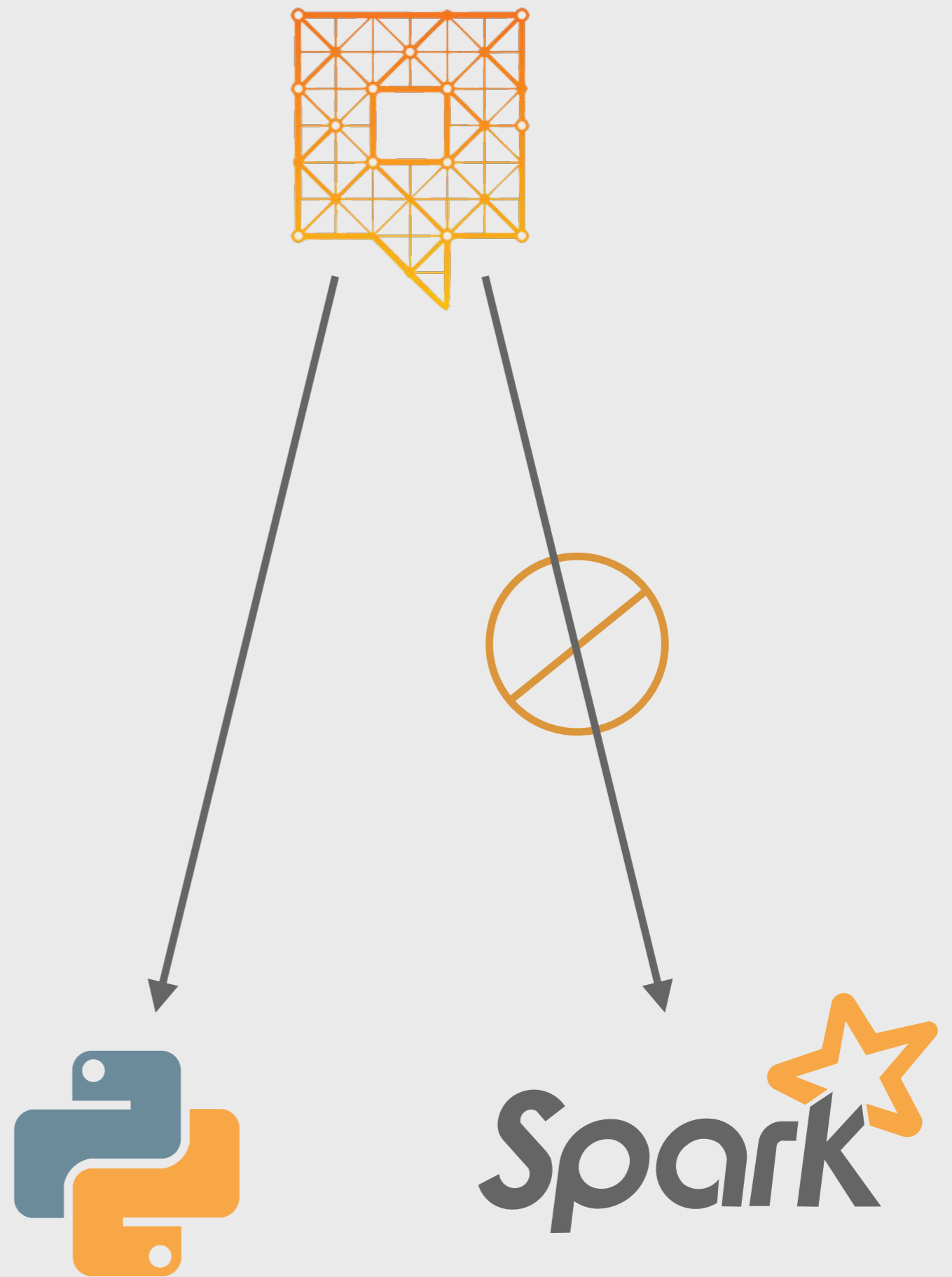
# Understand



Current client  
supports Python on  
single machine



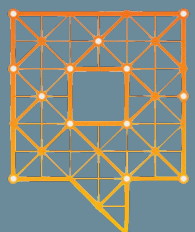
Quiltdata



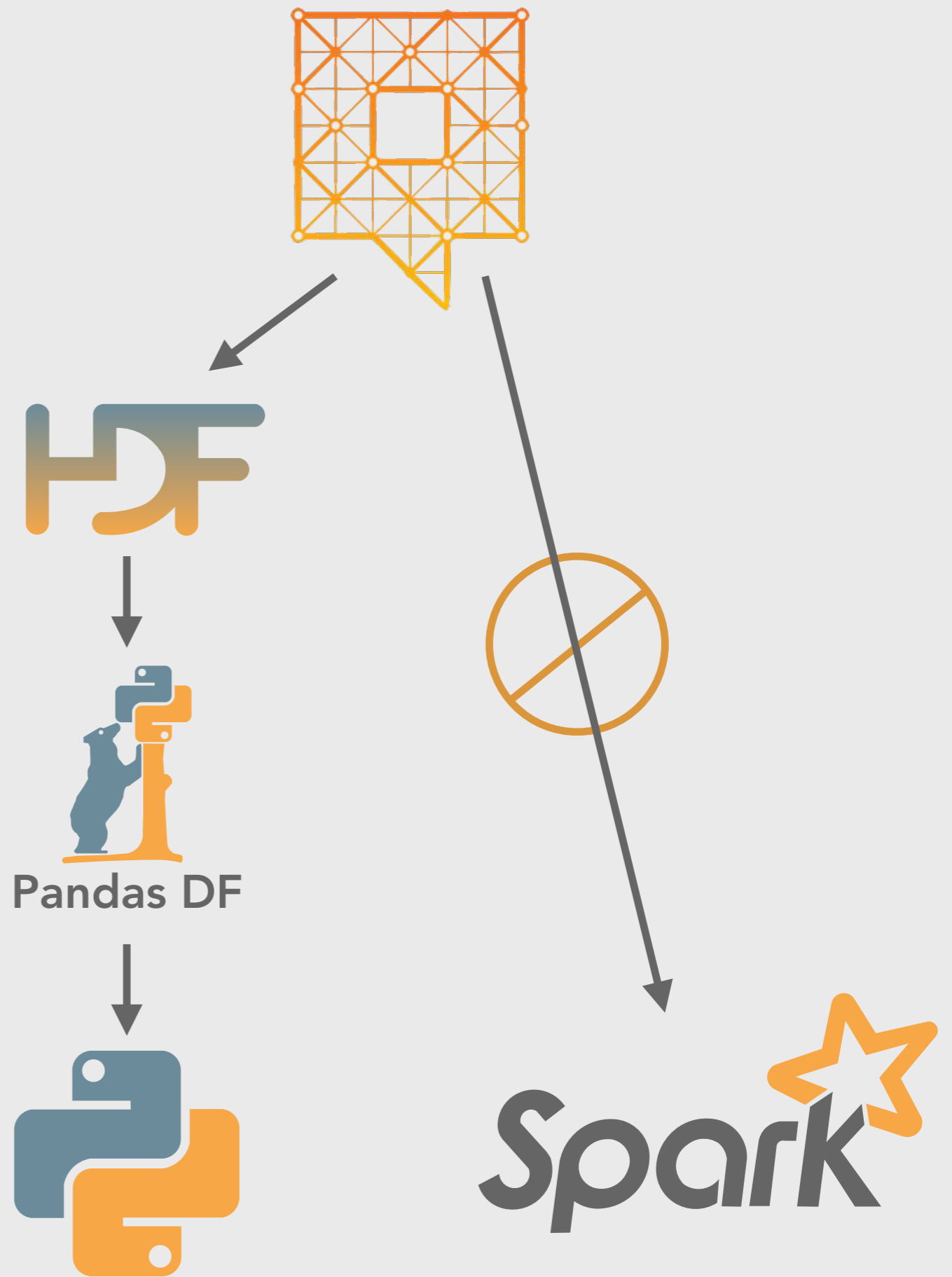
# Understand



Current client supports Python on single machine



Quiltdata



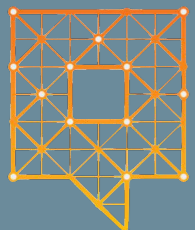


# Understand

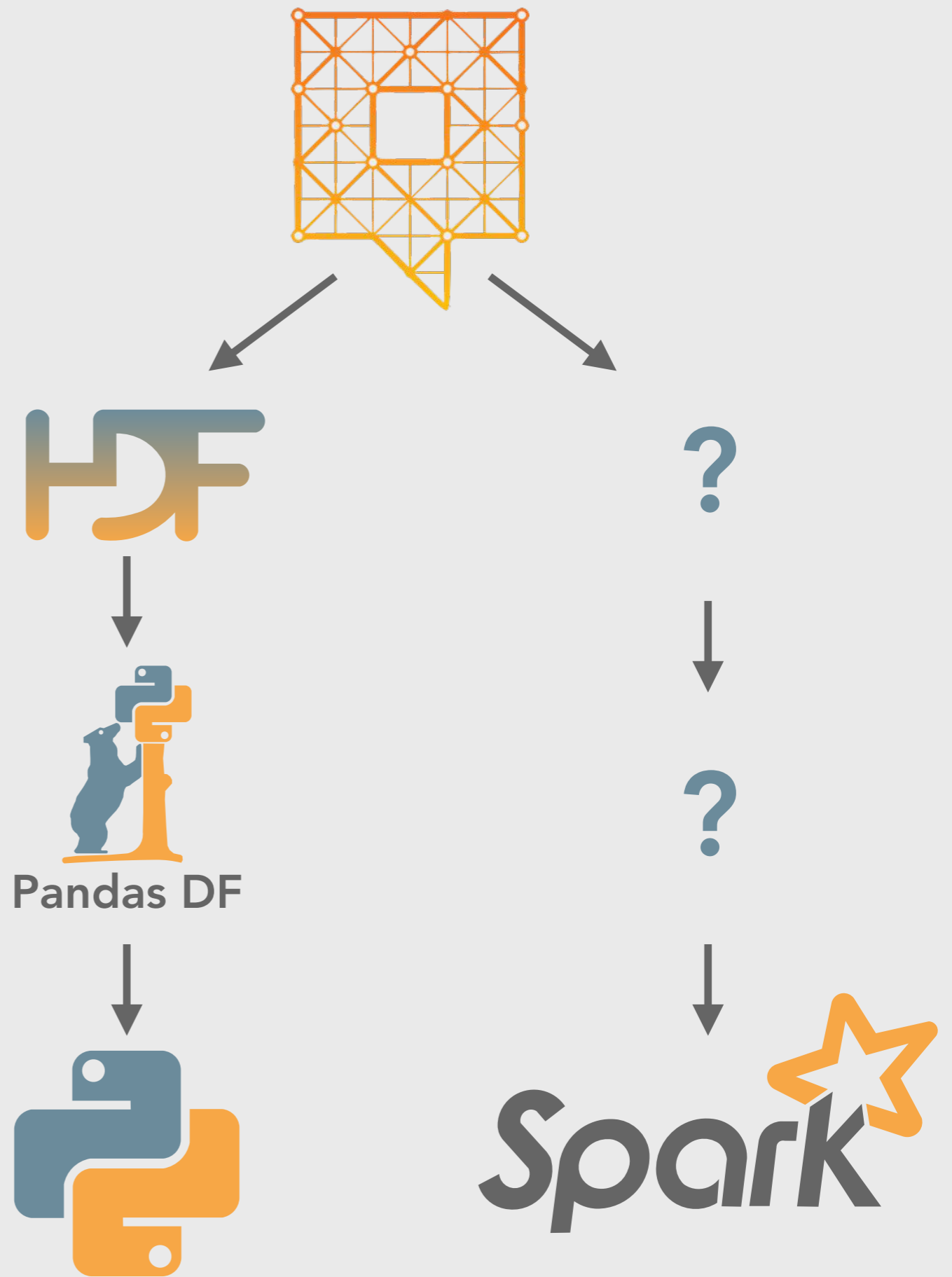


Current client supports Python on single machine

Extend client support to Spark



Quiltdata

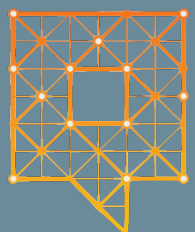


# Understand

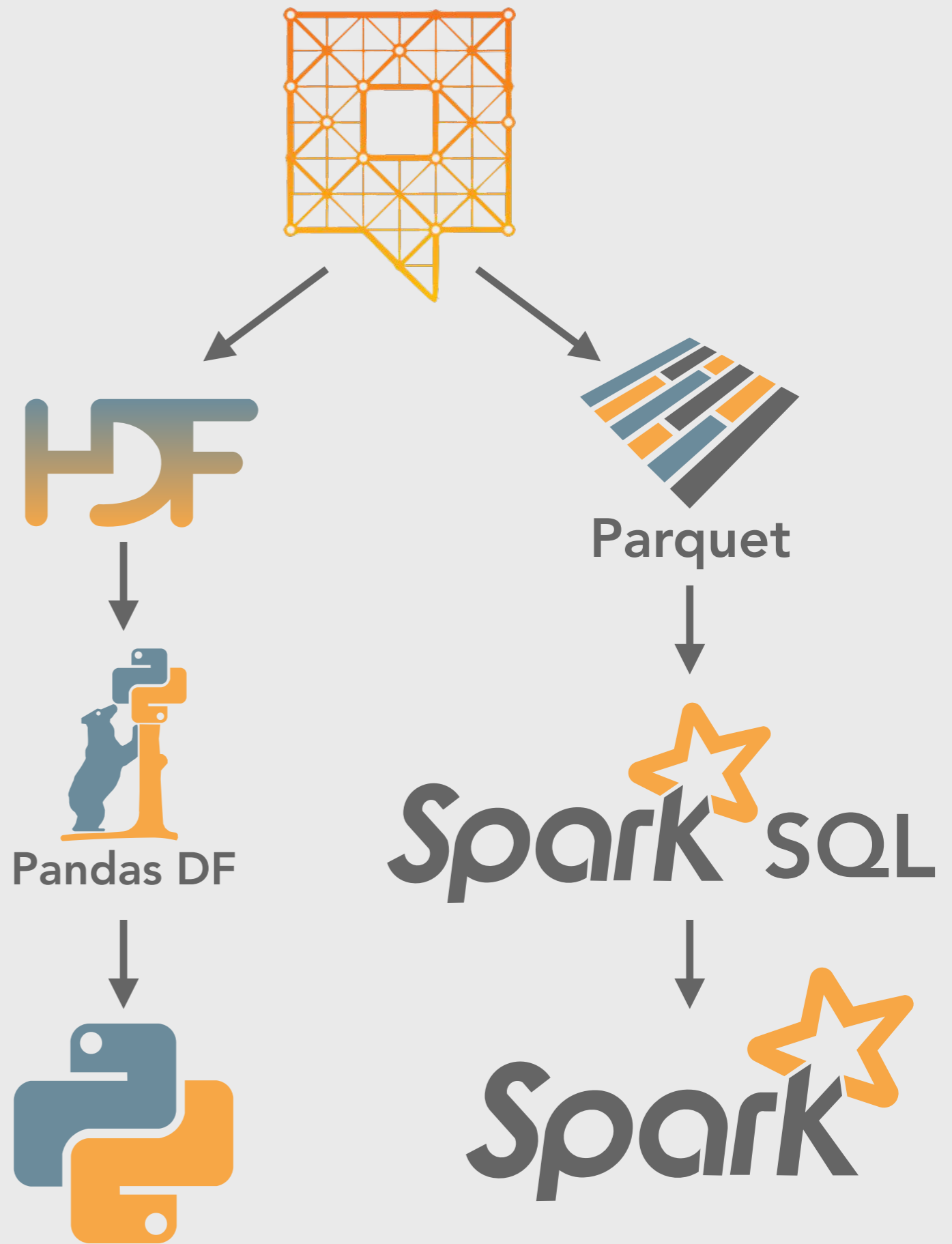


Current client supports Python on single machine

Extend client support to Spark



Quiltdata

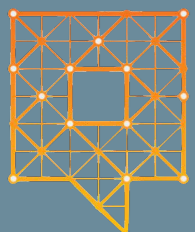


# Understand

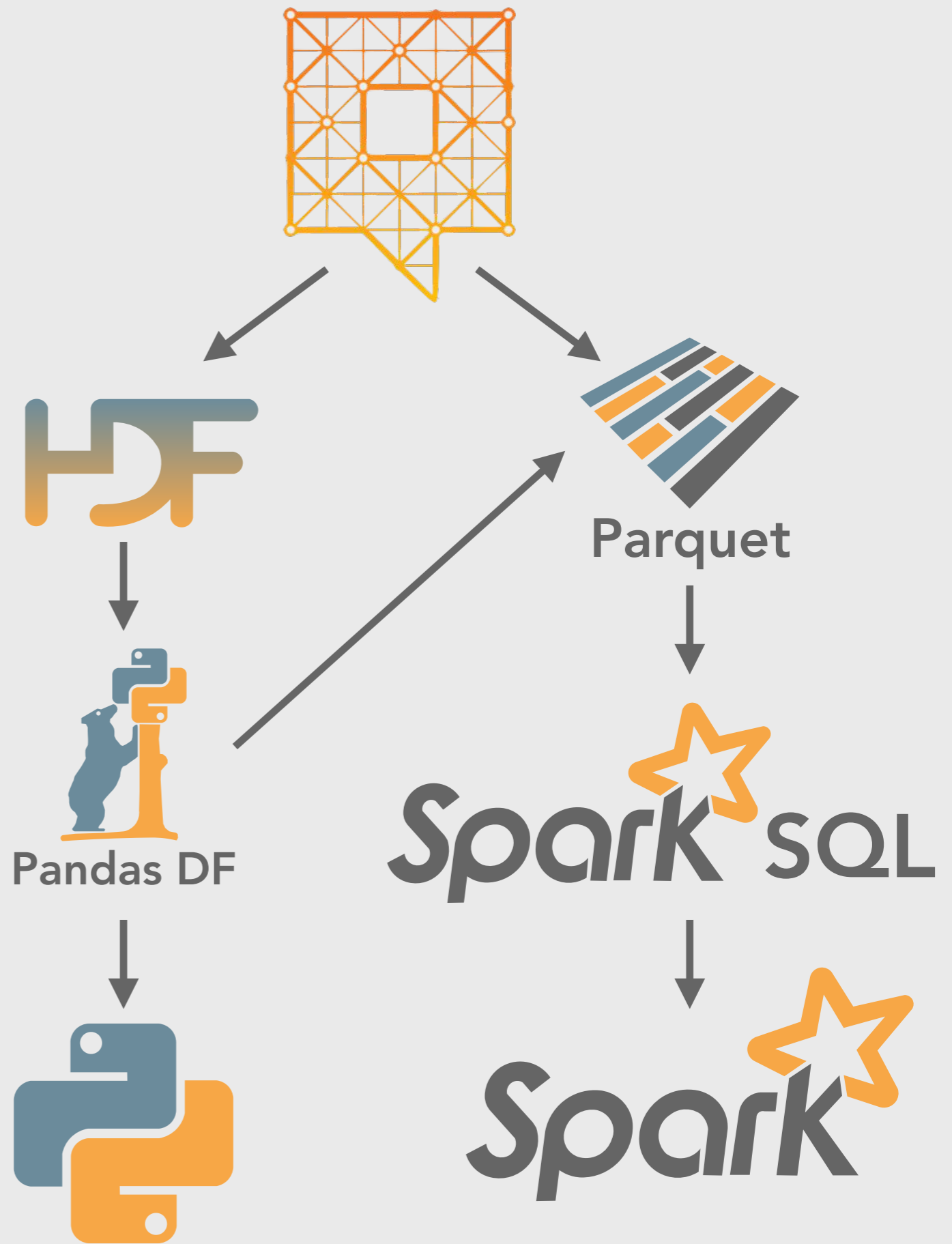


Current client supports Python on single machine

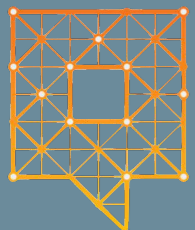
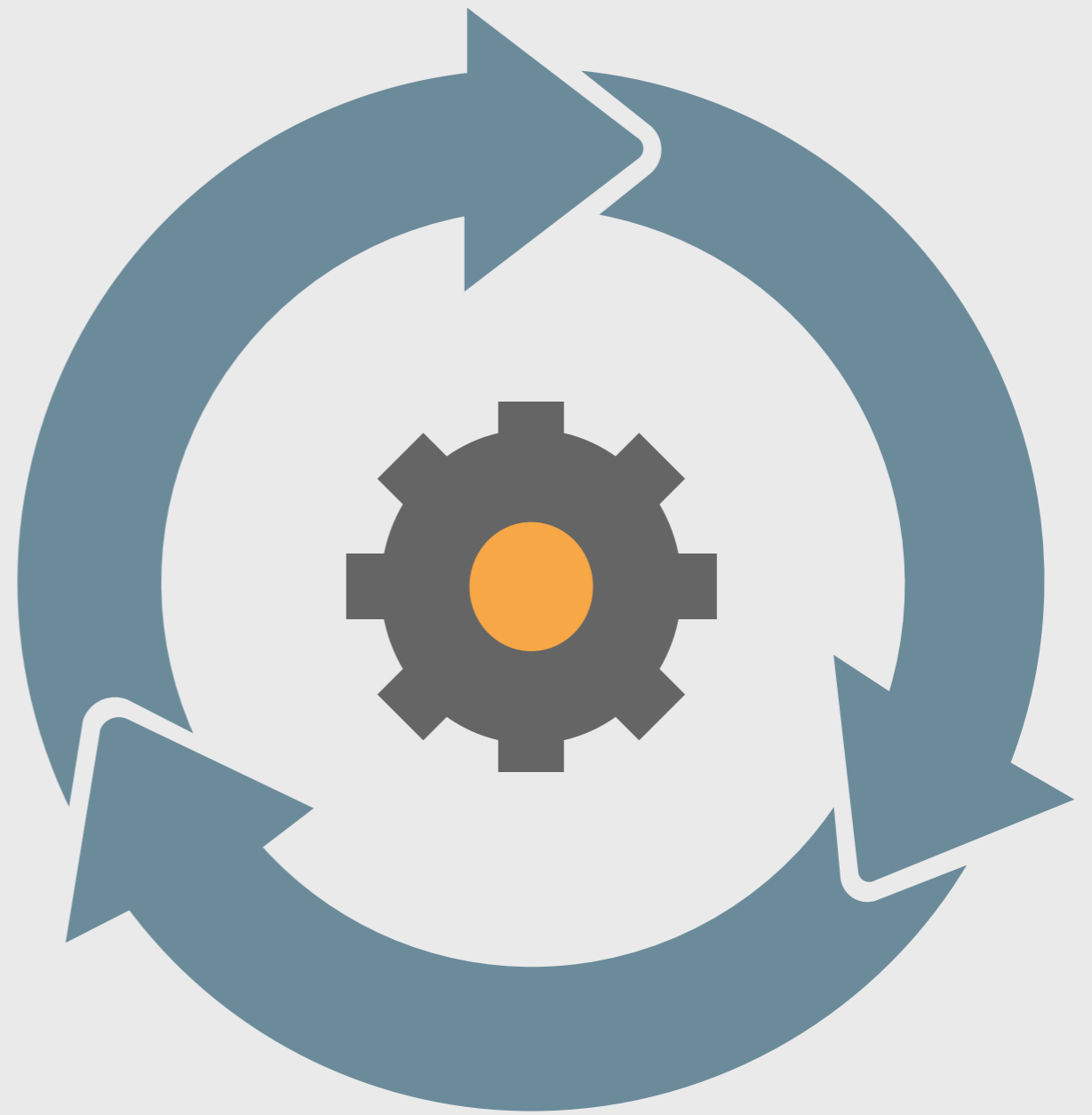
Extend client support to Spark



Quiltdata



# Implement



Quiltdata

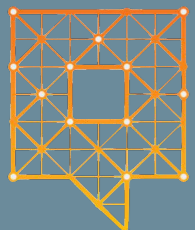
# Implement



High Level Build



Build Code



Quiltdata

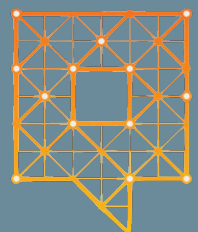
# Implement



High Level Build

Build Consults YAML

Choses Store



Quiltdata



Build Code



HDF5 Store



Filesystem



Deserialized  
Pandas

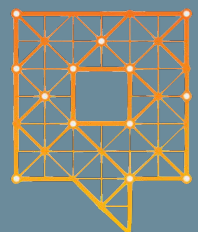
# Implement



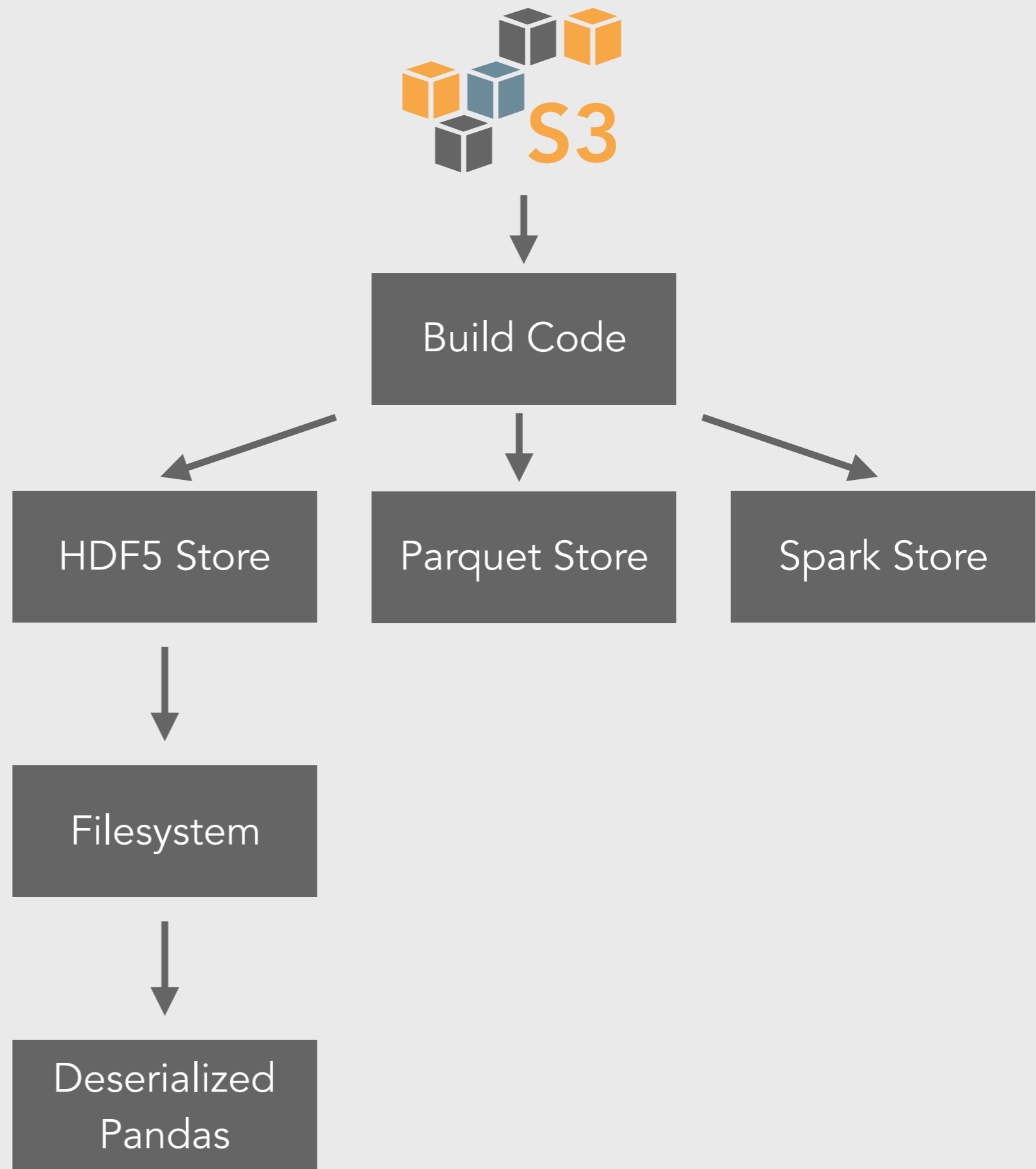
High Level Build

Build Consults YAML

Choses Store



Quiltdata



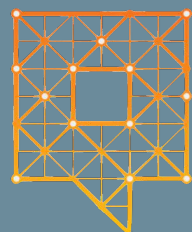
# Implement



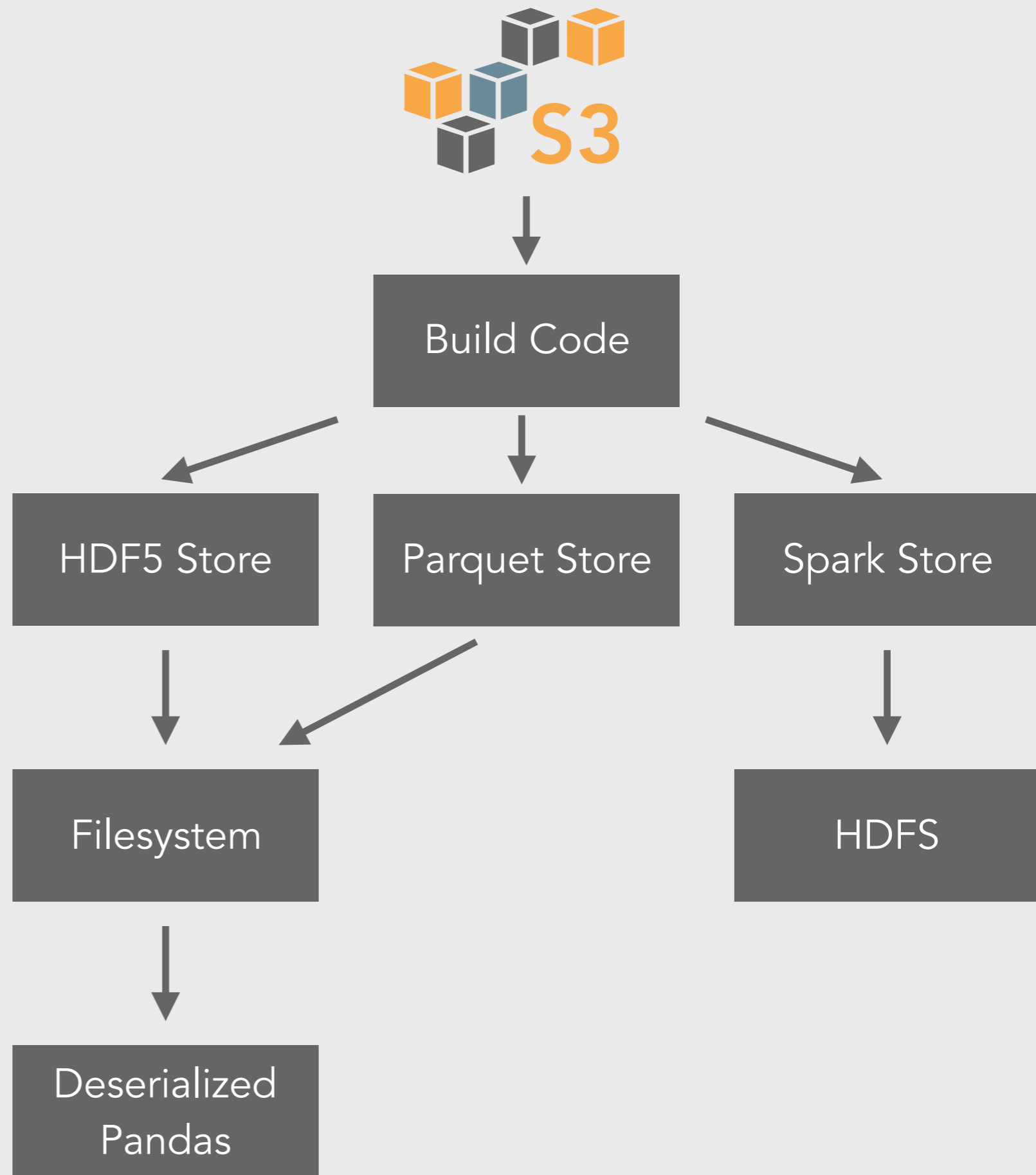
High Level Build

Build Consults YAML

Choses Store



Quiltdata





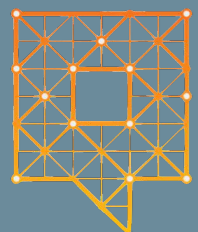
# Implement



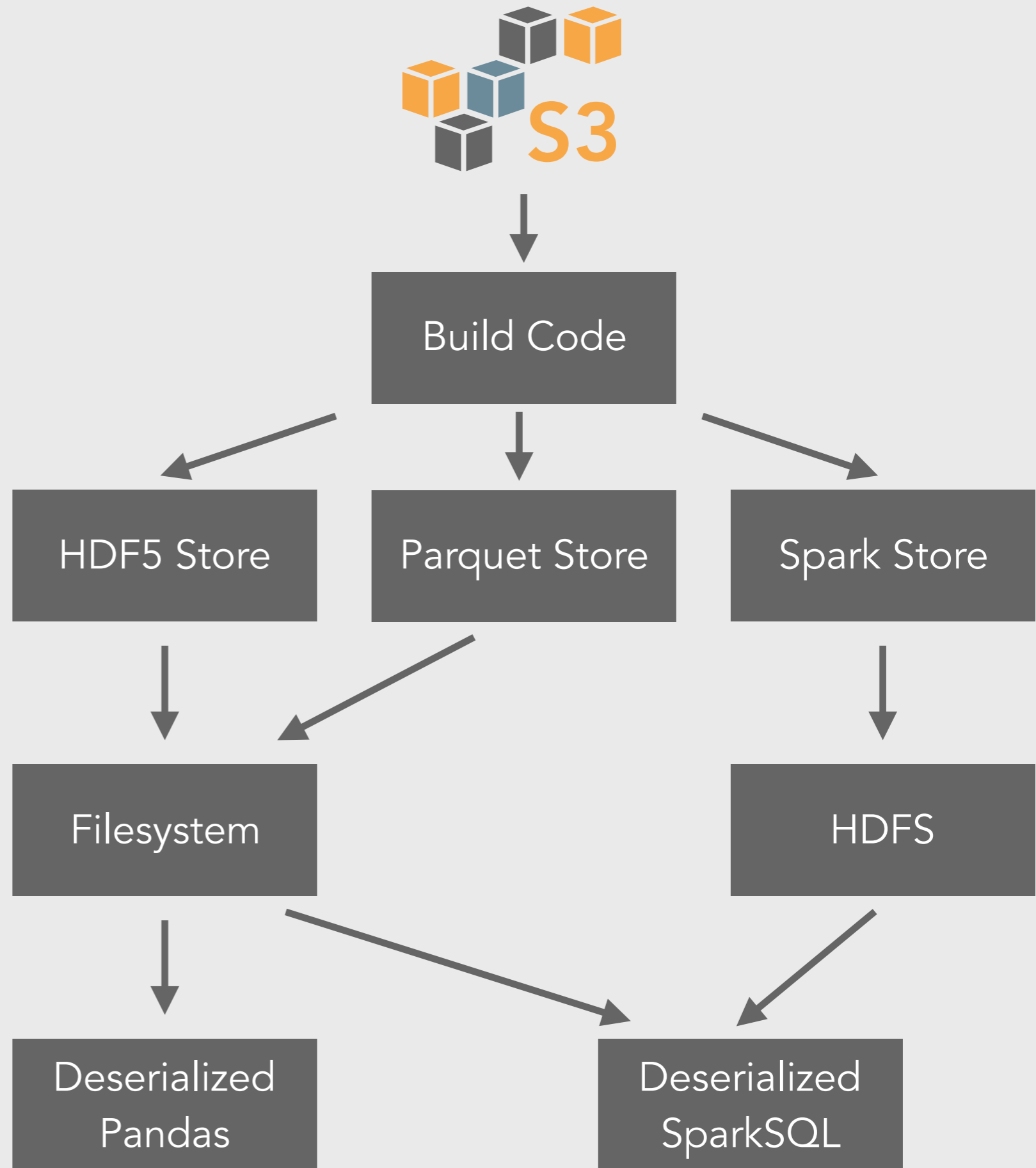
High Level Build

Build Consults YAML

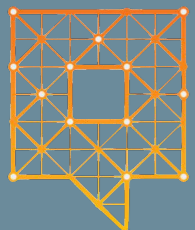
Choses Store



Quiltdata



# Test & Dist.



Quiltdata

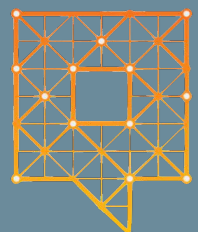
# Test & Dist.



Tested on various  
platforms



Amazon EC2



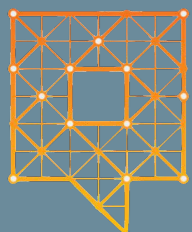
Quiltdata

# Test & Dist.



Tested on various  
platforms

Multiple Versions



Quiltdata



Amazon EC2



1.6 & 2.1

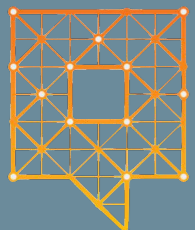
# Test & Dist.



Tested on various  
platforms

Multiple Versions

Changes Pulled into  
Master



Quiltdata



Amazon EC2



1.6 & 2.1

SparkPackageStore #4

Merged

kevinemoore merged 11 commits into `quiltdata:master` from `rinman24:master`



# Rich Inman

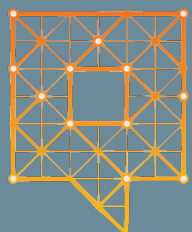
PhD in Mechanical Engineering UCSD

I Love to...

Hike Up Mountains

Travel

San Francisco Sports



Quiltdata



UC San Diego

