



# Spark Compute as a Service @ Paypal

Prabhu  
Paypal, Inc.

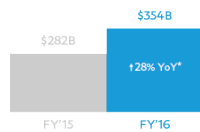


# Scale

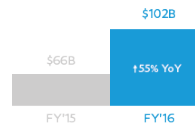
# Paypal Scale

## Business

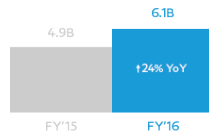
- One of the world's largest internet payment companies
- 203+M active accounts on 200 markets around the world
- PayPal platform includes Braintree, Venmo, Paydiant, PP Credit and Xoom



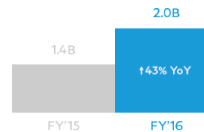
**\$354B**  
Total Payment  
up 28% YoY



**\$102B**  
Mobile Payment  
up 55% YoY



**6.1B**  
Total Transactions  
up 24% YoY



**2.0B**  
Mobile Transactions  
up 43% YoY

# Paypal Scale

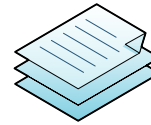
## Core Data Platform



**15+**  
Hadoop Clusters



**70+PB**  
Data



**40,000+**  
Yarn Jobs Per Day



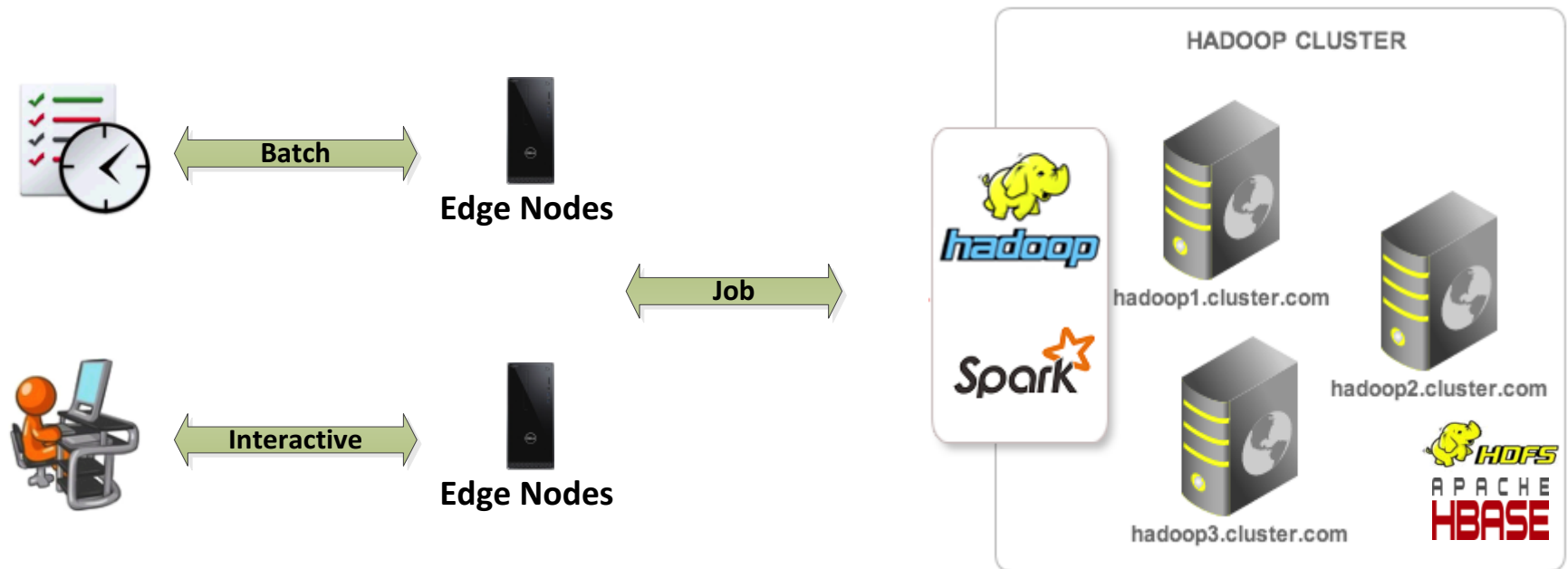
**5+**  
Compute



# Spark on Yarn

# Spark on Yarn

## Deployment

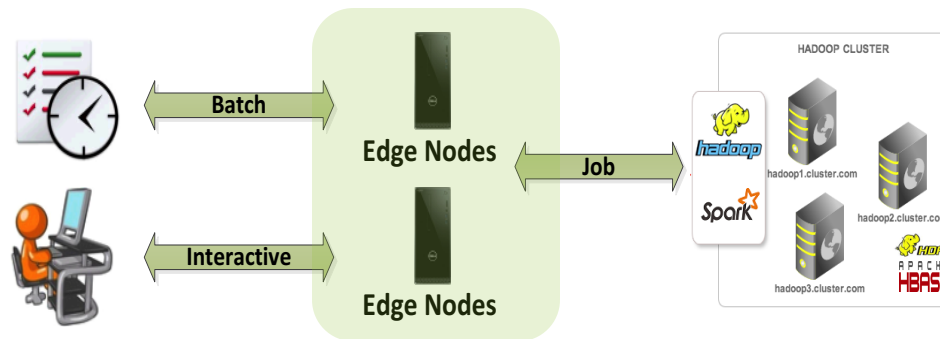




# Challenges

# Challenges

## Administrators

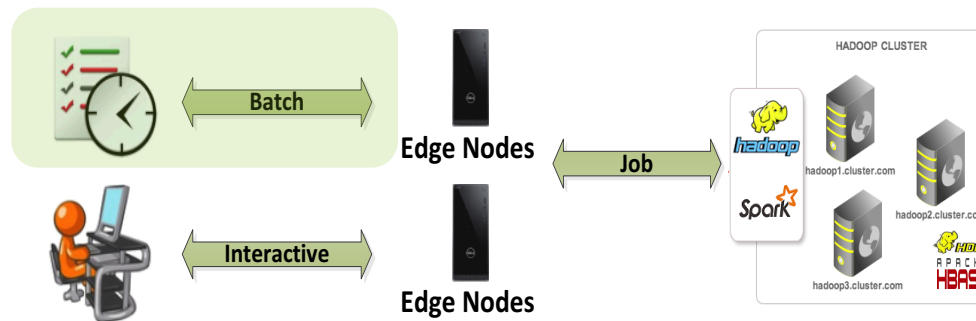


- Need extensive support and maintenance for CLI
- Need to deploy entire stack of software
- Need to sync configurations across systems
- Need extensive testing of jobs before any upgrade



# Challenges

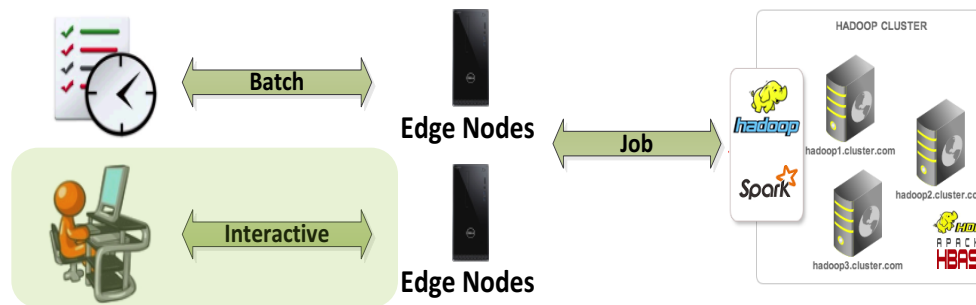
## Developers



- No REST-friendly
- No low-latency/sub-seconds execution
- No cache sharing across jobs
- No modularity and easy-restartability

# Challenges

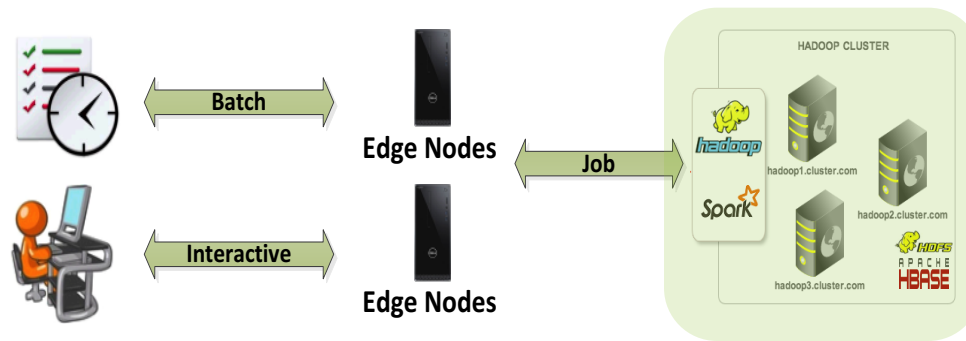
## Analysts/Scientists



- No easy way of interactive applications
- No multi-tenancy support and private workspace
- No direct spark sql execution
- No Kerberos integration

# Challenges

## Operations/Security



- Different ways of jobs execution and coding standards
- No uniform logging, monitoring and alerting
- Limited audit and control
- No statement level history or metrics

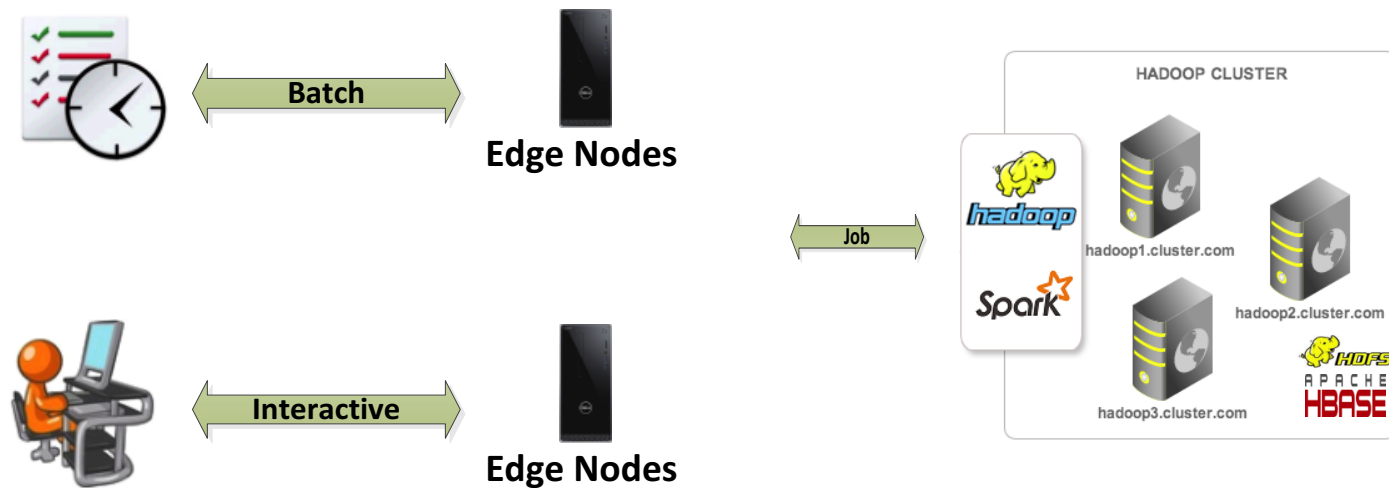


# Building SCaaS

Spark Compute Platform

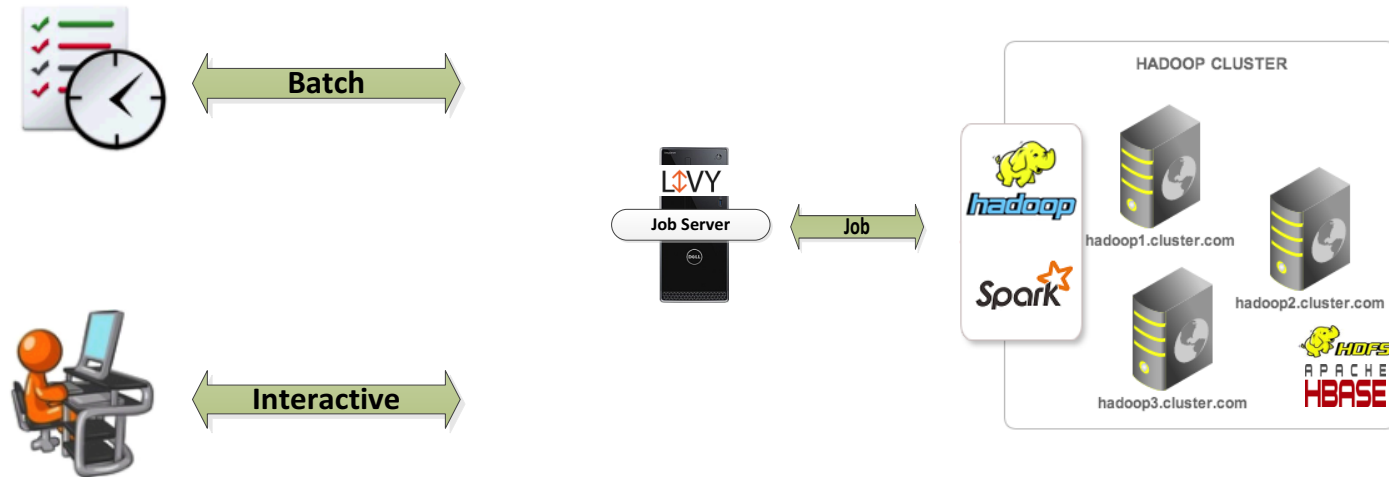
# Building SCaaS

Where we started!



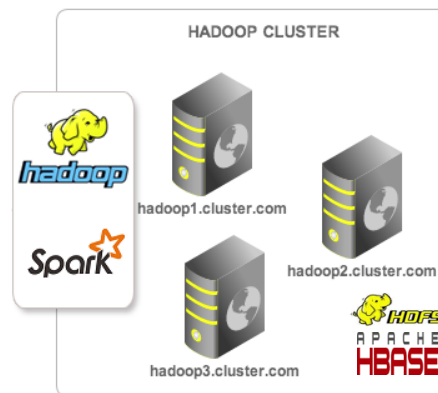
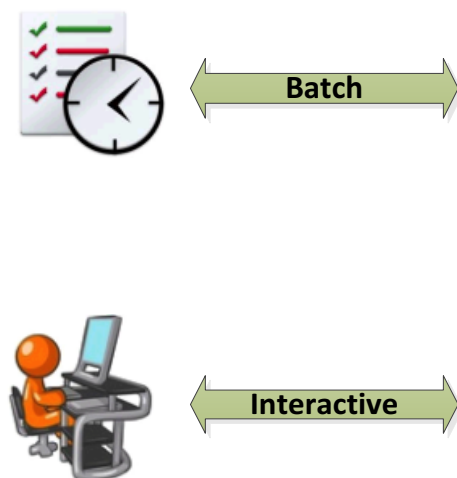
# Building SCaaS

## Adding REST Job Server



# Building SCaaS

Adding HA and Enhance Livy

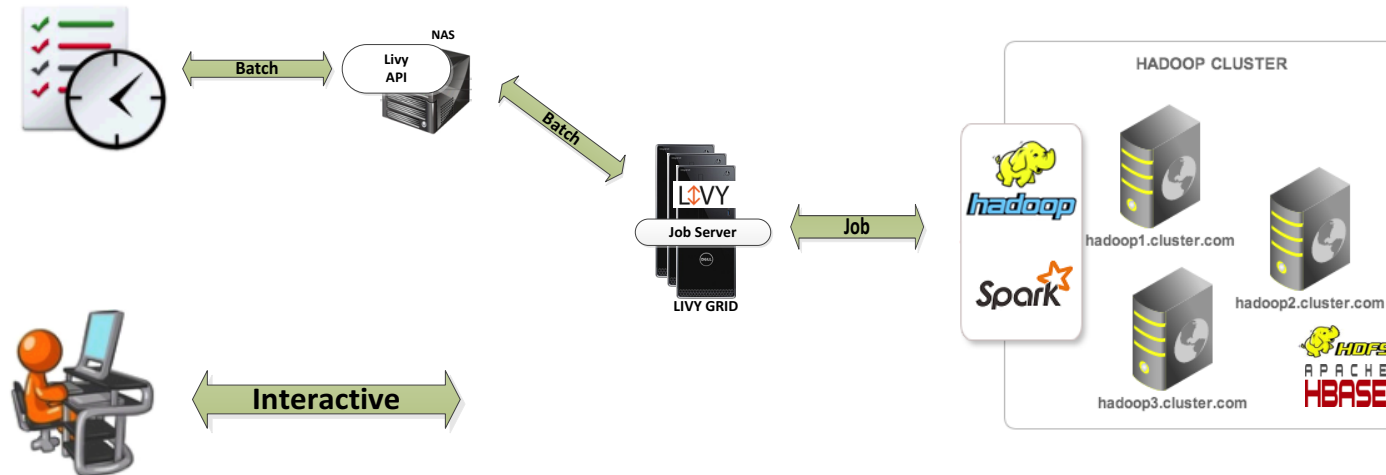


## PayPal Livy Version

- ✓ Multi-Nodes High Availability
- ✓ Kerberos Authentication Changes
- ✓ SQL Interpreter
- ✓ Session Manager Enhancements
- ✓ Session GC Improvements
- ✓ Plug-in Logger
- ✓ Yarn Poll Re-architecture
- ✓ Multiple Spark Versions Support
- ✓ White/Black list User Authentication
- ✓ Dockers
- ✓ Hbase Support
- ✓ Flink/Beam Support

# Building SCaaS

## Adding Livy API and Utilities



### Batch Utilities

- ✓ `startSparkBatch`
  - ✓ `stopSparkBatch`
  - ✓ `listSparkBatch`
- ✓ `startSparklingWater`
  - ✓ `stopSparklingWater`
- ✓ `startSparkSql`
  - ✓ `stopSparkSql`
- ✓ `startSparkSession`
  - ✓ `execSparkFile`
  - ✓ `execSparkCode`
  - ✓ `stopSparkSession`
  - ✓ `listSparkSession`
- ✓ `livy-spark.jar`

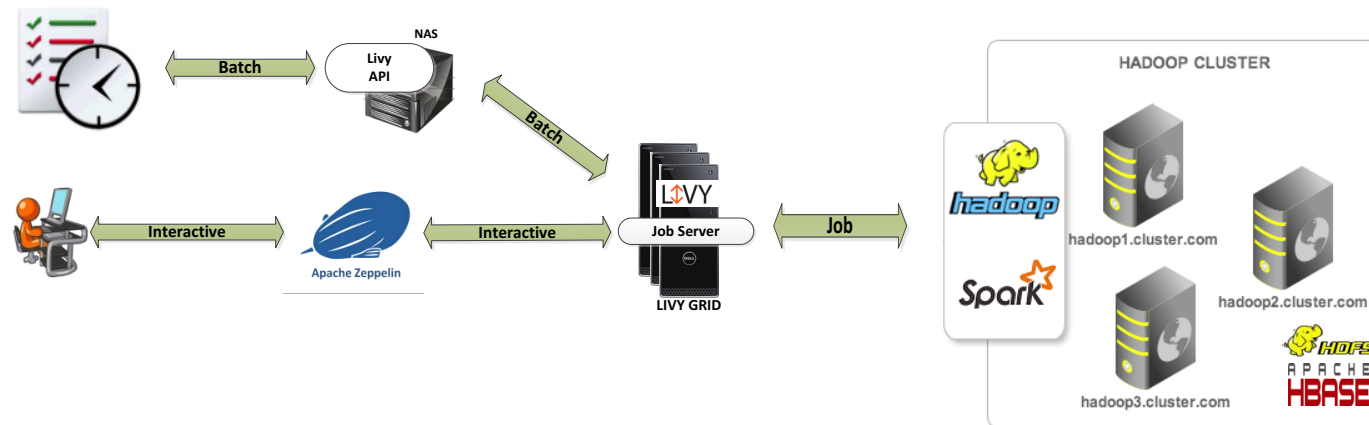
### Interactive Utilities

- ✓ `livy-spark`



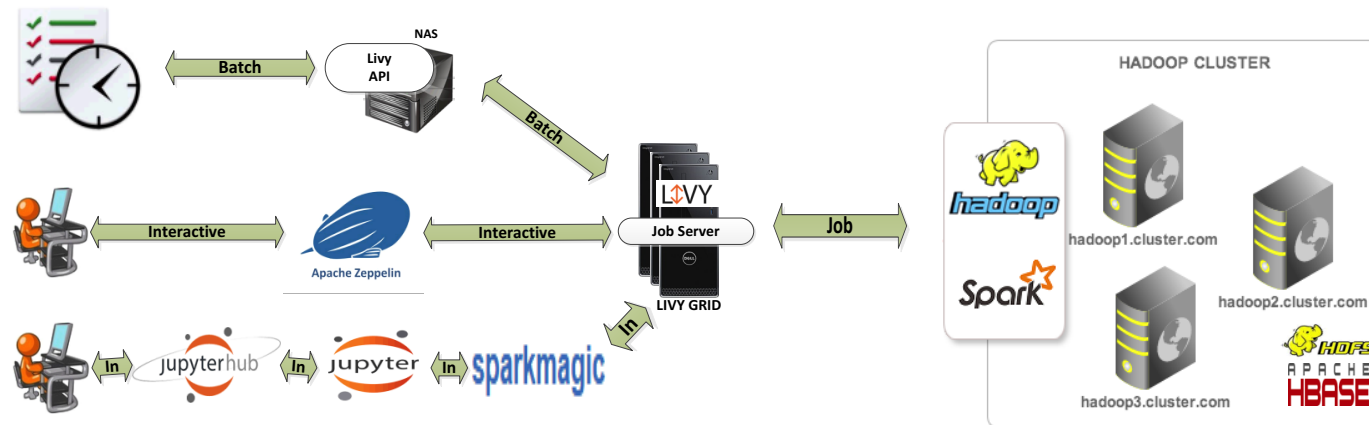
# Building SCaaS

## Adding Zeppelin



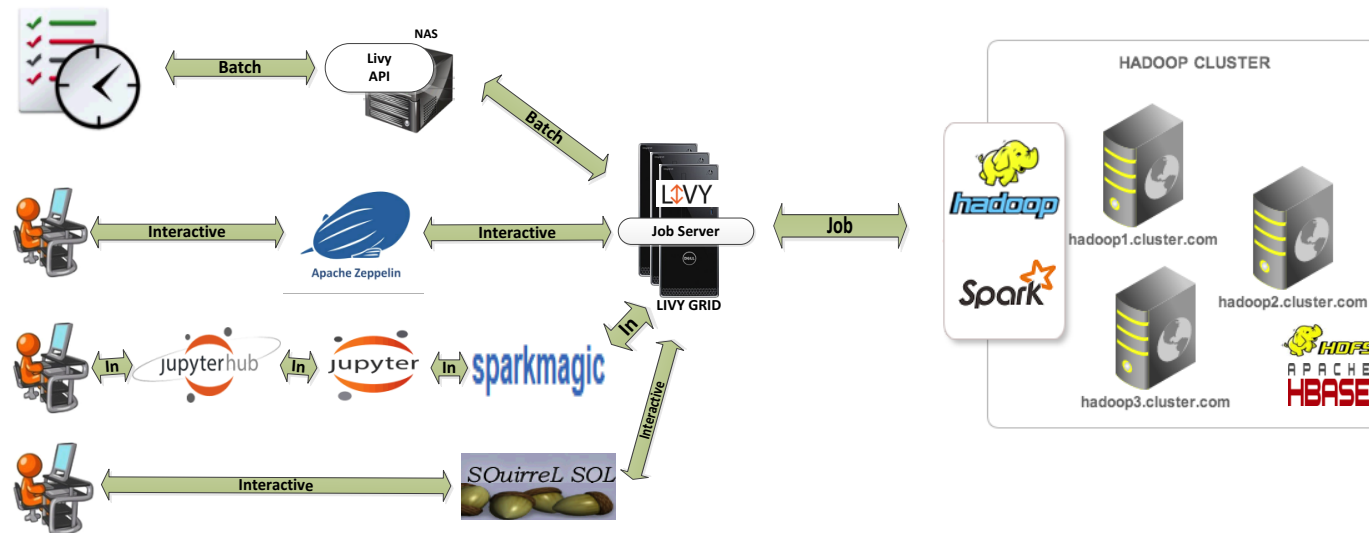
# Building SCaaS

## Adding Jupyter



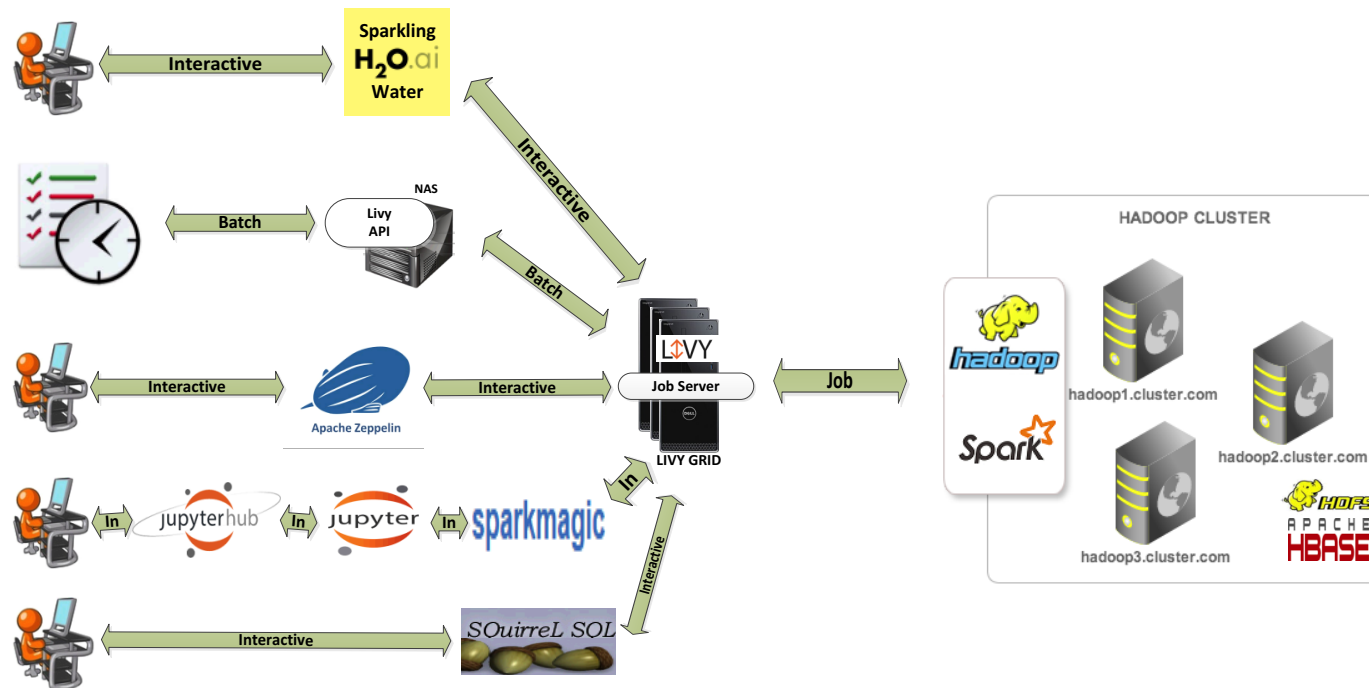
# Building SCaaS

## Adding SQL Client



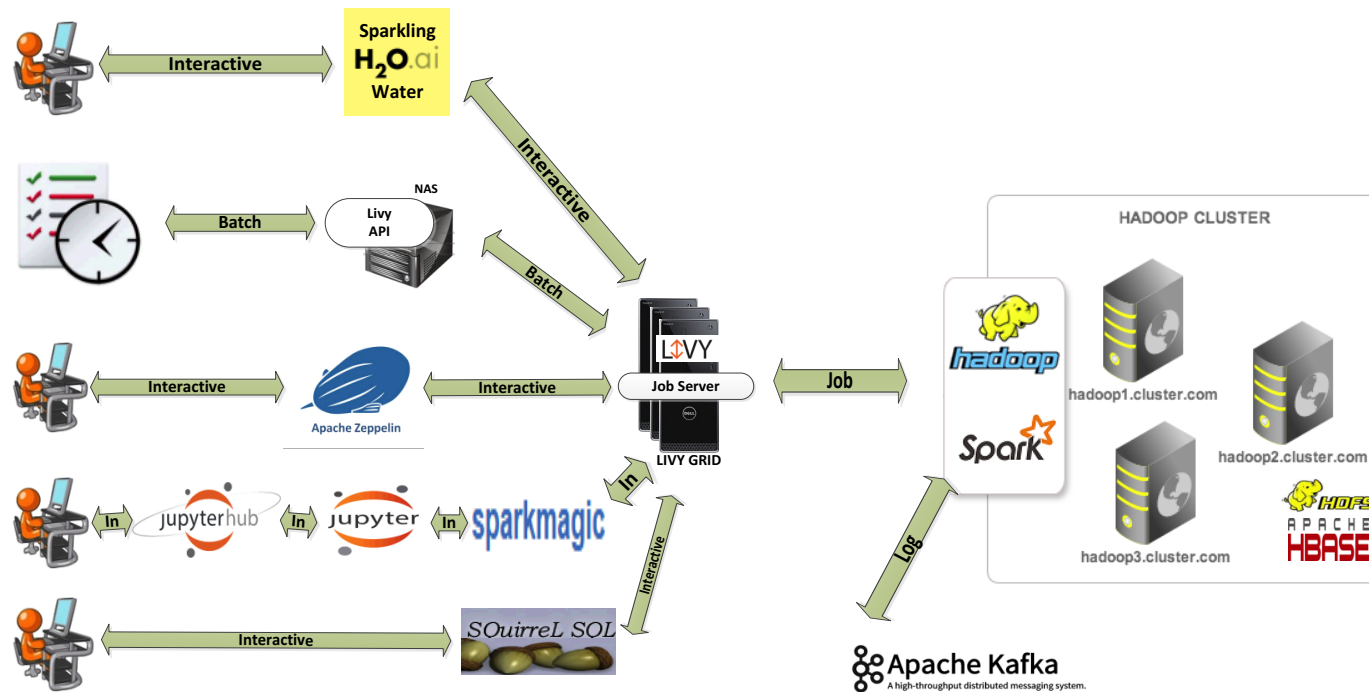
# Building SCaaS

## Adding Sparkling Water



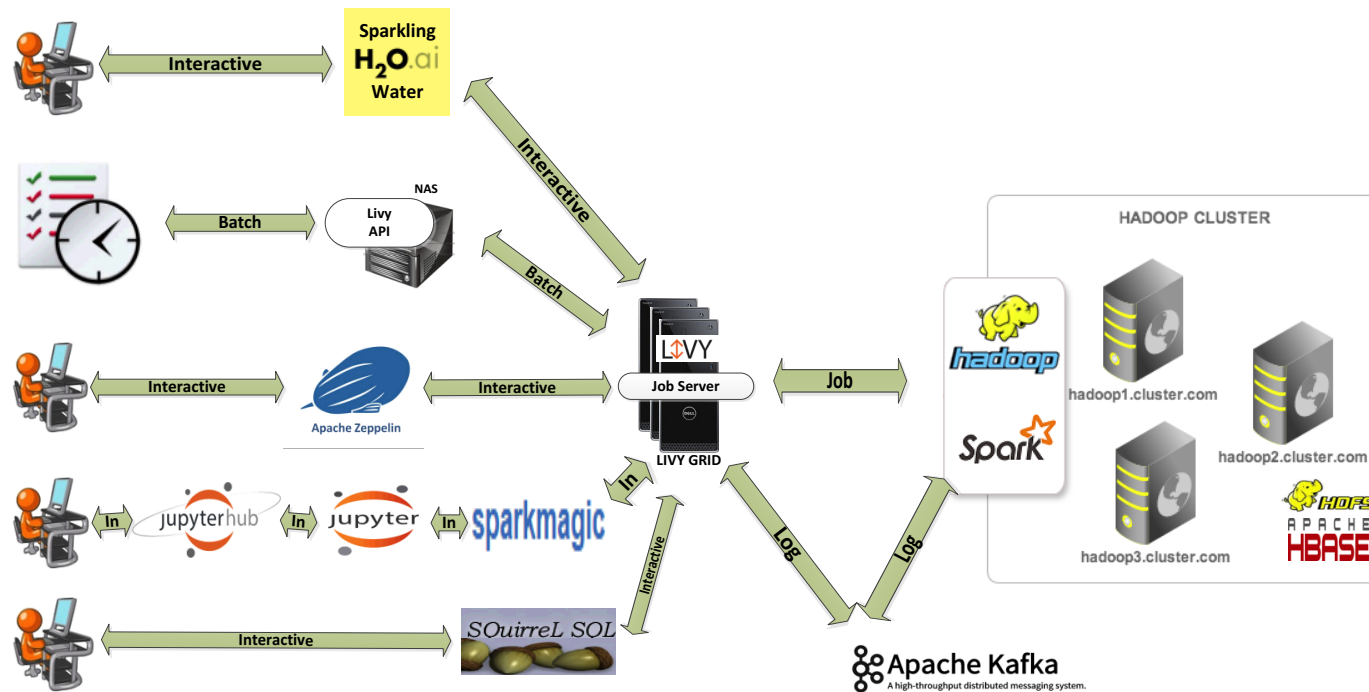
# Building SCaaS

## Adding Spark Logger



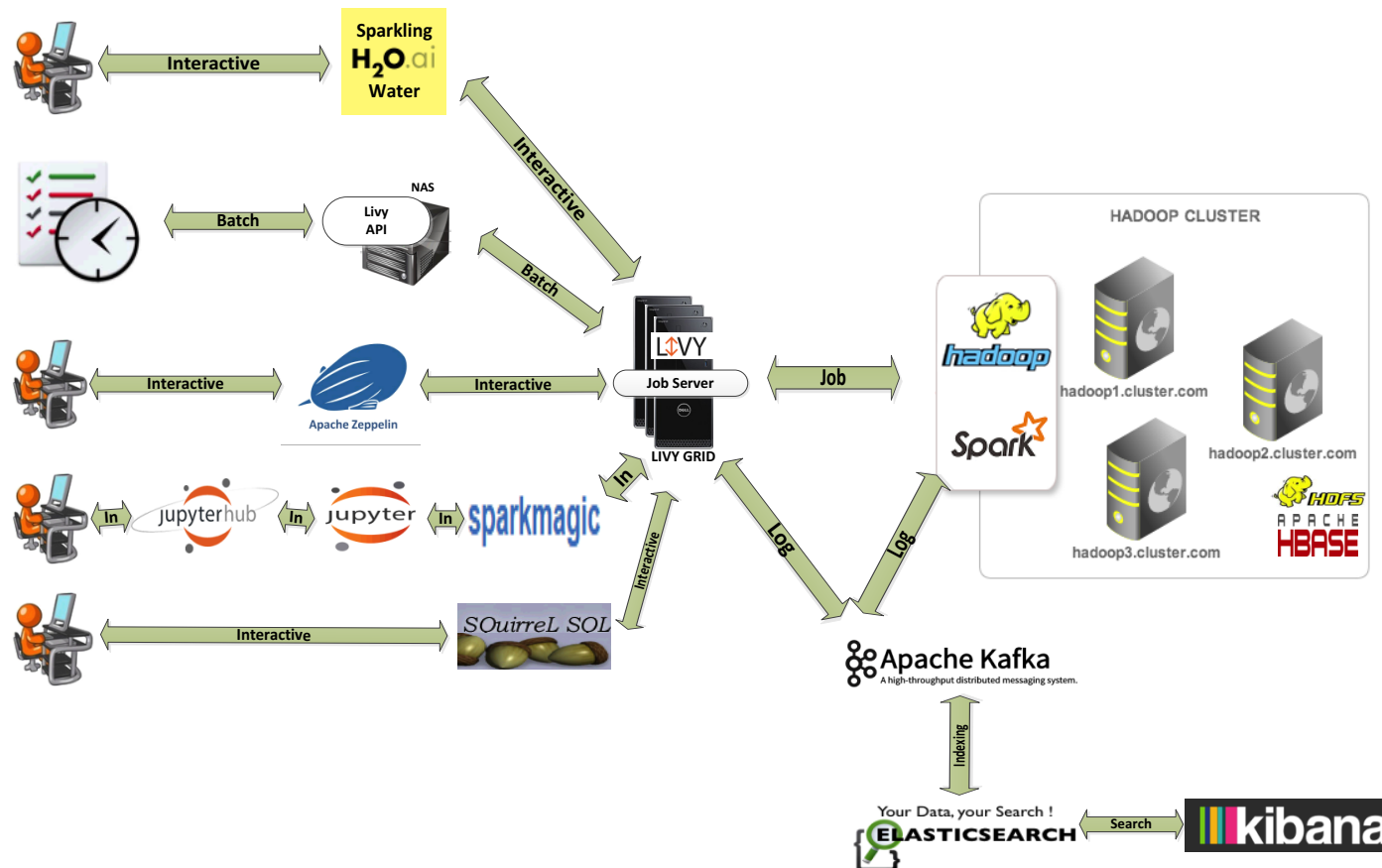
# Building SCaaS

## Adding Livy Logger



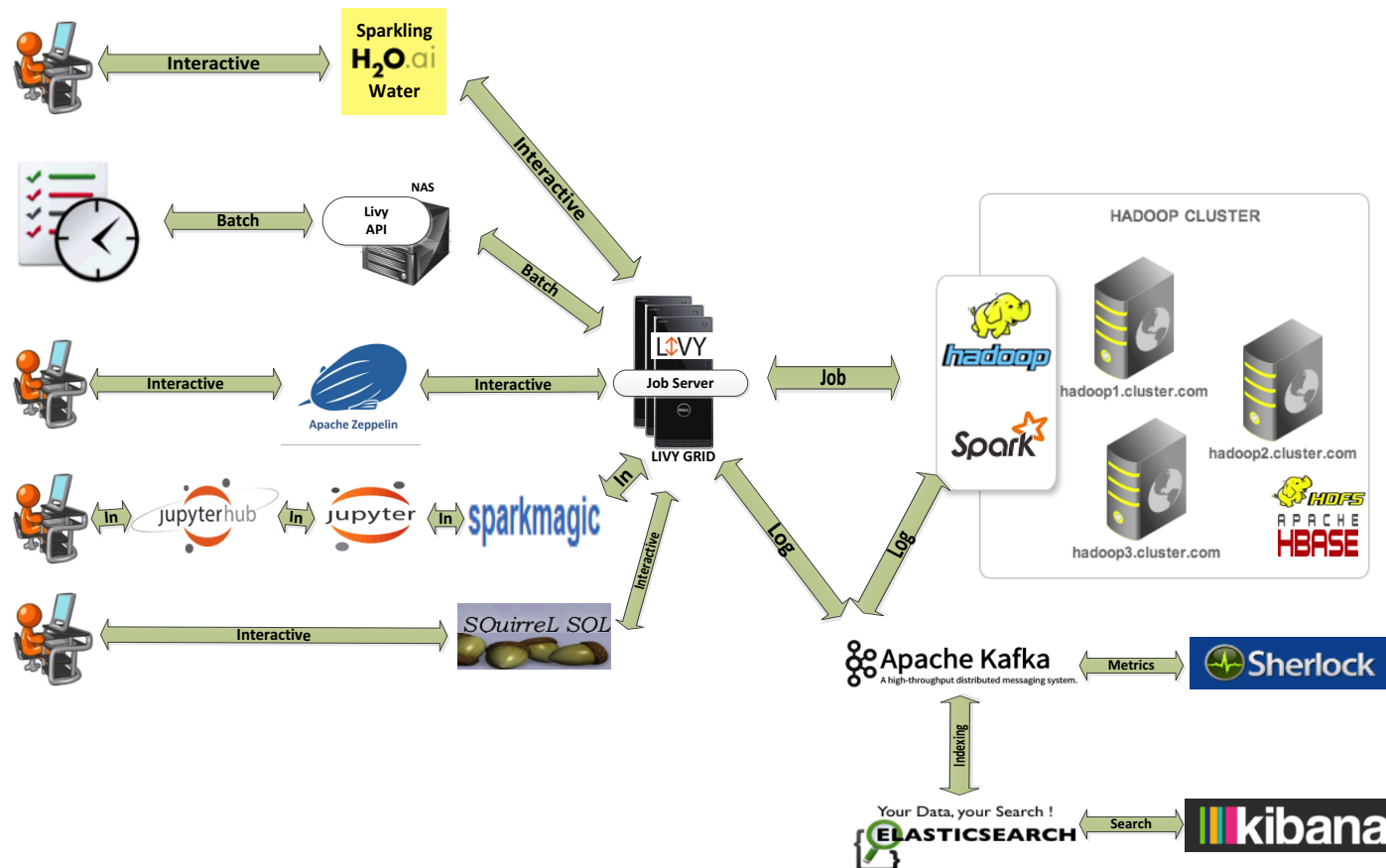
# Building SCaaS

## Adding Search Engine



# Building SCaaS

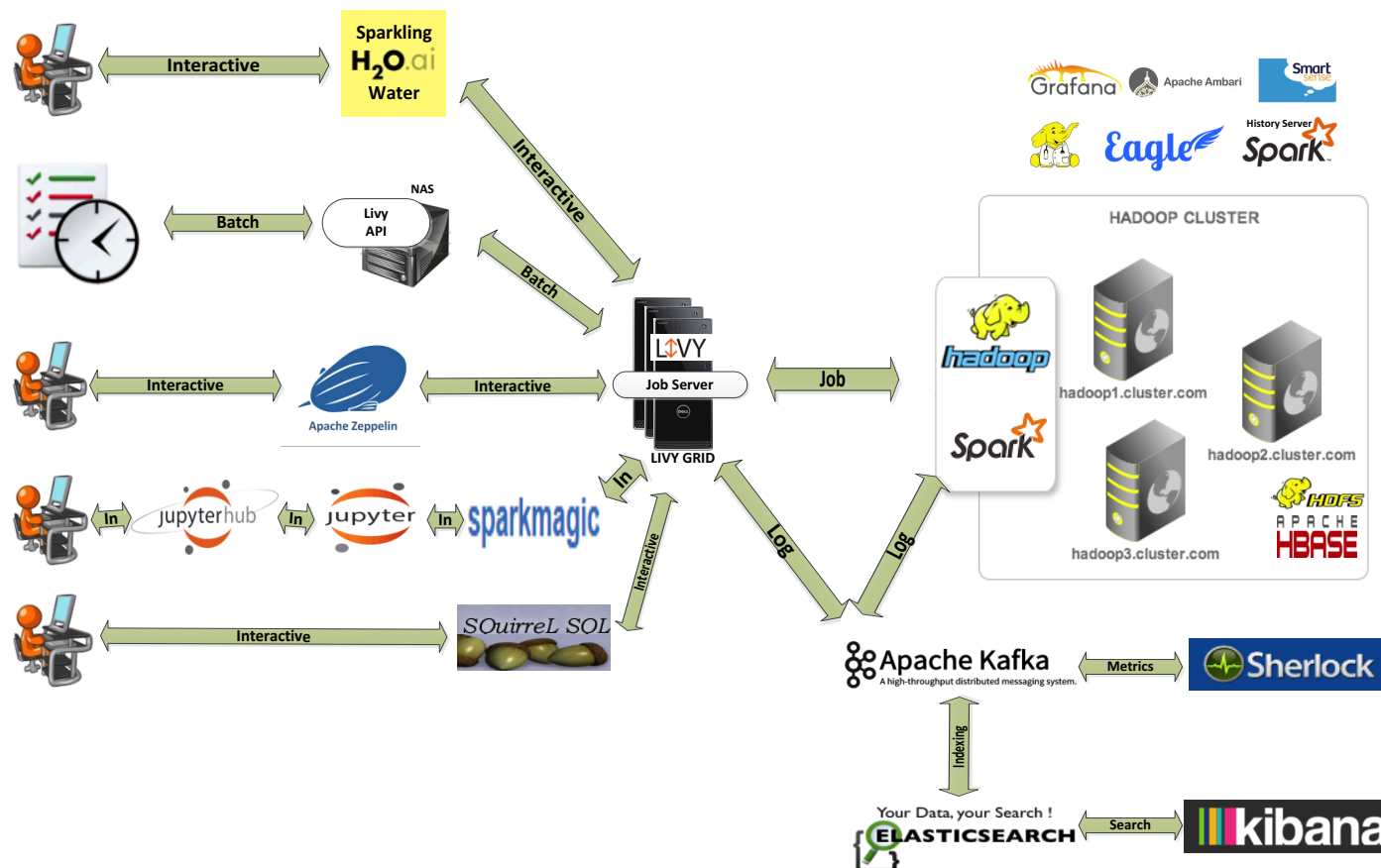
## Adding Monitoring/Alerting





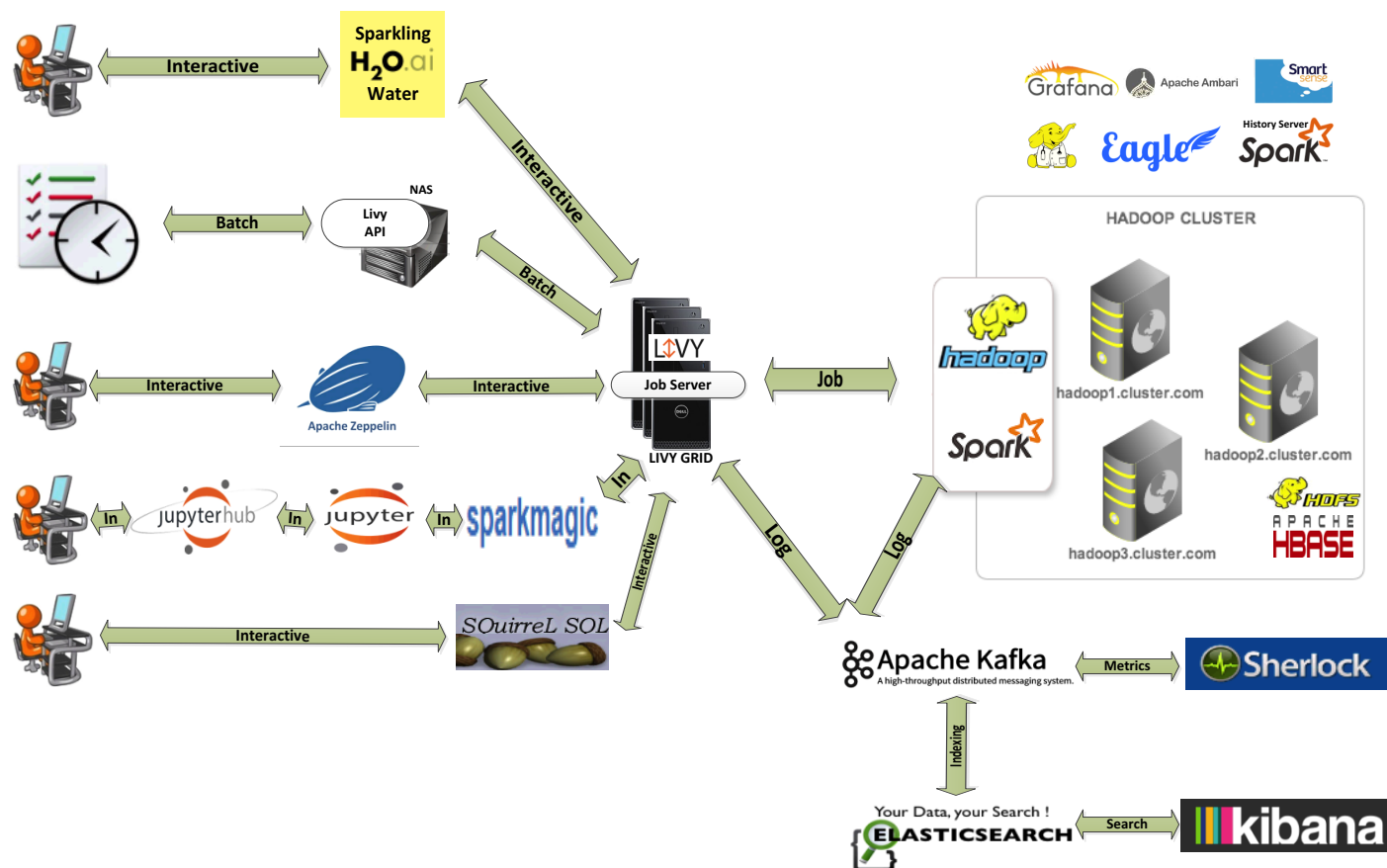
# Building SCaaS

## Adding Standard Tools



# Building SCaaS

Here, you go! The SCaaS Framework



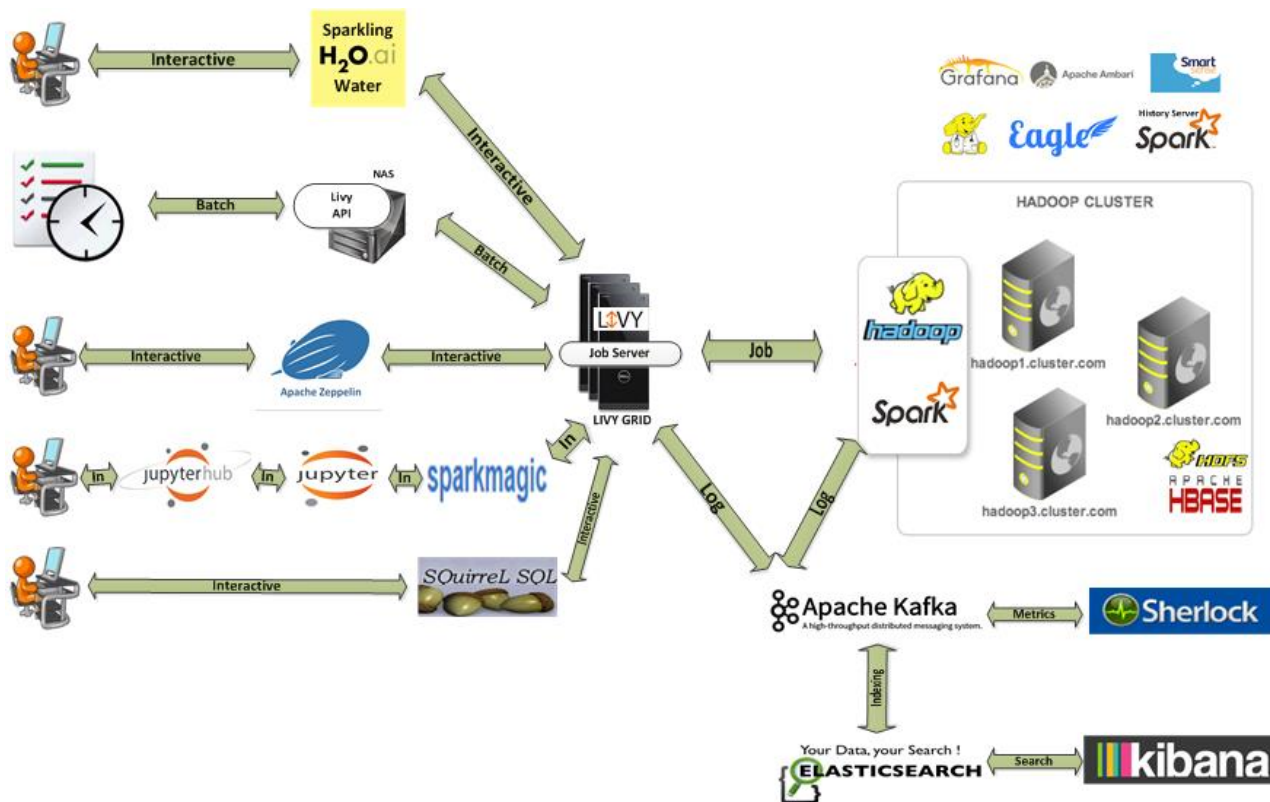


# SCaaS

## Benefits

# SCaaS

## Benefits



### Administrators

- ✓ Less maintenance on CLI
- ✓ Deploy software stack only on Job Server
- ✓ Configurations at one place
- ✓ Easy platform/software upgrade

### Developers

- ✓ REST-friendly and Docker-friendly
- ✓ Low-latency/sub-seconds execution
- ✓ Sharing cache across jobs
- ✓ Modularity and easy restartability

### Analysts/Scientists

- ✓ User friendly interactive applications
- ✓ Multi-tenancy and Private workspace
- ✓ Direct spark sql execution
- ✓ Kerberos Support

### Operations/Security

- ✓ Standardized coding and unified execution
- ✓ Uniformed logging, monitoring and alerting
- ✓ Fine-grained audit
- ✓ Complete statement level history and metrics



# SCaaS Demo



# Follow Us

<https://www.paypal-engineering.com>

 @paypaleng

# Thank You