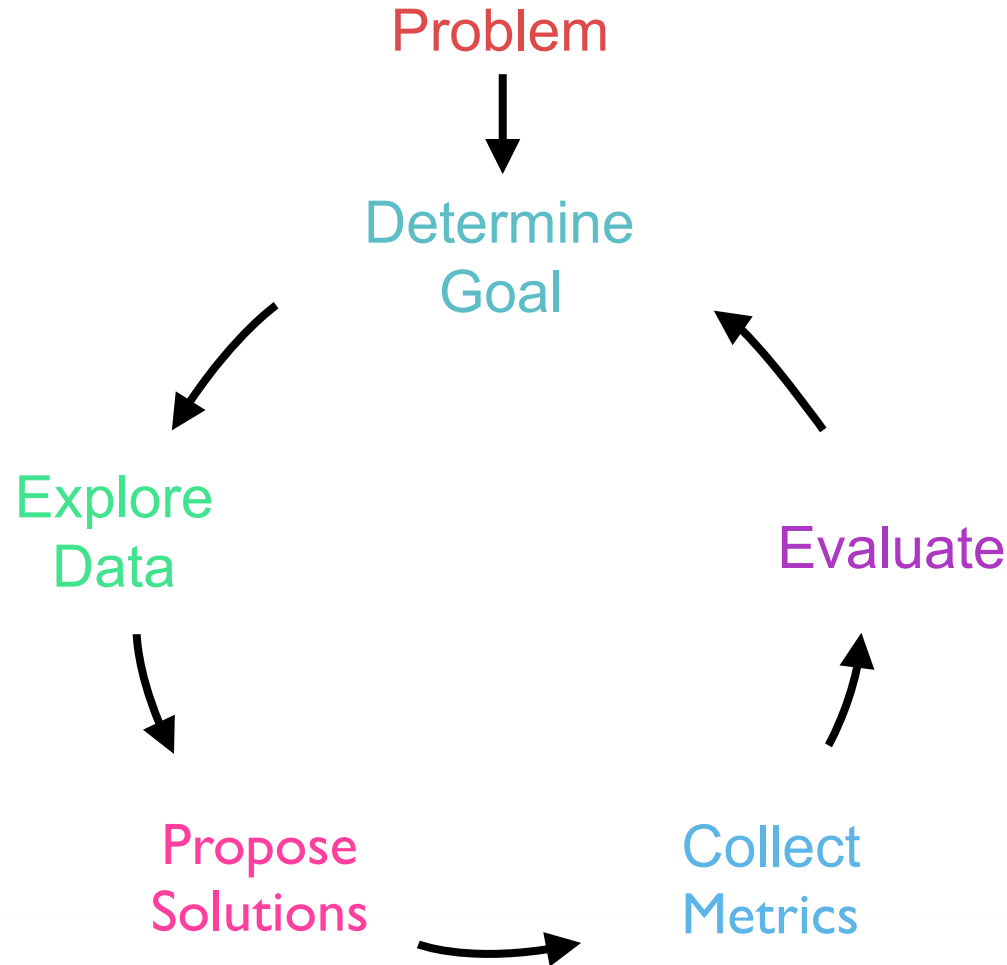


# Lesson 3: Your First Spark Application

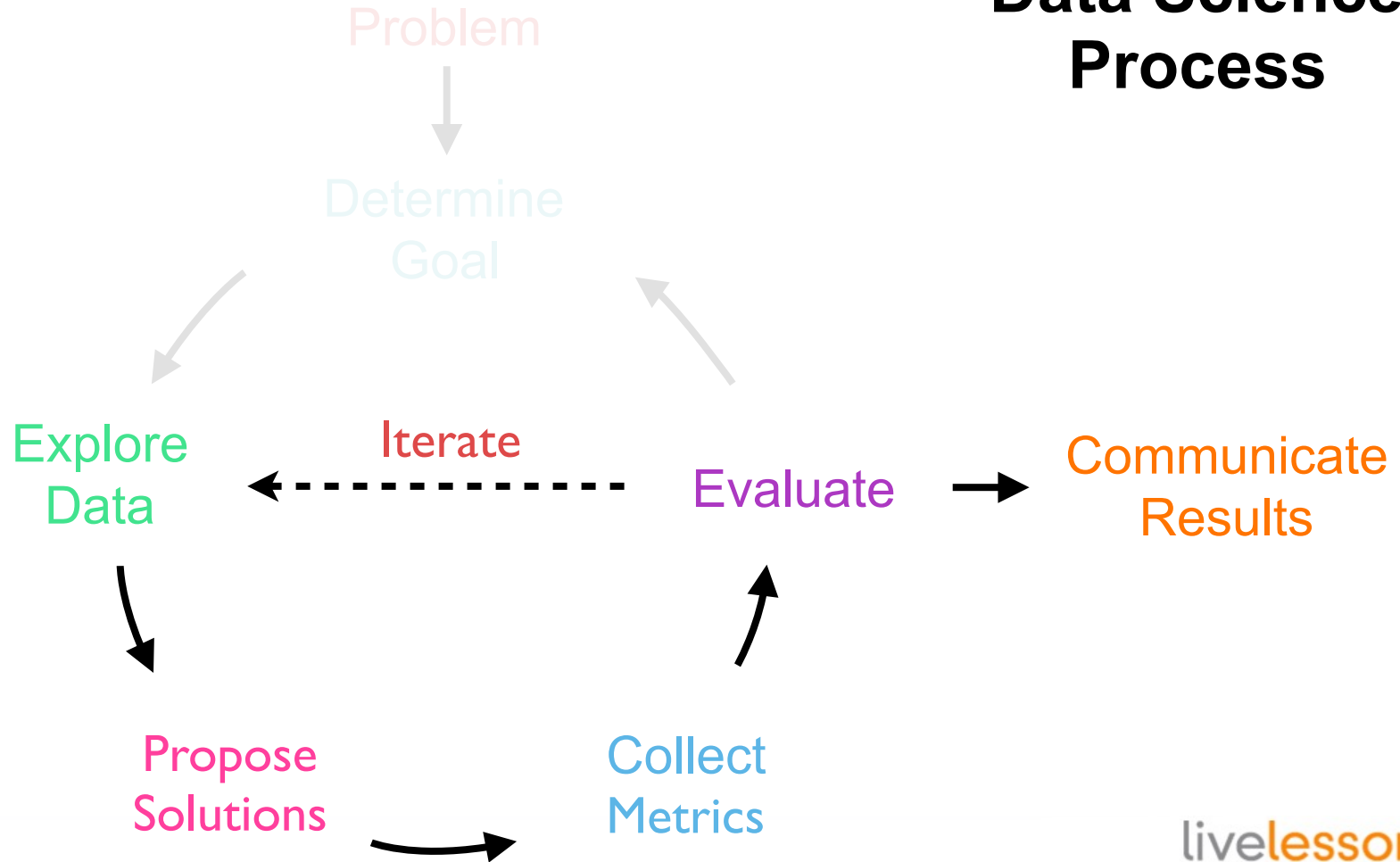
## 3.1 How Spark Fits into the Data Science Process



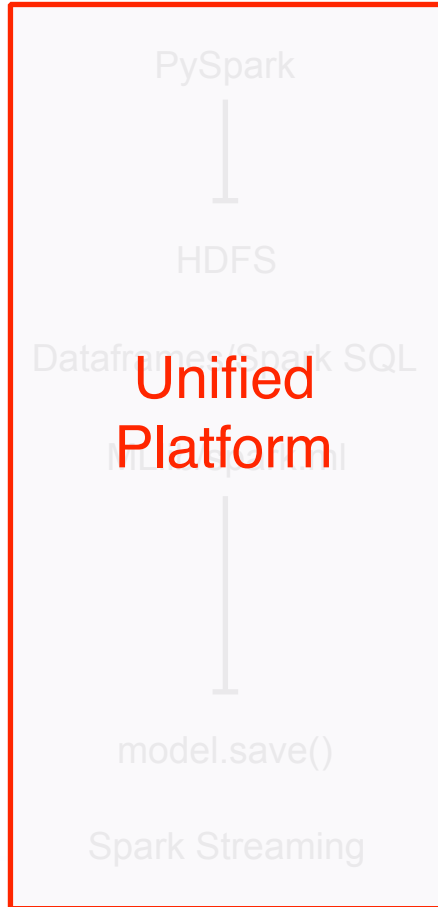
# Data Science Process



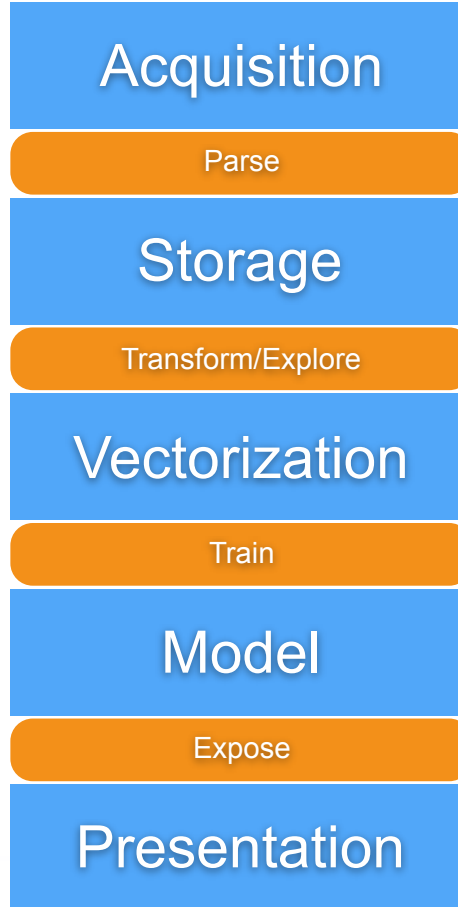
# Data Science Process



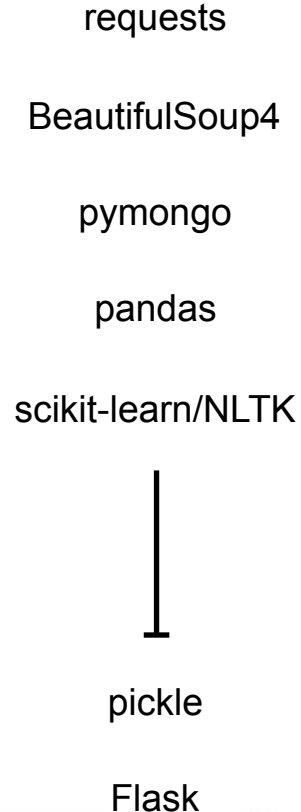
## At Scale



## Data Pipeline



## Locally



# Review

- The **data science** process is inherently **interactive**
- Spans many **scales** of data and computation
- Data **pipelines** require linking many **diverse tasks** (and data)
- **Quick** insights necessary for fast **iteration**

