

SPOT HERO

Data Discoverability

Maggie Hays

Senior Product Manager -- Data Services

DataHub Town Hall -- September 25, 2020

Agenda

- Overview of Teams & Data Stack
- Current State of Data Discoverability
- Data Catalog Evaluation
- DataHub POC - Hypotheses, Progress, and Next Steps
- Brief Demo!



SpotHero's Data-Focused Teams

Data Engineering

3 Engineers

SpotHero IQ

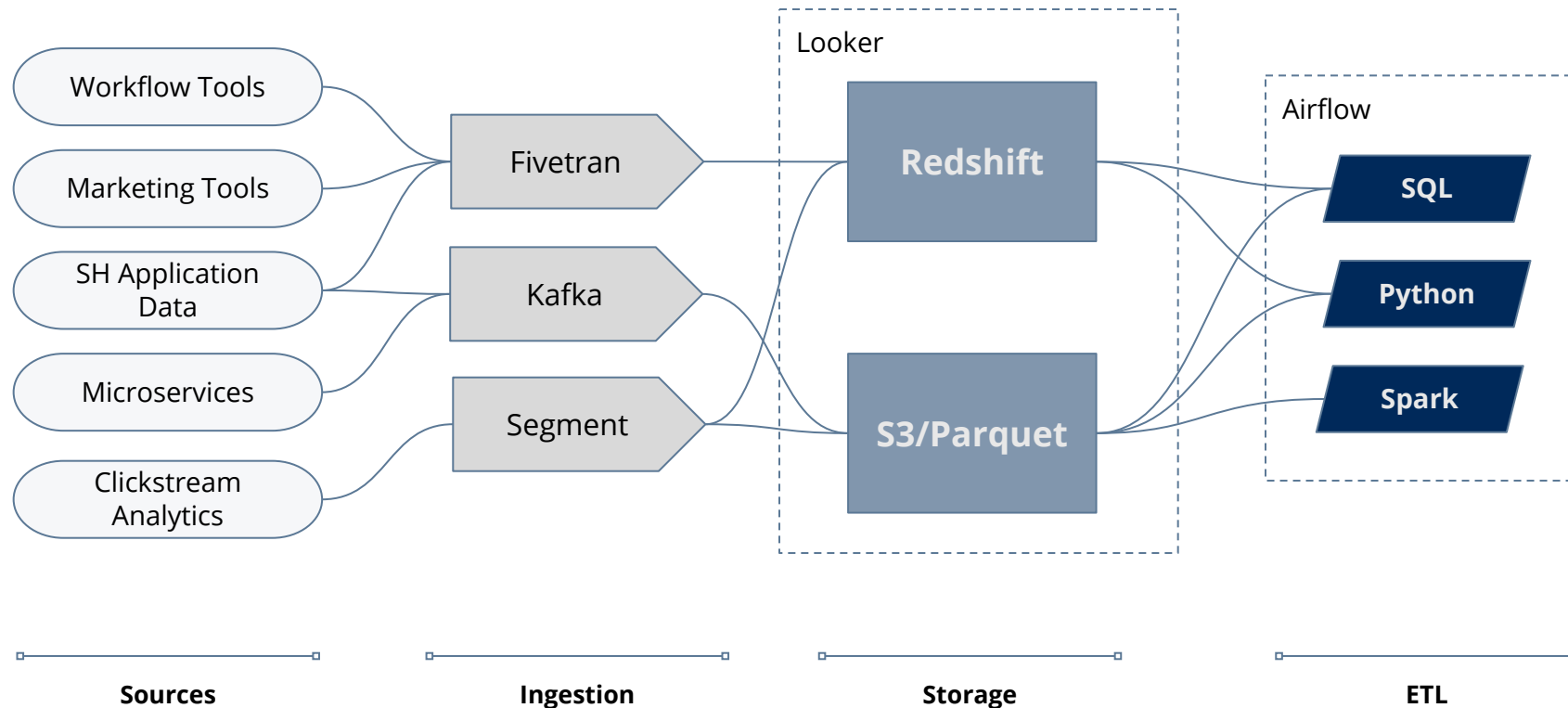
2 Engineers
1-3 Data Scientists
(We're hiring!!)

Analytics

5 Business Analysts



SpotHero's Data Stack



Current State of Data Discoverability

- 1 Data Lineage is difficult to discover and navigate, regardless of role or tenure**
 - Impact analysis is arduous; Engineers avoid breaking changes at all costs
 - Prolonged debugging/troubleshooting data issues
- 2 Difficult to discover what data exists and/or what it represents**
 - Reliance on tribal knowledge
 - Large burden on the Analytics team to answer any/all questions
- 3 Confidence in Data Accuracy is neutral, but room for improvement**
 - Once folks track down the data, they are relatively confident in its accuracy



Data Catalog Evaluation

	DataHub	Amundsen / Marquez	Apache Atlas	Alation
Ease of Integration	✓	🏆	✓	✓
Lineage Support	🏆	✗	✓	✓
Configurable Metadata	🏆	✓	✓	✓
Affordability	✓	✓	✓	✗



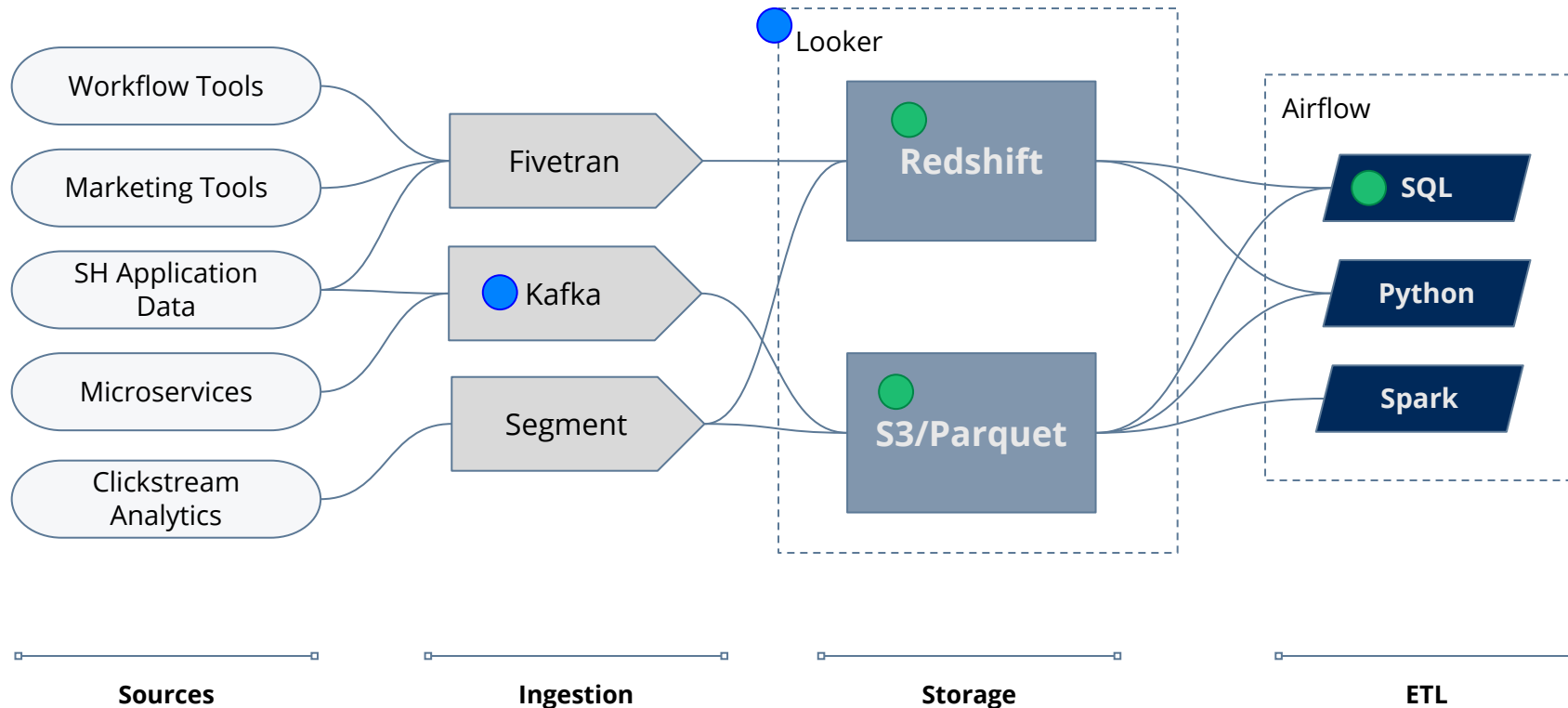
DataHub POC - Hypotheses

- 1 Increase visibility into what data exists, what it represents, and how it's used across the company
- 2 Decrease the effort required by SpotHero teammates to use and interpret data
- 3 Increase SpotHero teammates' confidence in the accuracy and/or relevance of data



DataHub POC - Progress & Next Steps

- Complete
- Q4 2020



Quick Demo!

