Building Fast, Scalable Machine Learning Pipelines

Vlad Giverts

Sr Director of Software Engineering, Workday


Identified

LinkedIn

Vlad Giverts

500+ connections

Sr Director of Engineering at Workday / Startup Investor & Advisor

San Francisco Bay Area | Internet

Current	Workday, Data Startups
Previous	Identified (acquired by Workday), Spitball Entertainment (acquired by Jumpstart), MarsFog (acquired by Mahoot)
Education	University of California, Berkeley
Recommendations	3 people have recommended Vlad

Experience

Sr Director of Software Engineering

Workday

February 2014 – Present (1 year 3 months) | San Francisco

Leading the former Identified team to create a new product category called Insight Applications (http://nyti.ms/1wxEYK2)

Making Workday's applications both predictive and prescriptive to help the world's largest companies make strategic decisions about their people and finances

Advisor & Investor

Data Startups

January 2010 – Present (5 years 4 months) | San Francisco

Helping startups with team building/hiring, technology selection & architecture and fundraising.

CTO

Identified (acquired by Workday)

December 2011 – February 2014 (2 years 3 months) | San Francisco

Built the predictive technology behind Identified Recruit to take on LinkedIn.
Ran the product/engineering team and scaled it to 30 people.
Oversaw the technology architecture and created SYMAN: http://tcrn.ch/144dUBU

Some of our press:

http://www.forbes.com/sites/joshbersin/2014/02/27/workday-acquires-identified-a-potential-disruptive-move-in-recruiting/

http://techcrunch.com/2013/05/08/identified-looks-to-solve-social-medias-dirty-data-problem-for-recruiters-with-help-from-former-linkedin-data-gurus/

Director of Engineering

Spitball Entertainment (acquired by Jumpstart)

August 2011 – December 2011 (5 months) | San Francisco Bay Area

Monetizing social games through content and merchandising.

Took over management of the engineering team. Re-architected the internal platform for rapid iteration and maintainability


facebook


twitter


Google+

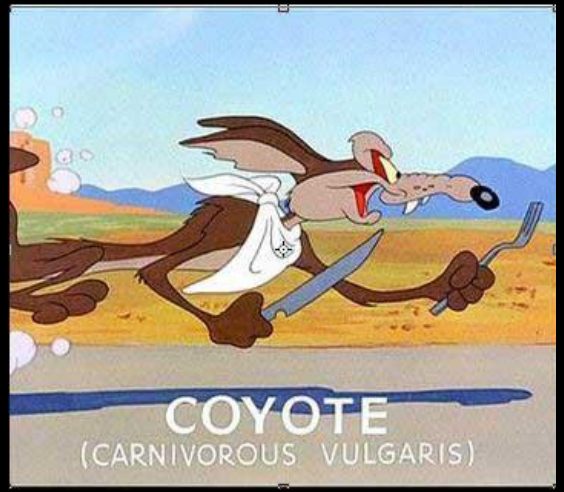

doximity


Academia.edu
share research


RP
RallyPoint


GitHub


ACME
ACME
ACME


COYOTE
(CARNIVOROUS VULGARIS)


ROAD RUNNER
(ACCELERATI INCREDIBILUS)

Identified

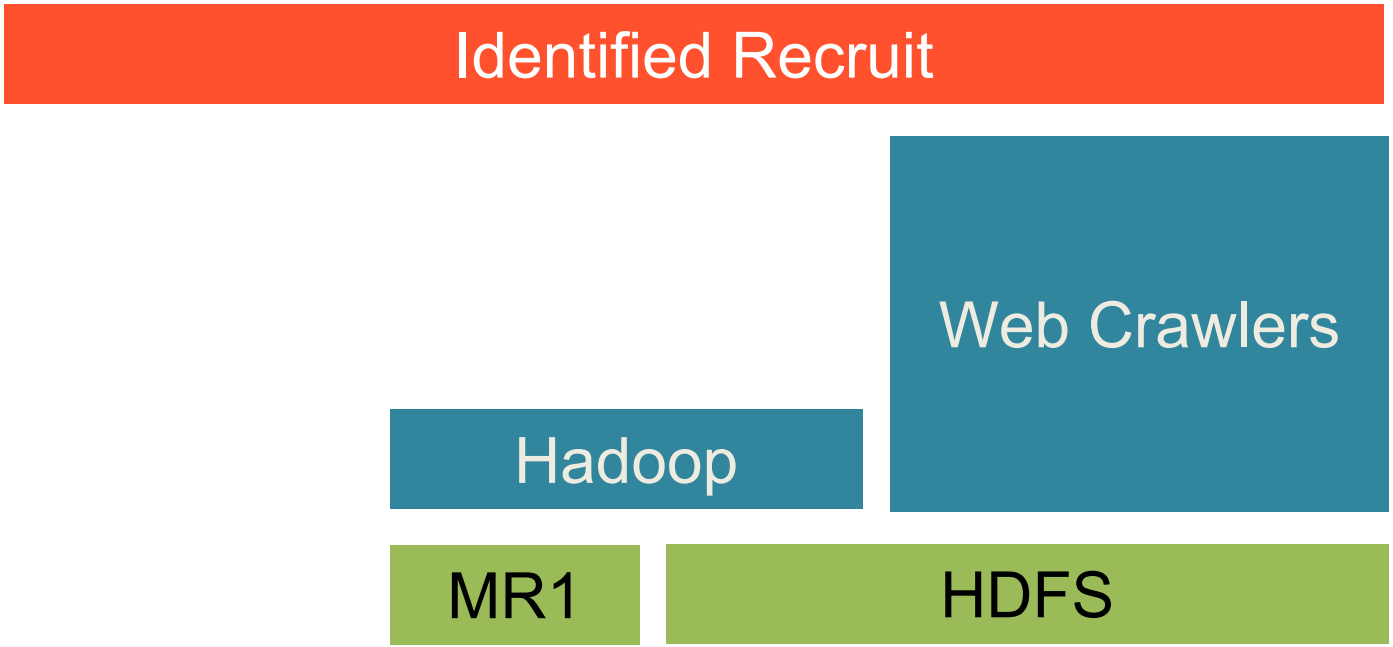
RECRUIT

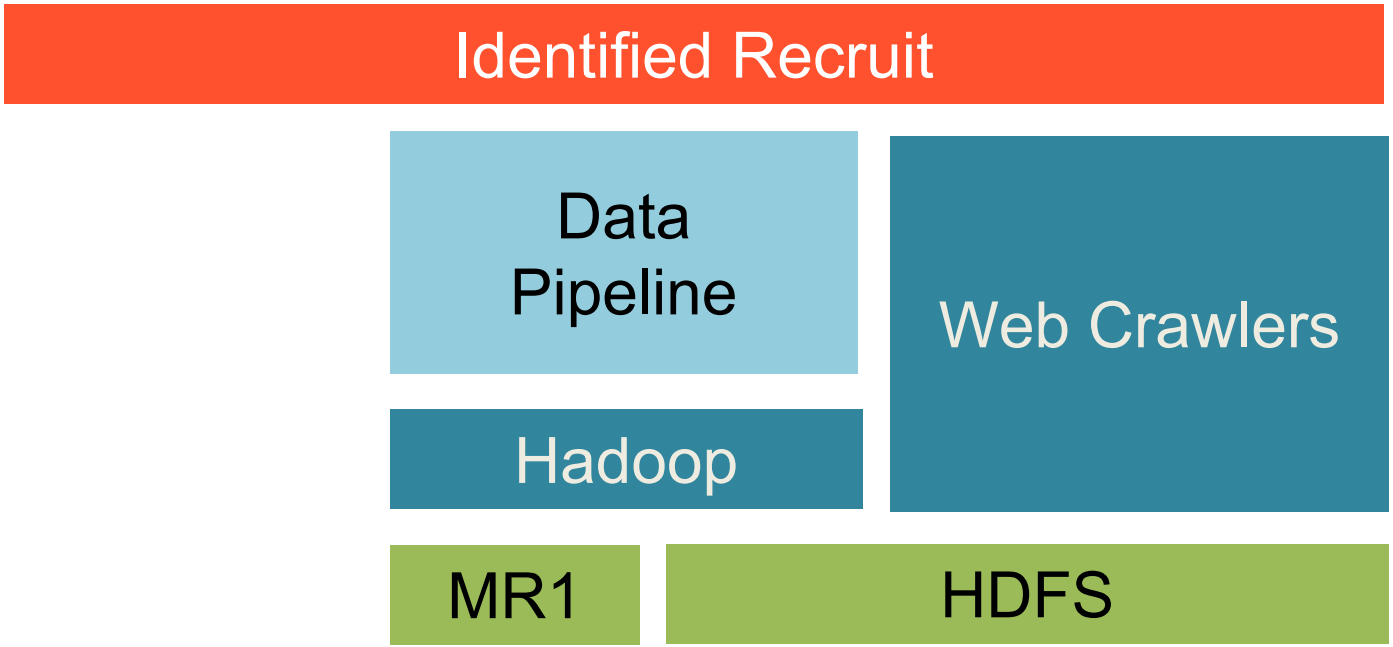
Identified Recruit

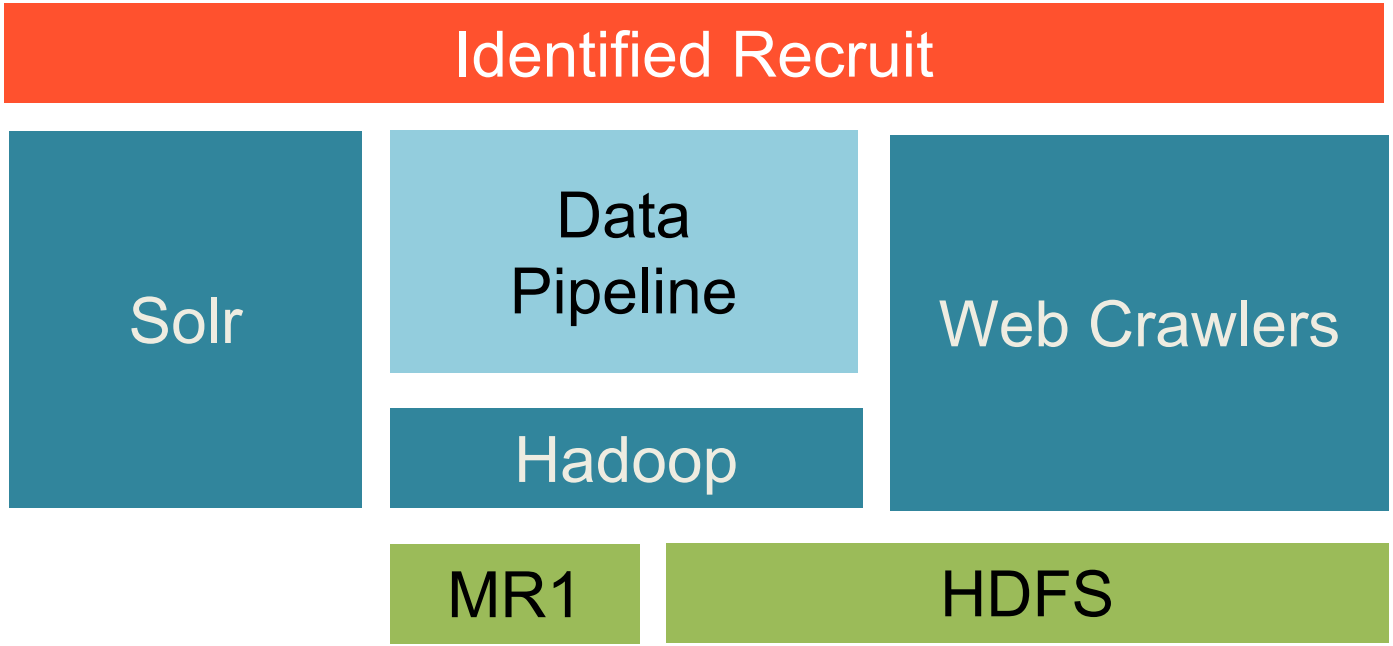
Identified Recruit

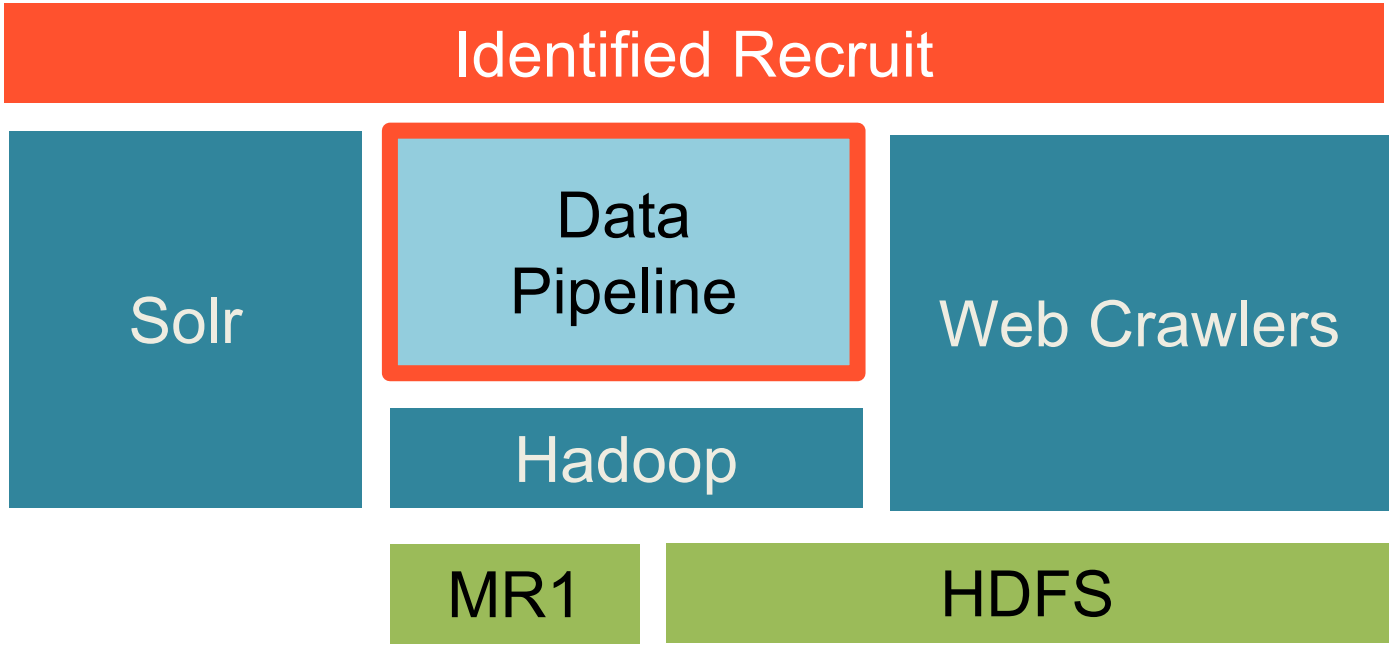
HDFS


Identified Recruit
Web Crawlers
HDFS


Identified Recruit
Web Crawlers
Hadoop
MR1
HDFS


Identified Recruit
Data Pipeline
Web Crawlers
Hadoop
MR1
HDFS

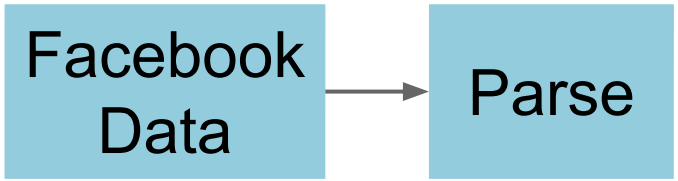

Identified Recruit
Solr
Data Pipeline
Web Crawlers
Hadoop
MR1
HDFS


Identified Recruit
Solr
Data Pipeline
Web Crawlers
Hadoop
MR1
HDFS

Identified Data Pipeline 1.0

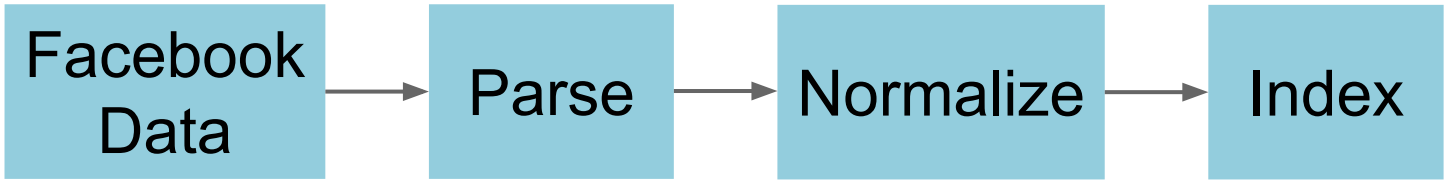
Facebook Data

Identified Data Pipeline 1.0



Identified Data Pipeline 1.0

Identified Data Pipeline 1.0



Identified Data Pipeline 1.0

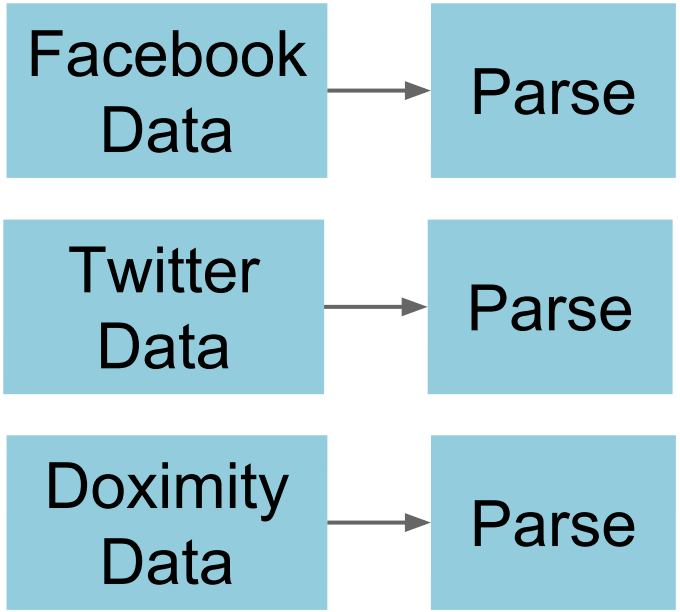
Identified Data Pipeline 2.0

Facebook Data

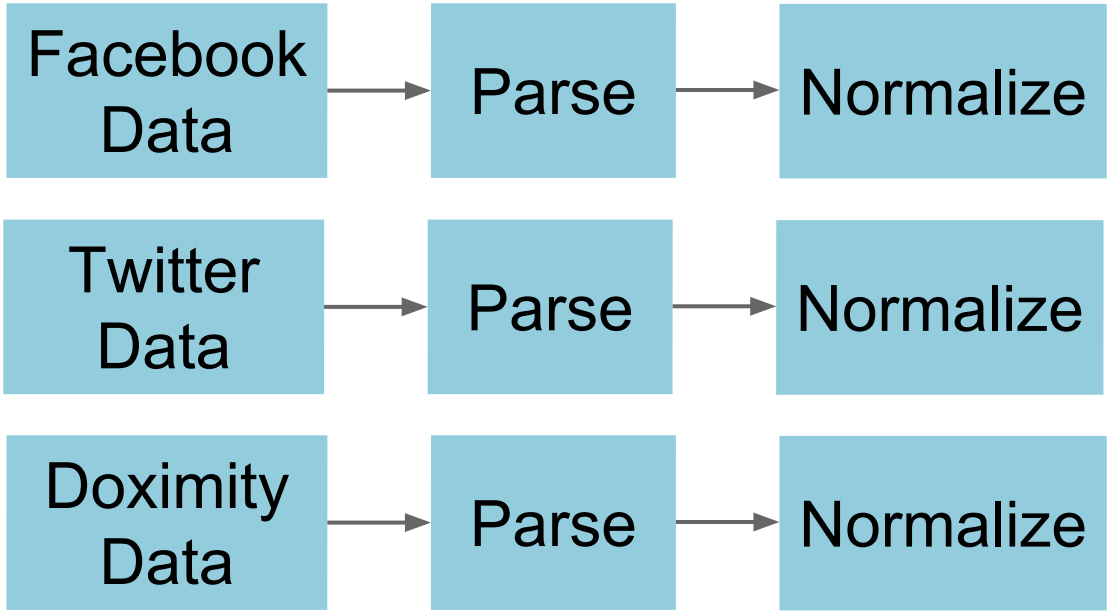
Twitter Data

Doximity Data

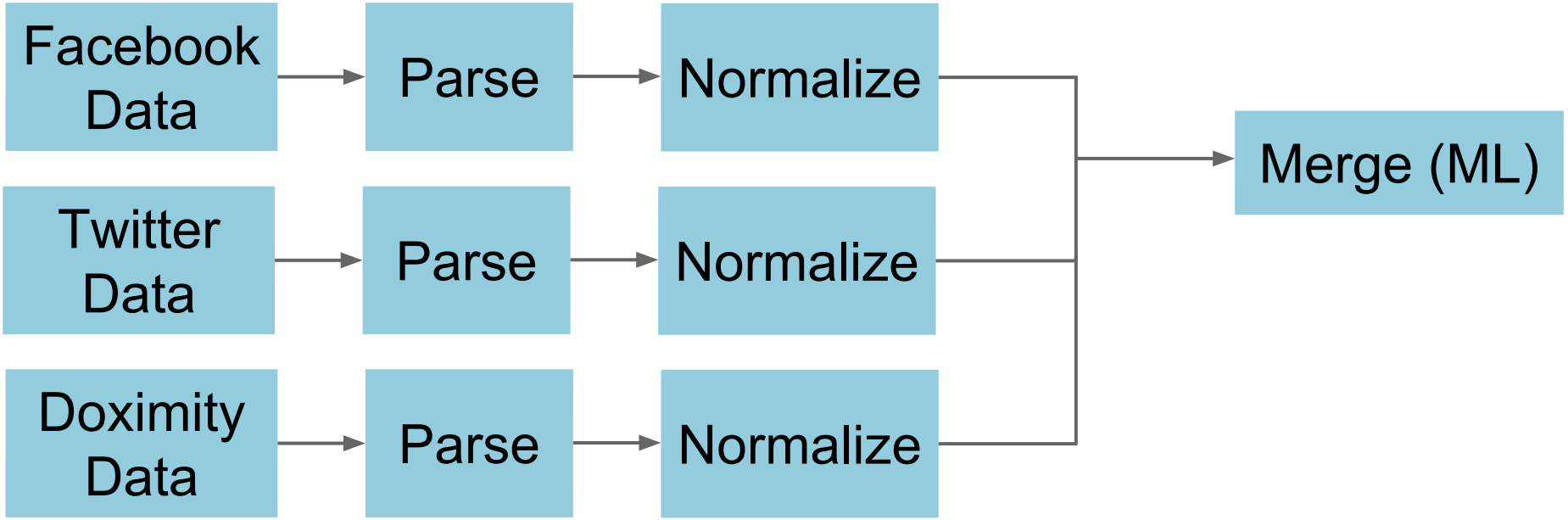
Identified Data Pipeline 2.0



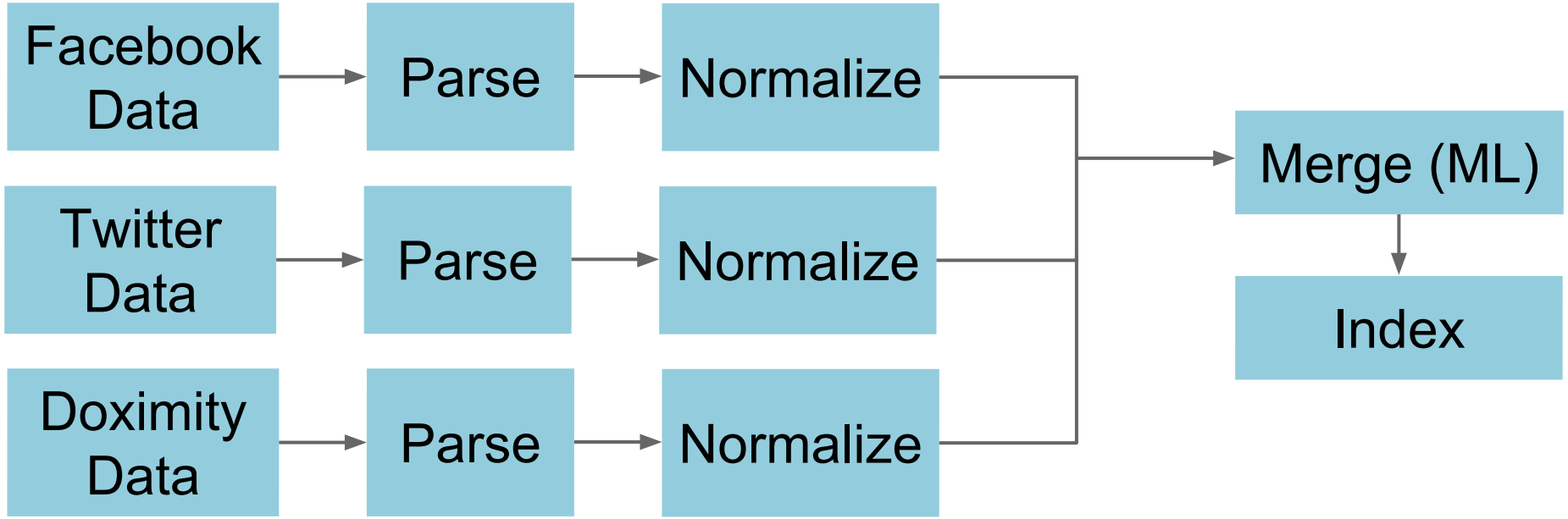
Identified Data Pipeline 2.0



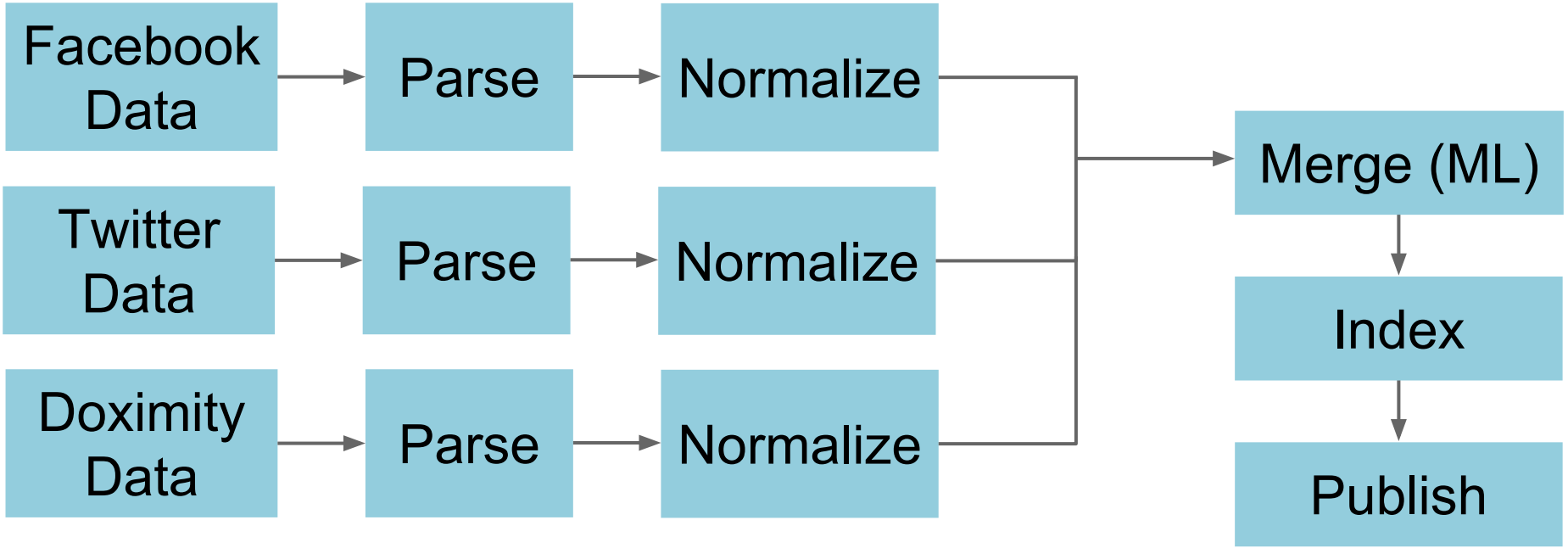
Identified Data Pipeline 2.0



Identified Data Pipeline 2.0



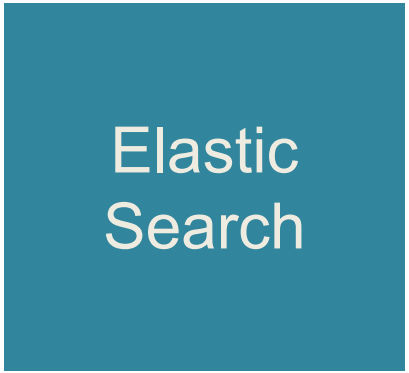
Identified Data Pipeline 2.0



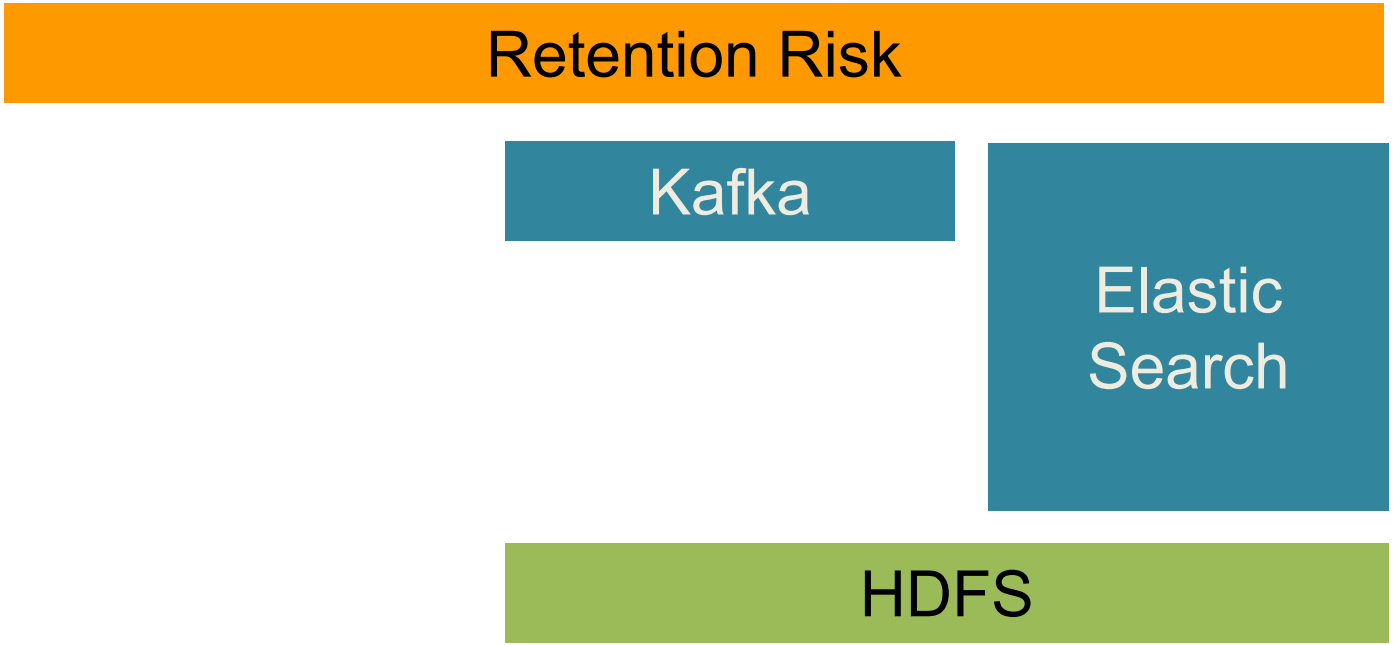

workday.

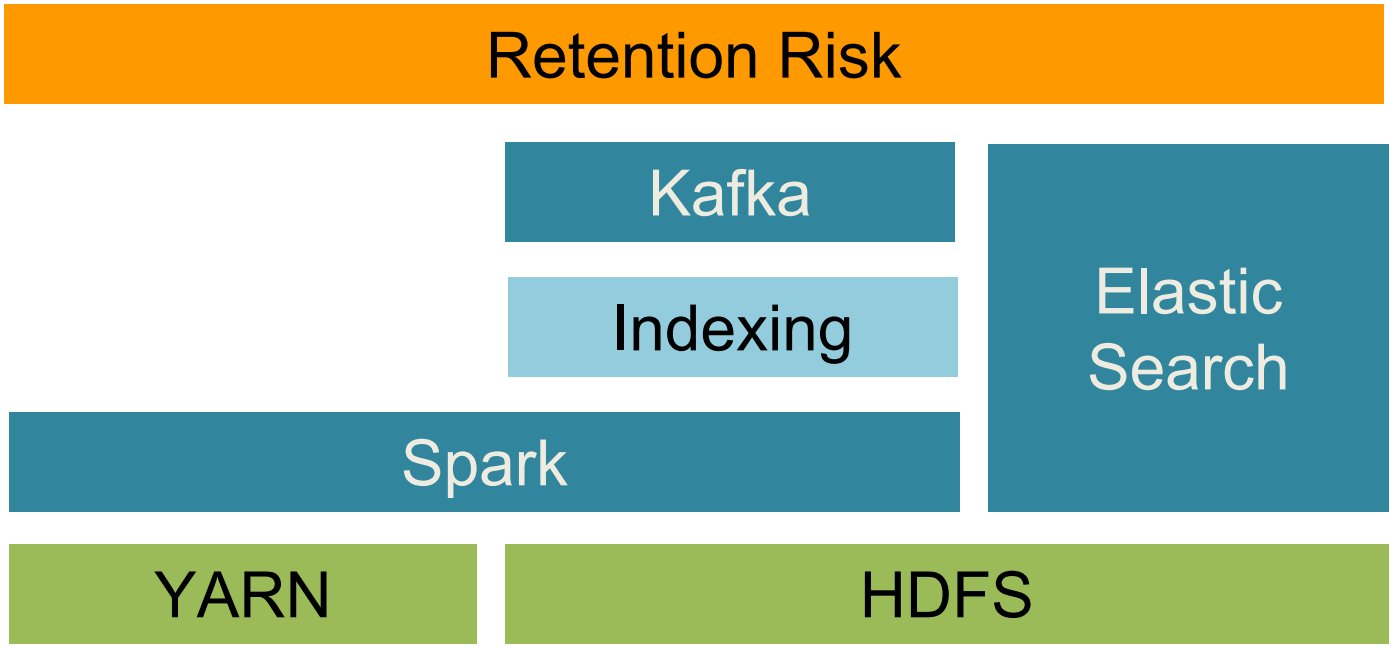
Retention Risk

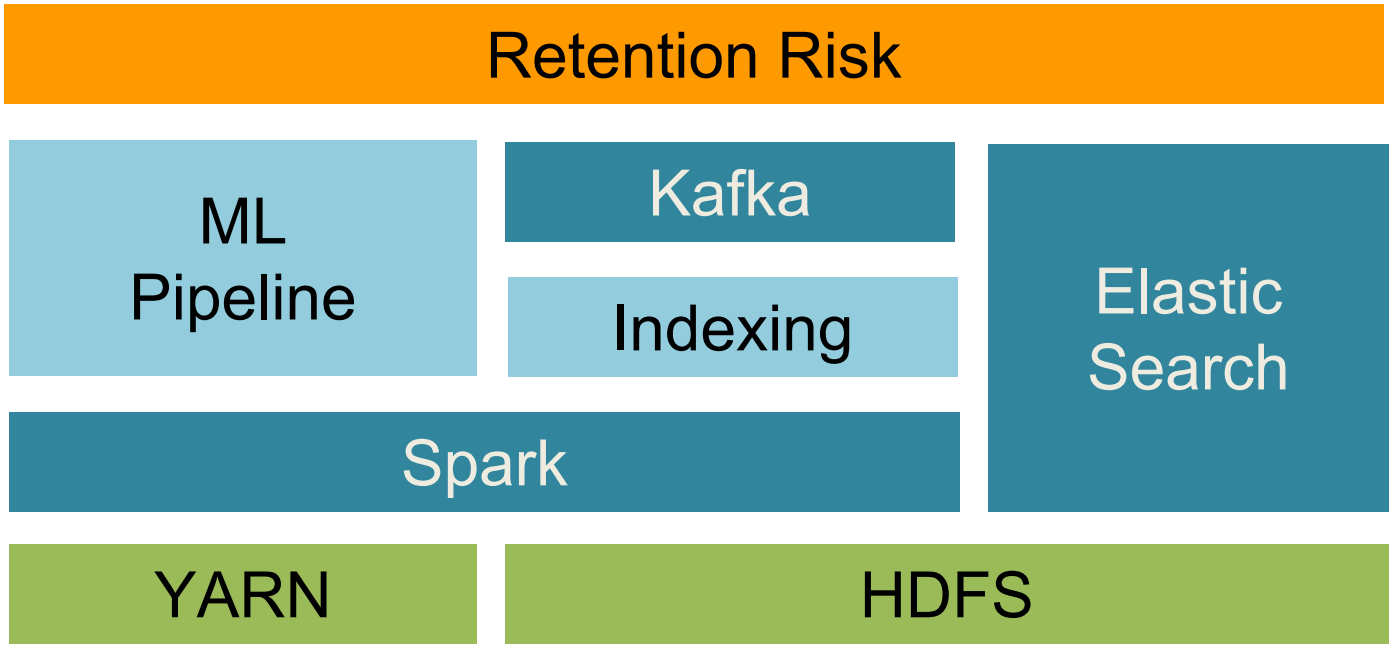
Retention Risk

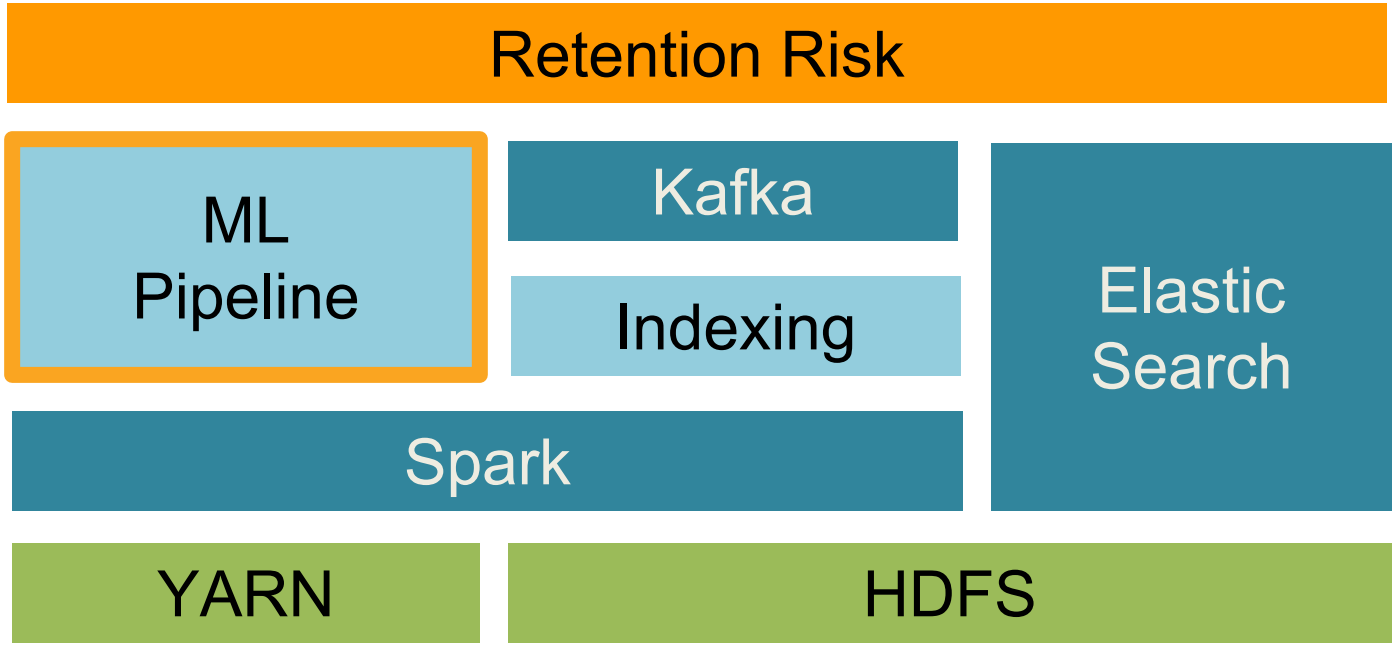



Retention Risk
Elastic Search
HDFS


Retention Risk
Kafka
Elastic
Search
HDFS


Retention Risk
Kafka
Indexing
Elastic Search
Spark
YARN
HDFS

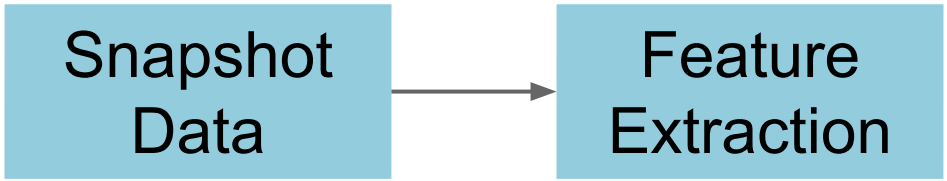

Retention Risk
ML Pipeline
Kafka
Indexing
Elastic Search
Spark
YARN
HDFS


Retention Risk
ML
Pipeline
Kafka
Indexing
Elastic
Search
Spark
YARN
HDFS

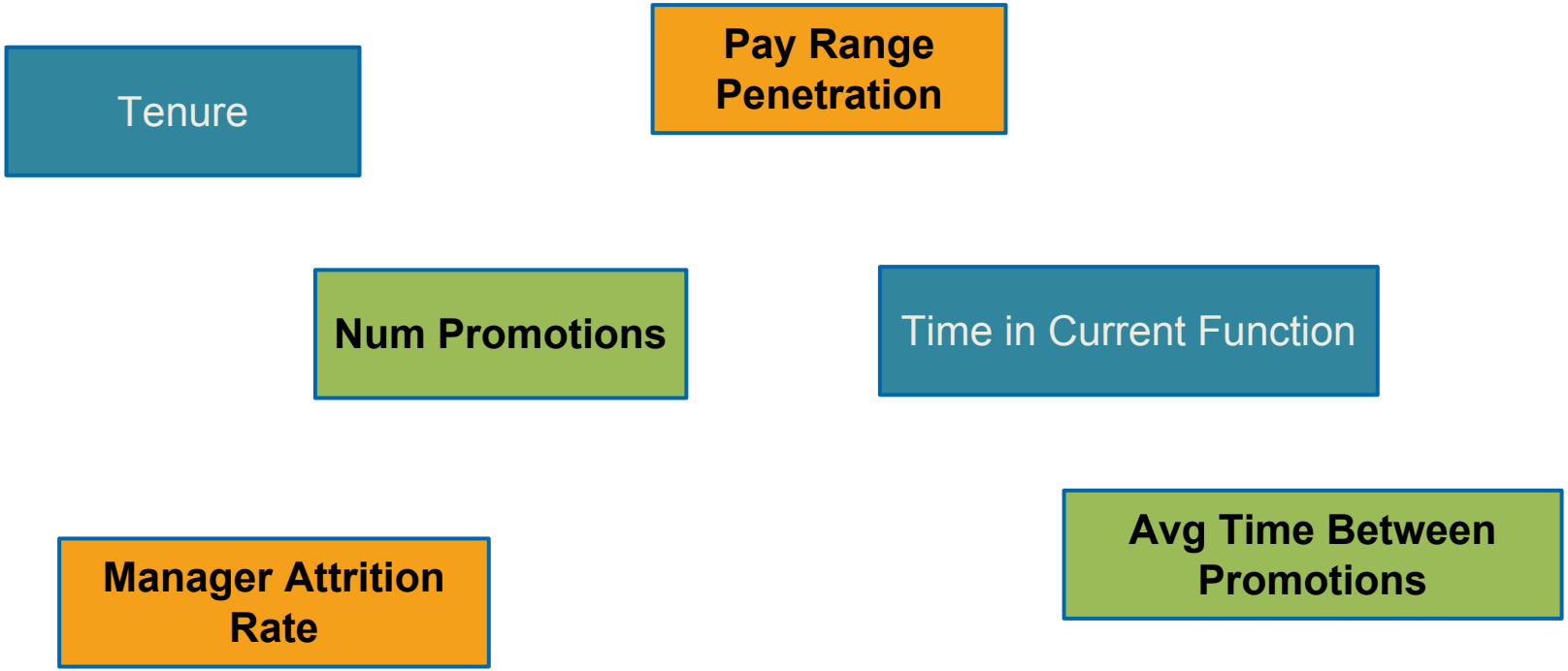
ML Pipeline

Snapshot Data

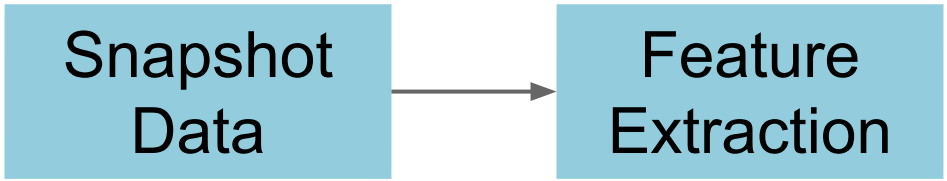
ML Pipeline



Data and “Features”

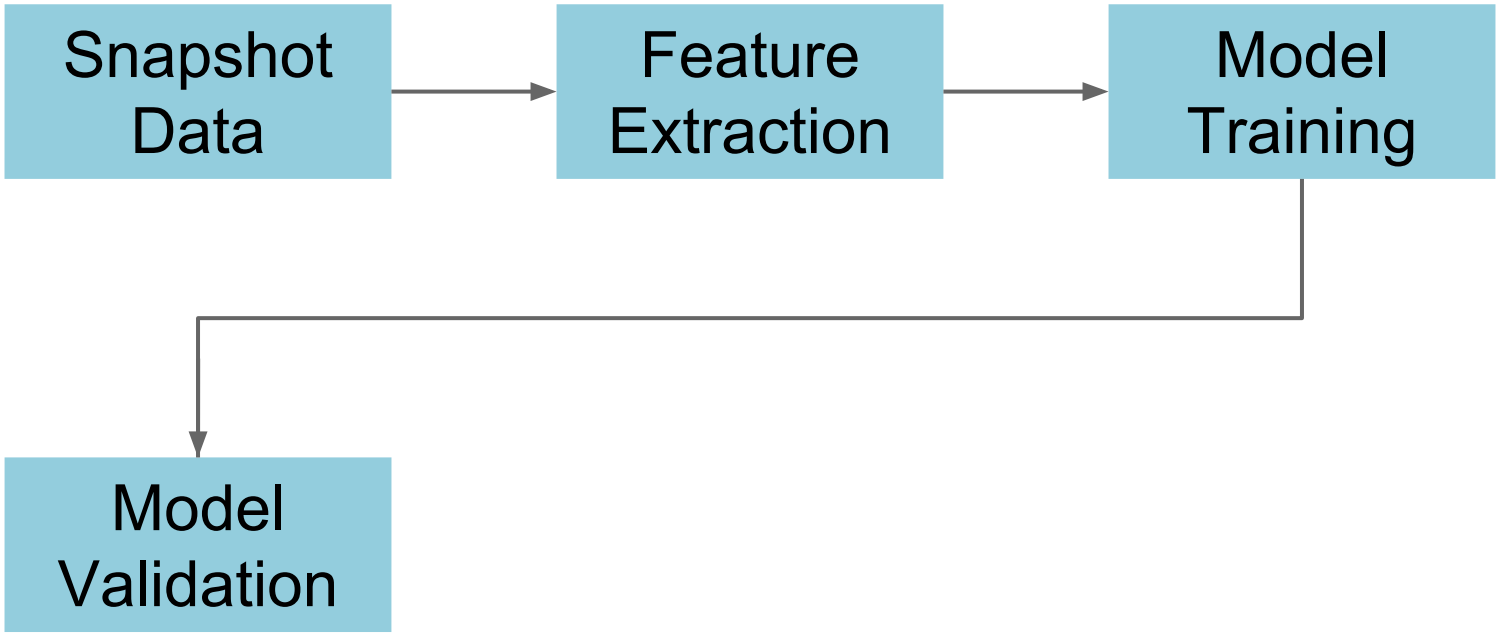


ML Pipeline

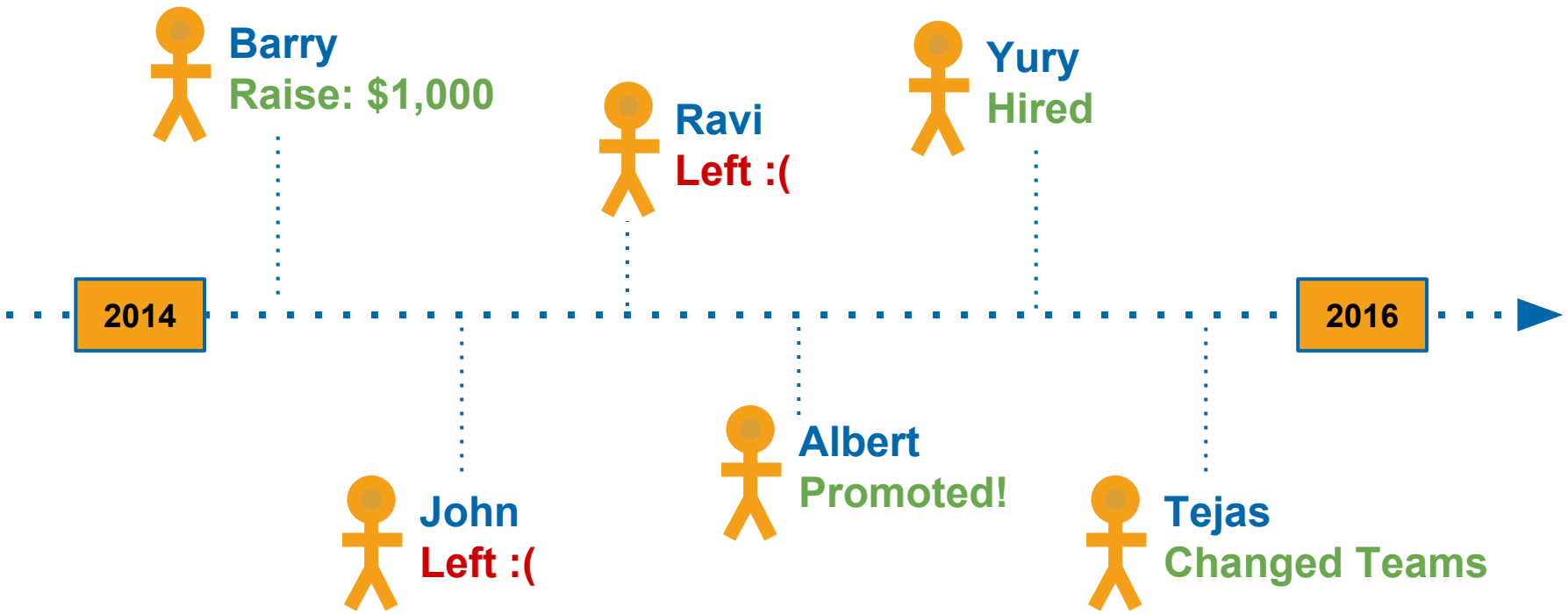


ML Pipeline

ML Pipeline

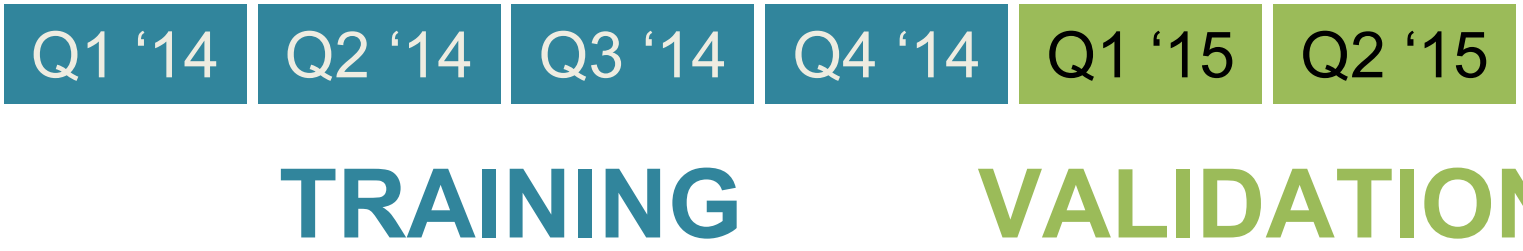

Snapshot Data
Feature Extraction
Model Training
Model Validation

Training and Validation

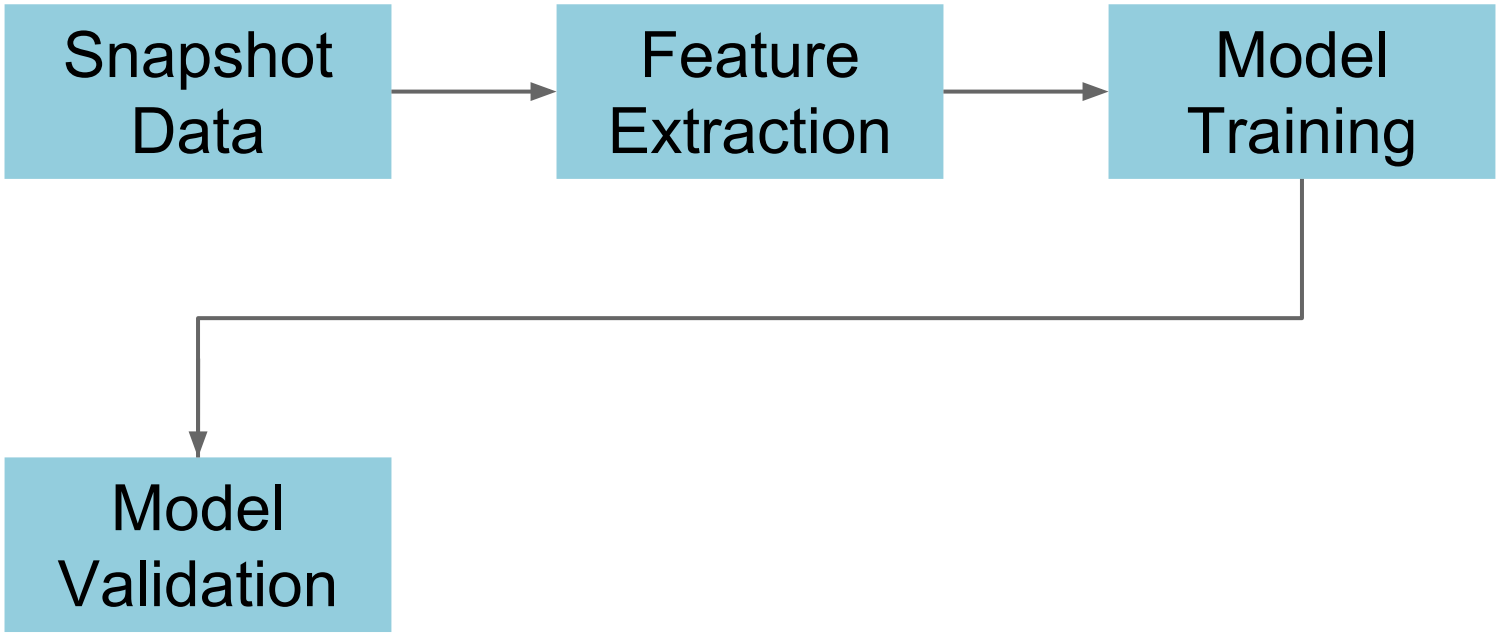


Training and Validation

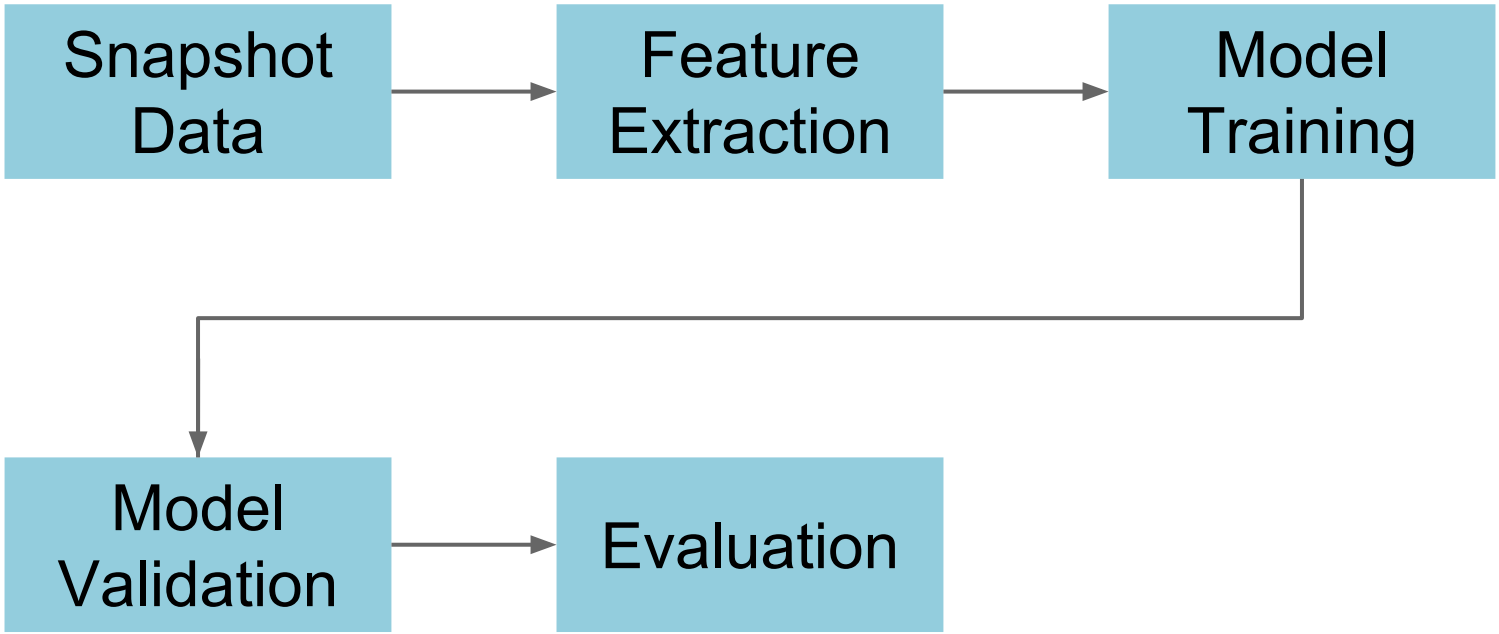
Training and Validation



ML Pipeline


Snapshot Data
Feature Extraction
Model Training
Model Validation

ML Pipeline

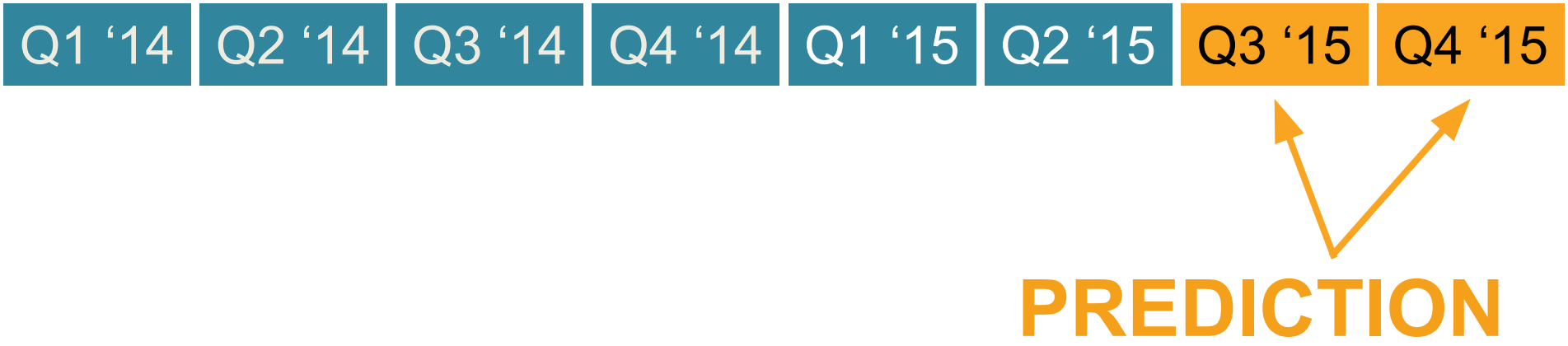

Snapshot Data
Feature Extraction
Model Training
Model Validation
Evaluation

Evaluation

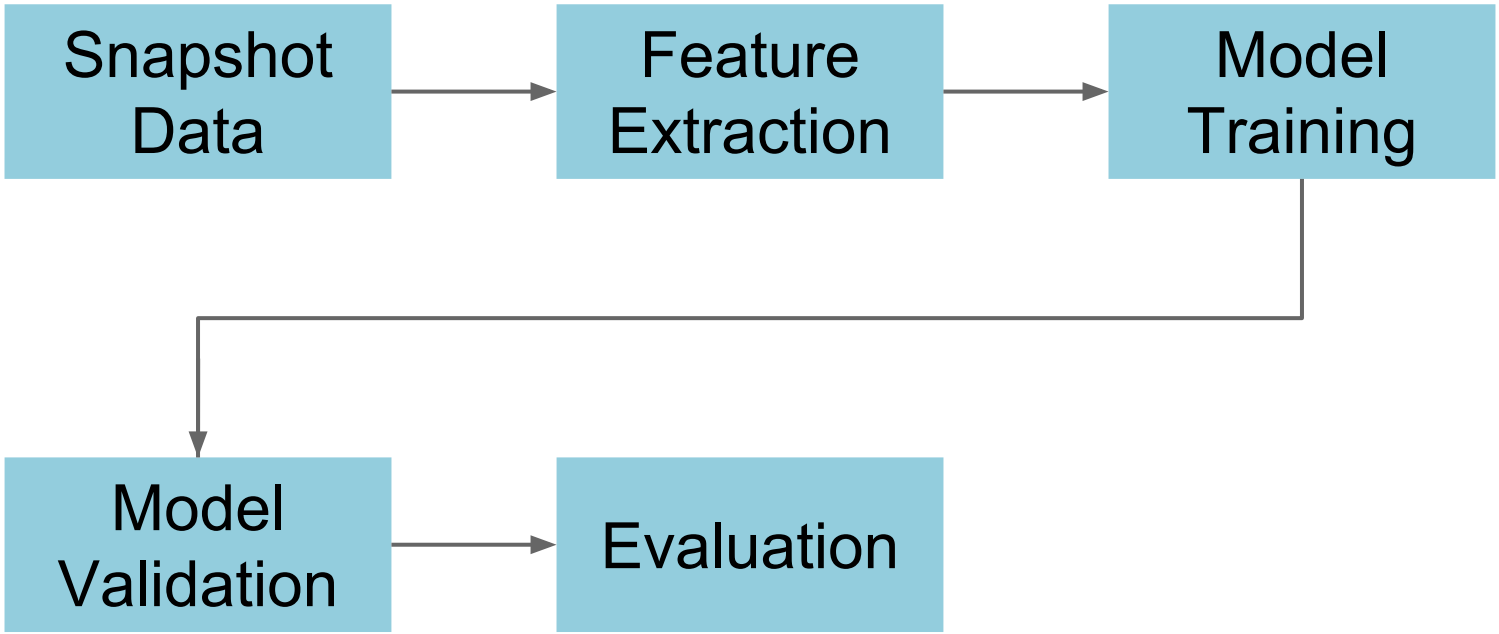
Q1 ‘14	Q2 ‘14	Q3 ‘14	Q4 ‘14	Q1 ‘15	Q2 ‘15

Evaluation

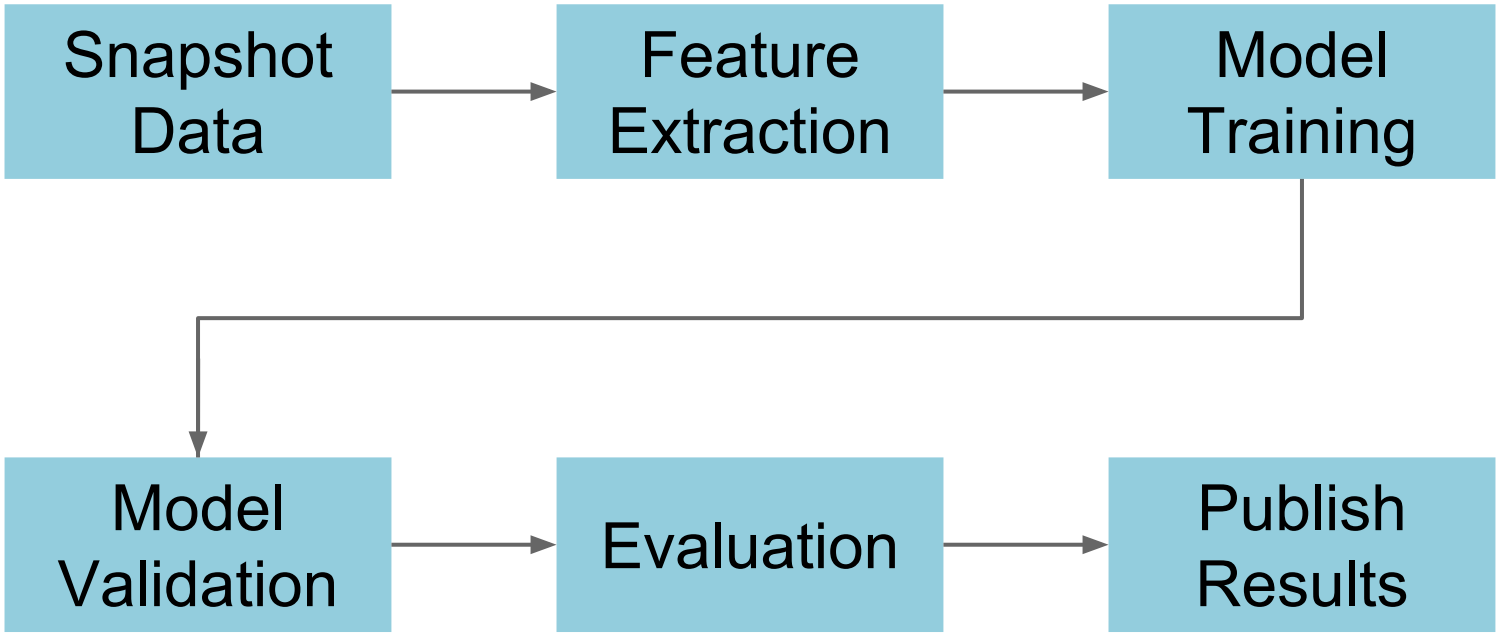
Evaluation



ML Pipeline


Snapshot Data
Feature Extraction
Model Training
Model Validation
Evaluation

ML Pipeline


Snapshot Data
Feature Extraction
Model Training
Model Validation
Evaluation
Publish Results


ROAD RUNNER
(ACCELERATI INCREDIBILUS)


W