

Technology Requirements for Stitch Fix

*Group 2: Claire Wu (jw3644), Mandy Chan (mc4528),
Sarah Garman (sag2213), Teresa Newton (ten2111)*

Content

- ❑ Executive Summary
- ❑ Business Requirements Review
- ❑ Nature of Stitch Fix Data
- ❑ Data Engines
- ❑ Data Lake Architecture
- ❑ Data Management
- ❑ Conclusion

Executive Summary

What: New data sources consisting of celebrity-related information

Why: Increase algorithm accuracy

How: Incorporation of MongoDB and Neo4j into a broader data lake configuration consisting of Hadoop, AWS, and Spark tools.

Fashion Influencers

A collection of fashion-related icons including a hat, perfume bottle, hairbrush, handbag, shoes, and a dress.

-

Nature of Stitch Fix Data

Stitch Fix IPO: It's About the Data, not the Fashion

Volume



- *User data*
- *Product data*
- *Fashion design*
- *Supply chain*

Velocity



- *Real-time*

Variety



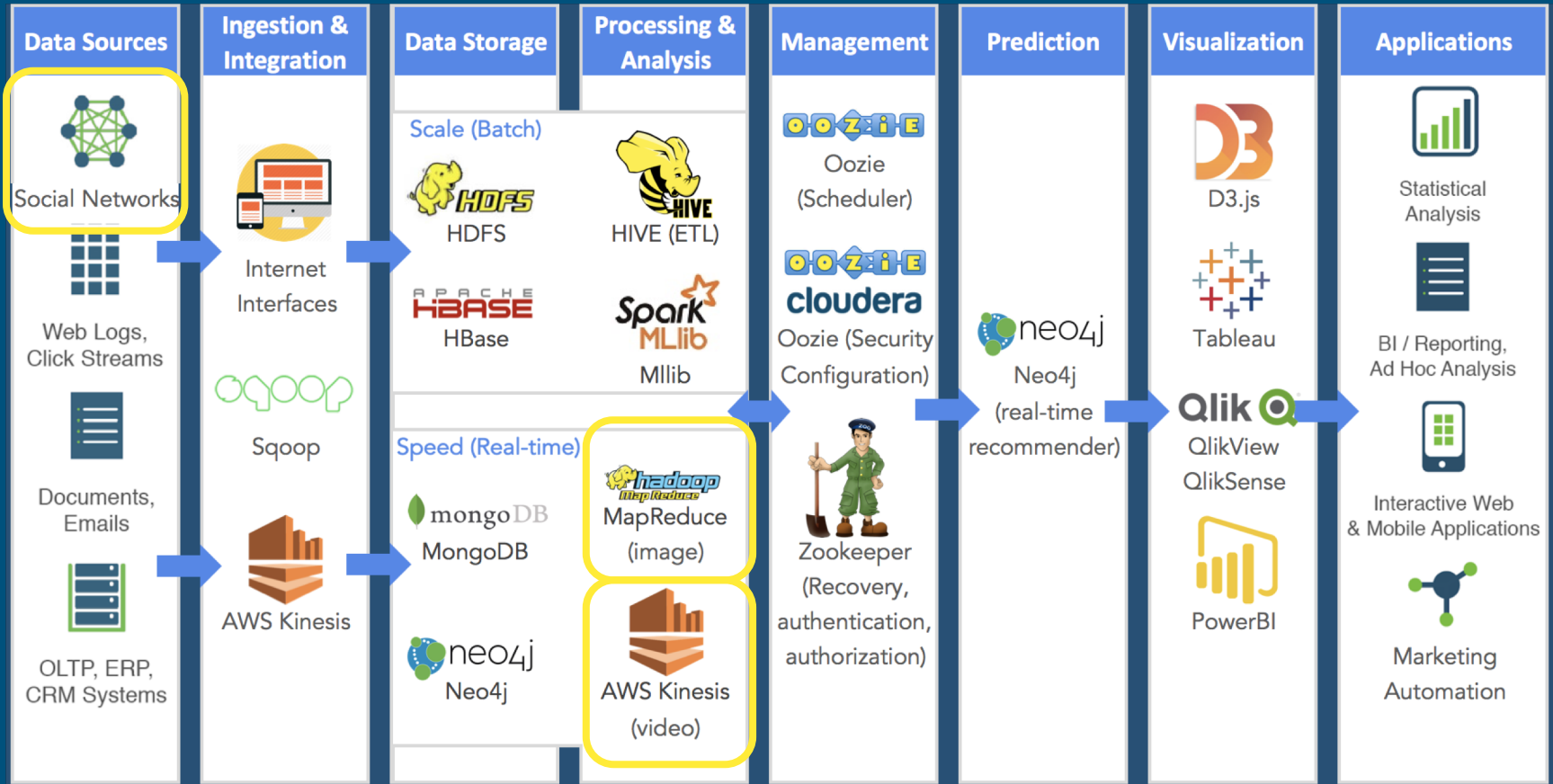
- *Structured*
- *Unstructured*

Data Engines



Benefits	<ul style="list-style-type: none">• NoSQL• Real-Time Analytics• Ability to store video and images• High-Speed	<ul style="list-style-type: none">• NoSQL• Real-Time Analytics• Deep Queries• User Item Graphs
Risks	<ul style="list-style-type: none">• Joins not Supported• High Memory Usage• Limited Data Size - cannot exceed 16 MB• Limited Nesting - cannot exceed 100 levels	<ul style="list-style-type: none">• Graph database can only be stored on one server• Storage Limitation - 34 billion nodes & 274 billions of attributes

Data Lake Architecture



Data Management

Recovery/Continuity of Business

- Up to \$75,000 - the cost of one hour of downtime
- **Risk Mitigation:** systems up-to-date & daily data backup
- **Disaster Recovery Team**
 - Updated stakeholder contact list
 - Operational Recovery options:
 - Data Replication
 - Alternate Network Routes
 - Working Remotely
 - Cloud-Based Services
- **Hadoop**
 - Hadoop Administrator
 - Apache Zookeeper

Data Governance

The following tools may be useful to supplement Stitch Fix's data governance strategy

- Cloudera Navigator
- SAS Data Governance
- IBM InfoSphere Information Governance Catalog

Conclusion

- **Requirement:** Improve Operational Efficiency Using Data for Predictive Customer Behavior
 - Design fusing business requirements and technology
 - Analytics-driven for predictive customized consumer data
 - Security considerations
- **Result:** Infrastructure promoting increased customer engagement and brand sustainability



Thank you!



Questions?

