

# LEVERAGING SPARK TO DEMOCRATIZE DATA FOR OMNI-COMMERCE

JUNE 7, 2017

---

SHAFaq ABDULLAH,  
DATA SCIENCE @HONEST



- ▶ Serial Tech Entrepreneur, Heading Data Engg at Honest.
- ▶ Cofounder & CTO at Insnap
- ▶ Led Realtime Analytics platform at GREE.
- ▶ Pivotal in Zenprise (acquired by Citrix)
- ▶ Cloud and Mobile Platform Engg at Nokia
- ▶ MS from Tampere Univ of Tech
- ▶ 10 patents in areas of Enterprise Security-aaS, BigData, ML



## The world's most valuable resource is no longer oil, but data

*The data economy demands a new approach to antitrust rules*



Print edition | Leaders >

May 6th 2017



A NEW commodity spawns a lucrative, fast-growing industry, prompting antitrust regulators to step in to restrain those who control its flow. A century ago, the resource in question was oil. Now similar concerns are being raised by the giants that deal in data, the oil of the digital era. These titans—Alphabet (Google's parent company), Amazon, Apple, Facebook and

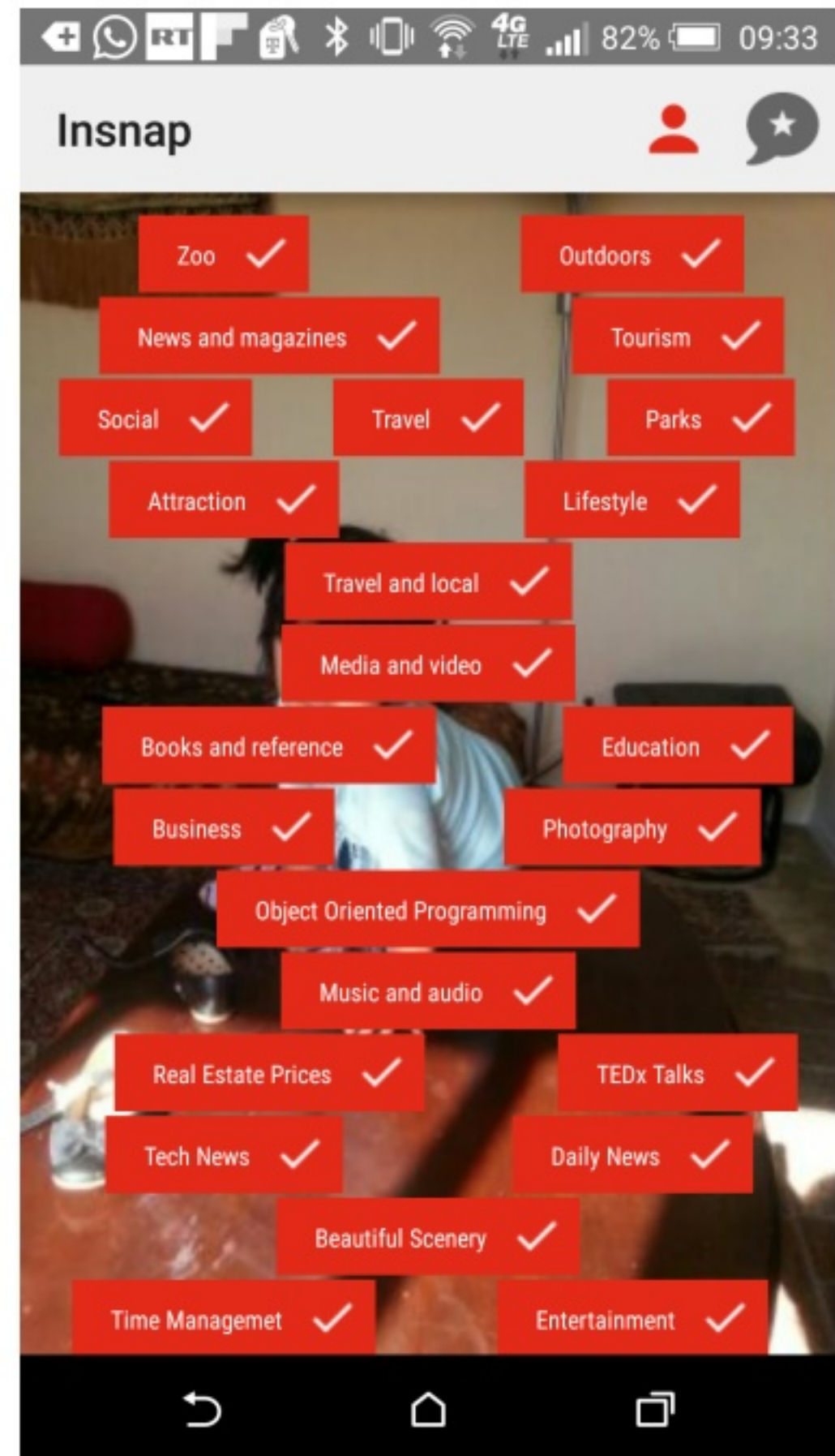
## AGENDA

- ▶ Status quo at Honest (pre Insnap)
- ▶ Data Initiatives
- ▶ Alignment of Team, Culture, Tech
- ▶ Sparking Tech
- ▶ Conclusions



# INSNAP- INSTANTANEOUS PERSONAS

- ▶ Problem: Beating Cold-Start
- ▶ Product: Multitenant SaaS ML API to drive monetization & engagement by interest inference



## THE MISSION-DRIVEN

- ▶ Honest is dedicated to providing trusted, effective products across a wide range of consumer categories, including baby, beauty, personal care, home care, and vitamins & supplements.
- ▶ Bundle based Subscription Model for e-commerce



Diapers & Wipes Bundle



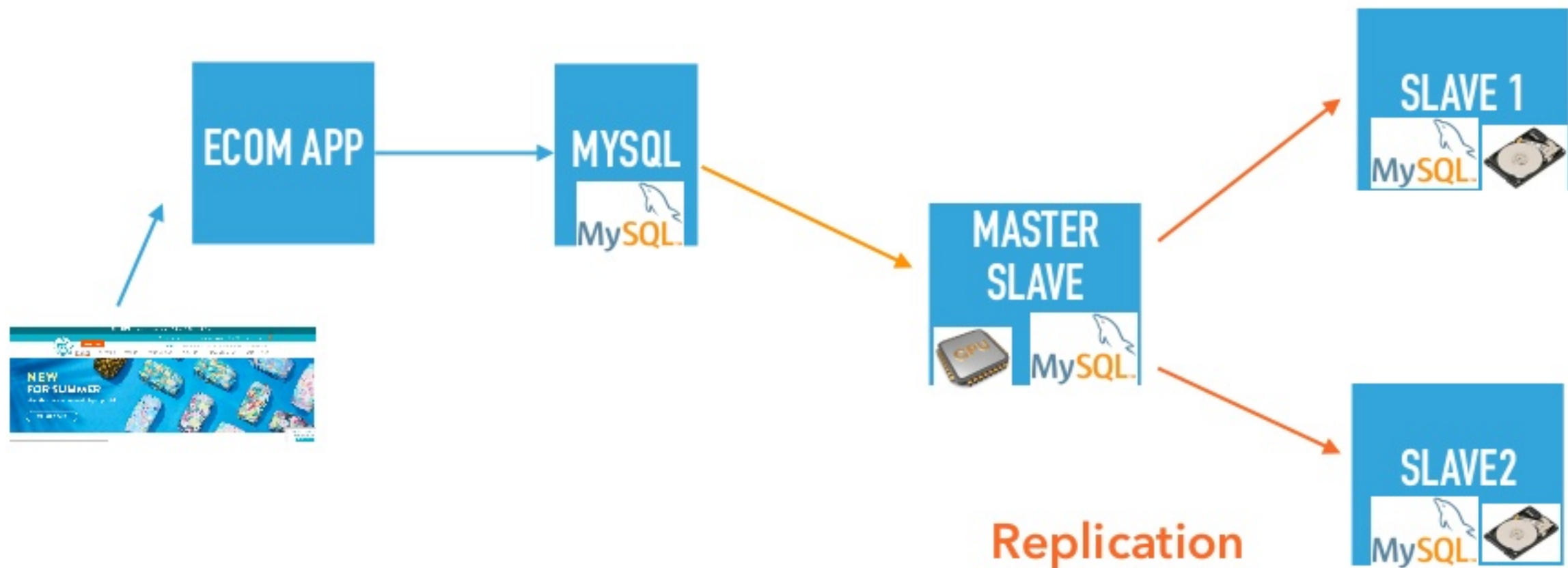
Honest Essentials Bundle



## STATUS QUO AT HONEST (PRE-INSNAP)

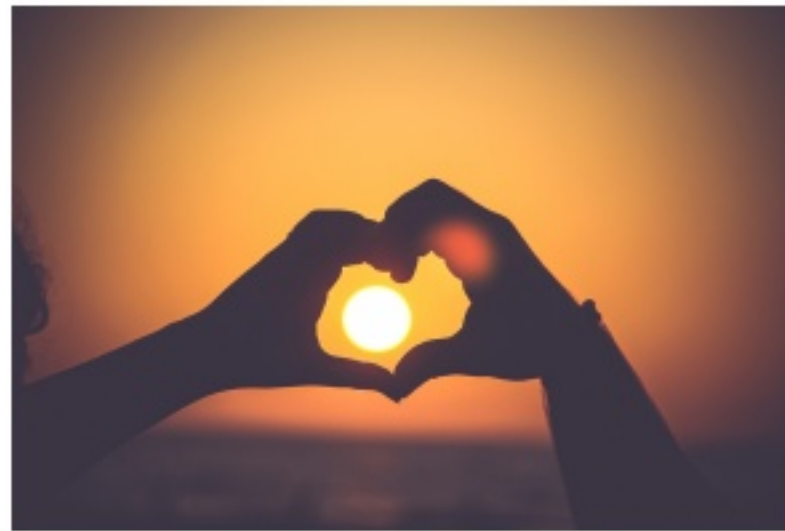
- ▶ Silo'd data and org
- ▶ Data Integrity ?
- ▶ Great potential to become data-driven







# INSNAP ACQUISITION BY HONEST



## DATA INITIATIVES

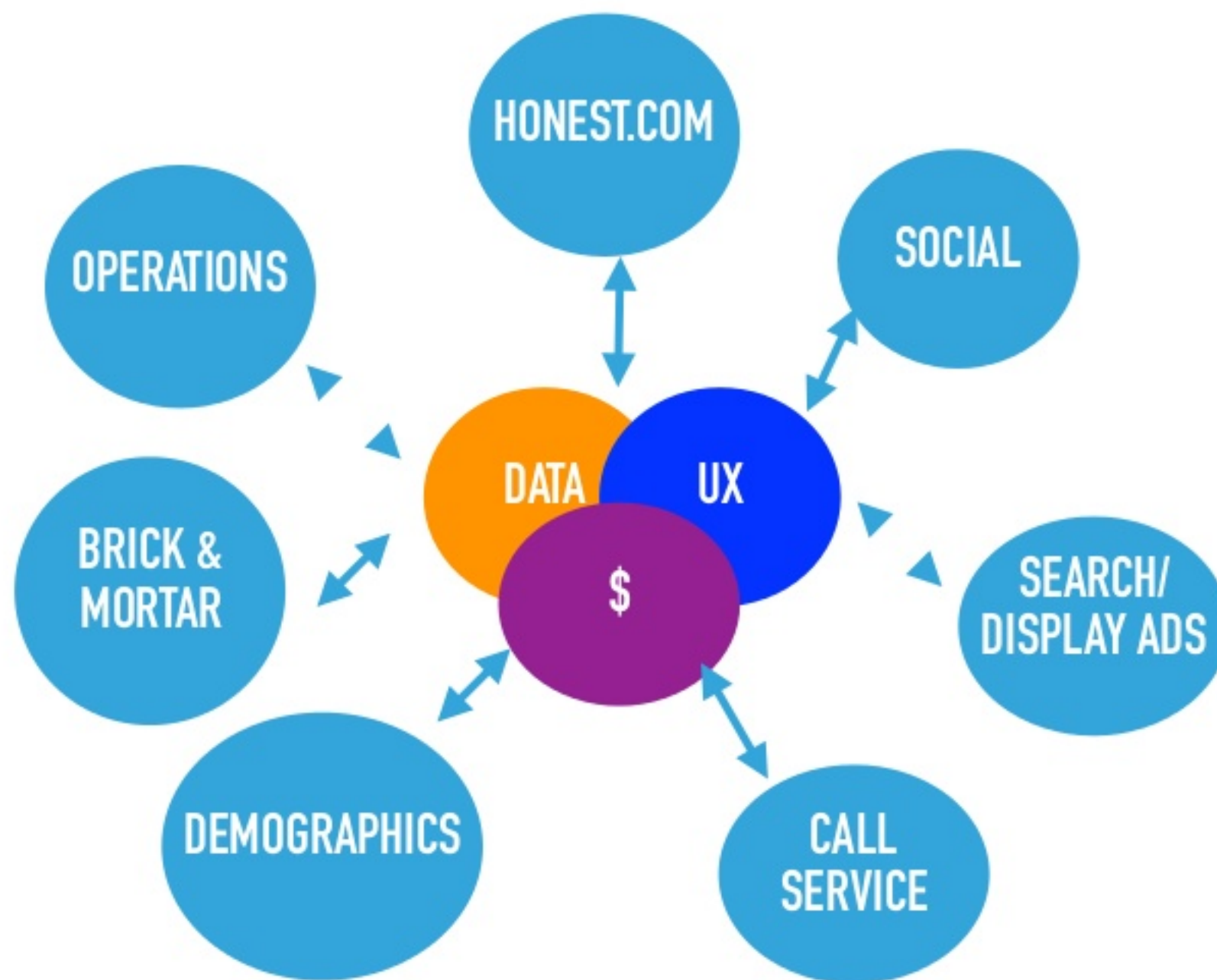
- ▶ Break away data and org silos
- ▶ Instrument logging
- ▶ Use ML to turn data into actionable insights
- ▶ XFunctional teams aligned to business goal
- ▶ Self-Served Model for decision-making folks

## CULTURE

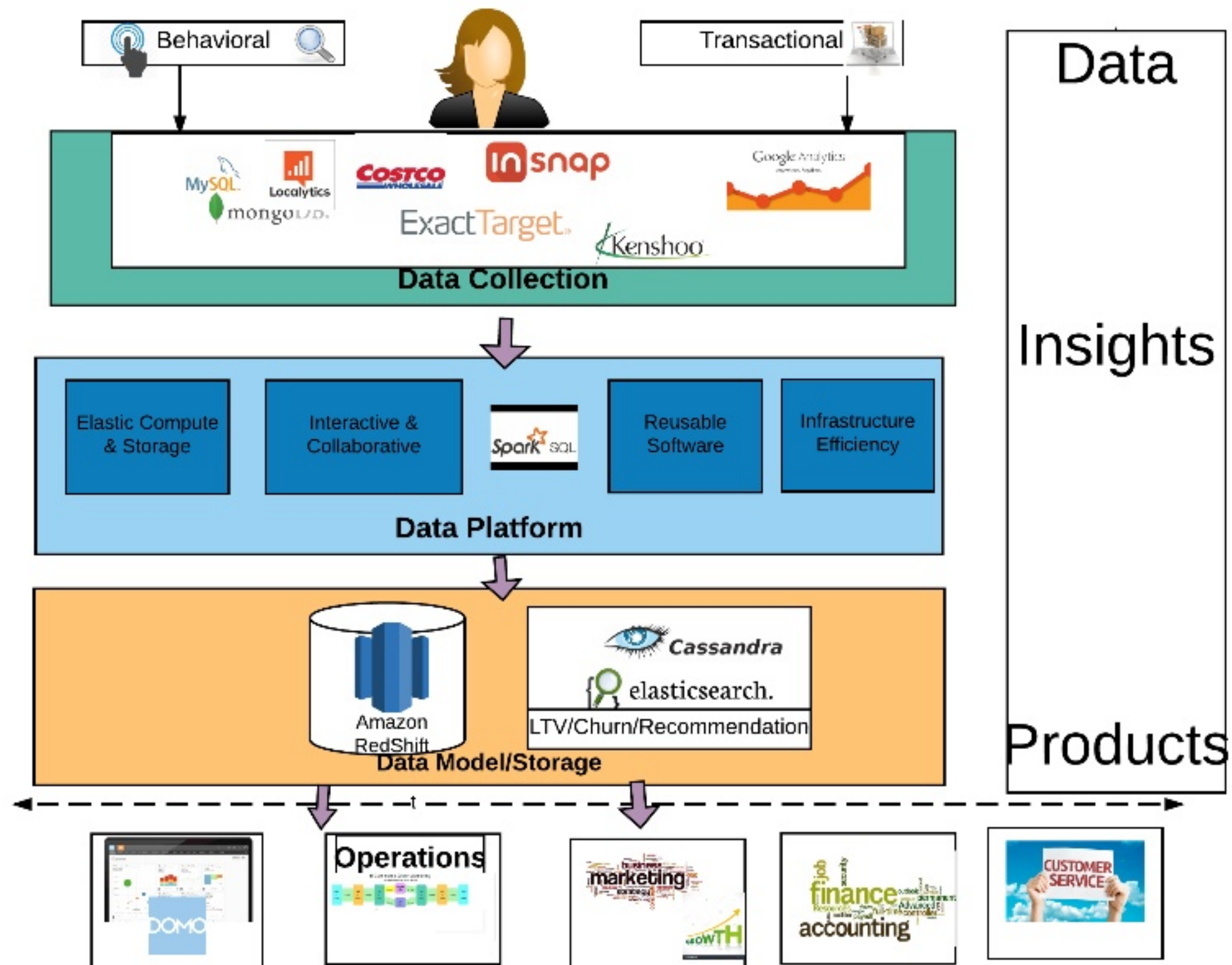
- ▶ What cannot be measured, cannot be improved
- ▶ Transparency around decision-making
- ▶ Empower users via self-served platform
- ▶ Fail-fast



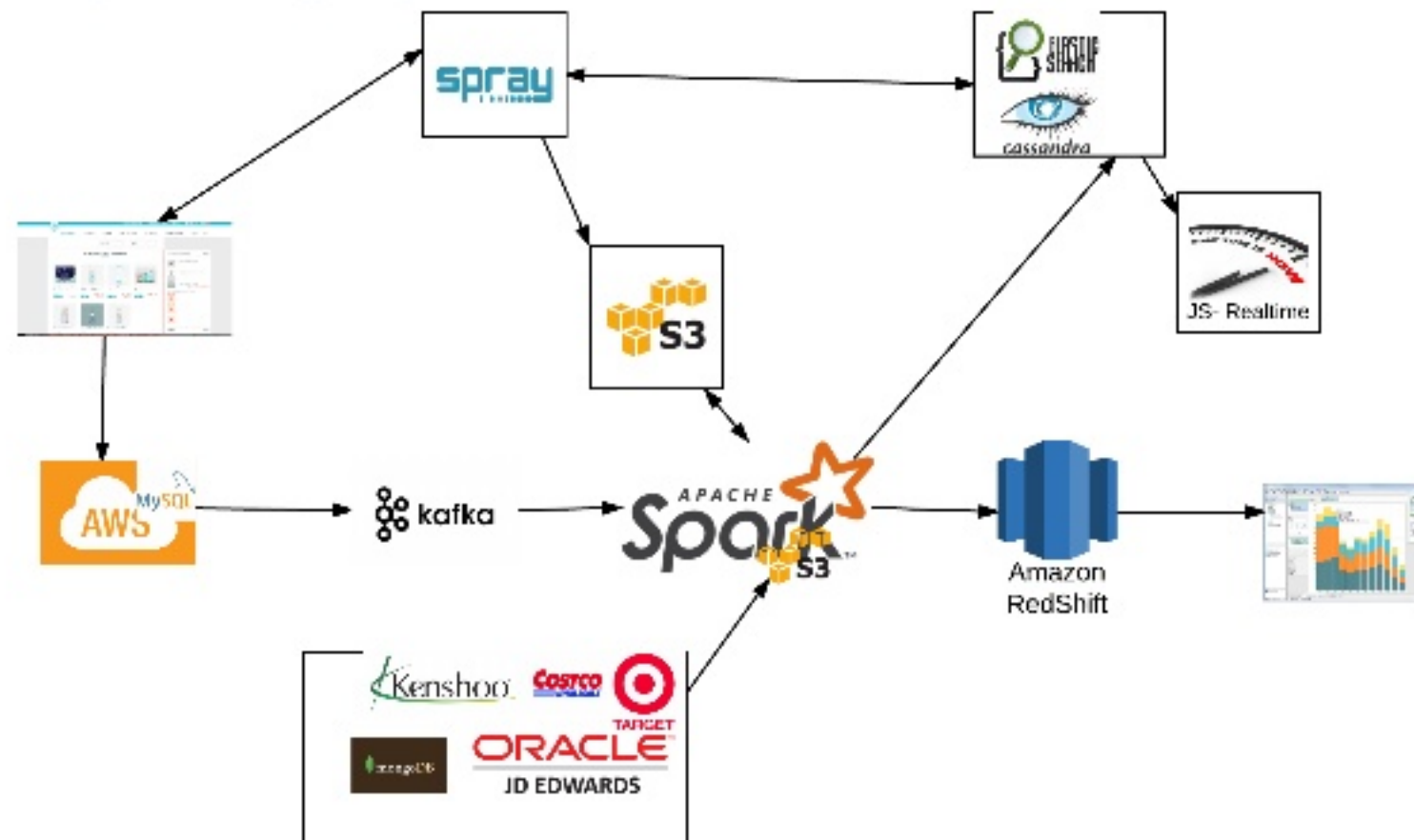
## DATA DRIVES OMNI-CHANNEL SOLUTION



# DATA PLATFORM



# NEW ARCHITECTURE



- 100M+ Monthly Events
- 100s Device Types
- S3 datalake 50TB

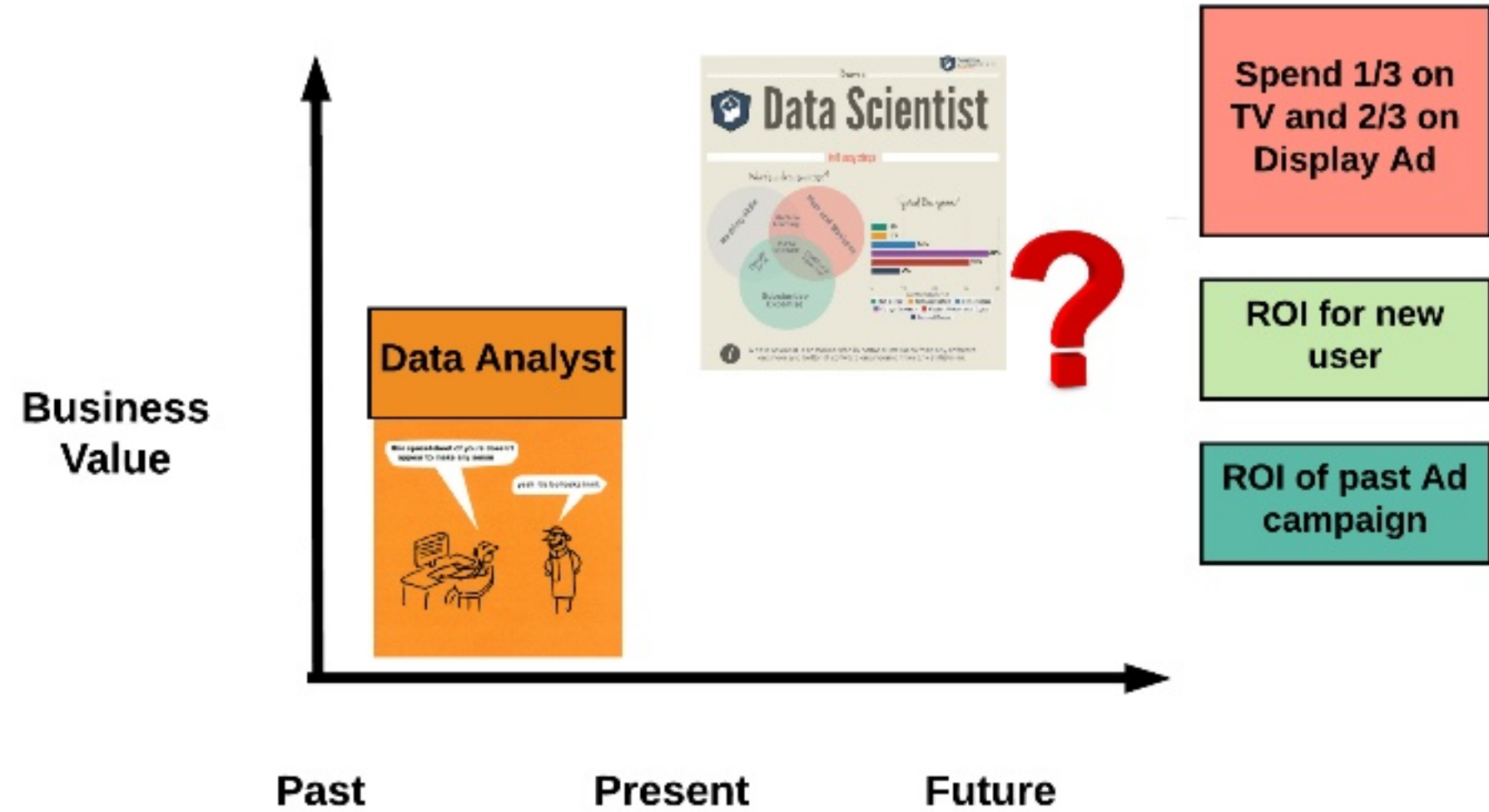
- Spark is 10x faster at computing than legacy
- Used for Scoring Models



## ADVANTAGES OF SPARK & DATABRICKS

- ▶ Scaling code, scaling teams
- ▶ Elastic Store & Compute - Spark
- ▶ Collaborative & Interactive
- ▶ Infrastructure Efficiency

DESCRIPTIVE -> PREDICTIVE -> PRESCRIPTIVE



## CONCLUSIONS

---



- ▶ Invest in Infrastructure & People
- ▶ Embed Analytics in Business & Org
- ▶ Keep calm and let the Data speak



---

THANK YOU -

@shafaq 110

Questions