THE DATA LAKE DREAM

Edd Dumbill • @edd
edd@svds.com • svds.com/StrataNY2014
WHAT IS A DATA LAKE?

A scalable, accessible repository of data

(in its natural or processed state)
CONVENTIONAL DATA STRATEGY
“WHAT YOU DO TO DATA”

CLEAN

VALIDATE

CONTROL

PROTECT
MODERN DATA STRATEGY

“WHAT YOU DO WITH DATA”

ATTRACT NEW CUSTOMERS

TARGET VIP CUSTOMERS

AUTOMATE
TOWARDS THE “DATA LAKE” — Step 1
TOWARDS THE "DATA LAKE" — Step 2
TOWARDS THE “DATA LAKE” — Step 3
TOWARDS THE “DATA LAKE” — Step 4
Different use cases put different demands on the data infrastructure. Increasing cost per unit of capability from scale-up architectures causes rationing of resources. Only the most valuable use cases are pursued.
THE DATA VALUE CHAIN
DRAW VALUE FROM YOUR STRATEGIC DATA ASSETS

Discover    Ingest    Process    Persist    Integrate    Analyze    Expose
• Make it **cheap**
  • Failure as a feature
  • Ask good questions
• Make it **quick**
  • Both learning and adaptation
  • Enable the feedback loop
• **Don’t break things**
  • Make operations a platform for innovation
  • APIs, platforms, simulation
THE EXPERIMENTAL ENTERPRISE

Data science allows us to observe our experiments and respond to the changing environment.

We need to both support investigative work and build a solid layer for production.

The foundation of the experimental enterprise focuses on making infrastructure readily accessible.
Edd Dumbill
edd@svds.com
@edd
@SVDataScience

Yes, we’re hiring!
info@svds.com

Want these slides? Go to:
svds.com/StrataNY2014