Developing a Modern Enterprise Data Strategy

John Akred | @bigdataanalysis
Edd Dumbill | @edd
OUR PHILOSOPHY

• Prioritize for highest business value when using emerging technology.
• Provide highly productive teams of data scientists and engineers.
• Design with outcomes in mind.
• Be agile: deliver initial results quickly, then adapt and iterate.
• Collaborate constantly with our customers.
OUR SERVICES AND CAPABILITIES

DATA STRATEGY
Defining your approach to data, driven by business priorities, resulting in a prioritized action plan.

ARCHITECTURE DESIGN
Designing architectures that enable rapid development of your data-driven platforms and applications.

AGILE ENGINEERING
Prototyping, piloting, and deploying capabilities and systems using cross-functional teams of data scientists and engineers.

AGILE DATA SCIENCE
Investigating hypotheses to expose and validate analytical narratives, and building actionable models from your data.
INTRODUCTION

TODAY’S SCHEDULE

Introduction

Why Have a Data Strategy?

Connecting Data with the Business

New Technology Potentials

The Data Platform Architecture

Break

Devising a Data Strategy

The Chief Data Officer

The Experimental Enterprise
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.
THE DATA VALUE CHAIN
DRAW VALUE FROM YOUR STRATEGIC DATA ASSETS

Discover    Ingest    Process    Persist    Integrate    Analyze    Expose
What’s on your mind?

What is preventing your organization from realizing its vision?
TODAY’S SCHEDULE

Introduction

Why Have a Data Strategy?

Connecting Data with the Business

New Technology Potentials

The Data Platform Architecture

Break

Devising a Data Strategy

The Chief Data Officer

The Experimental Enterprise

WHY HAVE A DATA STRATEGY?
Data strategy is not for the faint of heart

IS THERE AN ALTERNATIVE?

The alternative is to treat data as a cost of business, to be minimized.

In today’s competitive environment, data must serve the **strategic imperatives** of a business - the key strategic aspirations that define the future vision for an organization.

A modern data strategy is a roadmap to enable **data-driven decision-making** and **applications** that helps an enterprise achieve its strategic imperatives.

An effective data strategy helps an enterprise make technology choices, grounded in **business priorities**, to get the most value from their data.
CONNECTING TECHNOLOGY AND BUSINESS VALUE

If you find that you can’t articulate how the cost of your data systems relates to the benefits to your business, or if you can’t articulate how your technology philosophy enables your business aspirations, then your organization would almost certainly benefit from data strategy.
Poll

- Raise your hand if you feel like the technology leadership in your organization prioritizes investments to meet the ambitions of the business?
- Can your organization clearly articulate the business impact of the data and technology investments it makes?
TODAY’S SCHEDULE

Introduction

Why Have a Data Strategy?

Connecting Data with the Business

New Technology Potentials

The Data Platform Architecture

Break

Devising a Data Strategy

The Chief Data Officer

The Experimental Enterprise
CONVENTIONAL DATA STRATEGY

“WHAT YOU DO TO DATA”

CLEAN

VALIDATE

CONTROL

PROTECT
Conventional Wisdom: 10 THINGS A DATA STRATEGY SHOULD INCLUDE

1. What data should be collected?
2. How long should data be kept?
3. Where should the data be stored?
4. How will data privacy and security be managed?
5. From where can data be accessed?
6. What data can be displayed?
7. What level of detail should be retained?
8. Who is responsible for the data (governance)?
9. How is data integrated?
10. How will data be distributed? (virtualization?)

2 10 Key Elements of your Data Strategy; Mike Schiff – http://www.tdwi.org//Articles/2012/01/17/10-Elements-Data-Strategy.aspx?Page=1
MODERN DATA STRATEGY

“What you do with data”

ATTRACT NEW CUSTOMERS

TARGET VIP CUSTOMERS

AUTOMATE
NEW ORTHODOXY?

4 PRINCIPLES OF A SUCCESSFUL DATA STRATEGY

1. How does data generate value?
2. What are our critical data assets?
3. What is our data ecosystem?
4. How do we govern data?

NOT ALL DATA IS EQUAL

Conventional data strategy

EDW
Governance
Security
NOT ALL DATA IS EQUAL

Modern data strategy
WHAT IS A DATA STRATEGY?

Business strategic ambitions

Existing data & technology

Possible data & technology

Priorities

Constraints

Roadmap of investments

Tools to update and assess roadmap
BEGIN WITH THE BUSINESS

• First understand what drives your business
• Then make the leap from strategy to tactics

@Technologists: This can’t be done without the business leaders in the room

@BusinessLeaders: This can’t be done without the technologists in the room
Understand the strategic imperatives of your organization

- Annual report
- Investor updates
- Talk to leadership
REAL ESTATE MARKETPLACE: ZILLOW

Business Objectives

• Build and maintain best algorithms for pricing
  • Use Hedonic pricing method to incorporate multiple attributes and ‘nearest neighbors’ to create accurate Zestimate®
  • Deploy sophisticated and adaptive models, at scale (over 110 million homes) and at timely interval (3 times / week)
  • Use scalable infrastructure (cloud) for rapid analysis
• Build industry’s best real estate data sets
  • Increase completeness of data by include public data sets such as construction listings, foreclosure listings, market context
  • Capture unique data with customer reviews and feedback from real-estate firms
  • Manage scale of 110 million properties and growing

Strategic Imperatives

• Provide products and services to help consumers with every stage of home ownership – buying, selling, renting, borrowing, and remodeling
• Generate more subscription and ad revenue
• Drive more unique users to marketplace
• Become leading real estate and home-related information marketplace on mobile and web

NOTE: Zillow is not an SVDS client.
HEALTH PROVIDER: KAISER PERMANENTE

Strategic Imperatives
• Provide seamless, personalized care through an integrated team of care providers
• Enable members to manage their own care through easy-to-use channels
• Transform care and improve outcomes through investments in research and innovation

Business Objectives
• Increase data sharing with extended care teams through secure electronic health record access
• Provide quicker, better diagnoses through evidence-based medicine techniques
• Provide mobile access to scheduling, pharmacy interactions, and other related services
• Improve member satisfaction by analyzing web and mobile user interactions, behavior, and feedback data
• Share access to knowledge, innovation, and population data with the public and other health care leaders

NOTE: Kaiser Permanente is not an SVDS client.
TODAY’S SCHEDULE

Introduction

Why Have a Data Strategy?

Connecting Data with the Business

New Technology Potentials

The Data Platform Architecture

Break

Devising a Data Strategy

The Chief Data Officer

The Experimental Enterprise

NEW TECHNOLOGY POTENTIALS
Why Big Data?

1. New Capabilities
2. Economic Scalability
Existing revenue streams:

- Ads
- Price quotes (leads)

Shopping is the focus:

- Need real-time inventory
- Accurately described VIN’s
DATA PLATFORMS FOR ECONOMIC SCALABILITY at NetApp

NOTE: NetApp is not an SVDS client.
Different products and features put different demands on the data infrastructure.

Increasing cost per user from scale-up architectures causes a barrier to economic expansion of the product user base.
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.
TODAY’S SCHEDULE

Introduction

Why Have a Data Strategy?

Connecting Data with the Business

New Technology Potentials

The Data Platform Architecture

Break

Devising a Data Strategy

The Chief Data Officer

The Experimental Enterprise
towards a production enterprise data architecture
DATA PLATFORM

Data Acquisition
- Internal
- External

Data Ingest
- Offline Processing
- Real-Time Processing

Low Latency Access
- Persistence

Data Repository
- Batch Processing

Data Storage
- External Systems
- Data Services

Data Services
- Data Acquisition
- Internal
- External

Analytics

Data Management
- Security, Operations, Data Quality, Meta Data Management and Data Lineage
### CHOICES: TOOLS

#### Accounts
- **Accounts
  - **Accounts.csv**
  - Sample Size: 488,284
  - New Sample Ready

#### Transform Editor
- **Script**
  - `unnest col: votes before: votes` keys: `cool`, `funny`, `useful`

#### Data Preview
- **Columns**
  - `date`
  - `votes`
  - `votes_cool`
  - `votes_funny`
  - `votes_useful`
  - `review_id`

#### Data Table
- **2018-08-12**
  - `date`: 2018-08-12
  - `votes`: 6
  - `votes_cool`: 3
  - `votes_funny`: 1
  - `votes_useful`: 2
  - `review_id`: 63

---

@SVDataScience
CHOICES: DATABASES

Graph
Social networks
Ontologies
Knowledge, Property

Document
Logging
Document archive
Web content

Key-Value
Shopping Cart
Session Data

Columnar
Sensors
Network devices
Internet of Things

Technical Use Cases
CHOICES: DATABASES

SPECIALIZED

Graph
Social networks
Ontologies
Knowledge, Property

Document
Logging
Document archive
Web content

Key-Value
Shopping Cart
Session Data

Columnar
Sensors
Network devices
Internet of Things

DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED

CHOICES: DATABASES

SPECIALIZED
CHOICES: DATABASES
GENERAL PURPOSE

Graph
- Social networks
- Ontologies
- Knowledge, Property

Document
- Logging
- Document archive
- Web content

Key-Value
- Shopping Cart
- Session Data

Columnar
- Sensors
- Network devices
- Internet of Things
Batch:

• Using FFT transformed frequency data identify the train based around fundamental frequencies of train whistle.

• Construct the decision tree for train classifier based on minimum and maximum fundamental frequencies

Real-Time:

• Apply FFT to audio signal

• Extract min and max fundamental frequencies

• Classify the train into local or express

• Send data to the Event Detector to alert the APP

• Store results in HBase
[Amazon] do services because they've come to understand that it's the Right Thing. There are without question pros and cons to the SOA approach, and some of the cons are pretty long. But overall it's the right thing because **SOA-driven design enables Platforms.**

... You wouldn’t really think that an online bookstore needs to be an extensible, programmable platform. Would you?  
+[Steve Yegge](https://plus.google.com/112678702228711889851/posts/eVeouesvaVX)
CHOICES: DATA RESILIENCY

**Stovepipe:** One-to-one relationship from data source to product.

**Hard Failure:** If the data source is broken, so is the app.

**Multi-sourced:** Redundancy of overlapping data sources makes your products more resilient.

**Graceful Degradation:** If a data source breaks, there is a backup and your app continues to function.

Production data services abstract the probabilistic integration of overlapping data sources. We call this model a **Data Mesh**.
CHOICES:
External Systems:
applications, visualization, business intelligence
DEFINING SUCCESS

✓ Incremental revenue
✓ Time to market
✓ Economically viable implementation
✓ Cost avoidance
✓ Brand benefit
✓ Goodwill
TODAY’S SCHEDULE

Introduction

Why Have a Data Strategy?

Connecting Data with the Business

New Technology Potentials

The Data Platform Architecture

Break

Devising a Data Strategy

The Chief Data Officer

The Experimental Enterprise
HOW SVDS DOES DATA STRATEGY

• We work with your business, product, and technology leaders to analyze and articulate a data strategy.

• The data strategy will provide an actionable roadmap that generates immediate value and provides the foundation for future capability investments.

• Our focus is to understand your current business and technology landscapes. With this insight, we will provide an actionable technology roadmap designed to help you achieve your business ambitions. The implementation of this roadmap will unlock untapped business opportunities.

• Our collaborative approach ensures that your business, product, and technology teams will become an effective advocate for your data strategy within your organization.
A product research and recommendation company is transforming their core business from content and information services to a referrer of high-value transactions to partners.

SVDS devised a data strategy that enables new analytical capabilities core to their retail ambitions, addressing critical accuracy and timeliness issues with unstructured data.

Based on this data strategy, they are building a solution for near real-time product inventory that increases their value to partners in a complex, multi-tier market.
A media and entertainment company seeks to deliver personalized content directly to users on digital entertainment devices.

SVDS developed a data strategy and architecture that enables real-time data ingestion, deeper customer insight, and highly-personalized content recommendations.

The data strategy and architecture design now serve as the foundation for iterative, new product development and guide technology investments and acquisitions.
OUR METHOD FOR DATA STRATEGY

IDENTIFY STRATEGIC IMPERATIVES

DEFINE BUSINESS OBJECTIVES

DEFINE DATA REQUIREMENTS

IDENTIFY GAPS IN CURRENT SYSTEMS & TECHNOLOGY

MAP OBJECTIVES AND REQUIREMENTS TO USE CASES

RATIONALIZE USE CASES INTO WORKLOADS

PROJECT ACTION PLAN & ROADMAP
IDENTIFY YOUR STRATEGIC WORKLOADS

USE CASE
1
WORKLOAD A
WORKLOAD B
WORKLOAD C

USE CASE
2
WORKLOAD B
WORKLOAD C

USE CASE
3
WORKLOAD B
WORKLOAD D
AN EXAMPLE
DATA STRATEGY FOR THE DOGS
AN EXAMPLE
DATA STRATEGY FOR THE DOGS

“...We’ve been investing in new capabilities to help us capture and use customer and pet data, and this year, we will deliver on new methods to use this data to drive growth.

— David Lenhardt
PetSmart CEO

NOTE: PetSmart is not an SVDS client. This is a fictional example based on public information.
http://risnews.edgl.com/retail-news/PetSmart-Leverages-Analytics-for-Personalized-Experience91783
STRATEGIC IMPERATIVE

Connect with pet parents in a personalized way
BUSINESS OBJECTIVES

Capture and use customer and pet data
Deliver personalized recommendations and offers
Track and share real-time store inventory
USE CASES

Recommend new pet products based on past purchases at point of sale

Recommend upcoming store/community events based on customer preferences

…
WORKLOAD

Recommendation engine
FOCUS ON THE VALUE

PRIORITIES

DIMENSIONS

OVERCOME YOUR ASSUMPTIONS

© 2015 SILICON VALLEY DATA SCIENCE LLC. ALL RIGHTS RESERVED.
### Business Objectives

<table>
<thead>
<tr>
<th>Information</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Illustrative DATA CONSTRAINTS

- latency
- breadth
- frequency
- no gaps
- gaps
- not applicable
# Illustrative Development Horizons

<table>
<thead>
<tr>
<th>Technical Workloads</th>
<th>Use Cases</th>
<th>Business Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-time recommendations</td>
<td>10</td>
<td>Increase Customer Loyalty</td>
</tr>
<tr>
<td>Omnichannel data integration</td>
<td>10</td>
<td>Expand Brand Presence &amp; Reach</td>
</tr>
<tr>
<td>Predictive modeling</td>
<td>9</td>
<td>Predict Region-Level Supply / Demand</td>
</tr>
<tr>
<td>Unstructured text analysis</td>
<td>9</td>
<td>Target VIPs with promotions</td>
</tr>
<tr>
<td>Behavioral analytics</td>
<td>8</td>
<td>Understand customer buying experience</td>
</tr>
<tr>
<td>Data quality monitoring</td>
<td>8</td>
<td>Personalize support experience</td>
</tr>
<tr>
<td>Pattern recognition</td>
<td>7</td>
<td>Identify new customer segments</td>
</tr>
<tr>
<td>Heterogeneous data storage</td>
<td>6</td>
<td>Create services for local regions</td>
</tr>
<tr>
<td>Data Ingestion</td>
<td>5</td>
<td>Increase information accuracy</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Increase service relevance</td>
</tr>
</tbody>
</table>

**Highest combined priority**: I

**Second highest combined priority**: II

**Third highest combined priority**: III

**Lowest combined priority**: IV
### TECHNICAL WORKLOAD PRIORITIZATION

#### Illustrative

<table>
<thead>
<tr>
<th>TECHNICAL WORKLOAD</th>
<th>STRATEGIC VALUE</th>
<th>TECHNICAL FEASIBILITY</th>
<th>ACCESSIBILITY OF REQUIRED SKILLS</th>
<th>ARCHITECTURAL FIT</th>
<th>PROD ROLL-OUT EFFORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real time recommendations</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omnichannel data integration</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictive modeling</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstructured text analytics</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral analytics</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data quality monitoring</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pattern recognition</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneous data storage</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data ingestion</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
DEFINE YOUR ROADMAP
We define a project plan to build a specific capability.

For each capability, we describe a project to build technical workloads that implement use cases that address high-priority business objectives.

Silicon Valley Data Science employs an agile development processes as we work with our clients from planning and proof-of-concepts to pilot implementations and finally full scale production systems.
PATH FORWARD Illustrative

Horizon I

Horizon II

Horizon III

Horizon IV

PERSONALIZATION ENGINE

CAPABILITY ROADMAP

CAPABILITY ROADMAP

CAPABILITY ROADMAP

CAPABILITY ROADMAP

2-3 months

5-6 months

3-4 months

3-4 months

months

© 2015 SILICON VALLEY DATA SCIENCE LLC. ALL RIGHTS RESERVED.
DEVISING A PROJECT PLAN – INPUTS & APPROACH

Workload Rationalization

Horizons of Development

Data Gaps

Technical Workload Assessment

Project Roadmaps
LATHER

RINSE

REPEAT
MAKE SURE IT’S FLEXIBLE

• Technology moves incredibly fast, and competitive landscapes are highly dynamic.
• Your data strategy should be a living document, revisited often and revised as conditions change.
MAKE SURE IT’S ACTIONABLE

• If it isn’t clear how you’re going to execute your strategy, then you don’t have the right one.
• Must work within the realm of the possible and practical.
FROM IDEA TO PRODUCTION

We identify the business goals, distill those into use cases, and then work in short, iterative cycles to achieve tangible gains.
Aging data infrastructure and brittle application integration was inhibiting growth and business insight for a health management company.

Their data strategy focused on creating a concrete roadmap for migrating to a new data platform so that technology and infrastructure are no longer a barrier to growth and transparency.

Based on this data strategy, they are building a new data platform in stages that allows them to add new products and services to capture more market opportunity.
Case Study: Data Strategy
Major Pharmaceutical Company

Defined Data Strategy that will help enable business growth and enable expansion into new markets

Challenge
• Ongoing need to improve discovery and better predict new targets for drug development
• Difficulty to integrate new data sources into identification & discovery processes
• Inability to connect business strategy & aims with specific, tangible projects

Solution
• SVDS devised a data strategy with a concrete roadmap for migrating to a new data platform
• Recommended data technology & architecture which supports highest value projects
• Outlined cultural, technological, organizational, and collaboration challenges & objectives

Results
• Identified specific opportunity areas to increase GTM efficiency
• Prescribed Common Data and Analytics Platform for Commercial and R&D operations
• Recommended projects for Predictive Modeling & Data Exploration
DATA STRATEGY CHECKLIST

- Identify your business objectives
- Go from objectives to tactics
- Include all stakeholders in the conversation
- Look at how technology can support strategic workloads
- Exploit patterns and reuse
- Prioritize the possibilities to figure out where to start
- Define your roadmap with an end-point in mind
- Lather, rinse, repeat
TODAY’S SCHEDULE

Introduction

Why Have a Data Strategy?

Connecting Data with the Business

New Technology Potentials

The Data Platform Architecture

Break

Devising a Data Strategy

The Chief Data Officer

The Experimental Enterprise
DO YOU NEED EXECUTIVE HELP?
To download a free PDF, go to: svds.com/CDOreport
EMERGENCE OF THE CDO

• Started with heavily regulated industries such as government and finance
• Now becoming common in “disruptable” industries such as retail and telecommunications
RESPONSIBILITIES OF THE CDO

• Centralization:
  • Data from internal silos
  • Data from external APIs and real-time streams
  • The organization’s priorities
RESPONSIBILITIES OF THE CDO

• Evangelization:
  • Technical chops, business savvy, and the diplomacy skills to translate between the two
RESPONSIBILITIES OF THE CDO

• Facilitation:
  • Coordinate stakeholders across the organization
  • Free up resources and lower barriers
  • Offer tools and training to help others succeed
CHALLENGES FOR THE CDO

• Building technical bridges:
  • Working with data in different silos, formats, etc.
  • Mining for business value:
  • “If you don’t have good business questions it doesn’t matter what kind of technology you have.” — Joy Bonaguro
UNDERSTANDING THE CDO

“While technology is inevitably involved when working with data, the defining goal of the CDO is not technological, but business-oriented. The ideal CDO exists to drive business value.”
DECIDING TO HIRE A CDO

• Know why you want one
  • Are you part of a regulated industry?
  • Do you need to move from being product-centric to customer-centric?
  • Could you add products or services?
  • Could your current processes and outcomes be optimized even further?
  • Are there insights in one part of your company that could benefit others?
DECIDING TO HIRE A CDO

• Look for the right skill set
  • Technical chops
  • Business savvy
  • Diplomacy and political skills
  • Executive-level experience
THE AVAILABILITY GAP

“The spike in demand for Chief Digital Officers has been felt globally. In Europe, the number of search requests for this role has risen by almost a third in the last 24 months. The United States has seen the same growth in half that time.”

— Russell Reynolds Associates
PREPPING FOR SUCCESS

Companies that are eager and prepared for real change will be the most appealing to qualified CDO candidates.
TODAY’S SCHEDULE

Introduction

Why Have a Data Strategy?

Connecting Data with the Business

New Technology Potentials

The Data Platform Architecture

Break

Devising a Data Strategy

The Chief Data Officer

The Experimental Enterprise
“... let's seek to understand how the new generation of technology companies are doing what they do, what the broader consequences are for businesses and the economy.”

– Marc Andreessen
DIGITAL NERVOUS SYSTEM
Data is your business
Average company lifespan on S&P 500 Index (in years)

Projections based on current data

Year (each data point represents a rolling 7-year average of average lifespan)

DATA: INNOSIGHT/Richard N. Foster/Standard & Poor’s
SILICON VALLEY’S DATA MACHINE

- UBER
- Google Shopping Express
- Amazon Fresh
- NETFLIX
- TESLA MOTORS
- STITCH FIX
- Square
- FLATIRON
- evolv
- IFTTT
big data applications

well understood systems
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 legacy of big data is business agility
• 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.
LEAD A DATA REVOLUTION

- Can only win with situational awareness
- New architectures offer new opportunities
- Creation of data-driven value requires new approach
- Create an Experimental Enterprise
- Business must lead, and understand the potential of the technology
Yes, we’re hiring!

info@svds.com

john@svds.com
@bigdataanalysis

edd@svds.com
@edd
TODAY’S SCHEDULE

Introduction

Why Have a Data Strategy?

Connecting Data with the Business

New Technology Potentials

The Data Value Chain

The Data Platform Architecture

Break

Devising a Data Strategy

The Chief Data Officer

The Experimental Enterprise

THE DATA VALUE CHAIN

@SVDataScience
THE DATA VALUE CHAIN

DRAW VALUE FROM YOUR STRATEGIC DATA ASSETS

Discover    Ingest     Process     Persist     Integrate     Analyze     Expose
THE DATA VALUE CHAIN—DISCOVER

- Of all the ‘V’s, variety holds most value
- Getting data is hard
  - Silos, politics
  - Can’t search databases
  - External data sources can be complex & opaque
- Assessing utility: frequency, latency, breadth, depth
- Exploration needed to assess utility
- Value is aggregate over problems
- Consider data multi-sourcing
THE DATA VALUE CHAIN—INGEST

- As varied as your data sources
  - Formats, protocols, schemas, frequency, volume
- Critical to your entire data pipeline
- Streaming ingest
  - Capacity & reliability
  - Computing context
  - Data volume: store everything?
- Batch ingest
  - Semantics (update, replace, augment)
  - Acquisition, behavior on error
  - Consider uses—denormalize, disaggregate
THE DATA VALUE CHAIN—EXPOSE

- Expose data services
- Don’t build new silos
- Don’t create a free-for-all
- Make data available as building blocks
  - Abstract away complexity
- Create robust data sources