DEVELOPING A MODERN ENTERPRISE DATA STRATEGY

Edd Wilder-James, Scott Kurth

March 2017
TODAY’S SCHEDULE

Introduction

Why Have a Data Strategy?
Connecting Data with the Business
Understanding Data Gaps
The Data Platform Architecture

Break

Identifying Strategic Workloads
The Chief Data Officer
The Experimental Enterprise
To view SVDS speakers and scheduling, or to receive a copy of our slides, go to:

www.svds.com/StrataCA2017
Silicon Valley Data Science is a boutique consulting firm focused on transforming your business through data science and engineering.
WE DO DATA RIGHT

• We work in cross-functional teams made up of data scientists, engineers, and solutions architects.
• We combine enterprise know-how with custom methods derived from Silicon Valley best practices.
• We use an Agile Software Development approach to make rapid progress against difficult problems that require flexibility.
• We focus on delivering business value as early as possible, then iterating toward the larger goal.
OUR SERVICES

DATA STRATEGY

ARCHITECTURE

AGILE ENGINEERING

AGILE DATA SCIENCE
THE EXPERIMENTAL ENTERPRISE

Conducts experiments and responds to the changing environment.

Supports investigative work and builds a solid layer for production.

Makes foundational 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

Understanding Data Gaps

The Data Platform Architecture

Break

Identifying Strategic Workloads

The Chief Data Officer

The Experimental Enterprise

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

* Creating an Enterprise Data Strategy by Wayne Eckerson
The alternative is to treat data as a cost of business, to be minimized.

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
• you can’t articulate how your technology philosophy enables your business aspirations

then your organization would almost certainly benefit from data strategy.
ARTICULATING THE BUSINESS IMPACT OF DATA & TECHNOLOGY

Poll:

• Is 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

Understanding Data Gaps
The Data Platform Architecture

Break

Identifying Strategic Workloads
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?)

* 10 Key Elements of your Data Strategy by 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
A NEW ORTHODOXY?
FOUR 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?

* The 4 Principles of a Successful Data Strategy by Paul Barth
NOT ALL DATA IS EQUAL

Conventional data strategy

EDW

Governance
Security
NOT ALL DATA IS EQUAL
WHAT IS A DATA STRATEGY?

- **Business strategic ambitions**
  - Constraints
  - Plan to update capabilities
  - Roadmap of investments

- **Existing data & technology**
  - Priorities
  - Tools to update and assess roadmap

- **Possible data & technology**
A NEW NOTION OF MATURITY

Modern Role of Data:
Represents the new role data and analytics play in the enterprise.

Outcomes, not Operations:
A strategic notion of maturity should begin with value creation before addressing underlying operational processes.

Transforming Pragmatically:
Changes are grounded in the holistic view of the future state of your enterprise.
An organization’s ability to derive value from its data defines its maturity.

Illustrative
DIMENSIONS OF DATA MATURITY

Not just the technology!

• People
• Processes
• Systems
CURIOS WHERE YOU FALL?

Maturity Mini-Assessment

• 20Q survey (5-10 min)
• Identifies your stage and provides general recommendations
• Creates baseline for future performance and growth

Illustrative

dmm.svds.com
YOU NEED A DATA STRATEGY WHEN...

- Infrastructure is holding back growth
- Infrastructure is holding back development
- Analog to digital transformation
- Changing business models
- Unifying fragmented offerings
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

**Business Leaders:** 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
Break down the strategic imperatives to make them tangible, achievable, and measurable. These become your business objectives.

Business objectives provide the guide for many other analyses in building your data strategy.
REAL ESTATE MARKETPLACE: ZILLOW

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

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

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.
REAL ESTATE MARKETPLACE: ZILLOW

NOTE: Zillow is not an SVDS client.

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
REAL ESTATE MARKETPLACE: ZILLOW

NOTE: Zillow is not an SVDS client.

BUSINESS OBJECTIVES

1. 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
REAL ESTATE MARKETPLACE: ZILLOW

NOTE: Zillow is not an SVDS client.

BUSINESS OBJECTIVES

2. 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
HEALTH PROVIDER: KAISER PERMANENTE

NOTE: Kaiser Permanente is not an SVDS client.

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
HEALTH PROVIDER: KAISER PERMANENTE

NOTE: Kaiser Permanente is not an SVDS client.

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
HEALTH PROVIDER: KAISER PERMANENTE

NOTE: Kaiser Permanente is not an SVDS client.

BUSINESS OBJECTIVES

• 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
TODAY’S SCHEDULE

Introduction

Why Have a Data Strategy?

Connecting Data with the Business

**Understanding Data Gaps**

The Data Platform Architecture

*Break*

Identifying Strategic Workloads

The Chief Data Officer

The Experimental Enterprise
Commonly-asked questions:

- Do I have gaps in my data?
- How good is my data?
- Is my data clean enough?

None of these questions make sense unless you ask:

*For what?*
FOR WHAT?

• Do I have gaps in my data?
  ...for understanding customer purchase behavior

• How good is my data?
  ...for predicting quarterly sales

• Is my data clean enough?
  ...for automating production
• What are you trying to achieve as a business [with data]? *These are your business objectives.*

• How do you plan to achieve it [with data]? *These are your use cases.*

UNDERSTAND YOUR BUSINESS GOALS
UNDERSTAND YOUR AUDIENCE

Who is going to use this analysis and how?

• CDO? Heads of Business Units? Data Science Directors? DBAs?

• Project assessment? Operational dashboard? Continuous improvement plan?

Understanding stakeholders and expectations will dictate the level of technical analysis required.
UNDERSTAND YOUR AUDIENCE

What are the dimensions of requirements that matter to your audience?

• For a technical application, it might be depth, breadth, latency, frequency.

• For an executive perspective, it might be higher-order requirements like ease of integration or coverage.

What are the questions your audience needs answered? Select the dimensions that provide visibility into those questions.
• Start with an effective catalog of your data.
• Organize the data to be effective. Think about how data is produced AND how it gets used in your organization.
  • By data source?
  • By entity?
  • By organization?
  • By data owner?
LINK IT ALL TOGETHER

Business Objectives

Use Cases

Requirements

Data
VISUALIZE YOUR GAPS

Business Objectives

Information

Customer

Product

Internal

© 2017 SILICON VALLEY DATA SCIENCE LLC. ALL RIGHTS RESERVED.
SO... WHAT IS A "GAP"?

Two schools of thought:

- **Purists**: If a requirement isn’t met, it’s a gap.
- **Pragmatists**: If you can still get the job done, it *isn’t* a gap.

Both views can be valuable ways of looking at your analysis.
TODAY’S SCHEDULE

Introduction

Why Have a Data Strategy?

Connecting Data with the Business

Understanding Data Gaps

The Data Platform Architecture

Break

Identifying Strategic Workloads

The Chief Data Officer

The Experimental Enterprise
WHY BIG DATA?

1. New Capabilities
2. Economic Scalability
Edmunds.com wanted to reduce time-to-market by speeding creation of attribute data for new car models.

We developed a new capability to automatically extract vehicle features from specification guides and categorize the features into appropriate vehicle classes.
Existing revenue streams:

- Ads
- Price quotes (leads)

Shopping is the focus:

- Need real-time inventory
- Accurately described VINs
DATA PLATFORMS FOR ECONOMIC SCALABILITY at NetApp

NOTE: NetApp is not an SVDS client.

UP VS. OUT — SAAS EDITION

- Users
- Revenue
- Scale-out cost
- Cost to serve
- Loss
- Profit

$ ¥ € £
UP VS. OUT — ENTERPRISE EDITION

Data Resource Usage

Scale-up cost

Scale-out cost
BIG DATA
... it’s really about agility
BUYING AGILITY

- Linear scale-out cost
- Opex vs. capex
- Ease of purchase
Scale-out systems move us from managing scarcity to promoting utility.
• Architectural factors
  • Schema on read
  • Rapid deployment
  • Mirror production setup
  • Executes faster
• Programmer factors
  • Fun to program
  • Concision
  • Easier to test
  • Faster to write
WHAT IS DOCKER?

- Container technology: bundles every part of an application
- Provides isolation for each application without the overhead of running a virtual machine
  - Ships only the parts that are needed—leaves out the operating system
WHY SHOULD BUSINESS CARE?

• Better use of server resource than virtual machines
• A fast and reliable way of deploying applications
  • It’s the ideal packaging mechanism for scale-out distributed systems
• Easy for developers to work in an environment identical to production
  • Sharing containers leads to innovation
WHAT IS APACHE KAFKA?

- Scale-out fault-tolerant messaging system
- Comes from LinkedIn
- Supported by Confluent
USE CASES

• Stream processing
• Log aggregation
• Creating decoupled evented architectures
WHY SHOULD BUSINESS CARE?

- Scalability in a critical area of distributed applications
- Online reliability, compared to alternatives
- Will be a core building block of distributed data architecture
WHAT IS APACHE SPARK?

• In-memory distributed computing platform
• Comes from Berkeley AMPLab
• In production with early adopters, now integral to every commercial Hadoop distribution
• Doesn’t need Hadoop, but runs easily on top
USE CASES

• Managing a major retailer’s inventory across a diverse network of entities in near real time
• Managing and processing event streams for online gaming
• Supporting data science initiatives across massive data sets at a media analytics company
WHY SHOULD BUSINESS CARE?

• Enables use cases Hadoop didn’t provide, all in one platform
  • streaming, interactive analytics, machine learning, graphs
• Fast
  • Iteration time down, more productive
• Use existing cluster investment
  • Sits on HDFS, can run under YARN (or use Amazon S3, or Cassandra)
WHY SHOULD BUSINESS CARE?

• SparkSQL
  • Use SQL skills and tools, e.g. Tableau
  • Dataframes integrate external data sources into one context: RDBMS, Hive, JSON...

• Developer-friendly
  • Concise and fluid to program
  • Language integration: Scala, R, Python, Java
WHAT ARE NOTEBOOKS?

• Interactive documents that contain a program and its output
  • Long history: Mathematica
• Particularly successful with data science
• Projects to watch
  • Jupyter — https://jupyter.org/
  • Apache Zeppelin — https://zeppelin.incubator.apache.org/
Running average

If motion occurs, the image changes. Therefore detecting motion requires detecting differences in what has been seen previously. To reduce sensitivity to small-scale motion, we want to use a combination of frames to compare to our current frame. In OpenCV, the running average function relies on a parameter \( \alpha \) that indicates how many of the previous frames are used in the average and how heavily they are weighted.

\[
\text{Running average} = \alpha \cdot \text{current frame} + (1 - \alpha) \cdot \text{running average}
\]

Detect change

To detect change, we take the difference between the running average frame and the current frame.

For simplification, I will replace steps with their corresponding functions in action_detector.py, in the following code.
WHY SHOULD BUSINESS CARE?

• Easy collaboration and sharing of data science
  • Think “Docker for analysis”
• Easy access to data and compute resource
• A building block for more self-service analytical capabilities

Commercial version of Notebooks + Spark is the Databricks Cloud
Towards a production

ENTERPRISE DATA ARCHITECTURE
DATA PLATFORM

Data Management
- Security, Operations, Data Quality, Meta Data Management and Data Lineage

Data Acquisition
- Internal
- External

Low Latency Access
- Data Ingest
  - Offline Processing
- Real-Time Processing

Data Repository
- Persistence
- Batch Processing

Analytics

External Systems

Data Services
CHOICES: TOOLS
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

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

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

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

© 2017 SILICON VALLEY DATA SCIENCE LLC. ALL RIGHTS RESERVED.
@SVDataScience
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

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

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
CHOICES: VELOCITY

SVDS R&D TRAINS

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.

Production data services abstract the probabilistic integration of overlapping data sources. We call this model a **Data Mesh.**

**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.
CHOICES: EXTERNAL SYSTEMS

Applications, visualization, business intelligence
✔ Incremental revenue
✔ Time to market
✔ Economically viable implementation
✔ Cost avoidance
✔ Brand benefit
✔ Ecosystem friendliness
BREAK
TODAY’S SCHEDULE

Introduction
Why Have a Data Strategy?
Connecting Data with the Business
Understanding Data Gaps
The Data Platform Architecture

Break

Identifying Strategic Workloads

The Chief Data Officer
The Experimental Enterprise
HOW SVDS DOES DATA STRATEGY

• We work with your stakeholders to analyze and articulate a data strategy.

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

• We work to understand your current business and technology landscapes in order to unlock untapped business opportunities.

• Our collaborative approach ensures that your business, product, and technology teams become effective advocates 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 BUSINESS OBJECTIVES TO USE CASES

RATIONALIZE USE CASES INTO WORKLOADS

ACTION PLAN & ROADMAP

@SVDataScience
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 IMPERATIVES

Our strategy:
“To be the preferred provider for the lifetime needs of pets.”

- Connect with pet parents in a personalized way
- Attract and retain our most valuable customers
- Provide innovative products & services at fair prices
- Drive consistent execution in our stores
AN EXAMPLE

Connect with pet parents in a personalized way

Deliver personalized recommendations and offers

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

Recommend upcoming store/community events based on customer preferences

Recommendation Engine
BUSINESS OBJECTIVES

Connect with pet parents in a personalized way

1. Learn from consumer interactions
2. Optimize consumer journeys based on insights
3. Deliver personalized content to customers

Illustrative
Deliver personalized content to customers

1. Identify customers
2. Profile behaviors
3. Understand context
4. Anticipate behaviors
5. Optimize personalization

Illustrative
WORKLOADS

Data Value Chain | Example Workloads
---|---
Acquire | • Capture mobile app transactions
• Accessing streaming web activity data
Ingest | • Flexible data ingestion
• Ingest unstructured data
Process | • Data validation
• Omnichannel data integration
Persist | • Heterogeneous data storage
• Scalable data storage
Analyze | • Probabilistic data integration
• Predictive modeling
Expose | • Service based data access
• Interactive visualization
1. Identify customers

- Customer data (Acquire, Ingest, Persist)
- Identity resolution
- Data cleansing
- Householding
- Relationship context
- Life-time Value

Technical Workload
- Acquire multiple data sources & formats
- Flexible data ingestion
- Flexible & scalable data storage and processing
- Probabilistic data integration
- Data validation
- Probabilistic data integration
- Detailed views of entities
- Feature engineering
### Illustrative

#### 2. Profile behaviors

| 360 degree view of customer | Detailed views of entities |
| Views of historical transactions | Time series analysis |
| Determination of ‘favorites’ | Predicting customer behavior |
| Map to archetype | Stream processing |
| Evaluate previously unseen transactions and classify | Stream processing |
| Update archetypes | Feature extraction |
| | Analyze customer behavior |
3. Understand context

Characterize temporal customer behavior

Determine goal of next interaction

Categorize content needs

Technical Workload

• Feature engineering
• Analyze customer behavior

• Predictive modeling
• Predictive modeling
4. Anticipate behaviors

Score product offers with likelihood to respond

Score content options with likelihood to respond

Identify next best action

Technical Workload

- Integrate internal systems
- Service based data access
- Integrate internal systems
- Service based data access
- Third party structured data integration
- Business rules execution
5. Optimize personalization

Apply business rules, constraints to personalization options
Select optimal personalization to achieve goal

Technical Workload
- Business rule execution
- Optimization execution
FOCUS ON THE VALUE

PRIORITIES

DIMENSIONS

OVERCOME YOUR ASSUMPTIONS
## DEVELOPMENT HORIZONS

**Illustrative**

### Technical Workloads

<table>
<thead>
<tr>
<th>Workload</th>
<th># of Use Cases Impacted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-time recommendations</td>
<td>10</td>
</tr>
<tr>
<td>Omnichannel data integration</td>
<td>10</td>
</tr>
<tr>
<td>Predictive modeling</td>
<td>9</td>
</tr>
<tr>
<td>Unstructured text analysis</td>
<td>8</td>
</tr>
<tr>
<td>Behavioral analytics</td>
<td>7</td>
</tr>
<tr>
<td>Data quality monitoring</td>
<td>6</td>
</tr>
<tr>
<td>Pattern recognition</td>
<td>5</td>
</tr>
<tr>
<td>Heterogeneous data storage</td>
<td>3</td>
</tr>
<tr>
<td>Data ingestion</td>
<td>3</td>
</tr>
</tbody>
</table>

### Development Horizons

<table>
<thead>
<tr>
<th>Horizon</th>
<th># of Use Cases Impacted</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>10</td>
</tr>
<tr>
<td>II</td>
<td>9</td>
</tr>
<tr>
<td>III</td>
<td>8</td>
</tr>
<tr>
<td>IV</td>
<td>6</td>
</tr>
</tbody>
</table>

### Business Objectives

1. Increase Customer Loyalty
2. Expand Brand Presence & Reach
3. Predict Region Level Supply / Demand
4. Target VIPS with promotions
5. Build product recommendations
6. Understand customer buying experience
7. Personalize support experience
8. Identify new customer segments
9. Create services for local regions
10. Increase information accuracy
# 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>Green</td>
<td>Yellow</td>
<td>Red</td>
<td>Yellow</td>
</tr>
<tr>
<td>Omnichannel data integration</td>
<td>10</td>
<td>Green</td>
<td>Yellow</td>
<td>Red</td>
<td>Yellow</td>
</tr>
<tr>
<td>Predictive modeling</td>
<td>9</td>
<td>Green</td>
<td>Yellow</td>
<td>Red</td>
<td>Yellow</td>
</tr>
<tr>
<td>Unstructured text analytics</td>
<td>8</td>
<td>Green</td>
<td>Yellow</td>
<td>Red</td>
<td>Yellow</td>
</tr>
<tr>
<td>Behavioral analytics</td>
<td>7</td>
<td>Green</td>
<td>Yellow</td>
<td>Red</td>
<td>Yellow</td>
</tr>
<tr>
<td>Data quality monitoring</td>
<td>6</td>
<td>Green</td>
<td>Yellow</td>
<td>Red</td>
<td>Yellow</td>
</tr>
<tr>
<td>Pattern recognition</td>
<td>5</td>
<td>Green</td>
<td>Yellow</td>
<td>Red</td>
<td>Yellow</td>
</tr>
<tr>
<td>Heterogeneous data storage</td>
<td>3</td>
<td>Green</td>
<td>Yellow</td>
<td>Red</td>
<td>Yellow</td>
</tr>
<tr>
<td>Data ingestion</td>
<td>3</td>
<td>Green</td>
<td>Yellow</td>
<td>Red</td>
<td>Yellow</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

Horizon I

Horizon II

Horizon III

Horizon IV

Illustrative

- Horizon I: 2-3 months
- Horizon II: 5-6 months
- Horizon III: 3-4 months
- Horizon IV: 3-4 months
DEVISING A PROJECT PLAN: INPUTS & APPROACH

Workload Rationalization

Development Horizons

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.

What can we do with data?
MODERNIZING DATA TECHNOLOGY
HEALTH MANAGEMENT COMPANY

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
Understanding Data Gaps
The Data Platform Architecture

Break

Identifying Strategic Workloads

The Chief Data Officer

The Experimental Enterprise
DO YOU NEED EXECUTIVE HELP?
To download a free PDF, go to: www.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.”

— Julie Steele
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
Understanding Data Gaps
The Data Platform Architecture

Break

Identifying Strategic Workloads
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 Index in years (rolling 7-year average)

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
UP VS. OUT

Data Resource Usage

Scale-up cost

Scale-out cost

UC1

UC2

UC3

UC4

UC5

$ € ¥ £
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

BUILD FOR EXPERIMENTS
THE EXPERIMENTAL ENTERPRISE

Conducts experiments and responds to the changing environment.

Supports investigative work and builds a solid layer for production.

Makes foundational infrastructure readily accessible.
LEAD A DATA REVOLUTION

• You 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
To view SVDS speakers and scheduling, or to receive a copy of our slides, go to:

www.svds.com/StrataCA2017
THANK YOU

Edd Wilder-James (@edd)
Scott Kurth (@ScottWKurth)
March 2017