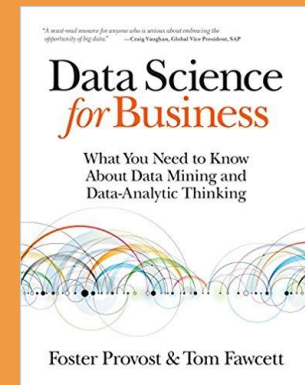
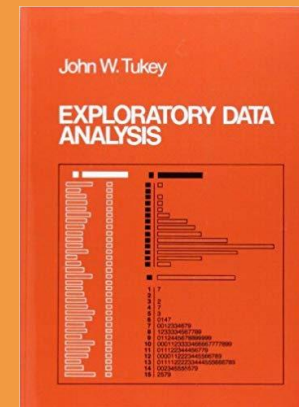
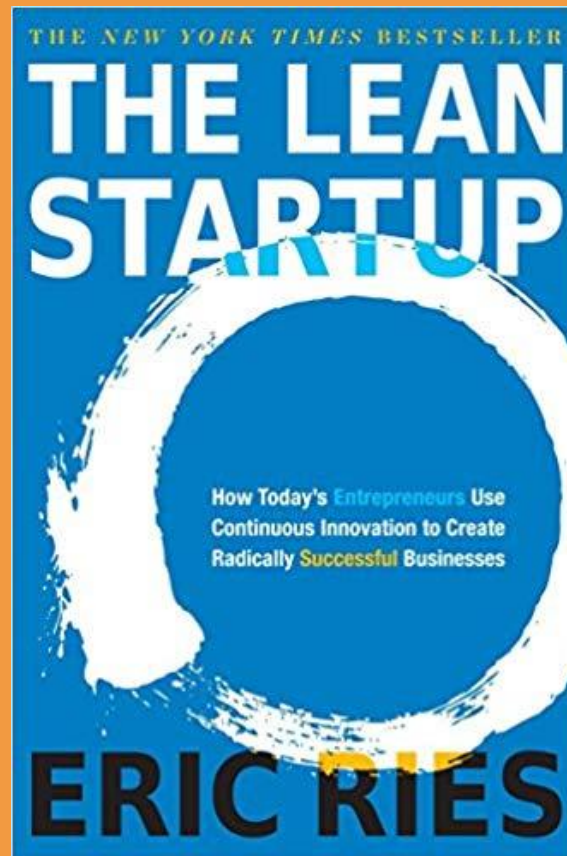
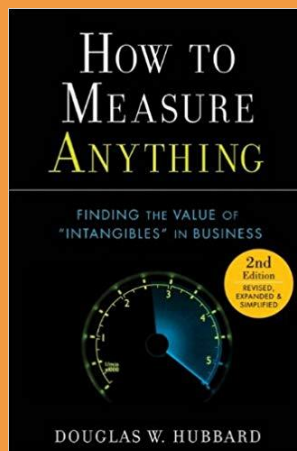
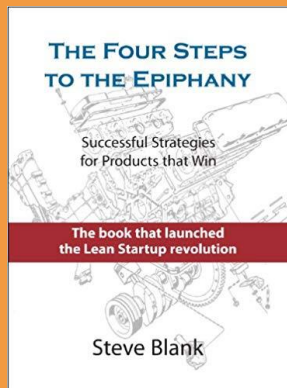
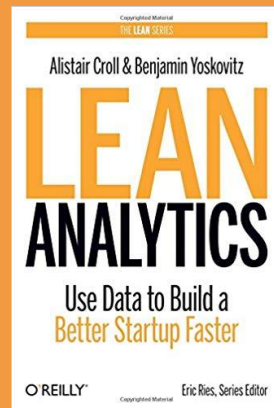


Building a Data-Centric Modern Quality Culture

Brent Jensen
Principal Data Scientist Manager
Azure@MSFT

Podcast: <https://www.angryweasel.com/ABTesting/>
MTP: ModernTesting.org
Twitter: [brentmjensen](https://twitter.com/brentmjensen)





Books recommended

Problem Overview

Why care

Test World is Changing

A recent study* expects:

- QA/Dev roles eventually merge globally

Proposed Modern Test concerns:

- Business requirements over Functional requirements
- Quality over Code Correctness
- Agility/Speed over Robustness

Biggest Challenge:

- Tests not reflecting production



*[Panaya's State of Functional Testing – 2017 Survey](#)

---Internal Use---

Business = Priority #1

Your work today. Is it ...?

Increasing

- Profit
- Market Share
- Customer Retention
- Efficiency

The Big 7

Decreasing

- Cost
- Risk
- Time to Market

How do you know?

"In God we trust, everyone else bring data"
-W. Edwards Deming



Welcome to the *Age of the Customer*

Assume there are 2 pieces of code shipping in the same product, both of the same size and complexity, which one is higher quality?

☐ The one that is certified as bug free (God tested it), but no one is using.

☐ The one that is buggy as heck, but is getting tons of active use.

[Vote](#) [View Results](#) [PollDaddy.com](#)

Quality Redefined?

“No one in the history of mankind has ever wanted software... They want their problems solved”

– Tom Poppendieck



- "Bet the farm" approaches are common
- Avoid Customer Pain
- Prevention justifies delays and costs
- Assume requirements are correct
- Assume bugs matter to customer
- Provide Information

Tradition

But....

Business needs to scale!

- Lots of decisions to be made
- Lots of data
- Lots of threats
- But so little time
 - Calendaring
 - Engineering



More than 50% of my current team are ex-testers.

Why? Customer Empathy and Risk training is invaluable

Solution

Remove the theory.

Quantify risk.

Automate the decision.

Take Action.

Tangent: What is a Data Scientist job?

Help a business understand its present in the future

action by relationships decisions

Top Reasons why I

- "All I do is report data"
- "I produce brilliant insights"
- "Management doesn't listen"
- "I've spent the last 6 months on a project that failed"
- "I spend all my time on requesting data"



Most common request from managers:

"Brent, I want to hire **A** Data Scientist to do all that Data stuff. What should I look for?"

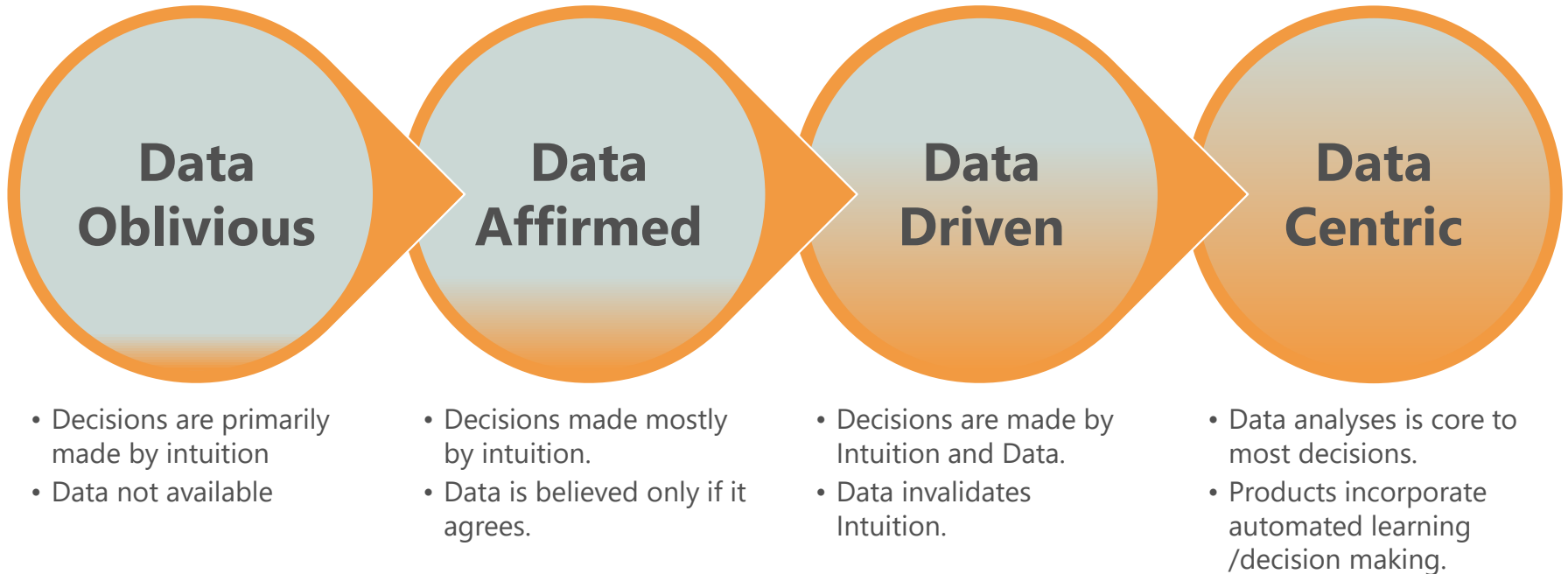
Data Culture Maturity



Intuition is the source of Truth



Data is the source of Truth



The Formula



Align

Connect to Business KPI
Generate Hypothesis
Determine User



Collect

Connect to Data
Curate



Inform

Build Visuals
Democratize



Recommend

Pareto Order
Build Recommender

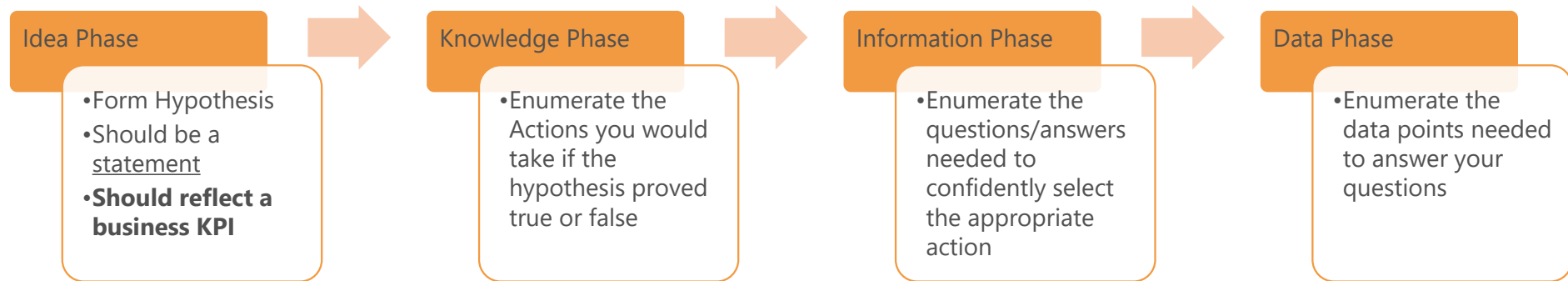


Act

Automate
Productize
Improve accuracy

Align

- Use only Actionable metrics – Not Vanity!!!
 - Simple start: divide your metric by any KPI from the Big7
- Generate Hypotheses from top down



- Get Users
 - Assuming your hypotheses held true, who would act?
 - DO NOT TARGET EXECUTIVES
 - Interview several of your decision makers to validate

Example

(more detail in Episode 82)

Hypothesis: Users sharing dashboards by email will cause a notable increase in Acquisition and Engagement

Possible Actions: (not complete)

- True
 - No action
 - Optimize – make it really easy for users to share by email
 - Enhance – improve the email content to entice receivers
- False
 - Abandon future feature development or cut
 - If acquisition improves, but not engagement, develop new hypothesis.
 - If engagement, but not acquisition, improves, develop new hypothesis.

Possible Questions: (not complete)

- Which users share by email?
- Which users receive shared email?
- Which receivers entered the service due to a shared email?
- How does acquisition via sharing compare to other acquisition means?
- How does sharers engagement level compare with those who don't?
- How does receivers engagement level compare with those who don't?

Data Needed:

- Users who share & receive
- Receivers' acquisition date & means
- Acquisition rate correlated with receiving rate
- Engagement rate correlated with sharing rate
- Engagement rate correlated with receiving rate

Collect

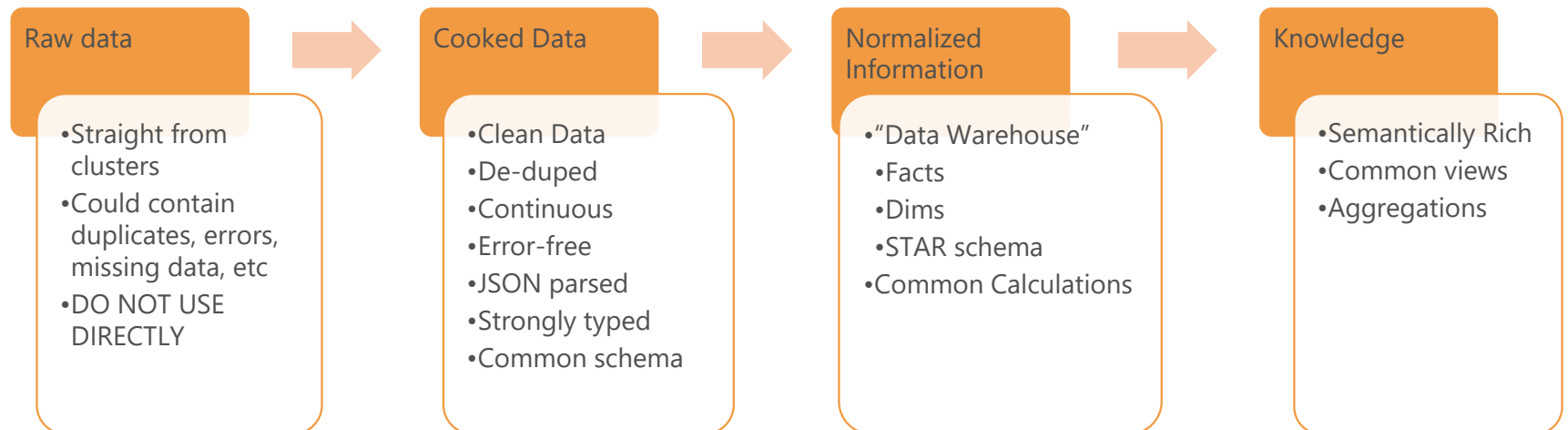
- Always start with product logs
 - Add instrumentation yourself, if need be
- Invest in ELS tech
 - The Cloud is your friend; it has everything you need
 - Open Source goodness:
 - FluentD – Instrumentation Stack
 - ElasticSearch – Store & Query Engine
 - Kibana – Visualization (charts) stack
 - No Silver bullet: research solutions for your context
- Retain Monitor Telemetry
- Curate your data via ETL jobs

Curation

Definitions

- Data – Raw facts that are meaningless by themselves (names, numbers, etc)
- Information – Data organized in a meaningful way
- Knowledge – Information in action; the ability to use information

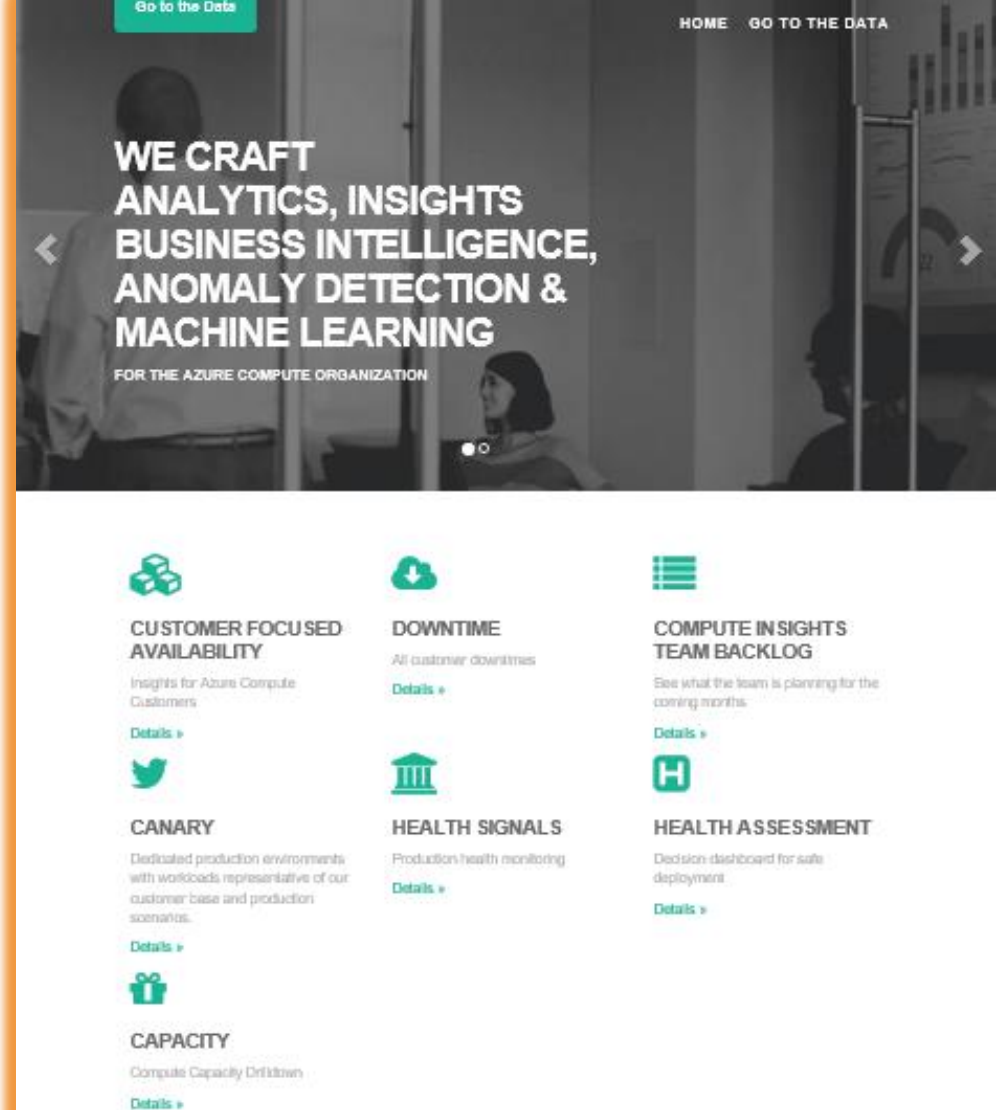
Organize your data into distinct stores/views



Inform

- Create a Portal site and brand it: ours - "Monocle"
- Create Visuals and link
- Instrument all of this
- Avoid 'Data puke' or 'Stoplights'
- Democratize the source data for the visuals
- Schedule Cadence meetings
- STRONGLY consider making data transparent to Business Customer
- Answer Questions
- Exploratory Data Analysis

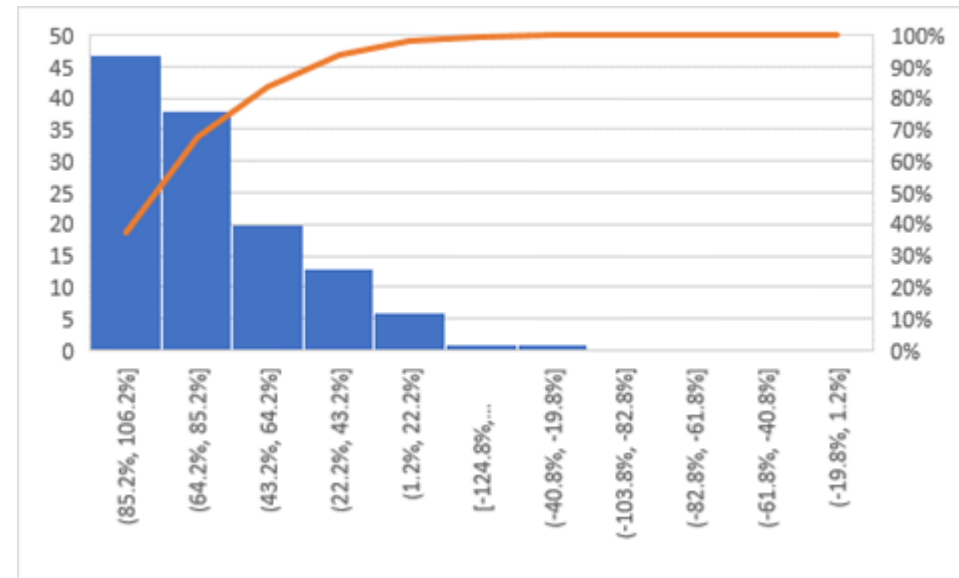
The point of this is to build trust in the data with the current decision makers



Recommend

- Score and Rank by most valuable first
- [Hypothesis testing](#)
- Publish your test stats!
 - F1, MASE, AUC, ROC, Conf Intervals, etc.
- Connect to Action list
 - Stats/ML
- Automate notifs to users based on best action to take

Pareto Chart



The point of this is to build trust in using the data to make decisions

Act

- When Decision makers show that they are just doing what is recommended, then it is time to productize (welcome to Dev)
- Considerations:
 - BCDR
 - Re-tuning
 - Monitoring
 - Fallbacks (Defaults system should use when ML is down)
 - Data/Usage Scale

Examples

How data are solving Quality problems

Bootstrap

- Start with QoS Concerns:
 - Reliability: measures failure rates
 - Engagement: measures usage
 - Availability: % Time spent failing
 - Latency: Time spent succeeding
- Pre-prod
 - automation drives load only
 - Validation done through data
- Production
 - repurpose your validation code as monitors
- Bridge to QoE:
 - Does [metric] impact revenue or growth?

CustomerImpactView

R.E.A.L. Drilldown Report - Last 7 Days

ItemPath: [R](#) / [Azure Compute Insights/Incident Situation/CustomerImpactView](#)

Time Histograms (by Log10 of Milliseconds)



Real World Advanced Examples:

ML has been successfully used to:

- Dynamically determine signal/noise thresholds for alerts
- Automate Deploy to Prod/rollback decisions
- Detect Service Outages
- Detect Memory Leaks
- Determine Customer Pain sequences
- Optimize Timeouts
- Classify likely lines of code causing “the bug”
- Refine Black box Test Suites – reduce, add test cases
- Train Test Oracles

Closing

Final Guidance

Target Actionability over Accuracy

- Focus on the goal, not the model
- Your insights must be interesting, valuable, and within the means of the business
- If recommendations would not change with increased accuracy, defer improving accuracy

Credibility is key

- Objective Truth is *everything*
- To the outside observer, Data Science is very similar to magic. Speak “human”.

Relationships are critical – human and data

- Successful ML is not offline. Part of the product.
- Cognitive limitation – you can not know all Domain knowledge.

Ethics

Do it right – Huge Impact and Fun

My Data Science team in Azure Compute directly contributed to nearly 2 orders of magnitude improvement in Availability (the One Metric that Matters!)

