Create Your Roadmap to Data-Driven Quality

Seth Eliot
Principal Knowledge Engineer, Test Excellence

Microsoft

April 15, 2014
About Seth

Digital Media Services

A/B Testing of Services

Petabytes Processed

Services and Cloud and Data-Driven Engineering and...

Microsoft
In This Session, You’ll Learn...

...how to create your “roadmap” to DDQ

Roadmap + Application (service or product) + Environment Engineering Processes = At destination

Roadmap + You and your... = DDQ Strategy
Data-Driven Quality

What is it? Why is it important?
Data, where have we been?

The HiPPO

Engineering Data
Test pass/fail results
Bug counts
Delivery cadence
Code coverage
Code Churn

**Highest Paid Person’s Opinion**
Data, where are we going?

Engineering data
Test results → Scoring engines using Bayesian analysis

Production-quality data
You may have heard of TiP
Not as difficult as you might think
Lots of solutions for lots of application types
“If you have to kiss a lot of frogs to find a prince, find more frogs and kiss them faster and faster” - Mike Moran
Virtuous cycle of DDQ

Deploy → Detect → Fix → Deploy → Detect
The roadmap

Determine your questions
Design for production-quality data
Select your data sources
Use the right data tools
Get answers to your questions
Learn new questions
Repeat
“Big Data” is hot, let’s start with that

Start at the beginning

GQM - 1994
Determine your questions

What do you need to know about quality?
Why not just get data and look for answers?

Does sunscreen increase chance of drowning?

Do night-lights cause near-sightedness in children?
What questions does EXO ask?

“Downtime” is defined as any period of time when users are unable to send or receive email via all supported mailbox access.

<table>
<thead>
<tr>
<th>Monthly Uptime Percentage</th>
<th>Service Credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 99.9%</td>
<td>25%</td>
</tr>
<tr>
<td>&lt; 99%</td>
<td>50%</td>
</tr>
<tr>
<td>&lt; 95%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Is it available?

Often a Pri 0

Is the application (product, service) there for the user?

“Dialtone”

Critical work stoppage for user
How do users perceive availability?

Application works, but feature does not Occasionally does not render/load properly
How is performance?

Huge Impact

Power of Production Data

- Real users
- Multiple environments
- End to end
- Scale & geo-diversity
What do users do?

Customer Experience Improvement Program (CEIP)

Do you want to participate in the Windows Customer Experience Improvement Program?

The program helps Microsoft improve Windows. Without interrupting you, it collects information about your computer hardware and how you use Windows. The program also periodically downloads a file to collect information about problems you might have with Windows. The information collected is not used to identify or contact you.

Read the privacy statement online

- Yes, I want to participate in the program.
- No, I don’t want to participate in the program.

This setting is managed by your system administrator. Why can’t I change some settings?
Your turn - Determine your questions

What type of answer are you looking for?
Availability, performance, usage?
Prioritize

When does availability NOT come first?
Your turn - Determine your questions

What are the key scenarios for the type you selected?

What is high priority?

SharePoint
Design for production-quality data

Get data from or near production
Two types of data to acquire

Active: synthetic
Passive: real (RUM)

Active data in prod
For services only?
Client: is the service there?
Staged data acquisition mitigates risk

**Service**

- **Stage 1:** In prod, no users
- **Stage 2:** Dogfood
- **Stage 3:** Some servers in prod
- **Stage 4:** Some more servers in prod
- **Stage 5:** World-wide prod

**Deployment Validation**

**Service Validation**

**Scale Validation**

**Real-time service quality**

**Product (client, on-prem server)**

- **Stage 1:** Partial or whole product team
- **Stage 2:** Dogfood
- **Stage 3:** Technology Adoption Programs (TAP)
- **Stage 4:** Some clients in production
- **Stage 5:** All Customers
Staged data acquisition - Netflix

1B API requests per day

Canary Deployment
Staged Data Acquisition - Facebook

- Dogfood
  - In prod, no users (except internal ones)
  - Some servers in Production
  - World-wide deployment
  - Feature flags

- code development and testing by engineer
- review by peer engineer using phabricator
- checkin and automated regression testing
- "latest" (internal use)
- release and perflab testing
- Sunday afternoon
- Monday
- H1 deploy (internal)
- H2 deploy (1% users)
- H3 deploy (full)
- Tuesday afternoon (contributing engineers on call)
- Gatekeeper switch on (full or selected subset of users)
Speed of deployment

Deploy

Usually easy for services
Client apps may have a deployment capability
Or may make use of feature flags

EaaSy - Everything is now connected and thus updatable
Data acquisition for clients

Filtering and aggregation at client

Be kind to the client

Don’t abuse user resources:
  - Bandwidth
  - Battery
  - Disk
Bucket-ized data is aggregated and uses less bandwidth.
Data acquisition for clients

Filtering and aggregation at client

Be kind to the client

Pipeline to collect and process data

Make it easy

Staged Data Acquisition
Your turn - Design for production-quality data

What might be your stages for risk mitigated data acquisition?

Role of active and passive monitoring?

How can you engineer for EaaSy deployment?
Select your data sources

Determine the data necessary to answer your questions
Application data

Failure Rate

Average Time
Hang and crash data

Specialized application data
Most frequently encountered conditions
Bucket the Big Data and find the offenders
Get offending code and function calls (stack)

But it worked in Dogfood
Culprit was a old version add-in
Harden against this
Usage data

Client-side instrumentation
Proprietary
Javascript: clicks, hovers [web apps]
Get 1x1 GIF: Page Views [web apps]

Combine into more complex scenarios
How did user get to shopping cart checkout?
Feedback data
Feedback data also includes...

Customer Support Data

Social Media Mining

<table>
<thead>
<tr>
<th>Service</th>
<th>Severity</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lync Online</td>
<td>Sev1</td>
<td>Service Degradation</td>
</tr>
<tr>
<td>Exchange Online</td>
<td>Sev2</td>
<td>Investigating</td>
</tr>
</tbody>
</table>

Top Communities
- Gaming Software Devs Bloggers

Top Locations
- USA, Japan, Australia, Puerto Rico, Malaysia, Singapore

Sentiment
- 7%
- 1%

Gender
- 62%
- 38%
Active monitoring data

Data Center X

AM1

Service

AM2

Dependency Service(s)

Data Center Y

AM3

Client
Ease of detection

Again services have it easy
Many clients are always connected
On prem servers (enterprise) require partnerships

EaaSy – Rich near real-time telemetry
Data handling - privacy

Transparency and Control

Collection and Retention

**Depends on Type**
- anonymous Data
- pseudonymous
- personally identifiable info (PII)
- sensitive PII

**Depends on Purpose**
- provide the service
- improve the current service
- improve a future version service
- improve non-associated services
- content personalization
- ad targeting
Data handling – non-service products

Client and on-prem server considerations

User owned resources:
- bandwidth
- battery
- disk, cpu, etc...

Correlations
- end-to-end across clients and services
  by user, by session
Your turn - Select your data sources

- Infrastructure data
- Application data
- Hang and crash data
- Usage data
- Feedback data
- Active monitoring data
- Design to handle your data
Use the right data tools

A 50,000 foot view
## Data storage and processing systems

<table>
<thead>
<tr>
<th>Database</th>
<th>Table storage (SQL)</th>
<th>Optimized for CRUD – (create, read, update, delete) of single records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Warehouse</td>
<td>Table storage (SQL)</td>
<td>Table structure optimized for queries and bulk insert</td>
</tr>
<tr>
<td>OLAP Cubes (Online Analytical Processing)</td>
<td>Multidimensional Cube (MDX)</td>
<td>Aggregations of measures for multiple dimensions</td>
</tr>
<tr>
<td>Hadoop / Map-Reduce</td>
<td>Distributed File System (HDFS)</td>
<td>Big Data TB-PB-EB Unstructured Data</td>
</tr>
</tbody>
</table>
Hadoop in 60 Seconds

Hadoop

HDFS – Hadoop Distributed File System

Map-Reduce

ADD
FCA
ADD
ADD
ADD
FCA
ADD
ECC
ECC
ECC
ECC
1xA
1xC
2xD
1xF
1xEA
3xC
2xD
1xE
1xF
A common data flow

Telemetry System → SQL DW → Cube

Real-time Monitoring → Other Visualizations

Programmatic access

Power BI in Excel

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Weight</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dave</td>
<td>43</td>
<td>136</td>
<td>168</td>
</tr>
<tr>
<td>Pat</td>
<td>25</td>
<td>149</td>
<td>186</td>
</tr>
<tr>
<td>Fred</td>
<td>32</td>
<td>170</td>
<td>150</td>
</tr>
<tr>
<td>Connie</td>
<td>37</td>
<td>177</td>
<td>151</td>
</tr>
<tr>
<td>Feng</td>
<td>37</td>
<td>171</td>
<td>170</td>
</tr>
<tr>
<td>Monica</td>
<td>30</td>
<td>145</td>
<td>184</td>
</tr>
<tr>
<td>Declan</td>
<td>31</td>
<td>164</td>
<td>140</td>
</tr>
<tr>
<td>Brendan</td>
<td>33</td>
<td>169</td>
<td>179</td>
</tr>
</tbody>
</table>
Web Apps integrated with OneDrive, FB, web mail, etc.

Real-time monitoring
Feedback

Top Linked Sites:
- slickdeals.net: 2,243
- youtube.com: 1,254
- detik.com: 1,217
- twitter.com: 895
- instagram.com: 644

Top Hashtags:
- #kinect: 1,437
- #giveaway: 762
- #xbox: 515
- #xbox360: 376
- #microsoft: 376

Sentiment:
- Neutral: 31,076
- Positive: 8,338
- Negative: 37

Top Clusters:
- Kinect Fusion: 446
- Microsoft Research: 402
- Microsoft Kinect: 345
- Fusion Estara: 306
- Fusion Estara En La Proxima: 305

Top Regions:
- United States: 1,527
- LATAM: 1,324
- APAC: 827
- UK: 610
- Western Europe: 451

Top Languages:
- English: 30,833
- Spanish: 5,282
- Japanese: 4,359
- Portuguese: 74
- French: 11
Your turn - Use the right data processing tools

- DB, DW, Cube, Big-Data platforms
- Put it all together

Do you need real-time monitoring?

...sentiment analysis?
Get answers to your questions
and learn new questions
Outlook.com prioritizes performance

View Inbox – Page Load Time (PLT) by Browser
As experienced by actual users

JSI
JavaScript Instrumentation

500 Million measurements per month
Prioritizes availability

Predict 75% of dips 24 hours ahead of time
Netflix prioritizes perceived availability & performance.
Yammer prioritizes usage

What happens to new user retention when you shorten the signup flow?

It goes down! Don't ship that feature.
Find new questions and repeat

Xbox recommends
Try algorithm
Collect data
Adjust algorithm
Collect data
Repeat
Find new questions and repeat

Microsoft Visual Studio
Preparing Solution...

In Visual Studio 2012 we added asynchronous loading for solutions

We also added Telemetry
We can see how it works

Loads w/ 2+ projects

- 62%
- 34%
- 15.1%
- 6%
- 2.5%
- 0.30%

25th Percentile
50th Percentile
75th Percentile
% Loaded Async
Modal Load Time (sec)
Your turn –
Get answers; find more questions

Enjoy the rest of STPCon :-)  
Get back to work  
Implement your roadmap  
Get answers to your quality assessment questions
Summary

Determine your questions
Design for production-quality data
Select your data sources
Use the right data tools
Get answers to your questions
Learn new questions
Repeat

The most recent version of this deck can be found at http://setheliot.com
Special thanks to these folks

Ravi Vedula  David Brooks  Monica Catunda
Andrea Jesse  Mike Tholfsen  Lynette Skinner
Bill Hodghead  Jodie Draper  Joe Schumachner
Danny Thayer  Brian Mueller  Donny Luu
Joseph Sefair  Tara Roth  John Hoegger
Kitty Thomas  Dror Cohen  Alain Anyouzoa
Amanda Reinke  Nathan Halstead
Heather Lader  Lori Oviatt

Any questions?
Please fill out the survey

Create Your Roadmap to Data-Driven Quality
Session 101

Seth Eliot