Gameplay Data Analysis

Asking the Right Questions
Ian Thomas

- Database/ Web Development
  - Epic Games, Inc
Intended Audience

• Server and Database Programmers
• Analysts
• Gameplay Programmers
• Producers
• Designers
Too Much or Too Little?

• Collect Too Much Data
  – Expensive
  – Can’t Analyze

• Collect Too Little Data
  – Cheap
  – Can’t Analyze
Collecting the Right Data

• Gears 3 Data Pipeline
  – Flexible
• (Gears of ) War Story
• How to Determine Which Data to Collect
  – And How To Store It
GEARS 3 DATA PIPELINE
Gears of War 3

• 18 M. Franchise
• 5.5 M. Gears of War 3
Peak Data Sizes

- **Storage**
  - 65 GB / Day
  - 21 Million Rows / Day

- **Traffic**
  - 180 GB / Day
  - 25 Mbps
Flexible Storage Pipeline
Network Infrastructure
Network Infrastructure
Network Infrastructure
Network Infrastructure
Network Infrastructure

Unreal MCP
- Database Mirror (Epic HQ)
- Database Servers
- Data Collectors
- Analysis Databases
- Post Processors
- Analysis

The World
- Teh Intertubes
- Xbox 360 Clients
Flexible

• Mitigate Costs
• Able to Send Anything
• Can Add or Remove Data Points
(GEARS OF ) WAR STORY
Priorities for Beta

1. Matchmaking
2. Weapon Balance
3. Achievement Progression
4. Anything that might come up
The One Thing we Weren’t Concerned with Was ...

DEDICATED SERVER MONITORING
Are the Dedicated Servers Up?

- Forums Say They’re Down
- IT Says They’re Up
- We Play Some Games
- ... And They’re Down
- WTF !?!
Healthy Servers; Unhealthy Game

• Dedicated Servers ARE Up
• Logs Look Healthy
• Users are Playing
  – On Listen Servers
Are the Dedicated Servers Up?
AHA!

• Matchmaking Results Can Be Used to Infer Game Health
Matchmaking Process

1. Request Datacenter
2. Read Skill Data
3. Search for Matches
4. Rate Matches
5. Join Match
Analysis -> Monitoring

• 25 – 2,500 Results per Matchmaking Request
• Full Dataset: 1 Day Turnaround
• Small Sample: 15 Minutes
Ranked Join Action Results Trend

Avg Ranked Join Action Results

- Failed Client: 0.14%
- Failed Full: 0.16%
- Failed Timeout: 68.85%
- Succeeded: 24.95%
Ok, so what’s wrong?
Comparing Apples to Oranges

- Different Granularities
- No Cohesive Picture
- Only Experts Could Divine Answers
- We Needed ...
The One True Chart

• One View That Showed Everyone What Problems Servers Were Having
Datacenter Status
Lessons

- Priorities Change
- Schema Affects Analysis
- Iteration is Key
- Knowing What You Want Allows You to Optimize
How do you only collect the data that you’re going to use?

ASK THE RIGHT QUESTIONS
Everybody Wants Something
PHASE 1  PHASE 2  PHASE 3
Collect
underpants
DATA
Profit
Collect Everything; Answer Anything

• Holy Grail
  – Don’t Know What You Don’t Know
  – Can Only Ship Once

• Easier Than Trying to **Answer** Everything
Answer Anything You Want

- Client Generated
- Raw
- Compressed
- Post Processed
- Analyzed

GB / Day

1  10  100  1,000
Selecting Questions to Answer

- Define Priorities
- Categorize
- Structure Data Accordingly
- Analyze
- Iterate
What are the most important things you’re expecting to get out of your data?

DEFINE PRIORITIES
Cost / Benefit
You Want the **Hot** Information Without Spending **Crazy** Amounts of Effort to Get It.
Who Needs to Be Involved?

- Gameplay Programmers
- Designers
- Producers
- Operations
  - Dedicated servers
- External partners
Actionable Data

• What can you Change?
• When can you **Use** this Information?
  – Config files
  – Beta
  – Title Update
  – Next Game
  – Press Release
  – Forums
What Will Knowing Get You?

Multiplayer Character Preference:

<table>
<thead>
<tr>
<th>Cog MP Character</th>
<th>Anya</th>
<th>Baird</th>
<th>Cole</th>
<th>ColeThrasball</th>
<th>Dom</th>
<th>Marcus</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20.18%</td>
<td>7.98%</td>
<td>4.78%</td>
<td>18.94%</td>
<td>13.78%</td>
<td>34.34%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Locust MP Character</th>
<th>Drone</th>
<th>Flamer</th>
<th>Kantus</th>
<th>SavageHunter</th>
<th>SavageTheron</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>33.35%</td>
<td>8.06%</td>
<td>14.40%</td>
<td>25.34%</td>
<td>18.85%</td>
<td>33.35%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Default MP Weapon</th>
<th>Hammerburst</th>
<th>Lancer</th>
<th>RetroLancer</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>41.58%</td>
<td>43.56%</td>
<td>70.69%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Default Secondary</th>
<th>Gnasher</th>
<th>SawedOffShotg</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>29.31%</td>
<td>70.69%</td>
<td>70.69%</td>
</tr>
</tbody>
</table>
What Will Knowing Get You?

Active Reload Success Rates
What Will Knowing Get You?

Downing Distances
What Will Knowing Get You?

Avid Fans
What Will NOT Knowing COST You?
What Will **NOT** Knowing **COST** You?

Dedicated Server Outages
What Will NOT Knowing COST You?

Nerd Rage
Gears 3 Priorities

1. Matchmaking Analysis
   - Minimize User Wait Times

2. Actionable Gameplay Measurements
   - Weapon Balance
   - Achievement and Level Progress
   - Game Types
Gears 3 Priorities

1. Dedicated Server Health Monitoring
2. Matchmaking Analysis
   – Minimize User Wait Times
3. Actionable Gameplay Measurements
   – Weapon Balance
   – Achievement and Level Progress
   – Game Types
CATEGORIZE QUESTIONS
Brainstorm

- Work With Everyone
- Gather Large List of Potential Questions
- Filter These Through Priorities
Categories Will Emerge

• Game Info
  – Games by Hour of the day, Games by Game Type, Average Match Length

• Weapons
  – Kills by Weapon, DBNO by Weapon, Starting Weapon by Game Type

• User Settings
  – Resolutions, Control Scheme, Gore, Subtitles
Now is the Time to Think About How to

STRUCTURE DATA
Storage Basics

You Try Counting All Those Kills
# Events

<table>
<thead>
<tr>
<th>Event</th>
<th>Damage Type</th>
<th>Source Player ID</th>
<th>Target Player</th>
<th>MapName</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kill</td>
<td>Lancer</td>
<td>P1</td>
<td>P2</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Death</td>
<td>Lancer</td>
<td>P2</td>
<td>P1</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Kill</td>
<td>Gnasher</td>
<td>P1</td>
<td>P2</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Death</td>
<td>Gnasher</td>
<td>P2</td>
<td>P1</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Kill</td>
<td>Sawed-Off</td>
<td>P1</td>
<td>P2</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Death</td>
<td>Sawed-Off</td>
<td>P2</td>
<td>P1</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Kill</td>
<td>Gnasher</td>
<td>P1</td>
<td>P2</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Death</td>
<td>Gnasher</td>
<td>P2</td>
<td>P1</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Kill</td>
<td>Gnasher</td>
<td>P1</td>
<td>P2</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Death</td>
<td>Gnasher</td>
<td>P2</td>
<td>P1</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Kill</td>
<td>Gnasher</td>
<td>P1</td>
<td>P2</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Death</td>
<td>Gnasher</td>
<td>P2</td>
<td>P1</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Death</td>
<td>Gnasher</td>
<td>P2</td>
<td>P1</td>
<td>MP_Hotel</td>
</tr>
</tbody>
</table>
# Counting Kills

<table>
<thead>
<tr>
<th>Event</th>
<th>Damage Type</th>
<th>Source Player ID</th>
<th>Target Player</th>
<th>MapName</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kill</td>
<td>Lancer</td>
<td>P1</td>
<td>P2</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Death</td>
<td>Lancer</td>
<td>P2</td>
<td>P1</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Kill</td>
<td>Gnasher</td>
<td>P1</td>
<td>P2</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Death</td>
<td>Gnasher</td>
<td>P2</td>
<td>P1</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Kill</td>
<td>Sawed-Off</td>
<td>P1</td>
<td>P2</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Death</td>
<td>Sawed-Off</td>
<td>P2</td>
<td>P1</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Kill</td>
<td>Gnasher</td>
<td>P1</td>
<td>P2</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Death</td>
<td>Gnasher</td>
<td>P2</td>
<td>P1</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Kill</td>
<td>Gnasher</td>
<td>P1</td>
<td>P2</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Death</td>
<td>Gnasher</td>
<td>P2</td>
<td>P1</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Kill</td>
<td>Gnasher</td>
<td>P1</td>
<td>P2</td>
<td>MP_Hotel</td>
</tr>
<tr>
<td>Death</td>
<td>Gnasher</td>
<td>P2</td>
<td>P1</td>
<td>MP_Hotel</td>
</tr>
</tbody>
</table>

**Beta Stats**

- 23 billion bullets fired
- 4.9 billion ribbons
- 927 million kills
- 291 million medals earned
- 131 million executions performed
- 23 million chainsaw deaths
- 249 years worth of matches played
- 145 different countries
Event Driven Heat Maps
Events

• Granular
  – 1,000 rows/game
  – 7,500,000 sessions/Day
  – 7,500,000,000 rows/Day!
    • If we were to all expand it all

• Expensive to Store and Analyze
  – 36 GB/Day Raw Compressed Data
    • Just for Game Events
<table>
<thead>
<tr>
<th>DateReceivedUT</th>
<th>MapName</th>
<th>WinningTeam</th>
<th>Kills</th>
<th>Deaths</th>
<th>DBNOs</th>
<th>ShotsFired</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/21/11 5:00 PM</td>
<td>MP_Sandbar</td>
<td>LOCUST</td>
<td>2988</td>
<td>1671</td>
<td>1402</td>
<td>83980</td>
</tr>
<tr>
<td>9/21/11 5:00 PM</td>
<td>MP_Hotel</td>
<td>COG</td>
<td>2864</td>
<td>1717</td>
<td>1477</td>
<td>83209</td>
</tr>
<tr>
<td>9/21/11 5:00 PM</td>
<td>MP_Hotel</td>
<td>LOCUST</td>
<td>3344</td>
<td>2046</td>
<td>1624</td>
<td>88872</td>
</tr>
<tr>
<td>9/21/11 5:00 PM</td>
<td>MP_Sandbar</td>
<td>COG</td>
<td>3872</td>
<td>2140</td>
<td>1950</td>
<td>108691</td>
</tr>
</tbody>
</table>
Counting Kills

<table>
<thead>
<tr>
<th>DateReceivedUT</th>
<th>MapName</th>
<th>WinningTeam</th>
<th>Kills</th>
<th>Deaths</th>
<th>DBNOs</th>
<th>ShotsFired</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/21/11 5:00 PM</td>
<td>MP_Sandbar</td>
<td>LOCUST</td>
<td>2988</td>
<td>1671</td>
<td>1402</td>
<td>83980</td>
</tr>
<tr>
<td>9/21/11 5:00 PM</td>
<td>MP_Hotel</td>
<td>COG</td>
<td>2864</td>
<td>1717</td>
<td>1477</td>
<td>83209</td>
</tr>
<tr>
<td>9/21/11 5:00 PM</td>
<td>MP_Hotel</td>
<td>LOCUST</td>
<td>3344</td>
<td>2046</td>
<td>1624</td>
<td>88872</td>
</tr>
<tr>
<td>9/21/11 5:00 PM</td>
<td>MP_Sandbar</td>
<td>COG</td>
<td>3872</td>
<td>2140</td>
<td>1950</td>
<td>108691</td>
</tr>
</tbody>
</table>

**BETA STATS**
- 23 BILLION BULLETS FIRED
- 4.9 BILLION RIBBONS
- 927 MILLION KILLS
- 291 MILLION MEDALS EARNED
- 131 MILLION EXECUTIONS PERFORMED
- 23 MILLION CHAINSAW DEATHS
- 249 YEARS WORTH OF MATCHES PLAYED
- 145 DIFFERENT COUNTRIES
Roll-ups

• Less granular
  – 1000 events -> 30 events

• Less Space
  – 7.5 Billion -> 480 rows/day
  – 2GB /Day Session + Rollup data

• Quick Reports
Things to Consider

- Timeliness
- Sample Size
- Outliers
- Discovery
- Dimensionality
Decision Driven

• Your Job is to Help People Make Decisions
• Worry About Accuracy
  – So Decision Makers Don’t Have To
Remember your Viewers

• People Tend to Confirm Beliefs

• Data Can be Difficult to Analyze
  – Nuanced Data Should be Simplified
Every Question Leads to Another

ITERATE
Iteration Data

- Generated
- Game QA
- Playtest
- Beta
- Live
Constantly

- Evaluate Priorities
- Solidify Categories
- Tweak Storage Layout
- Improve Charts
Coming up with questions without real data is HARD

Answering unexpected questions on Launch day is HARDER
Conclusion

The More You Know About What Knowledge You Want From Your Data
The More Efficient Your Collection and Analyses Will Be.
Gameplay Data Analysis
Collecting the Right Data

Special thanks to:
Wes Hunt
Josh Markiewicz
Joe Graf
Sam Zamani
Paul Oliver
Dana Cowley

Ian Thomas
Epic Games, Inc.