

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

The background features a large, faint, blue watermark of the Epic Games logo, which consists of the word "EPIC" in a stylized font above the word "GAMES" in a smaller font, all enclosed within a shield-like border. The background is a dark blue gradient with some abstract circular patterns.

GEARS 3 DATA PIPELINE

Gears of War 3

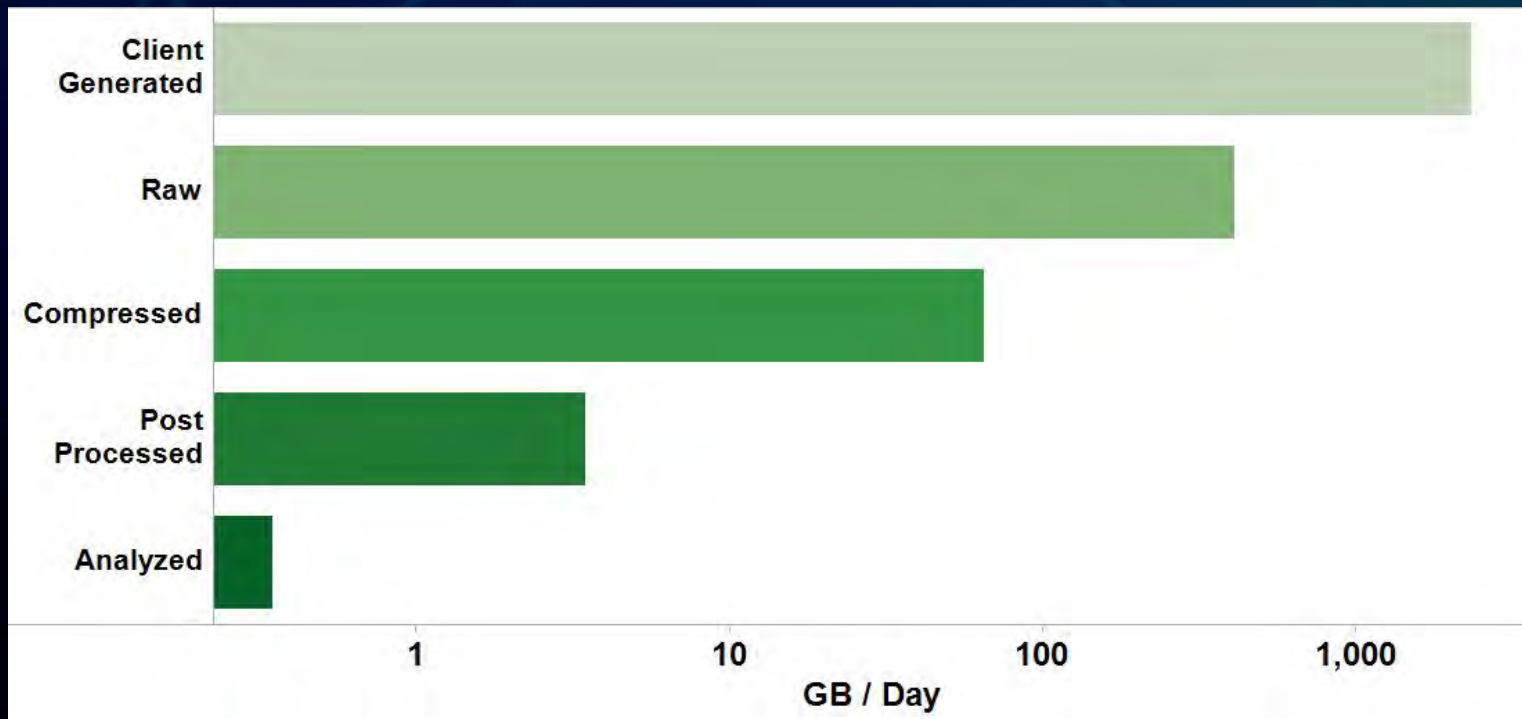
- 18 M. Franchise
- 5.5 M. Gears of War 3



Peak Data Sizes

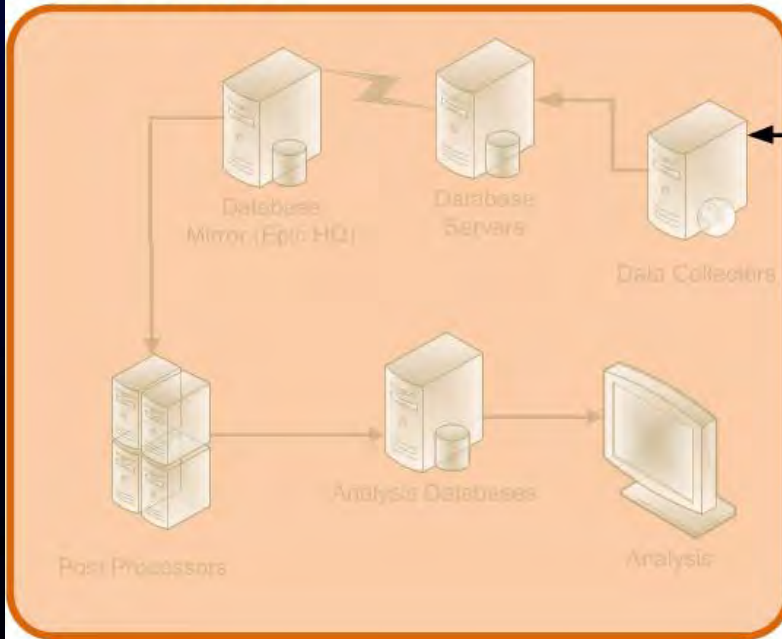
- Storage
 - 65 GB / Day
 - 21 Million Rows / Day
- Traffic
 - 180 GB / Day
 - 25 Mbps

Flexible Storage Pipeline



Network Infrastructure

Unreal MCP

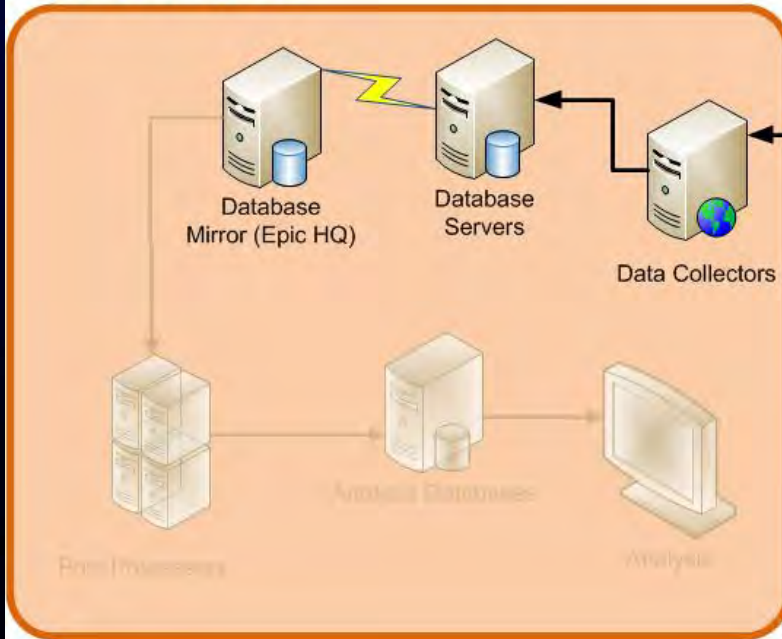


The World

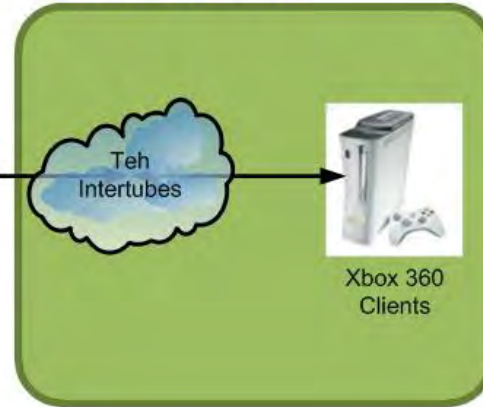


Network Infrastructure

Unreal MCP

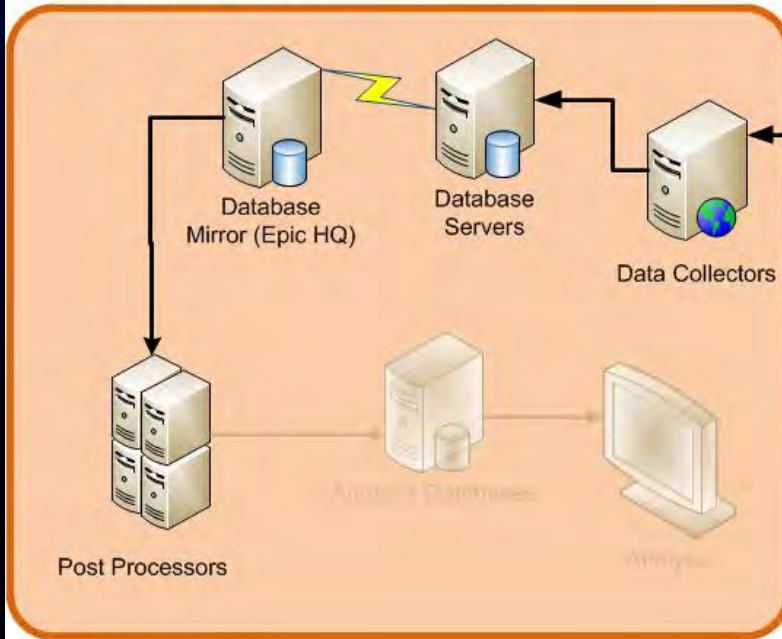


The World



Network Infrastructure

Unreal MCP

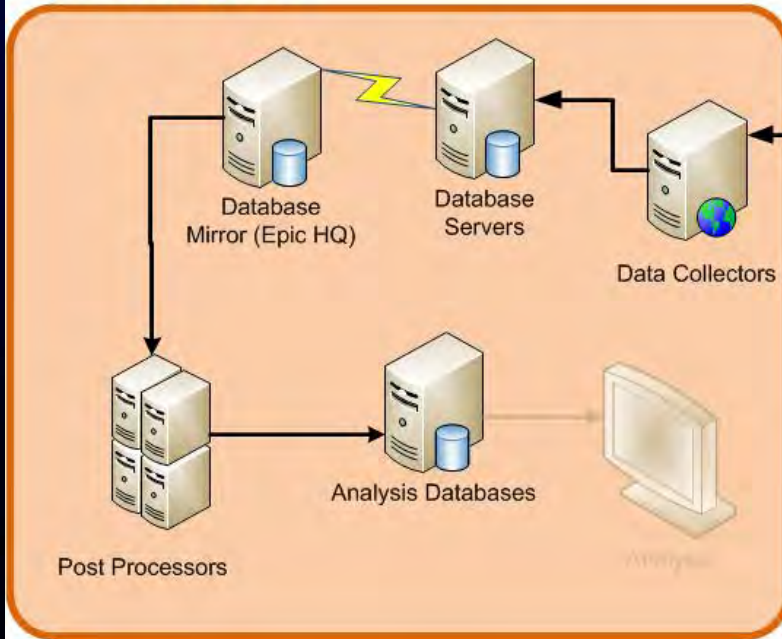


The World



Network Infrastructure

Unreal MCP

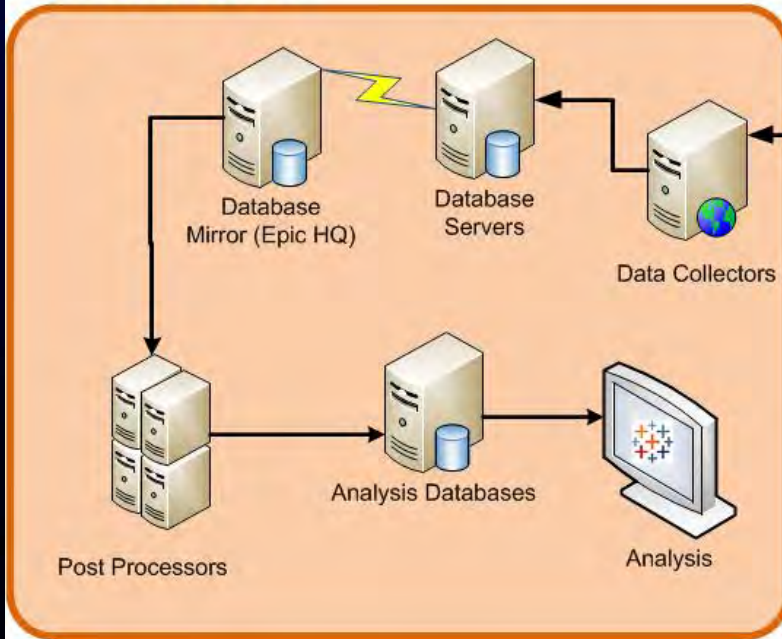


The World



Network Infrastructure

Unreal MCP



The World



Flexible

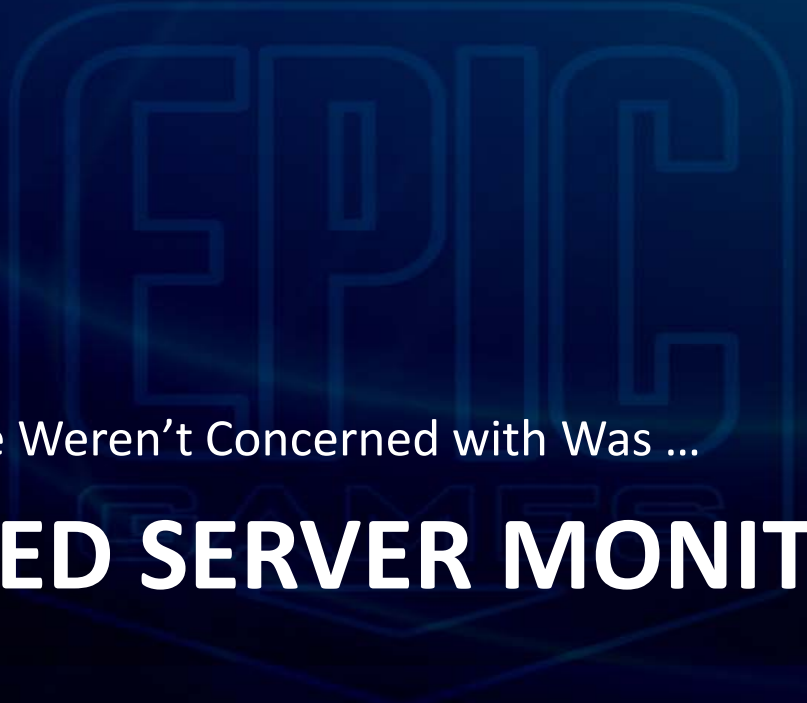
- Mitigate Costs
- Able to Send Anything
- Can Add or Remove Data Points

A large, faint, blue watermark of the Epic Games logo is centered in the background. The word "EPIC" is in a large, stylized, outlined font, and "GAMES" is in a smaller, similar font below it. The background is a dark blue gradient with a large, curved, lighter blue shape on the right side.

(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?

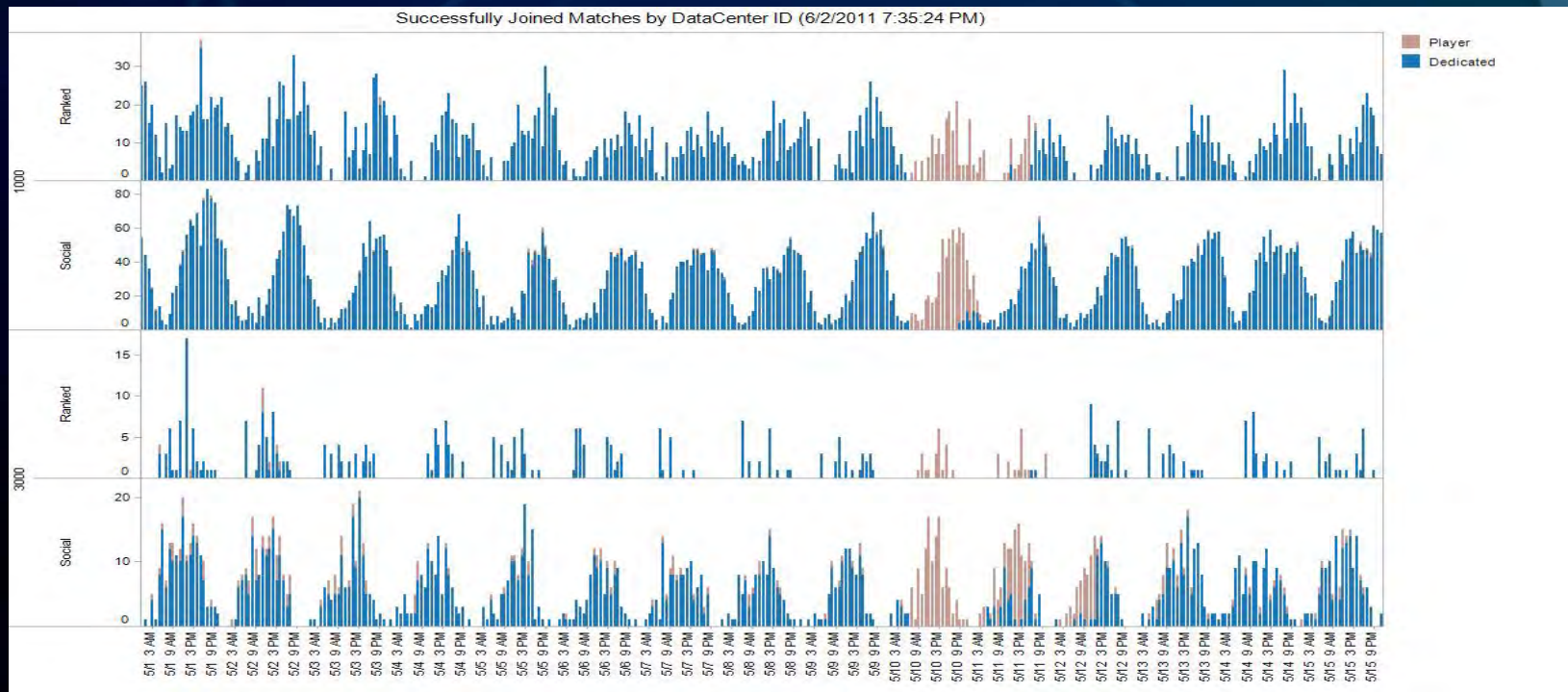


AHA!

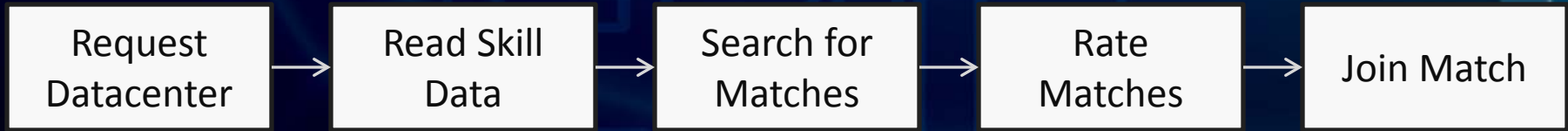
- Matchmaking Results Can Be Used to Infer Game Health



Inferred Datacenter Health



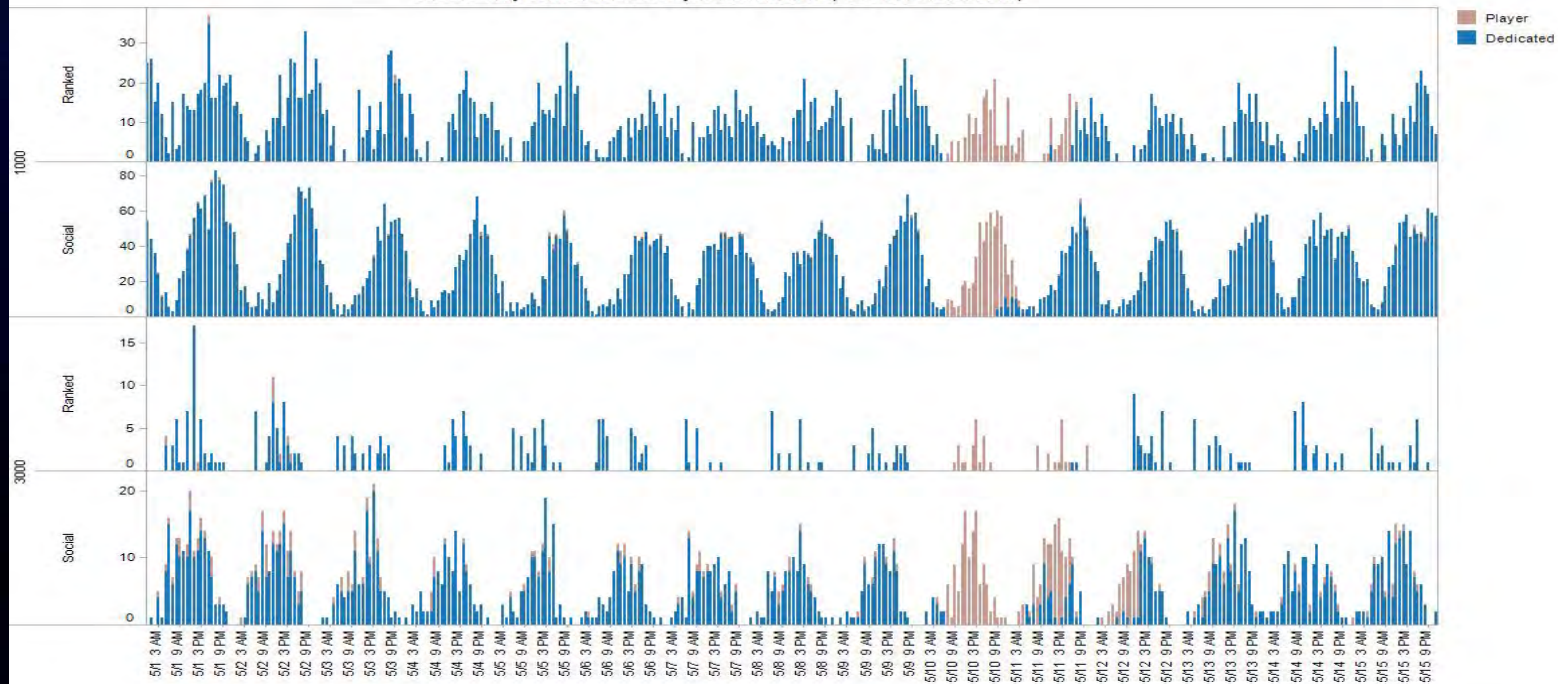
Matchmaking Process



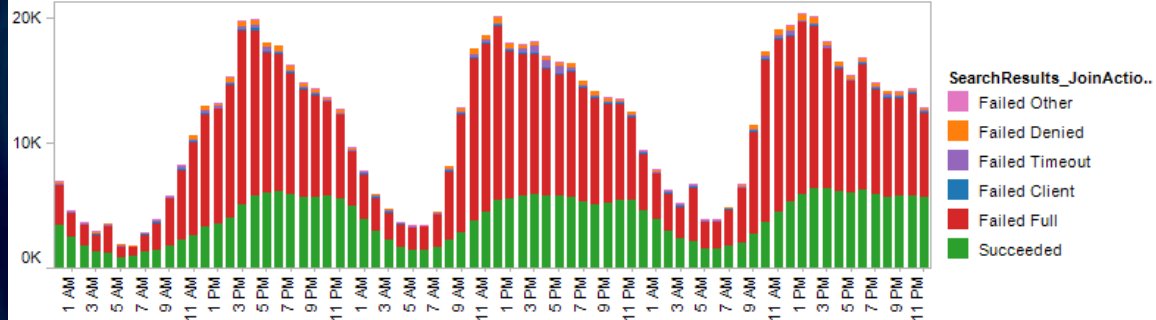
Analysis -> Monitoring

- 25 – 2,500 Results per Matchmaking Request
- Full Dataset: 1 Day Turnaround
- Small Sample: 15 Minutes

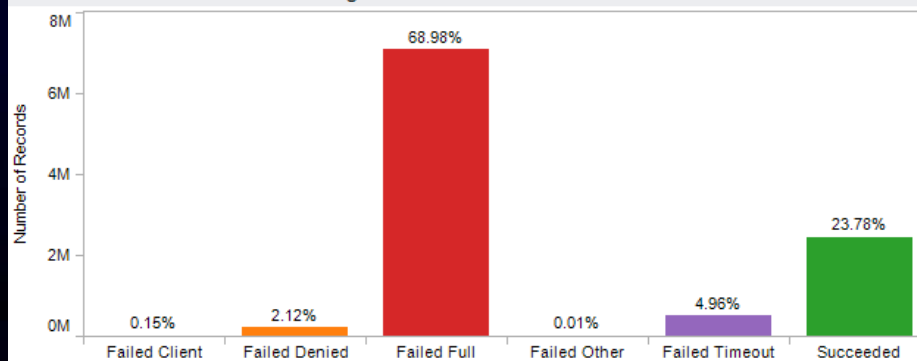
Successfully Joined Matches by DataCenter ID (6/2/2011 7:35:24 PM)



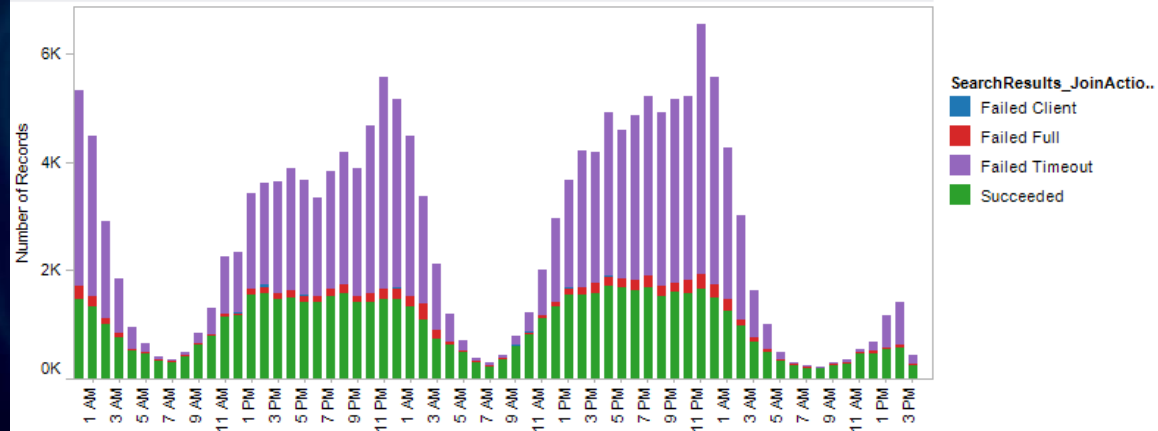
Social Join Action Results Trend



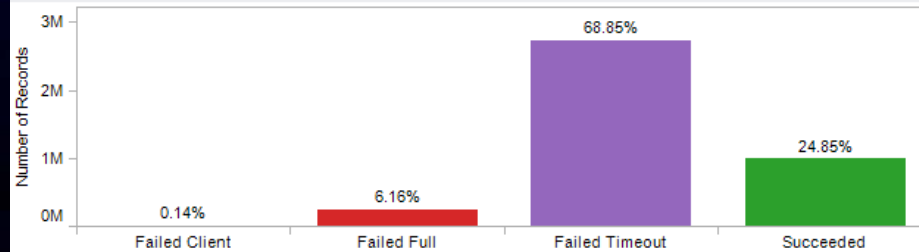
Avg Social Join Action Results



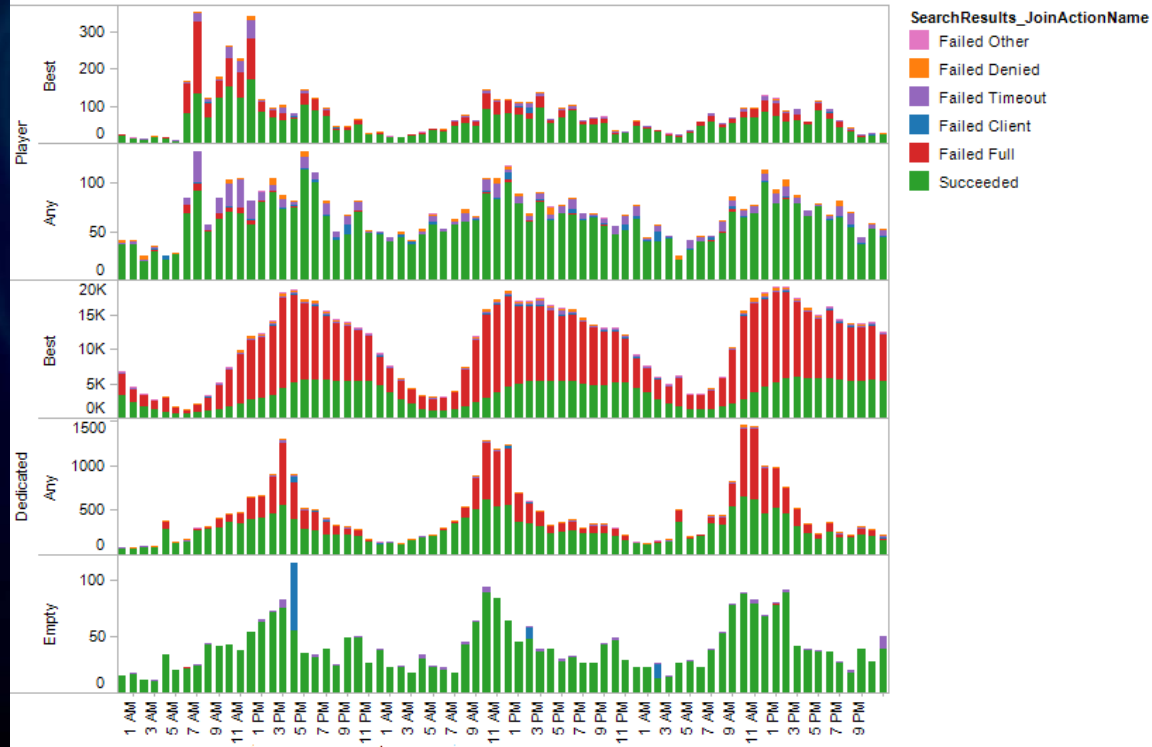
Ranked Join Action Results Trend

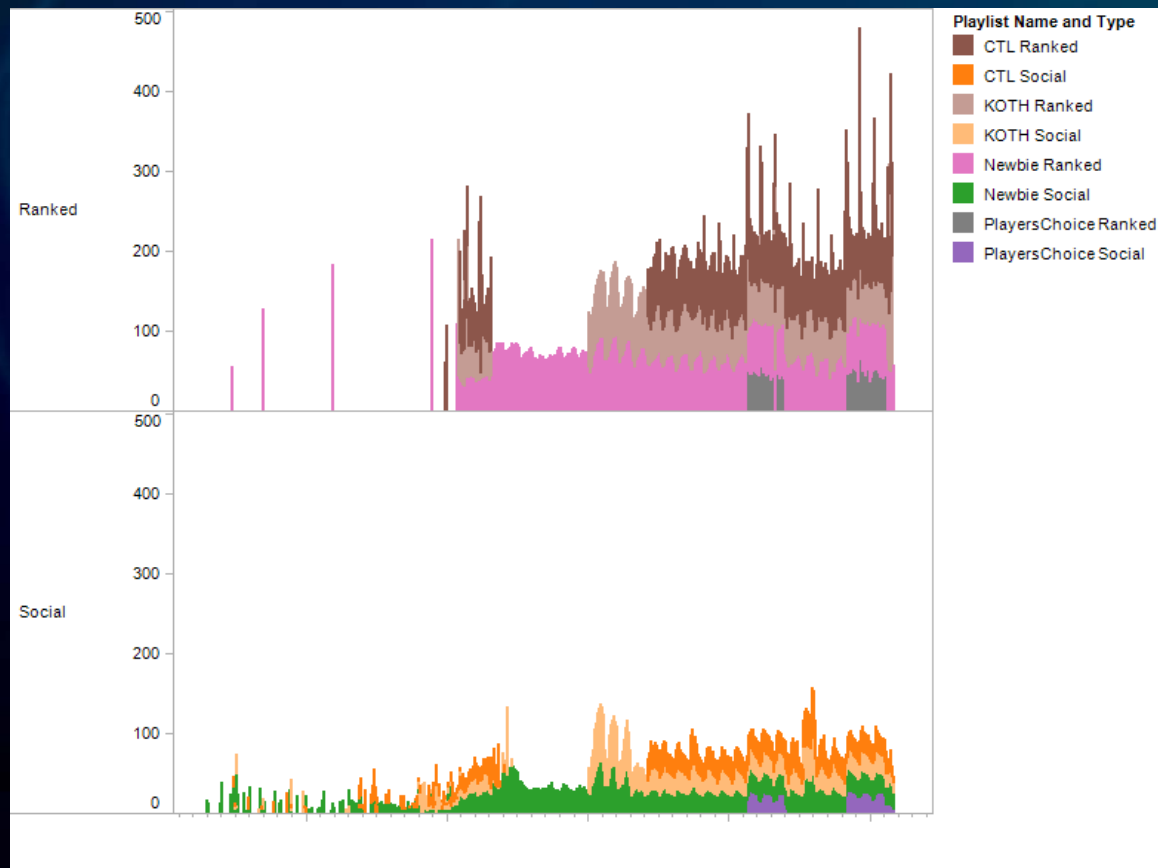


Avg Ranked Join Action Results

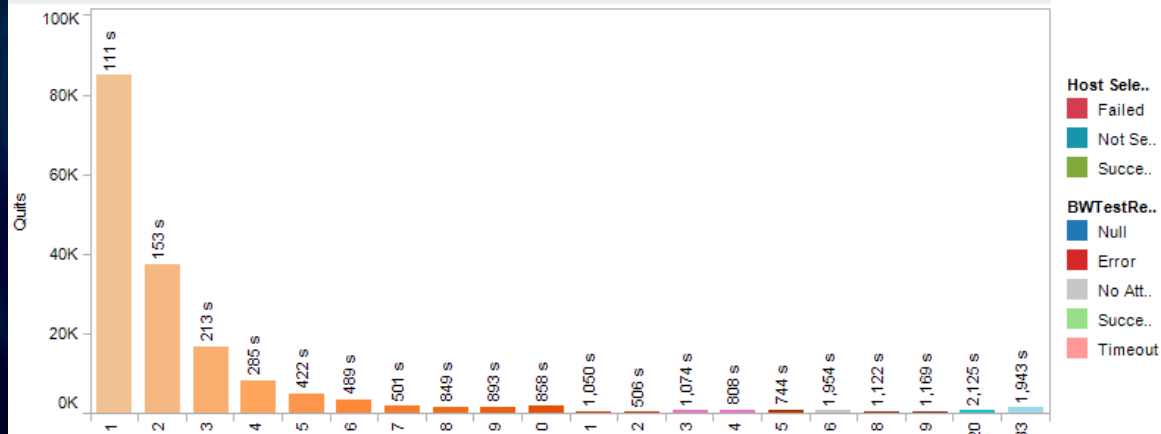


Social Join Action Results Trend by Search Type

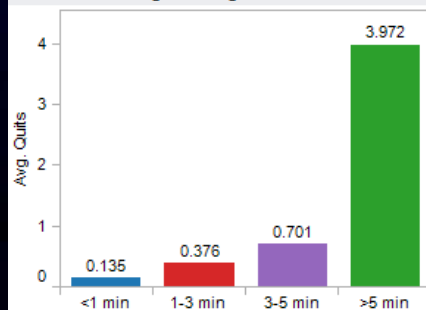




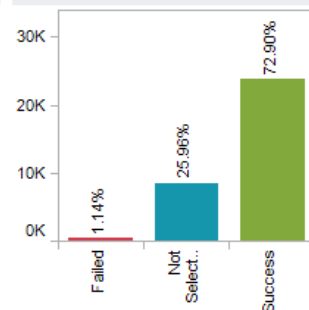
Avg Elapsed Gather Time by Number of Quits



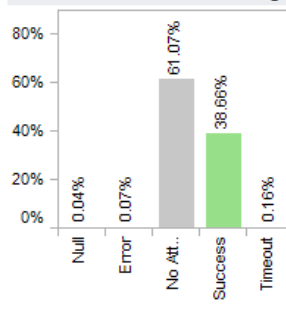
Average number parties that quit from a gathering host



Optimal Host Selection Results



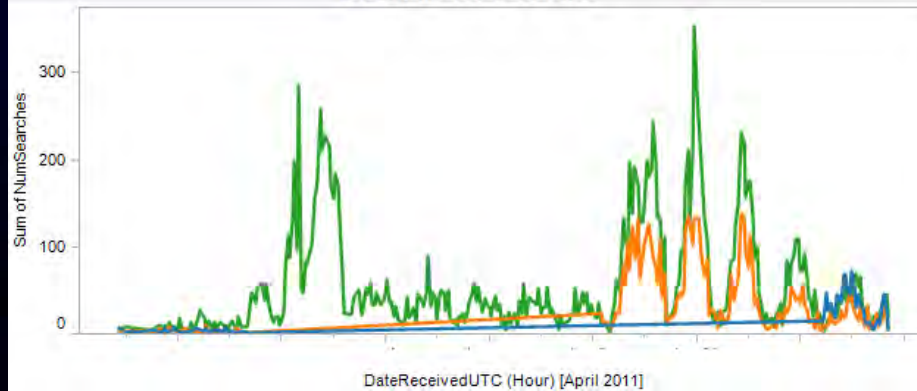
Player Hosted Reservations That Had Bandwidth Testing



Successful Social Matchmaking Times by Region



Total Raw Social Searches



Comparing Apples to Oranges

- Different Granularities
- No Cohesive Picture
- Only Experts Could Divine Answers
- We Needed ...

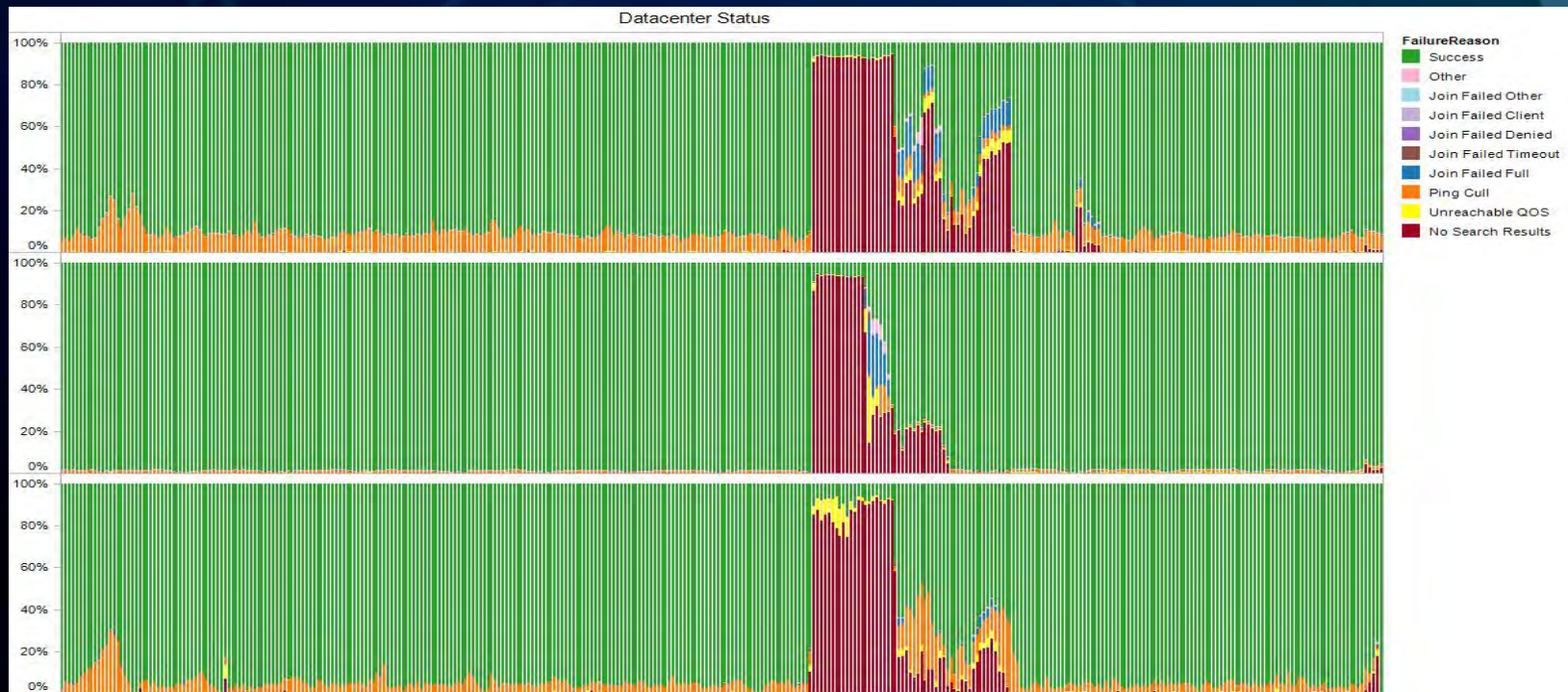


The One True Chart

- 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

A large, faint, blue watermark of the EPIC GAMES logo is centered in the background. The logo consists of the word "EPIC" in a large, stylized, outlined font, with the word "GAMES" in a smaller, similar font directly below it. The background is a dark blue gradient with abstract circular and curved lines.

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

?

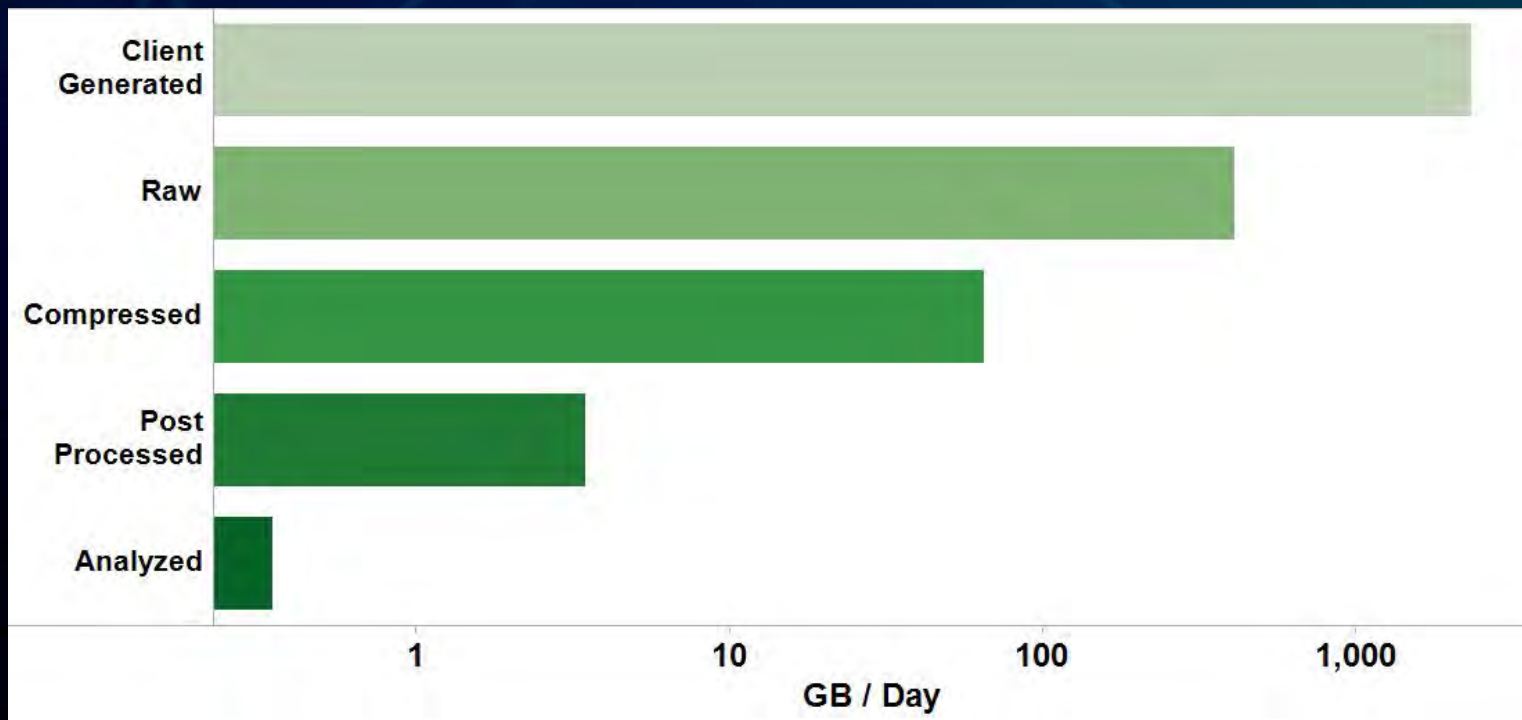
Profit



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

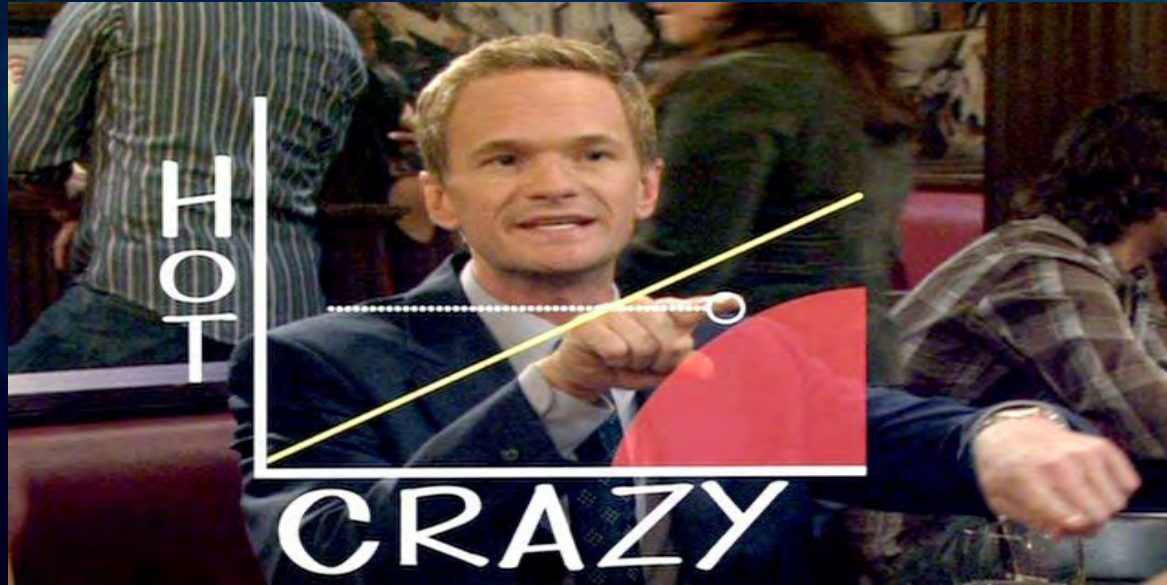


Selecting Questions to Answer

- 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



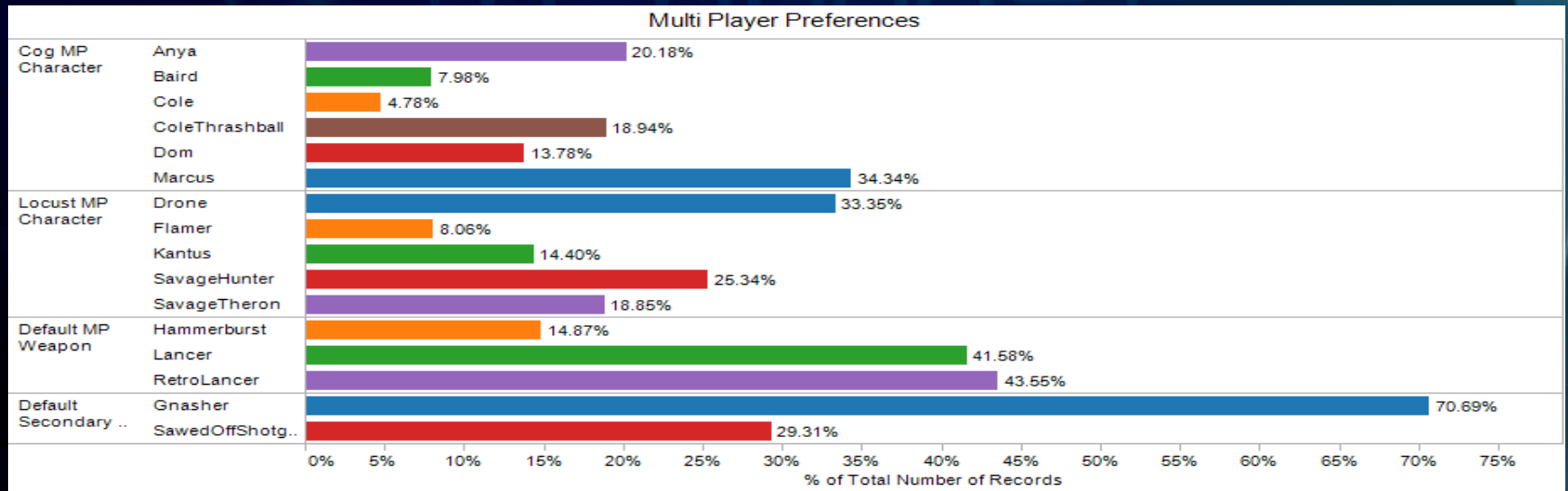
Actionable Data

- What can you Change?
- When can you **Use** this Information?
 - Config files
 - Beta
 - Title Update
 - Next Game
 - Press Release
 - Forums



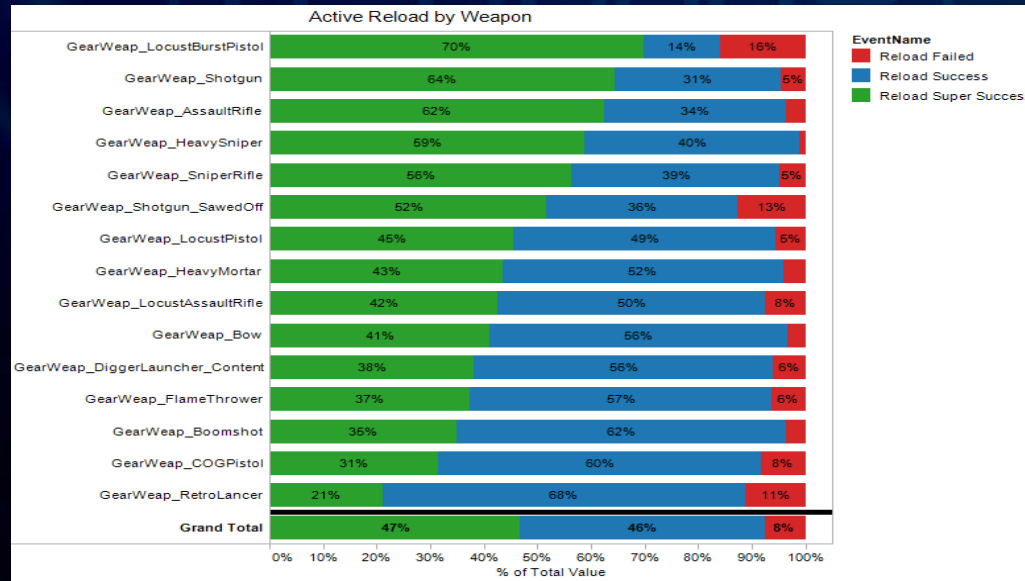
What Will Knowing Get You?

Multiplayer Character Preference:



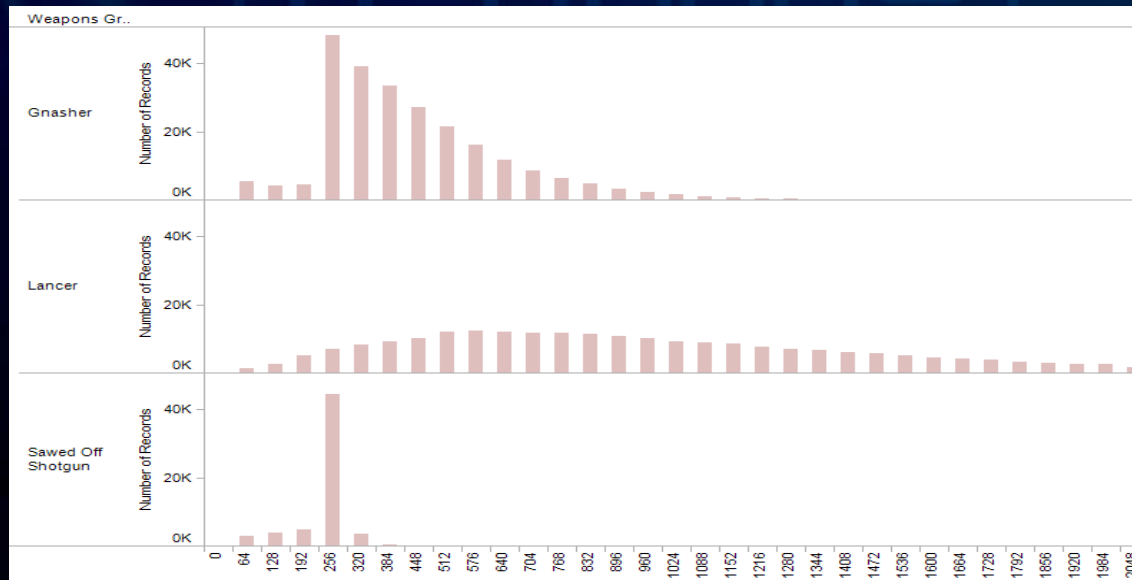
What Will Knowing Get You?

Active Reload Success Rates



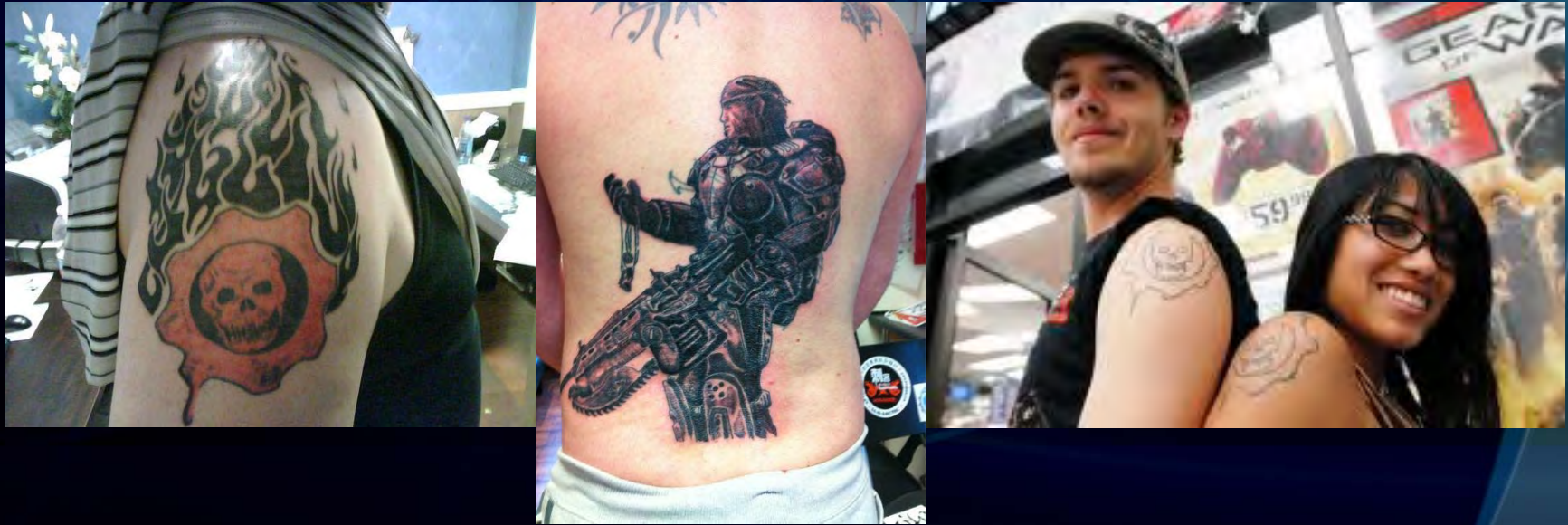
What Will Knowing Get You?

Downing Distances



What Will Knowing Get You?

Avid Fans

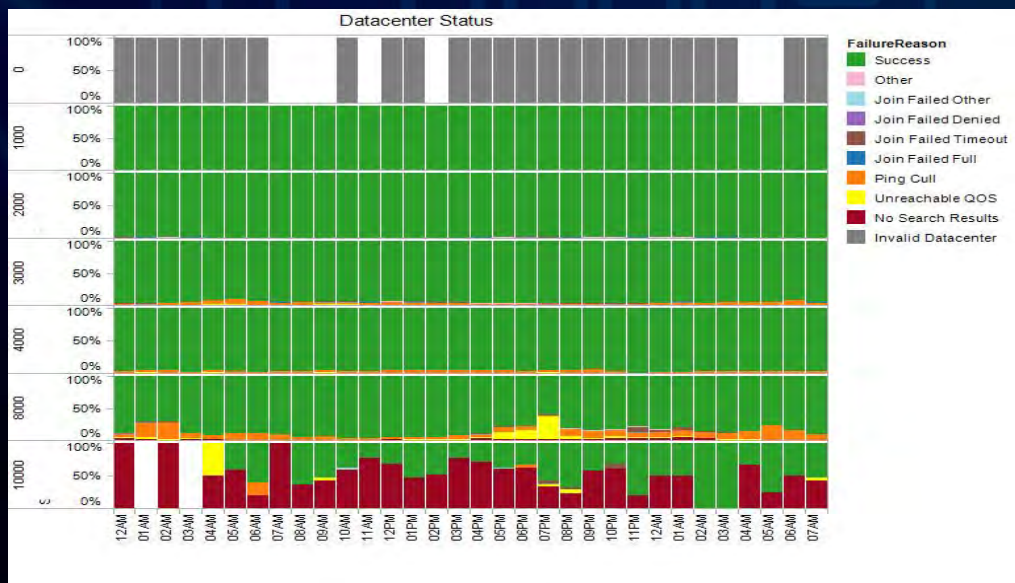


What Will **NOT** Knowing **COST** You?



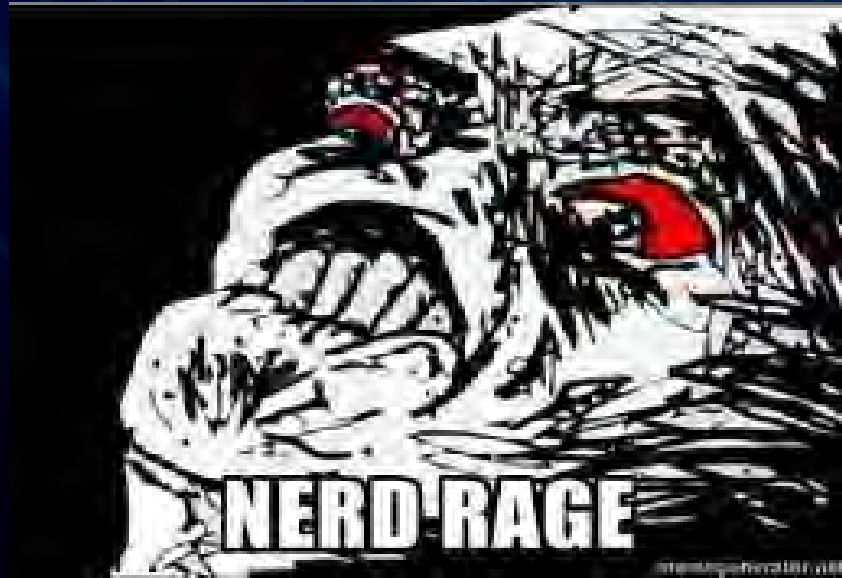
What Will **NOT** Knowing **COST** You?

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

A large, faint, blue watermark of the EPIC GAMES logo is centered in the background. The logo consists of the word "EPIC" in a large, stylized, outlined font, with the word "GAMES" in a smaller, similar font directly below it. The background is a dark blue gradient with abstract, curved light blue lines and a large, semi-transparent circular shape on the right side.

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

GEARS OF WAR 3

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

Storage Basics

You Try Counting All Those Kills

Events

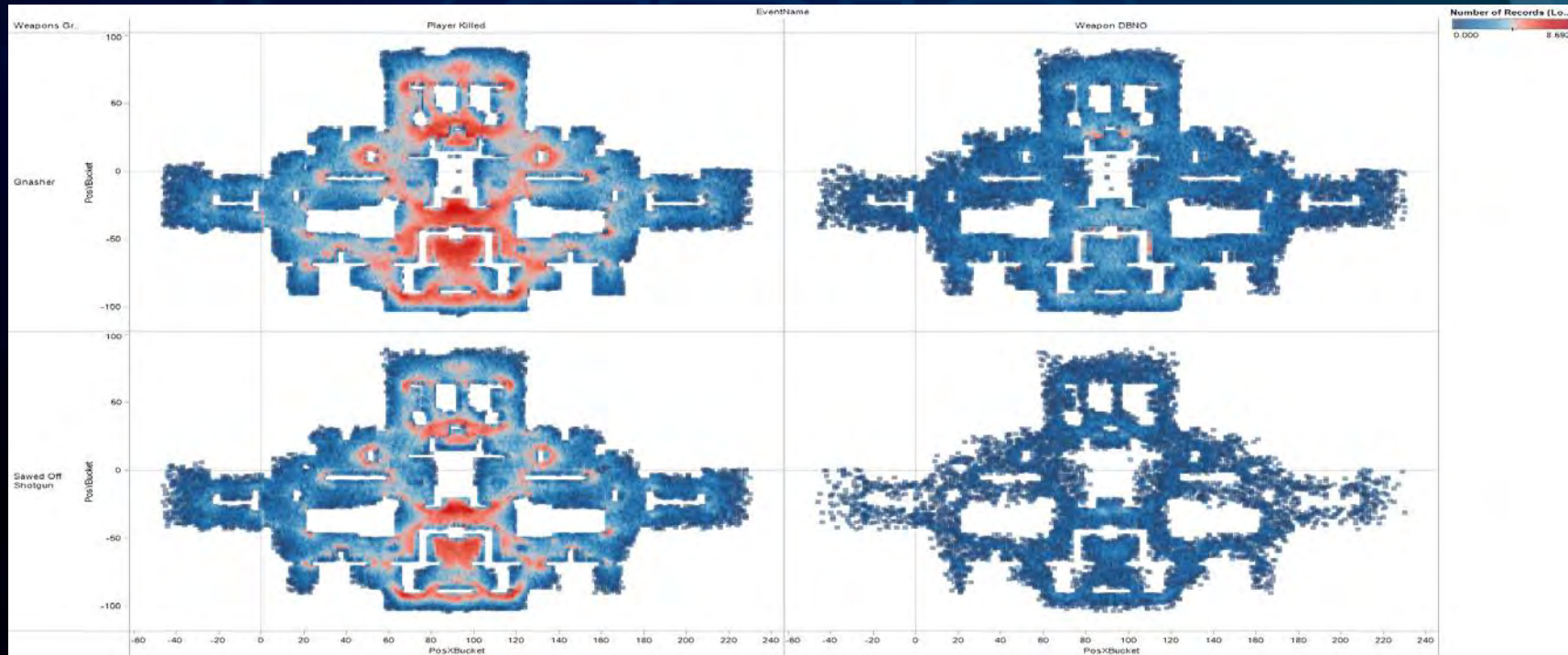
Event	Damage Type	Source Player ID	Target Player	MapName
Kill	Lancer	P1	P2	MP_Hotel
Death	Lancer	P2	P1	MP_Hotel
Kill	Gnasher	P1	P2	MP_Hotel
Death	Gnasher	P2	P1	MP_Hotel
Kill	Sawed-Off	P1	P2	MP_Hotel
Death	Sawed-Off	P2	P1	MP_Hotel
...
Kill	Gnasher	P1	P2	MP_Hotel
Death	Gnasher	P2	P1	MP_Hotel
Kill	Gnasher	P1	P2	MP_Hotel
Death	Gnasher	P2	P1	MP_Hotel
Kill	Gnasher	P1	P2	MP_Hotel
Death	Gnasher	P2	P1	MP_Hotel

Counting Kills

Event	Damage Type	Source Player ID	Target Player	MapName
Kill	Lancer	P1	P2	MP_Hotel
Death	Lancer	P2	P1	MP_Hotel
Kill	Gnasher	P1	P2	MP_Hotel
Death	Gnasher	P2	P1	MP_Hotel
Kill	Sawed-Off	P1	P2	MP_Hotel
Death	Sawed-Off	P2	P1	MP_Hotel
...
Kill	Gnasher	P1	P2	MP_Hotel
Death	Gnasher	P2	P1	MP_Hotel
Kill	Gnasher	P1	P2	MP_Hotel
Death	Gnasher	P2	P1	MP_Hotel
Kill	Gnasher	P1	P2	MP_Hotel
Death	Gnasher	P2	P1	MP_Hotel



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

Roll-ups

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

Counting Kills

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



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



ANALYZE



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



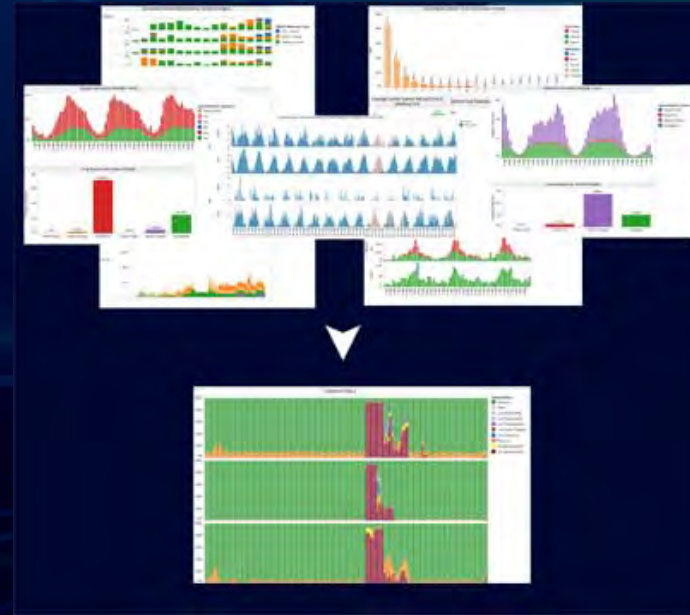
Iteration Data

- Generated
- Game QA
- Playtest
- Beta
- Live



Constantly

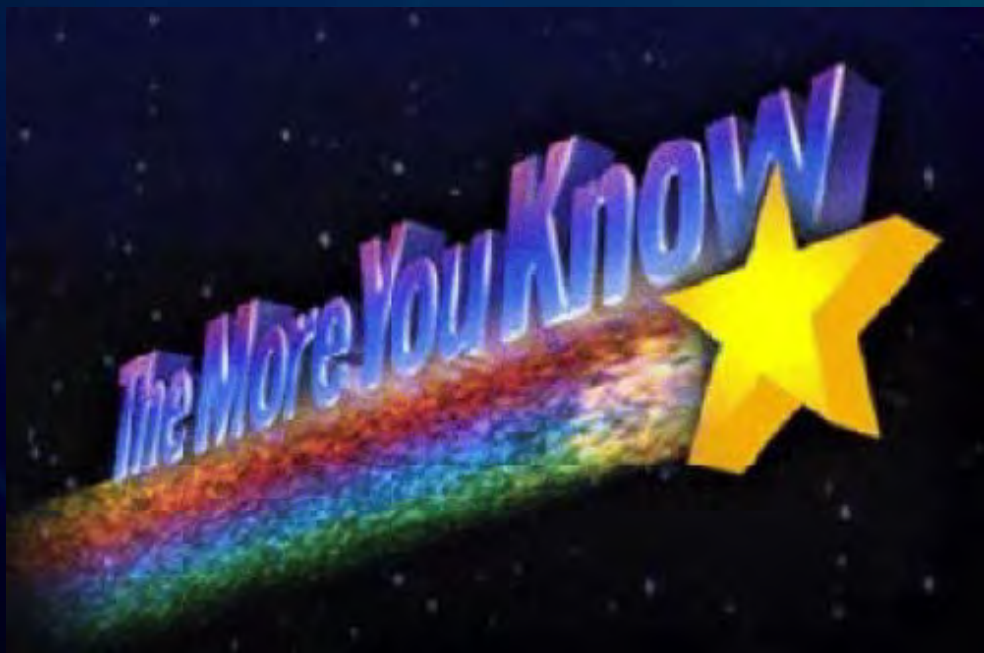
- Evaluate Priorities
- Solidify Categories
- Tweak Storage Layout
- Improve Charts



A large, faint, blue watermark of the Epic Games logo is centered in the background. The logo consists of the word "EPIC" in a stylized, blocky font above the word "GAMES" in a similar but slightly smaller font. The background is a dark blue gradient with some abstract circular shapes.

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.