



Connecting the World Through Games

# Zynga Analytics

Leveraging Big Data to Make Games More Fun and Social

*Daniel McCaffrey*

*General Manager, Platform and Analytics Engineering*

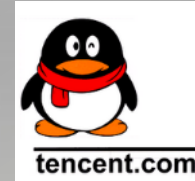
# World's leading social game developer



...

And growing rapidly web and mobile 3<sup>rd</sup> party games on the Zynga Platform

Built on global platforms,  
and our own Zynga API (services) and platform



- Play with your friends
- Synchronous or Asynchronous play
- Cooperative or Competitive play



# By The Numbers

## Users

- ~260 million MAUs
- ~60 million avg DAUs worldwide

## Game Data

- Vertica driven
- ~60 billion rows/day
- ~10TB daily semi-structured data
- ~1.5PB source data
- Largest 230 2U nodes

## Server Data

- Splunk
- 13TB per day raw logs from server and app logs
- Vertica or Hadoop for archives

# Starting Core Concepts

What helped made Analytics successful at Zynga

# Metrics Driven Culture

- Management desire to track goal progress by metrics

# Analytics Everywhere

- Wanted open data access as much as possible
  - Freely accessible reports by everyone
  - Open Ad-hoc SQL access
  - Easy external service integration

# Ease of Use and Integration

- Wanted easy/standard tool integration
  - ETL/ELT tools
  - Analysis tools/DB visualizers
  - Reporting
- External service integration via SQL
- Control data structure at moment logged via an API
- “Semi-structured” data capture for flexibility
- Centralized data schemas for easier analysis



# Organizational Structure

- Centralized
  - Data/BI (centralized data schema and aggregation)
  - Data Infrastructure (centralized data flow)
  - Network level Data Analysts
- Centralized but embedded
  - Game and partner group Data Analysts
    - Schema, architecture and data knowledge
    - Share insights company wide

# Art + Science, not Art vs Science

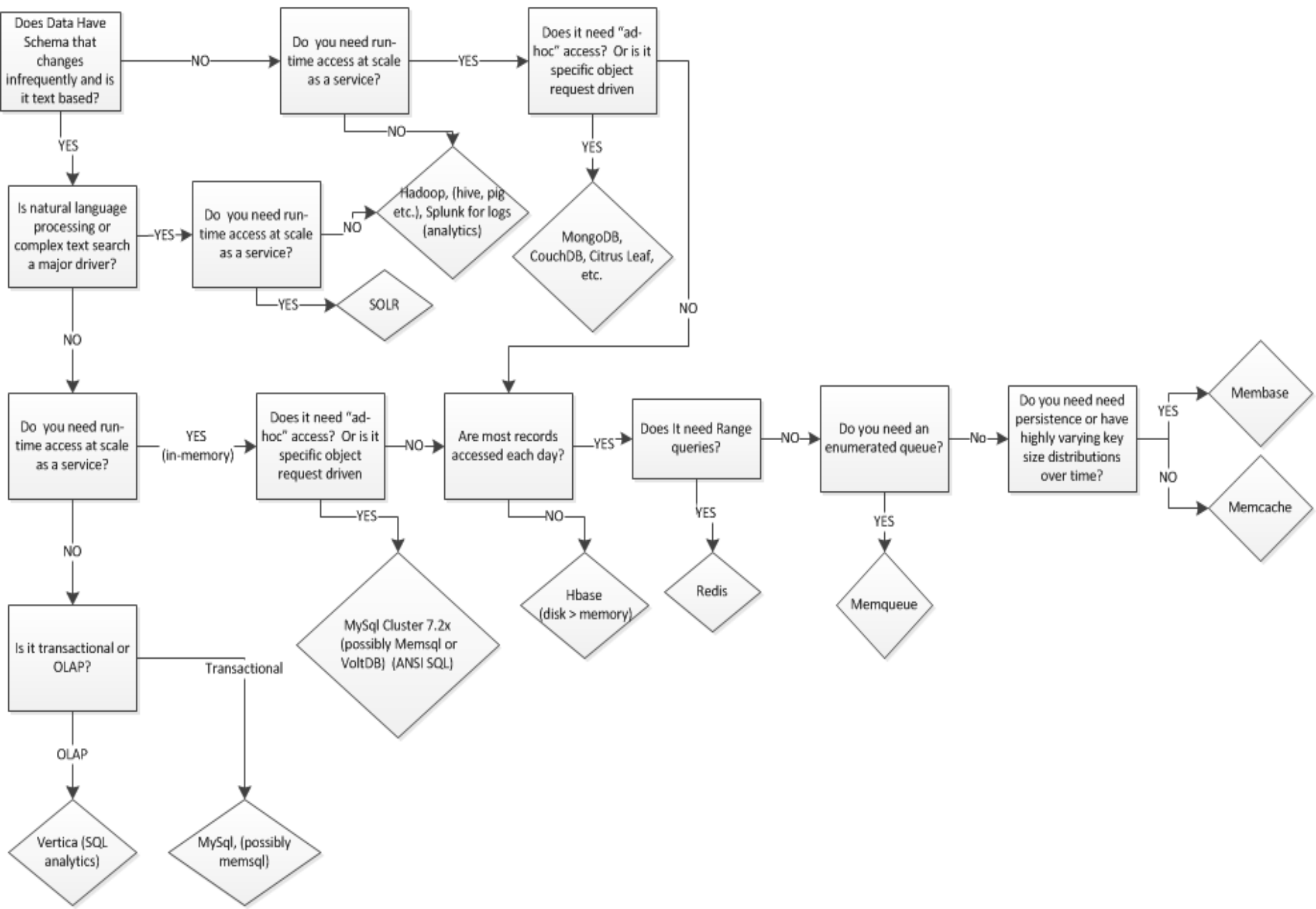
- Art: Generate the game idea and implement
- Science: Find out if it's good/fun. Listen to the players.

# SERVICES AND ARCHITECTURE

# DATABASES IN USE

- Vertica: primary game/user analytics stores. 10TB/day,70 billion rows/day
- Splunk: primary log analytics stores 13TB/day
- MySQL Cluster 7.2x: streaming event DB. 70 nodes,650million rows/day
- MySQL: many single node and sharded transactional DB's
- Membase: memory store with persistence. Replaced memcache+mysql.
- HBase: Messaging service store. Disk>memory
- Memcache,Memqueue: Service stores when persistence isn't needed.
- SOLR: Run-time text search needs
- Redis: Service stores with ranged queries.
- Oracle: Finance
- *Memsql: being looked at*

# Dan's Scalable Database Decision Matrix at Zynga, 2012



# ZTRACK API

- Simple to use logging API
  - PHP, Java and Ruby
  - REST API as part of Zynga API
- Backend Leverages:
  - FB Scribe for scalable, fast worldwide message forwarding
  - Custom Java “ETL” database loaders
- Semi-structured data logging (flexible taxonomy)

PHP example

```
ztrack_count($user_id, "myevent", $value1, "kingdom", "phylum", "class", "family", "genus");
```

# Vertica Data Warehouses

- MPP Compressed Column Store, Full ANSI SQL
- 6-9x compression on data, extremely fast bulk loading
- Stats
  - >60Billion rows/day, trickle-in/real-time from ZTrack
  - >10TB/day
  - Largest is 230 2U nodes, next generation will be 560 to 1,000
- Clusters
  - Production and Mirror, Social Graph, Sample, Virtual Goods tracking for revenue recognition/sox, Poker hands, International, Test and Staging...

# Reporting and Analysis

- stats.zynga.com
  - Over 6,000 distinct report types
  - ~1080 DAU, ~1,480 WAU
  - ~3,000 report runs per day, 500-600 distinct reports each day
  - ~15,000 ad-hoc queries from users per day
  - Taxonomy slicer reports
- Ad-hoc SQL access clusters for analysis
- Analysts, Product Managers, Engineers and BI team work to create new insights, metrics and profiles and operationalize



# Data Services

- Allows for run-time decisions in game or services
- API backed by fast in-memory data access (membase)
  - Network level data across data centers using membase sync
  - PHP and Java and REST API as part of Zynga API
- Access to real-time and daily aggregated user and game data, network level
- Some uses include:
  - Personalization
  - Targeting
  - Profiling
  - Matchmaking

# Experiment Platform, A/B Testing

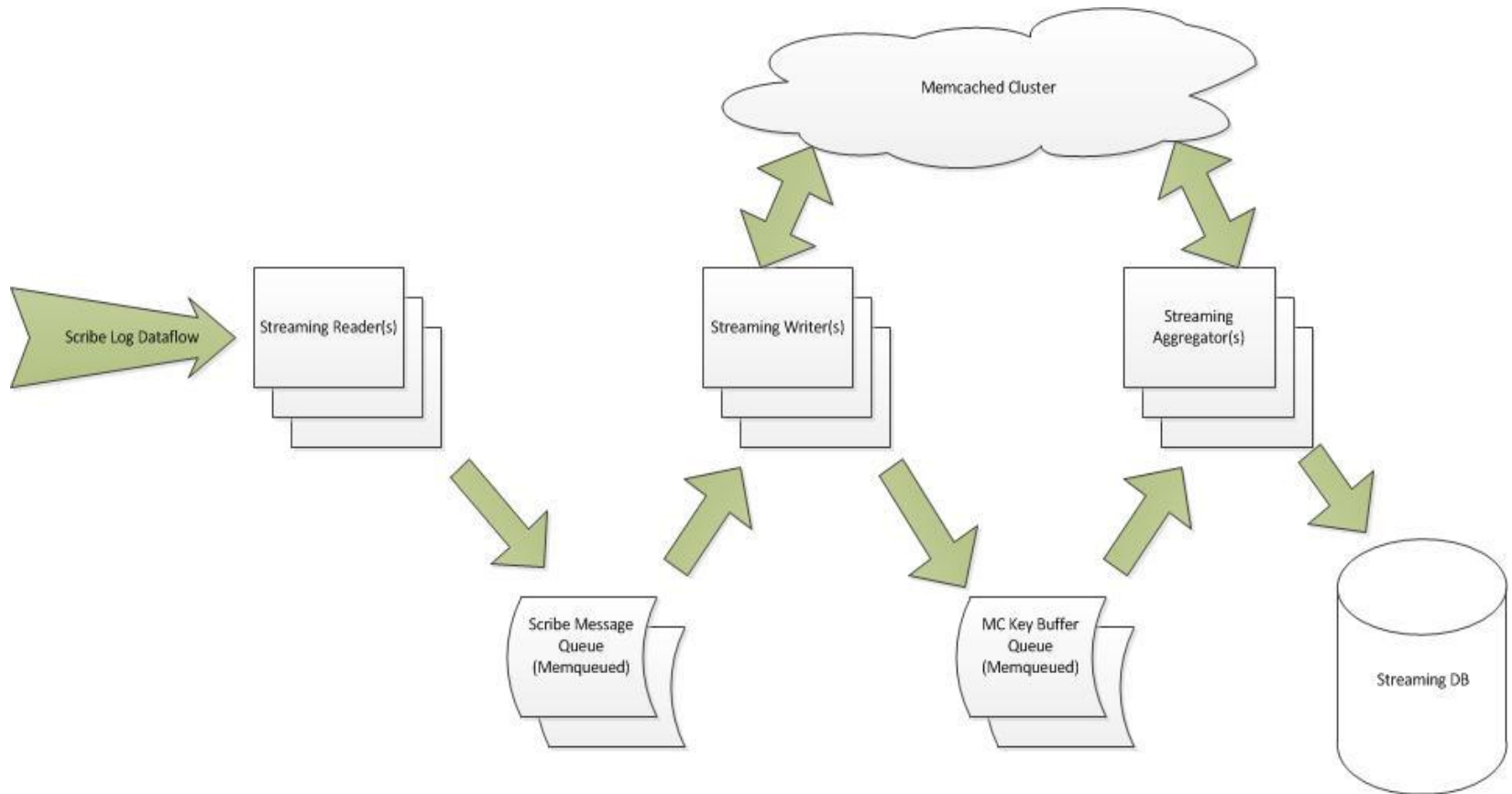
- Provides real-time:
  - Controlled Experimentation via web UI and game hooks
  - Reporting
- Simple API
  - Java, PHP and REST API
- Impact Game and Platform Design
  - Ability to see what happened in real time
  - Lots of experiments. Many fail. ~3-5K active at any time atm.

# Real-Time Streaming Data Events

- Real-time *scalable* event aggregation, using time windows
- Presently handles over 70 billion events per day
- Technology:
  - Custom java for processing
  - Memcaches
  - Memqueues (“enumerated” memcache)
  - MySql Cluster 7.2x: ~70 nodes, 624mil rows/day, 300k query/day

*Please see Michael Fan and Rushan Chen’s lightning talk on Streaming later today*

# Streaming Architecture

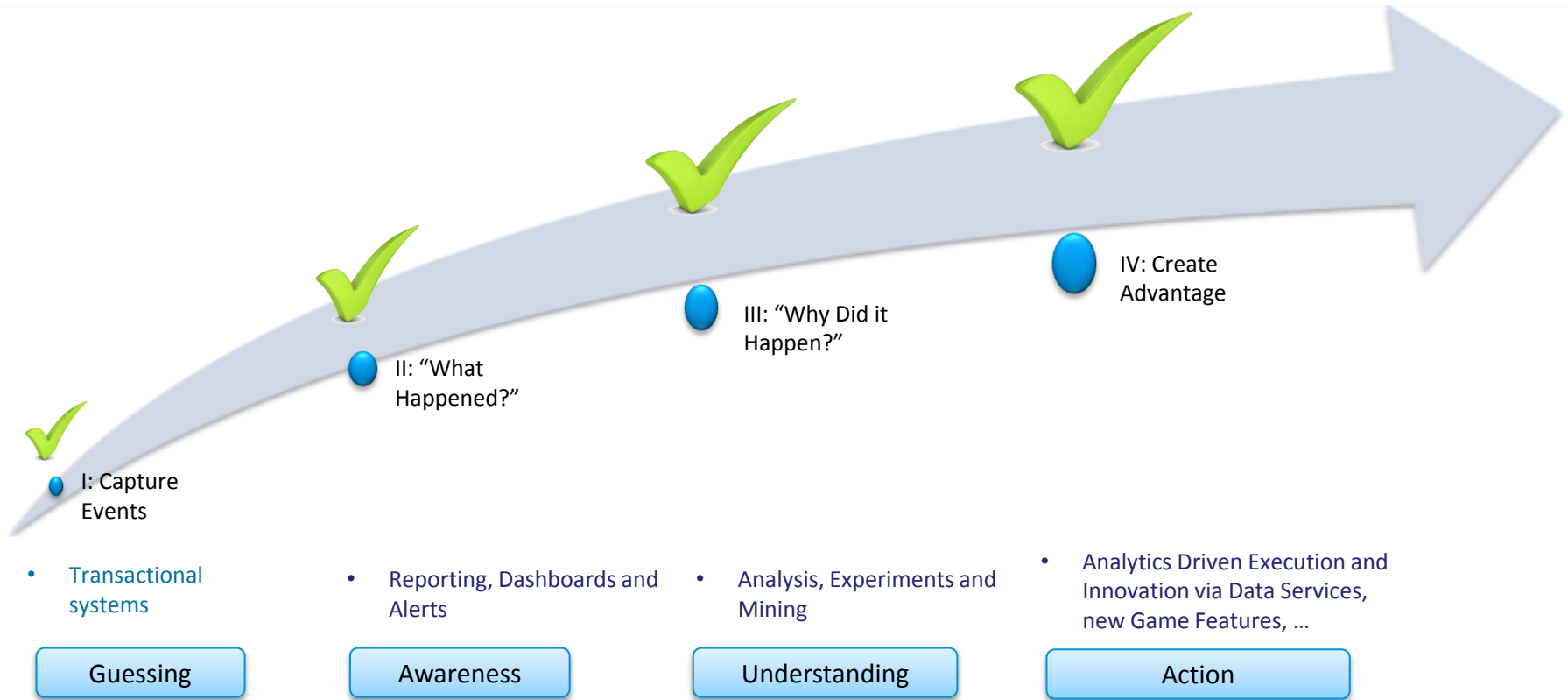


# Streaming Uses

- Operation health monitoring/reporting and alerting
- Fast key metric reporting, offloading from Vertica
- Data Validation– Compare to Vertica
- Future: more timely run-time decisions in-game

*Please see Michael Fan and Rushan Chen's lightning talk on Streaming later today*

# Analytics Maturity at Zynga



# APPENDIX/SUMMARY

# Services Summary

- Centralized network level tracking, reporting and warehouses
- Embedded analysts as a service
- Centralized network analysts, BI and data infrastructure
- Data Services– Run-time data access
- Experiment service – Easy A/B testing
- Streaming event service – real-time scalable event aggregation



# Zynga Core Concepts Summary

- Commitment to a metric driven mindset
- Open Access to data– reports, ad-hoc and external services
- Ease of use – tools, external services, schemas
- Art + Science
  - Experiment on ideas, analyze and make changes.
  - Use analytics to listen to the players and make changes
- Log with an API, add some structure data as possible at moment logged
- Standardize taxonomies quickly and enforce once mature



Connecting the World Through Games