



Architecture at Scale

About the speaker

- Manny Pelarinos- Sr. Director of Engineering, ESPN
- manole.pelarinos@espn.com
- Been with ESPN for 5 years
- First time at QCon!



Agenda

- High Level Architecture
 - Overview
 - Data Ingest
 - Core Service Architecture
- Advanced Systems
 - Content Management System
 - Template Language
 - Live Scores
 - Personalization
- Q&A

ESPN.com Facts & Figures



ESPN.com Facts & Figures

- ESPN.com has the 5th largest audience of men 18 to 54 across the entire Internet, behind only Facebook, Google, Yahoo and Microsoft!
- Serves over 100K RPS with few servers
- Almost entirely Java based
- ESPN digital properties include
 - ESPN.com
 - Fantasy games
 - Mobile
 - WatchESPN
 - ESPN the Ocho
 - No, not yet but others (Deports, W, HS, etc)



ESPN.com Mission

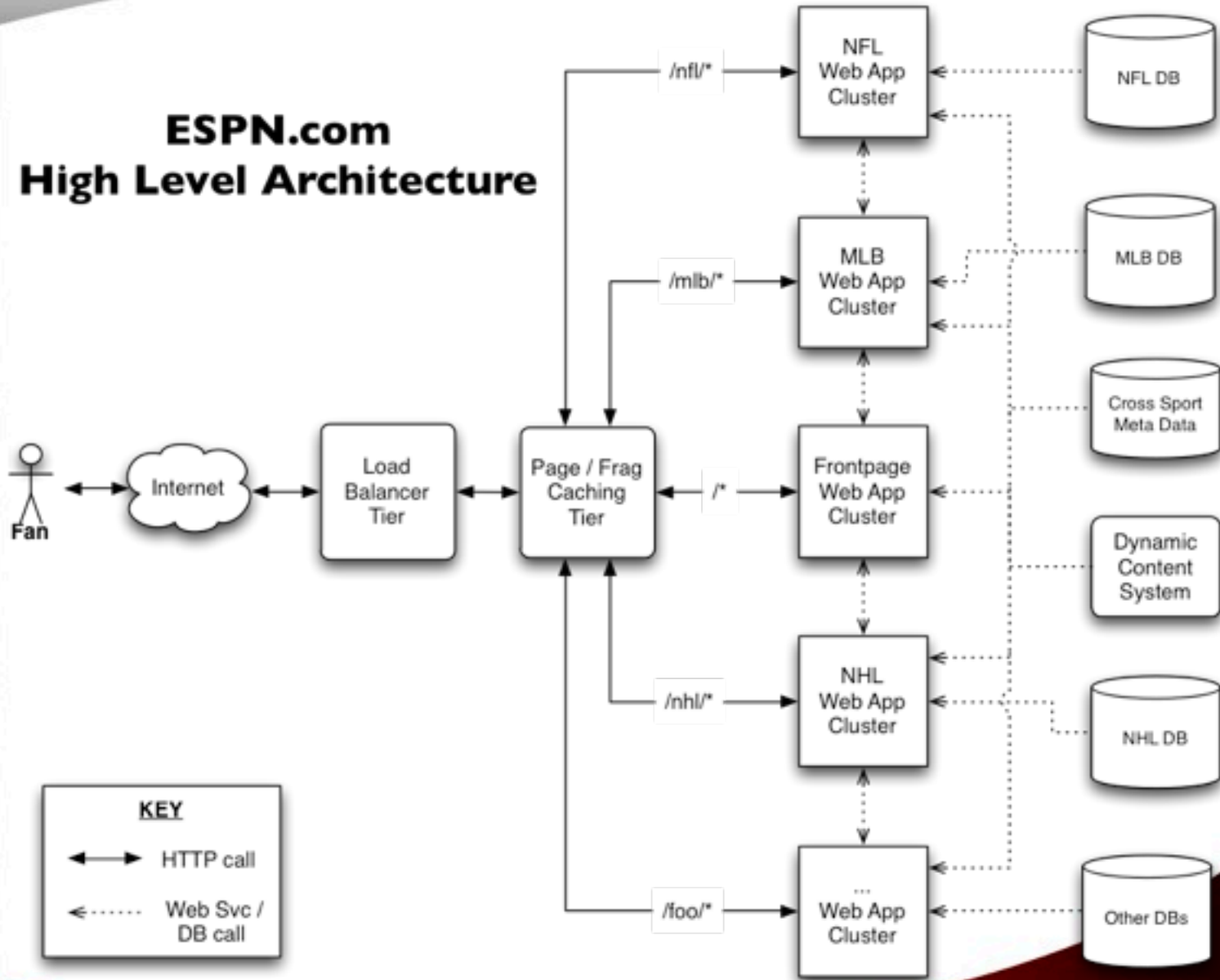


ESPN.com Mission

- Serve sports fans anytime, anywhere on any device
- Availability & accuracy of the utmost important
- Bring fan all stats and scores plus deeper content



ESPN.com High Level Architecture



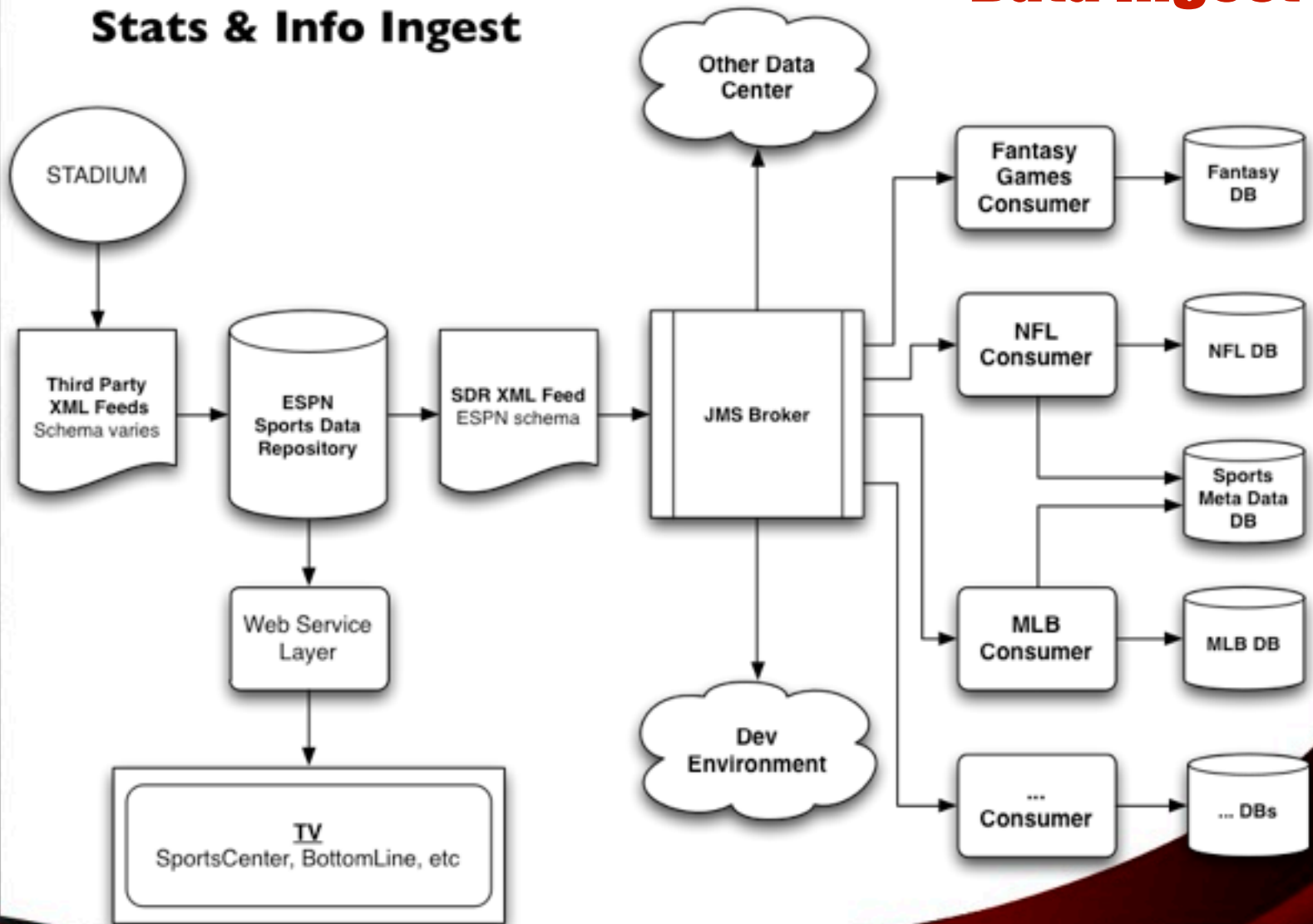


Data Ingest



ESPN.com Stats & Info Ingest

Data Ingest

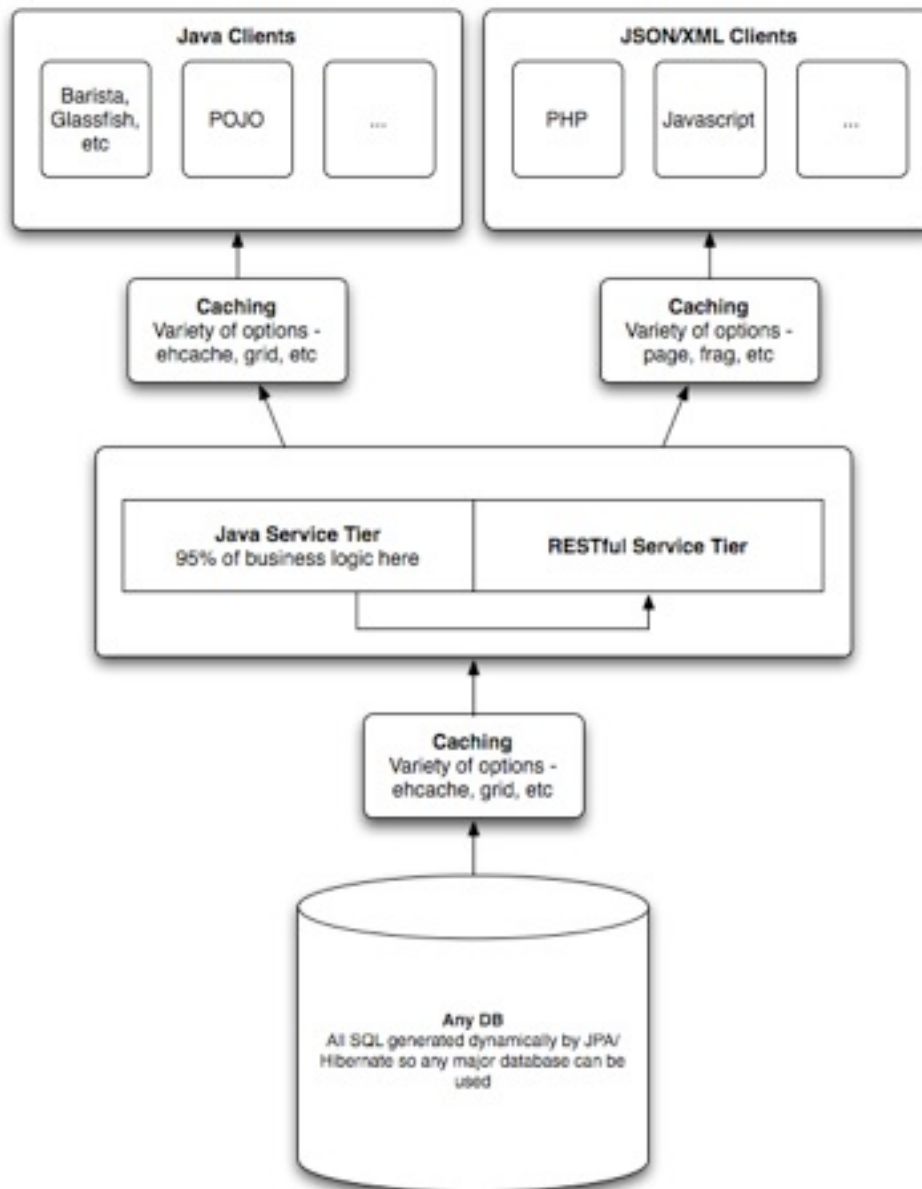




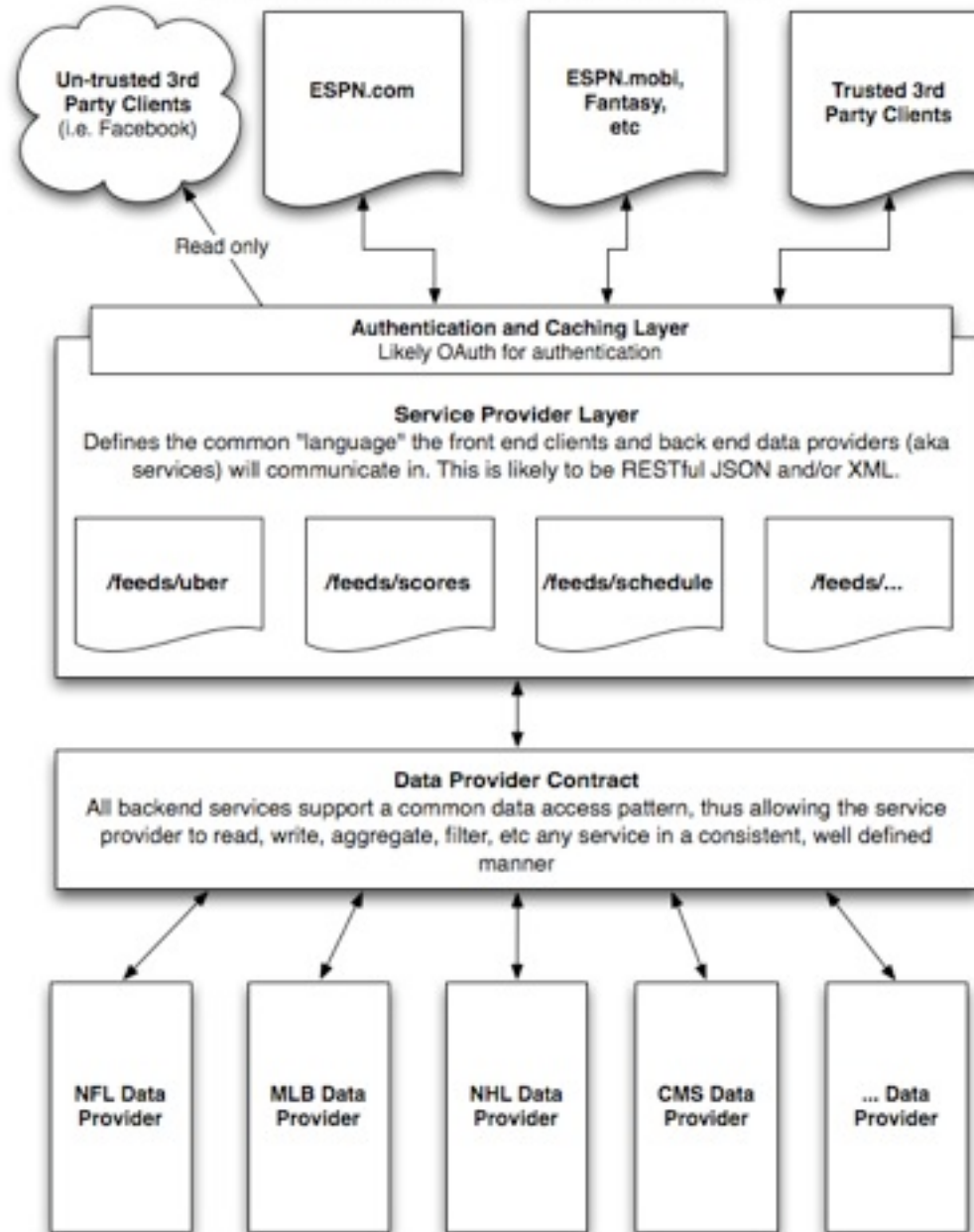
Application Services



Application Services



Any JSON or XML aware client can theoretically be supported





Cache Push

- Moving to cache push model
- Data ingest process has converts incoming XML message into JPA POJO
 - Inefficient to have DB send expire
 - MDBs and webapp both talk POJO so just push it
 - Remove biggest bottle neck (our DB) from the equation
- All live event data gets delivered to the web application by ingest (MDB) process
 - Eliminates millions of DB calls per hour during peak times
 - Once an application has primed its caches with historical & meta data, DB could theoretically be turned off



Hibernate Caching

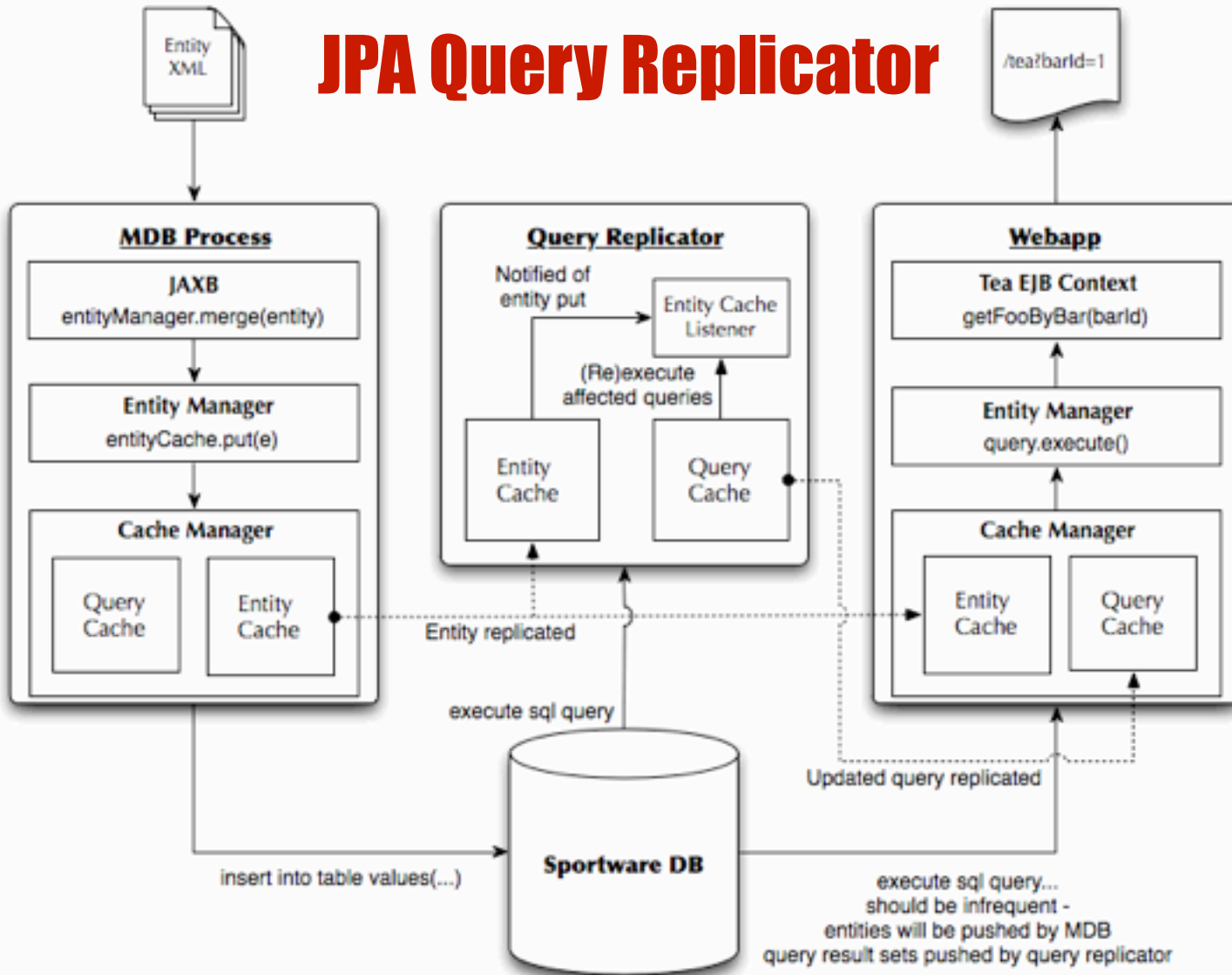
- Using Ehcache as 2nd level cache provider with cache replication enabled
- Works great for entity updates
- **BIG PROBLEM: almost everything on .com looked up via query...**
 - Hibernate support for query caching inefficient
 - Can't have getPlaysForGame(1234) query banging on DB all day b/c query cache only supports TTL or dopey last update timestamp
 - What to do....



JPA Query Replicator



JPA Query Replicator

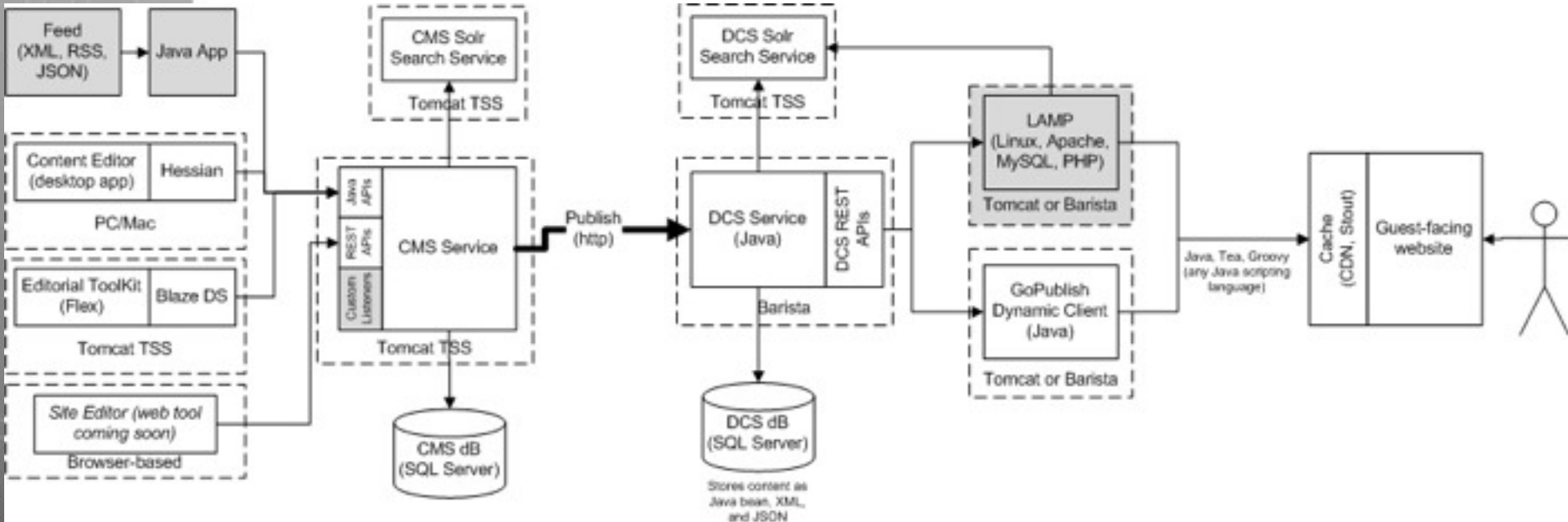




Content Management System



Content Management System





ESPN.com Web Template Language

- Proprietary, high performance template framework
 - <http://teatrove.org/>



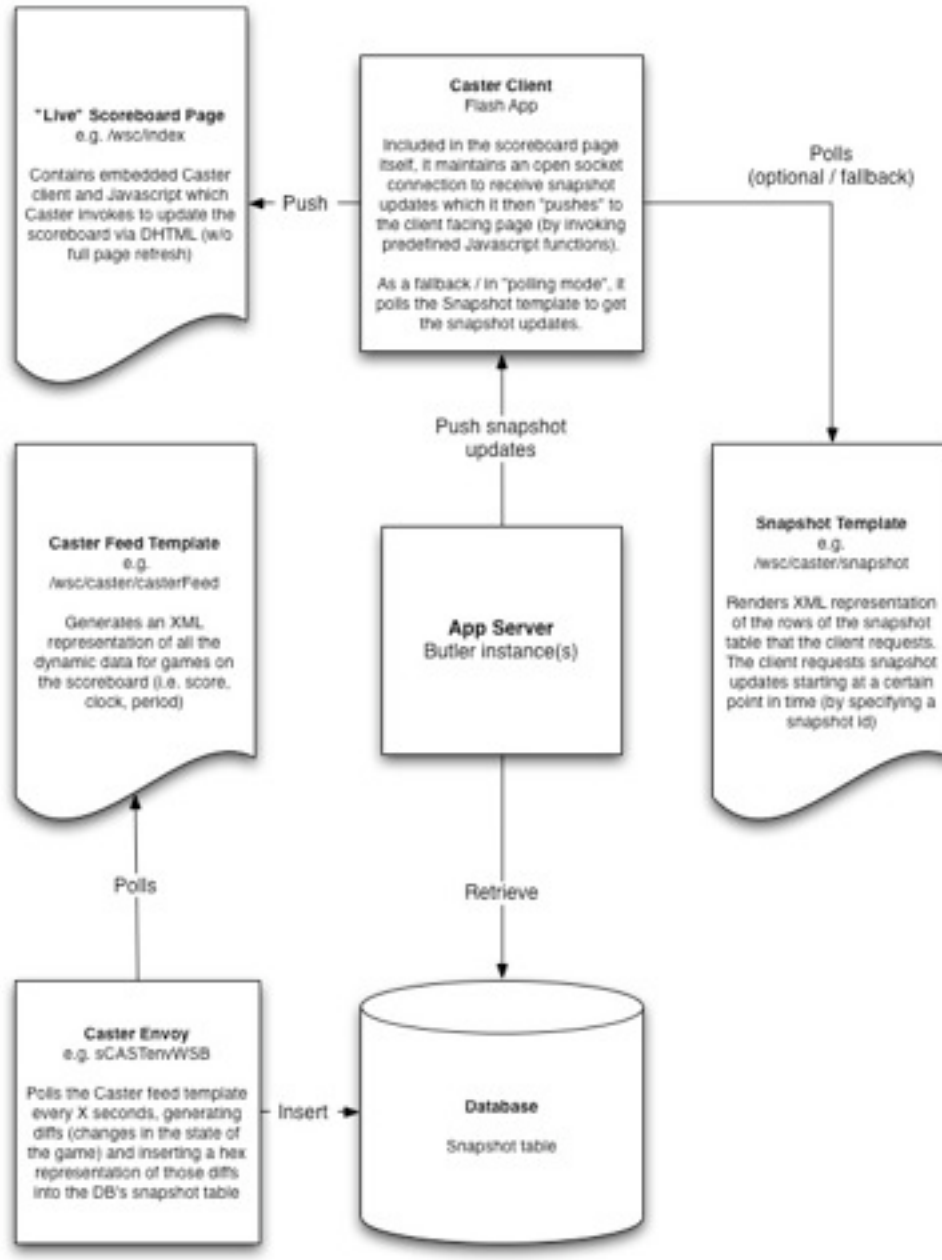
- Stout
 - High performance page caching
 - Peer fetch enabled
 - Per URI, TTL based expiration
 - Blocking and non-blocking source fetch
 - Automatically demotes unresponsive source servers
 - Runs on cheap / low end hardware



ESPN.com Live Scores

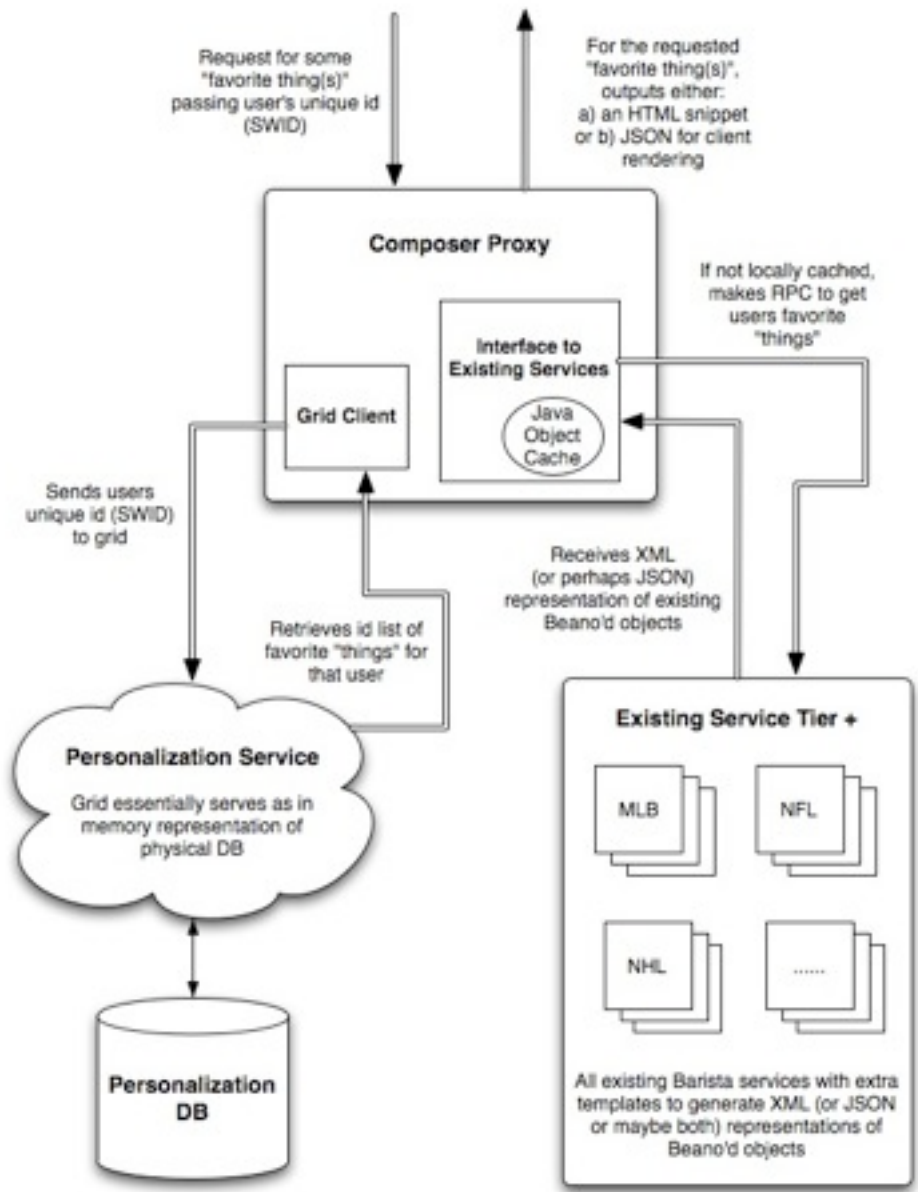


ESPN.com Live Scores



Personalization

- Fan desire to have personalized content and presentation
 - And have it follow you everywhere (.com, mobile, TV, etc)
- Need to build scalable, high performance, distribute cache
 - >500 GB of data
 - Constrained tree map - Lookup primarily by ID
- But site is heavily page cached... not conducive to personalized user experience



APIs ARE FOR NERDS

(Nerds who build freaking awesome sports stuff.)



REGISTER

You'll need a developer key to use our API. Sign up!

REQUEST A KEY

BROWSE

Access exclusive data, sports content, scores & more.

EXPLORE APIs

BUILD

Have a key and ready to start using our data?

GET STARTED

FROM THE SHOWCASE [See All >](#)



foursquare

Check in to sports events with the ESPN experience on Foursquare.
Leverages: [Schedule API](#), [Research Notes API](#)



Pulse

The latest ESPN news and headlines in the Pulse reader app.
Leverages: [Headlines API](#)



Flipboard

Browse your own personal sports magazine with ESPN content in Flipboard.
Leverages: [Headlines API](#)



BYU

Follow the BYU Cougars, including the latest team scores, on BYUtv Sports.
Leverages: [Scores API](#)

MOST POPULAR FAQs

What is an API?

Why has ESPN decided to make its content/data available for outside users?

Do I have to pay to use the ESPN API?

POPULAR DATA

News & Headlines

Research Notes

Scores & Schedules

Athletes

QUICK LINKS

Getting Started

Using the ESPN API is as easy as 1-2-3.

Branding and Attribution Policy

Read the rules on how to credit us for using the ESPN API.



Challenges

- Designing APIs is hard
 - Fine grained vs. coarse grained debate
 - More generic APIs are more reusable, but often require the client to make many calls and caching is tough
 - More specific APIs are more efficient but the variations get out of hand really fast
- What did we settle on
 - Generic APIs for now and build use case driven ones as needed
 - ‘Coherent TTL’ cache on the way to help with mobile

Q&A



- **We are hiring!!!**
- Go apply at <http://espncareers.com> ... or better yet talk to me afterward
- Questions?
 - Feedback on anything you heard?
 - Suggestions for ESPN.com features?
 - Click that little happy face for my rating and I'll fix your fantasy football scores!