JUDCon 2013: United States
JBoss Users & Developers Conference
Using Drools and other JBoss projects to build real time mobile app performance analytics

Prabhat Jha
@prabhatjha
Who is Prabhat Jha?

• Loves to Build Stuff ...of value.
• Co-founder of InstaOps
• Director of Software Development at Apigee
• Member of JBoss Data Grid (Infinispan), JBoss EPP, Richfaces team.
• Founder of Eejot, a non profit to help improve computer literacy in Nepal
• @prabhatjha LinkedIn: prabhatkjha
To do List:

- A brief history of InstaOps
- Quick Demo
- Different technology stacks used to build InstaOps
- Architecture Evolution - good, bad and ugly 😊
- Lessons Learned

@prabhatjha
InstaOps

• A software start up from Austin, Texas
• Founders - Alan Ho and Prabhat Jha
• Bootstrapped
• Goal : Help developers build “better” mobile apps based on experiences from building “better” web apps.
Factors Affecting App Perf

- Device Platform
- Device Model
- Network Carrier
- Network Type
- Location
- App Store Policy
Demo

- Crash Detection
- Log capture
- Network Performance
- Alarm and Notification
- Configuration Management
Why it’s not an easy problem?

- Multi-tenant
- App with handful of users to millions of users.
- App with average usage less than a minute to few minutes to hour
- Flaky network connection
- Battery Optimization
- Limited Data Plan
- Not chronological data - Some data could arrive after few weeks.
- SDK needs to be fail safe
1st Iteration: Network Performance

SDK
RESTEasy
Hibernate
GateIn + Richfaces + PortletBridge
Custom Amazon EC2 AMI
Amazon RDS
2nd Iteration: + Log Capture

SDK
RESTEasy
Hibernate
GateIn + Richfaces + PortletBridge
Custom Amazon EC2 AMI
Amazon RDS
3rd Iteration: Configuration Management

SDK
RESTEasy
Hibernate
Seam*
Richfaces
Amazon Elastic Beanstalk*
Amazon RDS
Got the Min Viable Product (MVP)

- SDK sending data in real time.
- Visualization
- Closing the loop
Hibernate

- `hibernate.hbm2ddl.auto`
- Batch insert
- Flexibility to switch between HQL and native SQL
- ResultTransformer, `aliasToBean`
- Problem with Projection and where clause (i.e this.blah)
- Explicitly identifying attributes type (.addScalar ..)
- Enum - little bit of problem
Richfaces

• Savior for those who don’t know JavaScript
• Partial page refresh and granular control
• Good support for pagination and search to tie to Hibernate Criteria
• Good community
Seam

- Better support for conditional page flow
- Helps tie Front End and Back End together
- Custom Authentication, Custom Authorization, SSO, User Sign up Customization
RDBMS (MySQL)

- Familiar
- ORM
- Tooling
- Cloud
- Easier Custom Report
- FB, Twitter etc used and still use.
- “Easier” White labeling of the product
Performance & Availability

- Loss of data when REST server is down.
- Problem with high volume apps.
- Aggregation for UI/Visualization painfully slow.
Introduce Queue

• Amazon SQS
  • High availability
  • Not expensive
  • Read lock
  • Fine grain access control
• From REST to Polling SQS
• Buffer hence increases availability
Batching and CEP

- Batching & Read Lock from SQS
- Aggregation using stateful CEP with Drools
  - Local aggregation
  - Optimal active sessions, session length calculation
  - Less stress on database
  - DB stores local aggregates
Problems with Stateful CEP

- Data Loss
- OOM
- Need Clustering & Synchronization for auto scale in and out
- Not mature clustering/replication tech for public cloud
- Lacking tools & resources to automate
Stateless CEP

- Local Aggregation using stateless CEP using Drools
- High volume app does not affect other apps
- Some Database overhead
- Some “action moved” to SDK such as new session detection
- Amazon Beanstalk Auto Scaling
rule "NetworkMetricsCarrierRule"
    no-loop true
when
    $nm : ClientNetworkMetrics( $networkCarrier : networkCarrier ) from entry-point "clientNetworkMetricsStream"
    not (exists NetworkMetricsCarrier(type == $networkCarrier))
then
    insert( new NetworkMetricsCarrier($networkCarrier));
end
rule "CompactNetworkMetricCarrierGenerator"
when

$networkCarrier : NetworkMetricsCarrier()
$nm : ClientNetworkMetrics(endMinute != null, networkCarrier == $networkCarrier.type) from entry-point "clientNetworkMetricsStream"
not (exists (CompactNetworkMetrics(appId == $nm.appId, endMinute == $nm.endMinute,
    networkCarrier == $networkCarrier.type) from entry-point "compactNetworkMetricsStream_ByCarrier"))
then
    log.info("Detected new ClientNetworkMetrics to trigger new local aggregation")
Drools Code Snippet for Local Aggregation

$metric : CompactNetworkMetrics() from entry-point "compactNetworkMetricsStream_ByCarrier"

$sumNumSamples : Number()
from accumulate (ClientNetworkMetrics(appId == $metric.appId,
endMinute == $metric.endMinute, networkCarrier == $metric.networkCarrier,
$numSamples : numSamples )
from entry-point "clientNetworkMetricsStream" ,
rule "SimpleAlarm"
   salience -600
   no-loop true
when
   AppErrorsAppIds($appId : type);
   $assertLogs : ArrayList(size > 0) from collect (ClientLog(logLevel == "A", appld == $appId) from entry-point "clientLogStream")
then

   log.info("Detected critical errors for " + $appId);
   alarmService.sendCriticalErrors($assertLogs, $appId);
end
Denormalized Data

- No Joins
- Optimized for Insert
- Not really schema less but sort of helps towards that
Minimal Indexing

- Optimized for Insert
Data Pruning
Problems with Current Architecture

• Near real time - about 2 min delay.
• New aggregation criteria requires new “everything”
• Cloud based only.
• Wasted cycles in polling
• Data Pruning is expensive.
• n entries in DB for local aggregation where n is number of injector machines.
FAQs

• Why no NO-SQL databases?
Lessons Learned

- Tools/Services used properly can solve “new” and “unknown” problem space
- Don’t boil the ocean.
- Cloud is awesome !!
Lessons Learned

- Developers have difficulty making up their minds and like stuff free. 😊
- Early to market - may be.
- Multi-tenancy is hydra with new tentacles popping up all the time.
- Big data solution does not have to be based on NoSQL, Hadoop and other buzz words.
Give it a shot !!

http://apigee.com/about/mobile-analytics

It’s Free.

@prabhatjha
Future Changes

• Full Text Search using Elastic Search or Solr or LogStash (Still evaluating)
• Use No-SQL likely Cassandra for Raw Data

@prabhatjha
Personal Appeal!!

http://www.eejot.org
Thanks !!

@prabhatjha