Working with Big Data, Large Memory, and Low Latency

Eric Theriault

Apptio

Senior Software Engineer

February 21, 2013

About me:

Apptio: core calculation and data management

Amazon: S3

AutoDesk Media and Entertainment Division: Networking and Storage

About Apptio:

Work with big data

-use metrics

-use models

-create a report

Sharing data across multiple nodes is a problem

Each server maps the data slightly differently

Study your data

Know your hardware

Dave Broka

About:

Apptio: Tech lead, next-gen IT Planning application

Microsoft: Team Foundation Server

Trilogy: E-commerce and financial services

HP/Agilent: EMC compatibility measurement

Object Relational Mapping

-Mapping an object to a database

-flat:csv, xml

-object bases: versant, cache, objectDB

-NoSQL databases: Mongo, CouchDB

-Relational databases (tables)

-oracle IBM DB2 MS SQL Server, MySQL, PostgresQL

-Relational databases (object types)

Why ORM?

-OO is a proven s/w engineering practice

-relational is a proven mathematical principle

-built in features (isDirty flag = object changed) (conflict checking)

-cross-platform support

-schema modularization/integration

-not a panacea

OO vs relational:

-concept:

=objects vs rows

=classes vs tables

=references and collections vs foreign keys

-datatypes and operations:

=pretty similar to each other

-location

=in memory vs remote

-atomicity

=transactions vs synchronizations

Object-relational

-granularity

-inheritance

-identity

-references and referential integrity

-data navigation

-DBA vs. ORM culture

How we chose an ORM:

-20+ for each of Java and .Net

-Mix of open-source and commercial

-Standards evolution

Finally chose Ebean because of easy save queries and explicit queries

-uses JPA

-simpler API

-lazy loading

-but it has some bugs

-can’t model