Rails and NoSQL

Great combination for building scalable apps for data analysis (big and small) and visualization

- Tom Zeng
- Director of Engineering, Intridea
- tom@intridea.com
- @tomzeng

Agenda

- Intro (Intridea – Rails, Mobile, Big Data)
- Rails and NoSQL for scalable apps
- Development Environment
- Deployment Options
- Some Sample Usage Scenarios
- Demo and Q/A

Ruby on Rails

- Productive framework for web apps and services
- Popular among startups and now enterprises
- Programmer happiness leads to productivity and faster time-to-market
- Mature(since 2005), “there’s a gem for that”

Client Side MVC Stack

- HTML5/CSS3(Sass) / Twitter Bootstrap / Responsive Design
- CoffeeScript / JavaScript
- Backbone.js / Ember.js
- Template Engines (Mustache, eco, ejs)
- D3, Highcharts, TileMill/MapBox
- Jasmine (CS/JS testing)

Server Side Stack

- Ruby/JRuby on Rails – web apps & services
- Behavior Driven Development(BDD): RSpec + Cucumber
- API Services for Client Side MVC (responsive)
- Server Side MVC (search friendly)

- NoSQL(Mongo) / SQL(Postgres) backend

NoSQL

- Key-value Stores (Redis, Riak)
- Column Family Stores (Cassandra, HBase)
- Document Databases (MongoDB, CouchDB)
- Graph Databases (Neo4J)

Rails + NoSQL

- Will focus on MongoDB and Riak
- MongoDB and Riak are easy to use
- Works well with Rails
- Built in Map/Reduce capability
- Integrate with Hadoop

MongoDB

- Document oriented, JSON format
- Very high read and write throughput
- Rich query capabilities (aggregation framework), flexible indexes
- Scale with auto-sharded replica sets
- Map/Reduce in JavaScript

Riak

- Key value store (Dynamo inspired)
- Web native data types (json, html, image)
- Every node is writable (master)
- Very easy to scale (just add new nodes)
- Riak CS – Amazon S3 compatible storage
- Map/Reduce in Erlang and JavaScript

Distributed Processing

- When built-in Map/Reduce is not enough
- MongoDB Hadoop connector
- Riak + Hadoop
- Streaming data to/from HDFS or S3
- HBase/HDFS (better speed) + Hadoop M/R
- S3 (lower cost) + AWS Elastic Map/Reduce

Distributed Processing

- Map/Reduce (the foundation, mappers/reducers in Java, Python, Ruby)
- Pig (Python, JRuby, Java UDFs)
- Hive (SQL like queries on HDFS/HBase)
- Impala (Hive like but much faster)
- JRuby + JDBC => Hive/Impala + HBase/HDFS

Distributed Processing

- Cloudera
- HortonWorks
- MapR
- Amazon Elastic Map/Reduce
- IBM, Oracle, EMC, Microsoft

Sample Usage Scenarios

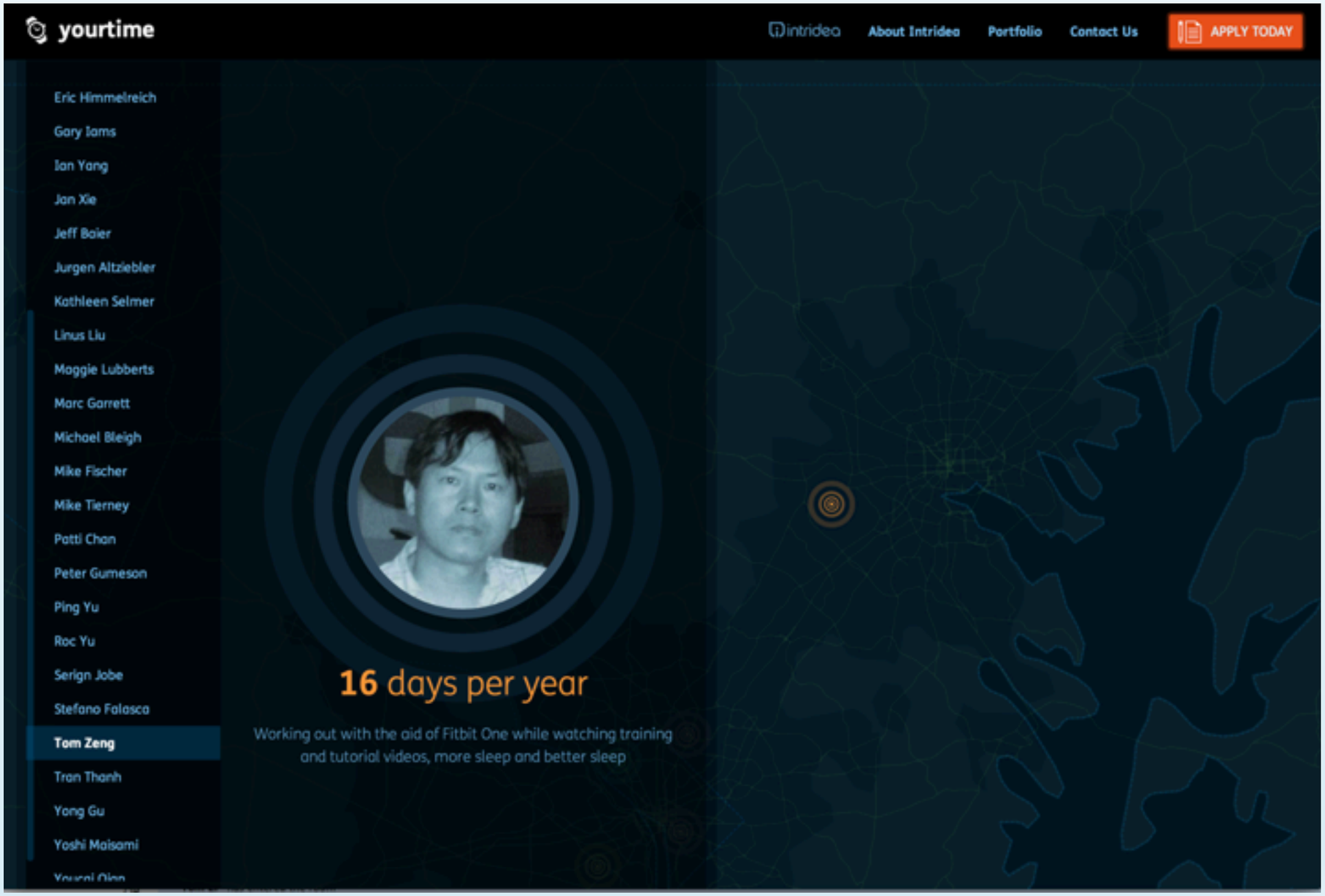
- Rails + Mongo + Oracle
- Rails/Sinatra + Riak
- Rails + Mongo + Pig/Hive on EMR/S3
- Rails + JRuby + Pig/Hive/Impala + HBase/HDFS

Demo

- Sample apps
- Rails + Mongo, Rails + Riak
- HortonWorks Sandbox
- Pig and Hive, JRuby/Python/JS UDFs
- Cloudera Demo VM
- Impala vs Hive

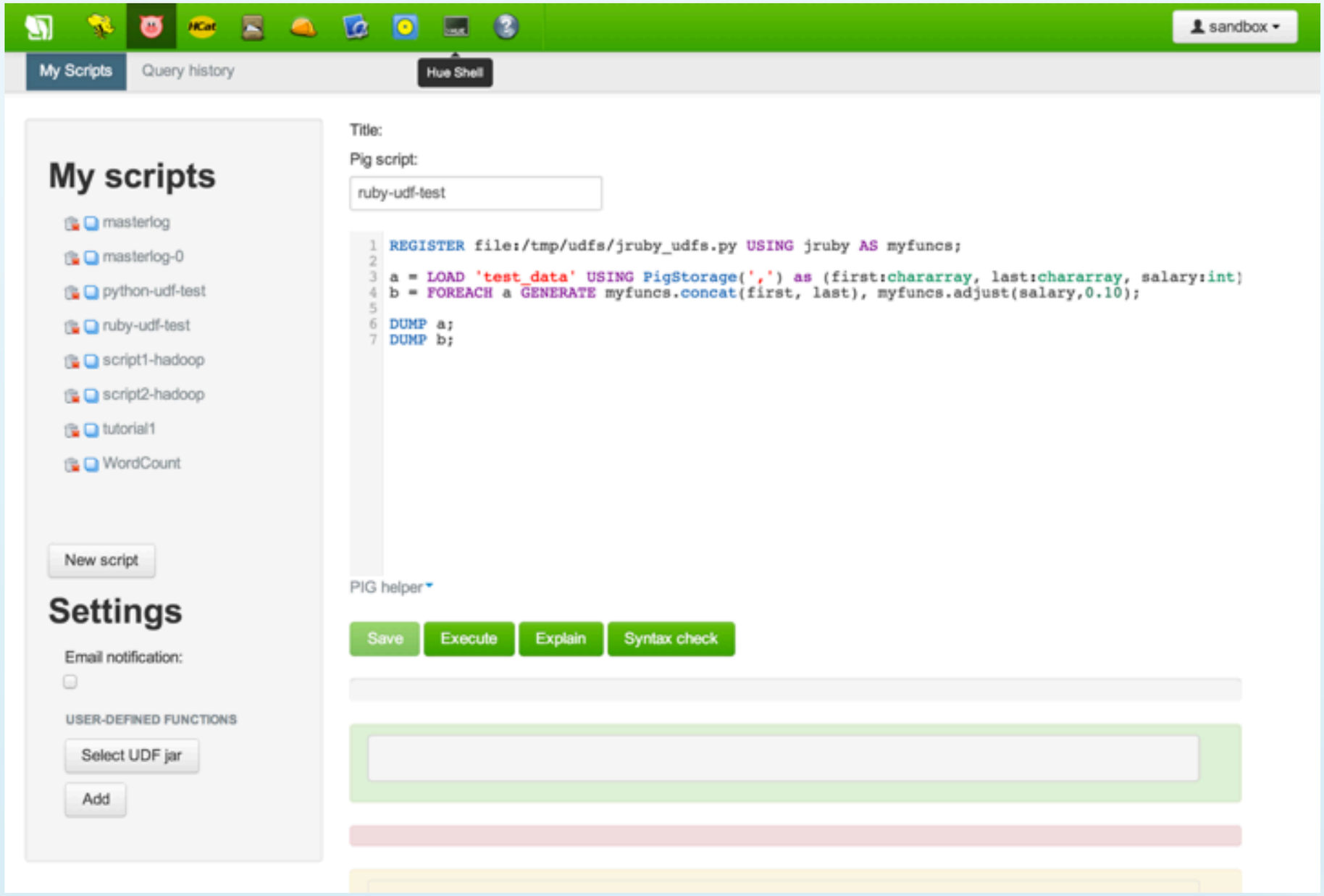
yourtime.intridea.com

- Client side MVC app using Backbone/CoffeeScript/Rails/TileMill/MapBox



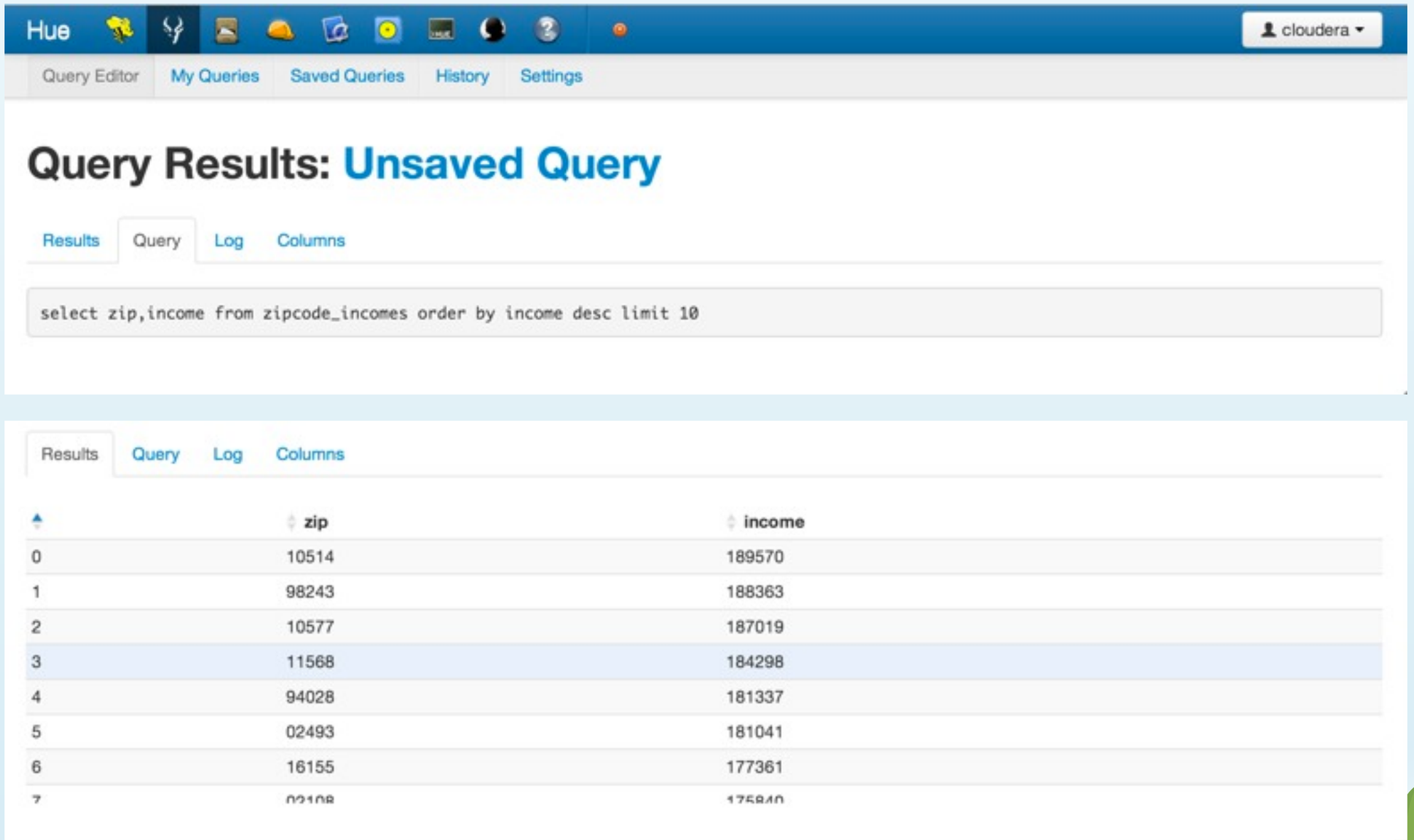
HortonWorks Sandbox

- JRuby User Define Function for Pig



Cloudera Impala

- Hive like but much faster

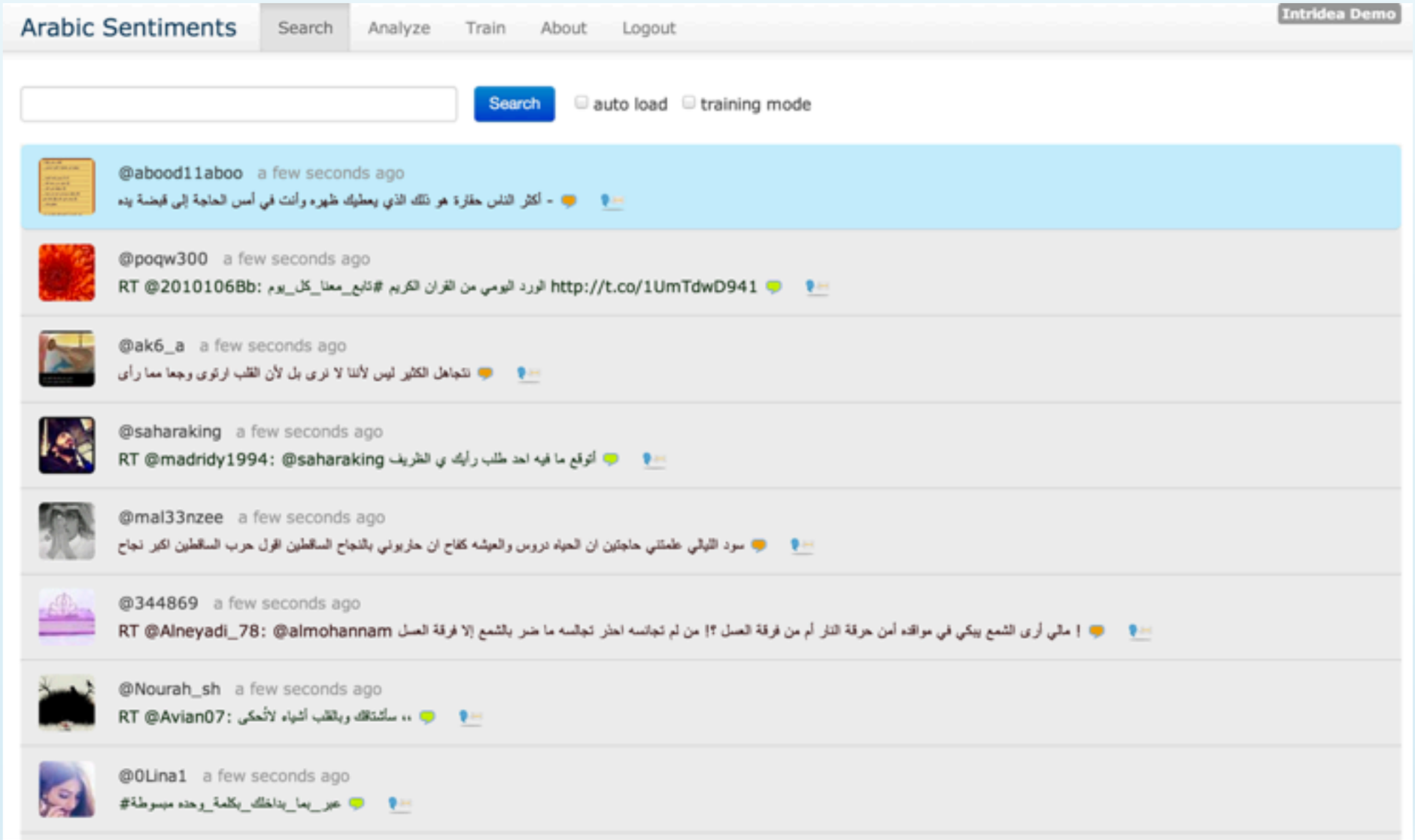


	zip	income
0	10514	189570
1	98243	188363
2	10577	187019
3	11568	184298
4	94028	181337
5	02493	181041
6	16155	177361
7	02108	175840

Arabic Sentiments – Client Side MVC

- Backbone.js, CoffeeScript, Twitter Bootstrap, Rails, MongoDB –

  R&D project put together quickly by backend engineer with not much UI experience



Resources – Rails

- http://rubyonrails.org/
- http://railscasts.com/
- http://ruby.railstutorial.org/
- http://www.codeschool.com/paths/ruby
- http://pragprog.com/categories/ruby_and_rails

Resources – NoSQL

- https://en.wikipedia.org/wiki/NoSQL
- http://www.mongodb.org/
- http://mongoid.org/en/mongoid/index.html
- http://basho.com/riak/
- http://basho.github.io/ripple
- https://github.com/braintree/curator

Resources – BigData

- http://hadoop.apache.org/
- http://pig.apache.org/
- http://hive.apache.org/
- http://www.cloudera.com/content/cloudera/en/home.html
- http://hortonworks.com/
- http://www.mapr.com/

Thank You Merci

Gracias Danke

ありがとう 谢谢