TickSmith Corp.

Flash – Cloudera Impala on Docker

Big Data Montreal #29
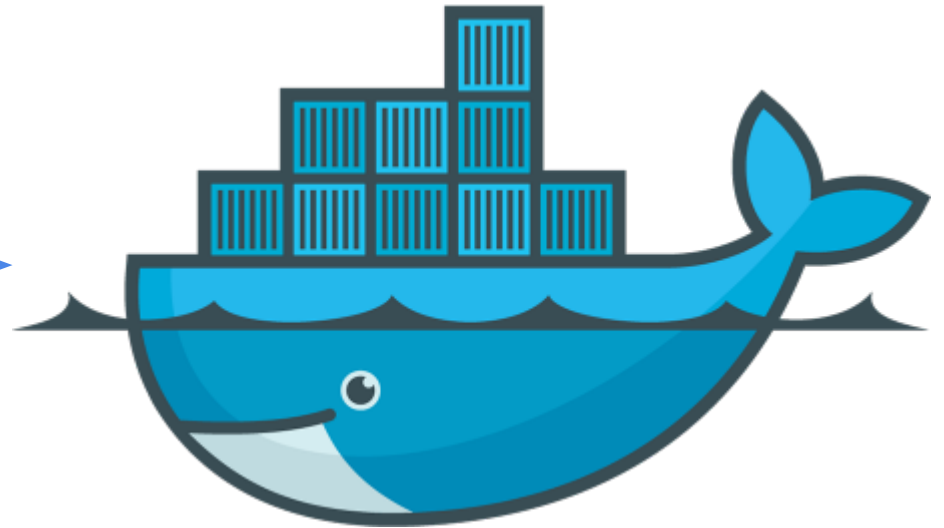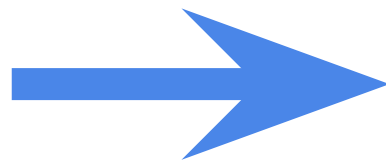October 7th, 2014

Tony Bussieres, SVP Dev & co-founder   **tony.bussieres@ticksmith.com**

Running Impala in a Docker container

http://impala.io

http://docker.com

What is Docker

- A portable and lightweight application container.

- It's not a VM

- Open source, works on most modern OS.

- Images are easy to create and maintain with Dockerfiles

- Images are easy to share with developers / sysadmin using the Docker Hub

- More info on Docker at **https://docs.docker.com/**

- **It's not a VM! (It shares the same kernel & resources as the host)**

TickSmith Corp.

Motivation

- It makes it easy to use Impala on your desktop

- Test queries without a connection to a cluster

- More lightweight than using Cloudera's QuickStart VM

- Wanted to prove that Impala could run in Docker

- Docker is cool ! :)

TickSmith Corp.

Quick Demo

Docker image **codingtony/impala** is based on CDH5.
It is configured as a Single-Node Hadoop cluster

It runs a namenode, a datanode, the impala-state-store, the impala-catalog and the
impala-server

The Docker Image is available on Docker Hub :
docker pull codingtony/impala

This presentation is available :
https://github.com/codingtony/docker-impala-bdm29

The recipe for building **codingtony/impala** is available here :
https://github.com/codingtony/docker-impala

Any Questions ???