



Oliver Meyn

Hadoop and Big Data Consultant

26 Maughan Crescent, Toronto, Canada, M4L 3E5
<https://elephant.tech> - oliver@elephant.tech - (416) 524-2240

Summary

I have been working primarily in Hadoop and Big Data since 2009, with Java and SQL development extending back to 1999. I have learned some hard-won lessons in the complexity of distributed system development and management, and I am happy to help others benefit from my experience. I am based in Toronto, Canada.

Skills

Big Data (HDFS, HBase, Hive, Spark, NiFi, Kafka)	★★★★★★★★★★
Java	★★★★★★★★★★
SQL (including Oracle, Mysql, PostgreSQL)	★★★★★★★★★★
Leading (Team Lead and Certified ScrumMaster)	★★★★★★★★★★
DevOps (Linux, scripting, log analysis)	★★★★★★★★★★

Experience

RBC I&TS, ACE PROGRAM – Toronto, Canada April 2017 - Present

Hadoop and Big Data Consultant

Contributing to the architecture, design, and some development towards ingesting data from several large legacy systems into the Data Lake, transforming with Spark, NiFi, and pure Java, and exposing through APIs and a web GUI built on microservices.

Established a 'Big Data Platform' as guiding architecture for data applications, and introduced Kafka with Avro to the program, establishing near real-time flows as a key component of the Platform.

Involved in a PoC and eventual integration of Palantir Foundry.

Technology used: Spark, NiFi, Kafka, Yarn, Hive, HDFS, krb5, Palantir Foundry, HDP 2.4/2.6, Confluent 4/5, RedHat7

T4G – Toronto, Canada Feb 2016 - April 2017

Hadoop and Big Data Consultant

Telus

Converting a batch-based system for TV customer experience modeling to near real-time. Hive over HDFS for static data, NiFi, Kafka, and Spark Streaming for near real-time, SparkML for customer modeling. Presented work at DataWorks Summit 2017 in San Jose: <https://www.youtube.com/watch?v=3UrH4ZU5aTI>

Technology used: Spark, NiFi, Kafka, Yarn, Hive, HDFS, krb5, HDF 2.0, HDP 2.2/2.4, RedHat6

CPP Investment Board

HBase optimization for time series data, Spark over HBase for processing, Hive over HDFS for static data.

Technology used: HBase, Spark, Yarn, Hive, Tez, HDFS, HDP 2.3, RedHat 7

EYERETURN MARKETING – Toronto, Canada

Jan 2016

Hadoop and Big Data Consultant

Converted a large number of Oozie and Pig scripts from MRv1 to YARN. Helped with performance and best practice tuning and configuring of Cloudera clusters.

Technology used: Pig, Oozie, HDFS, MapReduce, YARN, HBase, krb5, CDH 5.5, Ubuntu 12

GBIF – Copenhagen, Denmark

2010 – 2015

Senior Software Developer & Scrum Master

Worked as a developer and Scrum Master for a team of 4-8 developers to support the transition of gbif.org from a batch-oriented, MySQL based system to a real-time processing system based on Hadoop. The system went live in 2013 and has been running on two in-house Cloudera (CDH) clusters. I was responsible for cluster installation, upgrades, maintenance, performance tuning, and was part of the team that designed the architecture of the overall system. I wrote the portion of the system that speaks directly to HBase, including architecting the key and column structure, region sizes, etc. Contributed to all aspects of the Hadoop development, including Hive UDFs, Sqoop-ing in and out of the cluster, Oozie workflows, custom MapReduce jobs in both versions 1 and 2 (yarn), Spark for fast analytics, as well as interactions with Zookeeper and Solr. Served as the primary DevOps liaison between System Administrators and the development team.

Additionally worked to build RESTful, JSON-based webservice in Java to deliver data from both CDH and traditional SQL databases (MySQL, PostgreSQL, PostGIS). These were managed in a continuous build environment using Jenkins, Maven, and Nexus. Helped build the analytics portion of the site in a combination of R and Hadoop.

Technology used: Hadoop (HDFS, Hive, Zookeeper, HBase, Mapreduce, Yarn, Spark, Sqoop, Oozie, SolrCloud), CDH 3/4/5, Java 6/7, R, RabbitMQ, Maven, Jenkins, Nexus, Varnish, Puppet, Ansible, Ganglia, Elasticsearch, Kibana, Git, IntelliJ, JIRA

ZEROFOOTPRINT – Toronto, Canada

2008 – 2010

Software Architect & Team Lead

Responsible for designing and developing services, messaging infrastructure, web and api clients, within an SOA, for a suite of enterprise environmental products. Notable among them are the Velo enterprise carbon management package, and the TalkingPlug energy management hardware devices and software.

Technology used: Java 5/6, SOA, ESB, Mule, JMS, ActiveMQ, Hadoop/HBase, SaaS, Spring MVC, Hibernate, Maven, Eclipse, Web Services, Ruby, JRuby, Rails

TSOT – Toronto, Canada 2007 – 2008

Technical Lead & Senior Developer

Maintained and augmented a private social networking application in the fraternity, sorority and university market, written in Ruby on Rails. Designed and implemented a staged release process around a Subversion server. Technical Lead for a team of 5 comprising front-end, back-end and QA. Prioritized and organized work across the team to meet business goals.

Technology used: Ruby on Rails (1.2 & 2), MySQL, Mongrel, Subversion, Apache, TextMate, OS X, Linux

PENSON FINANCIAL SERVICES – Toronto, Canada 2007

Senior Java Developer

Built a Trade Order Management System for a fixed income trading platform. Built as a service in an SOA environment with connections to Bloomberg, both incoming trades via BTS, and outgoing trades via their Consolidated Message Feed (CMF).

Technology used: Java 5, J2EE, SOA, ESB, Mule, Websphere MQ/IBM MQSeries, JBoss, Spring, Hibernate, Maven, Eclipse, Subversion, Drools, XML, REST

Education

MCMASTER UNIVERSITY – Hamilton, Canada 1999
M.Eng Civil Engineering (Sustainability)

MCMASTER UNIVERSITY – Hamilton, Canada 1997
B.Eng Civil Engineering
B.Sc Mathematics (Computer Science)

Certifications

HORTONWORKS 2016
HDP Certified Administrator (HDPCA)

CLOUDERA 2015
Cloudera Certified Administrator for Apache Hadoop (CCA)

CLOUDERA 2015
Cloudera Certified Specialist in Apache HBase (CCSHB)

SCRUM ALLIANCE 2014
Certified ScrumMaster
