Synopsis

ETL Pipeline for Alerts into Hive

SUBMITTED
BY

Vatsal Saraswat   130905226   vatsal12117@gmail.com

Under the Guidance of:

Internal Guide

Mr. Siva Selvan
Assistant Professor,
Department of Computer Science and Engg.
Manipal Institute of Technology,
Manipal University
Manipal-576104

External Guide

Mr. Pratham Khanna
Associate,
APG-CP-CSO-System Ops,
BlackRock Services India Pvt. Ltd

DETAILS OF THE ORGANIZATION:

BlackRock Services India Pvt. Ltd.,
Building #14, 14th Floor, DLF Building, Tower C, DLF Cyber City Phase-III,
Gurugram, Haryana, 122002, India
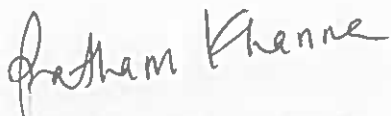
NAME OF EXTERNAL GUIDE WITH CONTACT DETAILS:

Mr. Pratham Khanna,
Associate,
APG-CP-CSO-System Operations
BlackRock Services India Pvt. Ltd.
Phone No: 919818111220
Email ID: pratham.khanna@blackrock.com

DATE OF COMMENCEMENT OF THE PROJECT:
25th February, 2017

PROJECT DURATION:
24 Weeks

Signature of External Guide

BLACKROCK®

CONFIDENTIAL   September 28, 2016

Vatsal Saraswat
Manipal Institute of Technology
Manipal

Dear Vatsal

We are pleased to offer you internship with BlackRock Services India Pvt. Ltd ("the Company")
as an Intern from January 16, 2017 until July 14, 2017 This letter, together with the enclosed
written statement of terms and conditions, constitutes our offer to you

This offer is confidential and should not be disclosed to others. It will remain open for ten
working days, after which time it will be deemed to have been withdrawn if you have not
accepted by then.

Please signify your acceptance of this offer by signing and return one copy of the written
statement of terms and conditions of internship, along with the documents listed in the
attachment using the BlackRock envelope provided.

If you have any queries regarding this letter or the enclosures or if you wish to receive copies of
the various policies and schemes referred to in the written statement of terms and conditions,
please contact email contactHR@blackrock.com.

We are very excited about you joining the Company as an intern and hope you would benefit
substantially from your learning at BlackRock.

Yours sincerely

_28/9/2016_

Ramneesh Mohan
Director
Human Resources
For and on behalf of
BlackRock Services India Pvt. Ltd

1. Introduction:

1.1 Topic

An Extract, Transform and Load pipeline is commonly used to integrate data from multiple applications by transforming it for storing in a proper format or structure for the purposes of querying and analysis. The Service Operations team will use this data to analyse the various alerts and use it for deriving patterns in the data.

1.2 Organization

BLACKROCK®

BlackRock, Inc. is an American global investment management corporation based in New York City. Founded in 1988, initially as a risk management and fixed income institutional asset manager, BlackRock is the world's largest asset manager with $5.1 trillion in assets under management. BlackRock operates globally with 22 investment centers, 70 offices in 30 countries and clients in 100 countries.

In 2000, BlackRock launched BlackRock Solutions, the risk management division of BlackRock, Inc. It is the in-house investment analytics and process engineering department for BlackRock which works with their portfolio management teams, risk and quantitative analysis, business operations and every other part of the firm that touches the investment process. One major department is the APG (Aladdin Product Group). Aladdin (Asset Liability and Debt and Derivative Investment Network) is the platform on which BlackRock runs its complete business. The team is organized to provide scalable investment and client systems which are built on a resilient technical infrastructure. The APG group mainly focusses on providing the best technology platform for investment management by developing and building BlackRock's core platform and supporting all the various units of BlackRock.

I am currently a part of the CSO-System Operations team. This team is responsible for the day-to-day operations of core systems and focuses on ensuring a fully stable, scalable and cost efficient infrastructure across all the systems supported by the Aladdin Product Group. The group manages business-critical batch workloads, real-time processing and data transfer services for BlackRock, BlackRock subsidiaries and BlackRock Solutions clients.

3. Objective:

- Imbibe the BlackRock Culture and demonstrate understanding of CSO and CSO-System Operations

- Complete AJDC and gain thorough understanding of BlackRock's coding standards

- Develop basic understanding of Hadoop Infrastructure in Understand Aladdin platform that is used by the organization

- Build an ETL pipeline to load alert data into Hive tables in Hadoop

- Develop high level understanding of various functions of System Operations with focus on BAU Operations

4. Methodology:

The project is carried out in the following parts:

➢ **Understanding Big Data Essentials**

- o Getting familiar with the core Hadoop stack components including the HDFS file system, Apache Hive, Apache Sqoop etc.
- o Understand the architectural components and programming models used for scalable big data analysis.

➢ **Understand Current Database Architecture of Alert Mechanism**

- o Make use of Database Explorer to understand the relationships between different entities in the database related to alerts.
- o Understand the different types of alerts and analyse the various attributes associated with them.

➢ **Propose and Create a New Database Architecture in Hive**

- o Create an efficient database architecture while keeping in mind the conventions, keys, integrity checks, indexes etc.

- o Test the pipeline flow from the Database into Hive.

➢ **Automate Data Upload from Current Database Tables to New Hive Tables**

- o Load the data into the proposed architecture in Hive for analysis.
- o Use incremental import approach to update the tables.
- o Visualize the data in the form of graphs and charts to derive frequents patterns or repeated alerts.

5. Project Schedule:

January 2017	O Intern Onboarding programme with 3-day HR induction. O Aladdin Certification training overview. O Overview of compliance and ethics at BlackRock.
February 2017	O Completion of Aladdin Java Certifications - 3 levels. O Completion of team specific trainings. O Basic training on prerequisites required for the project.
March 2017	O Develop understanding of the working of Hadoop Core stack. O Develop basic understanding of the functional structure, operating model and regional presence of CSO as a team. O Creation of Database Architecture in Hive. O Initiation of the implementation.
April 2017	O Familiarizing with Apache Sqoop and leveraging it to export data to HDFS O Initiation of development of the ETL Pipeline.
May 2017	O Implementation of the incremental update pipeline. O Testing of the newly developed pipeline.
June 2017	O Documentation of the end-term report. O Review of the project.

6. References:

- BlackRock Internal Documents and Documentation
- Tutorials Point
- HortonWorks
- Coursera

Project Details

Student Details

Student Name	Vatsal Saraswat		
Register Number	130905226	Branch / Roll No	CSE D - 20
Email Address	vatsal12117@gmail.com	Phone No (M)	7259646439

Project Details

Project Title	ETL Pipeline for Alerts into Hive		
Project Duration	24 weeks	Date of reporting	23/01/2017

Organization Details

Organization Name	BlackRock Inc.		
Full postal address with pin code, telephone number, Fax and email	DLF Building No. 14, 14th Floor, Tower C DLF Cyber City Phase-III, Gurugram122002, Haryana, India Phone: +91 124 6780000 Email: contact.us@blackrock.com		
Website address	https://www.blackrock.com		

External Guide Details

Supervisor Name	Pratham Khanna		
Designation	Associate		
Full contact address with pin code	Building #14, 15th Floor, DLF Building, Tower C, DLF Cyber City Phase-III, Gurgaon, Haryana 122002, India		
Email address	pratham.khanna@blackrock.com	Phone No (M)	919818111220

Internal Guide Details	
Faculty Name	Siva Selvan
Full contact address with pin code	Dept. of Computer Science & Eng., Manipal Institute of Technology, Manipal University, Manipal. 576104
Email address	siva.selvan@manipal.edu