Hemant Kumar

Senior Associate Level 2 at Sapient

Summary

Started my career with Sapient Consulting on July 2010. Currently working as a senior big data developer, having strong technical experience on Hadoop ecosystem. I have more than 6.5 years of experience out of which last 3.5 years of experience in Big Data space plus BI/DW and initial 3 years of experience on Business Intelligence and Data Warehousing solution. I am responsible for successful Big Data /Hadoop implementation and PoCs/Pilot project for internal projects as well as customers. Majority of the projects are providing data warehousing solutions using big data platform. Currently i am working as a Technical Lead for one of the Big Data project where we are processing around 200 GB of data daily using Apache Spark on MapR cluster.

Areas of expertise include:
~ Hadoop Technlogy: HDFS, Maprfs, MapReduce, Hive, Pig, Sqoop and Oozie
~ Relational Databases: Oracle, and MS SQL Server
~ NoSQL Databases: HBase
~ Graph Databases: Neo4j
~ In-memory/MPP/Search: Impala and Spark(Python API)
~ ETL/Data Integration: Talend Big Data and Informatica
~ Languages: Python,SQL, PL/SQL and Core Java
~ Operating Systems: Windows, UNIX, and Linux
~ Other Tools: Toad, SQL Developer, Putty, Mputty and Winscp
~ Hadoop Distributions: Cloudera and MapR

Experience

Senior Associate Level 2 at Sapient
August 2016 - Present (8 months)

Senior Associate Technology Level 1 at Sapient
August 2014 - Present (2 years 8 months)

Big Data Developer/Technical Lead for Pulse Big data project for a Leading US Home Automation for Security Systems. This project caters to build a big data platform to process more than 100 GB of log data generated on a daily basis from pulse application and populate it to a dimension/fact table in big data environment. This will be used to analyze data to mine for valuable signals of customer behavior and other analysis.
~ This system is capable of storing more than 100 TB of data.
~ Maintain at-least 12 months of data snapshot for data mining/analysis.

~ Capable of processing large volume of data within reasonable time.
~ Integrate with tableau for Business Intelligence and Reporting.
~ Should be able to process ad-hoc query on entire data set.

Responsibility:
~ Create framework for data cleansing and profiling which can be used across all the tables.
~ Develop reusable function using python library for data standardization.
~ Develop SQOOP jobs to ingest data from oracle (pulse application) into Maprfs
~ Develop Spark jobs for processing source data (present in maprfs) and populating it into Dimension/Fact in form of parquet files.
~ Performance Optimization of Spark Jobs.

Environment Details:
~ MapR: Apache Hadoop Distribution
~ Apache Spark 1.2.1
~ Apache Hive
~ Drill/Impala (query engine)
~ Tableau

Associate Technology Level 2 at Sapient
October 2012 - July 2014 (1 year 10 months)

Work with internal team to build the capabilities in big data and build a framework for processing social media data and integrate it with the existing/new e-commerce clients. This framework is being developed in big data platform to process any amount or type of data whether its structured/semi-structured or unstructured. This framework is listening to social media API's (Facebook/Twitter/Bestbuy) and extracts relevant data and process it for Business Intelligence & Reporting. This framework will integrate e-commerce data with social media data and provide some basic reporting along with Text Analytics, Social Media Analytics and Web Analytics.

Responsibility:
~ Data Modeling for Social Media and ecommerce.
~ Track Lead for Social Media.
~ Explore Facebook and Twitter API’s
~ Understanding the attributes and data access policy for Social Media.
~ Design the Hbase data model for storing Twitter and Facebook Data.
~ Extract the data from Hbase and load it to hive staging using Talend Big Data.
~ Cleanse the data and format it in dimensional modeling using Talend Big Data.
~ Load the data into hive/Impala dimension and fact tables using Talend Big Data

Environment Details:
~ Cloudera Hadoop Distribution
~ Apache Hive
~ Talend Open Studio (Big Data Edition)
~ Hbase (NoSQL)
~ QlikView

Associate Technology Level 1 at Sapient

January 2011 - August 2012 (1 year 8 months)

ETL Developer to build data warehousing solution for on of the US Telecommunication Major. The project catered to data warehousing and reporting requirement for Regulatory bodies and analysis for managers and business owners. It involved dimension and fact table on the basis of which BO reports and excel macro reports are generated. It includes monitoring and fixing production bugs, Support and enhancements of existing PL/ SQL procedures, BO Reports creating new Informatica mappings, scheduling of workflows for loading new data from different source systems, handling of ad-hoc requests by the users.

Responsibility:
~ Designed and developed data warehousing solution for a new Weekly reporting platform.
~ Requirement gathering, Analysis and estimations for weekly reporting platform.
~ Analysis of the current system and draw correct inferences, in keeping the objectives of the analysis.
~ Involved in the complete life cycle of ETL software development: analysis, design, documentation, reporting and testing.
~ Develop ETL logic for weekly reporting using technologies like Informatica PowerCenter, Oracle PL/SQL, UNIX and BO.
~ Support member of the CR and Ad-hoc track.
~ Optimization of ETL process and PL/SQL code.

Environment Details:
~ Informatica Power Center
~ Oracle 10 G
~ Business Object (Reporting)
~ SAS

Associate Trainee at Sapient

July 2010 - December 2010 (6 months)

Junior ETL developer for one of the US e-commerce major. Working as a shadow resource in migration of an e-commerce platform from its legacy storage system to Oracle. I was working as a Informatica developer to develop mapping/session/workflows to read the source data (legacy system) and populate it to the new target tables after cleansing and standardization. With in a very short span of time i was able to get good understanding of the requirement and started contributing independently.

Responsibility:
~ Understand the requirement
~ Develop mapping/session/workflow
~ Testing of the mapping/session/workflow
~ Write Unit Test cases for the developed mapping/session/workflow

Skills & Expertise

Big Data
Hadoop
Apache Spark
Scala
Python
Hive
PL/SQL
Sqoop
Apache Pig
Data Warehousing
Informatica
Oozie
MapReduce
Data Analytics
apache drill
Talend Open Studio
Oracle PL/SQL Development
Unix Operating Systems
Data Integration
Microsoft SQL Server
ETL Tools
Basics of Unix
Microsoft Office
Microsoft Excel
Microsoft Word
Unix
SQL
Oracle
Core Java
ETL
Java
Impala
Extract, Transform, Load (ETL)

Education

Vellore Institute of Technology

Master of Computer Applications (MCA), Computer Software Engineering, 2007 - 2010

Ranchi University

Bachelor of Science (B.Sc.), Mathematics, 2003 - 2006

Languages

English
Hindi

Honors and Awards

Domain Connect - Idea Innovation Fest 2012

Sapient

December 2012

Winner of the Idea Innovation fest 2012.

All the sapient employees are requested to submit their innovative ideas to solve some real time problems or new innovation. Out of the total submission only 30 ideas goes to the next level for developing the prototype. Best idea is being chosen by the team of CTO/Executives/VP's based on the prototype developed/presented.

Publications

Data Processing Performance Optimization Using Spark

Authors: Hemant Kumar

Certifications

Developer Certification for Apache Spark

O'Reilly Media License 1.x-0470 March 2016

CP100A: Google Cloud Platform Fundamentals

ROI Training License 10184987 April 2016

Hemant Kumar

Senior Associate Level 2 at Sapient

Contact Hemant on LinkedIn