

# Dhananjay Kumar

## Summary

**Working as Lead Data Scientist for BFSI and SCM domain.**

**Data Scientist/Big Data Consultant/ DWH/Business Intelligence Professional with 10+ years of experience.**

**Nearly 6 years of experience in People Management activities and hands on Hadoop/Big Data tools.**

**Top analytical & problem-solving skills combined with excellent communication and interpersonal skills and good Team Player.**

**Adept at client interaction and leading teams while embracing accountability and delivering high quality results**

**5 years of onsite experience and working at client location.**

## Phone

+91 – 70195 73007

## Email

dkumar.soft@gmail.com

## Present Address

**#224, 5<sup>th</sup> Block, 4<sup>th</sup> Cross,  
KHB Colony, Koramangala  
Bengaluru - 560095**

## Experience Summary

- Certified Data Scientist from IIML (Indian Institute of Management Lucknow).
- Experience in Statistical modeling, Predictive modeling and Prescriptive modeling and NLP (Natural Language Processing)
- Implemented end to end ML platform for various clients across the globe on advance analytics.
- Experience in Data Science, Machine Learning, Google Cloud Platform (GCP), Big Query, Data lab, AWS.
- Experience in Big Data, Hadoop, Spark, Python, Hive, Impala, Sqoop, Flume, YARN, Pig.
- **Deep** knowledge of Supervised Learning, Unsupervised Learning and Reinforcement Learning.
- Experience in implementing Machine Learning algorithms in Spark MLlib and Google AutoML, CDAP, OpenRefine, Miflow, H2O AutoML
- Experience in Preparing ETL/DWH/BI Architect plan in coordination with project stack holders.
- Experience in development and designing of ETL methodology for supporting data.
- Transformations using Informatica Power Center, Power Exchange Highly skilled in Designing, developing and deploying Data Warehouses / Data Marts.
- Strong hands on experience in designing developing data model, Conceptual Data Model, Logical Data Model and Physical Data Model using Erwin tool.
- Very good knowledge of Database designing Transactional Database (OLTP) and Dimensional Database (OLAP).
- Strong analytical skills used for requirement analysis and root cause analysis.
- Possess a good understanding of the various business processes in healthcare.
- Effective implementation skill of project based on Onsite-Offshore model.
- Sound conceptual knowledge about ERP, ERP application and their end to end implementations.

## People Management Activities:

- Team Management including work distribution and tracking.
- Performance evaluation and providing feedback to other managers for appraisals.
- Approving time sheets for the team members and managing vacation tracker.
- Active participation in recruitment for the organization across various projects.
- Raising request for charge codes for projects.

## – Key Projects

### Atena Analytics

Client	:	Aetna Inc.
Duration	:	Mar. 2020– Till Date
Environment	:	AWS, JupyterHub, Mlflow, H2O AutoML
Role	:	Lead Data Scientist
Team Size	:	8

Aetna Inc. is an American managed health care company that sells traditional and consumer directed health care insurance and related services, such as medical, pharmaceutical, dental, behavioral health, long-term care, and disability plans, primarily through employer-paid (fully or partly) insurance and benefit programs, and through Medicare. Since November 28, 2018, the company has been a subsidiary of CVS Health.

#### Responsibilities:

- Created solution architect for various analytics use cases and evaluated open source tools.
- Worked with pre-sales team and created RFP to design ML solution for different stake holders.
- Responsible to develop ML and Data Science platform for Emergency Room Visitor
- Exploratory Data Analysis (EDA) on CVS data, Univariate and Bivariate plot on feature variables.
- Applied Machine learning methodologies for imbalanced data using SMOTE techniques.
- Predictive analytics for ER visit in next month and Quarter and year.
- Created NLP platform for RSS news feed, Topic Modeling (LDA), Stanza (Sandford NLP Group)

### CoE Advance Analytics

Client	:	Blume Global
Duration	:	Dec. 2019– Mar.2020
Environment	:	Google Cloud
Role	:	Data Scientist
Team Size	:	16

Blume Global is committed to an ethical, purpose-driven supply chain ecosystem and creating a more sustainable environment through providing an open, neutral platform across all trading partners to drive asset optimization, predictive visibility and traceability.

Blume Global committed to an ethical, purpose-driven supply chain that benefits all parties. A purposeful supply chain is optimized, efficient, effective, responsible and employs environmentally sustainable practices. The potential of an orchestrated, collaborative supply chain that addresses environmental and social challenges is profound.

Blume helps optimize moves to reduce carbon emissions, provides visibility and traceability, empowers small businesses and emerging economies, and orchestrates the delivery of goods where they are needed most. This is the Blume Global promise to our customers and partners.

#### Responsibilities:

- Responsible to develop Data Science platform for Shipment services.
- Created Big Data Platform on Google Cloud to Ingest data from heterogenous sources.
- Developed Data pipeline and road map for Advance Analytics team.
- Meeting with different project stakeholder and review progress on regular interval.
- Active interaction and weekly status meeting with team management.

### Vedanta Group

Client	:	Sesa Goa
Duration	:	Jun. 2018 to Dec. 2019
Environment	:	Anaconda, Spyder.
Role	:	Data Scientist
Team Size	:	2

Vedanta Resources Plc is a globally diversified natural resources company. Vedanta extract and process minerals, oil

and gas, engage more than 65,000 employees and contractors, primarily in India, Africa and Australia. Vendata group products are sold worldwide. Vendanta are headquartered in London, United Kingdom. Vedanta Resources Plc is the holding company for Vedanta Limited and Konkola Copper Mines, which in turn have multiple subsidiaries. We are run by a unified board and management.

SESA's Energy Recovery Coke Making Technology is an environment-friendly technology that is characterized by low capital and operating cost, high energy recovery, capable of producing high quality metallurgical coke. This technology is available for license globally in standard modules of 0.3 MTPA with a potential to generate 21 MW of electrical energy. Developed through in-house efforts, SESA holds patents in US, India, Brazil and Europe for this technology. This patent is a reaffirmation of SESA's commitment to development of innovative and cutting Edge solutions.

#### **Responsibilities:**

- Worked as Data scientist to optimize business processes and make machine learning predictive model.
- Created Regression model and non-linear regression model for Coal blend process.
- Created deep learning model to identify CSR (Coal Strength after Reaction) and CRI (Coal Reactive Indexes).
- Meeting with different project stakeholders and providing them update on progress being done.
- Active interaction and daily status meeting with the onshore team to provide progress update and work being done

#### **Analytics Center of Excellence (ACOE)**

Client	:	Bank of Baroda
Duration	:	Oct. 2017 – May 2018
Environment	:	Hortonworks, Data science Workbench, Hive, Hadoop, Spark, Impala
Role	:	Data Scientist
Team Size	:	8

Bank of Baroda is one of the largest public sector bank (PSU) in India with a branch network of over 5500+ branches in India and 100+ branches/offices overseas including branches of our subsidiaries, distributed in 25 countries. Bank of Baroda desires for Selection of knowledge partner for Analytics Center of Excellence (ACOE).

#### **Responsibilities:**

- Worked as Data scientist and lead senior technical consultant for Big Data.
  - Responsible to architect and development enterprise wide big data lake (BDL) which include setup of data environment including data ingestion across internal and external sources. Initially, Establish Federation Between EDW and BDL
  - Created predictive models for NPA, Vintage, Delinquency.
  - Meeting with different project stakeholders and providing them update on progress being done.
  - Revalidation of estimates and delivery dates as per issues being reported.
  - Active interaction and daily status meeting with the onshore team to provide progress update and work being done
  - Working on providing solutions to technical issues existing in Informatica Big Data Management as compared to Informatica Power Center edition.
- Onboarding of new resources and raising all necessary requests.

#### **Farmers EDBI Migration Project**

Client	:	Zurich Farmers
Duration	:	April 2015 – Sep. 2017
Environment	:	Informatica Big Data Management 10.1.0, DB2, Hive, Hadoop, Spark
	:	Role
	:	Senior Consultant
Team Size	:	13

Zurich Farmer's is one of the three core business segments of Zurich Insurance group. General Insurance and Global Life are other two core business segments. Zurich's Farmers segment includes Farmers Management Services, which provides non-claims related management services to the Farmers Exchanges (not owned by Zurich), as well as the Farmers RE business which includes reinsurance assumed from the Farmers Exchange by the Group. Zurich's Farmers Insurance Group is the third largest insurance group in the United States.

Farmers EDBI Migration project is offloading Specific ETL Processes from current processing path to a new Hadoop Processing path. Currently Informatica PowerCenter is being used along with DB2. Client will now be moving the processing to Hadoop. Informatica Big Data Management Edition is being used along with Hive as target database.

Project involves setting up of Hadoop cluster & converting PowerCenter ETL Mappings to Informatica BDE ETL Mappings. There are around 1800 ETL jobs in scope and 600 are being moved to Hadoop processing environment as part of initial project phase.

### **Responsibilities:**

- Working as Senior Consultant for Informatica Big Data Management (BDM) at offshore.
  - Meeting with different project stakeholders and providing them update on progress being done.
  - Updating the weekly progress tracker and providing ETL offload metrics to the client.
  - Revalidation of estimates and delivery dates as per issues being reported.
  - Active interaction and daily status meeting with the onshore team to provide progress update and work being done
  - Working on Informatica Developer Tool and finding workaround for limitations of tool for Hive environment.
  - Working on providing solutions to technical issues existing in Informatica Big Data Management as compared to Informatica Power Center edition.
  - Working with Informatica support team & raising support tickets related to technical issues/problems associated with BDM tool.
  - Onboarding of new resources and raising all necessary requests.
- Providing estimates for conversion of code in Informatica PowerCenter to BDM edition.

### **Asset & Resource Analytics**

Client : Network Rail  
Duration: November 2014 – March 2015  
Environment: Hortonworks, Sqoop, PIG, Hive, Tableau  
Role : Senior Consultant  
Team Size: 10

Network Rail is the owner and infrastructure manager of most of the rail network in England, Scotland and Wales. Network Rail's main customers are the separate and mostly private train operating companies (TOCs), responsible for passenger transport, and freight operating companies (FOCs), who provide train services on the infrastructure that the company owns and maintains.

SPEED: Weekly data is pulled from ORACLE database. This software runs on every consultant's machine capturing their activities like software tools being used.

STIME: Weekly data is pulled from SQL Server database. It contains shift timing and location information of an employee.

WORKDAY: Weekly data is pulled from DB2 database. The data contains full information of any employee.

STAGEL: Data is pulled from sources into STAGE Layer on Hadoop tables with minimal transformation.

Data Store: Data is transferred from STAGE layer to DS layer after application of business and transformation logic.

RAL: This contains views created on Data Store tables. Reporting done on views using Tableau

### **Project Responsibilities:**

- Worked as Senior Consultant and coordinating with offshore development team as well. Worked onshore at client location for a year.
- Creation of project related documents and metrics tracker to share the improvement with the client.
- Interacting with various stakeholders to get information required.
- Interacting with Business Analysts and Subject Matter Experts to get the requirements.
- Onboarding of new resources and raising all necessary requests.
- Revalidation of estimates and raising CR as needed.
- Data ingestion into Hadoop using Sqoop
- Creation and loading of data into Hive tables.
- Assigning work to team members and tracking the progress.
- Raising and tracking tickets with Hortonworks in case of any environment related issues.

### **Guardian Life Insurance Company**

Client : GLIC  
Duration: June 2014 – October 2014  
Environment: Power Center Informatica, Oracle, DB2, Unix  
Role : Team Lead  
Team Size: 8

LMER project provides a staging area for reporting environments to analyze policy information at a detailed level on daily basis for any changes (transactions) on the policy. This will include required data elements for the policy, location, coverage and associated risk attributes.

**Responsibilities:**

- Worked as Team Lead to implement ETL strategy.
- Contributed as an ETL Lead, developer and implementation of the following business process Life, Dental and WVR.
- Read source data from the DB2 system as part of extractions process using Power Center Informatica tool.
- Adhere with best practices to optimize the code in the Informatica mapping level and Informatica workflow level.
- Designed and developed ETL process for life data using Informatica tool.

## **CML Integration**

Client : Deutsche Bank, Frankfurt  
Duration: January 2013–April 2014  
Environment: SAP BW, Power Center Informatica 9.1.0 Hotfix, Informatica Power Exchnage, UNIX Role  
 : ETL Consultant  
Team Size: 4

Consumer and Mortgage Loan to finance the purchase of real estate and an amount of money lent to an individual (usually on a non-secured basis) for personal, family, or household purposes, usually with specified payment periods and interest rates. The borrower gives the lender a lien on the property as collateral for the loan. The mortgagor's lien on the property expires when the mortgage is paid off in full.

The legacy application PHD is processing and storing its data file based on the mainframe. Its data model is distributed over several files for storing bookings, dunning etc.

The PHDB file is the main format for data to interfacing applications. It provides a demoralized, flat, view on all accounts, with current past and future conditions, turnovers, booking, dunning data and secure. This integrated data mart is provided daily.

This output format is used by several applications consuming loan data including finance systems.

A general assumption of the project is to maintain old interfaces as much as possible, to keep the CML related impact for depending applications as low as possible.

Because of this ETLF will provide transformations based on CML data available in SAP BW open hubs, to derive data as close as possible to a super-set of existing legacy system interface requirements (provided as temporary buffer, referred to as Transformation Area and Delivery Area).

The Transformation Area and Delivery Area will be used to provide data feeds to interfacing applications with only the relevant columns (or substitution values if any).

### **Responsibilities:**

- Contributed for CML project as ETL architect and ETL developer. Accountable for end to end implementation of the following business process Transformation Area and Delivery Area.
- Configured SAP BW connection on ETL platform to read source data from the OH tables as part of extractions process using Informatica Power Center 9.1.0.
- Performance tuning at Informatica level, adhere with best practices to optimize the code in the Informatica mapping level and Informatica workflow level.
- Designed and developed ETL process for the Integration of Consumer and Mortgage data from SAP BW (OH Tables) transform and delivered to respective interfaces to Mainframe using Informatica tool.
- Reviewing deliverables that are to be delivered to the business.

## **Reconciliation Business Partner and SPAR**

Client : Deutsche Bank, Frankfurt  
Duration: September 2011 – December 2012  
Environment: SAP BW, Power Center Informatica 9.1.0 Hotfix, Informatica Power Exchnage, UNIX Role  
 : ETL Consultant  
Team Size: 4

Reconciliation is the process the reconcile the data from the different systems and generate the business report of deviation for each system. The following process has been evaluated to check data consistency and data completeness.

- End to End
- Daily Check
- Checksum

E2E: To compare source and target system on content an end to end Reconciliation. There are source data has been extracted from the various sources and transformed based on the business logic into target system. Then the transformed source data and the transformed target data can be compared. If data is missing in source or target a deviation should be reported.

Daily Check: During the daily check the list of changes are created via DSNTIAUL from PD. There is a list from the change log and a list from the refresh. The logs are loaded to the migration database via ETL. SAP BW is daily loaded from SAP BP.

Checksum: Checksums are used to indicate completeness of a data migration. Objects are counted and numeric attributes are summed over at several steps in the process and compared with the corresponding values at other steps. Since the transformation logic is often non-trivial - i.e. objects are filtered out, some may be split or merged – checksums at different steps should be compared via formulae.

A typical formula would be: Loaded records in target system = records (to be migrated) – records (not to be migrated) – error records

### **Responsibilities:**

- Contributed for Reconciliation project as a developer. Accountable for end to end implementation of the following business process E2E, Checksum and Daily Check.
- Created copy book to read source data from the mainframe system as part of extractions process using Power Exchange tool.
- Performance tuning at Informatica level, adhere with best practices to optimize the code in the Informatica mapping level and Informatica workflow level.
- Designed and developed ETL process for the migration of Business partner data from mainframe applications and reconcile using Informatica tool.

### **SPAR (DM) Migration**

Client : Deutsche Bank, Frankfurt  
Duration: April2011 – December2012  
Environment: SAP BW, Power Center Informatica 9.1.0 Hotfix, Informatica Power Exchnage, UNIX Role  
: ETL Consultant  
Team Size: 4

Bank is a financial institution and financial intermediately that accepts deposit and channels those deposits into lending activities either directory or through capital markets. A bank connects customers that have capital deficit to customers with capital surplus.

One of the main functions of Deutsche bank is to accept deposits. Deposits may be fixed, saving, current etc. Deutsche Banks should pay interest to the customers based on the amount deposited by them in different types of product e.g. FESTZINSSP, SPARBUCH, ZIELSPAREN, SPARCARD, TOPZINS etc.

SPAR is a legacy system being used by Deutsche bank to make business more dynamic and ease to operate business globally, Deutsche bank has migrated all their business data from legacy system to SAP DM (Deposit Management) system.

### **Responsibilities:**

- Achieved incredible milestone after successful implementation of SPAR release 03 with zero defect. Immensely applaud by the client and all the stack holders after successful accomplishment of release 03.
- Created copy book to read source data from the mainframe system as part of extractions process using Power Exchange tool.
- Delivered successfully csv files to the SAP team to accommodate migrated data into SAP system.
- Achieved incredible result more than 98% accounts migrated into SAP system after successful implementation.

### **TECHNICAL SKILL SET**

ETL Tools : InformaticaPowerCenter,  
Informatica Big Data Management (BDE)  
Database : Microsoft SQL Server, Oracle, DB2, Hbase  
Analytical tools : SAP Business objects XI R3  
BigDataTools : Informatica Big Data Edition, Hortonworks HD  
Programming : PL/SQL  
Hadoop Skills : Pig, Hive SQL, HBase, Sqoop, Flume

## EMPLOYMENT DETAILS

Organization	Designation	Duration	Location
Infosys Limited	Lead Data Scientist	June 2018-Till Date	Bangalore
Accenture India Pvt. Ltd	Associate Manager	April 2018- June 2018	Mumbai
CSC India Pvt. Ltd.	Professional 2 – Application Delivery	February 2010- March 2018	Bangalore
CSC United Kingdom	Senior Consultant	April 2015- April 2016	London
CSC Deutschland Solutions gmbh	ETL Consultant	April 2011 to April 2014	Frankfurt
Keane India Ltd. (NTT Data)	Senior Software Engineer	March 2007 to February 2010	Hyderabad
Vista Infotec	Software Engineer	August 2006 to March 2007	Bangalore

## EDUCATION DETAILS

Program	Institution (Board)	Years
EPDS01	Indian Institute of Management (IIML)	2018-2019
MBA-IT	Sikkim Manipal University, Hyderabad	2006-2008
Bachelor of Technology (Information Technology)	Manipal Institute of Technology, MAHE	2001-2005
BSEB	S.U College, Hilsa	1999
BSEB	Ram Babu High School, Hilsa	1997

## RECOGNITIONS

- Recognized by leadership multiple times for the zeal to deliver quality work and consistent impressive performance.
- Granted project appreciation award for exemplar performance in Deutsche Bank project
- Recognized as best technical trainer for Informatica Power center.
- Guided and coached new joiners by sharing business knowledge and experience thus helping them develop technical and professional skills

## PERSONAL DETAILS

Name : Dhananjay Kumar  
Father's Name : Ram Swarup Prasad  
Date of Birth : 20 December 1981  
Gender : Male  
Marital Status : Married  
Hobbies : Listening Music, Watching movies & TV Series.  
Permanent Address : #224, 5<sup>th</sup> Block, 4<sup>th</sup> Cross, Koramangala, Bangalore-560095