



Anand Nagar, Krishnankoil - 626126, Srivilliputtur (via), Virudhunagar District, Tamilnadu.

APPLICATION FOR ADMISSION TO Ph.D. PROGRAMMES

Date of Application:20-11-2025

Department	INFORMATION TECHNOLOGY	Application No.	2025011808
Area of Research	DATA ENGINEERING	Research Mode	PART TIME

Name :PRABHU BALAN JOSEPH
Date of Birth / Age :01-01-1998 / 27 Years
Gender :MALE
Category :BC
e-Mail ID :prabhubalanjoseph@outlook.com
Mobile :9884125299



Father's/Husband's Name	JOSEPH	Father's/Husband's Occupation	RETIRED SUB-INSPECTOR OF POLICE
Family Income	1000000	Residential Type	RURAL
Birth Place	KARAIKUDI	Mother Tongue	TAMIL AND ENGLISH
Religion	CHRISTIAN	Marital Status	MARRIED
Aadhaar No.	517804036630	PAN No.	BDPPP4572D
Physically Challenged	NO	Type of Disability	-

Address for Communication:
 S-1 2ND FLOOR
 5TH CROSS STREET, VIDUTHALAI NAGAR EXTN, S.KOLATHUR
 CHENNAI
 KANCHIPURAM DISTRICT
 TAMILNADU
 INDIA
 Pin-600117

Permenant Address:
 S-1 2ND FLOOR
 5TH CROSS STREET, VIDUTHALAI NAGAR EXTN, S.KOL
 CHENNAI
 KANCHIPURAM DISTRICT
 TAMILNADU
 INDIA
 Pin-600117

Qualification						
Degree	Discipline	College/university	Year Passed	AVG/CGPA	Class	Mode
B.TECH	INFORMATION TECHNOLOGY	VELAMMAL ENGINEERING COLLEGE	2009	73	FIRST CLASS	REGULAR
B.TECH	INFORMATION TECHNOLOGY	VELAMMAL ENGINEERING COLLEGE	2009	73	FIRST CLASS	REGULAR

Experience				
Organization	Designation	Experience From	Experience TO	Work Nature
WALMART	SR DATA ENGINEER	2020-01-06	1970-01-01	WORKING AS DATA ENGINEER TO BUILD PIPELINES FOR WALMART
COGNIZANT TECHNOLOGY SOLUTIONS	ASSOCIATES OF PROJECTS	2010-03-29	2019-10-04	WORKED AS DEVELOPER IN SOFTWARE DOMAIN

Payment Details				
Transaction ID	Reference	Date of transaction	Amount	Status

2025011808_251120103211	BHD51WO12SVQP0	20-11-2025	600	SUCCESS
-------------------------	----------------	------------	-----	---------

Anna University Chennai



Reg. No.12005205055/RG

The Syndicate of the Anna University Chennai hereby makes known that
PRABHU BALAN J *has been admitted to the* **DEGREE OF BACHELOR OF TECHNOLOGY** *in* **INFORMATION TECHNOLOGY** *under the*
Faculty of Information and Communication Engineering, having completed the prescribed programme of study and having been certified by the duly appointed examiners to be qualified to receive the same, and has been placed in **FIRST CLASS** *at the Examination held in* **APRIL 2009.**

Given under the Seal of the University



Chennai 600 025
India
February 2010
FLM420505809

V. Jayaram
Controller of Examinations

Sanyal
Registrar

Srinivasan
Vice-Chancellor
REGISTRATION OFFICE

Sl.No. 465509

Read by	J. Sugathan
Checked by	K. Chitra

Anna University Chennai



Reg. No.12005205055/RG

The Syndicate of the Anna University Chennai hereby makes known that **PRABHU BALAN J** has been admitted to the **DEGREE OF BACHELOR OF TECHNOLOGY** in **INFORMATION TECHNOLOGY** under the **Faculty of Information and Communication Engineering**, having completed the prescribed programme of study and having been certified by the duly appointed examiners to be qualified to receive the same, and has been placed in **FIRST CLASS** at the Examination held in **APRIL 2009**.

Given under the Seal of the University



Chennai 600 025
India
February 2010
FLM420505809

V. Jayaram
Controller of Examinations

S. Jayaram
Registrar

S. Srinivasan
Vice-Chancellor

Sl.No. 465509

Read by	J. Sugathan
Checked by	K. Chitra



CONSOLIDATED STATEMENT OF MARKS

NAME OF THE CANDIDATE		PRABHU BALAN J					REGISTER NO.	12005205055					
COLLEGE OF STUDY		120. Velammal Engineering College					MONTH & YEAR OF LAST APPEARANCE	April 2009					
PROGRAMME & BRANCH		B.Tech. Information Technology					REGULATIONS	2004					
SEM.	SUBJECT CODE	SUBJECT TITLE	MAX	MIN	MARKS SECURED	MONTH & YEAR OF PASSING	SEM.	SUBJECT CODE	SUBJECT TITLE	MAX	MIN	MARKS SECURED	MONTH & YEAR OF PASSING
01	CY1101	Chemistry-I	100	50	078	NOV 2005	05	GE1352	Communication Skills Laboratory	100	50	082	NOV 2007
01	GE1101	Engineering Graphics	100	50	093	NOV 2005	06	CS1253	Visual Programming	100	50	058	APR 2008
01	GE1102	Fundamentals of Computing	100	50	062	NOV 2005	06	IT1351	Network Programming and Management	100	50	057	APR 2008
01	HS1101	English-I	100	50	070	NOV 2005	06	IT1352	Cryptography and Network Security	100	50	064	APR 2008
01	MA1101	Mathematics-I	100	50	078	NOV 2005	06	IT1353	Embedded Systems	100	50	064	APR 2008
01	PH1101	Physics-I	100	50	064	NOV 2005	06	MA1251	Numerical Methods	100	50	067	APR 2008
01	CY1102	Chemistry Laboratory	100	50	095	NOV 2005	06	MG1351	Principles of Management	100	50	062	APR 2008
01	GE1103	Engineering Practices Lab	100	50	097	NOV 2005	06	CS1255	Visual Programming Lab	100	50	096	APR 2008
01	GE1104	Computer Practice-I	100	50	092	NOV 2005	06	CS1305	Network Lab	100	50	095	APR 2008
01	PH1102	Physics Laboratory	100	50	093	NOV 2005	07	CS1334	Graphics and Multimedia	100	50	063	NOV 2008
02	CS1151	Data Structures	100	50	056	APR 2006	07	IT1401	Component Based Technology	100	50	057	NOV 2008
02	CY1201	Environmental Science and Engineering	100	50	063	APR 2006	07	IT1402	Mobile Computing	100	50	070	NOV 2008
02	GE1151	Engineering Mechanics	100	50	067	APR 2006	07	MG1401	Total Quality Management	100	50	055	NOV 2008
02	HS1151	English - II	100	50	068	APR 2006	07	IT1001	Distributed Systems	100	50	073	NOV 2008
02	MA1151	Mathematics - II	100	50	060	APR 2006	07	IT1002	Information System Design	100	50	070	NOV 2008
02	PH1157	Physics-II	100	50	072	APR 2006	07	CS1355	Graphics and Multimedia Lab	100	50	086	NOV 2008
02	CS1152	Data Structures Lab	100	50	092	APR 2006	07	IT1403	Software Components Lab	100	50	090	NOV 2008
02	CS1153	Electronic Devices and Circuits Lab	100	50	093	APR 2006	08	IT1451	Web Technology	100	50	063	APR 2009
03	CS1202	Digital Principles and Systems Design	100	50	071	NOV 2006	08	CS1014	Information Security	100	50	064	APR 2009
03	CS1203	System Software	100	50	057	NOV 2006	08	IT1009	Web Services	100	50	070	APR 2009
03	CS1204	Object Oriented Programming	100	50	059	NOV 2006	08	IT1452	Project Work	200	100	198	APR 2009
03	IT1201	Signals and Systems	100	50	063	NOV 2006							
03	IT1202	Principles of Communication	100	50	055	NOV 2006							
03	MA1201	Mathematics - III	100	50	079	NOV 2006							
03	CS1205	Object Oriented Programming Lab	100	50	092	NOV 2006							
03	CS1206	Digital Lab	100	50	092	NOV 2006							
03	CS1207	System Software Lab	100	50	083	NOV 2006							
04	CS1251	Computer Architecture	100	50	050	APR 2007							
04	CS1304	Microprocessors and Micro Controllers	100	50	060	APR 2007							
04	CS1353	Software Engineering	100	50	064	APR 2007							
04	IT1251	Information Coding Techniques	100	50	074	APR 2007							
04	IT1252	Digital Signal Processing	100	50	072	APR 2007							
04	MA1253	Probability and Statistics	100	50	065	APR 2007							
04	CS1306	Microprocessors and Micro Controllers Lab	100	50	099	APR 2007							
04	IT1253	Software Engineering Lab	100	50	090	APR 2007							
04	IT1254	DSP and Communications Systems Lab	100	50	092	APR 2007							
05	CS1252	Operating Systems	100	50	072	NOV 2007							
05	CS1301	Database Management Systems	100	50	069	NOV 2007							
05	CS1302	Computer Networks	100	50	060	NOV 2007							
05	CS1402	Object Oriented Analysis and Design	100	50	060	NOV 2007							
05	BC1391	Telecommunication Systems	100	50	060	NOV 2007							
05	GE1301	Professional Ethics and Human Values	100	50	056	NOV 2007							
05	CS1254	Operating Systems Lab	100	50	094	NOV 2007							
05	CS1307	DBMS Lab	100	50	092	NOV 2007							
05	CS1403	Case Tools Lab	100	50	098	NOV 2007							

End of Statement
 Classification : **FIRST CLASS**
 Total Marks (from 3rd to 8th semester): 3582 / 4900
 Percentage (rounded to nearest integer) : 73



Sl. No.: UA 282047

Read by	J. Suganthi
Verified by	K. Chelva

VELAMMAL ENGINEERING COLLEGE

Approved by AICTE and Affiliated to Anna University, Chennai

(Governed by Velammal Educational Trust, Chennai.)

Velammal Nagar, Ambattur - Redhills Road,

Chennai - 600 066. Ph : 2659 0579, 2659 1860



TRANSFER CERTIFICATE

No. **4795**

1. Name and initials of the student PRABHU BALAN . J
2. Name of Parent / Guardian S. JOSEPH
3. Nationality, Religion & Caste Indian - Christian - Nathammar BC
4. Date of Birth (in words and figures) as entered in the Admission Register 01.01.1988 (First January Nineteen Eighty Eight)
5. Date of admission in this college 02.09.2005
6. Class and Course of study to which admitted I - Year B.Tech / IT
7. Semester studied / completed at the time of leaving VII - Semester
8. Whether qualified to proceed to the next semester / class Refer to the Mark Sheets
9. Whether the student has paid all the dues to the college -yes-
10. Whether he/she has completed the course -yes-
11. Date on which application for transfer certificate was made on behalf of the student by his/her parent or guardian 03.08.2009
12. Date on which the student left the College 25.05.2009
13. Reason for his/her leaving Course Completed
14. The student's Conduct and Character Good
15. Date of issue of Transfer Certificate 18.8.2009


Principal
18/8/09

Velammal Engineering College,

Ambattur Redhills Road,

Chennai - 600 066.

Phone : 2659 0579, 2659 1860

PRABHU BALAN JOSEPH

LEAD DATA ENGINEER

Lead Data Engineer with 14 plus years of experience with a proven track record in Architecting and implementing robust data solutions. Specialized in Apache Spark, Scala, Python, Tableau and a diverse range of Cloud technologies (Google & AWS). Adept at leading cross-functional teams and driving innovative data strategies to meet business objectives. Having Expertise knowledge in Insurance & Retail domain.

Informatica / Talend Certified Developer
AWS & Google Certified

WORK EXPERIENCE

Lead/Sr Data Engineer

[Walmart, Texas | Jan 2020 – Present](#)

- Design & Built various Spark/Scala pipelines for Data commercialization product of Walmart Data ventures data asset that helps Merchant and Suppliers community to take better decisions on item inventory and provide better Customer behavior insights and Product performance analysis into actions.
- It helps the business to utilize the Walmart data asset to generate new profitable revenue of 200 million plus every year.
- Implemented the Data Quality rules framework using Pyspark to calculate the key metrics and publish in Data Governance tool Collibra for Data owners and Business team.
- Lead the implementation of CCPA, CPRA and MHMD(My Health) privacy laws into Data engineering solutions for the application to stay be complaint.
- Lead the Data engineering team of 8+ people and working closely with Product team and Business team on Roadmap, LOE, data requirements and come-up with Architecture solution for new initiatives.
- Helping the Business team on re-writing any existing and New Third party data assets like Dun & Bradstreet, Nielson across the Global data teams and negotiate contracts by creating centralized 3P Data asset for Walmart teams for cost reductions etc.
- Building the functionality that helps the Data science team on building the models on top of Third-party data assets and Data commercialization projects.
- Expertise experience in Agile methodology and facilitates the scrum ceremonies like Sprint planning, backlogs etc.
- Maintaining the CICD process and Approval of code migration/review process end to end.
- Experienced in handling/coordinating between teams across multi-region efficiently.



860-709-5688

prabhubalanjoseph@outlook.com

Bentonville, AR

linkedin.com/in/prabhubalanjoseph

EDUCATION

B.Tech (IT)

Anna University, Chennai
2005 – 2009

EXPERTISE

TECHNICAL

- Spark, Scala & PL/SQL
- Talend, Informatica and Nifi
- BigData, Delta Lake, HUDI, Presto, Snowflake, BigQuery & Oracle
- Google & AWS Cloud
- Cloud break, CloudSQL, GCS, S3, EC2.
- Autosys, Airflow & Automic
- JIRA, Confluence, Rally.
- CICD (Looper & Concord)
- Tableau & Looker

PROFESSIONAL

- Collaboration
- Adaptability
- Problem Solving
- Strategic Thinking
- People Skills

AWARDS

Award of Awesomeness

Cognizant (EIM Practice)

2015

COMMUNITY INVOLVEMENT

Cognizant Outreach

Volunteering Food camps,
school teaching etc

- Experienced in handling multi-tera bytes of Data processing and transferring to external third parties via Google Cloud.
- Ownership of taking care end-to-end pipeline executions in Production and coordinating with Upstream teams on any data quality issues etc..

Module Delivery Lead (American Insurance Group)

[Mphasis Corporation, Connecticut | Oct 2019 – Dec 2019](#)

- Leading the team on building Life information factory platform for Data management and Governance process to handling the policies information and used by Actuarial system and finance team.
- Built Talend Bigdata jobs for Data standardization service by transforming data according to the Business rules and loading to Clean Lake (Amazon EMR).
- Developing Talend jobs for Data Acquisition services by pulling the flat files, fixed width separated files and XML's from S3 bucket to Murky Lake environment (Amazon EMR)
- Implemented the Data quality framework as part of data pipeline automation which saved manual data validations.

Sr Talend/Informatica Lead (The Hartford Insurance)

[Cognizant, Connecticut | Mar 2010 – Oct 2019](#)

- Built end to end Data pipelines using Informatica and Talend tool for Group Benefits and Commercial domain and store the data in On-Prem and Amazon S3 for downstream consumption.
- Hands on experience on Informatica 10.2 BigData Management tool to ingest Data into Hadoop platform.
- Build and maintaining the Master metadata using Informatica MDM tool for data lineage and reporting purpose.
- Building the Talend Ingestion framework to extract the Clickstream Data from BigQuery and store in Hadoop environment for Data science team modeling.
- Involved in building Nifi process to convert JSON files into AVRO format in Hadoop environment for Data Science team consumption.
- Leading the team of 10+ members (Onshore/Offshore) and coordinating with Application owners on new initiatives etc.
- Extensively used Talend components tMap, tDie, tConvertType, tFlowMeter, tLogCatcher, tRowGenerator, tOracleInput, tOracleOutput, tFileList , tReplicate, tSortrow, tDenormalize, tNormalize etc
- Creating reusable Audit framework for the Data commercial domain using Talend tool using context and global variables.
- Played active role in carrying out source data analysis, volumetric analysis, performance testing of the Datamart.
- Actively participating in the Tableau dashboard development along with the business team for data gathering and profiling.
- Handling Maintenance and Production support teams in Onshore and responsible for Application availability etc.
- Expertise in Ticketing tools like Rally, JIRA and creating knowledge sharing documents for LI support team.
- Lead the Conversion project of Informatica to Talend tool and converting 1000+ mappings to Talend jobs in Bigdata which reduced the Licensing cost to Business and decommission of separate compute for Informatica tool.

आयकर विभाग
INCOME TAX DEPARTMENT

भारत सरकार
GOVT. OF INDIA

J PRABHU BALAN
JOSEPH
01/01/1988

Permanent Account Number
BDPPP4572D

Prabhulan J.
Signature



In case this card is lost / found, kindly inform / return to :
Income Tax PAN Services Unit UTITSL
Plot No. 3, Sector 11, CBD Belapur,
Navi Mumbai - 400 614.

इस कार्ड के खोने/पाने पर कृपया सूचित करें/लौटाएं :
आयकर पैन सेवा यूनिट, UTITSL
प्लॉट नं: ३, सेक्टर ११, सी.बी.डी. बेलापूर,
नवी मुंबई-४०० ६१४.



India Driving Licence(Tamilnadu)
Form 7

DOI 10/01/2009

D.L.No : TN-0520090000289
Name : PRABHU BALAN J
S/D/W of : JOSEPH S
Address :
NO 5/420 NEWNO 2 MUTHAMIL NAGAR
BLOCK 5 CHENNAI
600118
Temp.Addr:



D.O.B. : 01/01/1988 B.G. :

Punishments:

Licensed to drive throughout India, vehicles of the following descriptions
MC-WG 10/01/2009 TN-05

Non-Transport Veh. Valid upto **09/01/2029**



Prabhu J
Signature/L.T.I
of the Holder

R. Anandaraman
10/1/09
Asst. Licensing Authority
Chennai (North)

R.Dis. 16163 / 2005 04.06.2005

BC

சான்றிதழ் எண் :
Certificate No. :

6319618

மாவட்டக் குறியீடு எண் :
District Code :வட்டக் குறியீடு எண் :
Taluk Code :கிராமக் குறியீடு எண் :
Village Code :

01	
03	
004	

சாதிச் சான்றிதழ்
COMMUNITY CERTIFICATE

ஆணை மாவட்டம் ஏ.டி. வட்டம்
எலத்தி கிராமம்/நகரம், திரு./திருமதி/செல்வி/
செல்வன் த. ஜி. முருகு மூலன் தகப்பனார்/கணவர்
பெயர் சி. ஜே. சி. இராஜ் கந்தன்

வகுப்பைச் சார்ந்தவர், அரசு ஆணை நிலை எண் 28, பிற்பட்ட மற்றும்
மிகவும் பிற்பட்ட பிரிவின் நலத் துறை, நாள் 19-7-1994 வரிசை எண் 142 படி,
பிற்பட்ட பிரிவினைச் சார்ந்தவர் எனச் சான்றளிக்கப்படுகிறது.

This is to certify that..... Son / Daughter of
Thiru..... of.....
..... Village / Town,..... Taluk,.....
..... District of the State of Tamil Nadu belongs to
..... Community, which is recognised as
a Backward Class as per Government Order (Ms.) No. 28, Backward Classes
and Most Backward Classes Welfare, dated 19th July 1994
vide Serial No.....

2. திரு./திருமதி/செல்வன்/செல்வி. ஜி. முருகு மூலன்
என்பவரும் அவருடைய குடும்பத்தினரும் தமிழ்நாட்டில் 129 ச.மீ.
மாவட்டத்தில் ஏ.டி. வட்டத்தில் இராஜ் கந்தன், 10
கிராமத்தில்/நகரத்தில் வசித்து வருகிறார்கள் எனச் சான்றளிக்கப்படுகிறது.

2. It is certified that Thiru/Tmt./Selvan/Selvi..... and
his/her family ordinarily reside(s) at..... Village/
Town..... Taluk..... District of
Tamil Nadu.

முத்திரை.
Seal.கையொப்பம் :
Signature :நாள் :
Date :பதவி (Capital Letters) :
Designation :பதவி
Designation :

DEPUTY TAHSILDAR
(CERTIFICATES)
EGMORE-NUNGAMBAKKAM TALUQ
CHENNAI - 600 031.

Cross-Functional Frameworks for Petabyte-Scale Analytics: Design, Optimization, and Cost Modeling

Background:

Today, organizations generate petabytes of data in areas like retail, finance, healthcare, and IoT. Processing and analyzing this data requires frameworks that bring together data engineering, distributed computing, and cost optimization strategies. Most current solutions focus on separate parts, such as storage or computing, instead of taking a complete approach to build scalable systems, improve performance, and model costs in hybrid environments (cloud plus on-prem).

Problem Statement

Petabyte-scale analytics faces challenges such as:

- Fragmented architectures lacking interoperability between data engineering, ML, and BI teams.
- High operational costs due to inefficient resource allocation.
- Performance bottlenecks in distributed query execution and ETL pipelines.
- Limited predictive cost modeling for multi-cloud and hybrid deployments.

Objectives

1. Design a cross-functional framework that unifies data engineering, analytics, and ML workflows for petabyte-scale datasets.
2. Develop optimization strategies for query execution, storage tiering, and compute resource allocation.
3. Create predictive cost models for hybrid and multi-cloud deployments using AI-driven simulations.
4. Benchmark performance and cost trade-offs across different architectures (e.g., Spark, Presto, Snowflake, BigQuery).

Proposed Approach

- **Architecture Design:** Define modular components for ingestion, transformation, and analytics using distributed systems (Apache Spark, Presto, Dask).
- **Optimization Techniques:**
 - Adaptive query optimization using ML.
 - Intelligent caching and tiered storage strategies.
- **Cost Modeling:**
 - Build AI-driven models to predict cost under varying workloads.
 - Simulate scenarios for hybrid cloud vs. on-prem deployments.
- **Evaluation:**
 - Use real-world datasets (e.g., retail transaction logs, IoT streams).
 - Metrics: Query latency, throughput, cost per TB processed.

Methodology

- A reference architecture for petabyte-scale analytics.
- An optimization toolkit for performance tuning.
- A predictive cost modeling framework for enterprise decision-making.
- Comparative analysis of cost-performance trade-offs across platforms.

Conclusion

This research will help enterprises scale analytics efficiently, lower costs, and improve teamwork between data engineering, analytics, and ML teams. It responds to the rising demand for frameworks that work together amid massive data growth and hybrid cloud use.