Kannan Kandasamy

+1 (425) 499-5580 | Email:
kannanvijay@hotmail.com

https://www.linkedin.com/in/kannankandasamy
http://stackoverflow.com/cv/kannankandasamy

Senior Data Engineer

PERFORMANCE HIGHLIGHTS

- ☑ **14+** years of professional IT experience which includes experiences in databases and **Data Warehousing** in large volume data processing and expertise in **ingestion, storage, querying, processing and analytics of structured and unstructured data** on SQL Server BI, Analysis Services, and Hadoop/Spark Ecosystem
- ☑ Expertise in **Python 3.5, pandas, numpy, scikit learn,** matplotlib, pyspark Cloudera Hadoop, HDFS
- ☑ Expertise in **Exploratory data analysis, model building** using machine learning techniques
- ☑ Extensive experience in **SQL SERVER** 2017/2016, SSIS, SSRS, SSAS**,** Azure SQL, Azure Analysis Services**, Azure Data Factory, Azure DataBricks**, Power BI
- ☑ **Cloudera Certified** Developer on Spark and Hadoop (**CCA175**)
- ☑ Experience in **Microsoft Azure** – Microsoft **certified Azure Architect** with specialization on data
- ☑ Experience in AWS cloud **- AWS Certified** Cloud Practitioner
- ☑ Contributor on **StackOverflow,** author on Mssqltips and my blog
- ☑ Expert in **dimensional modeling** and designing SSAS tabular cube as per Kimball group of dimensional modeling techniques
- ☑ Experience in Microservices architecture using **Python** and Azure services
- ☑ Experience in designing and implementing relational database model as per business needs using **Erwin Data Modeler** and MS Visio
- ☑ Proven ability to implement and adhere to **SDLC, AGILE**(SCRUM) methodologies
- ☑ Experience in Functional **Solution Architecture for ETL & Business Intelligence projects** in Enterprise Data Warehouse space
- ☑ Good understanding on streaming datasets using **Kafka and Spark Streaming**
- ☑ Extensive Consulting Experience, architectural capabilities with presales ability
- ☑ Experience in **costing/budgeting, pricing** of various Azure components related to Data
- ☑ Extensive experience in **sizing, scoping, estimation, risk management, release management** and change management on BigData and DataWarehousing projects

TECHNICAL SKILLS

Databases	SQL Server 2017/16/14/12/2008(R2)/2008, SQL Azure, DB2, Oracle
Big Data	Python 3.5, Pandas, Numpy, Scikit learn, Hadoop(2.6), Apache Spark (1.6.2, 2.4), Spark SQL, HDFS, Map Reduce
BI Tools	SSIS, SSAS, Azure Data Factory, Microsoft Internal ETL Tool MeTAL3.0, Azure Analysis Services, Azure Data warehouse, Azure DataBricks, Kafka
Data Visualization	SQL Server Reporting Services(SSRS), Power BI, Seaborn, Matplotlib
Data Modeling Tools	Erwin, Visio
NoSQL	Azure Cosmos DB
Version Control	Visual Studio Online, Source Tree, VSTF, Visual Source Safe, SVN, Git
Development Tools	Visual Studio 2013, 2015, 2017, Eclipse, Intellij, PyCharm, Jupyter, VSO

CAREER SNAPSHOT

July 2016 – Present	Senior Data Engineer	Yashco Systems
2005 – July 2016	Technology Architect	Infosys Ltd.,

Kannan Kandasamy

+1 (425) 499-5580 | Email:
kannanvijay@hotmail.com

2003 – 2005	Software Engineer	Covansys (CSC)

CERTIFICATIONS

Cloudera Certified Spark and Hadoop Developer (CCA175) (License code: 100-020-588)

Microsoft Certified - DP-200 Implementing an Azure Data Solution
(https://www.youracclaim.com/badges/2e3f01b7-40a3-4506-97e6-190e2e92c70e)

Architecting Azure Solutions - Microsoft (Microsoft Certification ID: 3127887)– 70-534
(https://kannankandasamy.blob.core.windows.net/kan/kannanazurecert.pdf)

AWS Certified Cloud Practitioner (Validation ID: NGQ5GRYK12F1QHSJ)

Microsoft Certified Technology Specialist (Microsoft Certification ID: 8814881) – 70-448 (**Microsoft SQL Server Business Intelligence development** and maintenance)

MAPR Academy Certified **Apache Spark** Essentials - DEV 360 (ID: m2fu3xko7sr7)
(http://verify.skilljar.com/c/m2fu3xko7sr7)

EDUCATION & QUALIFICATIONS

Masters in Computers and Applications (M.C.A) – (Madurai Kamaraj University) - 2003
Bachelor of Science (B.Sc.) (Computer Science) - 2000
Certification – Hadoop Application Framework - University of California, San Diego - 2016

PROFESSIONAL EXPERIENCE

KEY PROJECTS:

- **Excellus BCBS** – Data analytics for Pega — **Excellus BCBS**
  Technologies : Python, Numpy, Pandas, Erwin, Azure DataFactory
- **PeerPlace Networks – Advanced Analytics for HealthCare** — **PeerPlace Networks LLC**
  Technologies : SQL Server, SSIS, SSAS, SSRS, Python, Azure
- **Microsoft – Customer Data Enrichment Services** — **Microsoft Corporation**
  Technologies : SQL Server, SSIS, SQL Azure, Marketo
- **Microsoft – METaL ETL Framework** — **Microsoft Corporation**
  Technologies : C#, SQL Server, SSIS
- **Toyota – One Dealer Daily** — **Toyota Motor Sales - US**
  Technologies : DB2, Informatica, PWX

01 Project Name: Pega Marketing for Excellus Blue Cross Blue Shield (BCBS)

Client: Virtusa – Excellus Blue Cross Blue Shield

Description of the project

Analysis and design of data solution for marketing application in Excellus BCBS. We source data from multiple sources and feed the data into Pega Marketing

Role : **Data Engineer**

Duration : Mar 2020 to till date

Technologies : **Python 3.5, Pandas, Numpy, Scikit learn, Matplotlib, Seaborn, Jupyter notebook, Azure DataFactory, Azure SQL, SQL Server, SQL Server Analysis Services, Azure databricks, Azure SQL DataWarehouse**

Responsibilities

Kannan Kandasamy

+1 (425) 499-5580 | Email:
kannanvijay@hotmail.com

- End to end data architecture for getting data from different sources and feeding into Pega application
- Data Modeling based on business requirements for Marketing
- Data Harmonization of all the data points which are available on existing sources
- Detailed **Exploratory Data Analysis** using Pandas, matplotlib and seaborn libraries

02 Project Name Advanced Analytics for HealthCare

Client PeerPlace Networks LLC (Data management system for healthcare)

Description of the project

PeerPlace data management software streamlines identification and tracking by providing one Master Client Record that can be securely shared electronically among all care professionals that provide services to that consumer. This project has two different roles and responsibilities one to do traditional descriptive analytics using SQL Server and predicting future data using predictive analytics.

Predictive Analytics using Machine learning and Statistical way

Role : **Data Scientist**

Duration : July 2019 to Feb 2020

Technologies : **Python 3.5, Pandas, Numpy, Scikit learn, Matplotlib, Seaborn, Jupyter**

BigData : **PySpark, Spark framework, Cloudera Hadoop cluster, HDFS and MapReduce, Azure DataLake, Azure DataFactory, Azure DataBricks**

Responsibilities

- **End to End model building for a classification problem** to **predict** whether a senior citizen can be hospitalized based on the observations
- Data cleansing, cleaning, transformation using **Pandas and Numpy** libraries
- Detailed **Exploratory Data Analysis** using Pandas, matplotlib and seaborn libraries
- Feature engineering and Feature selection using machine learning approaches and healthcare domain knowledge
- Developed a Supervised Machine learning model for a regression and classification problems using **Linear Regression, Logistic Regression** and Decision tree, random forest algorithms
- Trained and tested the model using **train_test_split/k-fold cross validation** approaches
- **Hyperparameter tuning** using **GridSearchCV**, RandomSearchCV and traditional methods
- Verified Model accuracy using RMSE or Confusion Matrix/classification report depends on the algorithm
- Deployed model using Python and merged predicted data into traditional Java and DB2 application

03 Project Name Advanced Analytics for HealthCare - **Descriptive Analytics Datawarehouse approach**

Client PeerPlace Networks LLC (Data management system for healthcare)

Role : Senior Data Engineer

Duration : July 2016 to July 2019

Kannan Kandasamy

+1 (425) 499-5580 | Email:
kannanvijay@hotmail.com

Technologies : **SQL Server, SQL Server Analysis Services, Python, Pyspark, Azure Data Factory(ADF), Azure databricks, Azure SQL, Azure SQL DataWarehouse, SSIS, Azure Data Lake, Azure blog Storage, SSRS, Power BI, C#, Custom Security**

Responsibilities

- Design and development of end to end **Data warehousing** from data extraction, transformation and loading to DataMart along with reports generation
- Analysis of business requirement and generating Source to Target mapping
- Implemented dimensional modeling by defining dimensions, facts along with relationships based on reporting analytical requirements
- **Migration of On-premises BI application** pipelines to Azure cloud
- Built pipelines in **Azure data factory** to ingest data from transactional on-premises system to **Azure Blob Storage**
- Built pipelines in **Azure DataBricks** for doing faster in memory data transformation and sink data to Azure data Lake storage
- Analysis and design of slowly changing dimensions based on the requirements
- Design and Development of ETL using SSIS, cube using Analysis Services and Analytics reporting using SSRS and Power BI
- Implemented custom security (my blog) with single sign-on authentication based on SAML 2.0 using C# and SSRS

---

04 Project Name Customer Data Enrichment Services (CDES)

Client Microsoft Corporation

Description of the project

Customer Data Enrichment Services is a Lead Enrichment provider which gets leads from Marketo and enrich leads based on Company, Email, Country, phone number and provide the enriched leads back to Marketo.

Role : Sr. Software Development Engineer

Duration : Feb 2014 to July 2016

Technologies : **Python, Marketo, SQL Server 2014, SSIS, Azure Data Factory, VSO CI/CD deployment, PowerShell, SQL Azure, Azure Data Warehouse, SQL Server Analysis Services**

Responsibilities

- Developed end to end SSIS with Fuzzy lookup component for company matching with existing Microsoft Sales information for enrichment.
- Messages we receive from Marketo to Azure Service Bus and experience in processing messages in separate Services. Overall project is in **MicroServices** and one of the service is to enhance customer data

Kannan Kandasamy

+1 (425) 499-5580 | Email:
kannanvijay@hotmail.com

- Involved in designing high volume **Tabular model cube in SSAS 2016** to provide hot path analytics solution
- Implemented continuous Integration and continuous deployment using Powershell and Visual Studio Online and **DACPAC** for SQL and SSIS deployments
- Designed and **developed Archival framework for large volume of transactional data** which provides online/offline data movement and purging
- Designed and developed **solution for scrubbing data by masking** production PII information on development and test environments

---

05 Project Name **METaL ETL Engine Framework** and other Projects for Microsoft

Client Microsoft Corporation

Description of the project

METaL is an ETL engine framework built on top of metadata to centrally manage and execute their ETL packages. It's a standardized warehouse solution across BICOE utilizing SQL Server 2008/12 & SSIS, reducing development time for new and existing systems, as well as eliminating shadow ETL application development. METaL implements Blueox a general-purpose logging subsystem, built on SQL Server Service Broker and Event Notifications, designed to provide benefits such as Asynchronous local and remote logging, Minimal application resource requirements, Centralized log store for multiple applications, reduced application development time.

Role : Software Development Engineer

Duration : From 2005 to 2014

Software : Visual Studio 2012, C#, SQL Server, SQL Server Integration Services

Responsibilities

- Part of development team involving **Metadata design**, **engine design and development** using **C# and SSIS components**
- Metadata includes source to target mapping, filtering, aggregation transformations and engine is developed accordingly to implement the execution of metadata
- METaL engine **dynamically generates SSIS packages** based on source to target mapping metadata and generated package will be deployed on worker nodes and executed at worker nodes for data transfer
- Logging is implemented based on Service broker from worker nodes to master nodes
- Implemented METaL in different projects in BICoE group of Microsoft

---