Industrial Project for Sekar Chenniyappan

Area:	Big Data Analytics
Title:	Text Analytics of Ancient Indian Scriptures
	(I reserve the right to change the title suitably as we progress)
Company Name:	SRILA Labs, Srirangam and
	Radien Software Private Limited, Pondicherry
R&D Cell Details:	Radien Software is an advanced design and
	development company for specialised startups
	in Silicon Valley. Almost all the work we do is
	cutting edge design and development. And pushes
	the edge of technology. All the work involved research
	without publishing papers.
	Date of Incorporation: 1 March 2007
	Software Research Industrial Life Analytics Labs (SRILA Labs) is a
	startup founded by the founders of Radien Software
	for Big Data Analytics.
	Date of Incorporation: 1 January 2015
Copy of Registration:	https://opencorporates.com/companies/in/U72900PY2007PTC002034
Internship order:	Document Attached
Internship Locations:	Srirangam, Pondicherry, Chennai and Bangalore
External Supervisor:	Dinesh Thirumurthy
Internal Supervisor:	Venkatesan Pillai
Abstract:	The core idea of Sekar's project will be to analyse groups of words
	that occur repeated in english translations of the Bhagavad-gita.
	Identity anything interesting with Machine Learning techniques.
	If possible visualise the analysis through Visualisation technologies.
	And make available it all for the whole world to see and experience
	in an internet portal called http://bhagavad-gita.today
Technologies:	Python, Hadoop 2, SPARK, Map Reduce,
	Data Structures, and Text Processing Algorithms
Base paper:	See next page.
	This is not a research project.
Attendance:	Please kindly treat intern as attending college as he is
	reporting to my office since 27th December 2014. I kindly request
	that all meetings happen over skype instead of physical travel.
Single Point of Contact:	Dinesh Thirumurthy
Email of SPOC:	srila.labs@gmail.com
	dinesh.thirumurthy@acm.org

This document is to certify that Sekar Chenniyappan is doing the above project with me at my companies in their various branches. Thanks very much.

Dinesh Thirumurthy
VP, HR, Radien Software Private Limited.
Founder, SRILA Labs, Srirangam
Founder, Radien Software Private Limited, Pondicherry

ACM Base Paper:

Distributed Data Management using Map Reduce

- Feng Li, Beng Chin Ooi, M. Tamer Özsu, Sai Wu
January 2014 Computing Surveys (CSUR) , Volume 46 Issue 3
Publisher: ACM

Location: http://www.comp.nus.edu.sg/~ooibc/mapreduce-survey.pdf

Base Paper Abstract:

Distributed Data Management Using MapReduce

FENG LI, National University of Singapore
BENG CHIN OOI, National University of Singapore
M. TAMER ÖZSU, University of Waterloo
SAI WU, Zhejiang University

MapReduce is a framework for processing and managing large scale data sets in a distributed cluster, which has been used for applications such as generating search indexes, document clustering, access log analysis, and various other forms of data analytics. MapReduce adopts a flexible computation model with a simple interface consisting of map and reduce functions whose implementations can be customized by application developers. Since its introduction, a substantial amount of research efforts have been directed towards making it more usable and efficient for supporting database-centric operations. In this paper we aim to provide a comprehensive review of a wide range of proposals and systems that focusing fundamentally on the support of distributed data management and processing using the MapReduce framework.

_
End of Document v3