

GERÊNCIA DE INFRAESTRUTURA PARA BIG DATA

Prof. Tiago Ferreto – tiago.ferreto@pucrs.br



INTRODUCTION

Who am I?

- Prof. Tiago Ferreto
 - Associate Professor – School of Technology – PUCRS (since 2004)
 - Member of the Post-graduate Program in Computer Science (since 2012)
 - Advisor of Master and PhD students
 - Coordinator of LAD/PUCRS (High Performance Computing Laboratory)
- Research topics
 - High Performance Computing, Grid Computing, Cloud Computing, Edge Computing, Fog Computing
 - IT Infrastructure – virtual machines, containers, unikernels, serverless computing
 - Networking – Software Defined Networks, Network Functions Virtualization
 - DevOps – Infrastructure as Code, CI/CD, Deployment Automation

Overview

- Topics
 - Big Data
 - Apache Hadoop
 - HDFS
 - Data ingestion: Flume and Sqoop
 - YARN
 - MapReduce
 - Data analysis: Pig, Hive
 - Apache Spark
 - HUE (Hadoop User Environment)
 - Apache Ambari
 - Hadoop Administration and Monitoring

Agenda

#	Data	Assunto
1	SEX – 3/8/18 – 19:00-22:30	Big Data, Apache Hadoop Hadoop Cluster Architecture
2	SÁB – 4/8/18 – 8:30-12:00	Hadoop VM, HDFS Data Ingestion (Flume, Sqoop)
3	SEX – 10/8/18 – 19:00-22:30	YARN, MapReduce
4	SÁB – 11/8/18 – 8:30-12:00	Pig, Hive
5	SEX – 17/8/18 – 19:00-22:30	Spark
6	SÁB – 18/8/18 – 8:30-12:00	HUE, Ambari
7	SEX – 24/8/18 – 19:00-22:30	Administração e monitoração Hadoop, Cloud

DINÂMICA