Einext Blog

About Us

Apache Cassandra

Authentication

Cassandra - SASI Index

Cassandra Snapshot and Restore

Complex join queries using Spark

CQL (Cassandra Query Language)

Migrate data from RDBMS

Query Cassandra Tables using Spark

Related Technologies

SSTables

Stress Testing Cassandra

Apache Solr

Solr: Custom Request Handler

Solr: Index pdf, word etc (Tika)

Solr: Indexing using spark

Solr: Morphline for ETL

Solr: Query parameters

Solr: Text Analyzers

Apache Spark

Bucket By

Codegen in Spark 2.0

Compress Output Files in Spark

Convert String to Timestamp for SparkSQL

Create Spark Project in Eclipse

Create UDF

DAG (Directed Acyclic Graph)

Data Virtualization Using Spark

Dataframe Summary

File Format

Hive Metastore in Spark

HiveContext vs Spark SQLContext

Jupyter Notebook for Pyspark

Kryo Serialization

Loading Data into HBase using Spark

ML Using SparkR

PredictionIO for Machine Learning

Search all sites 💌

Apache Solr >

Solr: Custom Request Handler

defaults: These simply establish default values for various request parameters. Parameters in the request will override them. appends: For parameters that can be set multiple times, such as fq, this section specifies values that will be set in addition to any that may be specified by the request. invariants: This sets defaults that cannot be overridden. It is useful for security purposes. It can also be used to override what the client sends when you don't have control over the client application; for instance, if the application is deployed and you can't easily redeploy a new client.

Comments

You do not have permission to add comments.

Programming Language Support for Spark

Pyspark working with HBase

RDD Operations (Scala)

RDD Partition Behaviour

Running Spark on Windows

Sbt build manager

Scala for Spark

Scala UDF in Pyspark

Setup Spark Cluster

Setup Zeppelin

Simple Dataframe Operations

Simple Stream Producer

Spark Dataframe with Python (Pyspark)

Spark Memory Management

Spark SQL over REST API

Spark to Read from S3

Stream Processing RDBMS

Streaming RDBMS Tables

Thrift Service on Spark SQL and JDBC

Twitter Kafka Spark Streaming

Twitter Stream as Kafka Source

Window Functions in Spark SQL

Working with AWS S3 Storage Using Spark

Working with MySQL from Spark SQL

XML Doc and Blob Field

Datasources

AWS Hosted Datasources

Download Stock Prices

Extracting Text

Live Tweets using Streaming API

NLP

Hadoop

Apache Kafka

Apache Solr Basics

Big Data Use Cases

Data Analysis Using Pig

Drill for Interactive Query

Hadoop Logging

Hadoop MR Project Using Maven

6/20/2018, 10:07 PM

Hadoop Security

Hadoop Stress Test

HBase - Bulk Load Into HBase Table

HBase Fundamentals

HDFS Commands

Hive - Connecting to Hive JDBC

Hive - Optimize Joins

Hive - Table Partitions

Hive - Window Functions

Hive and Sqoop: CDC

Hive Bucketing Example

Hive File Format and Compression

Hive Join Example

Hive Table - Indexing

Hive Table Using Regex Serde

Import RDBMS Data Using Sqoop

Limitations of HDFS

Map Reduce Algorithm

Oozie - Incremental Table Load Workflow

Performance Enhancement of MR Jobs

Scheduling Job Using Oozie

Setup Eclipse for Hadoop MapReduce

Development

Storage Format and Compression

Submit MapReduce Job

Verifying Zookeeper

YARN Resource Allocation

Yelp Academic Dataset

Machine Learning

Anaconda on windows

CUDA, OpenCL and OpenGL

Learning resources

Stanford NLP

Tensorflow with GPU

Xgboost for Python in MacOS

Miscellaneous

AWS SSH Tunneling

Create A Big Data Sandbox

Create NFS Sharable Directory

Getting Started with Solr

Install Scala on CentOS

Passwordless SSH and SCP

Python Useful Commands

Scala / Java Commons

Setup the VM

Sign up for a dev account of twitter

Useful AWS Commands

Useful Git Commands

Useful Java Tips

Useful Linux Commands

Useful MySQL Commands

VirtualBox Commands

Statistical Data Analysis using R

- 01 CRAN packages
- 02 Data Sources and Getting Data into R
- 03 Slice and Dice in R
- 04 Joining Datasets using R
- **05 Create Composite Variable**
- 06 Grouping and Aggregation in R
- 07 Sampling using R
- 08 Statistics for Single Variable
- 09 Working with Missing Data
- 10 Correlation Analysis
- 11 Working with colors in R
- 12 Plotting Variables in R
- 13 Association Plots in R
- 14 Plotting Heatmap
- 15 Overlaying plots
- 16 Outlier Analysis
- 17 Data Transformation of a Variable
- 18 Tidy Data
- 19 Writing Function in R

Appendix 01 Benchmarking R Performance

Appendix 02 Machine Learning Resources

ggplot

- ML 01 Linear Regression
- ML 02 Classification Metrics

Sitemap

G+

Navigation

 $\underline{Sign\ in}\ \mid\ \underline{Recent\ Site\ Activity}\ \mid\ \underline{Report\ Abuse}\ \mid\ \underline{Print\ Page}\ \mid\ \underline{Powered\ By}\ \ \underline{\textbf{Google\ Sites}}$