Date functions

current_date()

date_add (date, int) - Return the date that is int days after date.

date_diff (date1, date2) - Return the number of days b/w date1 & date2.

year (date)
month (date) } Return the year, month, day part of date. (STRING)
day (date)

Integer functions -

abs (u) - Returns the absolute value of u.

round (u, d) - Rounds 'u' to 'd' decimal places.

ceil (u), floor (u): Returns the smallest (largest) integer not smaller (greater) than u.

rand([seed]): Returns a random number

pmod (u, m): Returns the positive modulo 'u modulo m'.

String functions

length (String A)

ltrim (String A) rtrim (String A) - left (or) right trimming

lower (String A) upper (String A) — Converts lower to uppercase.

ARCER

RDD

map((func)). transform the RDD by applying func to each element

filter((func))

reduce((func)) - Aggregate

count()

MERGE

Pandas

read_csv()

describe() - Generates descriptive statistics of dataframe.

groupby() - Group dataframe using a mapper (or) by a series of columns

merge() } merge (or) concatenate Dataframes along a
concat() } particular axis

RDDline

Numpy functions

numpy.array() - create a numpy array.

numpy. zeros(), ones(), empty() - Array filled with 0, 1, empty.

numpy.arange() - create an array with regularly incrementing

numpy.dot() - Dot product of 2 arrays.
          values

numpy.reshape() - Change the shape of an array without changing
          its data.

AZADI

Spark Sessions -

CreateDataFrame(): Creates a Dataframe from an RDD, list or a
          Pandas. DataFrame.

read(-) - Interface for reading data.

sql(query) - Returns a DF representing result of given sql query

table(tableName): Return DF representing associated with the
          given table.

stop(): Stop the underlying context.

List

append, remove, pop, count, sort

spark session API

RDD, Dataframe, Dataset, spark SQL, streaming