Writing User Defined Functions

Thomas M. Henson

@henson_tm | www.thomashenson.com

pluralsight

Overview

Grunt Shell

User Defined Functions

Piggy Bank

Sharpen Pig Latin Skills

pluralsight

Pig Editor

pluralsight

Sandbox = Pig Editor

Pig Grunt Shell

Interactive shell that allows users to enter Pig Latin scripts & interact with HDFS

pluralsight

How to Access

When using Pig locally or sandboxed version

Use: $pig –x local

When using Pig for multiple nodes

Use: $pig –x mapreduce

pluralsight

Basic Grunt Shell Commands

Help	List Directory Contents	Change Directory
>help	**>ls**	**>cd**

pluralsight

Grunt Shell

Command Prompt

SSH into Sandbox

pluralsight

Why Do This?

```
weather_raw = LOAD '/user/hue/weather.csv' USING
PigStorage(',') AS (id:chararray, name:chararray,
date:datetime, day_high:int, day_low:int);


X = FOREACH weather_raw GENERATE day_high * 1.8 +32;


DUMP X;
```

pluralsight

When We Can Do This!

```
weather_raw = LOAD '/user/hue/weather.csv' USING
PigStorage(',') AS (id:chararray, name:chararray,
date:datetime, day_high:int, day_low:int);


X = FOREACH weather_raw GENERATE ToFehrenheit(high_temp );


DUMP X;
```

pluralsight

Pig UDF

User defined functions are custom code written to specify custom processing in Pig Latin.

pluralsight

UDF Languages

- Java
- Python
- JavaScript
- Ruby
- Groovy

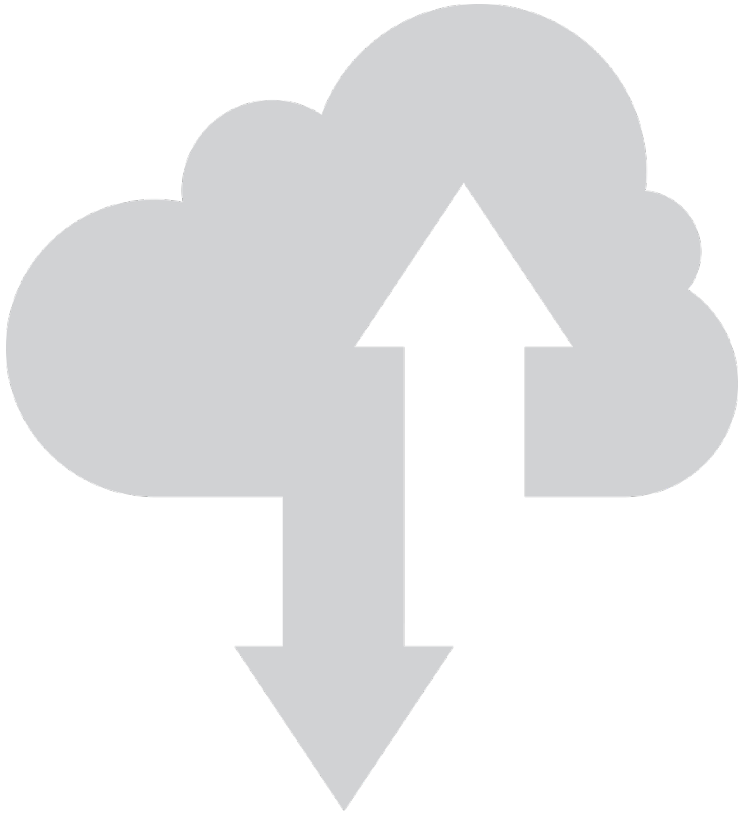
pluralsight

UDF Demo

Sample UDF

pluralsight

Piggy Bank



pluralsight

Piggy Bank

Pig.apache.org

SVN

pluralsight

Congratulations

Basic Pig Latin

pluralsight

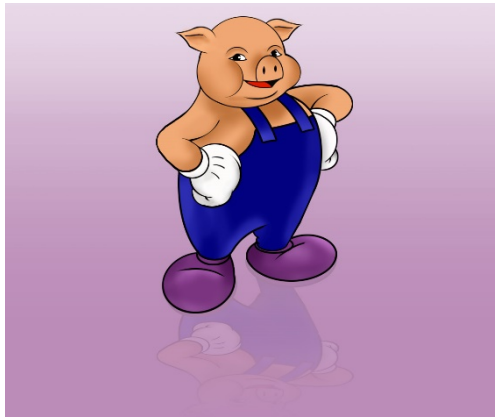
Keep Learning

- Join the mailing list
  - user-subscribe@pig.apache.org
- Documentation is your Friend
- Answer Questions on Discussion Post
- Commit to Piggy Bank
- Contribute to the Pig Project

pluralsight

Summary

That's all folks.



Grunt Shell Basics

Looked at Piggy Bank

Developed User Defined Function

How to get Involved

pluralsight