
DO WE HAVE ANY ACTIONABLE ANALYTICS FROM OUR BIG DATA IN THE CLOUD?
YES, THE DATA SHOWS THAT MY PRODUCTIVITY PLUNGES WHENEVER YOU LEARN NEW JARGON.
MAYBE IN-MEMORY COMPUTING WILL ACCEL-ERATE YOUR APPLICA-TIONS.
PLUNGE, PLUNGE, PLUNGE.
Dilbert.com DilbertCartoonist@gmail.com
1-9-13 ©2013 Scott Adams, Inc./Dist. by Universal Uclick

Easy Time Series Analysis
with
Riak TS, Python, Pandas, and Jupyter Notebook

riakTS

What is Time Series?

What is Riak TS?

Creating Tables

Populating with Data

Querying

Visualizing Queries

Log File Entries

Weather Data

Patient Vitals

Tidal Measurements

Sensor Data

Stock Quotes

Meter Readings

Scientific Observations

Etc.

Highly Available

Scalable

Fault Tolerant

Operationally Simple

And

Optimized for
time series data using
time quantization to
achieve data locality

Creating a TS Table

```
CREATE TABLE WaterMeterData (
    customer_id        varchar   not null,
    meter_id           varchar   not null,
    time_stamp         timestamp not null,
    water_pressure     double    not null,
    gallons_per_hour   double    not null,
    total_gallons      double    not null,
    fault_condition    boolean   not null,
    fault_message      varchar,
    PRIMARY KEY(
      (customer_id, meter_id, quantum(time_stamp, 30, 'd')),
       customer_id, meter_id, time_stamp)

)
```

Creating a TS Table

```
CREATE TABLE WaterMeterData (
    customer_id        varchar   not null,
    meter_id           varchar   not null,
    time_stamp         timestamp not null,
    water_pressure     double    not null,
    gallons_per_hour   double    not null,
    total_gallons      double    not null,
    fault_condition    boolean   not null,
    fault_message      varchar,
    PRIMARY KEY(
      (quantum(time_stamp, 30, 'd')),
       time_stamp)

)
```

Day
Hour
Minute
Second

varchar
double
sint64
boolean
timestamp

A subset of SQL designed to make range queries easy

```
SELECT
      *
FROM
      WaterMeterData
WHERE
      time_stamp >= 1464739200000 AND
      time_stamp < 1467334800000 AND
      customer_id = 'CUSTOMER-0001' AND
      meter_id = 'METER-0001'
```

COUNT()

SUM()

MEAN()

AVG()

MIN()

MAX()

STDDEV_SAMP()

STDDEV_POP()

Java
.Net
Ruby
Python
Node.js
Erlang

PHP

http://docs.basho.com/riak/ts/

Python

Riak Python Client

Jupyter – jupyter.org

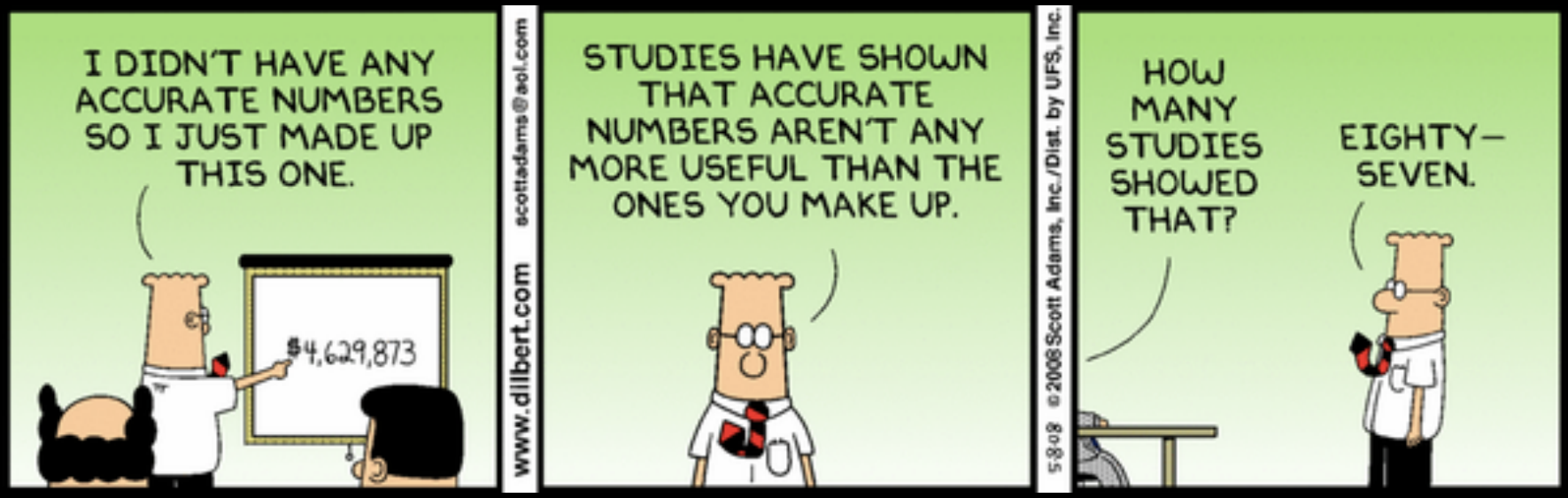
Pandas – pandas.pydata.org

Pandas – pandas.pydata.org

Python Data Analysis Library - an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

http://github.com/cvitter

http://bit.ly/24woJML


I DIDN'T HAVE ANY ACCURATE NUMBERS SO I JUST MADE UP THIS ONE.
$4,629,873
www.dilbert.com scottadams@aol.com
STUDIES HAVE SHOWN THAT ACCURATE NUMBERS AREN'T ANY MORE USEFUL THAN THE ONES YOU MAKE UP.
5-8-08 ©2008 Scott Adams, Inc./Dist. by UFS, Inc.
HOW MANY STUDIES SHOWED THAT?
EIGHTY-SEVEN.

Craig Vitter
Solutions Architect
cvitter@basho.com
@craigvitter