

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



What is Time Series?

What is Riak TS?

Creating Tables

Populating with Data

Querying

Visualizing Queries



#BOOM

Log File Entries

Weather Data

Patient Vitals

Tidal Measurements

Sensor Data

Stock Quotes

Meter Readings

Scientific Observations

Etc.



riakTS

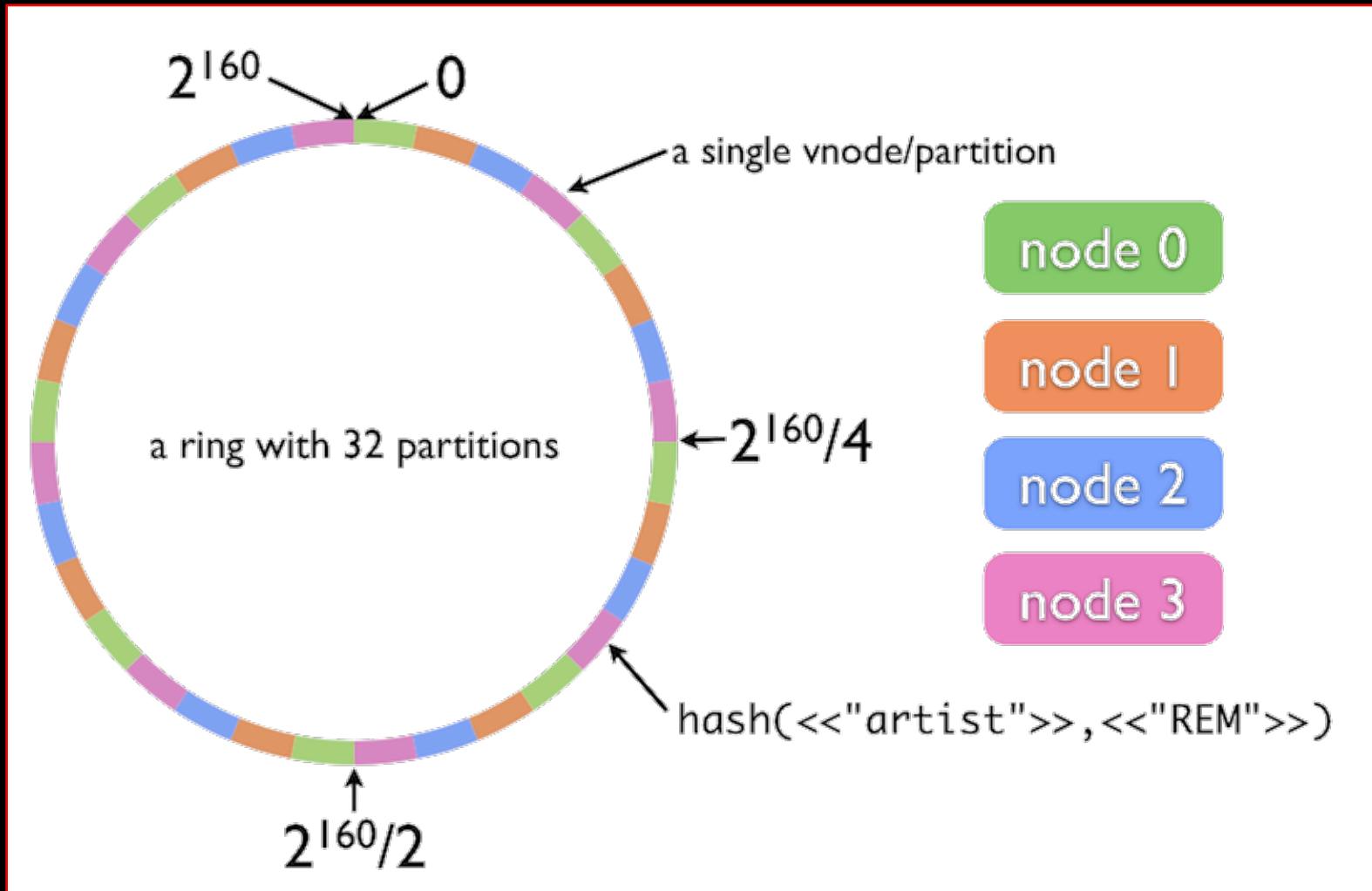
**Highly Available
Scalable
Fault Tolerant
Operationally Simple
And**

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

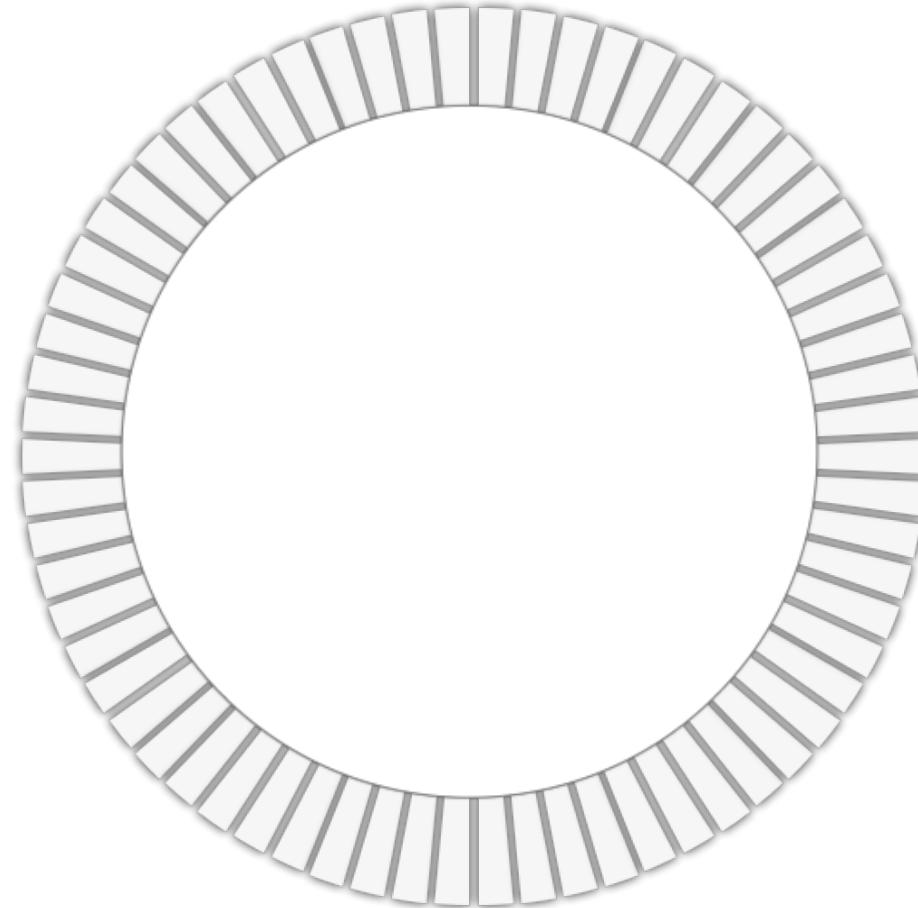
Creating a TS Table

```
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(
    (quantum(time_stamp, 30, 'd')),
    time_stamp, customer_id, meter_id)
)
```

Consistent Hashing

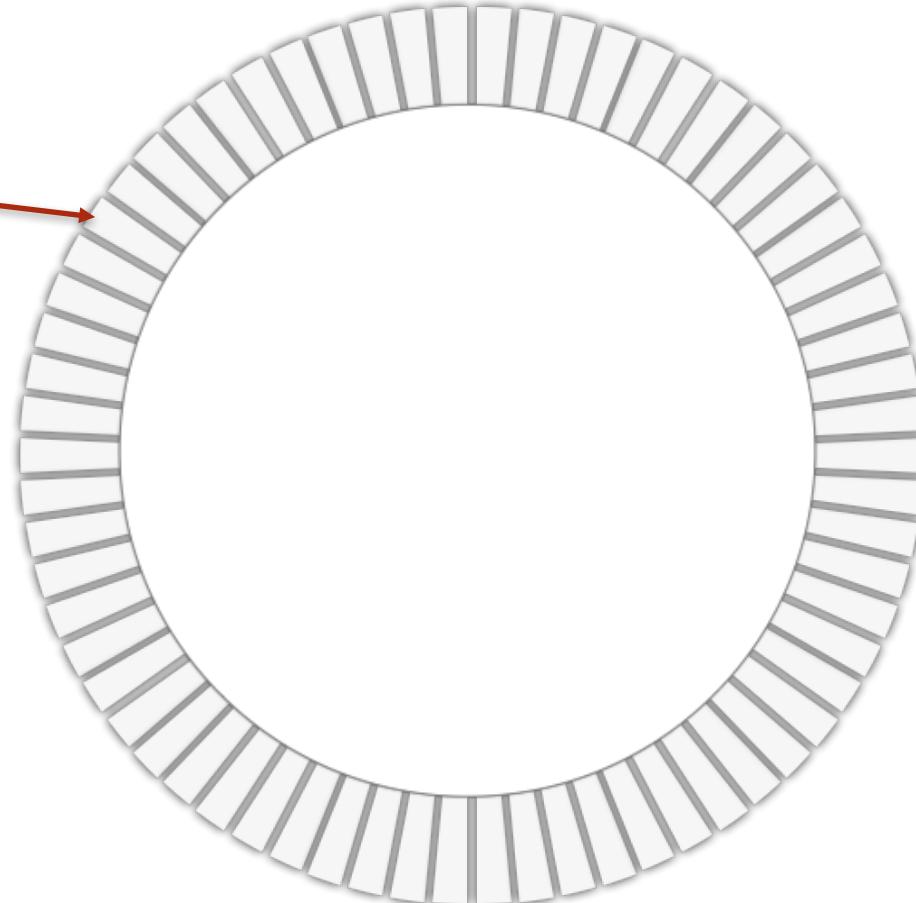


Consistent Hashing

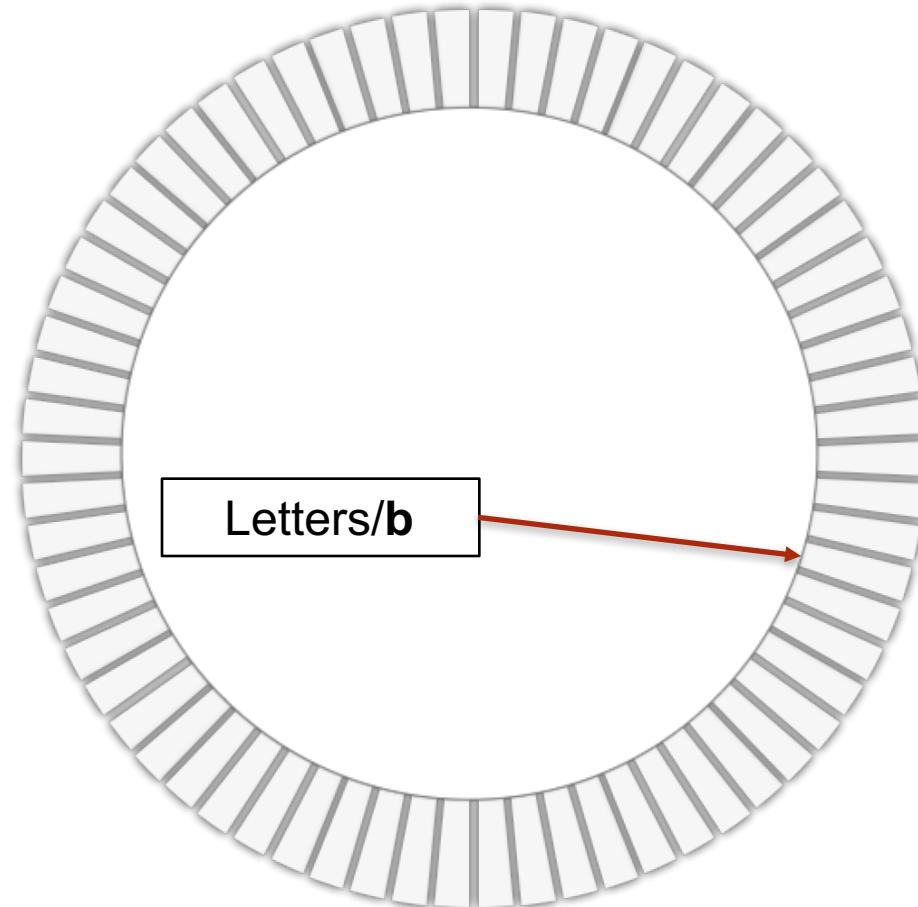


Consistent Hashing

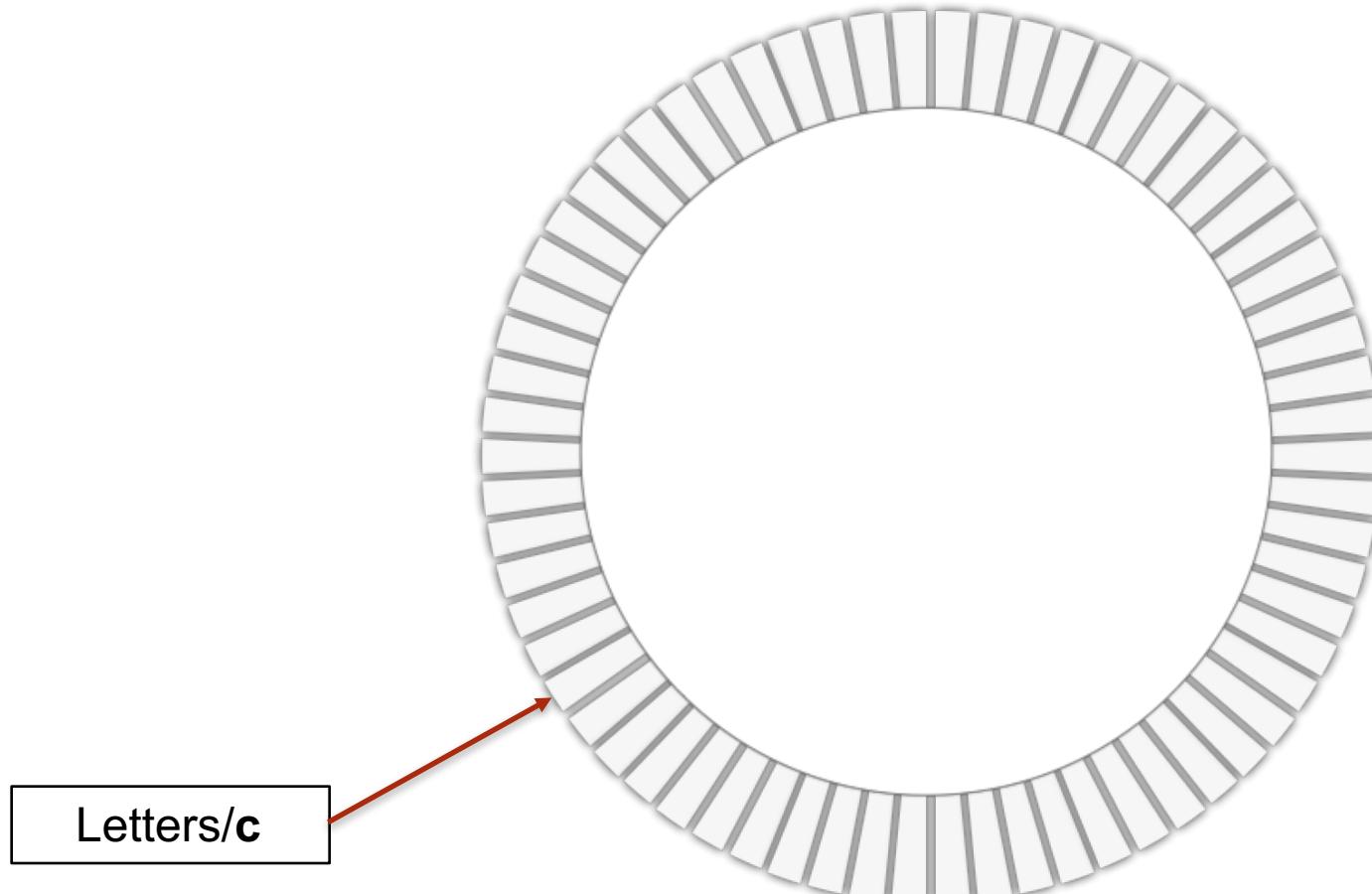
Letters/a



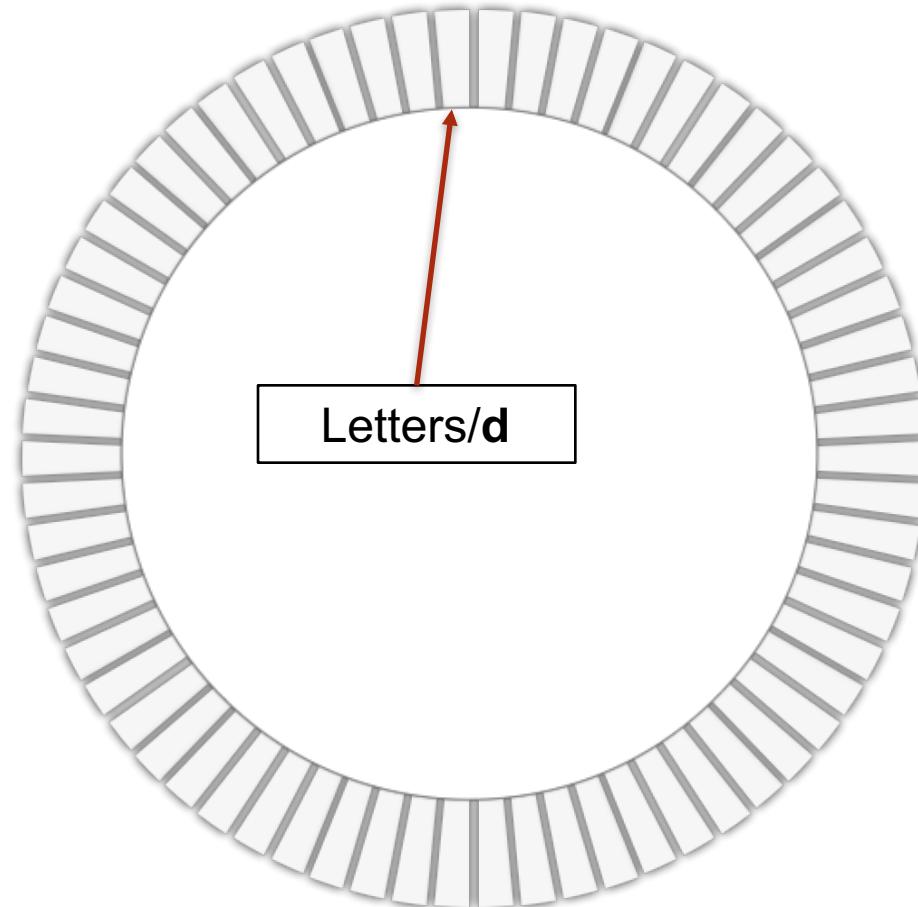
Consistent Hashing



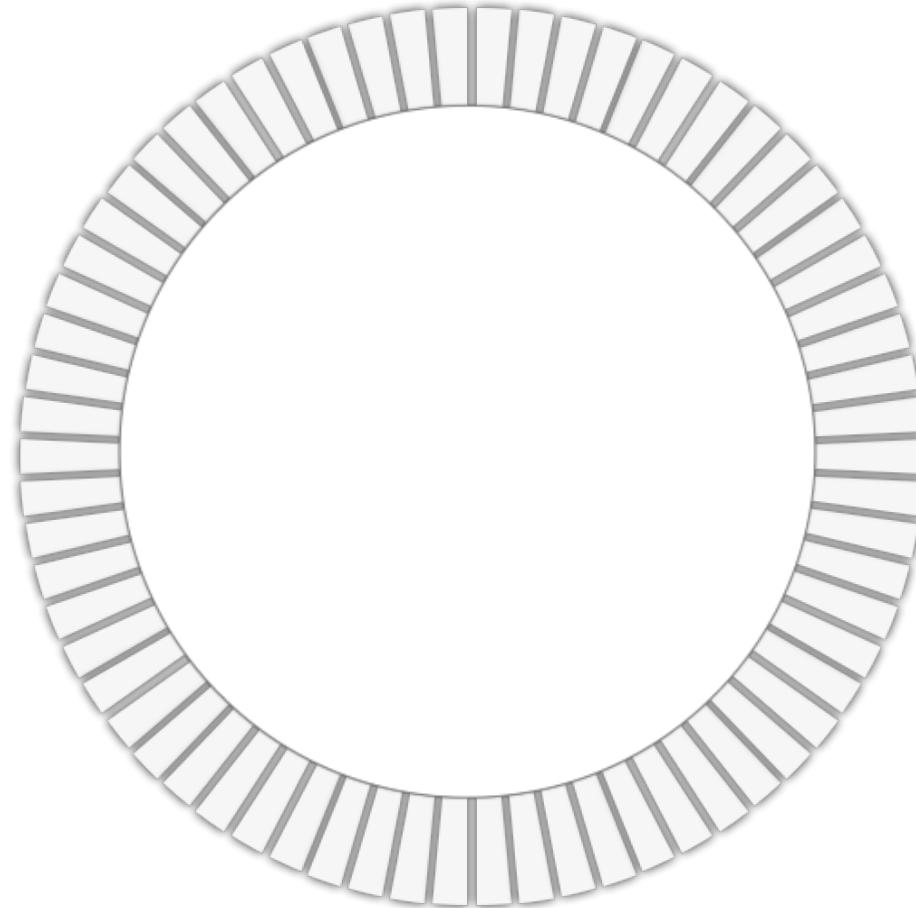
Consistent Hashing



Consistent Hashing

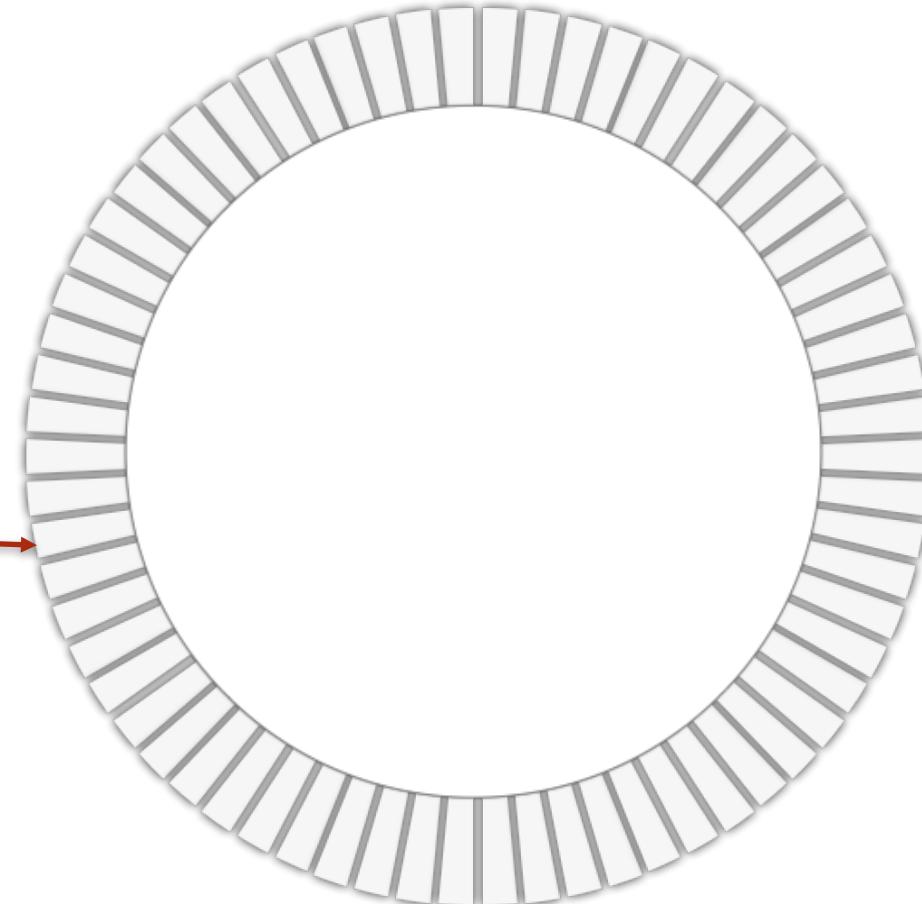


Data Co-Location



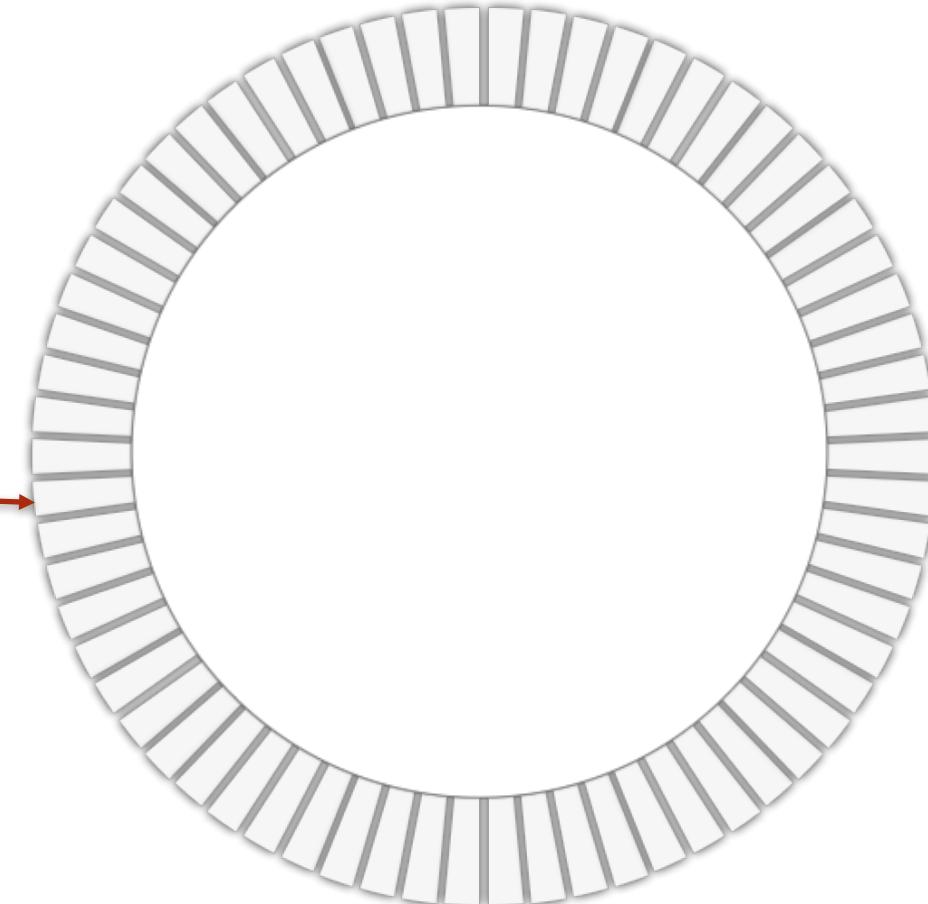
Data Co-Location

WaterData/12-01-2015



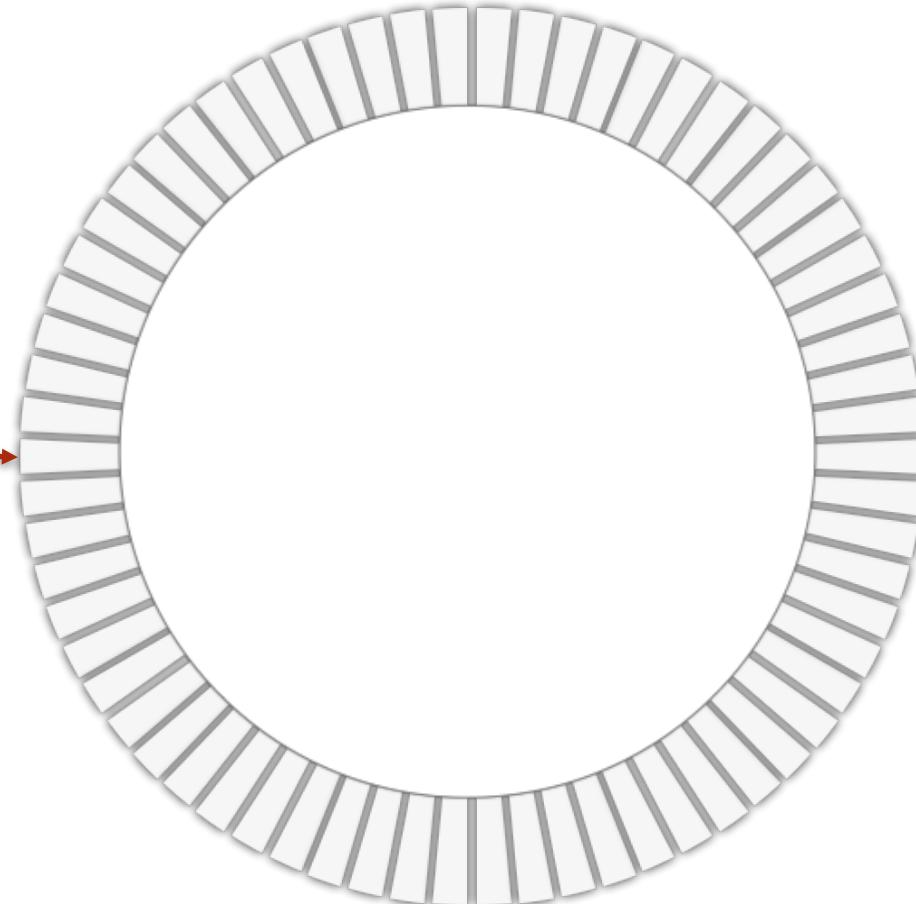
Data Co-Location

WaterData/12-02-2015



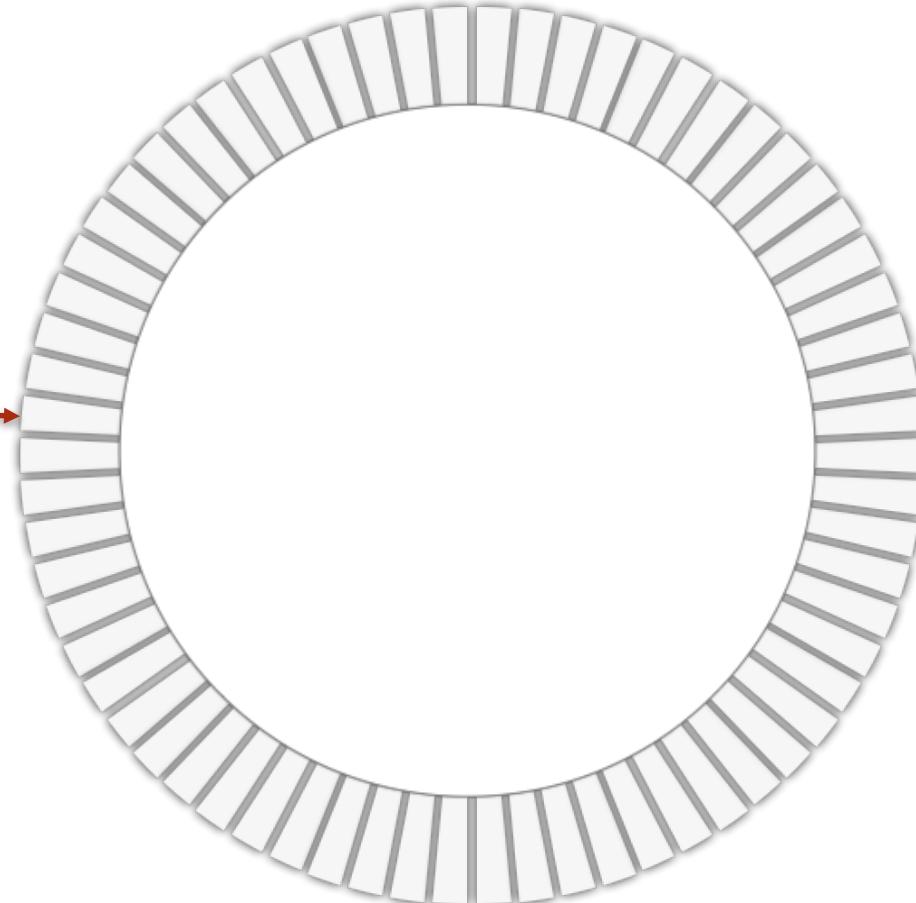
Data Co-Location

WaterData/12-03-2015



Data Co-Location

WaterData/12-04-2015



**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**

spark

<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>



Craig Vitter
Solutions Architect
cvitter@basho.com
[@craigvitter](https://twitter.com/craigvitter)