

# Squigglee Summary (Confidential)



## Squigglee

Contact:  
Dr. Ramesh K. Raghunathan  
ramesh.k.raghunathan@gmail.com  
214.620.1863





# What is Squigglee?

*...Time Series Done Right*

*Squigglee*

- A new time series platform
- Single offering to manage the entire lifecycle of time series data
- A collection of time series managers that store and share any data type at any frequency
- Lowest cost of ownership at large scale
- Seamless integration via data virtualization and APIs

# Squigglee's Pitch



*...Time Series Done Right*

## What is the market for time series data management?

- Enormous enterprise target market (e.g. 10000 customers X \$1000 per TB X 100 TB = \$1 billion)
- Currently large enterprises incur costs from \$10k to 60k per TB to manage such data
- Many domain specific vendors & software each with their own feature sets & product road-maps
- A lot of time series data is not being targeted for the cloud due to privacy & data locality challenges

## Why should you enter this market?

- No other services competitor has such a dedicated offering
- Disrupt the legacy vendors & rapidly gain market share
- Attract new customers, new / additional data from extant customers
- Improve margins, negate competitor margin pressure
- Be the first to introduce a generic time series management product

## Why should it leverage Squigglee to enter this market?

- Applicable for all domains & industries
- Addresses many unaddressed customer pain points - ownership, sharing, privacy, data locality, fast real-time response across multiple data centers, pattern matching, data synopses
- Fast time to market for since version 1 is already built

Dimensionality  
Easier Operation  
New Customers  
Storage  
Lower Total Cost of Ownership  
Indexing  
More Revenues  
Unlimited Scale  
Node New Workflows  
Diversification  
Hybrid Deployment  
Data Management  
Approximate Query Processing

Complex Topologies  
Location  
Patterns  
Sampling  
Sketching  
Cloud  
Cluster  
Motif  
Synapses  
Replication  
Retrieval  
Instant Retrieval

## Time Series

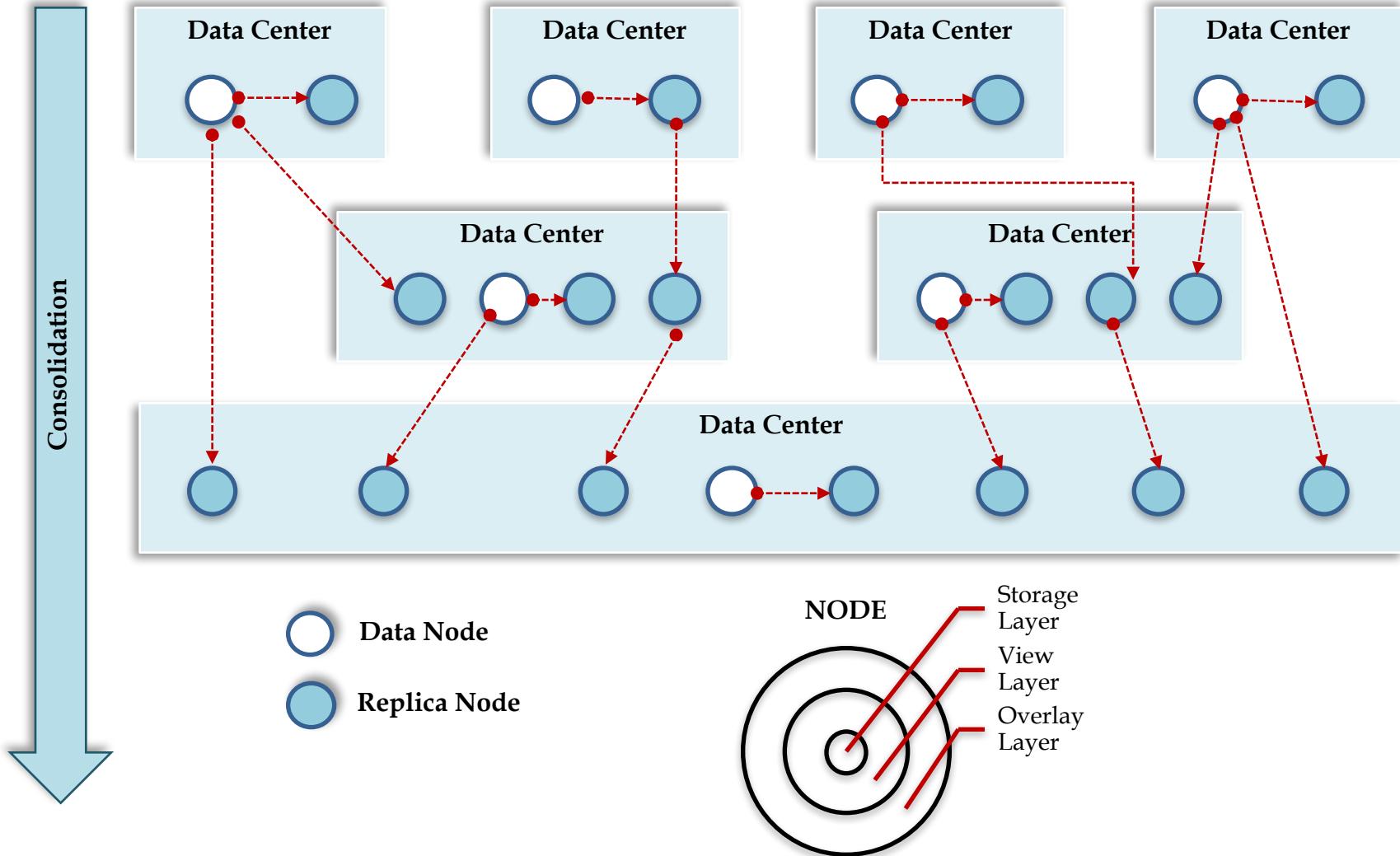
# Squigglee Basics





# Deploying Squigglee

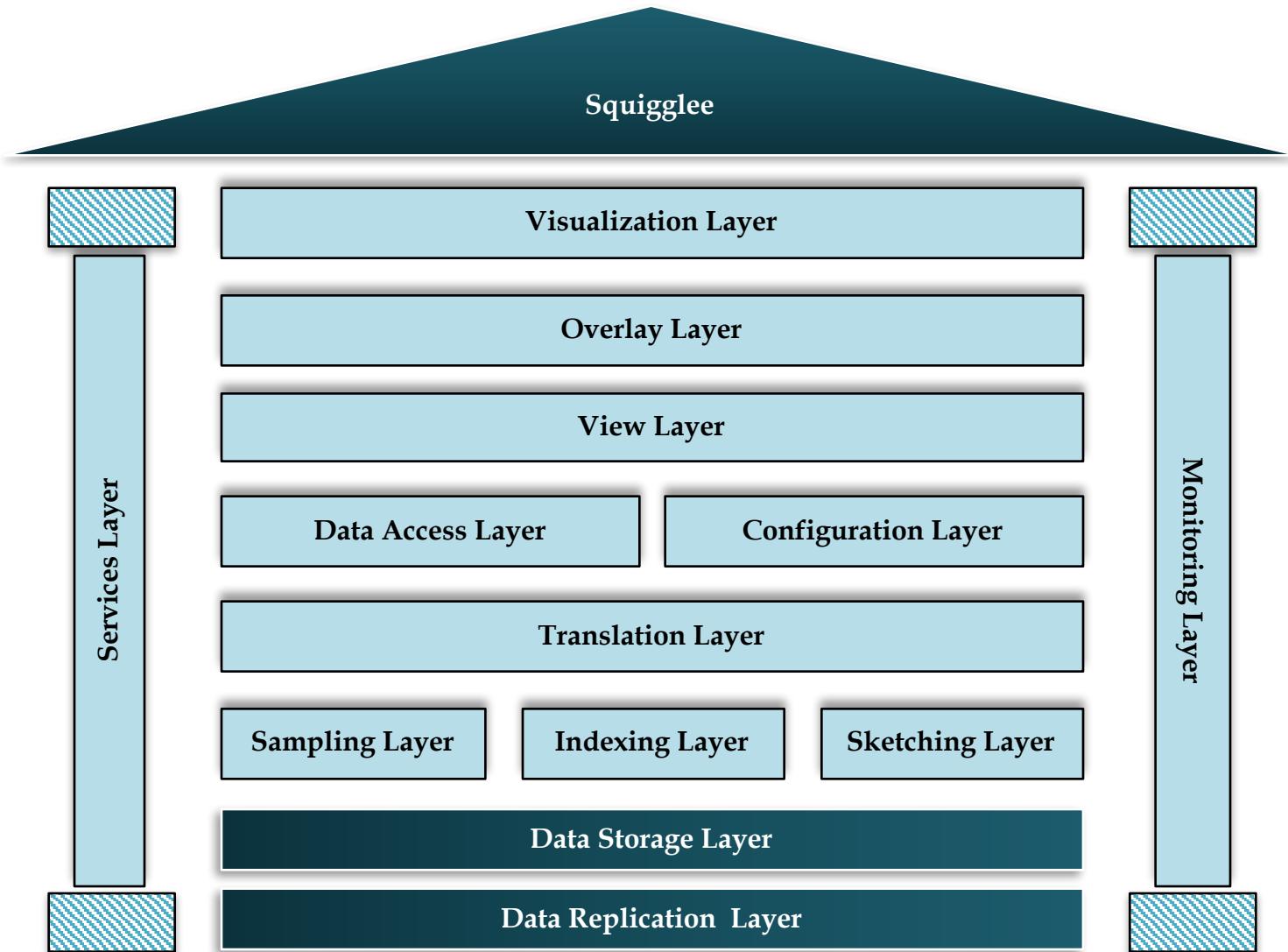
*...Time Series Done Right*



# Squigglee Architecture



*...Time Series Done Right*



# Fundamental Usage Patterns



*...Time Series Done Right*

## Deployment

- Users create time series managers across data centers
- Users group time series managers to create replica sets
- Users decide who can query or access their data (entitlements)
- Users manage data and time series manager lifecycle, including termination and removal

## Configuration

- Users configure time series data sets at a time series manager (unique name, frequency, data type, and duration)
- Users upload data into the cluster, update or delete existing data
- Users configure indexes for pattern matching
- Users configure sketches for data synopses

## Retrieval

- Users retrieve time series data from time series managers per entitlements
- Users query for matches for a given pattern within a search radius
- Users capture and store patterns of interest for use by entitled users
- Entitled users select multiple time series from multiple locations for matching and querying

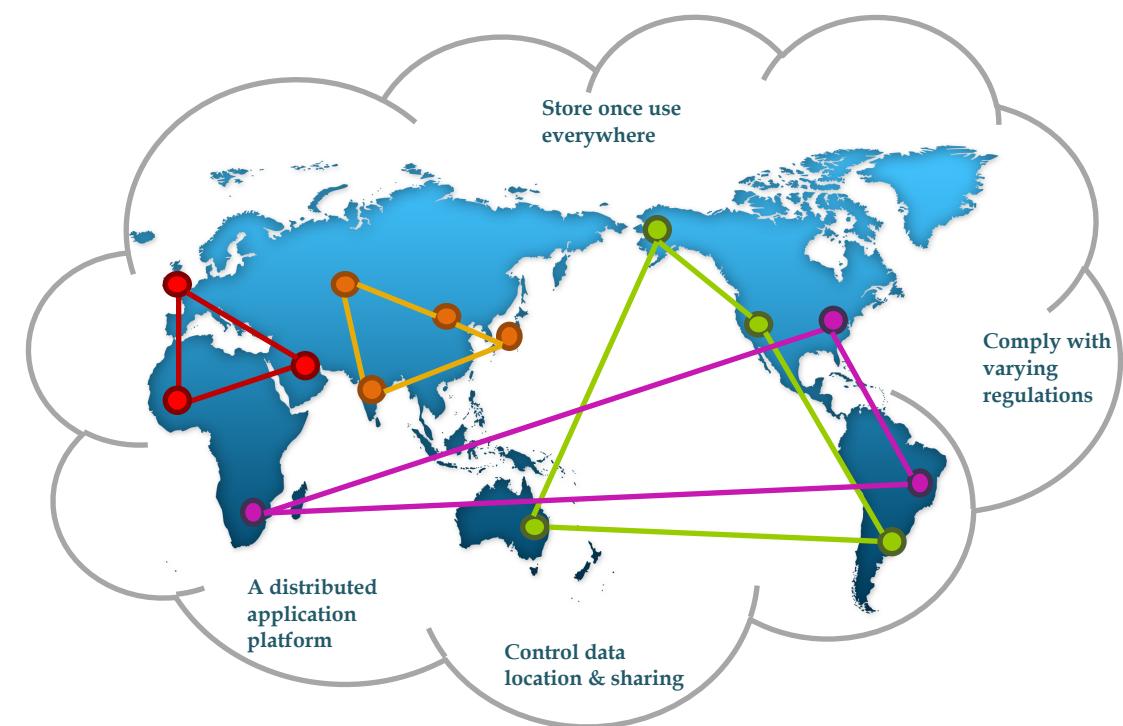
## Synopses

- Users obtain complete summary statistics for sketched time series
- Users issue queries on sampled data
- Users issue point, range, and inverse queries on sketched time series
- Users obtain the frequency distribution for sketched or sampled data



# Squigglee Deployment Scenarios

*...Time Series Done Right*



## Sample Industry Applications

1. A conglomerate can ensure data replication occurs only to permitted or desired locations across countries (pursuant to local government regulations) while still enabling entitlement driven shared access.
2. An oil field services company accessing its own shared data for queries from multiple client locations, each of whom may have different data center providers (e.g. BP or Shell data centers located in Amazon or Google cloud locations respectively)
3. Governmental agencies sharing data for queries while still maintaining ownership and control over their data.
4. Medical industry institutions sharing data for specific research purposes without having to violate any patient or country health regulation.

Dimensionality  
Easier Operation  
New Customers  
Storage  
Lower Total Cost of Ownership  
Indexing  
More Revenues  
Unlimited Scale  
Node New Workflows  
Diversification  
Hybrid Deployment  
Data Management  
Approximate Query Processing

Complex Topologies  
Location Patterns  
Sampling  
Sketching  
Cloud  
Motif  
Synapses  
Replication  
Cluster  
Retrieval  
Instant Retrieval

## Time Series White Space Innovation

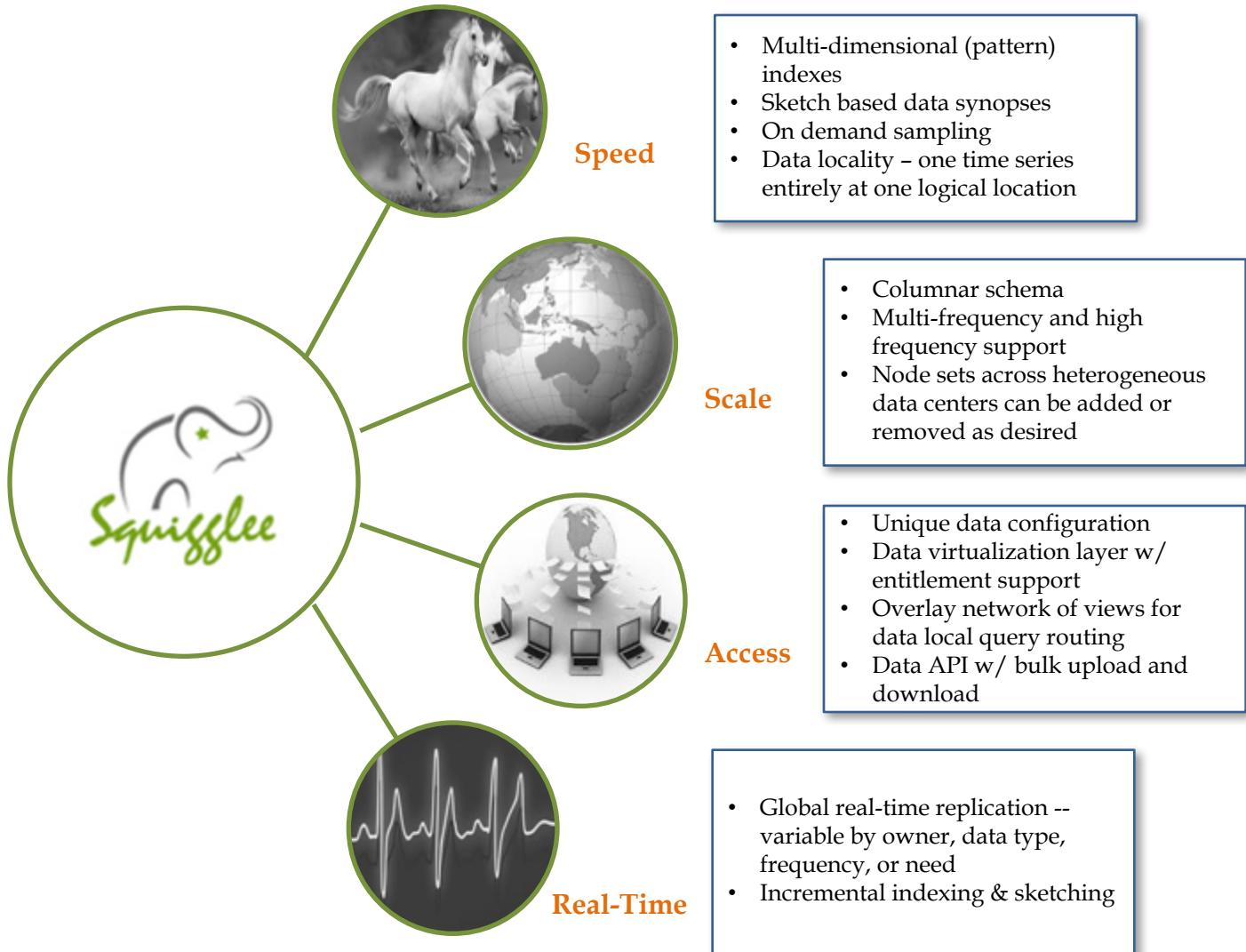
# Squigglee Value Proposition



# Squigglee Differentiators



*...Time Series Done Right*



# Value Proposition to Investors

*...Time Series Done Right*

- Time Series data management needed across all verticals
- Thousands of companies routinely manage such data in excess of 100 TB
- Companies license multiple products & employ numerous workarounds to manage such data

- Unique IP  
Patent Pending
- Numerous Differentiators  
Scale  
Speed  
Real-time  
Access

- Flexible Pricing Model  
Pass hosting costs to customer  
Metered cloud hosted instances
- Cloud, On-premise, and Hybrid licensing options

Huge Market



Strong Product



Right Pricing



Massive Revenue

On Cloud – can hit gross profits of \$20 million annually at pricing \$6000/TB & \$40 million at \$8000/TB with just 100 customers managing 100 TB

On Premise – can hit gross profits of \$20 and \$40 million annually at pricing of \$1600 and \$3300/TB respectively for 100 customers



# Value Proposition to Customers

*...Time Series Done Right*

## Top Line Benefits

Introduce new integrated workflows & capabilities

Share real-time data with full control over data lifecycle

Dramatically decrease response times for extreme scale time series data using indexing, sketching, and sampling

Experience benefits, in real-time, globally across heterogeneous data centers

## Bottom Line Benefits

Experience dramatic cost savings in managing high volume time series data with costs << \$10,000 per TB

Seamless integration with other repositories and systems

Deploy flexibly in internal, cloud, and hybrid data centers globally with any desired topology

Greatly ease the creation and management of networked time series managers



# Feature Comparison

*...Time Series Done Right*

|                             | Squigglee | KairosDB | Open TSDB | InfluxDB | Tempo DB | Historians | Notes  |
|-----------------------------|-----------|----------|-----------|----------|----------|------------|--|
| Storage & Basic Retrieval   | ✓         | ✓        | ✓         | ✓        | ✓        | ✓          | Squigglee is Cassandra based, KairosDB is either Cassandra or HBase, Open TSDB and InfluxDB are HBase, rest are proprietary                    |
| High Frequency Support      | ✓         | ✗        | ✗         | ✗        | ✗        | ✓          |  |
| Meta-data Framework         | ✓         | ✗        | ✓         | ✓        | ✗        | ✓          | In Squigglee data virtualization provides a meta-data framework far richer than tag-value pairs, which others utilize for grouping & selection |
| Non-numeric data types      | ✓         | ✓        | ✗         | ✓        | ✗        | ✗          | Only Squigglee supports blob time series e.g. for multi-media feeds  |
| Data Replication            | ✓         | ✗        | ✓         | ✓        | ✗        | ✓          | Only Squigglee provides automated configuration driven fine grained replication for complex topologies   |
| Multi-dimensional Retrieval | ✓         | ✗        | ✗         | ✗        | ✗        | ✗          | Exclusive to Squigglee   |
| Sketching & sketch queries  | ✓         | ✗        | ✗         | ✗        | ✗        | ✗          | Exclusive to Squigglee   |
| Summarization               | ✓         | ✗        | ✓         | ✓        | ✓        | ✓          | Squigglee provides a complete frequency distribution via sketches in addition to summary statistics like the others                            |
| Real-time indexing          | ✓         | ✗        | ✗         | ✗        | ✗        | ✗          | Exclusive to Squigglee   |
| On Demand Sampling          | ✓         | ✗        | ✗         | ✗        | ✗        | ✗          | Exclusive to Squigglee   |
| Data Virtualization         | ✓         | ✗        | ✗         | ✗        | ✗        | ✗          | Squigglee provides many view layers including an overlay network for data local query routing  |
| Data APIs                   | ✓         | ✓        | ✓         | ✓        | ✓        | ✓          |  |
| Bulk Data Operations        | ✓         | ✗        | ✓         | ✗        | ✗        | ✓          | Squigglee uses a binary serialization protocol to provide bulk operations and matching clients   |

Dimensionality  
Easier Operation  
New Customers  
Storage  
Lower Total Cost of Ownership  
Indexing  
More Revenues  
Unlimited Scale  
Node New Workflows  
Diversification  
Hybrid Deployment  
Data Management  
Approximate Query Processing

Location  
Patterns  
Sampling  
Sketching  
Cloud  
Cluster  
Motif  
Synapses  
Replication  
Retrieval  
Instant Retrieval

## Time Series

# Squiggle Economics



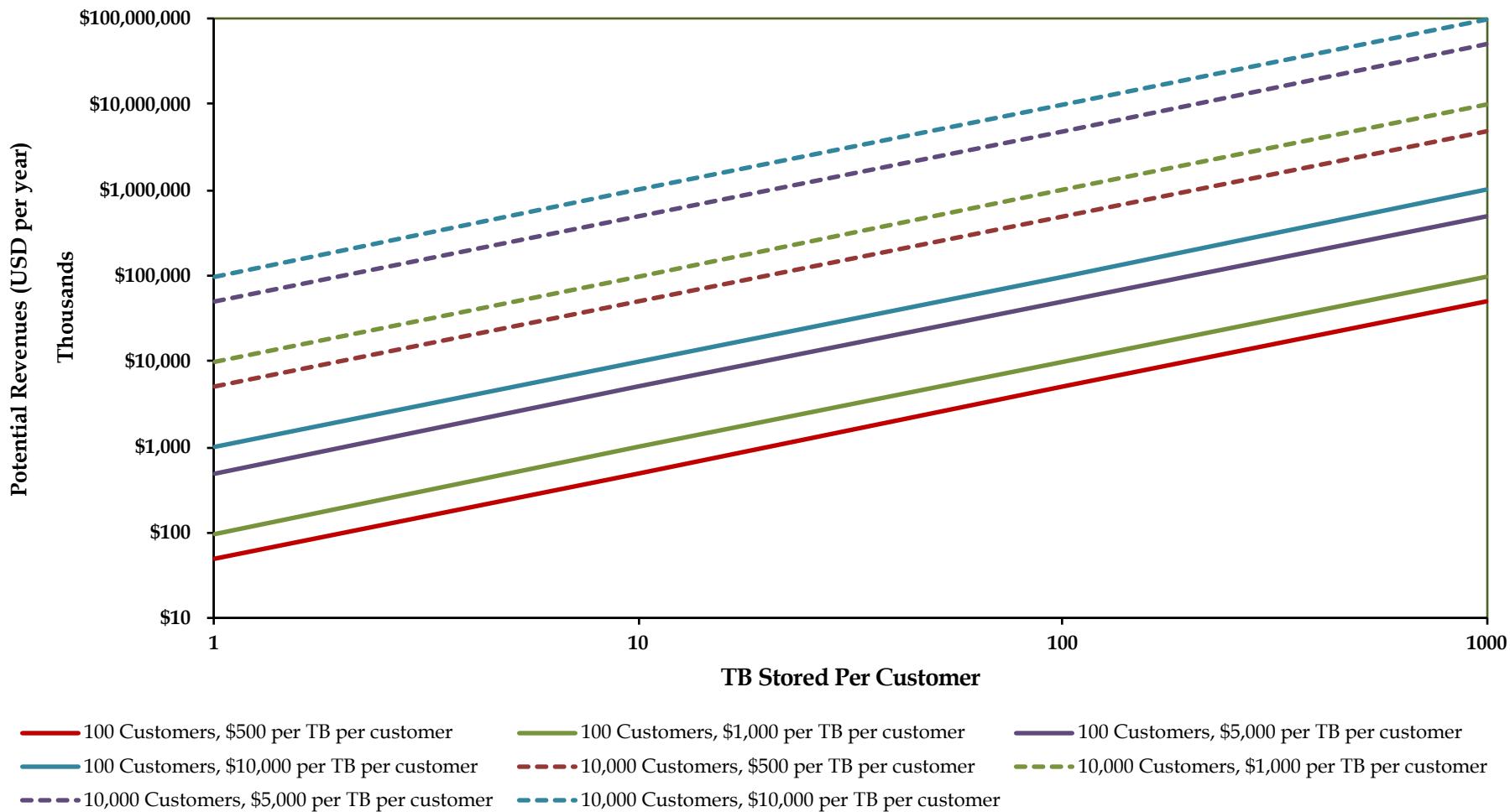
# Revenue Potential



*...Time Series Done Right*

Time Series Offering: Projected Annual Revenues

(Conservative Assumptions: 100 to 10,000 customers with each paying between \$500 - \$10,000 per TB per year)



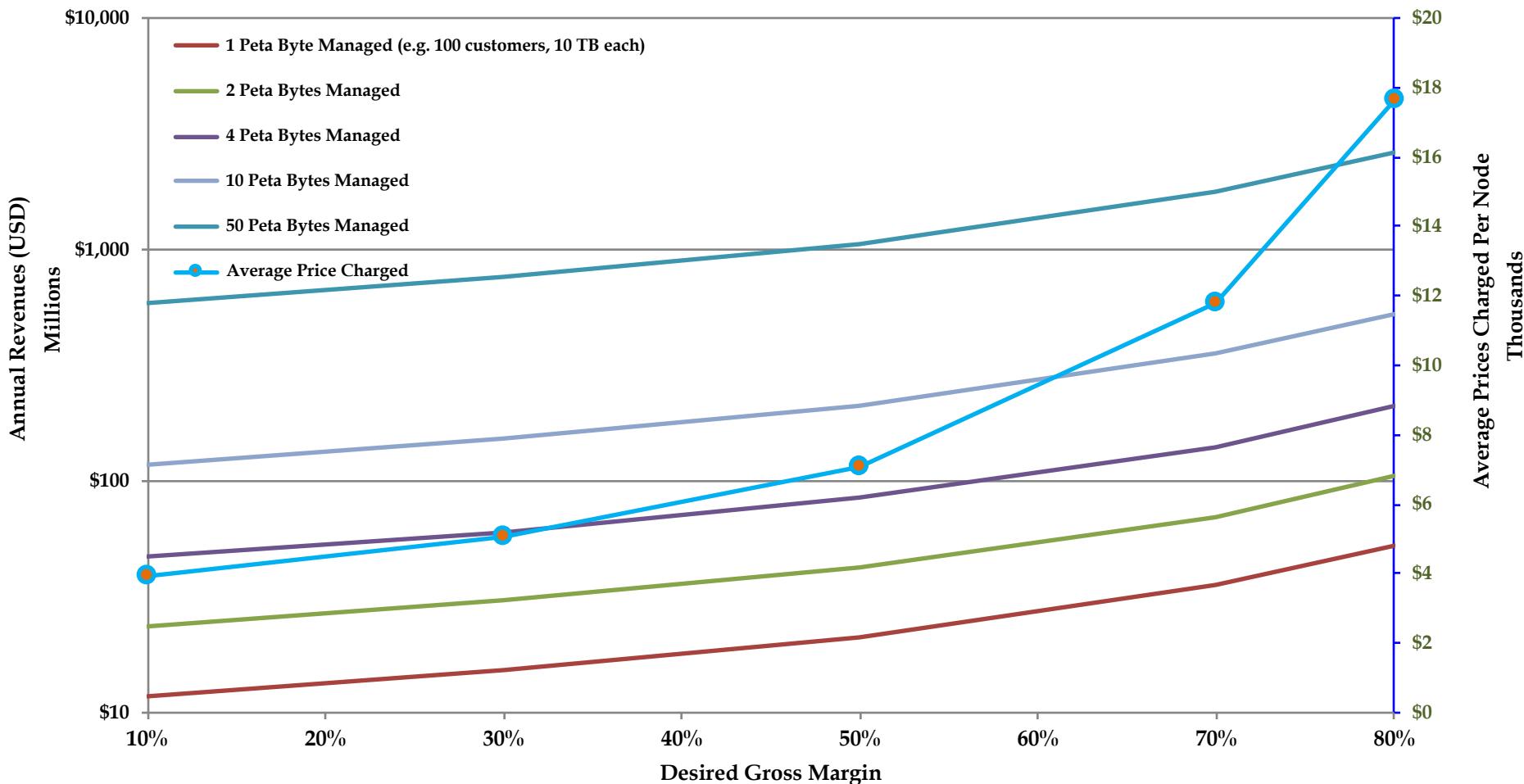
# Detailed Pricing Strategy



*...Time Series Done Right*

## Cloud Offering Pricing Strategy

(Assumptions: Hosting costs ranging from \$150-\$600 per server per month based on server and storage choices)



Dimensionality  
Easier Operation  
New Customers  
Storage  
Lower Total Cost of Ownership  
Indexing  
More Revenues  
Unlimited Scale  
Node New Workflows  
Diversification  
Hybrid Deployment  
Data Management  
Approximate Query Processing

Complex Topologies  
Location Patterns  
Sampling  
Sketching  
Cloud  
Motif  
Synapses  
Replication  
Cluster  
Retrieval  
Instant Retrieval

# Time Series

## White Space Innovation

# Squigglee In Action



# Home



...Time Series Done Right

Squigglee

Home

Operation

Configuration

Retrieval

Synopses



## Operation

- Deploy globally
- Use replica sets to support any topology
- Control data lifecycle
- Share data with others

## Configuration

- Add location as an attribute of time series
- Support multiple and high frequencies
- Support any data type

## Retrieval

- Perform multi-dimensional data retrieval (pattern queries)
- Configure pattern indexes
- Capture and store patterns of interest
- Retrieve any number of matches within a radius

## Synopses

- Configure sketches on large time series
- Query sub-samples on demand
- View summary statistics
- Perform point, range, and inverse estimates



# Operation

...Time Series Done Right

*Squiggle Operations*

Home    Operation    Configuration    Retrieval    Synopses

Add To List    Remove From List  
 Add To Cluster    Remove From Cluster  
 Restart Nodes    Restart Services

| Ln | Address                      | Data Center | Instance Id | Name              | Seed Node? | Bootstrap Node? | Replica Of | Type  | Size | Storage Status | Indexing Status | Overlay Status | Local View | Global View |
|----|------------------------------|-------------|-------------|-------------------|------------|-----------------|------------|-------|------|----------------|-----------------|----------------|------------|-------------|
| 0  | <a href="#">52.4.127.90</a>  | us-east-1   | i-8d95cb70  | TimeSeriesNode_0  | true       | true            | 0          | Small | 250  | Up             | Up              | Up             | Up         | Up          |
| 1  | <a href="#">52.4.16.100</a>  | us-east-1   | i-95c49a68  | TimeSeriesNode_1  | false      | false           | 1          | Small | 250  | Up             | Up              | Up             | Up         | Up          |
| 2  | <a href="#">52.6.226.26</a>  | us-east-1   | i-5fc799a2  | TimeSeriesNode_2  | false      | false           | 1          | Small | 250  | Up             | Up              | Up             | Up         | Up          |
| 3  | <a href="#">52.5.176.5</a>   | us-east-1   | i-aac29c57  | TimeSeriesNode_3  | false      | false           | 1          | Small | 250  | Up             | Up              | Up             | Up         | Up          |
| 4  | <a href="#">52.7.75.120</a>  | us-east-1   | i-76c39d8b  | TimeSeriesNode_4  | false      | false           | 1          | Small | 250  | Up             | Up              | Up             | Up         | Up          |
| 5  | <a href="#">52.4.148.30</a>  | us-east-1   | i-d0c39d2d  | TimeSeriesNode_5  | false      | false           | 1          | Small | 250  | Up             | Up              | Up             | Down       | Down        |
| 6  | <a href="#">52.6.234.36</a>  | us-east-1   | i-02e9b7ff  | TimeSeriesNode_6  | false      | false           | 6          | Small | 250  | Up             | Up              | Up             | Up         | Up          |
| 7  | <a href="#">52.5.63.207</a>  | us-east-1   | i-c4e6b839  | TimeSeriesNode_7  | false      | false           | 6          | Small | 250  | Up             | Up              | Up             | Up         | Up          |
| 8  | <a href="#">52.5.218.145</a> | us-east-1   | i-47e6b8ba  | TimeSeriesNode_8  | false      | false           | 6          | Small | 250  | Up             | Up              | Up             | Up         | Up          |
| 9  | <a href="#">52.4.3.92</a>    | us-east-1   | i-85e6b878  | TimeSeriesNode_9  | false      | false           | 9          | Small | 250  | Up             | Up              | Up             | Up         | Up          |
| 10 | <a href="#">52.1.99.193</a>  | us-east-1   | i-fde7b900  | TimeSeriesNode_10 | false      | false           | 9          | Small | 250  | Up             | Up              | Down           | Up         | Up          |
| 11 | <a href="#">52.5.216.152</a> | us-east-1   | i-0de6b8f0  | TimeSeriesNode_11 | false      | false           | 9          | Small | 250  | Up             | Up              | Up             | Up         | Up          |
| 12 | <a href="#">52.6.227.37</a>  | us-east-1   | i-f9e4ba04  | TimeSeriesNode_12 | false      | false           | 9          | Small | 250  | Up             | Up              | Up             | Up         | Up          |

Updated cluster status at Tue May 05 2015 20:00:07 GMT-0500 (CDT)

# Configuration

*...Time Series Done Right*

0

### Data Configuration

|  |   |                                       |
|--|---|---------------------------------------|
| Time Series ID   | Data Type   | Frequency                             |
| <input type="text"/>   | <input type="button" value="double"/>                 | <input type="button" value="MILLIS"/> |
| Start:   | <input type="text" value="2015-03-17T00:00:00.000Z"/> |                                       |
| End:   | <input type="text" value="2015-03-17T23:59:59.999Z"/> |                                       |
| <input type="button" value="Add"/> <input type="button" value="Remove"/> |   |                                       |

| ID             | Data Type | Freq    | Start                    | End                      |
|----------------|-----------|---------|--------------------------|--------------------------|
| S_MINUTES      | int       | MINUTES | 2000-01-01T00:00:00.000Z | 2999-12-31T23:59:00.000Z |
| S_YEARS        | int       | YEARS   | 2000-01-01T00:00:00.000Z | 2999-01-01T00:00:00.000Z |
| S_HOURS        | int       | HOURS   | 2000-01-01T00:00:00.000Z | 2999-12-31T23:00:00.000Z |
| S_MICROS       | int       | MICROS  | 2014-11-22T00:00:00.000Z | 2014-11-22T00:00:59.999Z |
| EKG_Sample1    | double    | MILLIS  | 2014-11-22T00:00:00.000Z | 2014-11-22T23:59:59.999Z |
| ZIPF_Sample1   | int       | MILLIS  | 2014-11-22T00:00:00.000Z | 2014-11-22T23:59:59.999Z |
| S_DAYS         | int       | DAYS    | 2000-01-01T00:00:00.000Z | 2999-12-31T00:00:00.000Z |
| S_SECONDS      | int       | SECONDS | 2014-01-01T00:00:00.000Z | 2014-12-31T23:59:59.000Z |
| aud_cad_bid    | double    | MILLIS  | 2014-01-01T00:00:00.000Z | 2014-01-31T23:59:59.999Z |
| S_NANOS        | int       | NANOS   | 2014-11-22T00:00:00.000Z | 2014-11-22T00:00:00.999Z |
| S_MILLIS       | int       | MILLIS  | 2014-11-22T00:00:00.000Z | 2014-11-22T23:59:59.999Z |
| aud_cad_ask    | double    | MILLIS  | 2014-01-01T00:00:00.000Z | 2014-01-31T23:59:59.999Z |
| aud_cad_spread | double    | MILLIS  | 2014-01-01T00:00:00.000Z | 2014-01-31T23:59:59.999Z |

### Index Configuration

|  |                      |                      |
|--|----------------------|----------------------|
| Time Series ID   | Dimensionality       | Projections          |
| <input type="text" value="S_MINUTES"/>                                   | <input type="text"/> | <input type="text"/> |
| Size   | Scalar               | Bucket Width         |
| <input type="button" value="Add"/> <input type="button" value="Remove"/> |                      |                      |

| Time Series ID | Index ID                |
|----------------|-------------------------|
| EKG_Sample1    | ptrn_16_1000_100_8_1000 |
| ZIPF_Sample1   | ptrn_32_1_100_8_1       |
| aud_cad_ask    | ptrn_24_1_100_8_10000   |

### Sketch Configuration

|  |                                   |                      |
|--|-----------------------------------|----------------------|
| Time Series ID   | Sketch Type                       | Cardinality          |
| <input type="text" value="S_MINUTES"/>                                   | <input type="button" value="CM"/> | <input type="text"/> |
| Size   | Top k                             | Counter Width        |
| <input type="text"/>   | <input type="text"/>              | <input type="text"/> |
| <input type="button" value="Add"/> <input type="button" value="Remove"/> |                                   |                      |

| Time Series ID | Sketch ID        |
|----------------|------------------|
| ZIPF_Sample1   | skchEX_1000_1_10 |

# Retrieval



# ...Time Series Done Right

The figure shows the Squiggle Retrieval interface with several panels and annotations:

- Top Navigation:** Home, Operation, Configuration, Retrieval, Synopses.
- Left Panel (6):** Time Series ID: SampleZipf, Window Size: 50. Dimensions: 24, DataType: int. Start: 2015-05-21T00:00:00.000Z, End: 2015-05-21T23:59:59.999Z. Stored Patterns: (dropdown) X. Distance: 2. Pattern Name: [empty]. Buttons: Add, Remove, Match, Capture. Table:

| Ln | Id          | Start                    | End                      |
|----|-------------|--------------------------|--------------------------|
| 1  | SampleZipf  | 2015-05-21T00:00:00.000Z | 2015-05-21T23:59:59.999Z |
| 3  | SampleZipf3 | 2015-05-21T00:00:00.000Z | 2015-05-21T23:59:59.999Z |
| 5  | SampleZipf5 | 2015-05-21T00:00:00.000Z | 2015-05-21T23:59:59.999Z |
- Center Panel (0):** A main visualization area showing a time series plot with orange dots and green lines. The x-axis is labeled "Data Frequency: MILLIS" and ranges from 0 to 45. The y-axis ranges from 0 to 800. Below the plot, five time intervals are highlighted with labels: 950 (2015/5/21 0:59:59 sec), 960 (2015/5/21 0:59:59 sec), 970 (2015/5/21 0:59:59 sec), 980 (2015/5/21 0:59:59 sec), and 990 (2015/5/21 0:59:59 sec).
- Right Panel (1):** Six smaller subplots arranged in a 2x3 grid, each showing a zoomed-in view of a specific time interval. Each subplot has two y-axes (left: 0-800, right: 0-800 or 0-500). The subplots are labeled:
  - Node1: SampleZipf
  - Node5: SampleZipf5
  - Node3: SampleZipf3
  - Node1: SampleZipf
  - Node1: SampleZipf
  - Node1: SampleZipf
Each subplot includes a timestamp and a value of r (e.g., r = 0.00, r = 0.95, r = 1.07, etc.).
- Bottom Panel (4):** A large, empty panel at the bottom of the interface.
- Annotations:** Numbered circles (1 through 6) are placed around the interface to highlight specific components: 1 points to the main plot, 2 points to the first subplot, 3 points to the second subplot, 4 points to the bottom panel, 5 points to the table in the left panel, and 6 points to the top-left corner of the left panel.

# Synopses

...Time Series Done Right

Squigglee Synopses
 Home Operation Configuration Retrieval Synopses

### Sketching

|                |            |            |        |
|----------------|------------|------------|--------|
| Time Series ID | SampleZipf | Data Type: | Int    |
| Min:           | 1.0        | Max:       | 1000.0 |
| First:         | 59.0       | Last:      | 17.0   |
| Value:         | 1          |            |        |
| Value Range:   | 1          | 1000       |        |
| Quantile (%):  | 50         |            |        |
| Bins:          | 100        |            |        |

Stats
Point Query
Range Query
Inverse Query
Histogram

### Sampling

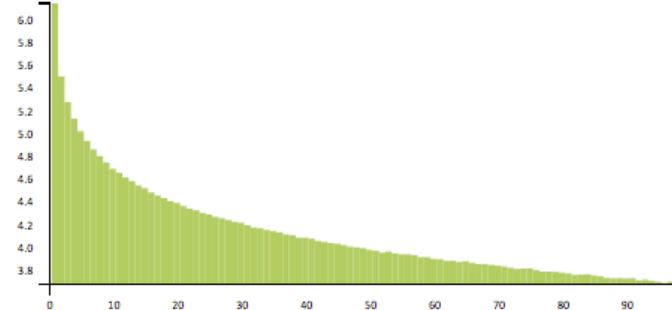
|                |                          |              |      |
|----------------|--------------------------|--------------|------|
| Time Series ID | SampleZipf               | Data Type:   |      |
| Start Date:    | 2015-05-21T00:00:00.000Z | Sample Size: | 1000 |
| End Date:      | 2015-05-21T23:59:59.999Z | Bins:        | 100  |

Add
Remove

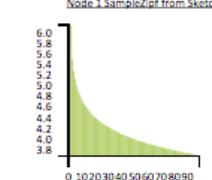
| Ln | Id          | Start                    | End                      | Histogram   |
|----|-------------|--------------------------|--------------------------|---|
| 1  | SampleZipf  | 2015-05-21T00:00:00.000Z | 2015-05-21T23:59:59.999Z | <span style="border: 1px solid #ccc; padding: 2px 10px; border-radius: 5px;">Histogram</span> |
| 3  | SampleZipf3 | 2015-05-21T00:00:00.000Z | 2015-05-21T23:59:59.999Z | <span style="border: 1px solid #ccc; padding: 2px 10px; border-radius: 5px;">Histogram</span> |
| 5  | SampleZipf5 | 2015-05-21T00:00:00.000Z | 2015-05-21T23:59:59.999Z | <span style="border: 1px solid #ccc; padding: 2px 10px; border-radius: 5px;">Histogram</span> |

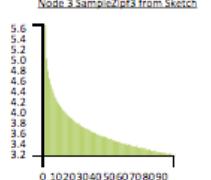
6
0
1

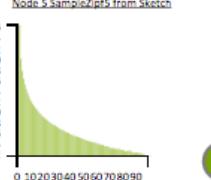
#### Sketch Data Distribution for Node 1 SampleZipf



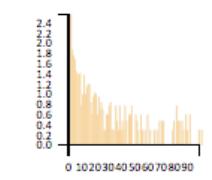
5
2
3

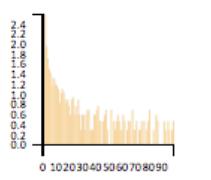
Node 1 SampleZipf from Sketch


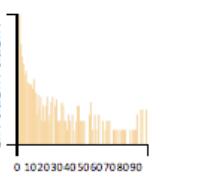
Node 3 SampleZipf3 from Sketch


Node 5 SampleZipf5 from Sketch


4
1
2

Node 1 SampleZipf from Sample


Node 3 SampleZipf3 from Sample


Node 5 SampleZipf5 from Sample


Updated status for cluster: Cluster1 at Thu May 21 2015 11:59:34 GMT-0500 (CDT)

Dimensionality  
Easier Operation  
New Customers  
Storage  
Indexing  
Lower Total Cost of Ownership  
More Revenues  
Unlimited Scale  
Node New Workflows  
Diversification  
Hybrid Deployment  
Data Management  
Approximate Query Processing

Complex Topologies  
Location Patterns  
Sampling  
Sketching  
Cloud  
Motif  
Synapses  
Replication  
Cluster  
Retrieval  
Instant Retrieval

## Time Series White Space Innovation

Thank You

For More Information Contact:

Dr. Ramesh K. Raghunathan

@ 214-620-1863

[ramesh.k.raghunathan@gmail.com](mailto:ramesh.k.raghunathan@gmail.com)

