

BBoxDB

A Key-Bounding-Box-Value Store

Demo Paper: Large Scale Spatial Data Processing with User Defined Filters in BBoxDB

Jan Kristof Nidzwetzki / Ralf Hartmut Güting Fernuniversität in Hagen, Germany Database Systems for New Applications {jan.nidzwetzki@studium., rhg@}fernuni-hagen.de

Outline



- Key-Value stores
- Partitioning data

BBoxDB – Architecture and User Defined Filters

- BBoxDB
- Architecture
- User defined filters



Key-Value stores Partitioning data

Basics

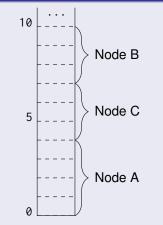
Key-value stores...

- have a simple data model: key and value.
- support at least the operations put(table, key, value) and get(table, key).
- don't support features like *joins* or *transactions*.
- can be implemented as a distributed system.

Key-Value stores Partitioning data

Partitioning data

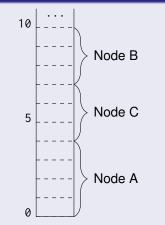
Range partitioning



Partitioning data

Key-Value stores Partitioning data

Range partitioning



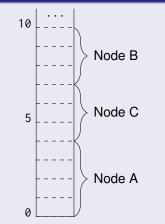
The key is important...

- to determine which node is responsible for a value.
- to locate the value on the node efficiently (e.g., *binary search* on sorted key-value pairs).

Partitioning data

Summary Partitioning data

Range partitioning



The key is important...

- to determine which node is responsible for a value.
- to locate the value on the node efficiently (e.g., *binary search* on sorted key-value pairs).

Data access path

Keys are the data access path in a key-value store. If a key is not known, an expensive full data scan is needed.

Key-Value stores Partitioning data

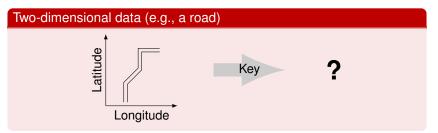
Multi-dimensional data in key-value stores



Key-Value stores Partitioning data

Multi-dimensional data in key-value stores





BBoxDB Architecture User defined filters



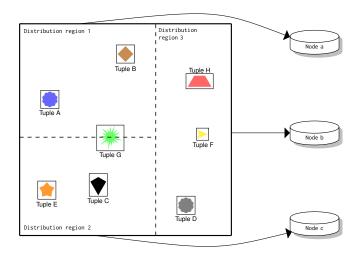
BBoxDB-A distributed key-bounding-box-value store

BBoxDB...

- is a distributed *key-bounding-box-value store*.
- stores each value together with a *bounding box*. The bounding box determines the location of the value in the *n*-dimensional space.
- can handle *n*-dimensional *point* and *non-point* big data.
- partitions the space dynamically and redistributes the data.
- stores data co-partitioned for efficient spatial joins.
- is freely available and licensed under the Apache 2.0 license.

BBoxDB Architecture

Partitioning the space



BBoxDB Architecture User defined filters

Supported Operations

The most important operations

• Store new data:

put(table, key, hyperrectangle, value)

Retrieve data:

getByRange(table, hyperrectangle)

Execute a spatial join:

```
join(table1, table2, hyperrectangle)
```

BBoxDB Architecture User defined filters

Supported Operations

The most important operations

• Store new data:

put(table, key, hyperrectangle, value)

Retrieve data:

getByRange(table, hyperrectangle)

• Execute a spatial join:

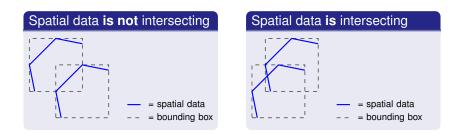
join(table1, table2, hyperrectangle)

Query processing in BBoxDB

BBoxDB is a generic datastore; values are plain arrays of bytes. The semantics of the stored values are unknown. The software performs operations (e.g., range queries or spatial joins) **only** on the bounding boxes of the data.

BBoxDB Architecture User defined filters

Spatial data with overlapping bounding boxes



Bounding boxes / Spatial join

Overlapping bounding boxes are a necessary but not a sufficient criterion for a spatial join on real geometries.

Multi-Dimensional Data in Key-Value Stores	BBoxDB
BBoxDB – Architecture and User Defined Filters	Architecture
Demonstration and Summary	User defined filters

User defined filters

BBoxDB Architecture User defined filters

Motivation for user defined filters

User defined filters...

 enhance the query processor, so that the stored values can be decoded (e.g., GeoJSON encoded values).

BBoxDB Architecture User defined filters

Motivation for user defined filters

User defined filters...

- enhance the query processor, so that the stored values can be decoded (e.g., GeoJSON encoded values).
- turn the generic data store into a specialized system for a specific data type (e.g., spatial joins on the real geometries of stored values become possible).

BBoxDB Architecture User defined filters

Implementation details

User defined filters in BBoxDB

- The UDF acts as a filter and refines the output of the query processor.
- The creation of a new UDF is simple: only the two methods of the interface UserDefinedFilter need to be implemented.
- The method filterTuple refines range queries; the method filterJoinCandidate refines join queries.
- The UDF is compiled into *Java bytecode* and is loaded dynamically at runtime.
- Existing Java libraries can be used (e.g., the *Esri Geometry API* for Java).

Demonstration Summary

Demonstration

Summary

BBoxDB...

- is a distributed key-bounding-box-value store.
- enhances the key-value data model with a bounding box.
- executes operations only on the bounding boxes of the data.
- can be enhanced by user defined filters. They turn BBoxDB into a system that is specialized on a certain data type.

Summary

• is written in Java and licensed under the Apache 2.0 license.

Demonstration Summary

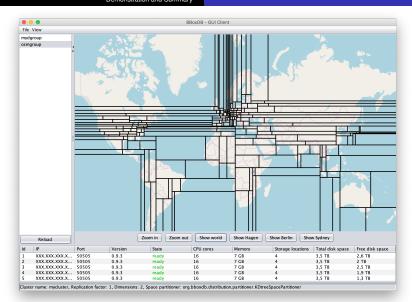
Questions?

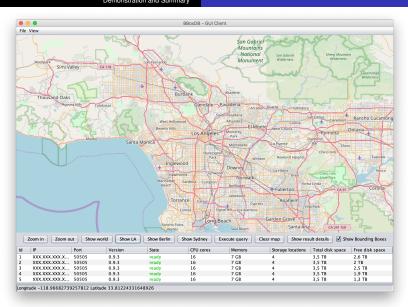
Further information

- Web: https://bboxdb.org
- Github: https://github.com/jnidzwetzki/bboxdb
- Twitter: @bboxdb
- Google Groups: https://groups.google.com/forum/bboxdb

Demonstration Summary

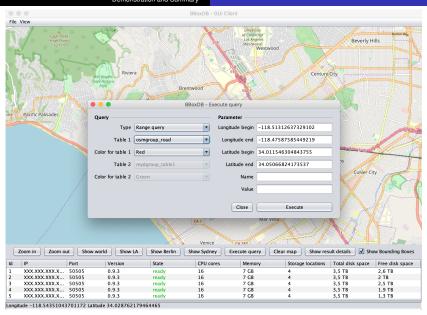
Backup Slides

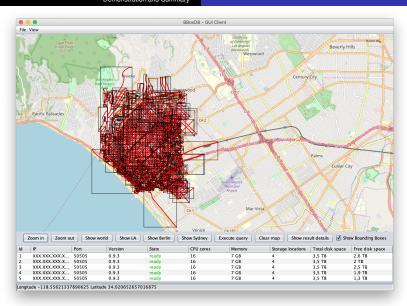




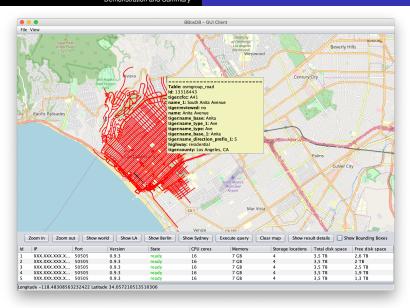
					BBoxDB - GUI Cli	ient			
File Vier	N								
	A case of pro- case of pro- sources of the second secon	27 10 10 27 10 10 29 10 10	Nogara		boomer Document	Diversity of Carlyong in a Bayes in the State is the Stat	nd Century		
		- 7	a Santa	Monica	and the second sec		538 534	Palms	< 13
Zoor			vorld Show LA	Show Berlin			Clear map Show re:	culver Cl	ow Bounding Boxe
IP		Port	vorld Show LA	Show Berlin State	Venice Show Sydney I CPU cores	Manucipal Arport Mar Vi FA 187 Execute query Memory	Clear map Show res Storage locations	Culver Cl	ow Bounding Boxe Free disk space
	xx.xxx.xxx.x	Port 50505	vorid Show LA	Show Berlin State ready	Venice Show Sydney	Mar Vi Aroari Mar Vi CA 187 Execute query Memory 7 CB	Clear map Show res Storage locations 4	Culter () suit details v Sho Total disk space 3,5 TB	ow Bounding Boxe Free disk space 2,6 TB
	XX.XXX.XXX.X XX.XXX.XXX.X	Port 50505 50505	vorld Show LA	Show Berlin State ready	Venice Show Sydney I CPU cores 16 16	Marvi Arport Mar Vi Execute query Memory 7 G8 7 G8	Clear map Show res Storage locations 4 4	Culver () sult details & Sh Total disk space 3,5 TB	ow Bounding Boxe Free disk space 2,6 TB 2 TB
	xx.xxx.xxx.x	Port 50505 50505 50505	vorid Show LA	Show Berlin State ready	Venice Show Sydney	Mar Vi Aroari Mar Vi CA 187 Execute query Memory 7 CB	Clear map Show res Storage locations 4	Culter () suit details v Sho Total disk space 3,5 TB	ow Bounding Boxe Free disk space 2,6 TB

				BBoxDB - GUI C	lient			
File View								
tie Vew		Rogisti Rogisti Poly		bentwood Co.2	Deserving C cale Applies C cale Applies Profiles Control of Control Control of Control of Control of Control Control of Control of Control of Control Control of Control of Control of Control of Control Control of Control of Control of Control of Control of Control Control of Control of Contro	of Century The university of		
		Santa	Monica	CA1 Venice Show Sydney	Smith down Maryon Argan Mar V r4 187 Execute query	\mathbb{X}	Palms Culver C	y w Bounding Boxe
Zoom in Zoom	out Show we	orld Show LA						
	Dut Show we	Version	State	CPU cores	Memory	Storage locations	Total disk space	Free disk space
I IP XXX.XXX.XXX.X	Port 50505	Version 0.9.3		16	7 GB	Storage locations 4	3,5 TB	2,6 TB
I IP XXX.XXX.XXX.X.X. XXX.XXX.XXX.X.	Port 50505 50505	Version 0.9.3 0.9.3	State ready ready	16 16	7 GB 7 GB	4 4	3,5 TB 3,5 TB	2,6 TB 2 TB
I IP XXX.XXX.XXX.X.X. XXX.XXX.XXX.X XXX.XXX.XXX.X	Port 50505 50505 50505	Version 0.9.3 0.9.3 0.9.3	State ready ready ready	16 16 16	7 GB 7 GB 7 GB	4 4 4	3,5 TB 3,5 TB 3,5 TB	2 TB 2,5 TB
I IP XXX.XXX.XXX.X.X. XXX.XXX.XXX.X	Port 50505 50505 50505 50505 50505	Version 0.9.3 0.9.3	State ready ready	16 16	7 GB 7 GB	4 4	3,5 TB 3,5 TB	2,6 TB 2 TB

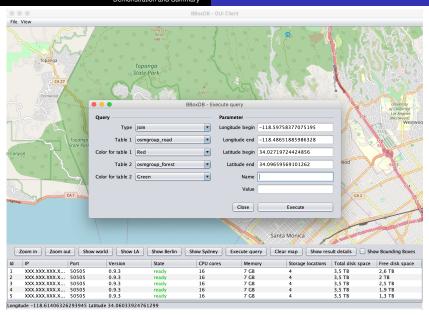




. 😐 😐				BBoxDB - GUI C	lient			
ile View								
e Vew				Brenwood 0.1	Burensin 2 Cardination 2 Cardi	cord Centur Van at even	City	y Hils
Zoom in Zoor IP XXX XXX XXX XXXX	Port 50505 50505 50505	world Show LA Version 0.9.3 0.9.3 0.9.3	Show Berlin State ready ready ready	CPU cores 16 16 16 16 16	Anna Joneo Manacpai Armar Armar Execute query 7 GB 7 GB 7 GB 7 GB	\mathcal{A}	sult details Total disk spa 3,5 TB 3,5 TB 3,5 TB 3,5 TB	Show Bounding Box ce Free disk spac 2,6 TB 2,7 B 2,5 TB 1,9 TB
XXX.XXX.XXX.X					7 GB	4		1.9 18
XXX.XXX.XXX.X XXX.XXX.XXX.X XXX.XXX.XXX		0.9.3	ready	16	7 GB	4	3.5 TB	1.3 TB



				BBoxDB - GUI Cli	ient			
ile Vi	w							
	Toerge A Toerge A Toerge Toerge Toerge Toerge		oppanga ate. Park	27 gg g Shrm Gash Pola Da rei Da rei		6		Nerring California La Appropria Wes
2			itelanmare Pacifi Mesa 105 m	c Palisades	Mill Rogens Stole History Park	B	rentwood CA2	55810
	min Zoom out 5 P Port	now world Show La	Show Berlin State ready	Show Sydney CPU cores 16	Execute query 7 CS	Inta Monica Inta Monica Clear map Show res Storage locations 4	Co.2 wit details Shi Total disk space 3,5 TB	ow Bounding Boxe 2,6 TB
Zo	m in Zoom out S P Port XXXXXXXXXX 55555	now world Show LM Version 0.9.3	Show Berlin State ready ready	Show Sydney CPU cores 16 16	Execute query 7 C8 7 C8	Clear map Storage locations 4	uit details Sho Total disk space 3,5 TB	ow Bounding Boxe Free disk spac 2,6 TB 2 TB
Zo	min Zoom out 5 P Port	now world Show La	Show Berlin State ready	Show Sydney CPU cores 16	Execute query 7 CS	Inta Monica Inta Monica Clear map Show res Storage locations 4	Co.2 wit details Shi Total disk space 3,5 TB	ow Bounding Boxe Free disk spac 2,6 TB



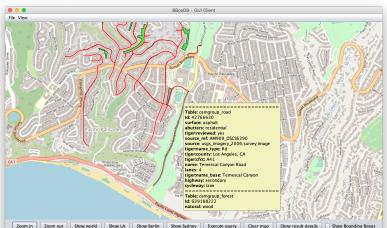
File View				BBoxDB - GUI Cli	ent			
Toppor	Topanga		tanga e.Park	abm Marine Sam	All Register Beer Angeles	Nep 3	reutwood	University of California Metabolic Metabolic West
	CAT	Mesa	lammar Pacifi Aca Om	Paisads		anta Monica	CA2	
Zoom in Zoom	out Show w	orld Show LA	Show Berlin	Show Sydney I	Execute query	Clear map Show res	sult details Sho	ow Bounding Boxe
Zoom in Zoom	out Show w	orld Show LA Version	Show Berlin State	Show Sydney E CPU cores	Execute query	Clear map Show res Storage locations	sult details Shu	Free disk spac
Zoom in Zoom IP XXX.XXX.XXX.X	out Show w Port . 50505	orld Show LA Version 0.9.3	Show Berlin State ready	Show Sydney E CPU cores 16	Execute query Memory 7 GB	Clear map Show res Storage locations 4	sult details Sho Total disk space 3,5 TB	Free disk spac 2,6 TB
Zoom in Zoom IP XXX.XXX.XXX.XX XXX.XXX.XXX.XX	out Show w Port - 50505 - 50505	orld Show LA Version 0.9.3 0.9.3	Show Berlin Show Berlin State ready ready	Show Sydney I CPU cores 16 16	Execute query Memory 7 GB 7 GB	Clear map Show res Storage locations 4 4	sult details Sho Total disk space 3,5 TB 3,5 TB	Free disk spac 2,6 TB 2 TB
IP XXX.XXX.XXX.X	out Show w Port . 50505 . 50505 . 50505	orld Show LA Version 0.9.3	Show Berlin State ready	Show Sydney E CPU cores 16	Execute query Memory 7 GB	Clear map Show res Storage locations 4	sult details Sho Total disk space 3,5 TB	Free disk spac 2,6 TB

Demonstration Summary

	•				BBoxDB - GUI Clie	ent			
ile Vi	ew								
			Arrest down Johnson Parister Bany -					Ville Roger State Histor	
Zor	om in Zoom o	ut Show wor						1	
_				State	CRIL coror				Eree dick space
	IP	Port	Version	State	CPU cores	Memory	Storage locations	Total disk space	Free disk space
	IP XXX.XXX.XXX.X	Port 50505	Version 0.9.3	ready	16	7 GB	4	3,5 TB	2,6 TB
	IP XXX.XXX.XXX.X XXX.XXX.XXX.X	Port 50505 50505	Version 0.9.3 0.9.3	ready ready	16 16	7 GB 7 GB	4 4	3,5 TB 3,5 TB	2,6 TB 2 TB
	IP XXX.XXX.XXX.X XXX.XXX.XXX.X XXX.XXX.XXX.X	Port 50505 50505 50505	Version 0.9.3 0.9.3 0.9.3	ready ready ready	16 16 16	7 GB 7 GB 7 GB	4 4 4	3,5 TB 3,5 TB 3,5 TB	2,6 TB 2 TB 2,5 TB
	IP XXX.XXX.XXX.X XXX.XXX.XXX.X	Port 50505 50505 50505 50505 50505	Version 0.9.3 0.9.3	ready ready	16 16	7 GB 7 GB	4 4	3,5 TB 3,5 TB	2,6 TB 2 TB

Longitude -118.55187892913818 Latitude 34.03957367053948

Demonstration Summary



Id	IP	Port	Version	State	CPU cores	Memory	Storage locations	Total disk space	Free disk space
1	XXX.XXX.XXX.X	50505	0.9.3	ready	16	7 GB	4	3,5 TB	2,6 TB
2	XXX.XXX.XXX.X	50505	0.9.3	ready	16	7 GB	4	3,5 TB	2 TB
3	XXX.XXX.XXX.X	50505	0.9.3	ready	16	7 GB	4	3,5 TB	2,5 TB
4	XXX.XXX.XXX.X	50505	0.9.3	ready	16	7 GB	4	3,5 TB	1,9 TB
5	XXX.XXX.XXX.X	50505	0.9.3	ready	16	7 GB	4	3.5 TB	1.3 TB

Longitude -118.53110790252686 Latitude 34.04512113757758

		BBoxDB -	GUI Client			
File View						
Planet Date			Pacific Palisades	And a start of a start		
		BBoxDB - Ex	ecute query		-	i i cesti da
sugar took a	Query		Parameter		JIE COL	The second for
Country out	Type Join	•		4711532592773		Toulon Drive
and the second second	Table 1 osmgro	up_road	Longitude end -118.5	1823329925537		
Bellevard	Color for table 1 Red	•	Latitude begin 34.034	381481654364		
	Table 2 osmgro	up_forest	Latitude end 34.052	62386500679		and the second se
CAI	Color for table 2 Green	•	Name .UserDe	efinedGeoJsonSpatialFilter		$17/n^{\circ}$
			Value			in the second
			Close	Execute		on international second
		Bacille Coast Highway	200	ANS		
Zoom in Zoom out Sho	w world Show LA S	Show Berlin Show Sydne	y Execute query	Clear map Show res	ult details 🗌 Sho	w Bounding Boxes
ld IP Port		State CPU co		Storage locations	Total disk space	Free disk space
1 XXX.XXX.XXX.X 50505		ready 16	7 GB	4	3,5 TB	2,6 TB
2 XXX.XXX.XXX.X 50505		ready 16	7 GB	4	3,5 TB	2 TB
3 XXX.XXX.XXX.X 50505 4 XXX.XXX.XXX.X 50505		ready 16 ready 16	7 GB 7 GB	4	3,5 TB	2,5 TB 1,9 TB
4 XXX.XXX.XXX.X 50505 5 XXX.XXX.XXX.XX 50505		ready 16	7 GB	4	3,5 TB 3,5 TB	1,9 TB 1,3 TB
Longitude -118.52823257446289 L						

. 😑 🕒				BBoxDB - GUI C	lient			
le View								
		Horace down						
Zoom in Zoom			Show Berlin	Show Sydney	Execute query			w Bounding Boxes
IP	Port	Version	State	CPU cores	Memory	Storage locations	Total disk space	Free disk space
		0.9.3	ready	16	7 GB	4	3,5 TB	2,6 TB
XXX.XXX.XXX.X			ready	16	7 GB	4	3,5 TB	2 TB
XXX.XXX.XXX.X								
XXX.XXX.XXX.X XXX.XXX.XXX.X	50505	0.9.3	ready	16	7 GB	4	3,5 TB	2,5 TB
XXX.XXX.XXX.X	50505 50505			16 16 16	7 GB 7 GB 7 GB	4 4 4	3,5 TB 3,5 TB 3,5 TB	2,5 TB 1,9 TB 1,3 TB

Demonstration Summary

	•		BBoxDB - GUI Clie	ent			
File			550XDD - 001 Cile				
CA1			Teleson de 30012 software as tigerreite ames 20 tigerreite ames 20 tigerreite tigerreite tigerreite tigerreite tigerreite tigerreite tigerreite tigerreite tigerreite	group_road 905 905 908 908 908 908 908 908 908 908 908 908			
7	Toom in Toom out Shown	and Chan A Chan	Partie Coart Hollow		e Dent		XX.
_	Zoom in Zoom out Show w			xecute query Clear	re Drift Show res		w Bounding Boxes
Z	IP Port	Version Stat	e CPU cores	xecute query Clear Memory	map Show res Storage locations	Total disk space	Free disk space
	IP Port XXX.XXX.XXX.X 50505	Version Stat 0.9.3 rea	e CPU cores dy 16	Xecute query Clear Memory 7 GB	map Show res Storage locations 4	Total disk space 3,5 TB	Free disk space 2,6 TB
_	IP Port XXX.XXX.XXX.X 50505 XXX.XXX.XXX.X 50505	Version Stat 0.9.3 rea 0.9.3 rea	e CPU cores dy 16 dy 16	xecute query Clear Memory 7 GB 7 GB	map Show res Storage locations 4 4	Total disk space 3,5 TB 3,5 TB	Free disk space 2,6 TB 2 TB
Ē	IP Port XXX.XXX.XXX.X 50505	Version Stat 0.9.3 rea	e CPU cores dy 16 dy 16 dy 16 dy 16	Xecute query Clear Memory 7 GB	map Show res Storage locations 4	Total disk space 3,5 TB	Free disk space 2,6 TB

Longitude -118.53076457977295 Latitude 34.046365711752664