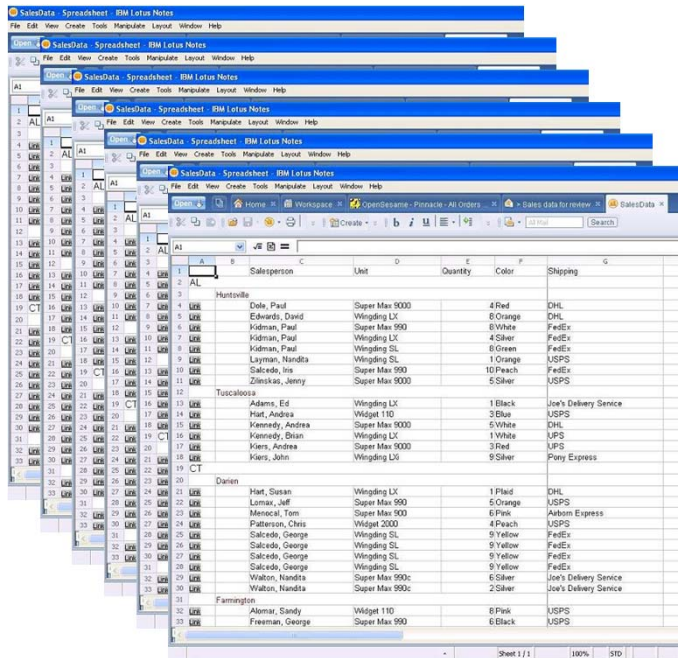


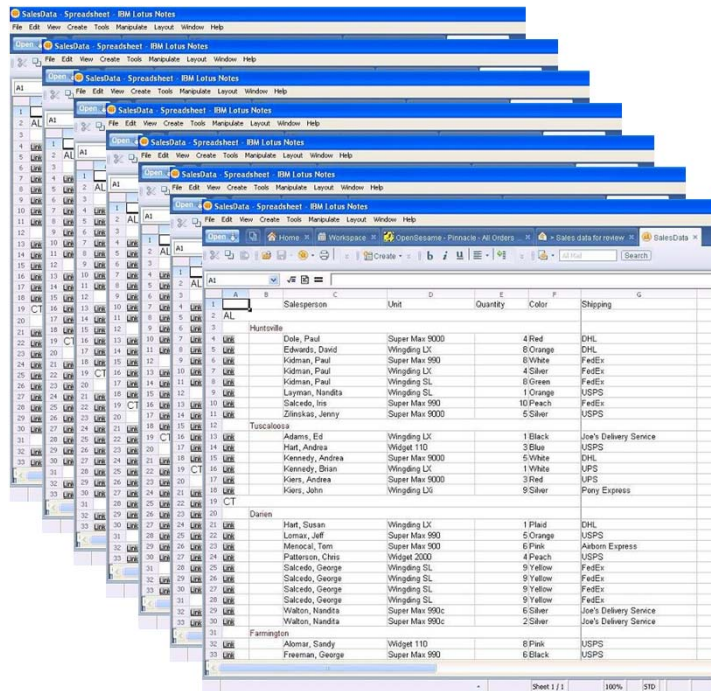
Can HBase Improve SpreaDooop?

Presenter: Mert Coskun

Quick Intro to SpreaDooop



Quick Intro to SpreaDoop



The screenshot shows a spreadsheet application with multiple overlapping windows, each titled "SalesData - Spreadsheet - BM Lotus Notes". The main window displays a table with the following data:

	A	B	C	D	E	F	G
		Salesperson	Unit	Quantity	Color	Shipping	
1	AL						
2	AL						
3							
4	1						
5	2						
6	3						
7	4						
8	5						
9	6						
10	7						
11	8						
12	9						
13	10						
14	11						
15	12						
16	13						
17	14						
18	15						
19	16						
20	17						
21	18						
22	19						
23	20						
24	21						
25	22						
26	23						
27	24						
28	25						
29	26						
30	27						
31	28						
32	29						
33	30						
34	31						
35	32						
36	33						
37	34						
38	35						
39	36						
40	37						
41	38						
42	39						
43	40						
44	41						
45	42						
46	43						
47	44						
48	45						
49	46						
50	47						
51	48						
52	49						
53	50						
54	51						
55	52						
56	53						
57	54						
58	55						
59	56						
60	57						
61	58						
62	59						
63	60						
64	61						
65	62						
66	63						
67	64						
68	65						
69	66						
70	67						
71	68						
72	69						
73	70						
74	71						
75	72						
76	73						
77	74						
78	75						
79	76						
80	77						
81	78						
82	79						
83	80						
84	81						
85	82						
86	83						
87	84						
88	85						
89	86						
90	87						
91	88						
92	89						
93	90						
94	91						
95	92						
96	93						
97	94						
98	95						
99	96						
100	97						
101	98						
102	99						
103	100						
104	101						
105	102						
106	103						
107	104						
108	105						
109	106						
110	107						
111	108						
112	109						
113	110						
114	111						
115	112						
116	113						
117	114						
118	115						
119	116						
120	117						
121	118						
122	119						
123	120						
124	121						
125	122						
126	123						
127	124						
128	125						
129	126						
130	127						
131	128						
132	129						
133	130						
134	131						
135	132						
136	133						
137	134						
138	135						
139	136						
140	137						
141	138						
142	139						
143	140						
144	141						
145	142						
146	143						
147	144						
148	145						
149	146						
150	147						
151	148						
152	149						
153	150						
154	151						
155	152						
156	153						
157	154						
158	155						
159	156						
160	157						
161	158						
162	159						
163	160						
164	161						
165	162						
166	163						
167	164						
168	165						
169	166						
170	167						
171	168						
172	169						
173	170						
174	171						
175	172						
176	173						
177	174						
178	175						
179	176						
180	177						
181	178						
182	179						
183	180						
184	181						
185	182						
186	183						
187	184						
188	185						
189	186						
190	187						
191	188						
192	189						
193	190						
194	191						
195	192						
196	193						
197	194						
198	195						
199	196						
200	197						
201	198						
202	199						
203	200						
204	201						
205	202						
206	203						
207	204						
208	205						
209	206						
210	207						
211	208						
212	209						
213	210						
214	211						
215	212						
216	213						
217	214						
218	215						
219	216						
220	217						
221	218						
222	219						
223	220						
224	221						
225	222						
226	223						
227	224						
228	225						
229	226						
230	227						
231	228						
232	229						
233	230						
234	231						
235	232						
236	233						
237	234						
238	235						
239	236						
240	237						
241	238						
242	239						
243	240						
244	241						
245	242						
246	243						
247	244						
248	245						
249	246						
250	247						
251	248						
252	249						
253	250						
254	251						
255	252						
256	253						
257	254						
258	255						
259	256						
260	257						
261	258						
262	259						
263	260						
264	261						
265	262						
266	263						
267	264						
268	265						
269	266						
270	267						
271	268						
272	269						
273	270						
274	271						
275	272						
276	273						
277	274						
278	275						
279	276						
280	277						
281	278						

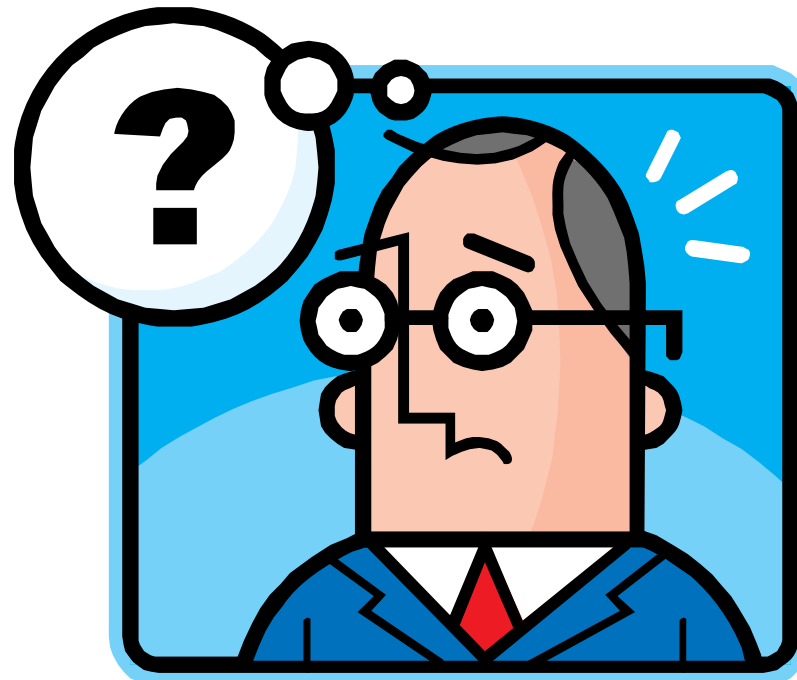
Why SpreaDoop?

- Goodness of Distribution
- Separate Functions from Data
- **Revolutionary Indexing/Naming**
- Optimisations

Indexing/Naming

- AVERAGEIF (AAAAA 16061987 :
AAAAA 42424242 = London; ...)

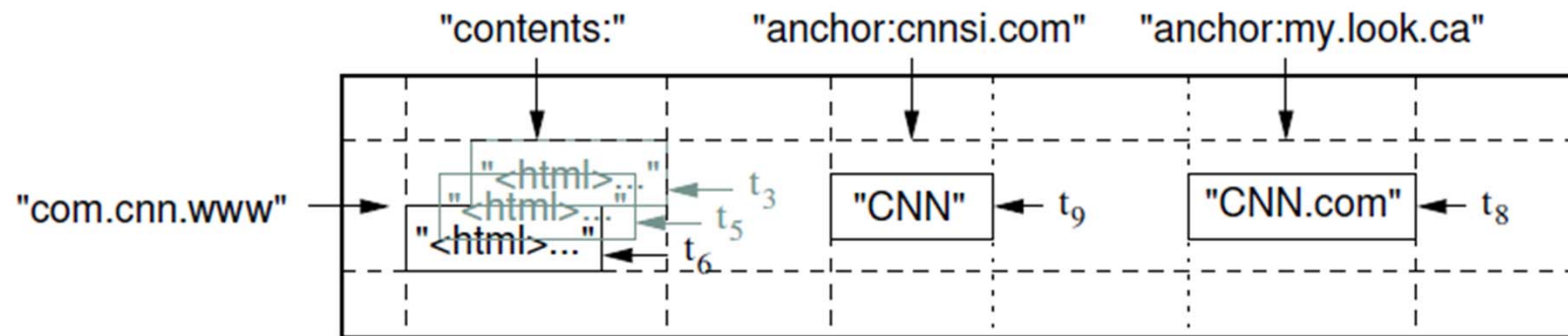
- 16061987
- 4242424242
- London
- AAAAA
- ZZKSA



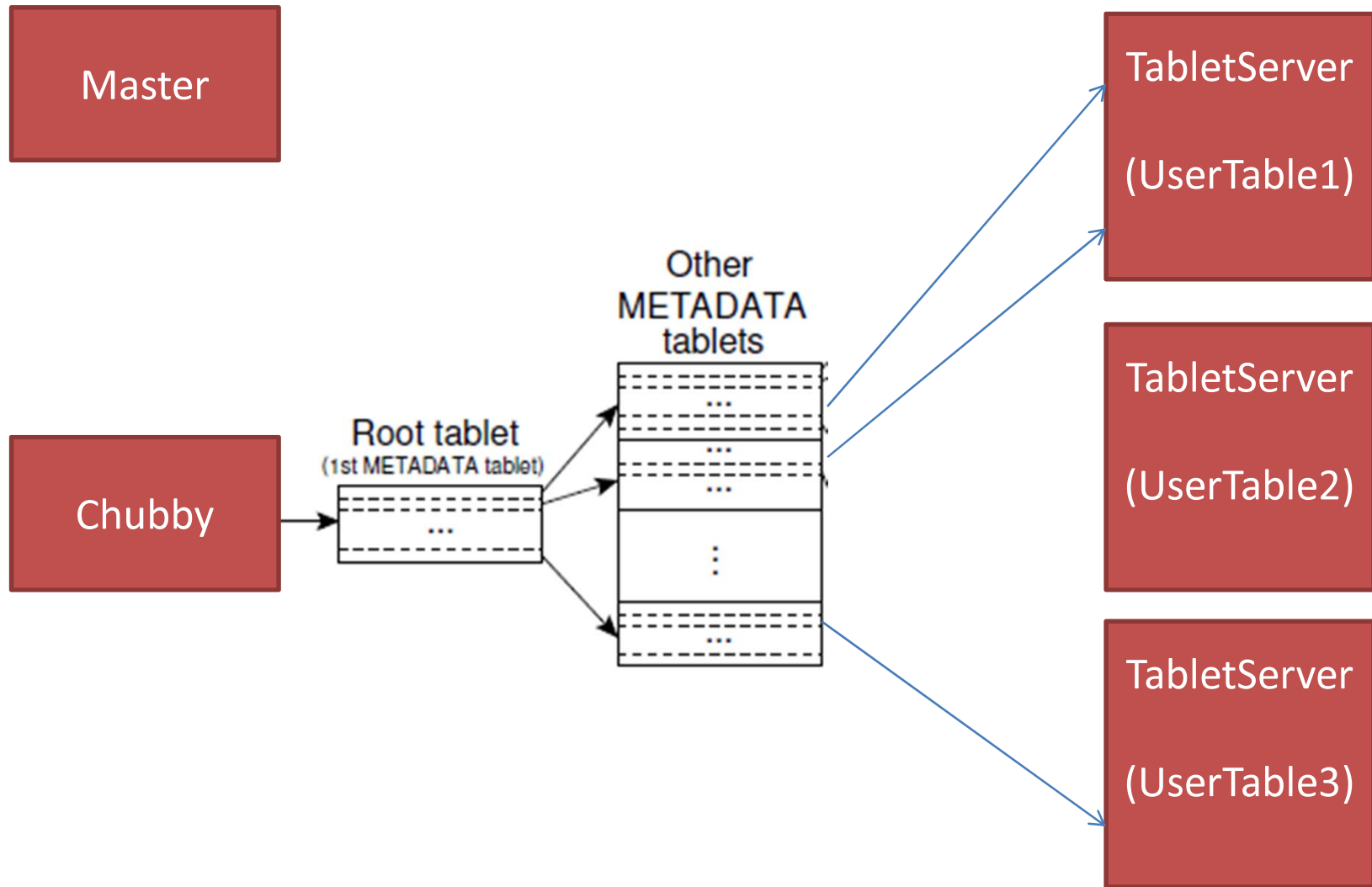
SpreaDoop Solution

- AVERAGE (City = London && Year = 2011,
January, March, May, July)

Hbase = OpenSource (BigTable)



BigTable



Why & Why not HBase

- Three Dimensional
- Random Reads
- Row Keys
- Translation

Experiment Design

- Antarctica Temperature Data (*)
 - Text vs. Hbase
- Common Functions
 - STDEV, AVERAGE, SUM all rows with Year = 2011

(*) <http://www.antarctica.ac.uk/met/READER/surface/stationpt.html>

HBase Evaluation Plan

- Hard Metrics
 - Computation Time (Access Time) (Scan) (Random)
 - Distributed Storage Requirements (Txt vs HBase)
 - Translation Time (Txt to HBase)
- Soft Metrics
 - Simplify Programming
 - Improve Extendibility
- How does it compare to other Hadoop optimisations, extensions such as
 - Nectar, Quincy, MapReduce-Merge, etc

Thank you!

