

BI / read / 1

BI 1	query	BI / read / 1																																						
BI 2	title	Posting summary																																						
BI 3	pattern	<table><tr><td>message: Message</td></tr><tr><td>creationDate < \$datetime</td></tr><tr><td>length year(creationDate)</td></tr></table>				message: Message	creationDate < \$datetime	length year(creationDate)																																
message: Message																																								
creationDate < \$datetime																																								
length year(creationDate)																																								
BI 4																																								
BI 5																																								
BI 6	desc.	<p>Given a datetime, find all Messages created before that moment. Group them by a 3-level grouping:</p> <p>1. by year of creation</p> <p>2. for each year, group into Message types: is Comment or not</p> <p>3. for each year-type group, split into four groups based on length of their content</p> <ul style="list-style-type: none">• 0: $0 \leq \text{length} < 40$ (short)• 1: $40 \leq \text{length} < 80$ (one liner)• 2: $80 \leq \text{length} < 160$ (tweet)• 3: $160 \leq \text{length}$ (long)																																						
BI 7																																								
BI 8																																								
BI 9																																								
BI 10																																								
BI 11																																								
BI 12																																								
BI 13																																								
BI 14																																								
BI 15																																								
BI 16	params	1	datetime	DateTime	For later microbatches, later datetime parameters are selected keep the variance low (<0.5%)																																			
BI 17		<table><tr><td>1</td><td>year</td><td>32-bit Integer</td><td>R</td><td>year(message.creationDate)</td></tr><tr><td>2</td><td>isComment</td><td>Boolean</td><td>M</td><td>True for Comments, False for Posts</td></tr><tr><td>3</td><td>lengthCategory</td><td>32-bit Integer</td><td>C</td><td>0 for short, 1 for one-liner, 2 for tweet, 3 for long</td></tr><tr><td>4</td><td>messageCount</td><td>32-bit Integer</td><td>A</td><td>Total number of Messages in that group</td></tr><tr><td>5</td><td>averageMessageLength</td><td>32-bit Float</td><td>A</td><td>Average length of the Message content in that group</td></tr><tr><td>6</td><td>sumMessageLength</td><td>32-bit Integer</td><td>A</td><td>Sum of all Message content lengths</td></tr><tr><td>7</td><td>percentageOfMessages</td><td>32-bit Float</td><td>A</td><td>Number of Messages in group as a percentage of all messages created before the given date</td></tr></table>				1	year	32-bit Integer	R	year(message.creationDate)	2	isComment	Boolean	M	True for Comments, False for Posts	3	lengthCategory	32-bit Integer	C	0 for short, 1 for one-liner, 2 for tweet, 3 for long	4	messageCount	32-bit Integer	A	Total number of Messages in that group	5	averageMessageLength	32-bit Float	A	Average length of the Message content in that group	6	sumMessageLength	32-bit Integer	A	Sum of all Message content lengths	7	percentageOfMessages	32-bit Float	A	Number of Messages in group as a percentage of all messages created before the given date
1						year	32-bit Integer	R	year(message.creationDate)																															
2						isComment	Boolean	M	True for Comments, False for Posts																															
3	lengthCategory					32-bit Integer	C	0 for short, 1 for one-liner, 2 for tweet, 3 for long																																
4	messageCount					32-bit Integer	A	Total number of Messages in that group																																
5	averageMessageLength					32-bit Float	A	Average length of the Message content in that group																																
6	sumMessageLength					32-bit Integer	A	Sum of all Message content lengths																																
7	percentageOfMessages	32-bit Float	A	Number of Messages in group as a percentage of all messages created before the given date																																				
BI 18																																								
BI 19																																								
BI 20																																								
sort	1	year	↓																																					
	2	isComment	↑	False < True, i.e. Posts come first and Comments second																																				
	3	lengthCategory	↑	order based on the lengthCategory value																																				
limit	n/a																																							
CPs	1.2, 3.2, 4.1, 4.2, 8.5																																							