

## 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 &lt; \$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								
BI 7	desc.	Given a datetime, find all Messages created before that moment. Group them by a 3-level grouping:						
BI 8		1. by year of creation						
BI 9		2. for each year, group into Message types: is Comment or not						
BI 10		3. for each year-type group, split into four groups based on length of their content						
BI 11		• 0: $0 \leq \text{length} < 40$ (short)						
BI 12		• 1: $40 \leq \text{length} < 80$ (one liner)						
BI 13		• 2: $80 \leq \text{length} < 160$ (tweet)						
BI 14		• 3: $160 \leq \text{length}$ (long)						
BI 15	params	1	datetime	DateTime	For later microbatches, later datetime parameters are selected keep the variance low (<0.5%)			
BI 16								
BI 17	result	1	year	32-bit Integer	R	year(message.creationDate)		
BI 18		2	isComment	Boolean	M	True for Comments, False for Posts		
BI 19		3	lengthCategory	32-bit Integer	C	0 for short, 1 for one-liner, 2 for tweet, 3 for long		
BI 20		4	messageCount	32-bit Integer	A	Total number of Messages in that group		
		5	averageMessageLength	32-bit Integer	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		
	sort	1	year	↓				
		2	isComment	↑	False < True, i.e. Posts come first and Comments second			
		3	lengthCategory	↑	order based on the lengthCategory value			
	CPs	1.2, 3.2, 4.1, 4.2, 8.5						