

## BI / read / 1

BI 1

BI 2

BI 3

BI 4

BI 5

BI 6

BI 7

BI 8

BI 9

BI 10

BI 11

BI 12

BI 13

BI 14

BI 15

BI 16

BI 17

BI 18

BI 19

BI 20

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