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

BI / read / 1

query	BI / read / 1
title	Posting summary

pattern

message: Message
creationDate < $dateTime
length year(creationDate)

desc.

Given a datetime, find all Messages created before that moment. Group them by a 3-level grouping:

1. by year of creation
2. for each year, group into Message types: is Comment or not
3. for each year-type group, split into four groups based on length of their content

   - 0: $0 \leq length < 40$ (short)
   - 1: $40 \leq length < 80$ (one liner)
   - 2: $80 \leq length < 160$ (tweet)
   - 3: $160 \leq length$ (long)

params

1	datetime	DateTime	For later microbatches, later datetime parameters are selected keep the variance low (<0.5%)

result

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

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