Anti-Skew: Single-Key Data Skew Mitigation for MapReduce

Yue Chen

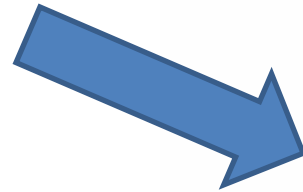Florida State University

Advanced Database Systems

Outline

- Background

- Data Skew

- Anti-skew Design

- Conclusion

- Related Work

Background

Big Data Trend

Big Data Trend

Mike Olson
@mikeolson

IBM seeks architect with 8 years Cloudera experience. Alas, I only have 7.
linkd.in/1B50API

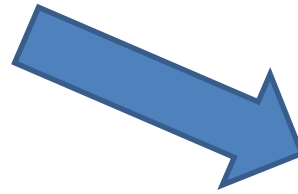| RETWEETS	FAVORITES
255	153

8:48 AM - 3 Mar 2015
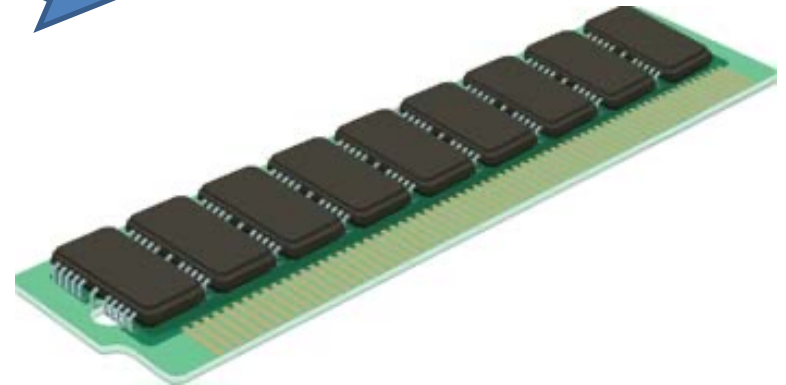
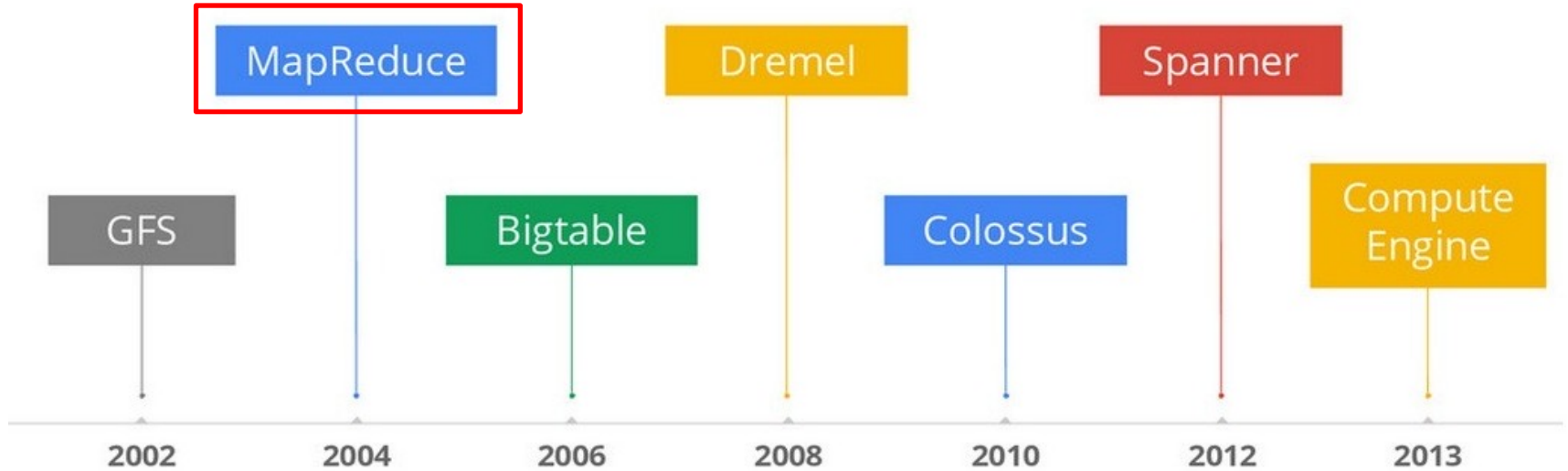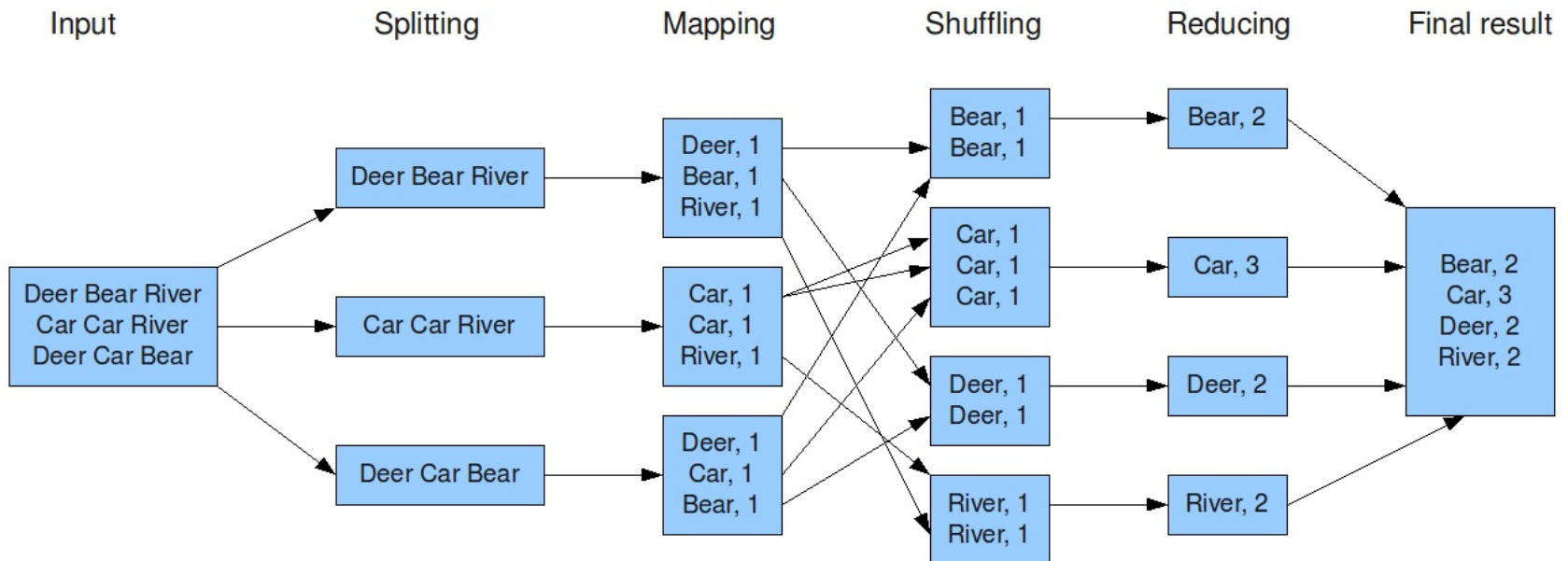Mike Olson is a co-founder and former CEO of Cloudera.
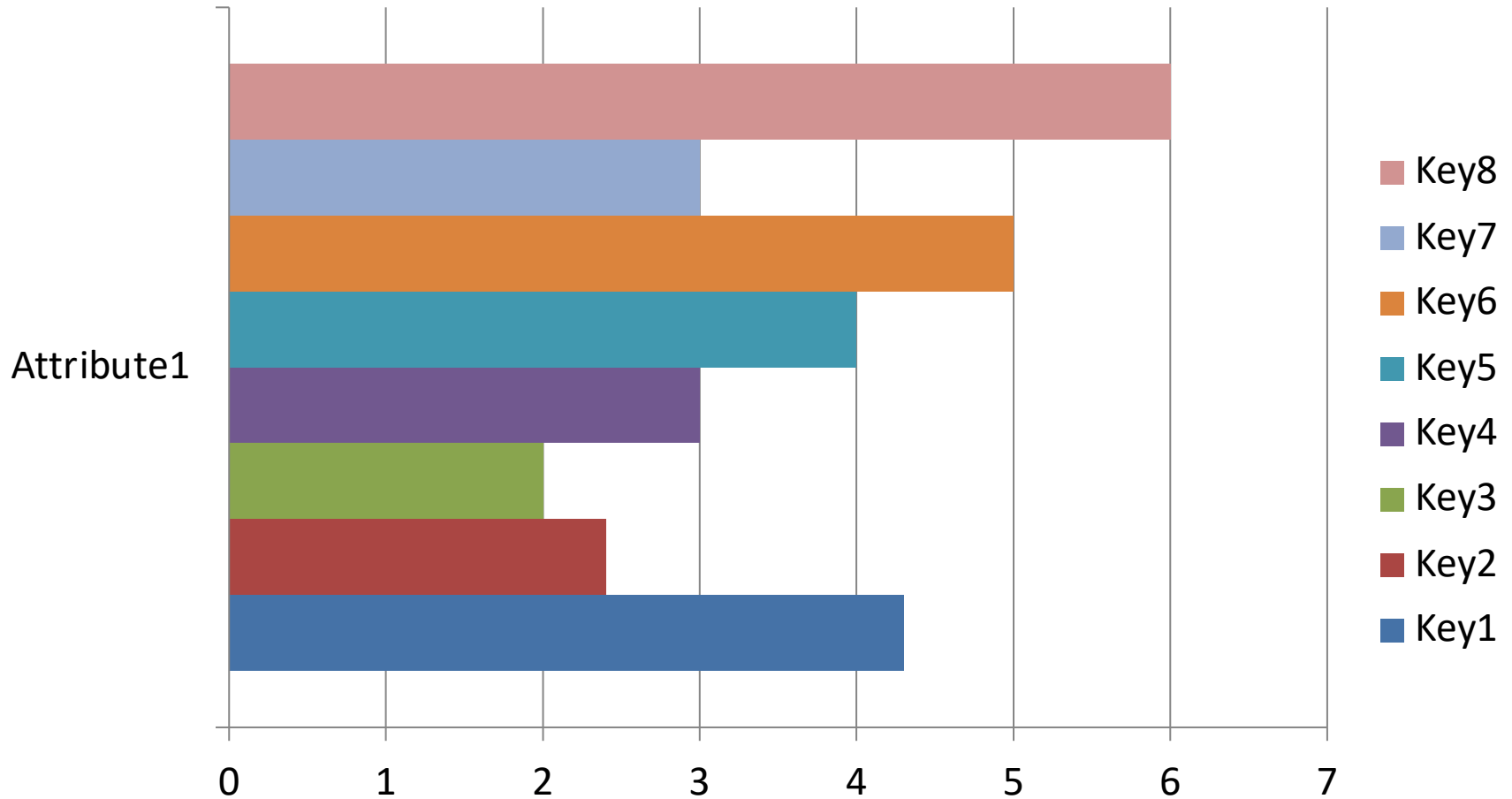
Big Data Trend

History

Review of MapReduce Word Count

The overall MapReduce word count process

Data Skew

What is data skew?
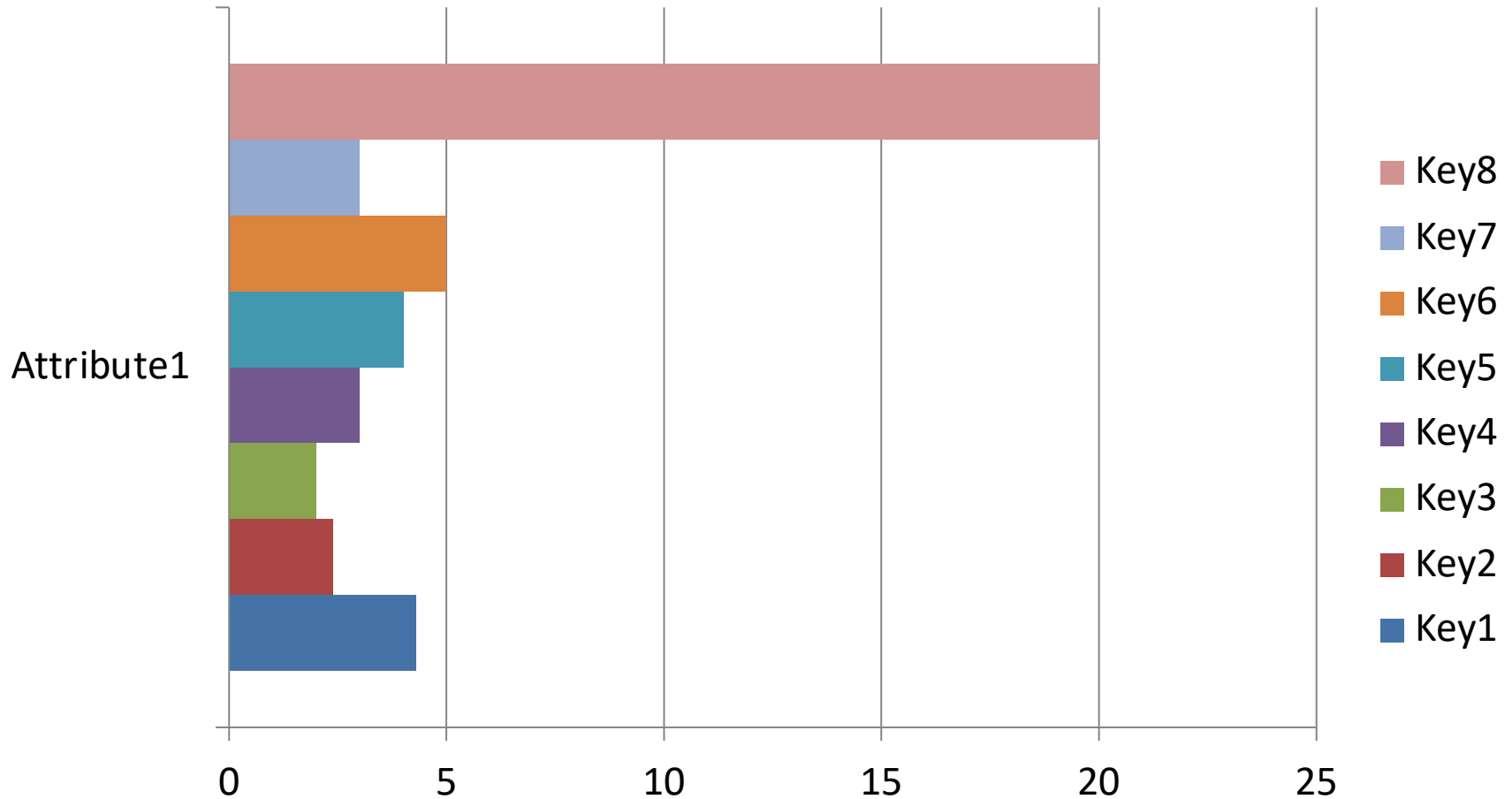
What is data skew?

What is data skew?
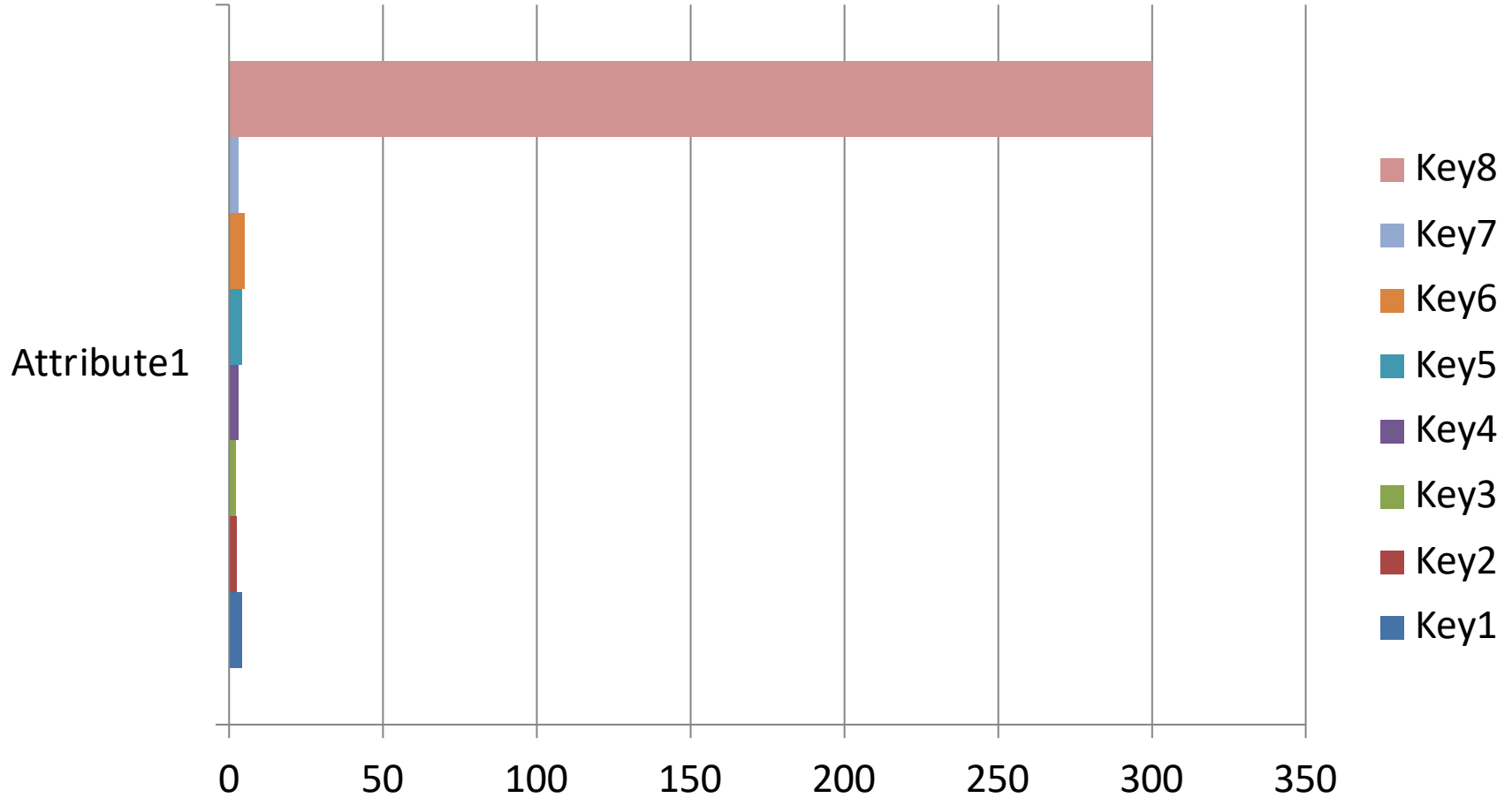
WORD	COUNT	PERCENT	bar graph
the	53.10 B	7.14%	the
of	30.97 B	4.16%	of
and	22.63 B	3.04%	and
to	19.35 B	2.60%	to
in	16.89 B	2.27%	in
a	15.31 B	2.06%	a
is	8.38 B	1.13%	is
that	8.00 B	1.08%	that
for	6.55 B	0.88%	for
it	5.74 B	0.77%	it
as	5.70 B	0.77%	as
was	5.50 B	0.74%	was
with	5.18 B	0.70%	with
be	4.82 B	0.65%	be
by	4.70 B	0.63%	by
on	4.59 B	0.62%	on
not	4.52 B	0.61%	not
he	4.11 B	0.55%	he
i	3.88 B	0.52%	i
this	3.83 B	0.51%	this
are	3.70 B	0.50%	are
or	3.67 B	0.49%	or
his	3.61 B	0.49%	his
from	3.47 B	0.47%	from
at	3.41 B	0.46%	at
which	3.14 B	0.42%	which
but	2.79 B	0.38%	but
have	2.78 B	0.37%	have
an	2.73 B	0.37%	an
had	2.62 B	0.35%	had
they	2.46 B	0.33%	they
you	2.34 B	0.31%	you
were	2.27 B	0.31%	were
their	2.15 B	0.29%	their
one	2.15 B	0.29%	one
all	2.06 B	0.28%	all
we	2.06 B	0.28%	we
can	1.67 B	0.22%	can
her	1.63 B	0.22%	her
has	1.63 B	0.22%	has
there	1.62 B	0.22%	there
been	1.62 B	0.22%	been
if	1.56 B	0.21%	if
more	1.55 B	0.21%	more
when	1.52 B	0.20%	when
will	1.49 B	0.20%	will
would	1.47 B	0.20%	would
who	1.46 B	0.20%	who
so	1.45 B	0.19%	so
no	1.40 B	0.19%	no

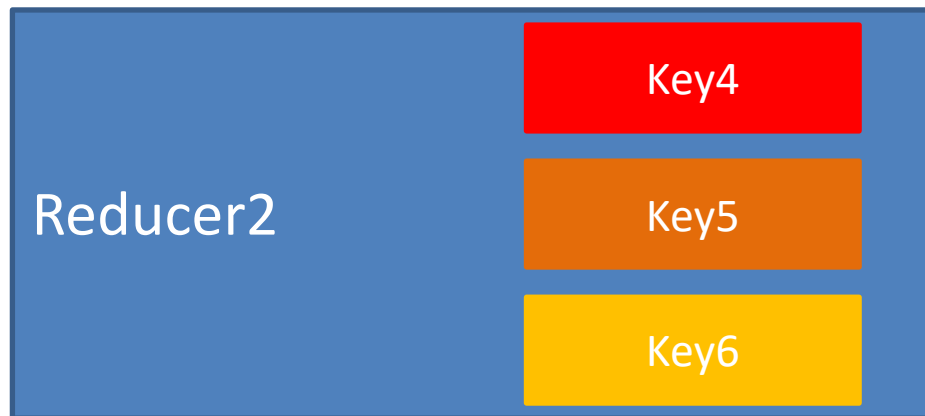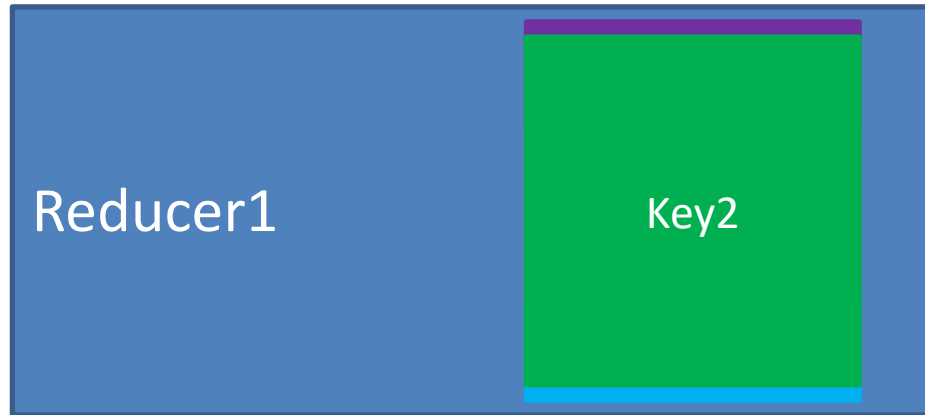# The Most Skewed Key?

NULL
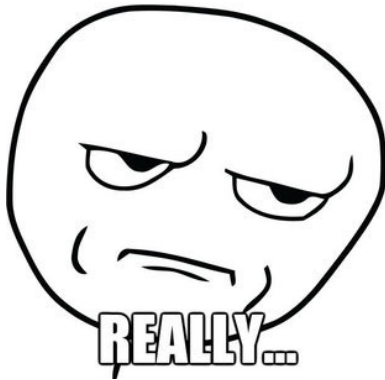
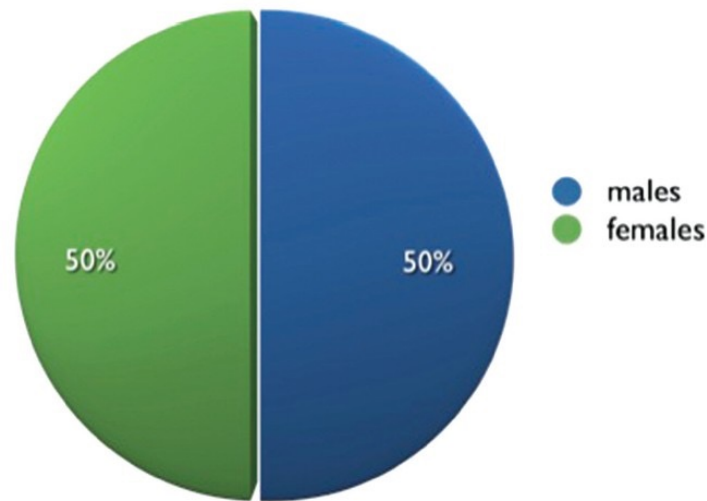# Anti-skew Design

Problem
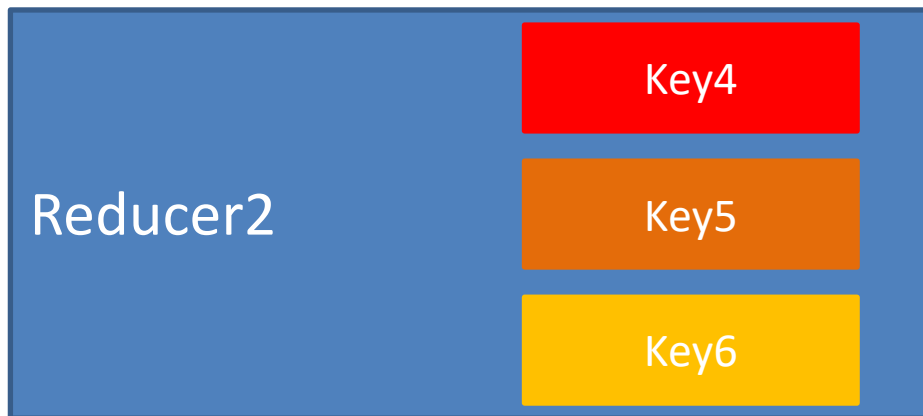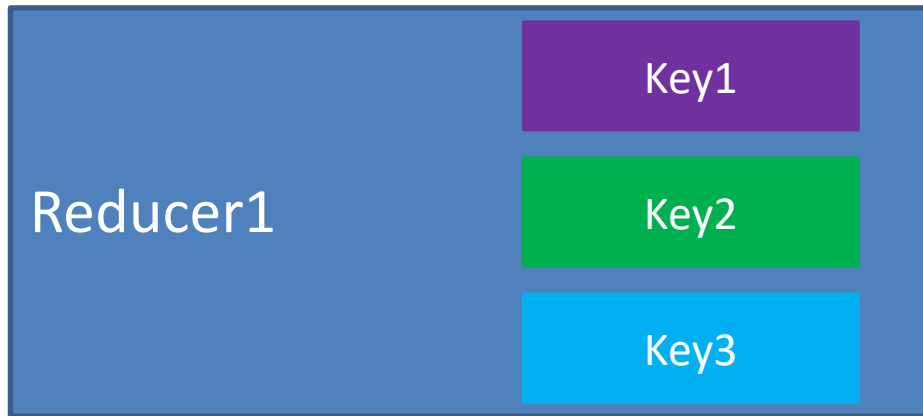
Problem

Assumption1

- The task can be divided into sub-tasks, and can be reassembled back to get the result in an easy way.
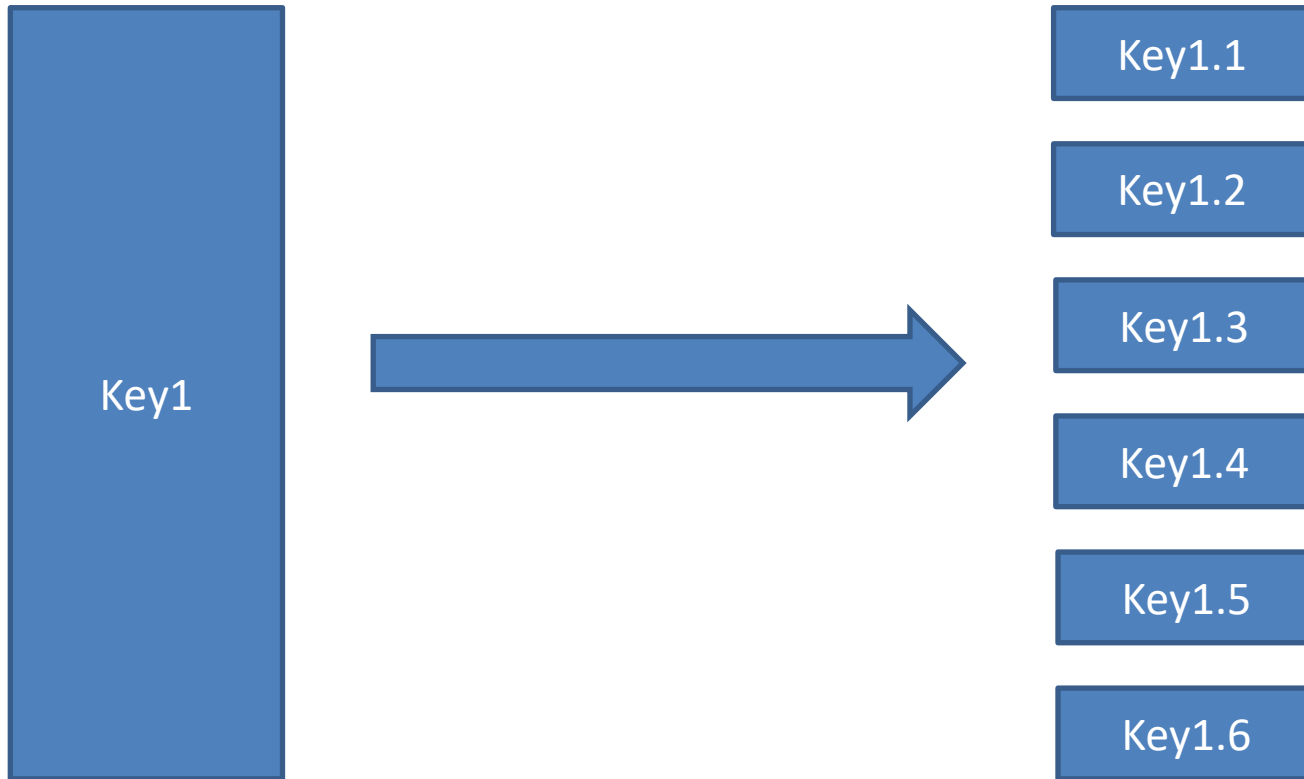
Assumption2

- The key-value pairs in input data are near-equally distributed, which means sampling would be effective; although pre-execution sampling is not required.

Basic Idea

Key1

Key1.1

Key1.2

Key1.3

Key1.4

Key1.5

Key1.6

Skew Perception

Kind	% Complete	Num Tasks	Pending	Running	Complete	Killed	Failed/Killed Task Attempts
map	99.87%	786	0	0	785	0	0 / 0
reduce	33.29%	45	0	1	44	0	0 / 0

All reduce tasks have completed except for one.

Needs visualization!
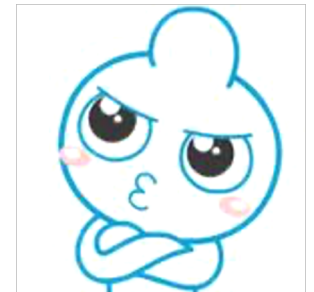
Skew Detection
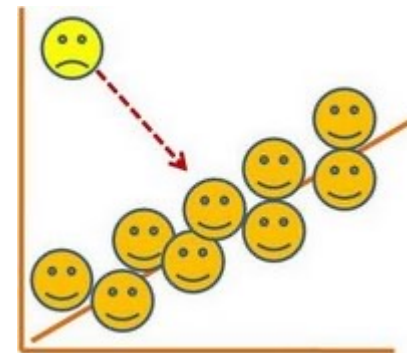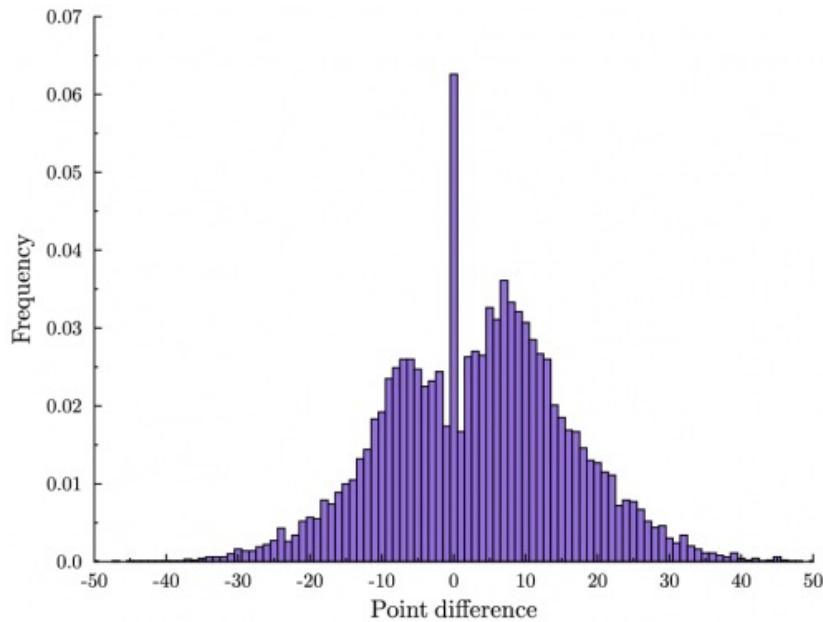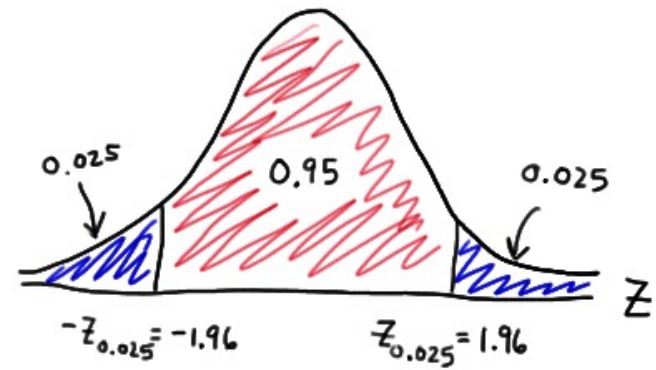
Straggler Identification (tentative)

- A certain key's count is more than 50% (100%? 200%?) of the median one.

Straggler

Straggler Identification (tentative)

Key Splitting

the ──Hash──▶ 8fc42c6ddf9966db3b09e84365034357

6e9b31333e61aad015fa16a3a5fe8e0d

2e20bfee9e4486f0ab651fc0bb988ffb
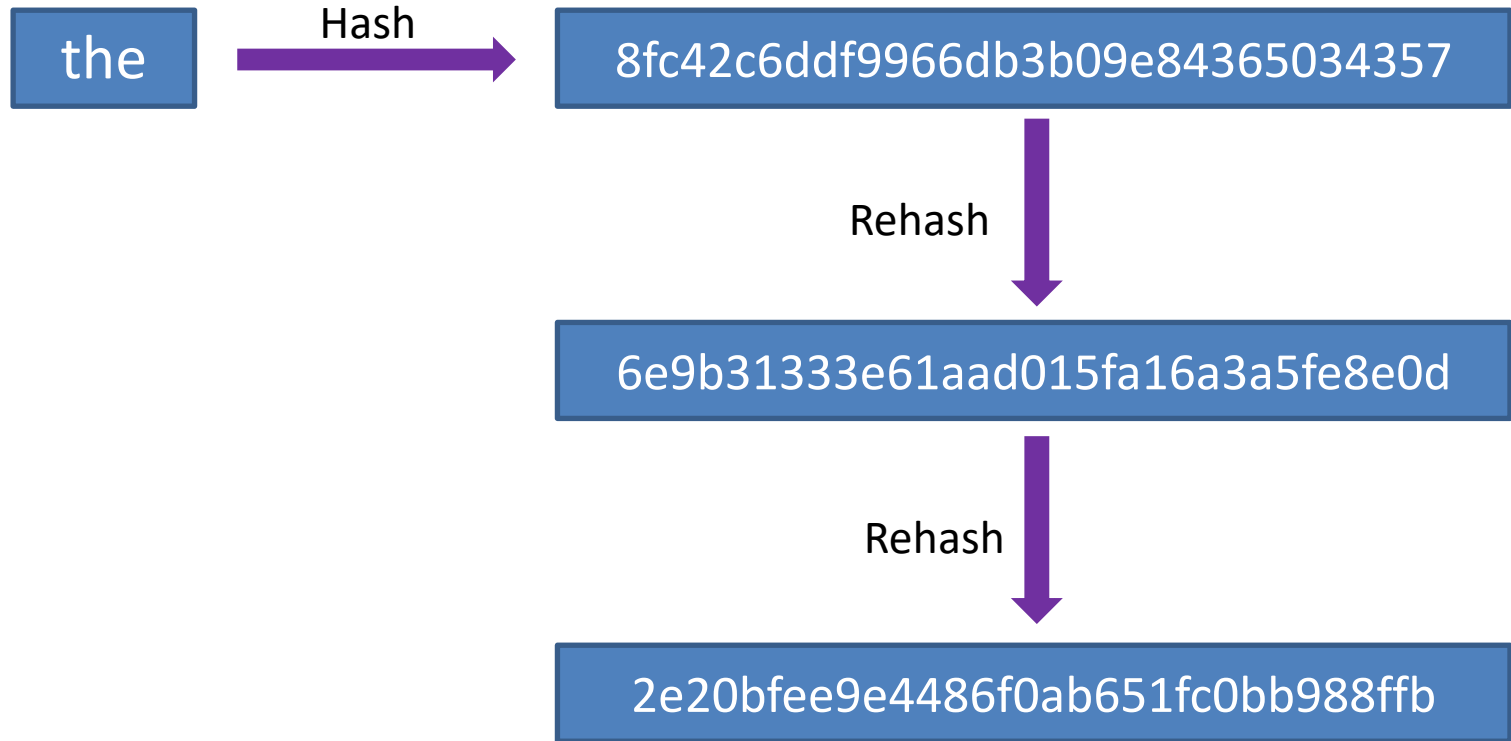
Rehash

Rehash

Key Splitting

Special Key

the

8fc42c6ddf9966db3b09e84365034357

6e9b31333e61aad015fa16a3a5fe8e0d

2e20bfee9e4486f0ab651fc0bb988ffb

Load Balancing (tentative)

- Can the hashing algorithm combined with the platform's partition algorithm evenly distribute the keys to reducers?

Partitioner Function

Privacy (if pre-processed)

Name	Net worth (USD)
Bill Gates	$79.2 billion
Carlos Slim	$77.1 billion
Warren Buffett	$72.7 billion
Amancio Ortega	$64.5 billion
Larry Ellison	$54.3 billion
Charles Koch	$42.9 billion
David Koch	$42.9 billion
Christy Walton	$41.7 billion

Privacy (if pre-processed)

Name	Net worth (USD)
8adc1a86f7	$79.2 billion
8ea0bb9a8f	$77.1 billion
9e640e0fe9	$72.7 billion
abf803fe43	$64.5 billion
bce5c74f58	$54.3 billion
4f589f4867	$42.9 billion
4867dbd572	$42.9 billion
e9ca9f808c	$41.7 billion

Conclusion

- A simple way to handle single-key skew in the MapReduce programming model

- No extra OS-level resources needed

- Implement it as a wrapper, no need to modify platforms' source code, can be used for online platforms (there are so many Hadoop distributions, versions and patches!)

Hadoop Distributions

very old versions of Hadoop

hadoop 0.18

hadoop 0.19

hadoop 0.20

hadoop 0.20.1   hadoop 0.21   hadoop 0.20.2

hadoop 0.20.1 (Greenplum)   hadoop 0.20.1 (IBM)   hadoop 0.20.2-append   MapR   hadoop 0.20.200 (Y! security)

EOL

CDH3

hadoop 0.20.203

hadoop 0.20.205   hadoop 0.20.203 (MS Azure)

hadoop 0.22   hadoop 1.0.0

hadoop 0.23   hadoop 2.0.2-alpha   hadoop 0.22_1   MR1

hadoop 3.0.0 (trunk)   CDH4

Hortonworks Data Platform

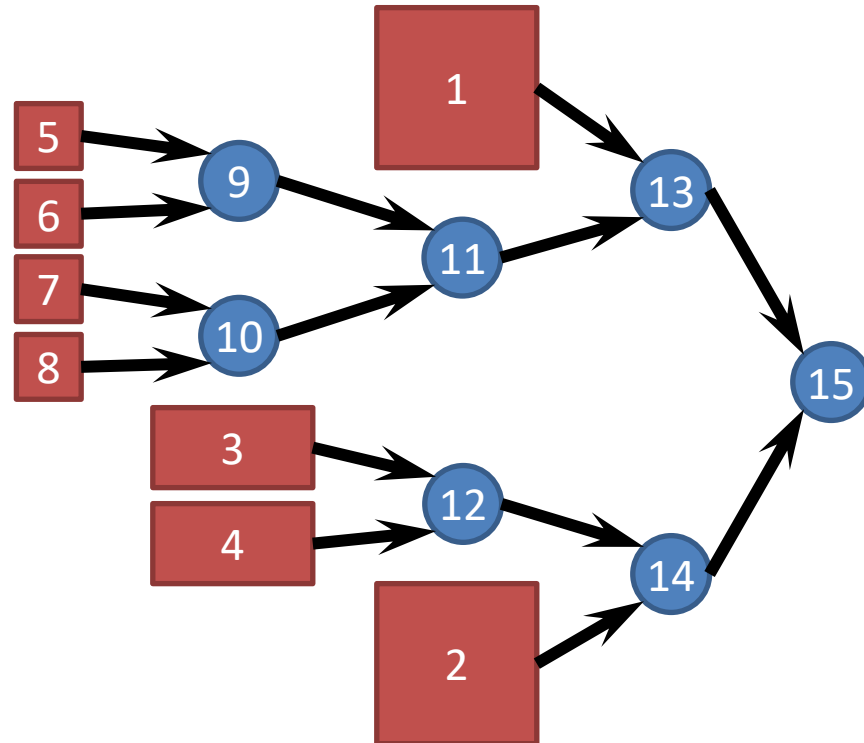|	Hadoop	Pig	HCatalog*	Hive	HBase	Sqoop	Oozie	Zookeeper	Mahout	Ambari
MAY 2013 — HDP 1.3	1.2.0	0.11	0.11.0*		0.94.6	1.4.3	3.3.2	3.4.5	0.7.0	1.2.3
FEB 2013 — HDP 1.2	1.1.2	0.10.1	0.5.0	0.10.0	0.94.2	1.4.2	3.2.0	3.4.5	0.7.0	1.2.0
SEPT 2012 — HDP 1.1	1.0.3	0.9.2	0.4.0	0.9.0	0.92.1	1.4.2	3.1.3	3.3.4		HMC1.1
JUNE 2012 — HDP 1.0	1.0.3	0.9.2	0.4.0	0.9.0	0.92.1	1.4.1	3.1.3	3.3.4		HMC1

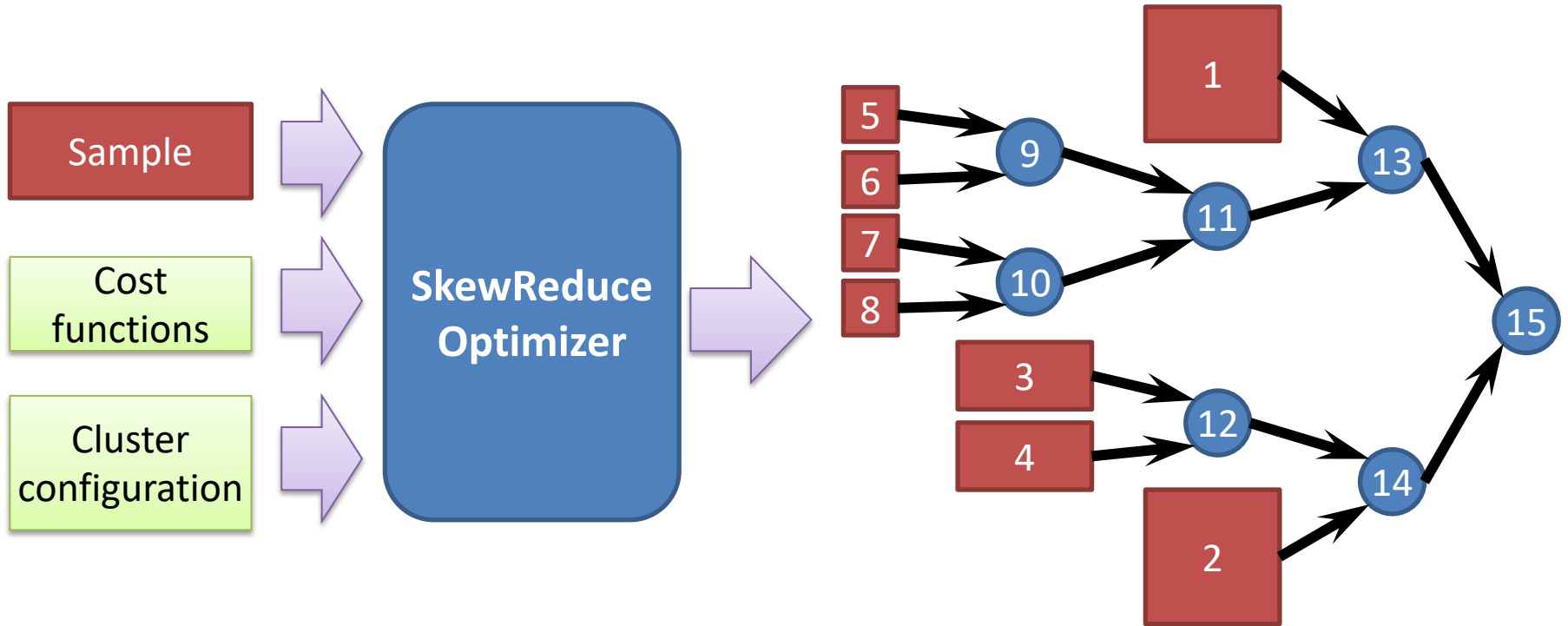*Apache HCatalog is subsumed by Apache Hive 0.11

Related Work

SkewReduce

- Varying granularities of partitions
- Can we *automatically* find *a good partition plan* and *schedule*?
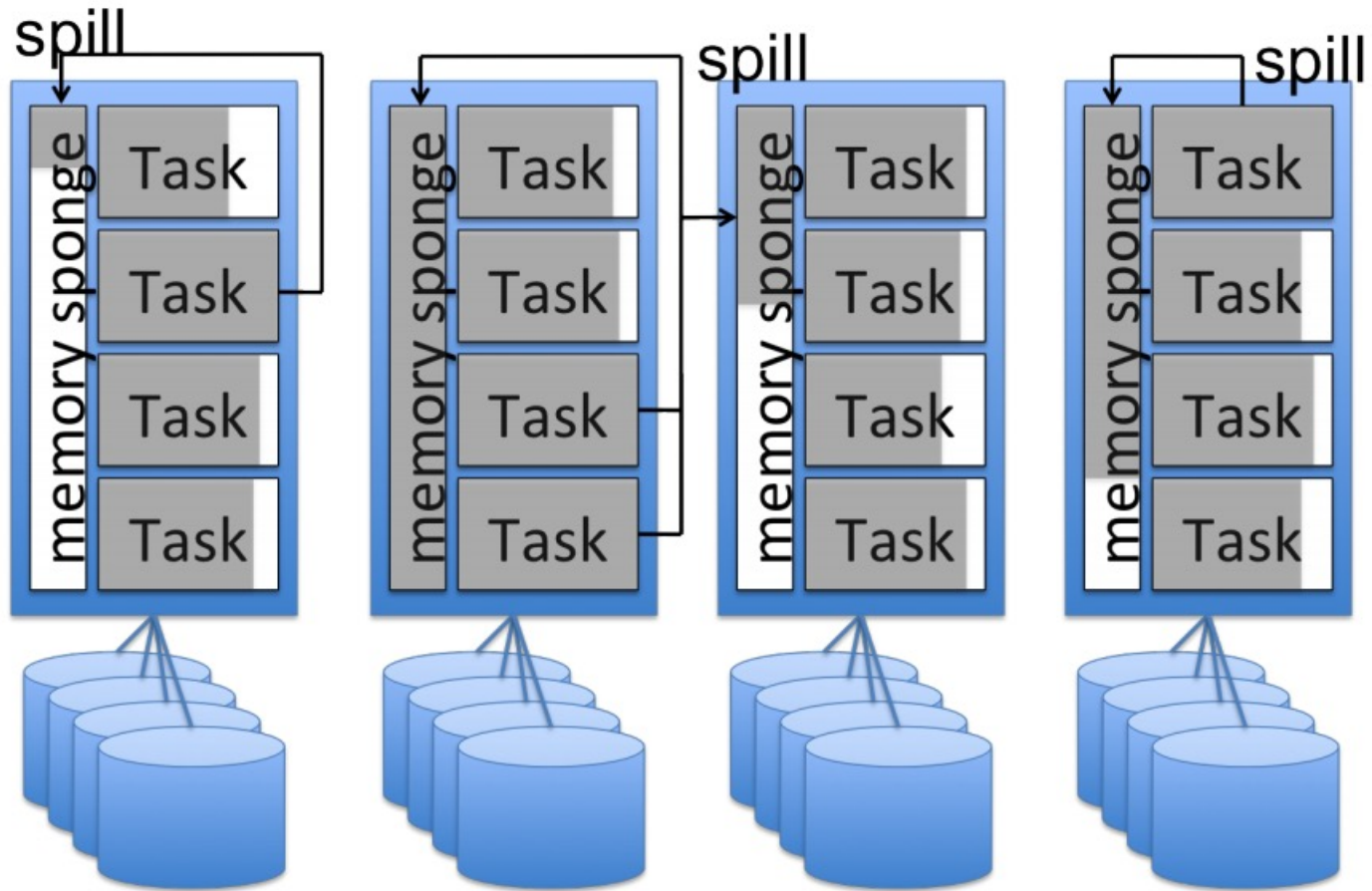
SkewReduce

- *Goal*: minimize expected total runtime

SkewTune

- Does what SkewReduce does when the program is running.

- Skew detected -> Stop -> Repartition -> Continue

SpongeFiles

Q&A
Suggestions?