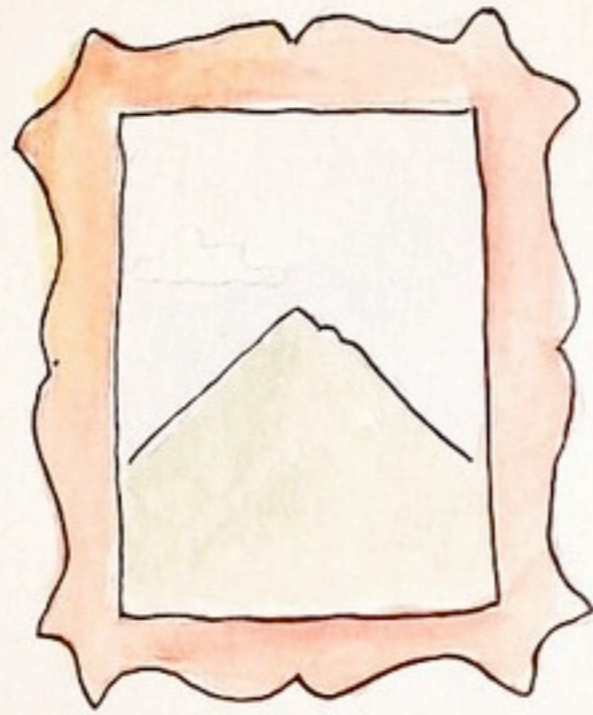
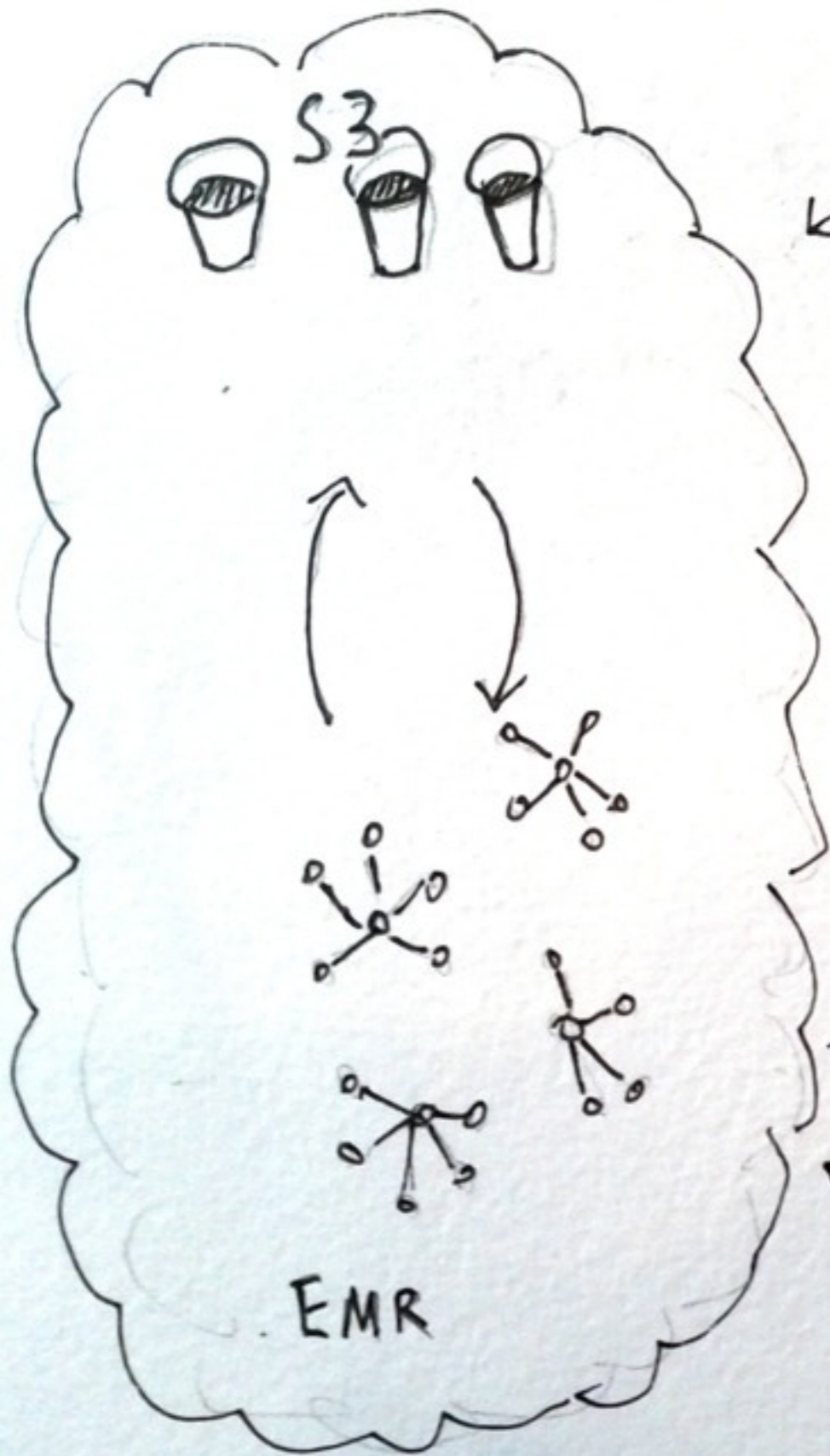


profiling and tuning

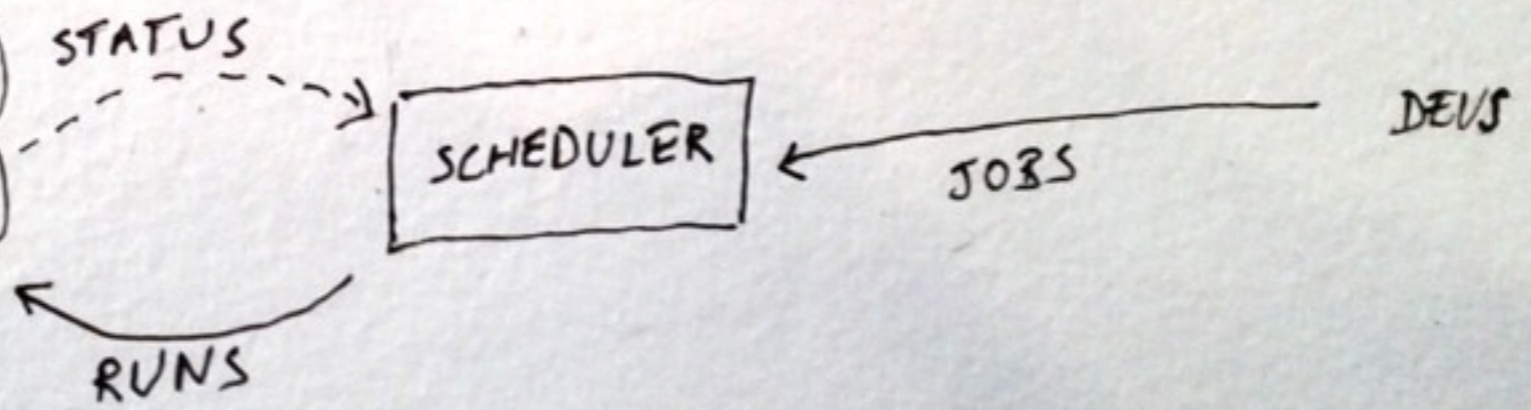
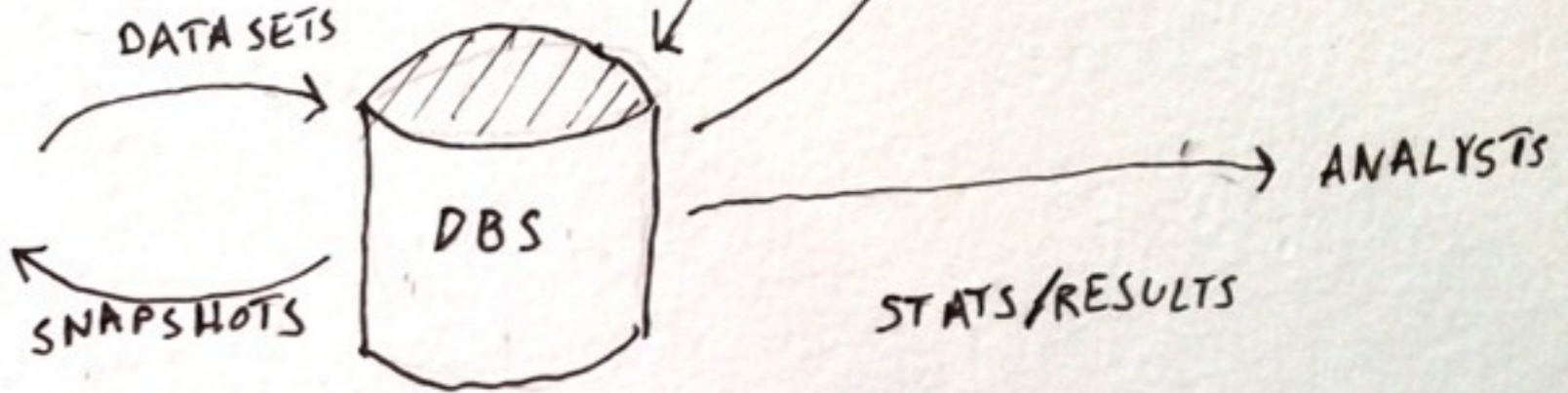
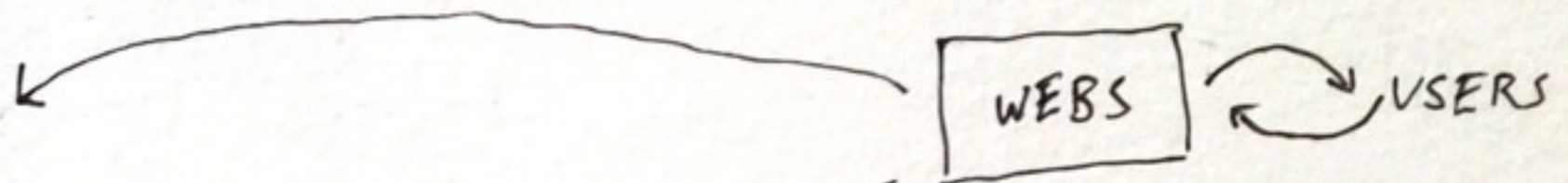
hadoop pipelines

FETTSY





LOG DATA



AWS



POOR VISIBILITY INTO
LOW LEVEL BEHAVIOR
CAUSES BLOAT

ON DEMAND IS OVERKILL
FOR REGULARLY SCHEDULED
(EG NIGHTLY) JOBS

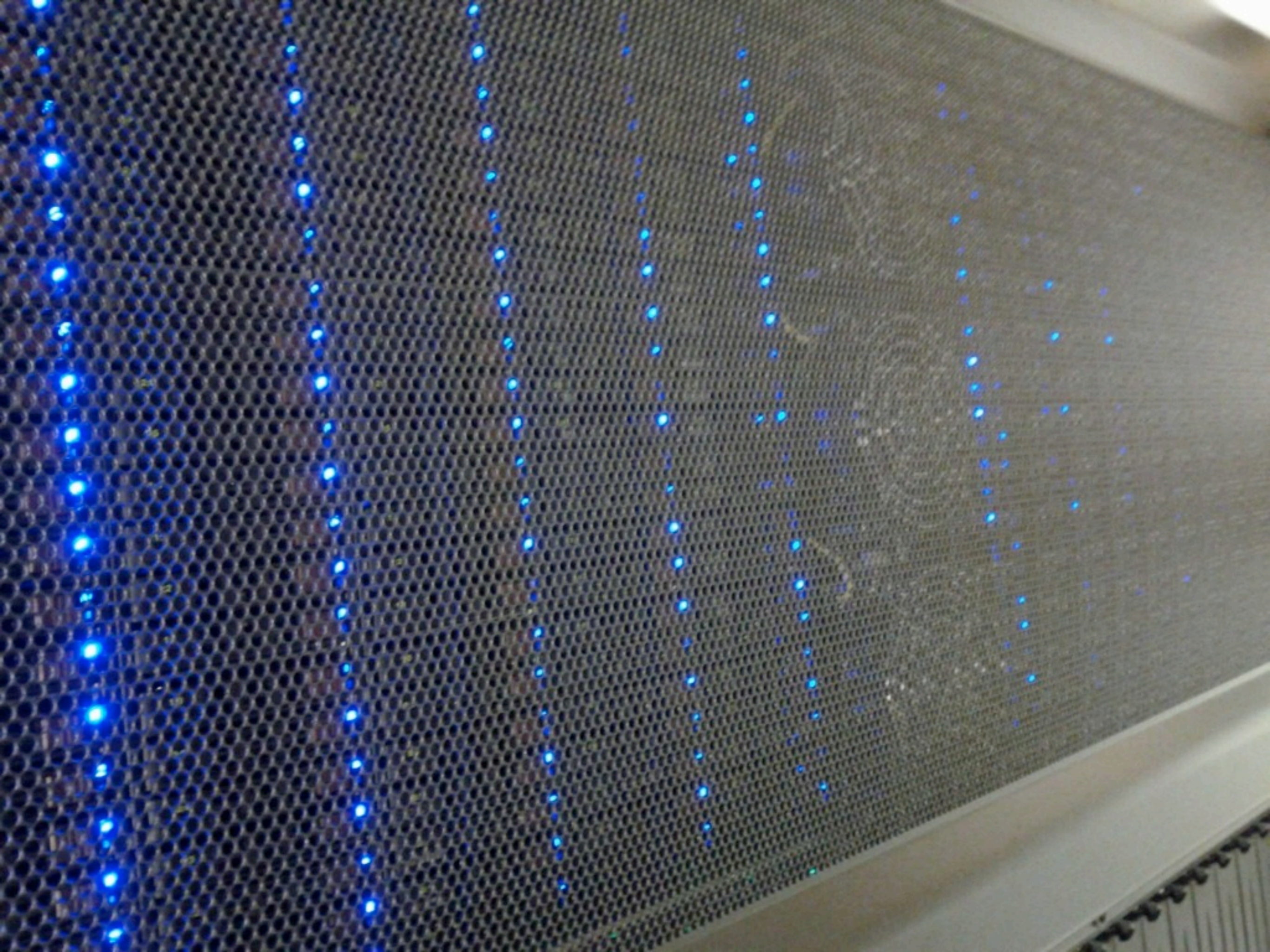
HIGH COST OF FAST FAILURE
POOR FOR DEVELOPMENT

PAYSETUP FOR EACH JOB

PIPE TO S3 IS BOTTLENECK

HOSTS ARE SLOW

AWS AVAILABILITY ISSUES



log parsing

should be lean

collected objects were allocated
 objects collected. Shallow size of collected objects (bytes): 18,746,060,528

Statistics

Class name: (syntax):

Name	GCed Objects	GCed Size	Existing Objects
int[]	51,862,518	3,820,781,088	13
java.lang.String	40,763,997	1,630,559,880	29
char[]	36,296,461	2,279,379,448	28
java.util.regex.Matcher	19,605,532	1,725,286,816	3
scala.util.parsing.input.OffsetPosition	16,364,860	654,594,400	6
scala.util.parsing.input.CharSequenceReader	13,709,425	438,701,600	2

Call tree (all threads together): \1 Call tree (by thread): \2 Method list: \3

Allocation call tree (all threads together) for objects selected in the upper table

Name	GCed Objects	GCed Size	Existing Objects
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$map\$1.apply(Reader)	41,717,551	3,091,998,472	81%
scala.util.parsing.combinator.Parsers\$anon\$2.apply(Reader)	41,717,551	3,091,998,472	81%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$flatMap\$1.apply(Object)	41,717,551	3,091,998,472	81%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$flatMap\$1.apply(Reader)	41,717,551	3,091,998,472	81%
scala.util.parsing.combinator.Parsers\$anon\$2.apply(Reader)	41,710,547	3,091,545,016	81%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$map\$1.apply(Object)	41,710,547	3,091,545,016	81%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$map\$1.apply(Reader)	41,710,547	3,091,545,016	81%
scala.util.parsing.combinator.Parsers\$anon\$2.apply(Reader)	41,710,547	3,091,545,016	81%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$flatMap\$1.apply(Object)	41,710,547	3,091,545,016	81%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$flatMap\$1.apply(Reader)	41,710,547	3,091,545,016	81%
scala.util.parsing.combinator.Parsers\$anon\$2.apply(Reader)	41,703,524	3,091,091,824	81%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$flatMap\$1.apply(Object)	41,703,524	3,091,091,824	81%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$flatMap\$1.apply(Reader)	41,703,524	3,091,091,824	81%
scala.util.parsing.combinator.Parsers\$anon\$2.apply(Reader)	41,435,778	3,071,297,912	80%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$flatMap\$1.apply(Object)	41,435,778	3,071,297,912	80%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$flatMap\$1.apply(Reader)	41,435,778	3,071,297,912	80%
scala.util.parsing.combinator.Parsers\$Success.flatMapWithNext(Function1)	41,292,372	3,060,881,640	80%
scala.util.parsing.combinator.Parsers\$anon\$2.apply(Object)	41,292,372	3,060,881,640	80%
scala.util.parsing.combinator.Parsers\$anon\$2.apply(Reader)	41,292,372	3,060,881,640	80%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$map\$1.apply(Object)	41,292,372	3,060,881,640	80%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$map\$1.apply(Reader)	41,292,372	3,060,881,640	80%
scala.util.parsing.combinator.Parsers\$anon\$2.apply(Reader)	41,292,372	3,060,881,640	80%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$flatMap\$1.apply(Object)	41,292,372	3,060,881,640	80%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$flatMap\$1.apply(Reader)	41,292,372	3,060,881,640	80%
scala.util.parsing.combinator.Parsers\$anon\$2.apply(Reader)	41,285,368	3,060,432,184	80%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$map\$1.apply(Object)	41,285,368	3,060,432,184	80%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$map\$1.apply(Reader)	41,285,368	3,060,432,184	80%
scala.util.parsing.combinator.Parsers\$anon\$2.apply(Reader)	41,267,955	3,059,206,184	80%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$append\$1.apply	41,267,955	3,059,206,184	80%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun\$append\$1.ap	41,267,955	3,059,206,184	80%
scala.util.parsing.combinator.Parsers\$anon\$2.apply(Reader)	41,267,955	3,059,206,184	80%
scala.util.parsing.combinator.Parsers\$anonfun\$rep1\$1.apply	41,267,955	3,059,206,184	80%
scala.util.parsing.combinator.Parsers\$anonfun\$rep1\$1.ap	41,267,955	3,059,206,184	80%
scala.util.parsing.combinator.Parsers\$anon\$2.apply(Re	41,267,955	3,059,206,184	80%
scala.util.parsing.combinator.Parsers\$Parser\$anonfun	41,267,955	3,059,206,184	80%
scala.util.parsing.combinator.Parsers\$Parser\$anon	41,267,955	3,059,206,184	80%
scala.util.parsing.combinator.Parsers\$Failure.a	30,996,971	2,394,728,168	63%
scala.util.parsing.combinator.Parsers\$Parse	30,996,971	2,394,728,168	63%
scala.util.parsing.combinator.Parsers\$Pa	30,996,971	2,394,728,168	63%
scala.util.parsing.combinator.Regexp	20,783,250	1,337,396,560	35%
analytics.CombinedLogFormat.h	10,677,377	685,729,768	18%
scala.util.parsing.combinator.F	10,677,377	685,729,768	18%
scala.util.matching.Regex.f	10,677,377	685,729,768	18%

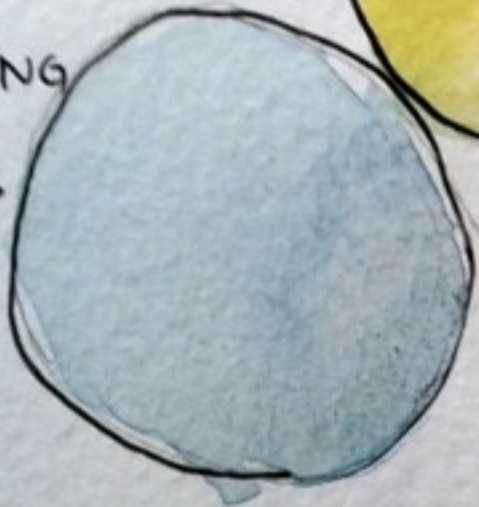
Actions

View for
 section

NATURAL LANGUAGE
PROCESSING



PROGRAMMING
LANGUAGE
IMPLEMENTERS



YOUR
LOGS

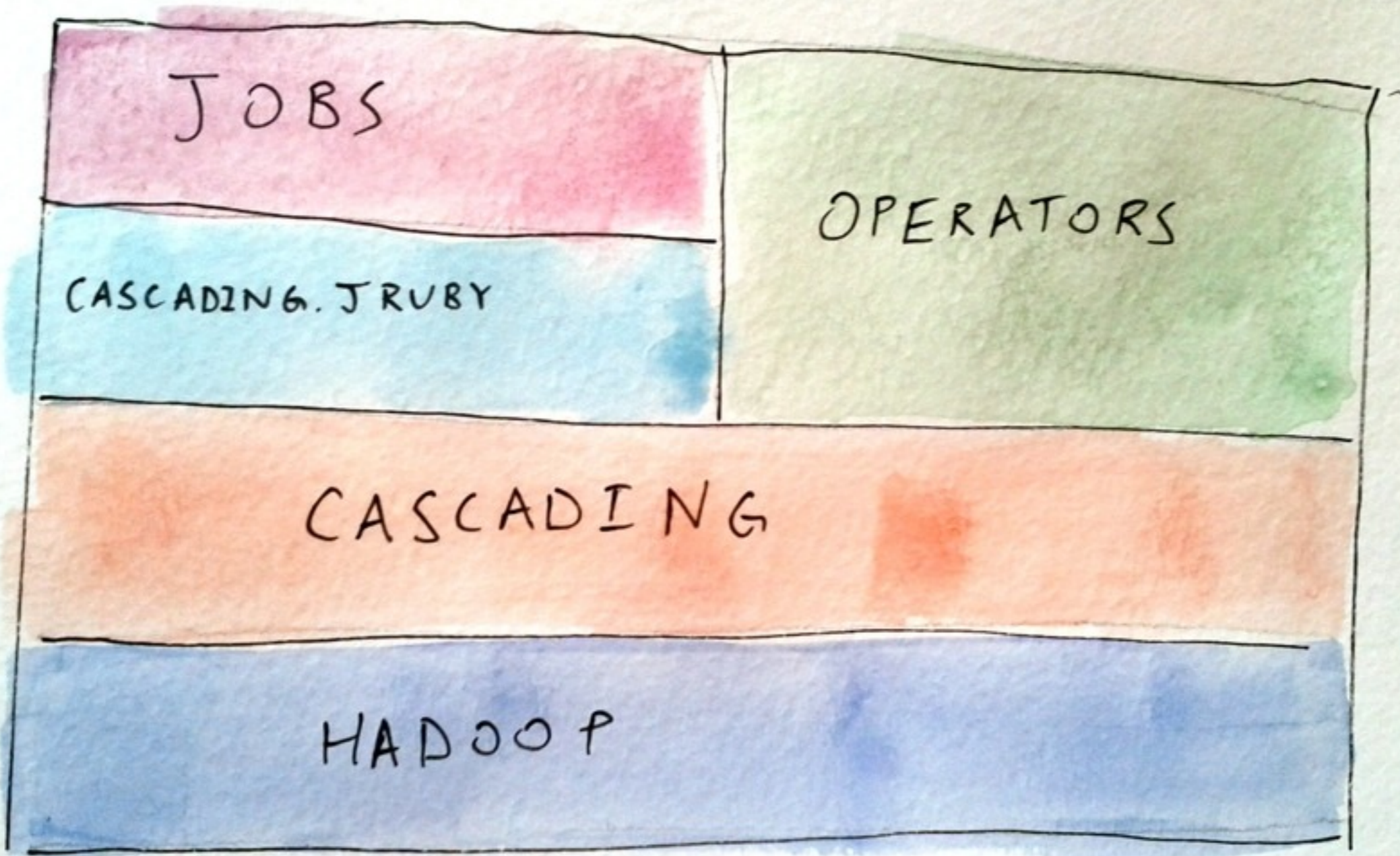
“PARSERS”

Know your
Writables



one canonical

representation



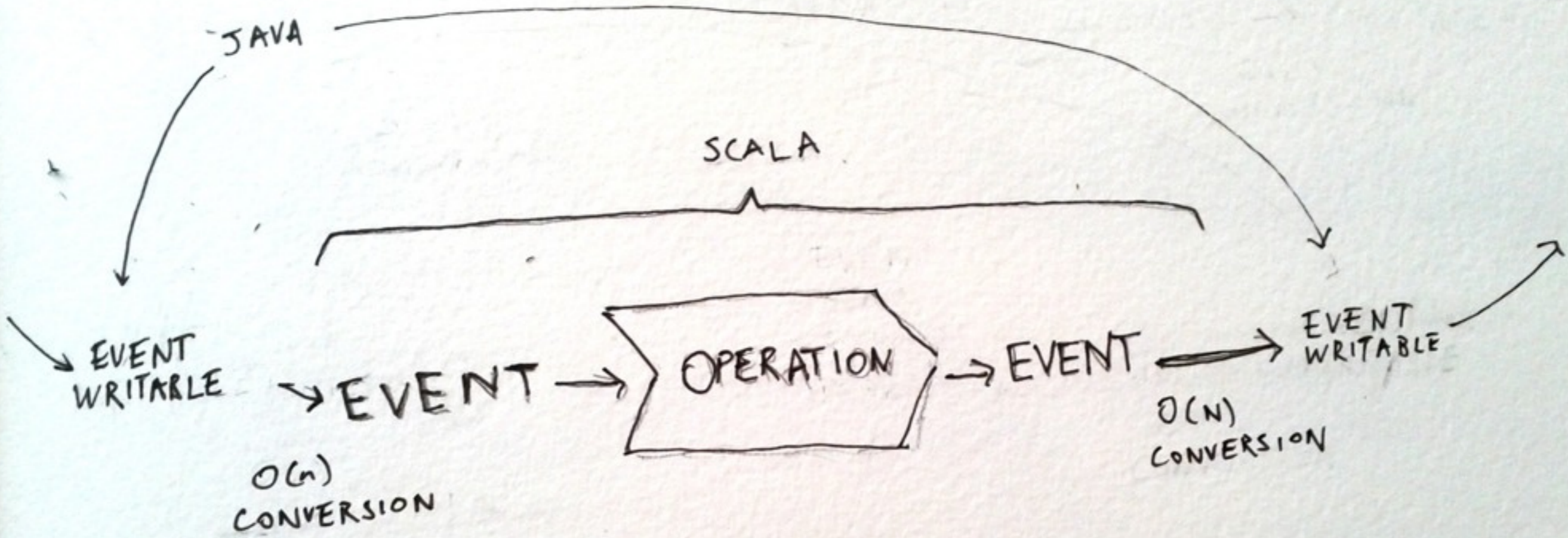
JOBS

OPERATORS

CASCADING. JRUBY

CASCADING

HADOOP



unconfiling

at: scale

metra

hadopp

questions?

contact

abeppu@etsy.com