J S ẻco (2013) 64:1038 1054
DOI 10.1007/ 11227-011-0682-5

P b hed o e: 23 Se e bẻ 2011
S gẻ Sc e ce+B e Med a, LLC 2011

I Ma Red ce ode , a ob d ded o a ẻ e of *map tasks* a d *reduce tasks*. The e ec o e of he ob o o ged b o e o a ẻ o , e - ec a heẻ oge eo e o e . To h he o a a oo a o b e, c e Ma Red ce ched ẻ a ch a bac a o o heẻ ode fẻ each of he o a . Ho e ẻ, ad o a Ma Red ce ched ẻ ca o de ec o a cẻ - ec ce he ca o e a e he oge e of a acc a e (Hadoo ho e age h ://hadoo .a ache.ẻ g/, 2011; Zahẻ a e a . 8 h USENIX o o ẻ a - g e de g a d e e a o , ACM, Ne Yẻ , . 29 42, 2008). To o e h ob e , h a ẻ o o e a H ẻ -ba ed A o-T g (HAT) Ma Red ce ched ẻ , h ch ca c a e he oge e of a acc a e a d ada o he co - o ẻ g e o e a o a ca . HAT e he e gh of each ha e of a a a a d a ed ce a accẻ d g o he a e of he h ẻ a a d e

Q. Che · M. G o (✉) · Q. De g · Y. She
De ẻ e of Co ẻ Sc e ce, Sha gha J ao To g U ẻ , Sha gha , Ch a
e- a : g o- @c . .ed .c

Q. Che
e- a : che - a @ .ed .c

Q. De g
e- a : de g- @c . .ed .c

Y. She
e- a : he _ ao@c . .ed .c

L. Zhe g
H a ho g U ẻ of Sc e ce a d Tech o og , W ha , Ch a
e- a : d8112104@ -a .ac.

L. Zhe g · S. G o
Schoo of Co ẻ Sc e ce a d E g eẻ g, The U ẻ of A , A a a a , Ja a

S. G o
e- a : g o@ -a .ac.

he acc a e e gh of he ha e o ca c a e he og e of c e a . Ba ed o he acc a e-ca c a ed og e of a , HAT e a e he e a g e of a acc a e a d f he a che bac a fo he a ha ha e he o ge e a g e. E e a e ho ha HAT ca g ca o e he e fo a ce of Ma Red ce a ca o o 37% co a ed h Hadoo a d o 16% co a ed h LATE ched e .

H o -ba ed a o- g · Sched g a go h · He e oge eo e o e · Ma Red ce

C e da a e ea ca o eed o oce a ge a d a ge da a e h he e o o of fo a o . D e o he a ge da a e of a ca o , e efe o e o e a d o e oce o o co e oge he o e e he e ec o e of he a ca o ea o ab e a d acce ab e. Th eed ha o o ed he de e o e of Ma Red ce, h ch o e of he o o a og a g a d ched g ode o oce a d ge e a e a ge da a e [8]. Ma Red ce e ab e e o ec f a a f c o ha oce e a e / a e a o ge e a e a e of e ed a e e / a e a , a d a ed ce f c o ha e ge a he e ed a e a e a oc a ed h he a e e ed a e e [8]. Ma Red ce ed C o d Co g he beg g [2, 3, 7, 23, 24]. I a ed b Goog e, oge he h GFS [12] a d B gTab e [4] co g bac bo e of Goog e' C o d Co g a fo . A a o he C o d Co g a fo , Ma Red ce a o o ed o o GPU a d oce o . I add o , a o e e ded o o e o e oo e-co g ob e [6, 9, 10, 18, 20, 22, 27].

Fo a Ma Red ce ob (.e., a a ca o ha e e ed ba ed o he Ma Red ce og a g ode), da a e d ded o a a da a e . Whe a Ma Red ce e a o e ec e a Ma Red ce ob, he Ma Red ce ched e [1] he e a che a a a fo each of he a da a e , a d a che a g o of ed ce a o co ec he e of a he a a . Af e he d o , he Ma Red ce ched e d b e he e a o o d ffe e ode acco d g o he oca o of he a ' da a e . I h a , a he ode (ca ed a *workers*) e ec e he a h ch a e a g ed o he a a e . S ce e e ode eed o e ec e a a , Ma Red ce ched e a che a *task scheduler* fo each ode o a age a . F he o e, ce a Ma Red ce ob o co e ed a he da a oce ed co e e , he e ec o e of he Ma Red ce ob dec ded b he a hed a (.e., he ea e effec).

I o a e o ob e ho oge eo e o e , ce ho oge eo ode e ec e a h he a e da a e e a e. I he e oge eo e o e , o he o he ha d, he e ec o e of a Ma Red ce ob o o ged b he a hed a (ca ed a *straggler tasks*) e o ce o e e e

[1] A Ma Red ce ched e a ched e ha ched e a a d ed ce a .

a o e acco h g e e he a e a d e o he d ffe e ce , ch a ca ac e of co a o a d co ca o , a ch ec e , a d e o e .

O e of he o o a o o of h ob e Ma Red ce a ch g bac a fo agg e a o fa ode. If a Ma Red ce ched e a che a bac a γ_b fo a agg e a γ, he a da a e of γ oce ed co e e he e he γ_b o γ he . I h ca e, f γ_b he befo e γ, he e ec o e of he ob ed ced.

A ho gh c e Ma Red ce ched e o a ch bac a fo agg e a , he fa o de ec agg e a co ec d e o he o g-e a ed e a - g e of a he a [13, 28]. The o g de ec ed agg e a ca e a ea o ob e . F , a ch g bac a fo he e o g agg e a ca o o e he e fo a ce of he Ma Red ce ob ce he ea agg e a o- o g he e ec o e. Seco d, he bac a h ch a e a ched fo he o g agg e a a e e e o ce . The co e o o he e e o ce e e deg ade he o e a e fo a ce of he Ma Red ce ob.

The a g of e e o ce o e of he a ob e of he bac a- eg . C e , Ma Red ce ched e c a f ode o fa ode a d o ode , o ha bac a ca be a ched o fa ode . Ho e e , o ode ca be f - he c a ed o *map slow nodes* a d *reduce slow nodes* a ea e , ce e o b e ha a ode oce e a a fa b oce e ed ce a o a d ce e a. We e a ed ce o ode o e e e he ode ha e ec e a ed ce a o ha o of o he ode . The d g h g be ee a o ode a d ed ce o ode a e e e o ce . Le a e a ed ce a γ ha eed a bac a fo e a e. C e Ma Red ce ched e o a ch he bac a o a o ode N_s. Ho e e , f N_s o a *map slow node*, a ch g he bac a of γ o N_s ca e e o ce o N_s ef c e a d o e he o e a e fo a ce, ce N_s ca oce ed ce a fa .

I o de o de ec agg e a a d a ch bac a ef c e , e o o e a *History-based Auto-Tuning* (HAT) Ma Red ce ched e . HAT co o a e h o - ca fo a o eco ded o each ode o e a a e e a d de ec agg e a d a ca . HAT f he c a e o ode o a o ode a d ed ce o ode . I h a , HAT ca a ch bac a fo a agg e a o ed ce o ode a d a ch bac a fo ed ce agg e a o a o ode .

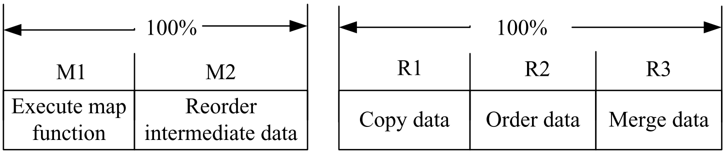
The o o a co b o of h a e a e h ee-fo d:

HAT de ec agg e a co ec ba ed o he h o ca fo a o eco ded o e e ode.

HAT c a e o ode o a o ode a d ed ce o ode f he .

The e e e a e ho ha HAT ched e ca ach e e a e fo a ce ga o 37% fo Ma Red ce a ca o .

The e of h a e o ga ed a fo o . Sec o 2 od ce he bac g o d fo a o a d he o a o of HAT Ma Red ce ched e . Sec o 3 e e he HAT Ma Red ce ched g a go h . Sec o 4 e o he e e a o de a of HAT ched e . Sec o 5 e a a e he e fo a ce of HAT. Sec o 6 de c be he e a ed o . Sec o 7 d a he co c o h o g o f e o .



(a) Map Task (b) Reduce Task

T o ha e of a a t a a d t hree ha e of a ed ce t a

k

C e t Ma Red ce ched e , ch a Hadoo ' ched e [13, 25] a d LATE [28], a ch bac a fo agg e t a t o o e e fo a ce of Ma Red ce a ca t o . The e a e o a o c e t o de ec agg e t a : t he ea t og e o c a d t he o ge e a g e o c . Fo e a e, Hadoo ' ched e e t he ea t og e o c a d LATE ched e e t he o ge e a g e o c . I t he ea t og e o c , a h t he ea t og e a e t he agg e t a . I t he o ge e a g e o c , t he a h t he o ge e a g e a e t he agg e a .

Bo t h o c e o o t he og e of e e a g *progress score* a ge f o 0 t o 1). I c e t Ma Red ce e , t he e ec o of a a a co e t o ha e a d t he e ec o of a ed ce t a co e t hree ha e a ho F g. 1. The efo e, t he og e co e of a a co e f o t he og e co e of e e ha e. C e t Ma Red ce ched e , ch a Hadoo ' ched e a d LATE, a e t ha t *M*1, *M*2, *R*1, *R*2, a d *R*3 a e 1, 0, $\frac{1}{3}$, $\frac{1}{3}$, a d $\frac{1}{3}$, e ec t e .

The og e of a ha e, de o t ed b *PS* ha e, ca be ca c a t ed (1). I t he e a t o , *M* t he be of e / a e a t ha ha e bee oce ed t he ha e a d *N* t he o e a be of e / a e a t ha t eeded t o be oce ed t he ha e.

$$PS_{\text{ha e}} = \frac{M}{N}. \tag{1}$$

Ta e a a γ fo e a e. If γ a a a , ce t he ha e occ e t he o e a og e co e, t he og e co e of γ t he og e *M*1. If γ a ed ce t a a d t he *K* ha e of γ ha hed, ce each ha e occ e $\frac{1}{3}$ of t he og e co e, PS_γ ca c a t ed b add g t he og e co e of t he hed ha e a d t he og e co e of t he c e t ha e. The efo e, t he og e co e of γ, de o t ed b PS_γ, ca c a t ed (2).

$$PS_\gamma = \begin{cases} PS_{\text{ha e}} & \gamma \text{ a a t a ,} \\ \frac{1}{3} * K + \frac{1}{3} * PS_{\text{ha e}} & K \in (0, 1, 2), \gamma \text{ a ed ce t a .} \end{cases} \tag{2}$$

Fo a Ma Red ce t e h *n* g t a $(\gamma_1, \gamma_2, \ldots, \gamma_n)$, t he a e age og e co e of t he *n* g t a , de o t ed b $PS_{\text{a g}}$, ca c a t ed (3).

$$PS_{\text{a g}} = \sum_{i=1}^{n} PS_i/n. \tag{3}$$

S o e a γ_j' og e co e PS_j a d ha T_j eco d $(j \in (1, 2, \ldots, n))$. If he ea og e o c ed o de ec agg e a , γ_j a agg e a o he $PS_j \leq PS_{a\ g} - 20\%$.

O he o he ha d, f he o ge e a g e o c ed, he e a g e of a he n a eed o be ca c a ed f he . The he ched e choo e he a h he o ge e a g e a agg e a . To ca c a e he e a g e of a γ_j, he og e a e of γ_j, de o ed b PR_j, ca c a ed (4). Ba ed o (4), he e a g e of γ_j, de o ed b TTE_j, ca c a ed, (5).

$$PR_j = PS_j / T, \tag{4}$$

$$TTE_j = (1.0 - PS_j)/PR_j = T \times \frac{1.0 - PS_j}{PS_j}. \tag{5}$$

I o ca e , he o ge e a g e o c o be e ha he ea og e o c [28]. Th beca e he a h he ea og e co e a e o a a he a hed a . Fo e a e, a Ma Red ce e ha ha a $(\gamma_1, \gamma_2, \ldots, \gamma_6)$, o e he og e co e a e 0.7, 0.5, 0.9, 0.9, 0.9, a d 0.9, e ec e . We f he o e ha he eed 100, 30, 10, 10, 10, a d 10 eco d o h he o . I h ca e, $PS_{a\ g} = (0.7 + 0.5 + 0.9 * 4)/6 = 0.8$. The ea og e o c a e γ_2 o be a agg e a . Ho e e , γ_1 he ea agg e a ce γ_1 eed o e e o h o .

O he o he ha d, f he og e co e of a he a ca be ca c a ed acc a e , he o ge e a g e o c ca a a de ec o he ea agg e a . Ho e e , c e Ma Red ce ched e ca o ca c a e he og e co e acc a e , ce $M1$, $M2$, $R1$, $R2$, a d $R3$ a e e o be co a a e 1, 0, $\frac{1}{3}$, $\frac{1}{3}$, a d $\frac{1}{3}$. I he ea e ec o , he a e of he a e o a d ffe e fo d ffe e ha d a e e g a d d ffe e Ma Red ce a ca o .

Fo e a e, g e a ode h $R1$, $R2$, a d $R3$ e a o 0.6, 0.2, a d 0.2, e ec e . S o e a ed ce a γ ha h he ha e a d ha T eco d o he ode, he e a g e of γ $T * \frac{1-0.6}{0.6} = 0.67T$ eco d . Ho e e , ce he a e of $R1$, $R2$, a d $R3$ a e a $\frac{1}{3}$ c e ched e , he ca c a ed e a g e of γ $T * \frac{1-1/3}{1/3} = 2T$ eco d . Ba ed o he o g e a g e, he o ge e a g e o c ca o d ea agg e a .

To h e d, e o o e HAT Ma Red ce ched e , h ch e a *history-based auto-tuning strategy*, o e he a e of $M1$, $M2$, $R1$, $R2$, a d $R3$ ba ed o he h o ca a e of he he co e ed a . Ba ed o he ec c a e of he fo he c e ha d a e fea e a d a ca o fea e, HAT ca e a e he og e co e of g a acc a e , a d he ce ca d ea agg e a . HAT e e ed g he o ge e a g e o c .

[illegible]

Th ec o e e HAT, a H o -ba ed A o-T g Ma Red ce ched e . F , e g e a o e e of HAT. The e od ce a h o ca -ba ed a o- g a eg fo g e a a e e ed b HAT. La , e e e he de a ed a go h HAT.

R e a go h of HAT

I : Ke /Va e a . O : S a ca e .
E o ead h o ca fo a o a d e a a e e g he h o -ba ed a o- g a eg a a ed Sec. 3.2.
E o co e oge co e of a he g a g he oge o o g a go h a a ed Sec. 3.3.
HAT oce e a a d de ec agg e a g he agg e de ec g a go h a a ed Sec. 3.4.
HAT de ec o ode (e he a o ode o ed ce o ode) g he o ode de ec g a go h a a ed Sec. 3.5.
HAT a che bac a o a o a e ode g he bac a a ch g a go h a a ed Sec. 3.6.
HAT co ec e a d da e h o ca fo a o o e ode.

3.1 O e e of HAT

HAT de ec agg e a ba ed o he acc a e oge co e a d ach e e be e e fo a ce co a ed h Hadoo a d LATE ched e . A go h 1 he e a go h of HAT.

Whe HAT a o e ec e a Ma Red ce ob, each o e ead he h o ca fo a o f o oca ode a d e he o be he defa a e of he a a e e h e ec o (o be de c bed ho). The h o ca fo a o co a he a e of $M1$, $M2$, $R1$, $R2$, a d $R3$. Ba ed o he d a c- ed $M1$, $M2$, $R1$, $R2$, a d $R3$, HAT ca co e oge co e of a o e acc a e, h ch he ba of agg e a de ec g. Mea h e, HAT de ec o ode acco d g o he a e age oge a e of a a a d ed ce a o e ode (o be de c bed ho). If he e a e a agg e a , HAT a che bac a fo he e agg e a . Af e a he da a e ha e bee oce ed, HAT e a e he Ma Red ce ob a d e o he a e .

3.2 H o -ba ed a o- g a eg

A e o ed befo e, d ffe e o e oce a d ffe e eed. The efo e, each o e ha d ffe e $M1$, $M2$, $R1$, $R2$, a d $R3$. To ob a he fo each o e acc a e , HAT e a h o -ba ed a o- g a eg o e he a e of he d a ca . I he a eg , each o e ead he h o ca a e of $M1$, $M2$, $R1$, $R2$, a d $R3$ f o he co e o d g ode a he defa a e he he o e a a ed. O ce a a he o he o e , $M1$ a d $M2$ a e da ed. O ce a ed ce a he o he o e , $R1$, $R2$, a d $R3$ a e da ed.

Fo a eco ded e gh of ha e (.e., $M1$, $M2$, $R1$, $R2$, $R3$), f he eco ded a e $V_{o d}$ a d he co e o d g a e of he hed a V hed, HAT da e he eco ded a e o V_e ha ca c a ed (6). HP e e e he e gh of o d a e of $M1$ he e a e of .

$$V_e = V_{o d} * HP + V_{hed} * (1 - HP) \quad (6)$$

A ho (6), f HP oo a ge (c o e o 1), V_{e} o de e d o $V_{\text{o d}}$. I h co d o , V_{e} ca o e ec he -o-da e fea e of he c e g a . O he o he ha d, f HP oo a (c o e o 0), he a o a e a e of he e gh a be de o ed b a do fac o , ce V_{hed} e o be e ced b a do e e .

I add o , he e o a add o a co ca o he a o ead a d da e h o ca fo a o , ce e e o e ead a d e h o ca fo - a o f o oca ode. So HAT ched e ca ab e.

3.3 P og e o o g a go h

D g he e ec o of a Ma Red ce a ca o , HAT co e he og e co e of a he g a e od ca (e e 100). G e a a γ ha g HAT. S o e K ha e of γ ha bee hed, (7) a d (8) co e he og e co e of γ, h ch $PS_{\text{ha e}}$ co ed acco d g o (1).

$$\text{Fo a a :}\quad PS_{\gamma}=\begin{cases} M1\times PS_{\text{ha e}} & \text{f } K=0,\\ M1+M2\times PS_{\text{ha e}} & \text{f } K=1.\end{cases} \tag{7}$$

$$\text{Fo ed ce a :}\quad PS_{\gamma}=\begin{cases} R1\times PS_{\text{ha e}} & \text{f } K=0,\\ R1+R2\times PS_{\text{ha e}} & \text{f } K=1,\\ R1+R2+R3\times PS_{\text{ha e}} & \text{f } K=2.\end{cases} \tag{8}$$

3.4 S agg e a de ec g a go h

Ba ed o he acc a e og e co e of a ha a e co ed acco d g o (7) a d (8), HAT de ec agg e a acco d g o A go h 2. A a γ a agg e a o he f he fo o g o e c o . F , γ a o a . Seco d, γ o e of he a h he o ge e a g e.

Le PR_{γ} a d $PR_{\text{a g}}$ g e e he og e a e of γ a d he a e age og e a e. γ a o a o he og e a e f (9). *Task_Cap* a ca of o o o o he e a o .

$$PR_{\gamma}<(1.0-Task\_Cap)\times PR_{\text{a g}}. \tag{9}$$

Acco d g o (9), f Ta _Ca oo a (c o e o 0), HAT ca f o e fa a o o a . O he o he ha d, f Ta _Ca oo a ge (c o e o 1), HAT c a f o e o a o fa a .

Fo a he o a , HAT co e he e a g e of he acco d g o (5). HAT choo e ho e o a h he o ge e a g e o be ag- g e a g he o ge e a g e o c . To ha d e he fac ha bac a of agg e a co e o ce , HAT he be of agg e a a d bac a . The efo e, a ca o he be of agg e a (.e., he be of bac a ce HAT o a o o e bac a fo each agg e a), de- o ed b *Strag_Cap*, ed. S o e he be of he o e a g a *Task_Num*, he -bo d of he be of agg e a , de o ed b *Strag_UB*, S t agg e t a de t ec t g a go t h

(t he ob t g)

E e d e co t e og e a e of e t a t ha a e g o t .
HAT co t e t he a e age og e t a e of a t he g t a .
E e d e de t e e o a acco d g t o (9).
E e d e e d t t he t of o a g o t .
HAT co t e t he e a g t e fo a t he o a acco d g t o (5) a d d de t he t a de ce d g d de acco d g t o t he e a g t e.
HAT co t e t he -bo d of t he be of t agg e t a , S t ag_UB.
(t he be of o a ≤ S t ag_UB)
A t he o a a e de t ec t ed a t agg e t a .
(t he be of o a > S t ag_UB)
HAT e ec t S t ag_UB o a h t he o ge t e a g t e a t agg e t a .
HAT e t a t he t agg e t a t o t agg e a ed ce t a t .
ee (100000); //HAT de t ec t t agg e t a e e 100 .

Strag_Cap ∗ *Task_Num*. If S t ag_Ca t oo a (c o e t o 0), o e ea t agg e t a o e oo ed b HAT. O t he o t he ha d, f S t ag_Ca t oo a ge (c o e t o 1), t he e ca be t oo a t agg e t a . Too a bac t a fo t he e t agg e t a co t a o t of t e e o ce . A go t h 2 t t he de t a ed t agg e de t ec t g a go t h .

3.5 S o ode de t ec t g a go t h

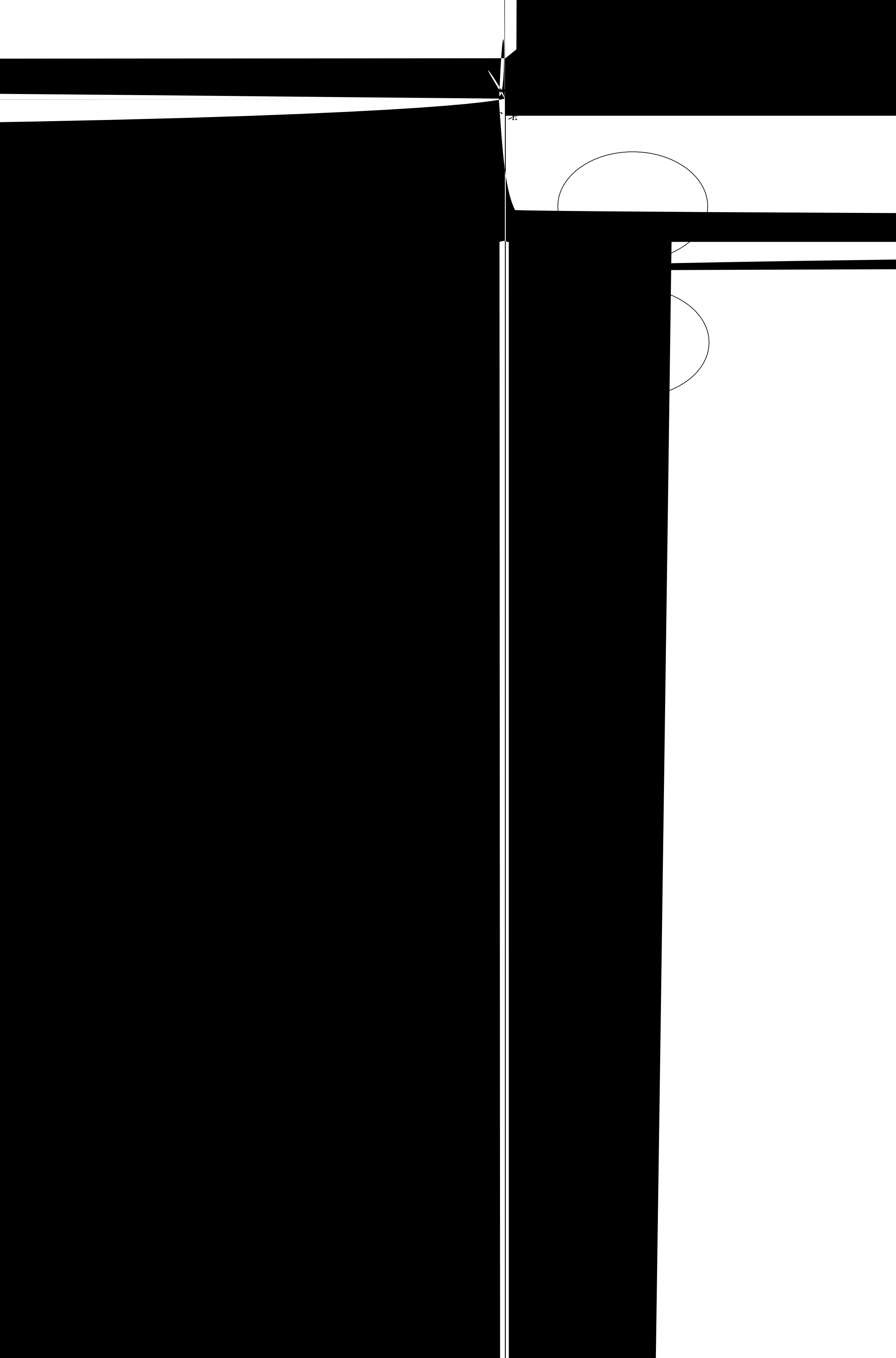
To de t ec t o ode t he t e , HAT e t he a e age og e a e of t he g a ed ce t a o a ode t o e e e t he a ed ce t a og e t a e of t he ode. The t ode h t he t a e t a ed ce t a og e a e a e a ed ce o ode . G e a ode Φ h M a a a d R ed ce t a . The a ed ce t a og e a e of Φ, de o t ed b MR_{Φ} a d RR_{Φ}, a e ca c t a ed (10). PR_i t he og e t a e of t he *i* t h a ed ce t a .

$$\begin{cases} MR_{\Phi} = \frac{\sum_{i=1}^{M} PR_i}{M}, \\ RR_{\Phi} = \frac{\sum_{i=1}^{R} PR_i}{R}. \end{cases} \qquad (10)$$

Fo ode Φ, f $MR_{\Phi} < (1 - Node\_Cap) * MR_{a\ g}$, t a a o ode. If $RR_{\Phi} < (1 - Node\_Cap) * RR_{a\ g}$, t a ed ce o ode. $MR_{a\ g}$ a d $RR_{a\ g}$ a e t he a e age a ed ce t a og e t a e of a t he ode . *Node_Cap* a ca of o o d t o of t he o ode.

T he efo e, f *Node_Cap* t oo a (c o e t o 0), HAT ca f o e fa t ode t o o ode . O t he o t he ha d, f *Node_Cap* t oo a ge (c o e t o 1), HAT c t a f e o t ode t o fa ode .

To t t he be of t o ode , a ca o t he be of o ode , de- o t ed b *SN_Cap*, t od ced HAT. S o e t he be of t he e a ode



```
<HAT>
        <MAP><M1>0.80</M1><M2>0.20</M2></MAP>
        <REDUCE><R1>0.59</R1><R2>0.19</R2><R3>0.22</R3></REDUCE>
</HAT>
```

A e a e of $M1$, $M2$, $R1$, $R2$, a d $R3$ t ha t a e eco ded XML fo a t

he co e o d g a oo . I h a , HAT a a a che bac a fo he agg e a h ch o o g he e ec o e o e o .

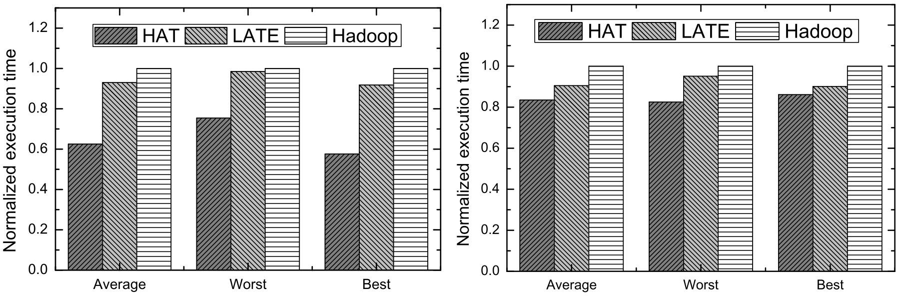
A e o ed befo e, e e ode eco d he a e of $M1$, $M2$, $R1$, $R2$, a d $R3$, h ch a e ec he e ec o fea e of a o he ode. Fo ea a a g, he a e a e o ed XML fo a , a ho F g. 3. E e ode e a XML a e o ead he o ed a e a d a e he a he defa a e of $M1$, $M2$, $R1$, $R2$, a d $R3$ he c e e ec o .

I HAT, a he ode efe e ec g oce ed a f o he *unprocessed task pool* a he ha a ch g bac a . D e o he a ge da a e of a , a he ode efe o e ec e a ho e da a e o ed o oca ode.

Th ec o e a a e he e fo a ce of HAT. D e o ha d a e , e e a c e ha co e f o e co e . I o de o a e he e oge eo e - o e , e ha e a ed d ffe e be of a ach e o he ho oge- eo co e . Each co e ha 1 GB RAM a d each a ach e

Se g of ha d a e e

Se g	VM/PC	No. of PC	No. of ode	W e a e (MB/)
Fa e g	1	1	1	2.87
	2	3	$2 * 3 = 6$	1.40
	ba e L	1	1	3.43
S o e g	1	1	1	2.87
	2	2	$2 * 3 = 4$	1.40
	2	1 e e e o PC	2	1.34
	ba e L	1	1	3.43



(a) Sort (b) WordCount

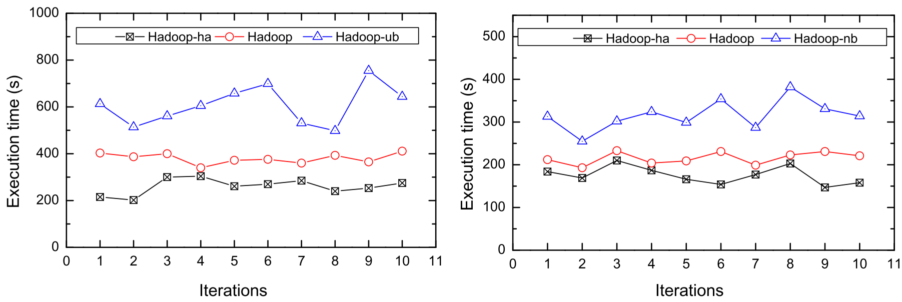
E ec o e of *Sort* a d *WordCount* he o e g

o he ha d, bo h *Sort* a d *WordCount* ha e ach e ed a gh e fo a ce ga h he LATE ched e , h ch e he o ge e a g e o c a e . The e fo a ce e of HAT a e co ec ed h be co g ed a a e e (.e., HP, Ta _Ca , Node_Ca , SN_Ca , a d S ag_Ca). We de be he a o choo e he a o a e a e fo he ho .

A ho F g. 4, bo h be ch a ach e e a gh be e e fo a ce LATE co a ed h Hadoo . The e fo a ce ga o g f o he o ge e- a g e a eg a ch g bac a [28]. S ce HAT e a e he og e of a acc a e , ca de ec agg e a o e acc a e. The efo e, *Sort* a d *WordCount* ach e e be e e fo a ce HAT co a ed h LATE.

5.2 Effec e e of bac a eg a d h o -ba ed a o- g a eg

I o de o e a a e he effec e e of bac a eg a d he h o -ba ed a o- g a eg HAT, e co a e he e fo a ce of *Sort* ha ched ed b Hadoo , Hadoo - b (Hadoo ho bac a eg) a d Hadoo -ha (Hadoo h h o -ba ed a o- g a eg) o bo h he fa e g a d he o e g. *Word-Count* ho a e . F g e 5 ho he e fo a ce of *Sort* ched ed b he hee ched e o he o e g , e ec e . F o he g e, e ca ee ha he bac a eg ca g ca o e he e fo a ce of Ma Red ce



(a) Slow setting (b) Fast setting

E ec t o t e of *Sort* o t he o e tt g a d t he fa t e tt g

The ecod ded ea a e of $M1$, $M2$, $R1$, $R2$, a d $R3$

	Ma t a		Red ce t a		
	$M1$	$M2$	$R1$	$R2$	$R3$
Node1	0.8/0.78	0.2/0.22	0.59/0.62	0.19/0.23	0.22/0.15
Node2	0.77/0.77	0.23/0.23	0.46/0.42	0.06/0.03	0.48/0.55
Node3	0.75/0.66	0.25/0.34	0.44/0.40	0.43/0.45	0.24/0.15
Node4	0.74/0.77	0.26/0.23	0.62/0.64	0.13/0.06	0.25/0.32
Node5	0.81/0.82	0.19/0.18	0.43/0.44	0.14/0.04	0.43/0.52
Node6	0.73/0.77	0.27/0.23	0.51/0.53	0.19/0.12	0.30/0.35
Node7	0.71/0.67	0.29/0.33	0.51/0.50	0.11/0.06	0.38/0.44
Node8	0.79/0.78	0.21/0.22	0.46/0.41	0.13/0.48	0.41/0.11

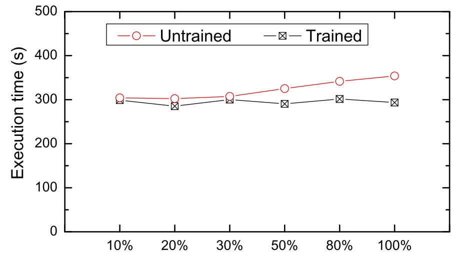
a ca t o h e t he bac t a eg ca e ha ce t he t ab t of e ec t o t e a e .

To de o t a e HAT ca t e t he a e of $M1$, $M2$, $R1$, $R2$, a d $R3$ acc t a e g t he *history-based auto-tuning strategy*, Tab e 2 t he eco ded a e a d t he ea a e of t he o e ode. Fo a a , t he d ffe e ce be ee ea a e a d t he eco ded a e a e e ha 5%. Fo ed ce a , o ca e , he d ffe e ce be ee ea a e a d t he eco ded a e a e e ha 10%. Bo h t he eco ded a e a d t he ea a e a e fa t o t he co a a e of t he e - o ed Hadoo ' ched e a d LATE ched e (.e., 1, 0, $\frac{1}{3}$, $\frac{1}{3}$, a d $\frac{1}{3}$). Ba ed o t he acc a e a e of t he , HAT de ec t he agg e a acc a e . The efo e, HAT ca o e t he e fo a ce of Ma Red ce a ca o t he e oge eo e - o e t .

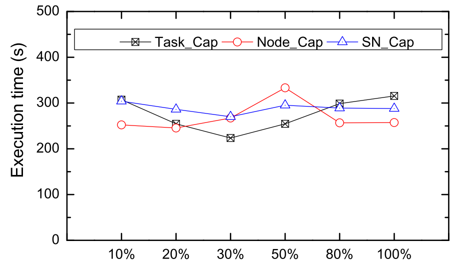
5.3 A o a e a a e e HAT

HAT e e a a e e (.e., HP, Ta _Ca , Node_Ca , SN_Ca , a d S ag_Ca) t o co g e t he ched e fo d ffe e t ha d a e a ch ec e a d d ffe e t a ca- t o . I o de o d a o a e a e fo t he a a e e , e e o e a a e e t h e ee g a t he o t he a a e e t a c. Fo o g t he t e a go t h of HAT

Pe fo a ce of *Sort* h d ffe e HP he a ed a d a ed ce a o



Pe fo a ce of *Sort* h d ffe e Ta _Ca , Node_Ca a d SN_Ca



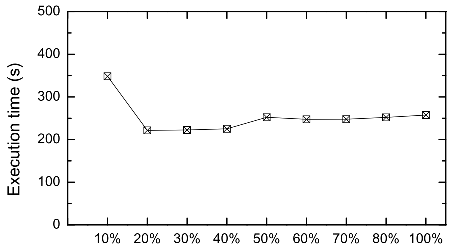
ha o o ed Sec . 3.1, e e a a e he e fo a ce of *Sort* h d ffe e HP, Ta _Ca , Node_Ca , SN_Ca , a d S ag_ Ca . E e e o *WordCount* ho a e .

F g e 6 ho he e fo a ce of *Sort* h d ffe e HP o ce a o : - a ed ce a o a d a ed ce a o. HP he e gh of he eco ded a e of *M*1, *M*2, *R*1, *R*2, a d *R*3 he e a e of he a de ed Sec . 3.2. We co c he a ed ce a o b e ec g *Sort* fo o e befo e he c e e ec o a d co c he a ed ce a o b e g he eco ded *M*1, *M*2, *R*1, *R*2, a d *R*3 o 1, 0, $\frac{1}{3}$, $\frac{1}{3}$, a d $\frac{1}{3}$ a a e he a o Hadoo a d LATE ched e . F o he g e, e ca ee ha he a e of HP doe o affec he e fo a ce of *Sort* oo ch he a ed ce a o. Ho e e , he e fo a ce of *Sort* deg ade h he c ea g of HP he a ed ce a o. The h gh a d a c e fo a ce of *Sort* he a ed ce a o e ed f o he e - a ed a e of *M*1, *M*2, *R*1, *R*2, a d *R*3. F g e 6 a o gge o e a HP f a a ca o e ec ed fo he e. I h a , he a e of *M*1, *M*2, *R*1, *R*2, a d *R*3 ca be ed ba ed o he c e e ec o a d . No e ha , f HP e a o 100%, he a e of *M*1, *M*2, *R*1, *R*2, a d *R*3 he c e o be a o he eco ded a e. I h ca e, HAT ched e he a e o he LATE ched e he a ed ce a o. I he fo o g e e , HP 20%.

F g e 7 ho he e fo a ce of *Sort* h d ffe e Ta _Ca , Node_Ca , a d SN_Ca he e g. F o he g e, e ca ee ha he be a e of *Task_Cap*, *Node_Cap*, a d *SN_Cap* a e 30%, 20%, 30%, e ec e .

Ta _Ca he e ce e of eed be o h ch a a be co de ed oo o o be a o a a de ed Sec . 3.4. A ho F g. 7, *Sort* ga be e fo a ce he Ta _Ca 30%. Ded ced f o (9), he a e Ta _Ca he o e a a e c a ed o o a . The efo e, f Ta _Ca a e ha 30%, o e fa a a e c a ed o o a a d e e agg e a . I h ca e, he a ch g of bac a fo he e o g-c a ed a co e a o of e

The ě fơ a ce of *Sort* h d ffě e t S t ag_Ca



e o ce , o t he o ě a e ec t e t e o o ged. O t he o t hě ha d, f Ta _Ca ǎ gě ha t 30%, o e o a a d e e agg ě t a ǎ e c a ed t o fa t a a d o e bac t a ǎ ě a ched fơ t he . The e t o a o o g t he e ec t e t e a e .

Node_Ca he ě ce t e of eed be o h ch a ode be co dě ed oo o o be a a ked ce t o ode a de ed Sec . 3.5. A ho F g. 7, *Sort* ga be ě fơ a ce he Node_Ca 20%. If Node_Ca a t ha 20%, o e fa t ode ǎ e ea ed a a ked ce o ode b fa . I t h ca e, t he co - t g t o ě of t he e o g-c a ed ode ca o be ed t o o e t he ě fơ - a ce b e ec t g bac a fơ agg ě a . O t he o t hě ha d, f Node_Ca ǎ gě ha 20%, o e a ked ce t o ode ǎ e c a ed t o fa ode b fa . I t he ca e, bac a a be a ched o t he e o ode . S ce t he bac t a o o ode be hed a t ě t ha t he ơ g a t agg ě t a , t he o ě a e ec t e e ca o be hơ e ed.

SN_Ca ed t o t he a bě of o ode a de ed Sec . 3.5. A ho F g. 7, *Sort* ga be ě fơ a ce he SN_Ca 30%. SN_Ca ef f Node_Ca ha a a o a e a e, ce SN_Ca t he bě of o ode . SN_Ca g ǎ a ee t ha t hě e ǎ e t o oo a ode ǎ e c a - ed t o o ode . If SN_Ca a ě t ha 30%, o e a ked ce o ode a be c a ed t o fa ode f Node_Ca oo a (e.g., a ě ha 20%). O t he o t hě ha d, f SN_Ca ǎ gě ha 30%, o e fa ode a be c a ed t o o ode f Node_Ca oo ǎ ge (e.g., ǎ gě ha 30%). The o g c a ca t o ead t o t he oơ ě fơ a ce a de ď bed t he a ǎ ag a h.

S t ag_Ca ed t o t he a bě of bac a a de ed Sec . 3.4. A ho F g. 8, *Sort* ga be ě fơ a ce t he S ag_Ca e a t o 30%. If S ag_Ca a ě t ha 20%, HAT ca o a ch bac a fơ a t he agg ě a beca e of t he a bě of bac a . O t he o t hě ha d, f S ag_Ca ǎ gě ha 0.2, oo a bac a co e a o t of e e o ce , o t he e ec o e of *Sort* o o ged.

Afě a ě e of e ě e , t he be ǎ a e ě fơ HAT ǎ e: HP = 20%, Ta _Ca = 30%, Node_Ca = 20%, SN_Ca = 30%, S ag_Ca = 20% o t e bed fơ *Sort*. The e ǎ a e ě be e ec ed fơ a e ce ǎ o .

k

Ma Red ce ď ea g o ǎ ǎ ge da a e oce g. Thě e ha e bee a o t of e eǎ ch ơ o t ada t o a d o e t e t [9, 20, 22, 27].

Ma Red ce ched g ha bee e e ded o a g ea a of a f , ch a ha ed- e c e a f , Ce oadba d e g e a f , GPU, FPGA, a d ob e a f . Phoe [18] a Ma Red ce a e o ha ed- e -c e a ch ec e. Ba ed o Phoe , [26] a d [14] o ed he f a ce of Ma Red ce o a c e a f . De K f a d Sa a a ga [6] a d Ra e a . [17] e e ed Ma Red ce a e f Ce oadba d e g e a d Ce -ba ed c e . Ma [6] ha e e he GPU co a o o e a d h gh e - ba d d h o acce e a e Ma Red ce a e , ch a Hadoo . I h ca e, Ma Red ce a ca o a e e ec ed o bo h CPU a d GPU . Sha e a . [21] o o ed FPMR f de e o e o c ea e Ma Red ce og a o FPGA. E e e a . [9] o o ed a Ma Red ce a e o he e oge eo ob e a f .

A o of ef c e Ma Red ce ched g a go h ha e bee o o ed o - o e he f a ce of Ma Red ce a ce a o . F che e a . [11] o- o e a dea ed a he a c ode o e a a e he co of a a g e a d de e o a o -ba ed a go h o o a a g a . I [16], a a c e a a e Ma Red ce ched e o o ed. The ched e o he a a d e a- a e he be e of g each a o d ffe e ode ea e. Ba ed o he e a a o , he ched e ca dec de he be d b o of a o ode acco d- g . I [5], T ed-Ma Red ce ched g a go h o o ed. T ed-Ma Red ce ched e a o a a ge Ma Red ce a o a be of a b a a d e a e oce e o e b a a a e h ef c e e of e o ce . I [27], a fa - ha g a go h f a e Ma Red ce e o o ed o a a ge e e o ce (a ed ce a o) f a e fa . I [1], a Ma Red ce ched g a go h o o ed o e he e ec o e a d o e he e e o ce a o . The a go h de e a ach e (VM) a d a o- ca e he VM o ob , a d o h ca ode . I [19], a D a c P (DP) a a e a ched e de g ed, h ch a o e o co o he a oca ed ca ac b d a ca ad g he dge . No e ha o h -ba ed a o- g a- eg co d be eg a ed o he e Ma Red ce ched g a go h o o e he f a ce of Ma Red ce a ca o f he .

De ec g agg e a he e ec o of c e ob a o he e e g - e. The e a e o o c e o de ec agg e a : he ea oge o c a d he o ge e a g e o c . F e a e, Hadoo [13] e he ea oge o c h e LATE [28] e he o ge e a g e o c o de ec agg e a a e o ed Sec . 2. Bo h o c e eed o e a e he oge of e a ed ce a acc a e . Pa aT e [15] a e- e ed oge d ca f a a e e e ha e e b e of Ma Red ce ob . Ho e e , he d ca ca o e a e he oge of SQL e e . O he o he ha d, o HAT ched e , h ch e a h -ba ed a o- g a eg , ca e a e he oge of a he a acc a e .

T ad o a Ma Red ce ched e ca o de ec o agg e a acc a e be- ca e he oge c e of a a e ca c a ed ba ed o acc a e e gh of each ha e t he o ėa ˋog̈e of a t a . To addȧe t he ˋob e , e ha e de g ed a d e e t ed HAT: a H t ȯ -ba ed A t o-T g Ma Red ce ˋched ė. HAT e t - a t e ˋog̈e of a t a acc ˋa e ce e t he e gh t of each ha e of a a t a a daˋed ce t a a t o a t ca accȯd g t o t he h ȯ ca a e of t he e gh t . HAT f ˋ t heȧ c a e o ode o a o ode a dˋed ce o ode . I t h a , HAT ca a ch bac a t fȯ ˋed ce ˋaggė t a o a o ode a d ˋce ė a. E ė e a ˋe de o ˋa e t ha t HAT ca ach e e t o 37% ėfȯ - a ce ga co ȧed t h Hadoo , a d t o 16% ėfȯ a ce ga co ȧed h LATE ched ė.

O e of o ˋ f t ˋe ȯ t o addȧe t he da t a oca t ˋob e t he a ch g bac t a , .e., a ch g bac a o ode h t he coˋe o d g da t a e t of t he ˋaggė t a . A o t heȧ o b e f t ˋe dˋec t o t o ˋo e a a - ca e of t he Ma Red ce a ca o t ˋ befȯe t he ˋea e ec t o . I t h a , t he ˋecȯded a - e of $M1$, $M2$, $R1$, $R2$, a d $R3$ ȧe ˋa ed a ˋo ˋ a e befȯe t he ˋea e ec o . We e a a t e o ˋ HAT ched ė t o ȯe a fȯ t , ch a t he Ce ḃ oadba d e g e a fȯ t a d t he ˋe t ed C o d Co t g a fȯ , t o e t t he ca ab t of HAT.

1. Abo aga A, Wa g Z, Zha g ZY (2009) Pac g t he o o o oˋ c o d. I : Ṗoceed g of t he ˋ ė a o a ȯ ho o C o d da t a a age e t . ACM, Ne Yȯ , 25 28
2. Baȧo o LA, Dea J, Ho e U (2003) Web eȧch fȯ a a e t : t he Goog e c t ė ȧch t ec t ˋe. IEEE MICRO 23(2):22 28
3. B a R, Yeo CS, Ve go a S, Ḃobėg J, Ḃa d c I (2009) C o d co g a d e ėg g IT a fȯ : o , h e, a d ˋea fȯ de ė g co g a t he 5 t h . F t ˋe Ge ė Co S t 25(6):599 616
4. Cha g F, Dea J, Ghe a a S, H eh WC, Wa ach DA, B ˋo M, Cha ḋa T, F e A, Ġ bė RE (2006) B g ab e: a d ˋ b ed ȯage e fȯ ˋ c ˋed da t a. I : Ṗoceed g of t he 7 t h USENIX o o o ėa g e de g a d e e t a o (OSDI 2006)
5. Che R, Che H, Za g B (2010) T ed-Ma Red ce: o gˋe o ˋce age of da t a- ȧa e a - ca o o coȧe h g. I : Ṗoceed g of t he 19 t h ė a o a co fėe ce o ȧa e ȧch t ec t ˋe a d co a o t ech e . ACM, Ne Yȯ , 523 534
6. De Ḱ f M, Sa ȧa ga K (2010) Ma Red ce fȯ t he ce ḃ oadba d e g e ȧch t ec t ˋe. IBM J Re De 53(5):10
7. Dea J, Ghe a a S (2010) Ma Red ce: a e b e da t a ˋoce g t oo . Co ACM 53(1):72 77
8. Dea J, Ghe a a S (2004) Ma ˋed ce: ed da t a ˋoce g o ȧge c ė . I : OSDI 2004: ˋoceed g of 6 t h o o o ėa g e de g a d e e t o . ACM Ṗe , Ne Yȯ , 137 150
9. E e ˋ P, Sha a S, M ḣa S (2009) Ma ˋed ce t e o ė he ėoge eo ob e de ce . I : Sof t ȧe t ech o og e fȯ e bedded a d b o t e , 168 179
10. Fa g W, He B, L o Q, Go dȧa NK (2010) Maȧ : acce ė a t g Ma Red ce h ġa h c ˋoce - ȯ . IEEE Ṫa Pȧa e D ˋ b S
11. F cheȧ MJ, S X, Y Y (2010) A g g a fȯ ef c e c Hadoo . I : Ṗoceed g of t he 22 d ACM o o ȧa e a gȯ t h a d ȧch t ec t ˋe . ACM, Ne Yȯ , 30 39
12. Ghe a a S, Gob off H, Le g S-T (2003) The Goog e t e e . I : SOSP 2003: ˋoceed g of t he 9 t h ACM o o o ėa g e ˋ c e . ACM, Ne Yȯ , 29 43
13. Hadoo (2011) Hadoo ho e age. h tt ://hadoo .a ache.ȯg/
14. J a g W, Ra VT, Aġa a G (2010) A a ˋed ce e h a a ė a e API fȯ -coȧe e - ˋo e . I : 2010 10 t h IEEE/ACM ė a o a co fėe ce o c t ė, c o d a d ġ d co g. IEEE Ṗe , Ne Yȯ , 84 93

15. Mo o K, Ba a a M, G o a D (2010) Pa aT e : a og e d ca o fo Ma Red ce DAG . I : P oceed g of he 2010 e a o a co fe e ce o a age e of da a. ACM, Ne Yo , 507 518
16. Po o J, Ca e a D, Bece a Y, To e J, A g ade E, S e de M, Wha e I (2010) Pe fo a ce a age e of acce e a ed Ma Red ce o oad he e oge eo c e . I : 39 h e a o a co fe e ce o a a e oce g (ICPP2010). Sa D ego, CA, USA
17. Ra e MM, Ro e B, B AR, N o o o o DS (2009) Ce MR: a f a e o fo o g a e-d ce o a e c ce -ba ed c e . I : IEEE e a o a o o a a e & d b ed oce g. IPDPS 2009. IEEE P e , Ne Yo , 1 12
18. Ra ge C, Ragh a a R, Pe e a A, B ad G, Ko a C (2007) E a a g a ed ce fo -co e a d oce o e . I : HPCA 2007: oceed g of he 2007 IEEE 13 h e a o a o h gh e fo a ce co e a ch ec e. IEEE Co e Soc e , Wa h g o , DC, 13 24
19. Sa dho T, La K (2010) D a c o o a ha e ched g hadoo . I : Job ched g a eg e fo a a e oce g. S ge , Be , 110 131
20. Scha MC (2009) C o dB : h gh e e ead a g h Ma Red ce. B o fo a c 25(11):1363
21. Sha Y, Wa g B, Ya J, Wa g Y, X N, Ya g H (2010) FPMR: Ma Red ce f a e o o FPGA. I : P oceed g of he 18 h a a ACM/SIGDA e a o a o o F e d og a ab e ga e a a . ACM, Ne Yo , 93 102
22. T a C, Zho H, He Y, Zha L (2009) A d a c Ma Red ce ched e fo he e oge eo o oad . I : P oceed g of he 2009 e gh h e a o a co fe e ce o g d a d coo e a e co g. IEEE Co e Soc e , Lo A a o , 218 224
23. Va e o LM, Rode o-Me o L, Cace e J, L d e M (2008) A b ea he c o d : o a d a c o d de o . Co Co Re 39(1):50 55
24. Va a J (2008) C o d a ch ec e . Wh e a e of A a o . e h a a. 3.a a o a .co / b c/ c o d a ch ec e - a a. df
25. Yahoo (2011) Yahoo! hadoo o a . h ://de e o e . ahoo.co /hadoo / o a /
26. Yoo RM, Ro a o A, Ko a C (2009) Phoe eb h: ca ab e Ma Red ce o a a ge- ca e ha ed- e o e . I : IEEE e a o a o o o oad cha ac e a o . IISWC 2009. IEEE P e , Ne Yo , 198 207
27. Zaha a M, Bo ha D, Sa a JS, E e eeg K, She e S, S o ca I (2009) Job ched g fo - e a ed ce c e . Tech ca e o , UCB/EECS-2009-55, U e of Ca fo a a Be e e
28. Zaha a M, Ko A, Jo e h AD, Ka R, S o ca I (2008) I o g a ed ce e fo a ce he e oge eo e o e . I : 8 h USENIX o o o e a g e de g a d e e a o . ACM, Ne Yo , 29 42