**Hadoop: MapReduce**

**Laporan Tugas 2 IF4031**

**Pengembangan Aplikasi Terdistribusi**

**Semester I Tahun 2014 / 2015**



oleh :

Iskandar Setiadi / 13511073

Sekolah Teknik Elektro dan Informatika (STEI ITB)

Institut Teknologi Bandung

Jl. Ganesha No. 10, Bandung 40132

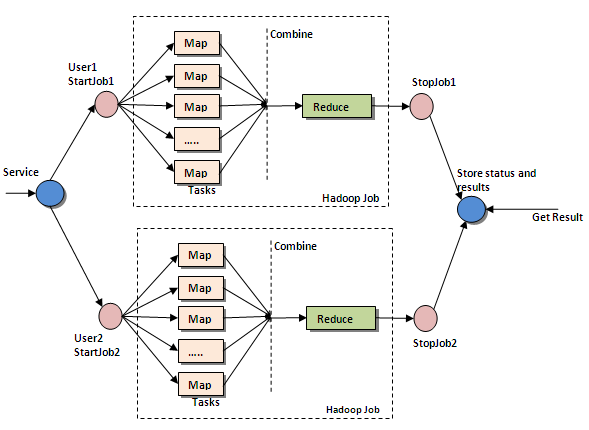
Tahun 2014

# **I. Deskripsi Singkat Algoritma**

Pada tahap awal, data yang tersimpan dalam *dblp.xml* dipotong-potong menjadi beberapa bagian untuk diproses di masing-masing *cluster*. Pembagian ini menggunakan *pre-processing* untuk menjamin keutuhan dari masing-masing *record* (agar tidak ada *record* yang terpotong sebagian).

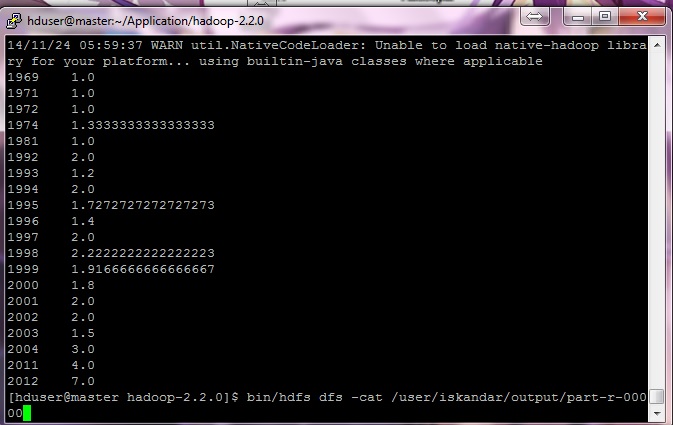
Data yang sudah dibagi menjadi beberapa bagian ini akan diproses dalam *mapper*. Algoritma yang dirancang menggunakan regex untuk melakukan pencocokan string. Hal ini didapatkan dari pengamatan sederhana bahwa setiap *record* akan dimulai dan diakhiri dengan kata kunci seperti *article, proceedings*, dsb. Algoritma yang dirancang mencari jumlah kemunculan *<author>* dalam satu *record*, dan menyimpannya kedalam pasangan <*key*, *value*> : <*year*, *jumlah\_author*> untuk diproses di tahap selanjutnya. *Key* (*year*) dicari menggunakan pola regex *<year>(\\w+)<year>*.

Pada tahap selanjutnya, data dari beberapa *mapper* digabungkan menggunakan *combiner* dan *reducer*. Setiap *key* yang sama akan dijumlahkan untuk dilakukan perhitungan *jumlah author* / *jumlah\_entity* dalam satu *key* tahun yang sama.

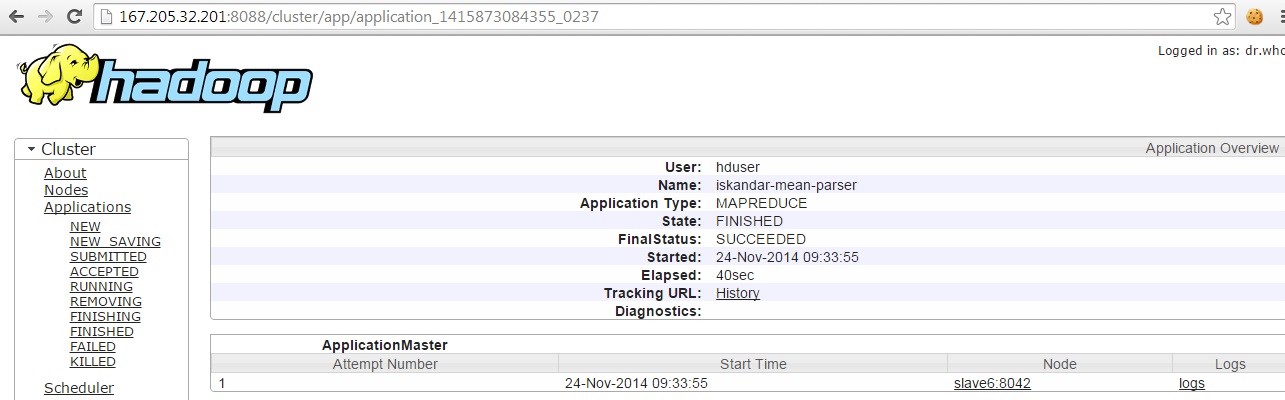


Gambar 1. 1 Proses Map-Reduce dalam Hadoop

# **II. Screenshot dan Hasil Eksekusi**



Gambar 2. 1 Hasil dari dblp-short.xml



Gambar 2. 2 Hasil dari *running* dblp.xml pada 167.205.32.201:8088

Tabel 2.1 Keluaran dari *running* dblp.xml

|  |
| --- |
| 0.4037677483369732  1936 1.0  1937 1.0  1938 1.1  1939 1.0  1940 1.2  1941 1.0  1942 1.0769230769230769  1943 1.0  1944 1.0  1945 1.125  1946 1.108695652173913  1947 1.2  1948 1.1333333333333333  1949 1.0217391304347827  1950 1.0476190476190477  1951 1.1176470588235294  1952 1.2  1953 1.0  1954 1.175  1955 1.2130387728213816  1956 1.2924419529732623  1957 1.1211192810457518  1958 1.1984792019574626  1959 1.4588027016120606  1960 1.2232861453916115  1961 1.243518995185662  1962 1.2272995534151718  1963 1.2514203215405928  1964 1.3017250309115511  1965 1.3275510933130457  1966 1.2237620839196512  1967 1.4267199274003504  1968 1.4093908607722556  1969 1.3734411091166594  1970 1.3598978422917363  1971 1.411490057805358  1972 1.4216357412802778  1973 1.436558943740863  1974 1.5585098719972446  1975 1.4699054887635854  1976 1.5583391571273055  1977 1.579201390524727  1978 1.5736810609757774  1979 1.6229476742470805  1980 1.5749848638227644  1981 1.6794401106344445  1982 1.6765771723155463  1983 1.7256304025272584  1984 1.7397445895008186  1985 1.7632087197035826  1986 1.8113809450558673  1987 1.8090567283593697  1988 1.8658272063670154  1989 1.8783648150211991  1990 1.9162155503467737  1991 1.975809529079837  1992 2.0058291300149675  1993 2.051608228690001  1994 2.107937912467992  1995 2.1646047822294374  1996 2.1724238476975932  1997 2.2312443885713695  1998 2.277433105383748  1999 2.3094024169193106  2000 2.355527423789749  2001 2.4117203452208362  2002 2.466234927435651  2003 2.5413088251010234  2004 2.5905312641390097  2005 2.6357331829037873  2006 2.7070136496833332  2007 2.77580350973301  2008 2.802001601123403  2009 2.834651748074778  2010 2.946405808624392  2011 2.987347708154229  2012 3.064171482109888  2013 3.1042282144345252  2014 1.7764079810005662 |