

## BAB 6

### KESIMPULAN DAN SARAN

#### 6.1 Kesimpulan

Setelah melakukan proses analisis, perancangan, implementasi, dan pengujian pada penelitian ini maka dapat diusulkan beberapa kesimpulan, yaitu:

1. *Incremental web crawling* merupakan salah satu jenis proses *crawling* yang melakukan *refresh* pada koleksi lokal dengan cara mengunjungi *webpage* dengan frekuensi waktu tertentu berdasarkan pengukuran seberapa sering sebuah *page* berubah (kontennya). *Incremental web crawling* yang diimplementasikan pada perangkat lunak menggunakan rancangan A.K. Sharma dan Ashutosh Dixit yang menerapkan algoritma *Self Adjusted Refresh Time Calculator Module* (SARTCM). Jenis *crawling* ini dapat dimanfaatkan untuk menyelesaikan permasalahan perubahan konten *web* seiring waktu.
2. NoSQL HBase merupakan salah satu basis data yang berada pada ekosistem Hadoop. HBase memiliki fitur *versioning* yang digunakan untuk menyimpan nilai *column qualifier* lebih dari 1 versi dikarenakan HBase tidak dapat melakukan pembaharuan informasi. Dalam penerapannya, perangkat lunak menerapkan fitur *versioning* yang salah satunya digunakan untuk menyimpan informasi URL setiap periode *crawling* yang dapat melebihi satu versi.
3. Terdapat dua buah cara dalam mengakses basis data HBase yaitu menggunakan HBase Shell melalui *command line interface* (CLI) atau menggunakan HBase API yang berjalan pada *platform* Java. Dalam penerapannya, perangkat lunak *incremental web crawling* dapat mengakses basis data HBase menggunakan HBase API.
4. *Incremental web crawling* telah berhasil diterapkan dan menghasilkan 2 perangkat lunak, yaitu situs induk J2EE dan agen *crawler* J2SE. Dari hasil pengujian yang sudah dilakukan seluruh fungsi pada perangkat lunak telah berjalan dengan baik.
5. Basis data HBase berjalan pada lingkungan Hadoop sehingga mampu memproses data dalam jumlah besar. Semakin besar lingkungan Hadoop, maka waktu transaksi yang dibutuhkan menjadi semakin kecil.

#### 6.2 Saran

Setelah melakukan proses analisis, perancangan, implementasi, dan pengujian pada penelitian ini maka dapat diusulkan beberapa saran, yaitu:

1. Penulis menyarankan penerapan perangkat lunak agen *crawler* agar dinyalakan lebih dari satu hari melebihi pengujian yang dilakukan penulis agar *refresh time* yang didapatkan lebih optimal.
2. Berdasarkan hasil penelitian dan pengujian yang telah dilakukan, penulis menyarankan agar ketika melakukan pengembangan aplikasi dengan penerapan basis data HBase ataupun dengan

Hadoop terdistribusi, dapat menggunakan spesifikasi komputer yang tinggi agar berjalan dengan baik dan lancar.

3. Admin perangkat lunak dapat melakukan pembaharuan informasi umum Admin, yang terdiri dari nama, *username*, dan *password*.
4. Informasi yang dapat dicari dari URL hasil proses *crawling* tidak hanya berupa tekstual, namun dapat berupa format data lain, contohnya adalah gambar.

## DAFTAR REFERENSI

- [1] Udupure, T. V., Kale, R. D., dan Dharmik, R. C. (2014) Study of web crawler and its different types. *IOSR Journal of Computer Engineering (IOSR-JCE)*, **16**, 1–5.
- [2] Sharma, A. dan Dixit, A. (2008) Self adjusting refresh time based architecture for incremental web crawler. *IJCSNS Internal Journal of Computer Science and Network Security*, **8**, 349–354.
- [3] Yakubovich, A., Lublinsky, B., dan Smith, K. B. (2013) *Professional Hadoop Solutions*. John Wiley and Sons, Inc, 10475 Crosspoint Boulevard Indianapolis, IN 46256.
- [4] George, L. (2010) *HBase - The Definitive Guide*, 2nd edition. O’Reilly Media, Inc, Inc, Gravenstein Highway North, Sebastopol, CA: 95472.
- [5] Godbole, A. dan Kahate, A. (2013) *Web Technologies: TCP/IP, Web/Java Programming, and Cloud Computing*, 3rd edition. Tata McGraw-Hill Education Pvt. Ltd, New York, NY Columbus.
- [6] Olson, C. dan Najork, M. (2010) Web crawling. *Foundation and Trends in Information Retrieval*, **4**, 176–180.
- [7] Cho, J. dan Garcia-Molina, H. (2000) The evolution of the web and implications for an incremental crawler. Technical Report Submitted to VLDB 2000. Stanford University, CA 94305.
- [8] Cho, J. dan Garcia-Molina, H. (1999) Estimating frequency of change. Technical report. Stanford University, CA 94305.
- [9] White, T. (2012) *Hadoop: the definitive guide*, 3rd edition. O’Reilly Media, Inc, Gravenstein Highway North, Sebastopol, CA: 95472.
- [10] Nathanael, H. D. (2017) Pengembangan aplikasi rukun tetangga (rt) online berbasis mobile cloud menggunakan platform nosql hbase. Skripsi. Universitas Katolik Parahyangan, Indonesia.