

*Abstrak*—Data warehouse adalah solusi yang sangat terkenal untuk menganalisis data bisnis dari sumber yang heterogen. Namun, data yang dapat dianalisis oleh data warehouse adalah data terstruktur, sedangkan saat ini, dengan semakin populernya media sosial, dan semakin mudah diaksesnya aplikasi web untuk membuat data oleh banyak orang dan organisasi, kita mengalami banjir data tidak terstruktur. Oleh karena itu diperlukan suatu pendekatan yang dapat “menstrukturkan” data yang tidak terstruktur menjadi data terstruktur sehingga dapat diolah oleh data warehouse dan menghasilkan fakta dan wawasan. Untuk melakukan ini, kami mengusulkan pendekatan ekstraksi yang akan menghasilkan skema dari data tidak terstruktur. Penelitian yang ada menggunakan Hadoop untuk mengekstrak skema, dan kami melihat beberapa keterbatasan dari pendekatan ini. Oleh karena itu kami mengusulkan menggunakan Google Cloud Platform untuk melakukan ekstraksi skema. Berdasarkan hasil eksperimen, pendekatan kami berhasil menghasilkan skema untuk data tidak terstruktur yang dapat diproses lebih lanjut oleh data warehouse. Kami juga membuktikan bahwa pendekatan kami membantu pengembang database untuk memahami data tidak terstruktur dengan lebih baik.

*Kata kunci: ekstraksi skema, data tidak terstruktur, data warehouse*