Analisis sumber data memainkan peran kunci dalam proyek data warehouse karena semua proses ekstraksi dan transformasi selanjutnya bergantung pada apa data yang benar-benar dimiliki sumber data. Sebelum Anda mulai mengekstrak data, luangkan waktu untuk melihat definisi struktural dari berbagai sumber data warehouse Anda:

* Tabel database dan definisi kolom, batasan, dan data lainnya Pernyataan Bahasa Definisi (DDL)
* Struktur file data sumber (file VSAM atau ISAM pada mainframe IBM, atau Tabel Relasional pada Database DB2 atau Oracle, misalnya), seperti yang dijelaskan dalam daftar program sumber
* Definisi "buku salinan" COBOL (jika digunakan - dan ya, untuk Anda semua yang muda di luar sana, COBOL masih eksis di dunia modern)
* Definisi yang mungkin disimpan dalam kamus data terpusat atau repository

Berikut langkah-langkah mengidentifikasi data:

1. Pada tahapan pertama, identifikasi sejumlah fakta yang berkaitan kemungkinan fungsionalitas data warehousing
2. Berdasarkan kumpulan fakta yang terkumpul (dan data yang diwakilinya), identifikasi sumber data yang diperlukan untuk menyediakan elemen untuk kemudian dibangun fakta-fakta di dalam data warehouse
3. Memulai tahapan desain data wharehouse