

Hadoop (4)

Flume dan Sqoop



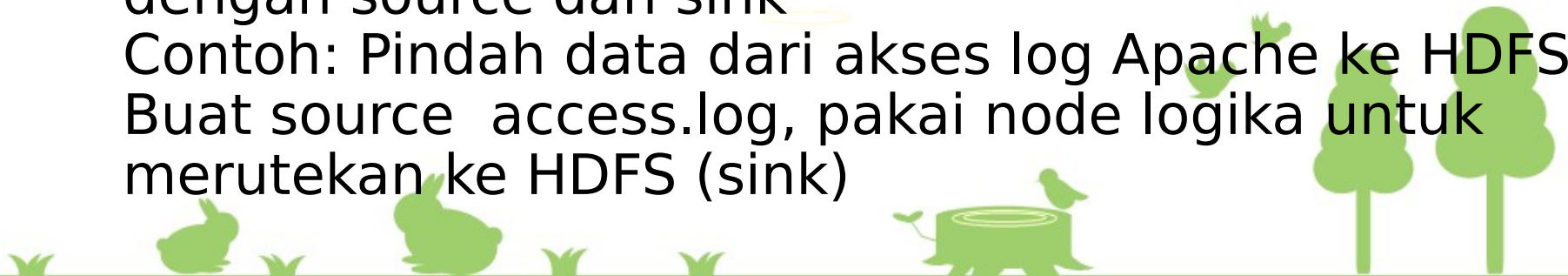
Garis Besar

- Memindahkan data ke Hadoop
- Mempelajari Flume
- Mempelajari Sqoop



Flume

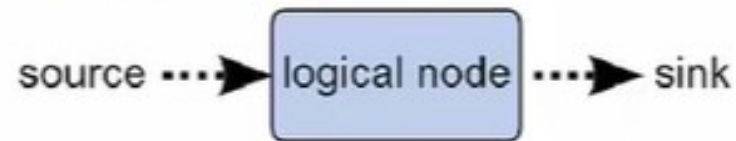
- Open source software – Cloudera
- Layanan memindahkan data jumlah besar di sekitar cluster setelah data dihasilkan
- Use case:
Mengumpulkan log files dari semua node di cluster
Mentransfer data ke tempat terpusat, HDFS
- Buat data flow rangkaian node logika terhubung dengan source dan sink
Contoh: Pindah data dari akses log Apache ke HDFS
Buat source access.log, pakai node logika untuk merutekan ke HDFS (sink)



Stream-oriented data flow

- Stream-oriented data flow

- Chains of logical nodes

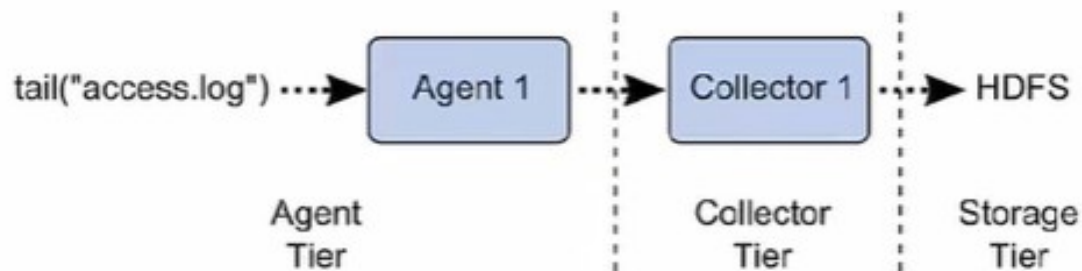


- Example:

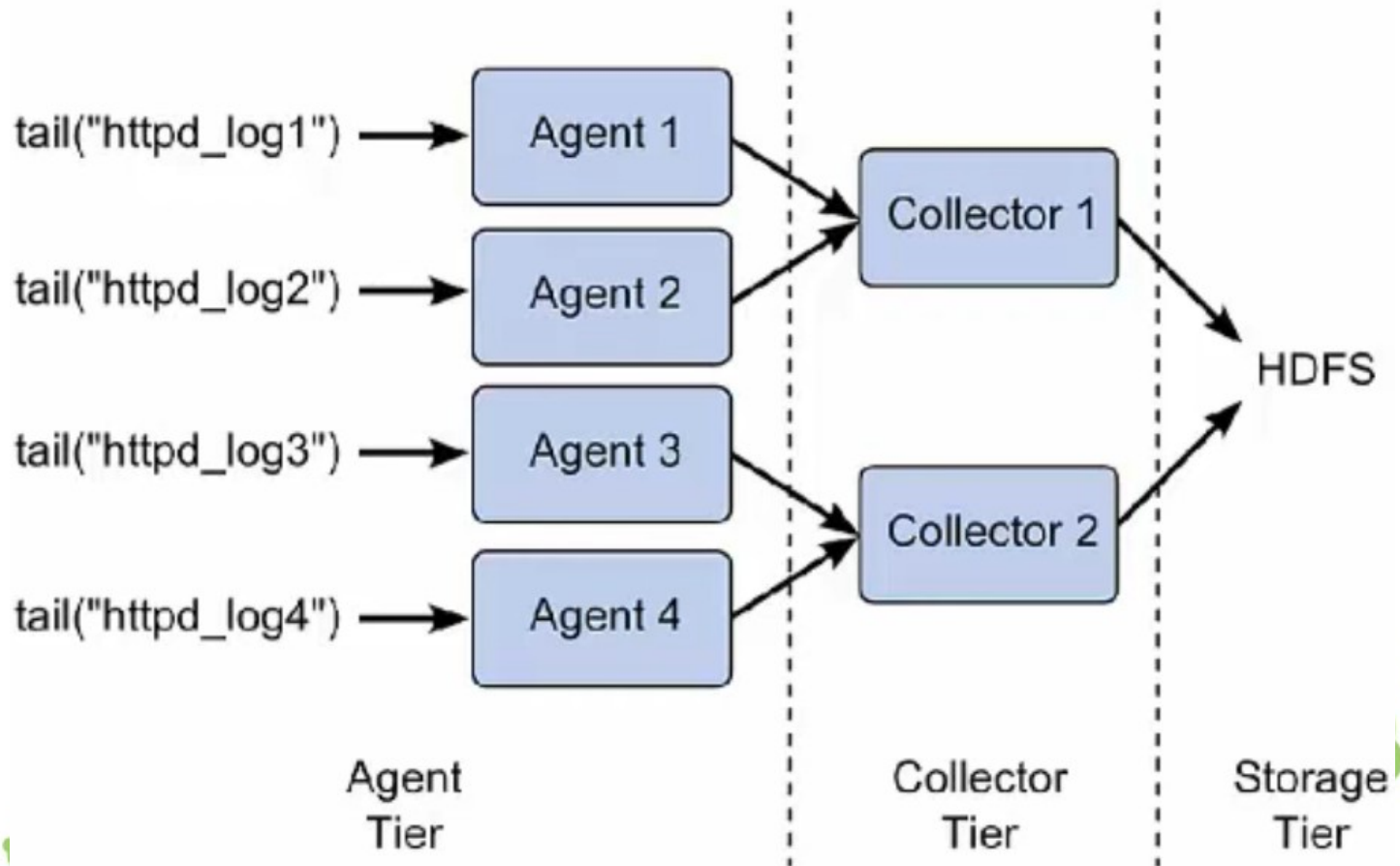


Arsitektur Flume

- Tiers
 - Agent – sumber (source) data
 - Collector – multiple collector dari multiple agent (multiple collector, meningkatkan paralelisasi)
 - Storage – sink data. seperti HDFS



Use Case



Sqoop

- Open source project transfer data antara Hadoop relational database
- JDBC akses sistem relational
- Akses database memahami skema data
- MapReduce impor dan ekspor data
- Import data ke Hadoop generate Java Class enkapsulasi 1 baris tabel yang diimpor untuk develop cepat aplikasi MapReduce lain yang menggunakan record yang Sqoop simpan di HDFS

