

Pendahuluan

- Pig dan Hive memiliki banyak persamaan
 - Translasi bahasa program tingkat tinggi (high-level) ke MapReduce job
Programmer bisa menggunakan selain Java atau bahasa tingkat rendah (low-level) dengan Hadoop streaming
 - Menawarkan ukuran jauh lebih kecil daripada kode Java
 - Menyediakan opsi ekstensi untuk mengisi kekurangan fungsionalitas
Contoh: fungsi tidak bisa dilakukan high-level, Pig dan Hive bisa tambahkan dengan user-defined function di Java
 - Interoperabilitas antar bahasa pemrograman

Pendahuluan (Lanjutan)

- Tidak mendukung random read/write atau low-latency query berjalan di atas Hadoop

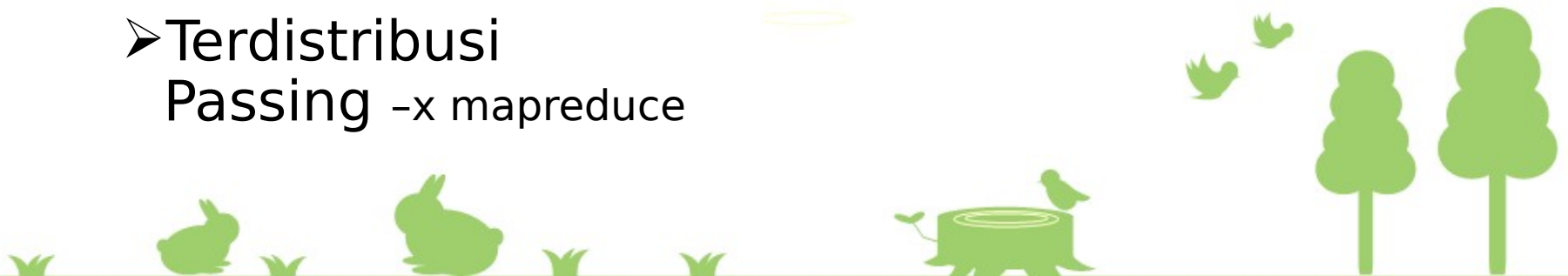


Pig

- Dikembangkan oleh Yahoo @2006
- Dihibahkan ke Apache @2007
- Bahasa data flow – PigLatin
Connecting things together
- Beroperasi di struktur data kompleks dan bertingkat (nested)
- Tidak seperti SQL, Pig tidak perlu skema (Schema optional)
Cocok untuk unstructured data
- Seperti SQL, relationally complete relational algebra
- Turing complete ketika di-extend dengan User-Defined Function (UDF)
*Turing completeness butuh looping construct, infinite memory model, dan conditional construct

Detail Menjalankan Pig

- Script `pig scriptfile.pig`
- Grunt `pig` (run CLI tool)
- Embedded Panggil ke Pig dari Java
- Execution environment
 - Local testing <-tidak punya Hadoop environment lengkap
Passing `-x local` option
 - Terdistribusi
Passing `-x mapreduce`



Hive

- Dikembangkan Facebook
- Mengubah Hadoop data warehouse dilengkapi declarative language (SQL dialect) untuk query HiveQL
- Tidak spesifik data flow tidak seperti PigLatin deskripsi hasil yang diinginkan Hive akan membuat data flow untuk mencapai
- Tidak seperti Pig Skema non-opsional HARUS Tidak terbatas 1 skema beberapa bisa
- Seperti SQL & PigLatin relationally complete
- Tidak Turing complete, sama seperti Pig

Detail Menjalankan Hive

- **Hive Shell**

- Interactive `hive`
- Script `hive -f myscriptpath`
- Inline `hive -e 'SELECT * FROM mytable'`

