



Formação Engenheiro de Dados

Pig

Pig



Mais uma solução a complexidade do Hadoop

Permite usar MapReduce sem precisar conhecer Java



Abstração do MapReduce



Linguagem de alto nível: Pig Latin para processamento de dados

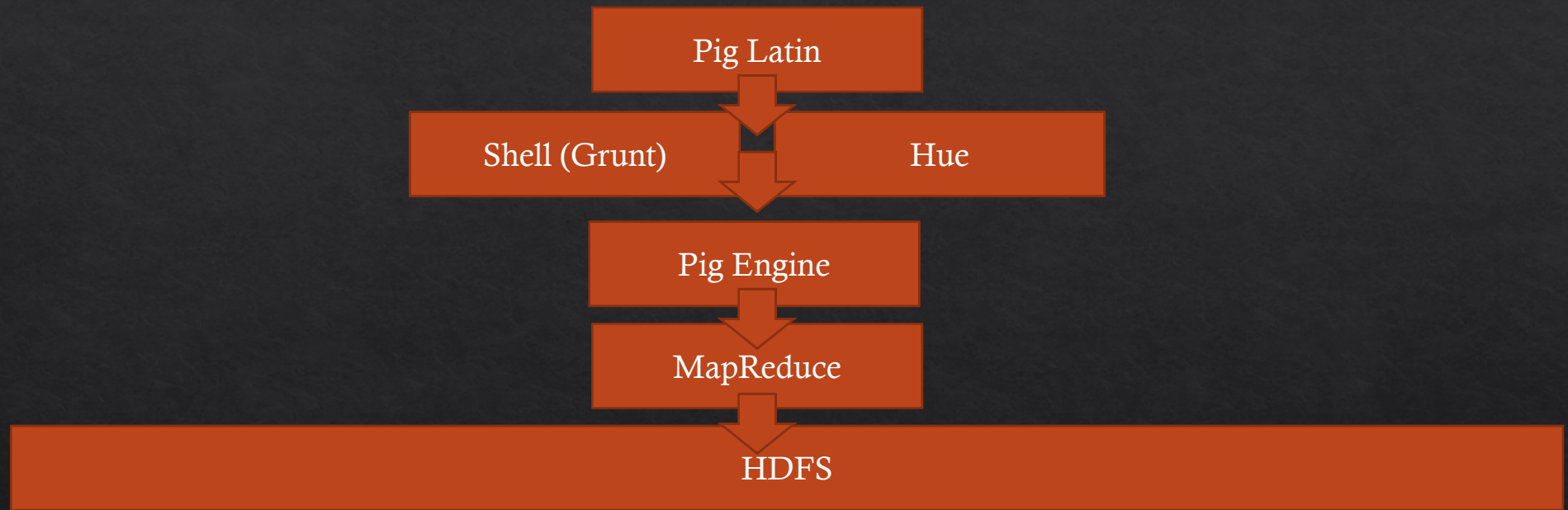


Transformações Lazy



Operações em Memória

Pig



Modelo de Dados



TUPLE: EQUIVALENTE
A UM REGISTRO



BAG: CONJUNTO DE
REGISTROS

Shell

- ◊ grunt
 - ◊ pig -x[f] local
 - ◊ **pig -x[f] mapreduce**
- ◊ Exemplos
 - `pig -x local`
 - `pig -f local '/user/script.pig'`

Formas de Execução

- ◇ Interativa (grunt)
- ◇ Batch: script com extensão pig
- ◇ Embedded: dentro de outras aplicações

Comandos

- ◇ sh : permite rodar comando do shell do Linux
- ◇ fs : permite invocar comandos do sistema de arquivos
- ◇ exec/run: executa script do shell
- ◇ history: histórico de comandos
- ◇ help
- ◇ quit

Tipo	Descrição
int	Inteiro de 32 bits
long	Inteiro de 64 bits. L
BigInteger	java big integer
BigDecimal	java decimal
float	Ponto flutuante de 32 bits. F
double	Ponto flutuante de 64 bits
boolean	Lógico
datetime	Data e Hora
chararray	String
bytearray	blob
Tuple	campos (fernando, 30)
Bag	Coleção de campos {(José,20),(Maria,12)}
Map	Conjunto de Chaves Valor [{"nome":"José","idade":30}]

Tipos