

# レコメンデーションエンジン データ仕様



## 購入ペアデータの仕様

- データは**HDFS**上の「/user/root/hadoop\_exercise/04/data/goods\_pair」に**テキストデータ**として配置されている
- 各行には、購入された商品のペアが**カンマ区切り**で記録されている

...

あずきキャラメル,チョコレートマシュマロ

あんドーナツ,あずきキャラメル

チョコレートマシュマロ,びわゼリー

あずきキャラメル,フィナンシェ

...



## 関連度分母データの仕様

- データはHDFS上の「/user/root/hadoop\_exercise/04/data/」に「denomination」という名前のテキストデータとして配置する
- 各行には次の順にデータがカンマ区切りで記録されている
  - ① 商品名
  - ② ①の商品が購入ペアデータのペア内に出現した回数

商品名

...  
あずきキャラメル  
びわゼリー,87  
カステラ,63  
...

「購入ペアデータ」  
内での出現回数



## 関連度分子データの仕様

- データは**HDFS**上の「/user/root/hadoop\_exercise/04/data/」に「**numerator**」という名前の**テキストデータ**として配置する
- 各行には次の順にデータが**カンマ区切り**で記録されている
  - ① カンマ区切りの商品名のペア（商品名は昇順）
  - ② ①のペアが購入データペア内に出現した回数（ペアの順序は問わない）

### 商品名のペア

あずきキャラメル,びわゼリー,23  
あずきキャラメル,カステラ,87  
フィナンシェ,マドレーヌ,53  
...

「購入ペアデータ」内  
でのペアの出現回数



## 関連商品データの仕様 (参考)

- データは**HDFS**上の「/user/root/hadoop\_exercise/04/data/」に「related\_goods」という名前の**テキストデータ**として配置する
- 各行には、次の順でデータがカンマ区切りに記録されている
  - ① 商品名(その1)
  - ② ①の商品と関連がある商品名
  - ③ ①の商品と②の商品の関連度 (0.025～1)

...

あずきキャラメル,チョコレートマシュマロ,0.032  
あずきキャラメル,草加せんべい,0.031  
あんドーナツ,あずきキャラメル,0.04  
チョコレートマシュマロ,びわゼリー,0.042  
チョコレートマシュマロ,きんつば,0.026

...