Dominion Datamining

Khaoula TAGNAOUTI, Liliana LOPEZ, Willian VER VALEM, Elmer BAYOL

February 19, 2016

Introduction

Le contexte du projet

- Un serveur ouvert d'Octobre 2010 à Mars 2013
- 12 Millions de logs à traiter (dont la taille générale est 13Go)
- Un Wiki offre des conseils au niveau de la stratégie

Dominion DataMining est un programme qui se compose de trois parties principales:

- Modélisation des donées
- Stockage des données
- Analyse des données

- Acteurs physiques:
  Utilisateur
- Acteurs de système:
  Parser
  Analyseur
  Base de données

Fonctionnalités

- Parser
- Stockage des données
- Calcul d'elo

Fonctionnalités

Parser

- 99.8% des logs sont parsés
- Seul le header est parsé
- Est capable de travailler avec les logs compressés

Fonctionnalités

Stockage des données

- Stockage des parties
- Stockage d'elo global pour chaque joueur
- Stockage de liste des parties d'un joueur

Exemple de structure stockée dans la base de données:


```
{ "_id" : ObjectId("56b370882db5d81f5287cebb"),
  "game-log" : {
      "date" : "Tue Feb 01 00:02:00 CET 2101",
      "winners" : [ [ "Noobular " ] ],
      "cardsgonne" : [ [ "Provinces" ] ],
      "market" : [ [ "Chapel", "Island", "Peddler", "Salvager", "Sea Hag", "Smugglers", "Steward", "Torturer", "Upgrade", "Wishing Well" ] ],
      "trash" : { "Coppers" : 2, "Estate" : 1, "Curses" : 2, "Sea" : 1, "Smugglers" : 1 },
      "players" : [ [
          {
              "name" : "Noobular",
              "Elo" : 0,
              "points" : 43,
              "turns" : 19,
              "victorycards" : { "Provinces" : 5, "Isls" : 5, "Estates" : 3 },
              "deck" : { "Coppers" : 5, "Provinces" : 5, "Silver" : 1, "Wishing" : 1, "Estates" : 3, "Islands" : 5, "Steward" : 1, "Torturer" : 1, "Golds" : 3 },
              "firsthand" : { "Coppers" : 3, "Estates" : 2 } },
          {
              "name" : "slemdog",
              "Elo" : 0,
              "points" : 26,
              "turns" : 19,
              "victorycards" : { "Provinces" : 3, "Isls" : 3, "Estates" : 2 },
              "deck" : { "Coppers" : 7, "Provinces" : 3, "Upgrades" : 3, "Peddler" : 1, "Wishing" : 1, "Estates" : 2, "Islands" : 3, "Silvers" : 2, "Salvager" : 1, "Torturer" : 1, "Golds" : 3 },
              "firsthand" : { "Coppers" : 3, "Estates" : 2 }
          }
      ] ]
  }
}
```

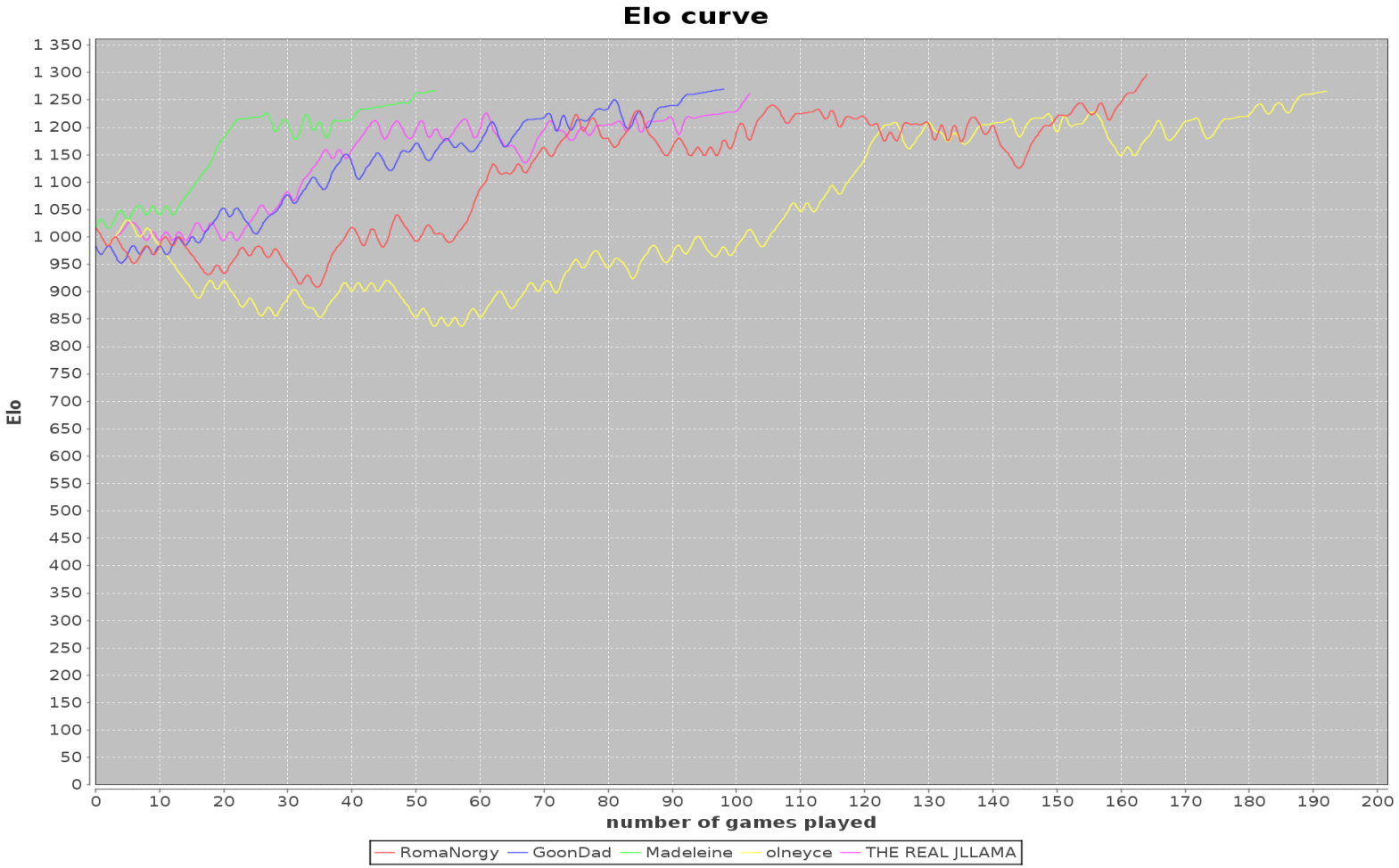
Fonctionnalités

Calcul d'elo

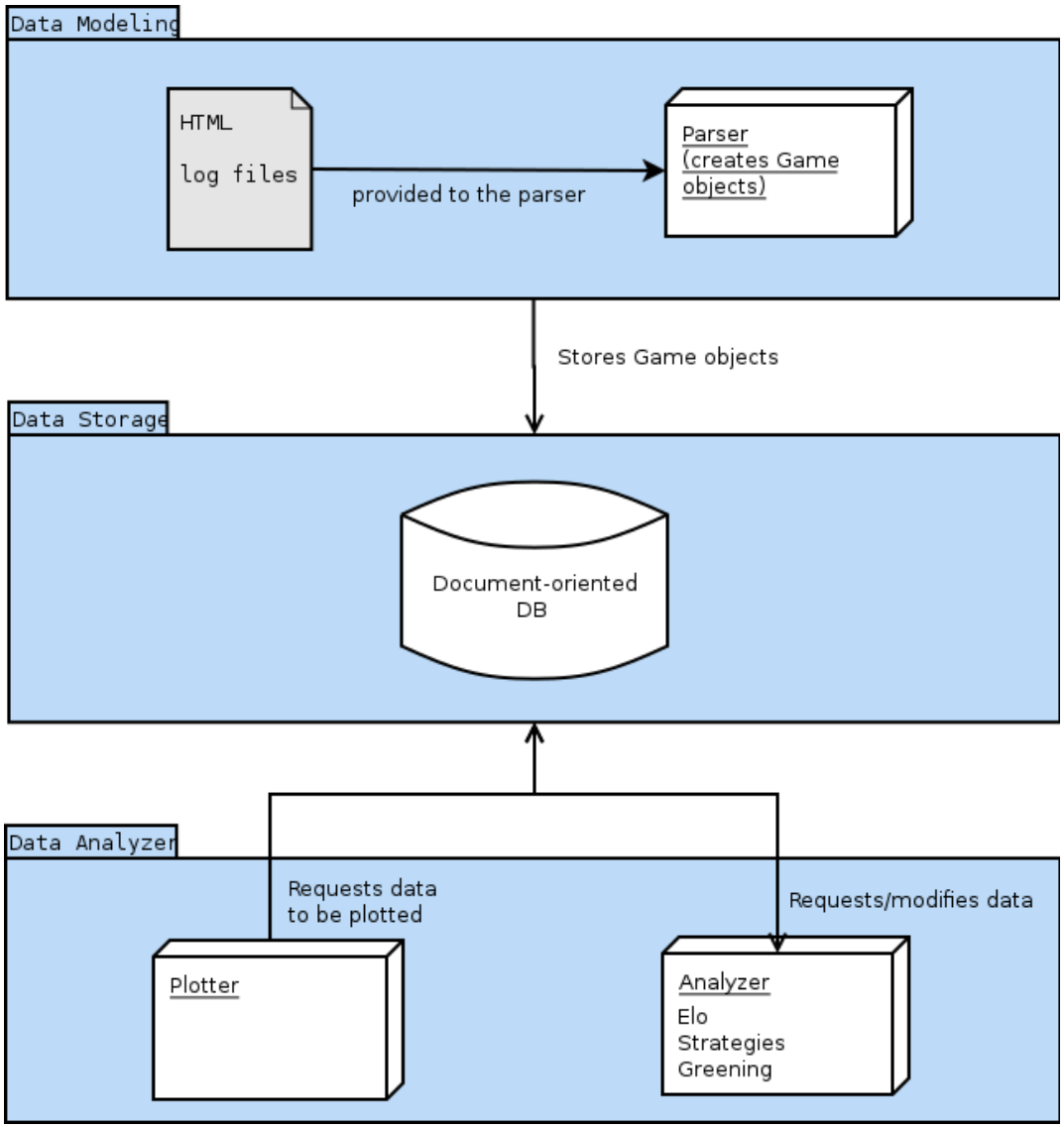
- Chaque partie est prise en ordre chronologique
- Prise en compte de l'elo global
- Mise a jour du nouvel elo global du joueur
- Ajout de l'elo du joueur au moment de la partie (ne varie pas)

Fonctionnalités

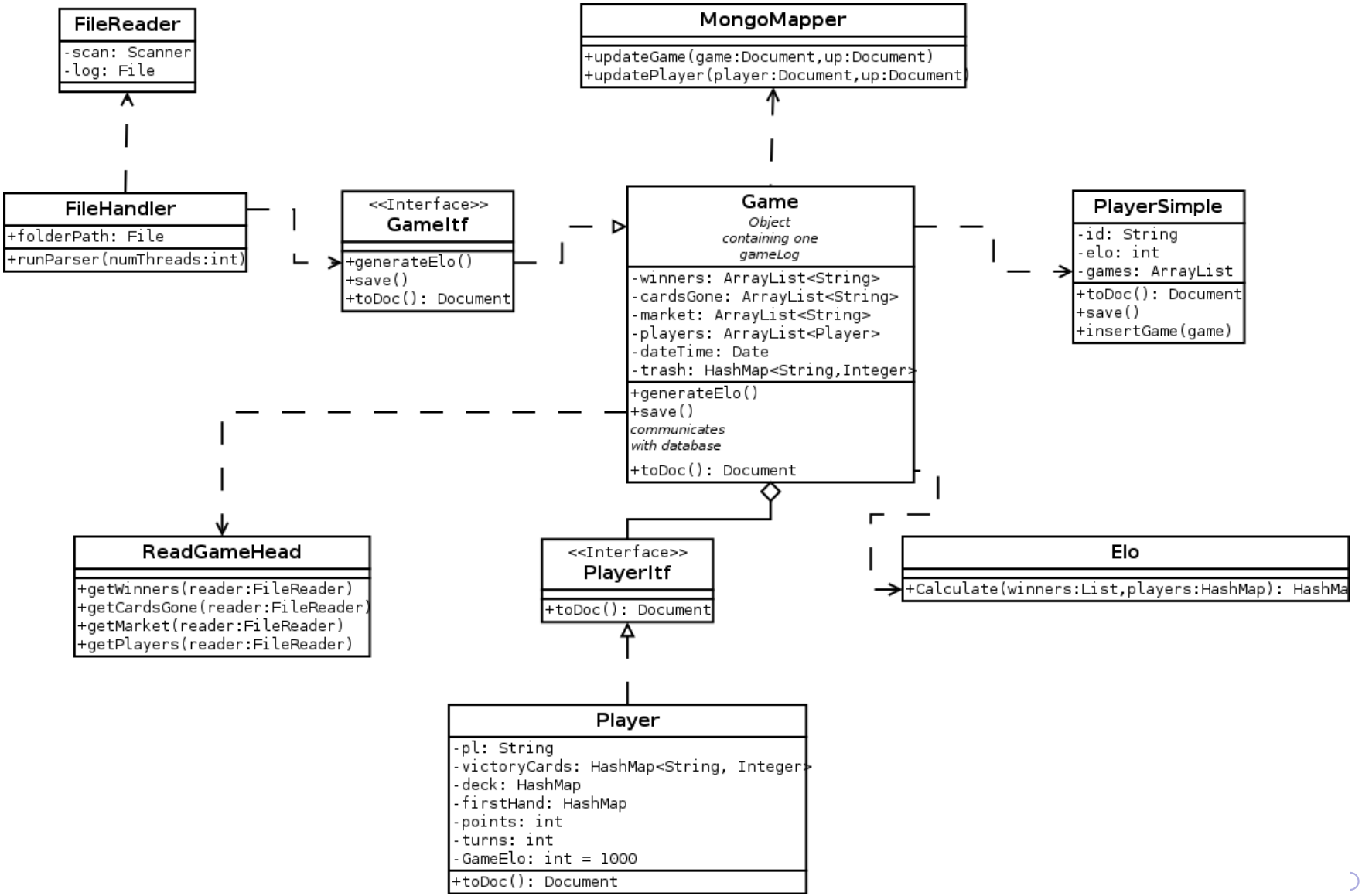
Calcul d'elo



Architecture



Architecture



Aspects techniques

- Choix techniques
- Algorithmes
- Bugs et problèmes

Choix techniques

Choix d'implémentation

Algorithmes

Parser

- un mélange jsoup + split
- et beaucoup de recherches regex

Elo

$$R_x = R_x + 32 * (v - \frac{10^{R_x/400}}{\sum_{n=0}^{np} Rn})$$

v = 1 si le joueur gagne et 0 si il perd

Bugs et Problèmes

Noms d'utilisateur

- cats and dogs living together
- gime all yo points
- hi, gl, hf, yada, yada, yada.
- (>^o ^)> <(*o*)> <(^o ^<)

Performance

- fichiers compressés
- grosse utilisation du disque

Questions