

Dominion Datamining

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

February 19, 2016

Introduction

Le but de projet

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

- ▶ Modélisation des données
- ▶ Stockage des données
- ▶ Analyse des données

Introduction

Le contexte de projet

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

Introduction

Classe de l'utilisateur et caractéristiques

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

- ▶ Parser
- ▶ Stockage des données
- ▶ calcul d'elo

- ▶ 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

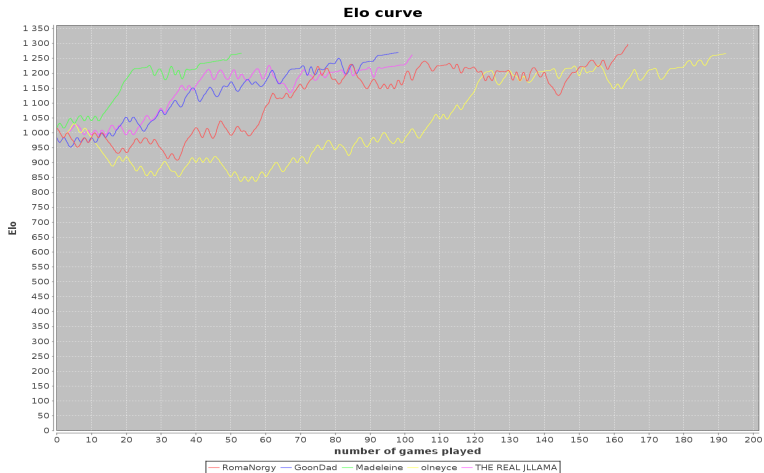
Fonctionnalités

Calcul d'elo

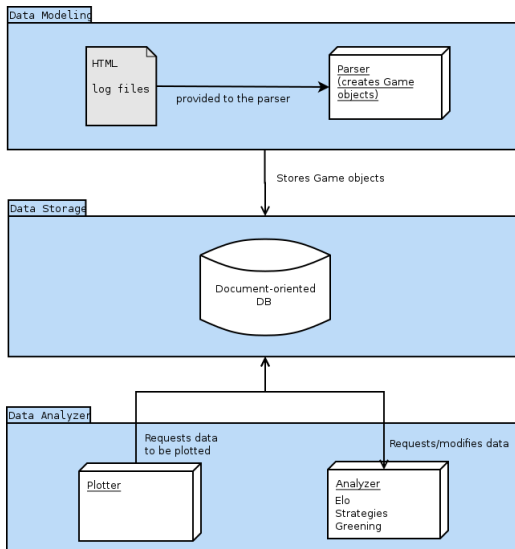
- ▶ Chaque partie est prise en ordre chronologique
- ▶ Prise en compte de l'elo global
- ▶ Mise a jour du nouvel elo global du joueur
- ▶ Ajout de l'elo du joueur au moment de la partie (ne varie pas)

Fonctionnalités

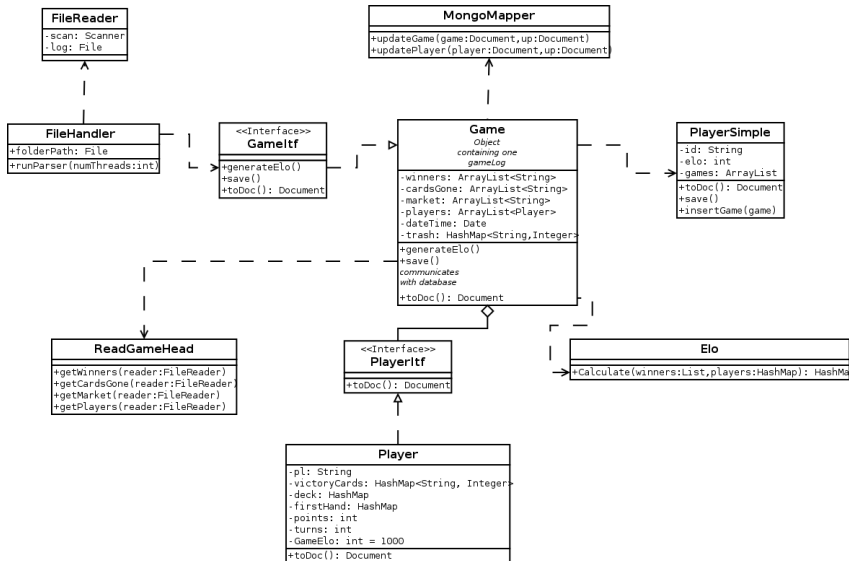
Calcul d'elo



Architecture



Architecture



Aspects techniques

- ▶ Technical choices
- ▶ algorithms
- ▶ bugs and problems

The Core



Parser

- ▶ a mix of jsoup + split
- ▶ and a lot of regex search

Elo

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

$v = 1$ if player win and 0 if loses

Bugs and Problems

usernames

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

Performance

- ▶ compressed files
- ▶ High disk usage