UNIVERSITÉ DE SHERBROOKE

MÉMOIRE PRÉSENTÉ AU

PROGRAMME DE MAÎTRISE EN ADMINISTRATION

Par

Philippe Lauzon, Candidat à la M. Sc. en Stratégie de l'intelligence d'affaires

Daniel Chamberland-Tremblay, Directeur de recherche

Modélisation multidimensionnelle vidéo dans le sport

Le cas Vert et Or Football

Le 30 juillet 2011

SOMMAIRE

La présente recherche a pour but d'étudier l'intégration de la vidéo dans l'entrepôt de données. Afin de nous accrocher à une problématique tangible, nous avons collaboré avec l'équipe de football du Vert et Or de l'Université de Sherbrooke. Selon Martin Montminy, entraineur de la ligne offensive, la vidéo est la donnée la plus importante dans le processus d'évaluation de la performance des joueurs. Dans le but d'améliorer ce processus, nous avons construit un modèle de données qui allie les données de jeu à de la vidéo.

Dans le but de mieux comprendre les différents enjeux liés à notre recherche, nous avons effectué une revue de littérature qui fait l'état des connaissances dans le domaine de la modélisation d'entrepôt de données et du multimédia.

Nous avons basé notre recherche dans le paradigme du « design science ». Afin de construire notre prototype, nous avons utilisé une méthodologie de développement d'entrepôt de données hybride basée sur l'approche guidée par les besoins et l'approche guidée les données.

Les résultats de l'étude montrent les différentes étapes réalisées afin d'en arriver à un modèle de données intégrant les données de jeux à la vidéo.

On y retrouve les choix de modélisation ainsi que les différents enjeux par rapport aux données.

Cette recherche contribue à l'avancement des connaissances dans le domaine de la modélisation multidimensionnelle multimédia en proposant l'utilisation de la vidéo pour appuyer les données factuelles alphanumériques.

Remerciements

La réalisation de ce mémoire de maîtrise n'aurait pas été possible sans l'aide, les conseils, la collaboration et le support de plusieurs personnes. Je tiens donc à remercier toutes les personnes qui ont été impliquées de près ou de loin dans cette recherche.

Je remercie tout particulièrement mon directeur de recherche Daniel Chamberland-Tremblay, qui m'a encadré tout au long de cette recherche. La réussite de ce projet est directement liée à l'encadrement, aux disponibilités, aux commentaires, aux idées, et au soutien de Daniel. Je remercie aussi mes lecteurs, la professeure Manon Guillemette et le professeur Olivier Caya pour le temps qu'ils ont accordé à la lecture de ce mémoire ainsi que pour les commentaires et critiques qu'ils ont su partager.

Je tiens à remercier l'équipe de football du Vert et Or et plus particulièrement M. Martin Montminy, entraineur de la ligne offensive lors du démarrage du projet, qui a su répondre à mes questions et fournir une problématique tangible au projet.

Je tiens aussi à remercier ma famille qui m'a supportée tout au long de mon cheminement scolaire. Mes parents qui ont, à leur façon, financé mon

projet de recherche et Audrey, qui a su me supporter dans les moments les plus difficiles.

Je remercie aussi mes partenaires de laboratoire pour leur soutien, commentaires et suggestions. Et plus spécialement, Geoffrey Rabault pour son soutien à « Team Fotress 2 », Alexandre Tardif pour ses conseils et son expérience, Sylvie Fréchette pour le partage de son expérience et Francis Dupont pour sa rigueur et son affiche inspirante.

Merci à vous tous!

Table des matières

Liste des Figures

Liste des Tables

Chapitre 1 : Introduction

1.1 L'intelligence d'affaires

L'intelligence d'affaires (IA) est définie comme un ensemble d'architectures, d'outils, d'applications et de méthodologies (Turban, 2008). Elle inclut la collecte, le stockage, l'analyse des données internes et externes dans le but de fournir des connaissances à la bonne personne, au bon moment et dans le bon format, afin de faciliter le travail du gestionnaire en améliorant la qualité et la rapidité du processus de prise de décision. (Negash 2004).

L'entrepôt de données est le point central de l'intelligence d'affaires. En effet, il permet d'intégrer les données en plus de les rendre accessibles à tous et facilement exploitables. De façon plus théorique, l'entrepôt de données est un ensemble de données qui contient de l'information d'un grand intérêt pour le management d'une organisation (Kemper 2000). De façon plus précise, Inmon (2005) le définit comme étant une collection de données orientées sujets, intégrées, non volatiles et historisées afin de soutenir les différents processus de prise de décision. Cette structure de données est à la source des différentes techniques d'analyse, par exemple du forage de données, rapports ad hoc et tableaux de bord (Turban 2008). En effet, l'entrepôt ou le comptoir de données, une version réduite de l'entrepôt traitant d'un sujet spécifique (Turban 2008),

est construit dans le but de fournir une source de données intégrée qui permettra à l'analyste d'avoir une vue valide et unique de l'organisation (Inmon 2008, Turban 2008).

1.2 Contexte sportif de la recherche

Dans la majorité des sports nord-américains professionnels, il devient de plus en plus difficile pour les dirigeants de construire une équipe qui aura du succès pendant plusieurs années consécutives (Lewis, 2003, Klein et Reif, 2001). En effet, avec l'instauration des plafonds salariaux et l'augmentation constante des salaires des joueurs, il reste de moins en moins de marge de manœuvre aux directeurs généraux pour construire une équipe gagnante que les amateurs veulent voir jouer.

L'exemple des Ducks d'Anaheim illustre bien ce problème. En effet, l'équipe a gagné la coupe Stanley en 2007 et depuis, elle ne cesse de baisser au classement général de la ligue nationale de hockey (LNH) chaque année. Pour la saison 2009-2010, l'équipe n'a pas fait partie des séries éliminatoires. Une des causes principales de cette chute au classement est que l'équipe a dû se départir de joueurs d'impact depuis la conquête de la coupe. En effet, Chris Pronger, Andy McDonald, Jean-Sébastien Giguère, Ilya Bryzgalov (finaliste au trophée Vézina 2009-2010), pour ne nommer que ceux-là, ont été échangés afin de respecter le plafond salarial et le budget de l'équipe.

1.3 Le sport et le monde des affaires

La structure d'une équipe sportive se compare à celle d'une entreprise : on y retrouve une hiérarchie composée d'un propriétaire, des gestionnaires et des employés. Au niveau stratégique se trouve l'équipe du directeur général, au niveau tactique, l'équipe d'entraineurs et au niveau opérationnel, les joueurs. Tout comme en entreprise, les équipes sportives tentent d'engager les meilleures personnes lorsqu'elles veulent combler un poste. Par contre, elles utilisent le repêchage, les échanges et le marché des joueurs autonomes afin de répondre à leurs besoins. Depuis quelques années, les équipes tentent de conserver les joueurs clés longtemps afin de construire le cœur de l'équipe et y parviennent en établissant avec ces joueurs des contrats à long terme. De plus, on utilise le ratio de roulement du personnel pour évaluer la santé de l'organisation, ce ratio pouvant aussi être appliqué au monde sportif. En effet, lorsqu'une équipe performe bien, on essaie le moins possible de faire des modifications à l'alignement. Finalement, la performance d'une équipe sportive se mesure entre autres par sa capacité à remplir les gradins.

1.4 L'intelligence d'affaires dans le sport

Dans le sport, comme dans toute entreprise, les gestionnaires doivent prendre des décisions importantes telles que l'échange et le repêchage de joueurs. Celles-ci peuvent influencer le quotidien et l'avenir de leur organisation. Environ 40% des décisions majeures prises par des cadres reposent sur

l'intuition de ceux-ci, plutôt que d'être basées sur des faits (Davenport, Harris et Morison, 2010). Bien que des décisions non factuelles puissent donner de grands succès, celles-ci ont tendance à être beaucoup plus risquées que celles basées sur des faits (Davenport et al., 2010).

Depuis quelques années, plusieurs équipes sportives optent pour une gestion de leur organisation à l'aide de différentes approches d'intelligence d'affaires. En effet, l'équipe de baseball majeur les A's d'Oakland est un exemple très connu. Ils ont allié les données de jeux aux « Sabermetric », métriques qui mesurent de façon quantitative et objective la performance des joueurs de baseball et qui ont été popularisées par Bill James[1]. Ceci leur a permis de se faire une place en série de fins de saison plusieurs années de suite, tout en ayant l'une des plus petites masses salariales de la ligue. L'exploitation des données par analyse statistique s'est avérée un outil indispensable pour l'évaluation des joueurs lors des repêchages, des échanges ainsi que pour développer des tactiques lors de parties. Depuis le milieu des années 1990, la ligue de basketball professionnelle NBA (National Basketball Association) utilise aussi des technologies d'IA. Plusieurs équipes, comme les Knicks et les Magics utilisent le forage de données pour trouver des patrons dans les données de jeu. Ceci leur permet de modifier les stratégies adoptées

[1] George William "Bill" James est un écrivain, et statisticien spécialisé dans le domaine du baseball. Il est le chercheur le plus connu de la Society of American Baseball Research, (SABR) ou sabermetrics.

durant les parties et créer de meilleures combinaisons de joueurs. Dans la ligue de football américain (NFL), les Patriotes de la Nouvelle-Angleterre utilisent depuis quelques années une approche analytique dans toutes les facettes de la gestion de l'équipe, ce qui leur a permis de gagner trois « Super Bowl » en quatre ans, tout en respectant le plafond salarial (Davenport et al., 2007).

1.4.1 Technologies vidéo dans le sport

De plus en plus de compagnies offrent des solutions logiciels d'analyse vidéo en tout genre aux équipes sportives afin d'améliorer leur performance. En effet, des outils d'annotation et de segmentation de vidéo sont utilisés pour analyser les stratégies employées durant une partie, comme « XoS digital »[2] par le Vert et Or. D'autres outils, par exemple Sport Motion[3], focalisent sur l'analyse vidéo ralentie pour détecter des erreurs dans les mouvements. Un frappeur au baseball l'utilise afin de corriger son élan et par le fait même réduire le risque de blessure. Cependant, bien que l'utilisation de ce type de technologie soit un atout pour la gestion de la performance d'une équipe sportive, il serait plus profitable d'intégrer ces données vidéo avec les données de jeu, afin de centraliser l'information utile pour la prise décision dans un entrepôt de données. Celui-ci permettrait d'avoir une vue d'ensemble sur toutes

[2] XoS digital www.xosdigital.com
[3] Sport Motion www.sportmotion.com

les données importantes de l'entreprise, en plus de simplifier grandement la complexité des requêtes à effectuer.

Nous n'avons pas été en mesure de trouver un produit commercial basé sur cette approche lors de nos recherches sur le Web. Cependant, il est difficile de savoir en détail ce que les équipes sportives adoptent comme stratégies pour améliorer la performance de leur équipe, puisqu'elles ne veulent pas divulguer leur secret.

Avec une structure de données multidimensionnelle multimédia, une équipe sera capable de cerner plus efficacement les forces et faiblesses des stratégies employées lors des entrainements et des parties ainsi que d'améliorer l'évaluation de la performance des joueurs tout en réduisant le risque lié à la prise de décision. Ainsi, une équipe qui saura se différencier dans son analyse des données aura l'opportunité de se bâtir un avantage sur les autres.

1.5 Question de recherche

Étant donné l'importance accordée aux données vidéo par les entraineurs dans les analyses de performance des tactiques de jeu et des joueurs, nous croyons qu'il est primordial de pouvoir intégrer ce type de données avec les données de jeu. Alors, dans l'objectif de fournir un modèle de

données multidimensionnel soutenant l'intégration de l'information multimédia dans les analyses de performance des joueurs d'une équipe sportive et de combler un manque dans la littérature sur le sujet, nous posons cette question de recherche:

Comment intégrer des données multimédias et des données de jeu au niveau factuel et dimensionnel dans la conception d'un modèle de données multidimensionnel?

1.6 Objectif de la recherche

L'objectif principal de cette recherche est de comprendre comment il est possible d'intégrer des données multimédias et des données de jeu dans un modèle de données multidimensionnel, tant au niveau des tables de faits que des dimensions. La conception d'un prototype de modèle de données multidimensionnel pour soutenir les analyses de performances sportives liées à des clips vidéo nous permettra d'atteindre notre objectif.

1.7 Conclusion et organisation du mémoire

Afin de réaliser cette étude, nous avons travaillé conjointement avec l'équipe de football universitaire de Sherbrooke, le Vert et Or. Au cours des dernières années, l'utilisation de la vidéo est devenue pour eux un facteur déterminant et indispensable dans l'évaluation des joueurs et des tactiques de

jeu. En effet, les entraineurs et joueurs ont accès en tout temps aux séquences vidéo afin de revoir certains détails techniques. Cependant, l'exploitation des données vidéo intégrant les données de jeu est presque inexistante. Ceci est dû à la difficulté d'intégrer des données de plusieurs sources et de formats différents. Alors, nous croyons que la modélisation d'un entrepôt de données intégrant la vidéo ne peut qu'améliorer la performance de l'équipe.

Dans le premier chapitre, nous avons présenté le contexte, l'objectif et la question de la recherche. Ce mémoire est composé de quatre autres chapitres. Le deuxième chapitre est constitué du cadre théorique qui permettra de présenter, d'éclaircir et d'approfondir le contexte de la recherche. Le cadre méthodologique de la recherche sera présenté dans le chapitre trois. Le chapitre quatre présente les résultats de la recherche. Ensuite, nous concluons cette étude en proposant des avenues de recherche et en énonçant les limites de celle-ci dans le chapitre cinq.

Chapitre 2 : Cadre théorique

Ce chapitre présente les différents concepts essentiels à la poursuite de notre recherche. Il nous permet de faire l'état des connaissances nécessaires afin de pouvoir mener à terme ce projet de recherche. Tout d'abord, le concept d'entrepôt de données est défini, puis il est suivi par la modélisation multidimensionnelle. Ensuite, nous traitons du multimédia en expliquant brièvement son importance. Par la suite, nous nous concentrons sur le média vidéo et son annotation. Finalement, nous parcourons la littérature traitant des entrepôts de données multimédias et en faisons un résumé.

2.1 L'entrepôt de données

L'entrepôt de données (EDD) est une base de données où celles-ci sont structurées de façon à faciliter leur accès aux utilisateurs finaux. Inmon (2005) le définit comme étant un ensemble de données orientées sujet, intégrées, non volatiles, historisées et structurées pour soutenir le processus de prise de décision. Toujours selon le même auteur, l'EDD intègre l'ensemble des données pertinentes à la prise de décision, provenant de multiples systèmes sources de données. Avant d'être intégré à l'entrepôt, l'ensemble des données doit être nettoyé et transformé. Les conflits de nomenclature ainsi que les

incohérences des données entre les systèmes sont des défis majeurs au niveau de l'intégration des données (Turban et al. 2007). L'EDD permet de consolider des données détaillées et des données agrégées. De plus, contrairement à un système transactionnel, l'EDD emmagasine plusieurs années de données pour permettre aux utilisateurs de faire des analyses de tendance, dans le but de faire des comparaisons et des prédictions (Turban et al. 2007). De plus, les données sont non volatiles, ce qui permet de retracer dans le temps la valeur exacte d'un fait à un moment désiré (Inmon, 2005). Un axe temporel est associé aux données afin de récupérer la valeur valide au moment voulu. La Figure 1 représente l'architecture de l'entrepôt de données. Les données sont extraites des sources de données internes (systèmes opérationnels) et externes (Web). Elles sont ensuite nettoyées, transformées (filtres, agrégations, calculs) et chargées dans l'EDD. Les utilisateurs finaux accèdent aux données par l'intermédiaire d'outils OLAP, de rapports et de tableaux de bord.

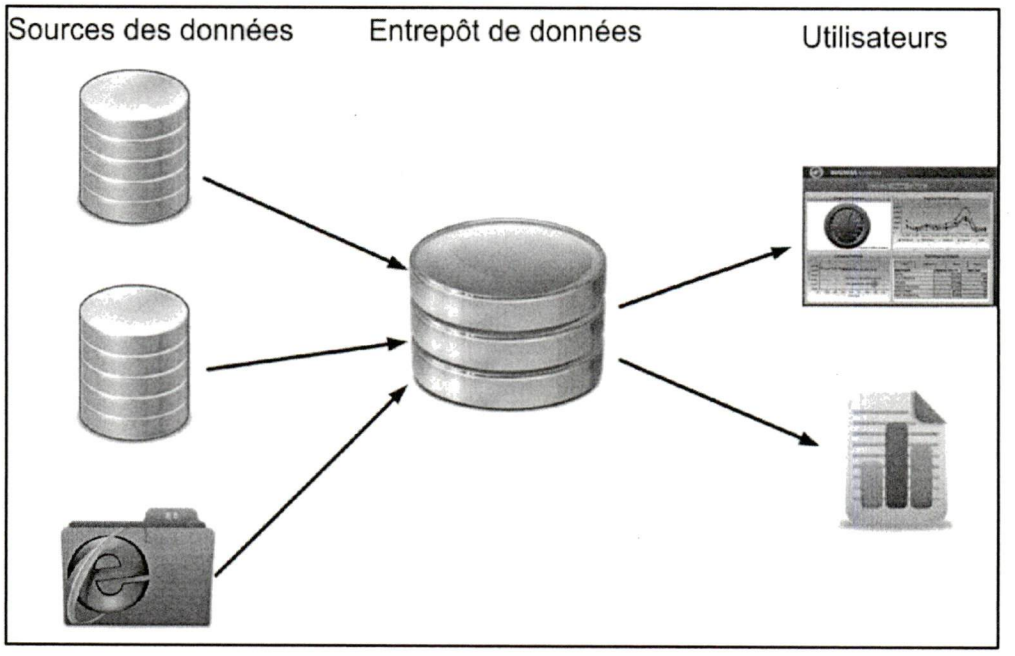


Figure 1: L'architecture de l'entrepôt de données

2.1.1 La modélisation multidimensionnelle

La mauvaise performance des systèmes transactionnels face aux requêtes complexes dans la prise de décision a amené une modélisation différente pour l'entrepôt de données. Trois types de modèle dimensionnel sont suggérés: en étoile, en flocon et la constellation (Figure 2) (Ballard, Farrel, Gupta, Mazuela et Vohnik, 2008). Le modèle en étoile est un modèle dénormalisé contenant une table de faits centrale liée à plusieurs dimensions. Le modèle en flocon est un modèle en étoile où une ou plusieurs dimensions ont été explosées pour créer une normalisation des dimensions afin de réduire

la redondance. La constellation est simplement un modèle dimensionnel composé de plusieurs étoiles. Par contre, Kimball et al. (2004) déconseille l'utilisation du modèle en flocon puisque la normalisation du modèle affecte grandement la performance des requêtes.

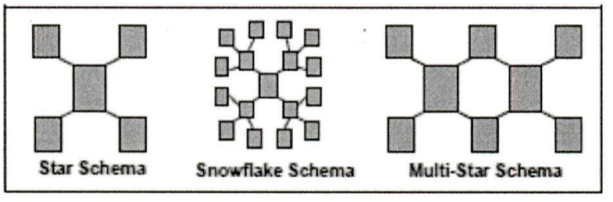


Figure 2 Types de modèles dimensionnels

Un modèle multidimensionnel est composé de deux éléments, des faits et des dimensions (Romero et Abelló, 2009). Les faits sont utilisés pour enregistrer des événements ou des mesures (Inmon, Strauss, et Neushloss, 2008; Kimball et Ross, 2002). Celles-ci sont des données numériques additives ou semi-additives, par exemple le nombre de verges de passe d'un quart-arrière et le nombre de sacs du quart, utilisées pour soutenir les gestionnaires dans la prise de décision de façon quantitative (Bonifati, Cattaneo, Ceri, Fuggetta, Paraboschi, 2001; Kimball et Ross, 2002, Romero et Abelló, 2009). De plus, chaque fait est mis en contexte par des tables de dimensions qui contiennent des descriptions textuelles (Kimball et al., 2002). En effet, les dimensions servent à contextualiser les événements en représentant différents axes d'analyse, par exemple la date du match et le nom des joueurs (Bonifati et

al., 2001). Elles rendent possibles les opérations d'agrégations telles que: « somme », « moyenne » et « nombre » (Schneider, 2008). La Figure 3 représente un exemple d'un modèle multidimensionnel en étoile qui ne fait pas partie de notre modèle final. On retrouve dans ce modèle six tables de dimensions : « d_joueurs », « d_equipe », « d_essais », « d_dates », « d_temps » et « d_distance » ainsi que la table de faits « f_jeux »

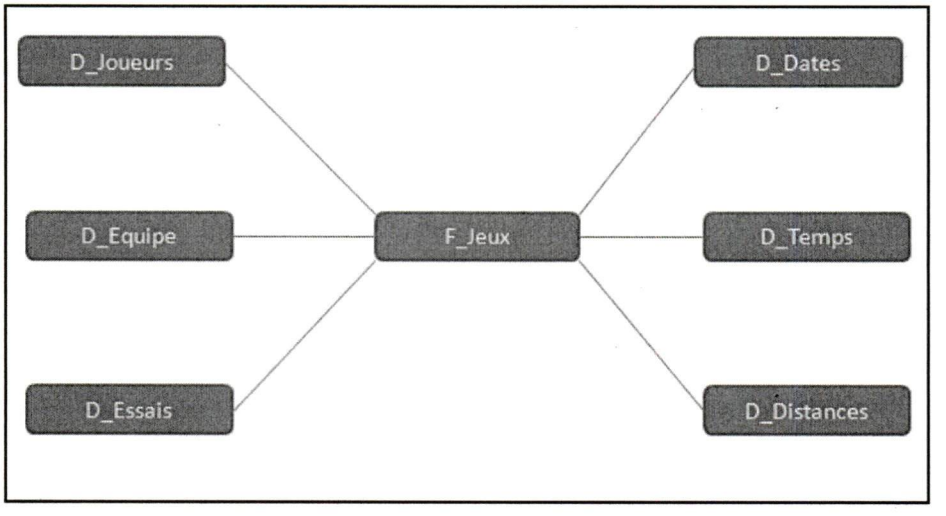


Figure 3: Modèle multidimensionnel en étoile

En conclusion, cette partie de la revue visait à mettre en place les connaissances nécessaires sur les entrepôts de données et les différents types de modélisation qui lui sont associés. Nous avons vu que l'entrepôt de données est utilisé pour remédier aux faiblesses des bases de données opérationnelles par rapport aux requêtes complexes. En suite, nous avons couvert les trois types de modélisation généralement utilisés dans la conception d'un entrepôt

de données. Des trois types de modèle, nous avons identifié que le modèle en étoile est plus performant puisqu'il requiert moins de jointures de tables lors des l'exécution des requêtes.

2.2 Le multimédia

Depuis plusieurs années, l'intérêt pour le multimédia a substantiellement augmenté. Ceci est dû en partie au fait que la diminution du coût des technologies reliées à la création et au stockage de contenu multimédia a chuté dans les dernières années (Figure 4). De plus, avec l'Internet et les nouvelles méthodes de compression, il est possible d'accéder facilement à ce type de contenu (Decleir, Kouloumdjian, et Hacid, 1999). Selon Carrer et al. (1997), l'utilisation des technologies de communication digitale comme la vidéoconférence, la visiophonie, les films sur demande et l'enseignement à distance sont petit à petit en train de rattraper l'utilisation de la télévision conventionnelle. Que ce soit sous forme de vidéo, d'image ou d'audio, les données multimédias peuvent fournir une grande quantité d'information. En effet, dans le domaine médical, on utilise le son de l'électrocardiogramme pour faire des analyses et on utilise les radiographies pour analyser l'état d'un patient (Arigon, Tchounikine, et Miquel, 2006). Dans le sport, l'utilisation de la vidéo est une pratique courante dans le but de permettre aux entraineurs et joueurs de revoir leur performance.

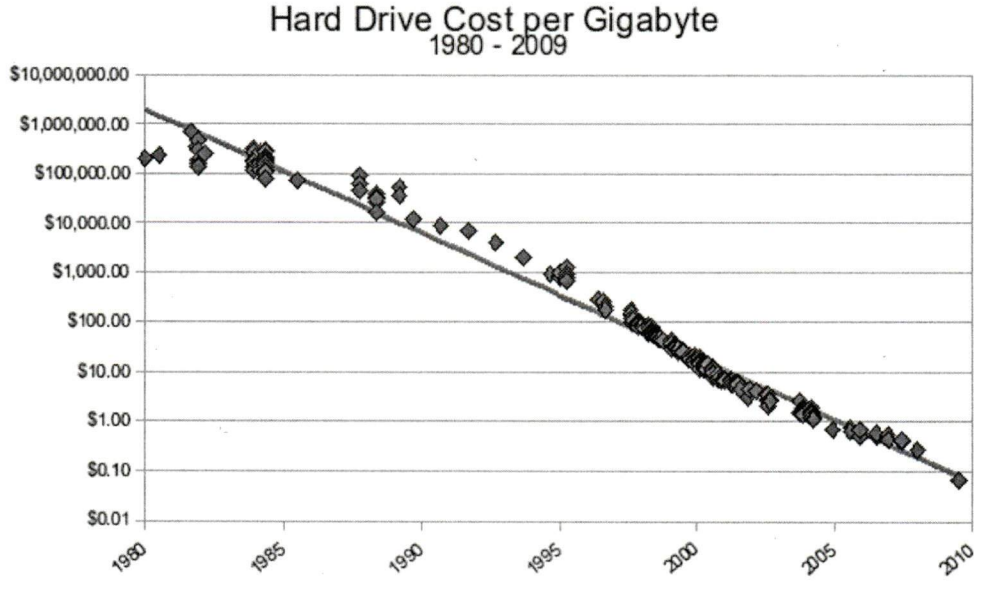


Figure 4 Coût par Gigabyte, figure tirée de Udayan (2011)

2.2.1 L'annotation des vidéos

Avec la quantité importante de données vidéos ainsi que leur grand volume, il devient de plus en plus compliqué d'en faire la gestion. Pour faciliter cette étape, il est possible d'associer des descripteurs aux séquences vidéo afin d'en décrire le contenu. L'annotation de vidéos se veut simplement une définition significative de la séquence vidéo pour en faciliter l'extraction de l'information (Decleir et al., 1999).

Il y a deux grandes approches pour faire l'annotation des vidéos selon Decleir et al. (1999), l'approche automatique et l'approche homme-machine. En

ce qui concerne l'approche automatique, il existe plusieurs systèmes qui, à l'aide d'algorithmes, vont fournir de l'information de bas niveau sur la vidéo, par exemple les formes, les textures et la résolution (Decleir et al., 1999). D'autre part, la technique homme-machine donne de l'information plus détaillée sur le contenu puisqu'il sera interprété par un humain en utilisant une liste de mots clés spécifiques au domaine pour en faire la description (Decleir et al., 1999). Cependant, il faut identifier l'endroit de la vidéo pour laquelle la description s'applique et étant donné la complexité du travail, la majorité des solutions existantes sont spécifiques à un type de vidéo, par exemple les nouvelles et les films (Carrer et al., 1997). Deux techniques sont souvent utilisées pour annoter une vidéo. La première technique, qui est la plus ancienne, est celle de la segmentation. Il suffit de segmenter la vidéo en plusieurs parties et d'annoter chacun des segments. La Figure 5 présente cette technique dans un exemple fictif, une séquence vidéo a été découpée en trois parties et chacune des parties est annotée. Une autre technique est celle de la stratification, proposée par Smith et Davenport (1993). Elle permet l'annotation des événements individuels tout en rendant possible le chevauchement des annotations. La Figure 6 démontre cette technique. Pour le jeu, il est possible de faire plusieurs annotations qui peuvent s'entrecouper.

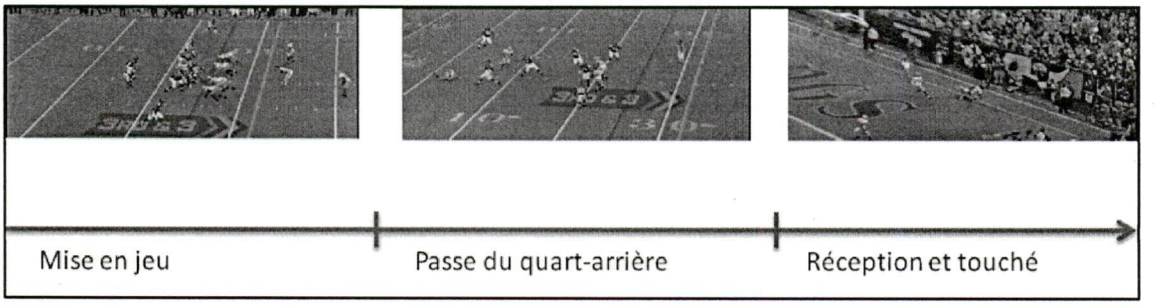


Figure 5 : Exemple de segmentation

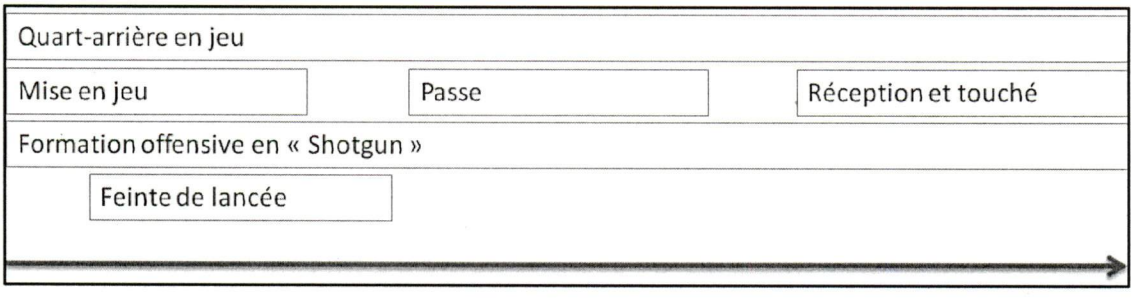


Figure 6: Exemple de stratification

2.2.2 L'entrepôt de données multimédias

La combinaison d'un entrepôt de données "traditionnel" à des données vidéo permettrait de faire des analyses de performance liées à des clips vidéo. Quelques études ont été faites sur la conception d'entrepôts de données multimédias.

Dans l'étude de You, Dillon, Liu et Pissaloux (2001), les auteurs proposent une structure d'entrepôt de données qui permet une grande flexibilité dans l'indexation et qui augmente la performance de l'extraction des données multimédias. Les auteurs proposent la combinaison du modèle en étoile et du modèle en flocon pour créer un "starflake". Les données multimédias sont insérées dans une table de faits et les dimensions correspondent aux différents types de données multimédias. Chacune des dimensions est normalisée, comme dans un modèle en flocon composé de différents descripteurs relatifs au type de données par exemple, les couleurs, la résolution et les textures. Au bout de la hiérarchie des dimensions, se retrouve une autre table de dimension qui elle est dénormalisée, comme dans un modèle en étoile.

Dans Wookey, Yongkyu, Yunsun et Jinho (1999), les auteurs portent leur étude sur les entrepôts de données intégrant la vidéo. Avant de charger les vidéos dans la table de faits, ceux-ci sont segmentés en regroupement d'images successives formant une scène. Ensuite, des mots clés sont associés à chaque segment. La dimension scène est construite à partir de ces descripteurs. D'autres dimensions sont ajoutées afin de venir compléter le modèle.

Zaïane Han, Li et Hou (1998) proposent un entrepôt de données multimédias construit à partir de vidéos et d'images du Web dans le but d'utiliser des techniques de forage de données. Les dimensions sont créées à

partir de descripteurs textuels retrouvés sur les pages Web et des descripteurs basés sur le contenu des données multimédias. Dans leur modèle, on retrouve par exemple des dimensions pour les couleurs, la taille et le format.

L'étude d'Arigon Tchounikine et Miquel (2006) traite de la gestion de multiples points de vue dans un entrepôt de données multimédia. Dans leur étude, les auteurs présentent un modèle multidimensionnel qui prend en compte les préférences des utilisateurs, selon, par exemple, leur niveau d'expertise. De plus, ils ont su utiliser des fonctions d'agrégation spécifiques aux données multimédias intégrées dans l'entrepôt de données. Les auteurs dénotent quelques difficultés rencontrées lorsque l'on veut intégrer des données multimédias à un entrepôt de données. Comparativement à des données alphanumériques, que l'on retrouve traditionnellement dans un entrepôt de données, les données multimédias sont beaucoup plus volumineuses, ce qui requiert une grande capacité de stockage et peut avoir une influence sur la performance. Ensuite, les fonctions d'agrégation traditionnellement utilisées sur les données numériques, comme « somme » et « moyenne », ne sont pas compatibles avec les données multimédias. Finalement, le langage d'interrogation de base de données SQL n'est pas optimisé pour faire des requêtes sur les données multimédias.

2.2.3 Conclusion sur le multimédia

De ce fait, nous avons constaté qu'il y a peu d'études faites sur les entrepôts de données multimédias. De plus, la majorité des études s'attardent à l'analyse technique des données multimédias, telles que la couleur, la texture, la résolution et la taille. Dans notre étude, nous voulons utiliser les données vidéo afin de soutenir visuellement les analyses produites par l'équipe du Vert et Or. Nous avons aussi remarqué que dans l'ensemble des études, les données multimédias sont uniquement insérées dans les tables de faits. Nous croyons qu'il serait intéressant d'avoir les données vidéo tant au niveau factuel que dimensionnel dans un même modèle afin d'améliorer la flexibilité des analyses.

2.3 Football Vert et Or

Dans cette section, nous couvrons les éléments nécessaires à la compréhension du mémoire au niveau du football et de la structure des entraineurs de l'équipe du Vert et Or.

Tout d'abord, la structure du Vert et Or est composée de dix entraineurs dans quatre différents groupes (Figure 7)[4]. Comme mentionné auparavant,

[4] Site web du Vert et Or football www.Usherbrooke.ca/football

Martin Montminy, que nous avons rencontré était à ce moment l'entraineur de la ligne offensive du club.

Ensuite, la Figure 8 présente les différentes positions offensives et défensives. Lorsque nous faisons référence à la ligne offensive, nous parlons du centre, des gardes et des bloqueurs. De l'autre côté, lorsque nous faisons référence à la ligne défensive, il est question des plaqueurs défensifs et des ailiers défensifs.[5].

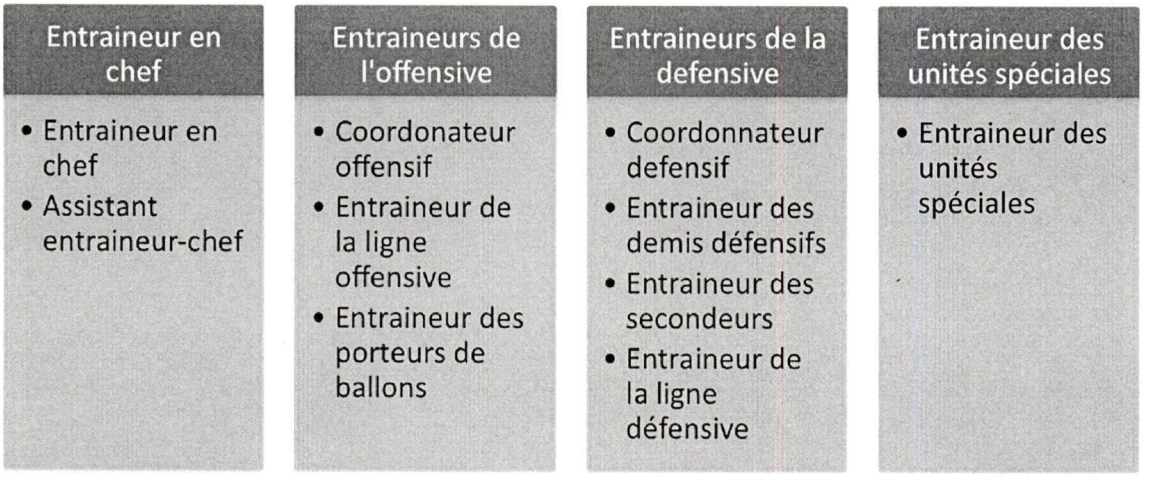


Figure 7 Liste des entraineurs par groupe d'entraineur

[5] Site web de la CFL www.CFL.com

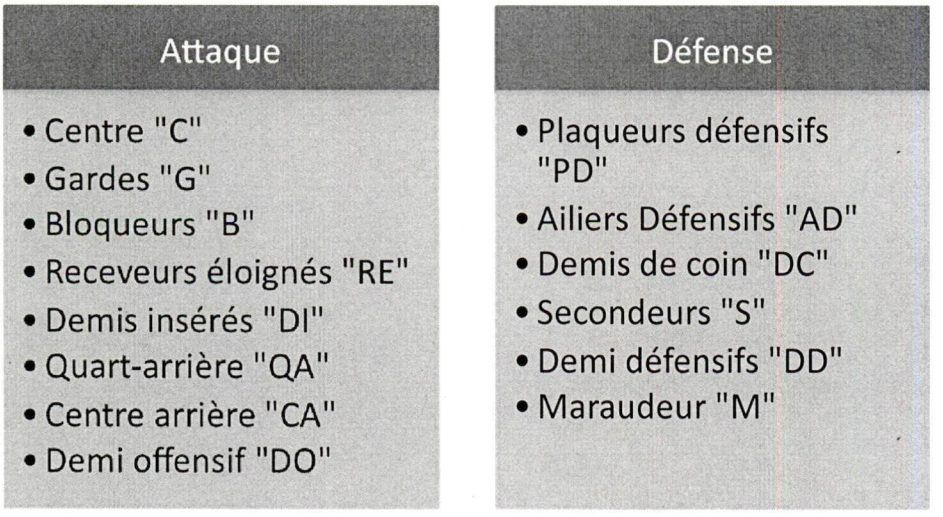


Figure 8 Positions offensives et défensives

2.4 Conclusion

L'objectif de ce chapitre était de mieux comprendre les différents concepts liés à notre étude. En premier lieu, nous avons identifié l'entrepôt de données comme le point central qui intègre les données des différents systèmes opérationnels afin de pouvoir faire des analyses. Ensuite, nous avons présenté le multimédia et approfondi l'annotation du média vidéo. Par la suite, nous avons montré les différentes études sur les sujets au cœur de notre recherche, l'entrepôt de données multimédias. Finalement, les bases requises du football ont été présentées afin de permettre à tous de comprendre les termes techniques de ce sport.

Chapitre 3 Cadre méthodologique

Dans ce chapitre, nous décrivons la méthodologie suivie afin d'atteindre notre objectif de recherche qui est de comprendre comment intégrer des données multimédias et des données de jeu dans un modèle de données multidimensionnel. La Figure 9 présente les différentes étapes de la méthodologie de recherche.

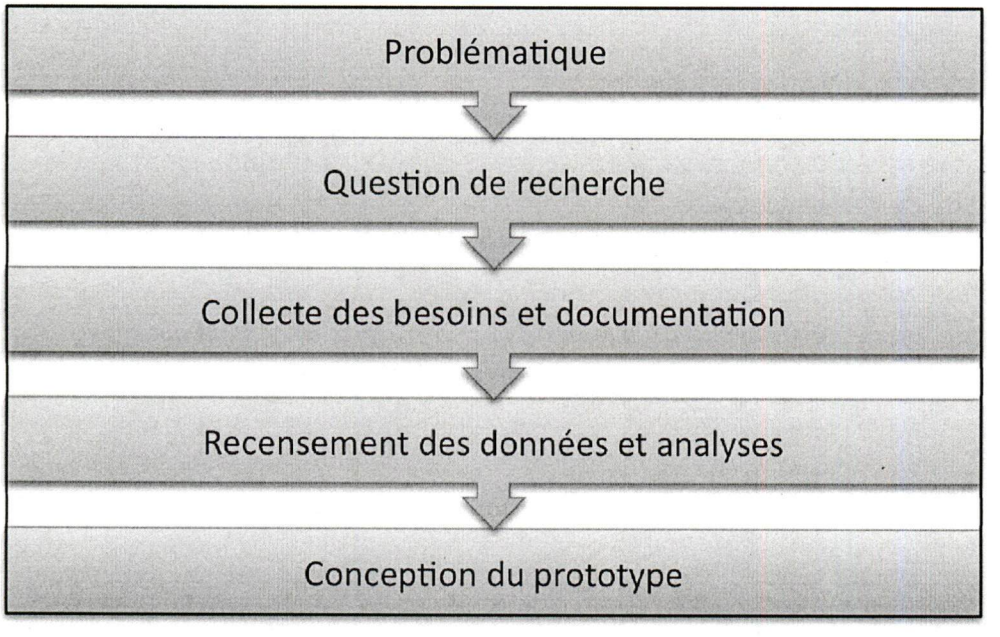


Figure 9: Étapes de la méthodologie de recherche

Dans le cadre de notre recherche, nous utilisons une méthodologie de développement par prototypage pour la conception de l'EDD. Cette méthodologie se retrouve dans un paradigme de « design science ». Celle-ci prend racine dans l'« engineering and the science of the artificial » (Simon,

1996 cité dans Hevner 2007). Celle-ci tend à créer des innovations technologiques définissant des idées, des pratiques et des produits (Hevner, 2004). L'objectif principal de cette méthodologie est la création d'un artefact technologique dans le but de répondre aux besoins d'une organisation (Hevner, 2004). L'artefact est représenté selon différentes formes : un construit, un modèle, une méthode et une instanciation (March et Smith, 1995).

Hevner (2007) propose un modèle de recherche en trois cycles : le cycle de pertinence, le cycle de rigueur et le cycle de conception (Figure 10).

Cycle de pertinence

La recherche doit être motivée par le désir d'améliorer l'environnement organisationnel par le développement d'un artefact (Simon, 1996, Hevner 2007). Selon Hevner (2007), une bonne recherche en « design science » doit être basée sur une problématique ou opportunité actuelle dans un environnement.

Cycle de rigueur

La recherche en « design science » doit prendre fondation sur les théories scientifiques, expertises, artefacts et processus déjà existants (Hevner, 2004, Hevner, 2007). Ce cycle apporte les connaissances nécessaires pour

s'assurer que le projet de recherche contribuera lui-même à cette base de connaissances (Hevner, 2007).

Cycle de conception

Le cycle de conception est au cœur de la recherche en « design science ». Ce cycle a pour objectif de comparer l'artefact aux exigences jusqu'à l'obtention d'une conception satisfaisante (Simon, 1996 cité dans Hevner 2007).

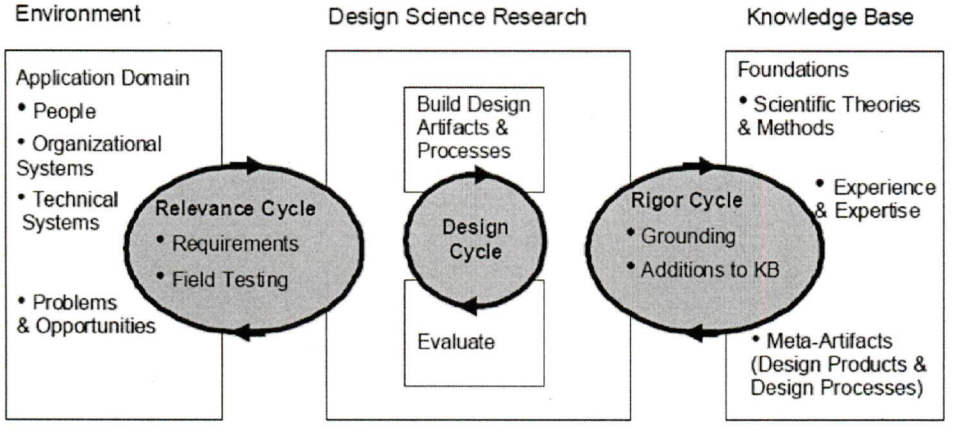


Figure 10 Les trois cycles de la recherche en "design science" de Hevner 2007

Nous baserons notre recherche sur cette méthodologie afin de nous assurer que nous utilisons une approche établie et acceptée dans le milieu de la recherche académique en système d'information. Tout d'abord, le cycle de pertinence est soulevé dans l'Introduction avec l'identification de la problématique ancrée dans la réalité des entraineurs du Vert et Or. Ensuite, le

cycle de rigueur de la recherche est assuré par le Cadre théorique qui fait l'état des connaissances nécessaires à la recherche et par le Cadre méthodologique. Finalement, la conception de l'artefact, sa contribution et les limites seront détaillées dans les chapitres Résultats et la Discussion.

3.1 Pertinence du projet

Tout d'abord, rappelons que l'objectif principal de cette recherche est de concevoir et d'analyser un prototype d'un modèle de données multidimensionnel dans le but de soutenir les analyses de performance liées à des clips vidéo tournés par les entraineurs de l'équipe de football universitaire le Vert et Or.

3.2 L'analyse des besoins et documentation

L'analyse des besoins est une étape importante dans le développement des systèmes d'information. Celle-ci vise à définir les besoins d'affaires du commanditaire de projet pour qui le système est développé. Dans le domaine de l'intelligence d'affaires et plus spécifiquement dans le développement d'entrepôts de données, il est dit que beaucoup de projets échouent parce qu'ils ne répondent pas aux besoins d'affaires (Giorgini, Rizzi et Garzetti, 2005). Il existe différentes méthodologies de développement d'entrepôts de données qui sont classées en deux grandes catégories :

- Guidé par les données : Ce type d'approche débute avec une analyse détaillée des sources de données. Du résultat de l'analyse, il y aura création du modèle de données (Romero et al., 2009).
- Guidé par les besoins : Ce type d'approche commence par la détermination de l'information requise par les utilisateurs de l'entrepôt. Ensuite, il faut identifier les sources de données nécessaires au développement pour finalement créer le modèle de données (Romero et al., 2009).

Il existe aussi plusieurs techniques hybrides qui combinent les deux approches. Bonifati et al (2001) et Cabibbo et Torlone (1998) en présentent une divisée en trois étapes :

1. Analyse guidée par les besoins
2. Analyse guidée par les données
3. Intégration des deux techniques

Dans le cadre de notre recherche, nous utilisons une méthode hybride. Cette méthode intègre et concilie les deux paradigmes. En effet, celle-ci combine les avantages des deux méthodes sans toutefois combiner les inconvénients. Premièrement, elle s'assure que les besoins des utilisateurs sont bien compris et intégrés dans la solution finale. Deuxièmement, elle

permet d'explorer les données des systèmes opérationnels et d'en déduire la sémantique des données. Finalement, l'étape d'intégration apporte une solution réalisable selon les données disponibles tout en tenant compte des besoins.

La première étape est d'identifier les besoins des utilisateurs, en plus de comprendre quelles sont les différentes analyses faites par ceux-ci. Par le fait même, un inventaire des systèmes sources utilisés par les utilisateurs est dressé. Ceci permet d'identifier l'ensemble des sources de données potentiellement exploitables lors de l'étape de l'inventaire des données. À la fin de cette étape, il faut être en mesure de dériver, à partir des besoins exprimés et des analyses faites, certains faits et dimensions qui iront dans le modèle de données.

La deuxième étape consiste en une analyse approfondie des données comprises dans les sources de données de l'organisation en se basant sur l'inventaire des systèmes utilisés par les utilisateurs. En effet, le but est d'identifier l'ensemble des dimensions et faits que l'on peut déduire directement des données contenues dans ces systèmes. De plus, cette analyse des données permet d'identifier plusieurs métriques qui peuvent être faites à partir de celles-ci.

La dernière étape en est une d'intégration. La première étape fait ressortir la structure de l'entrepôt de données basée sur les besoins des

utilisateurs. La deuxième étape fourni l'ensemble des modèles multidimensionnels qu'il est possible de créer à partir des données des systèmes sources. L'étape finale consiste en l'intégration des modèles potentiels qui ont été déduits à partir des données et des besoins, lors de l'exécution des deux premières étapes pour en faire un modèle final. Par le fait même, cette étape permet aussi d'identifier de nouvelles opportunités analytiques qui n'ont pas été détectées lors de la première étape. En effet, l'étape deux amène une panoplie de données et de métriques auxquelles les utilisateurs n'ont pas pensé lors de la cueillette des besoins.

3.3 Le développement du système

L'approche utilisée dans le cadre de notre recherche pour développer le modèle multidimensionnel est une méthode par prototypage. Cette approche itérative procure plusieurs avantages au niveau du développement. En effet, Tripp et Bichelmeyer (1990) définissent le prototype comme un système fonctionnel, non complet et temporaire qui contient les fonctionnalités essentielles du système final. Cette méthode permet un développement rapide et facilement modifiable. Nielson (1993) identifie deux différents types de prototype, le vertical et l'horizontal. Le prototype horizontal correspond au développement de l'interface homme-machine, généralement fait sous forme de maquette. Le prototype vertical met en place les fonctionnalités principales pour qu'un utilisateur puisse réaliser une tâche dans son ensemble. Le

processus de développement par prototypage se divise en quatre étapes (Trip et al., 1990) (Figure 11).

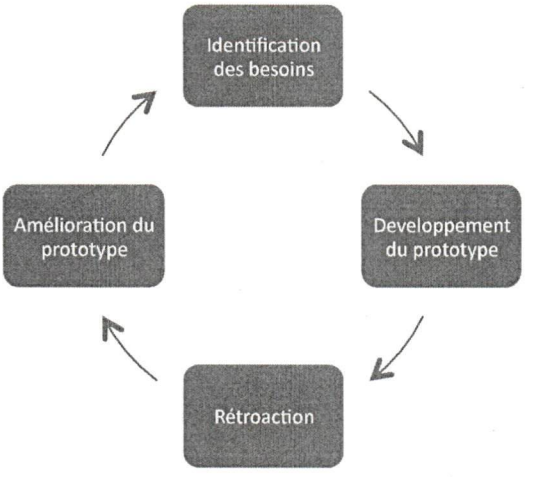


Figure 11 Étapes de développement d'un prototype

Dans le cadre de notre recherche, nous adoptons la technique de prototypage de type vertical pour quatre raisons.

1. Nous voulons construire un prototype fonctionnel et le prototypage vertical nous permet de développer une solution fonctionnelle.
2. La modélisation d'entrepôt multimédia en est encore à ses balbutiements. De ce fait, la flexibilité offerte par le prototypage nous permettra de faire les modifications requises en cas de problèmes.
3. Le temps requis pour développer un prototype fonctionnel est nettement inférieur au développement d'un système complet.

4. Permettre à l'utilisateur final de proposer des modifications tout au long du processus de développement et de les intégrer lorsque le projet s'y prête.

Il est à noter que dans le cadre de notre recherche, nous ne ferons qu'une seule itération pour des raisons de simplicité liée à des contraintes de temps. Cependant, il sera possible pour le Vert et Or de continuer à faire évoluer le projet s'il en voit le besoin.

3.4 Conclusion de la méthodologie

Dans ce chapitre, nous avons présenté le cadre méthodologique utilisé dans notre recherche. Premièrement, nous avons introduit le cadre de l'analyse des besoins en présentant une technique hybride qui se base sur les besoins de l'utilisateur ainsi que sur les données disponibles dans les systèmes sources. Deuxièmement, nous avons présenté la technique de développement par prototype que nous allons utiliser. Le chapitre suivant est consacré à l'élaboration du modèle de données multidimensionnel.

Chapitre 4 Résultats

À travers le présent chapitre, nous présentons le déroulement de la conception du prototype et par le fait même, les résultats de l'étude. Premièrement, nous effectuons un inventaire des systèmes et des données de l'équipe et identifions leurs besoins au niveau des analyses. Deuxièmement, nous présentons notre modèle de données découlant de l'inventaire des données et des besoins recueillis. Troisièmement, la création d'un processus d'extraction, de transformation et de chargement (ETC) nous permet d'extraire les données et de les charger dans notre entrepôt de données. Finalement, nous approfondissons le volet de l'intégration de la vidéo dans les tableaux de bord.

4.1 Inventaire des besoins et leur utilisation

Dans le cadre de notre projet, nous avons rencontré l'entraineur de la ligne offensive de l'équipe de football du Vert et Or à trois reprises afin de comprendre leur processus d'analyse des performances de l'équipe et de faire l'inventaire des systèmes et des données afin d'identifier leurs besoins au niveau des rapports.

La première étape est l'analyse des besoins de l'équipe au niveau de l'évaluation de la performance. Pour l'équipe, il est difficile d'identifier précisément ses besoins puisqu'elle est ancrée dans un processus fonctionnel et elle ne sait pas ce que l'IA peut lui apporter de plus. Aux yeux de l'entraineur, tout système pouvant améliorer le processus d'évaluation de la performance en automatisant celui-ci serait d'une grande aide pour son équipe. De ce fait, nous avons décidé de porter une plus grande attention à l'analyse du processus d'évaluation de la performance et des différents systèmes entourant celui-ci.

4.1.1 Processus d'évaluation de la performance dans le Vert et Or

Analyse des séquences vidéo

Dans un premier temps, nous portons une attention particulière au processus d'évaluation de la performance de l'équipe appuyé par la vidéo. Ceci nous permettra d'identifier les sources de données potentielles en plus de nous familiariser avec le type d'analyse fait par l'équipe. Actuellement, l'équipe utilise un processus complexe d'analyse des séquences vidéo. En effet, après chaque partie, une personne est chargée de transférer la vidéo de la partie dans une application spécialisée (XoS digital). Celle-ci leur permet de découper la partie en séquences vidéo jeu par jeu. Par la suite, des données complémentaires sont entrées manuellement pour chaque séquence vidéo, par exemple la formation, la distance et l'essai. L'entraineur de la ligne offensive, quant à lui,

regarde les séquences vidéo et note chacun de ses joueurs selon une mesure développée à l'interne. Le lendemain de chaque partie, l'équipe se réunit pour faire une analyse du match jeu par jeu et les entraineurs en profitent pour commenter la performance de chacun des joueurs. De plus, le Réseau du Sport Étudiant du Québec (RSEQ)[6] publie sur son site web les détails de chaque match. Ces données sont utilisées pour cumuler les statistiques des joueurs. Des ordinateurs sont mis à la disposition des joueurs afin qu'ils puissent accéder à l'application XOs pour revoir leurs performances. Cependant, cela requiert un effort non négligeable de la part des entraineurs et des joueurs puisque ce processus d'évaluation supporte mal l'intégration des données provenant de la RSEQ.

Suivi du conditionnement physique

La condition physique des joueurs est un aspect très important dans les sports de compétition. L'équipe de football du Vert et Or comporte un entraineur chargé de l'entrainement physique en salle des joueurs. Chaque joueur a sa routine d'entrainement à respecter et note le poids et le nombre de répétitions faites pour chaque exercice complété. De cette façon, les entraineurs peuvent suivre l'évolution physique des joueurs. Cependant, les données sont récoltées sur papier ce qui fait qu'elles sont difficilement exploitables pour des fins

[6] RSEQ assure la promotion et le développement du sport et de l'activité physique en milieu étudiant, de l'initiation jusqu'au sport de haut niveau.

d'analyse. De plus, ces données ne sont pas intégrées aux données de jeu par l'équipe.

Suivi de la Nutrition

La nutrition est un enjeu de plus en plus important dans le sport selon l'entraineur du Vert et Or. Présentement, aucun suivi n'est fait auprès des joueurs, mais les entraineurs sont bien conscients de l'importance de la nutrition sur la forme physique et sur la performance des joueurs sur le terrain. Il est donc envisageable de voir un processus de suivi de la nutrition des joueurs dans les années à venir.

Une fois que nous avons saisi globalement le processus d'analyse des séquences vidéo et d'évaluation des joueurs, nous constatons qu'il est possible d'améliorer l'intégration des données du RSEQ avec les données internes de l'organisation. Nous pensons être capables de réduire le nombre de tâches manuelles avec un processus d'extraction, de transformation et de chargement (ETC), telles que l'intégration des données de jeux à la vidéo. Ceci permet d'intégrer les données du RSEQ aux données internes et aux séquences vidéo dans un même endroit. Finalement, l'équipe aura la possibilité d'ajouter d'autres secteurs d'activités à son EDD, tel que les données de conditionnement physique et de nutrition, lorsque les processus seront mis en place.

4.1.2 Inventaire des systèmes et des données

Dans le but de bien comprendre l'ensemble des possibilités d'analyse, nous regardons plus attentivement les sources de données et leurs valeurs. L'inventaire des données nous permet d'identifier quelles sont les données disponibles, où on peut les retrouver, leur format et leur qualité. Dans notre recherche, nous concentrons nos efforts sur les données provenant du site web du RSEQ, les séquences vidéo ainsi que les données de pointage de la ligne offensive. Les données de conditionnement physique et de nutrition ne sont pas intégrées au système puisqu'il n'y a pas encore de processus de suivi de la nutrition des joueurs et l'effort d'informatiser les résultats d'entrainement physique est considérable. Nous aborderons ce sujet dans la discussion au Chapitre 5. L'étape de l'inventaire des données est très importante puisque celle-ci nous permettra de dresser un éventail de possibilités face à la modélisation des tables de faits et de dimensions.

Le RSEQ diffuse un grand nombre de statistiques pour chaque partie de football universitaire du Québec sur son site web[7]. Nous avons passé en revue l'ensemble des données disponibles sur celui-ci. Les données analysées vont du nom des joueurs au nombre de verges de passe par le quart-arrière en passant par la description de chaque jeu. Dans ces données, nous remarquons

[7] Site web du RSEQ www.sportetudiant.com

différents niveaux d'agrégations. En effet, on dénote quatre niveaux de données : jeu par jeu, par joueur, par équipe et par partie (Table 1).

Niveau agrégation	Nom des fichiers
Par jeu	Play by play , offensive line
Par joueur	Rushing, passing, receiving, punting, all returns, feild goal attemps, kickoffs, defensive
Par équipe	Team statistics, participation report
Par saison	Seasons game

Table 1 Niveau d'agrégation des fichiers

L'information critique que l'on retrouve à l'interne de l'organisation du Vert et Or est la vidéo et les données de jeu. En effet, chaque partie et plusieurs entrainements sont filmés avec deux caméras vidéo proposant des angles de jeu différents : un de la ligne de côté et un de la zone des buts. Les séquences vidéo sont au cœur de la prise de décision au sein de l'équipe puisque celle-ci apporte plus d'information qu'une simple statistique en fournissant l'ensemble du contexte d'une action. En effet, lorsque l'entraineur regarde simplement les statistiques, il peut constater qu'un joueur a fait un plaqué. Par contre, la vidéo permet de contextualiser l'action qui a été faite par

le joueur. Ceci permet à l'entraineur d'évaluer la technique ainsi que la décision prise par le joueur. La vidéo est une donnée de type non structurée qui renferme une panoplie d'information décodable par l'entraineur aguerri, mais difficile à formaliser et à encoder dans une base de données.

D'autre part, l'équipe du Vert et Or cumule une statistique spéciale pour les joueurs de la ligne offensive, composée de cinq joueurs : le centre (C), deux gardes (G) et deux bloqueurs (B) (Figure 12). Leur objectif principal est de défendre le quart-arrière lors de jeu de passe ou d'ouvrir un chemin pour le coureur lors des jeux au sol. Cette statistique a été conçue par un étudiant à la maitrise en kinanthropologie et elle consiste à donner, pour chaque jeu, une note variant de 0 à 2 en intervalle de 0,5 pour chacun des joueurs de la ligne offensive. Alors, après chaque partie, l'entraineur de la ligne offensive revoit les vidéos et note ses joueurs pour chaque jeu et cumule une note par joueur pour chaque partie ainsi qu'une note globale pour la ligne offensive.

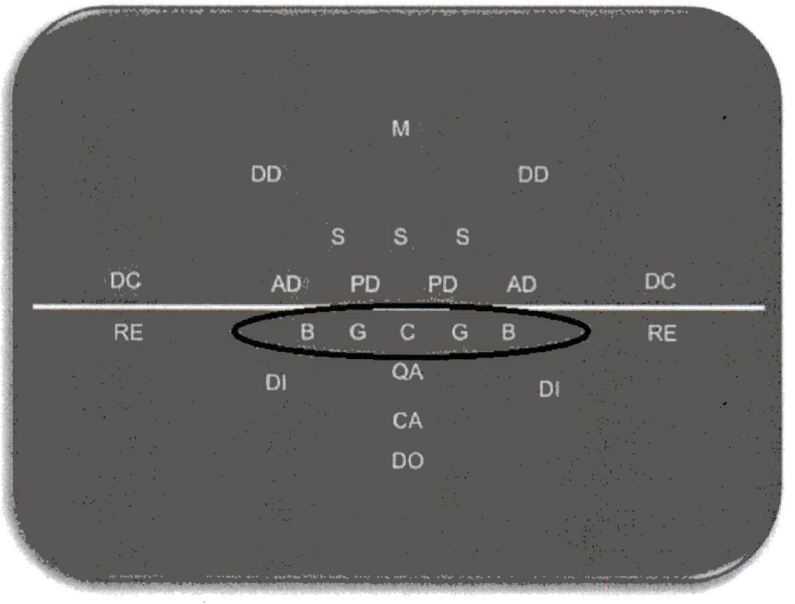


Figure 12 Joueurs de la ligne offensive dans une formation d'attaque de base tirée de wikipedia

4.1.3 Données

Maintenant que l'identification des données utiles à l'évaluation de la performance est terminée, nous passons à l'étape de la vérification des données. Le premier constat que nous faisons lorsque nous regardons les données du RSEQ est que le format des données est non structuré. En effet, les données sont publiées sur le site web en format de fichier plat de type texte. Ceci implique qu'il est impossible d'obtenir les données de façon tabulaire sans effectuer une manipulation importante des fichiers.

Il existe trois types de données : structurées, non structurées et semi-structurées. Lorsque l'on parle de données non structurées, on réfère

généralement à des données auxquelles on ne peut pas identifier de structure. On dit aussi qu'une donnée non structurée est une donnée que l'on ne peut pas insérer dans une colonne ou une rangée d'une base de données sans transformation. Par exemple, le fichier plat de type texte, la vidéo et le son sont des données non structurées. À l'opposé, les données structurées sont des données qui respectent une structure, par exemple une base de données relationnelle ou tabulaire. Entre les deux types de données, on retrouve les données semi-structurées. Celles-ci ne font pas référence à un schéma, mais ce sont des données qui contiennent généralement des métadonnées ou des balises. La Table 2 décrit quelques exemples pour chaque type de données.

Type de données	Exemples
Données structurées	Base de données, fichier plat,
Données non structurées	Image, vidéo, texte en libre
Données semi-structurées	XML, fichier Excel

Table 2 Exemples de données structurées, non structurées et semi-structures

Dans l'organisation du Vert et Or, on retrouve plusieurs données, des fichiers plats de type texte provenant du site web du RSEQ, des fichiers Excel

pour l'évaluation à l'interne de la ligne offensive et des vidéos pour l'évaluation des joueurs. Aucune des données utilisées par l'équipe n'est du type structuré. La Figure 13 présente quelques exemples de données non structurées utilisées par l'équipe. Nous traiterons plus en détail des données dans la section suivante lorsque nous aborderons le sujet du processus ETC.

Suite à l'analyse du processus d'évaluation de la performance des joueurs, l'identification des sources de données et l'analyse des données disponibles, nous pouvons construire le modèle multidimensionnel.

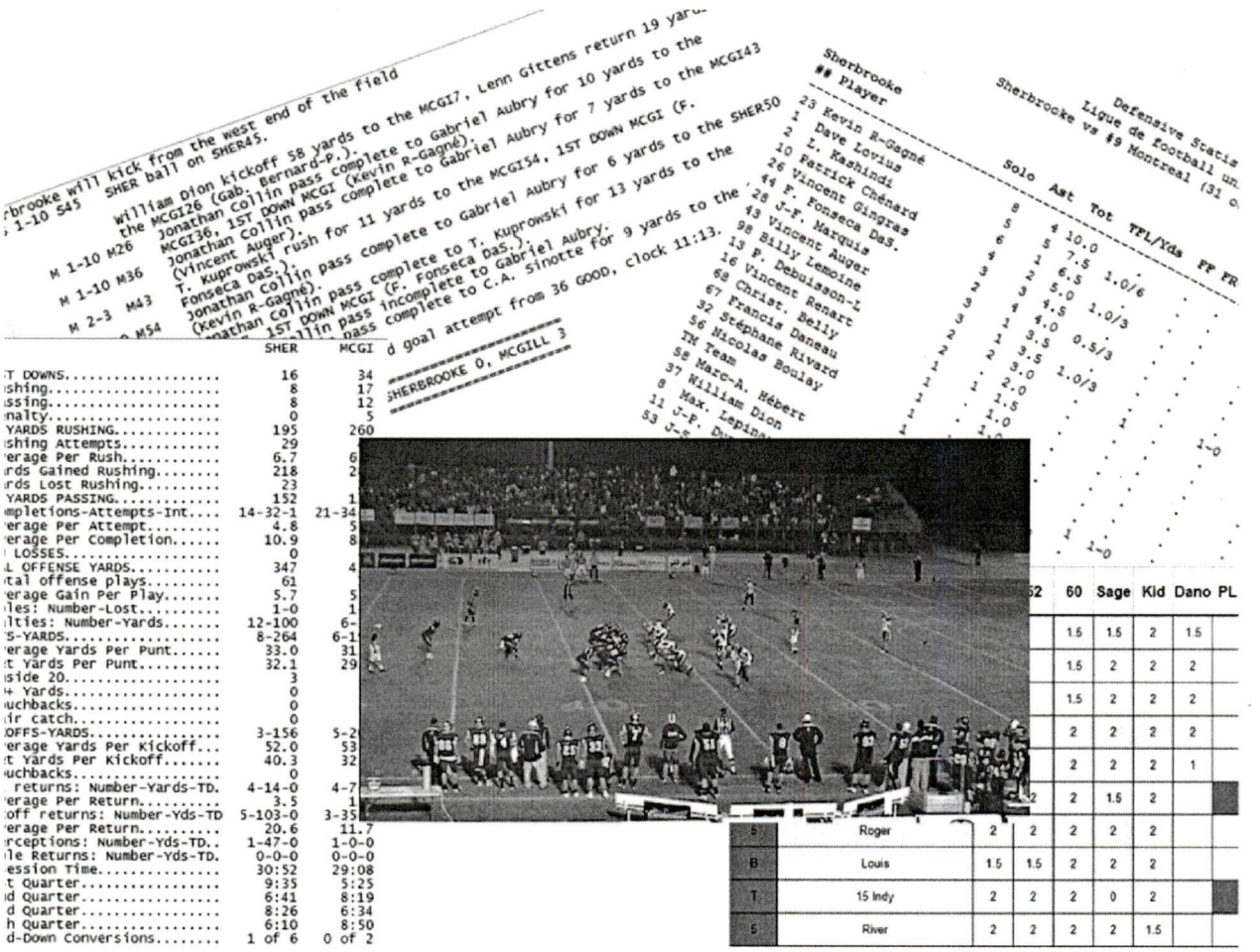


Figure 13 Exemple de données non structurées utilisées par le Vert et Or

4.2 Modélisation multidimensionnelle

Premièrement, nous présentons les tables de faits et les dimensions que nous avons déduites à partir de l'inventaire des données fait précédemment. Ensuite, nous identifions les différents choix de modélisation et justifions ceux-ci.

4.2.1 Faits et dimensions

Comme évoqué dans le Cadre théorique, un modèle multidimensionnel est composé de deux éléments principaux : les dimensions et les faits. Une dimension est un élément qui établit un contexte et un fait est l'élément que l'on mesure dans ce contexte selon Kimball et al. (2002). Tout d'abord, l'inventaire des données a fait ressortir une panoplie de mesures et de dimensions. Nous avons produit une liste des tables de faits et de dimensions ainsi que les attributs associés à chacune d'elles (Annexe 1). Pour en arriver à la création de cette liste de faits, nous avons pris les données par bloc. Nous avons d'abord découpé les données en trois catégories: attaque, défense et unité spéciale. Par la suite, nous avons fait une deuxième segmentation avec les données d'attaque. Celles-ci ont été séparées par type d'attaque, soit passe et course et par notes d'évaluation propre aux joueurs de la ligne offensive. Du côté de la défensive, il n'y a qu'un seul regroupement de données pour l'ensemble des joueurs défensifs. Finalement, les unités spéciales sont séparées en quatre parties : dégagement, retour de botté, botté de placement et botté d'envoi. De plus, nous avons quelques données qui n'entrent pas dans ces catégories, comme le classement des équipes pour la saison, les résultats des parties, la description des jeux et les séquences vidéo. Au niveau des dimensions, l'exercice est de faire ressortir le plus grand nombre de descripteurs retrouvés dans les données afin de contextualiser le plus possible les faits en offrant un

maximum d'axes d'analyse. Ensuite, nous avons simplement regroupé les descripteurs par thème.

4.2.2 Choix modélisation

Dans cette section, nous abordons les différents choix auxquels nous faisons face quant à la modélisation du modèle de données.

Type de modèle

Au niveau du modèle de données, nous avons opté pour un modèle de type constellation qui est en fait un modèle en étoile avec plusieurs tables de faits. Celui-ci a comme avantage d'être plus performant que le modèle en flocon. Le fait d'avoir qu'un seul niveau de dimension simplifie les requêtes pour les utilisateurs et par le fait même améliore la rapidité d'exécution de celles-ci.

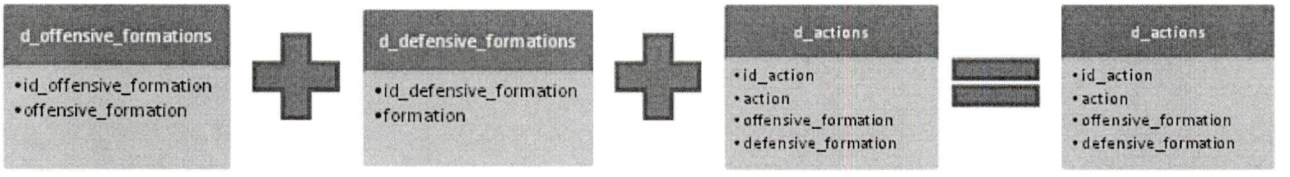
Nomenclature dans l'EDD

Avant d'amorcer la modélisation, nous devons établir certaines règles quant à l'assignation des noms des tables et des attributs. Premièrement, nous décidons de ne pas utiliser les lettres majuscules pour le nom des tables et des attributs. Deuxièmement, le soulignement « _ » remplace les espaces, alors chaque mot est séparé par ce symbole et aucun nom de table ni d'attribut ne contiendra d'espace. Troisièmement, les tables de dimensions et de faits sont

différentiées respectivement par la lettre « d » et « f » au début de leur nom. Finalement, tous les noms des attributs et des tables sont en anglais, puisque la grande majorité des données et des termes du RSEQ et de l'équipe du Vert et Or sont dans cette langue.

Dimension « rebut »

Lorsque les dimensions contiennent peu d'enregistrements, la façon la plus simple de créer une dimension « rebut » est de faire un produit cartésien avec l'ensemble des enregistrements et d'utiliser le résultat de ce produit pour créer cette nouvelle dimension. Au niveau des dimensions « rebuts», les dimensions formations offensives, formations défensives et actions possèdent peu d'attributs ainsi que peu d'enregistrements. Dans ce type de situation, il est préférable de créer une dimension « rebut » (Figure 14). Ceci a un impact sur la performance du modèle puisqu'il permet de réduire le nombre de clés dans la table de faits.



Dimension dégénérée

Au niveau des dimensions dégénérées, la table « d_quarters » n'avait qu'un seul attribut et est liée avec la table de faits « f_play_by_play ». Alors nous avons donc décidé d'incorporer la dimension « d_quarter » à l'intérieur de la table de faits « f_play_by_play » dans le but de réduire le nombre de dimensions dans notre modèle (Figure 15).

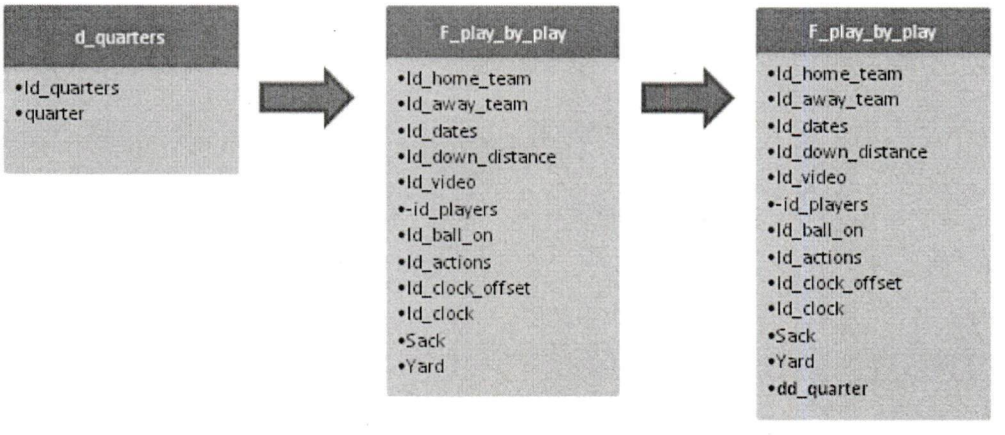


Figure 15 Exemple de dimension dégénérée

Dimension et fait vidéo

Au niveau de la vidéo, notre modèle de données contient une dimension et un fait vidéo. Ceux-ci sont composés d'attributs descripteurs tels que la référence de la vidéo, la taille, le format et la durée. La table de faits vidéo permet à l'utilisateur de faire des requêtes, avec la vidéo comme base d'analyse, en y intégrant plusieurs critères. Par exemple, on peut rechercher les vidéos pour les jeux dans lesquels un joueur X de la ligne offensive a obtenu une note de zéro lorsqu'il reste moins de deux minutes au quatrième quart. De l'autre côté, la dimension vidéo donne un axe d'analyse à la table de fait f_play_by_play. Chaque requête faite sur cette table peut être accompagnée de vidéo. Finalement, selon le type d'analyse voulu, il sera possible pour un membre de l'organisation du Vert et Or d'obtenir de la vidéo, ce qui est très important pour l'évaluation de l'équipe.

Faits et dimensions rejetés

Durant le processus de modélisation, nous avons écarté une dimension et un fait présentés dans l'Annexe 1. Premièrement, le fait pénalité ne sera pas comptabilisé puisque le fichier dans lequel l'information se trouve est un fichier texte qui contient beaucoup de problèmes de qualité de données. Deuxièmement, nous avons retiré la dimension « d_referees» que nous pensions être capables d'intégrer au modèle. L'information est difficile à trouver

et non systématique. Dans ce type de situation, il est préférable de régler ce problème de qualité de données directement à la source plutôt que dans du processus ETC (Kimball et al., 2004). La responsabilité de la correction de ces données est hors de notre portée puisque le RSEQ est propriétaire de celles-ci.

Après avoir géré les exceptions, nous poursuivons le processus de modélisation. Nous faisons les liens entre les tables faits et leurs dimensions associées. Le modèle conceptuel de données complet et le détail du moteur de la base de données choisi se retrouvent dans l'Annexe 2 et l'Annexe 3 la Figure X représente le modèle complet.L'étape suivante est la création de l'EDD par le chargement des données dans le modèle multidimensionnel.

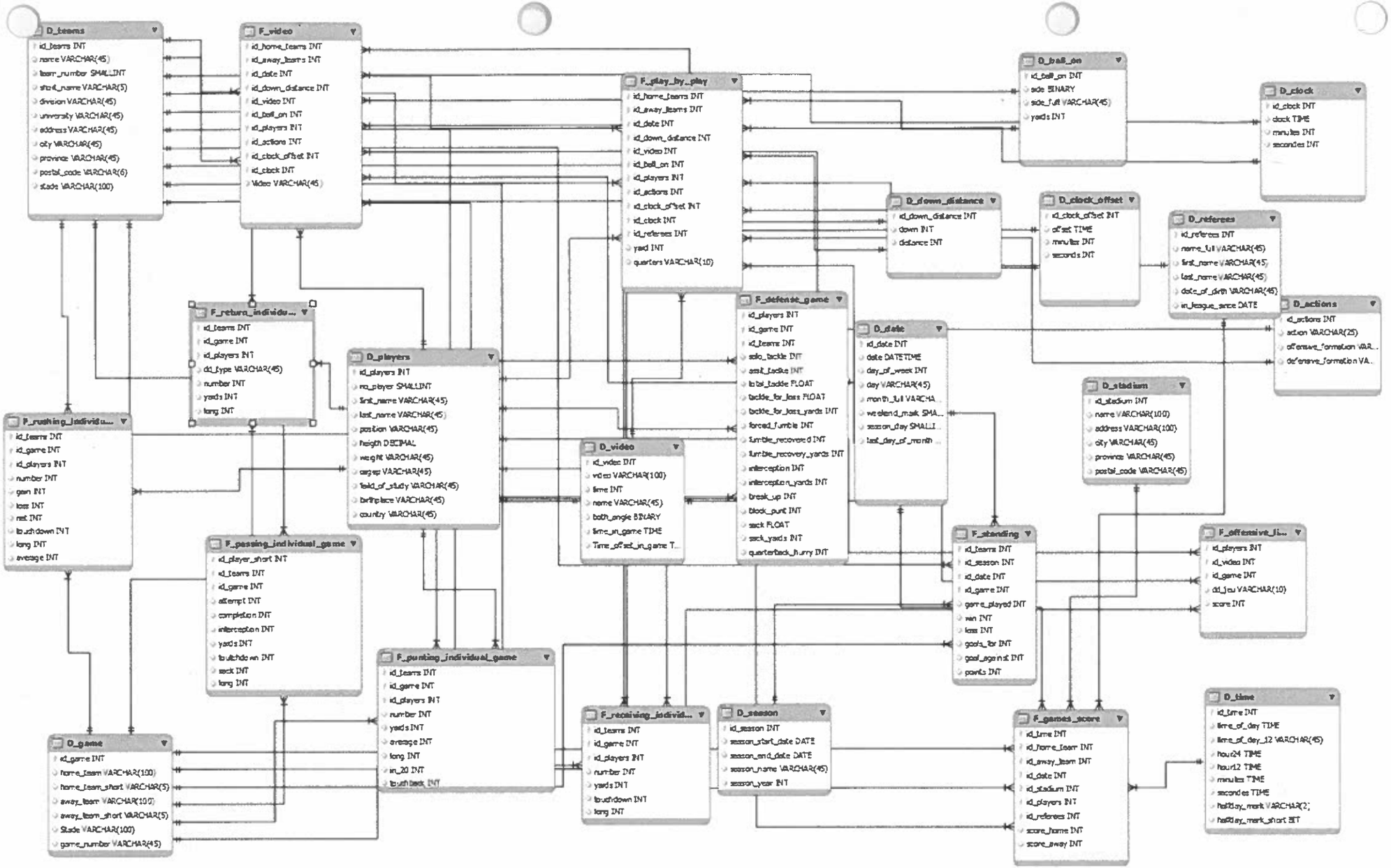


Figure 16 Modèle de données

4.3 Extraction, transformation et chargement

Le processus d'extraction, de transformation et de chargement est une série d'opérations de migration de données qui a pour but de peupler l'EDD à partir des données sources. Au cours du processus, une étape de transformation des données permet de nettoyer et de réconcilier celles-ci afin de les charger dans un format prévu par le modèle multidimensionnel. Dans cette section, nous traitons de la phase du développement du processus ETC qui nous a permis de validé notre modèle de données. Nous abordons le processus ETC mis en place dans le cadre de la recherche et des obstacles à surmonter.

Processus ETC pour le Vert et Or

Dans le but de fournir les données à l'EDD, nous devons construire un processus de chargement de données. L'Annexe 4 présente les routines ETC. Bien que l'élaboration d'un processus ETC ne soit pas l'enjeu principal de notre recherche, celui-ci représente un immense investissement en temps et en expertise. Nous pensons qu'il est important de construire le processus de chargement de données le plus structuré possible afin que le prototype puisse être réutilisable. Dans cette optique, nous avons développé un processus ETC à l'aide de l'outil « Talend Open Studio » (TOS). Le choix de cet outil est basé simplement sur l'expérience que nous avons sur celui-ci.

Pour l'organisation du Vert et Or, l'implémentation d'un processus ETC leur permettra de structurer et d'automatiser le transfert des données vers l'EDD, de corriger les erreurs, de combler les données manquantes et de documenter les problèmes de qualité de données dans une perspective d'ajout de nouvelles données.

4.3.1 Problèmes ETC

Dans la phase de création du processus ETC pour l'organisation du Vert et Or, nous faisons face à plusieurs obstacles.

Sources de données

Tout d'abord, les fichiers plats de type texte provenant du site web du RSEQ sont des fichiers non structurés très différents des données provenant d'une base de données normalisée. En effet, les fichiers n'ont aucun lien entre eux contrairement à une base de données. Il est donc difficile de lier les données d'un fichier à un autre. Bien que ce soit possible de le faire avec « Talend Open Studio » (TOS), notre manque de connaissances et de temps pour les acquérir nous force à parfois utiliser une autre solution. Afin de régler ce problème, nous ajoutons une structure aux données en intégrant une partie des données dans un tableur comme MS Excel. À l'aide du tableur et de

quelques manipulations, nous réussissons à améliorer la structure des données et par le fait même, nous minimisons le nombre d'opérations complexes à faire avec TOS. Ce type d'opération requiert de la programmation en Java dans TOS et ceci dépasse le contexte de ce projet. Nous sommes tout de même conscients que cette option n'est pas la meilleure solution à long terme, puisque celle-ci n'automatise pas complètement le processus ETC.

Nom des joueurs

Un autre problème auquel nous faisons face au niveau de la qualité de données est le manque de constance dans les noms des joueurs. En effet, certains noms de joueurs peuvent être écrits de plusieurs façons à différents endroits. Prenons l'exemple de Jean-Philippe Shoiry, le quart-arrière de l'équipe du Vert et Or. Nous remarquons que son nom est écrit de trois façons différentes soit J-P Shoiry, J P Shoiry et J.P. Shoiry. De plus, les noms sont rarement écrits au complet, comme l'exemple précédant le démontre. Nous pensons qu'il s'agit ici d'une limitation du nombre de caractères de l'application utilisée par la ligue pour cumuler les statistiques. Afin de contourner ce problème, nous utilisons une fonction de « TOS» appelé « Fuzzy match » qui permet, grâce à l'algorithme de Levenshtein, de faire un rapprochement entre les données similaires, comme les trois façons d'écrire J-P shoiry. Ceci nous a permis de régler la majorité, soit environ 90% des problèmes au niveau des noms. Les données restantes ont été corrigées manuellement.

Pour ce qui est des autres données alphanumériques, le processus ETC requiert moins de travail particulier. Nous utilisons simplement les opérateurs standards de manipulation de données de TOS, c'est pourquoi nous passons rapidement sur le sujet. Les données, comme les résultats de la saison 2009, n'ont pas exigé de manipulation ni d'opération complexe de transformation. En effet, celles-ci sont structurées de façon tabulaire, ce qui requiert moins de traitement. Une fois l'ensemble des processus ETC pour les données alphanumériques complété, nous concentrons nos efforts sur l'aspect des données vidéo.

4.4 La vidéo

Dans cette section, nous traitons des différents enjeux de la vidéo dans le cadre de notre projet de recherche. Nous abordons l'importance qu'a la vidéo pour l'organisation et son apport dans la prise de décisions. Par la suite, nous traitons de l'intégration de la vidéo dans les rapports. Finalement, nous analysons la structure et le stockage de la vidéo.

Tout d'abord, nous avons porté attention à un aspect technique de la vidéo utilisée par le Vert et Or. Celle-ci est dans un format « Audio Video Interleve » (AVI). Ce type de fichier créé par Microsoft est un fichier conteneur conçu pour stocker des données audio et vidéo. En plus de permettre de lire un flux vidéo et audio, il supporte aussi les descripteurs (métadonnées, sous-titres,

chapitre). Ce format vidéo est répandu et pris en charge par presque tous les lecteurs multimédias, ce qui veut dire que nous n'aurons pas besoin de le convertir dans un autre format.

4.4.1 Importance de la vidéo

Depuis quelques années, le Vert et Or utilise la vidéo comme information principale pour la prise de décision. En effet, la vidéo leur donne une grande partie de l'information dont ils ont besoin pour l'évaluation de la performance des joueurs, puisqu'il est possible de mesurer chaque action et d'identifier son contexte. Cependant, la dynamique de la vidéo fait que certaines informations sont difficilement assimilables. Il est pratiquement impossible de capter toute l'information incluse dans une vidéo en un visionnement. De plus, une partie de l'information contenue dans la vidéo est difficilement transposable en données alphanumériques. Par exemple, le choix d'un joueur de faire une action plutôt qu'une autre est difficile à transposer objectivement en données alphanumériques, puisqu'il y a plusieurs facteurs qui peuvent influencer cette décision. Il est donc plus utile de visionner la séquence vidéo dans laquelle l'action a été faite afin d'obtenir une vision globale du jeu.

Les statistiques cumulées par la ligue ne représentent pas toujours la véritable performance d'un joueur, puisque la dynamique du sport fait qu'il est difficile d'isoler la performance d'un seul joueur à la fois. Effectivement, l'intrant

d'un joueur est l'extrant d'un ou plusieurs joueurs. Par exemple, pour que la course du demi-offensif (DO) soit bien exécutée, le quart arrière doit lui remettre le ballon et les bloqueurs doivent lui faire un chemin. Alors, si le demi-offensif ne réussit pas à parcourir une grande distance, ce n'est peut-être pas dû à une mauvaise décision de sa part, mais plutôt à un mauvais intrant d'un coéquipier. Alors, un joueur avec de moins bonnes statistiques n'est pas nécessairement moins bon qu'un joueur avec des statistiques plus reluisantes. Cependant, il ne faut pas éliminer l'utilisation des statistiques dans l'évaluation de la performance puisque celles-ci représentent tout de même des faits accomplis par les joueurs. Il faut simplement en faire une utilisation juste et objective. L'intégration des données vient précisément répondre à ce besoin.

L'intégration des données alphanumérique et de la vidéo prend tout son sens lorsque nous combinons les forces de chacun des types de données. Les données alphanumériques jumelées à la vidéo fournissent l'ensemble de l'information nécessaire pour prendre une décision au niveau de l'évaluation de la performance des joueurs. La Table 3 résume les caractéristiques des deux types de données. Les données alphanumériques apportent des statistiques brutes et des faits, par exemple la distance d'une passe et le nombre de sacs du quart fait par un joueur. La vidéo quant à elle apporte la vue d'ensemble et contextualise les données alphanumériques. Il est donc possible de profiter d'une synergie entre les deux types de données.

Type de données	Caractéristiques
Vidéo	- Dynamique - Vue d'ensemble - Permet l'analyse de la technique des joueurs
Donnée alphanumérique	- Information facilement intégrable - Peut être croisée avec d'autres données - Peut être transformée

Table 3 Forces des données vidéos et alphanumériques

Maintenant que nous avons identifié l'importance de la vidéo dans l'organisation du Vert et Or et sa complémentarité aux données alphanumériques, nous descendons au niveau de la structure de stockage de la vidéo.

4.4.2 Stockage de la vidéo

Afin de pouvoir intégrer la vidéo dans les rapports, il faut d'abord établir la façon de stocker celle-ci. Nous optons pour une structure de stockage simple. En effet, nous rassemblons dans un même dossier l'ensemble des vidéos d'une même saison. Partitionner les vidéos par saison permet de simplifier la gestion de celles-ci tout en gardant un minimum de structure. Au

niveau de la nomenclature des vidéos, nous utilisons un format comme celui-ci « 02,Play.028 » (numéro de la partie du calendrier ,Play.numéro du jeu). Ceci permet de dissocier toutes les vidéos à l'intérieur d'une même saison et du même coup, éviter un conflit de fichier. De plus, l'utilisation du numéro de la partie nous offre la possibilité d'ajouter des vidéos.

4.4.3 Intégration de la vidéo dans les tableaux de bord

L'intégration de la vidéo avec les données alphanumériques dans les rapports peut se faire de plusieurs façons. Selon le type d'analyse que l'entraineur veut faire, la vidéo occupera une plus grande place que les données alphanumériques et vice versa. Il est à noter que plusieurs tables de faits ne sont pas liées à une dimension vidéo, ce qui empêche dans ce cas l'intégration de la vidéo dans l'analyse. La vidéo se retrouve seulement à un niveau de granularité jeu par jeu. Ceci pose un problème dans la création de tableaux de bord puisqu'il est impossible de lier la vidéo aux données à granularité différente.

Au niveau de la présentation des données, plusieurs présentations sont possibles selon le type de requête demandée et la personne qui construit et analyse le rapport. Nous proposons de diviser l'écran verticalement pour séparer les données alphanumériques de la vidéo, mais toute autre présentation reste bonne selon les goûts des utilisateurs (Figure 17). Selon

notre segmentation, la section de gauche contient les données alphanumériques présentées selon le format désiré (tableau, graphique). La vidéo est présentée dans la section de droite dans un lecteur multimédia supportant les fonctions généralement trouvées comme jouer, pause, avancement, recule et plein écran. De plus, l'information sur le vidéo se retrouve juste au dessus du lecteur, afin d'identifier facilement la vidéo en cours.

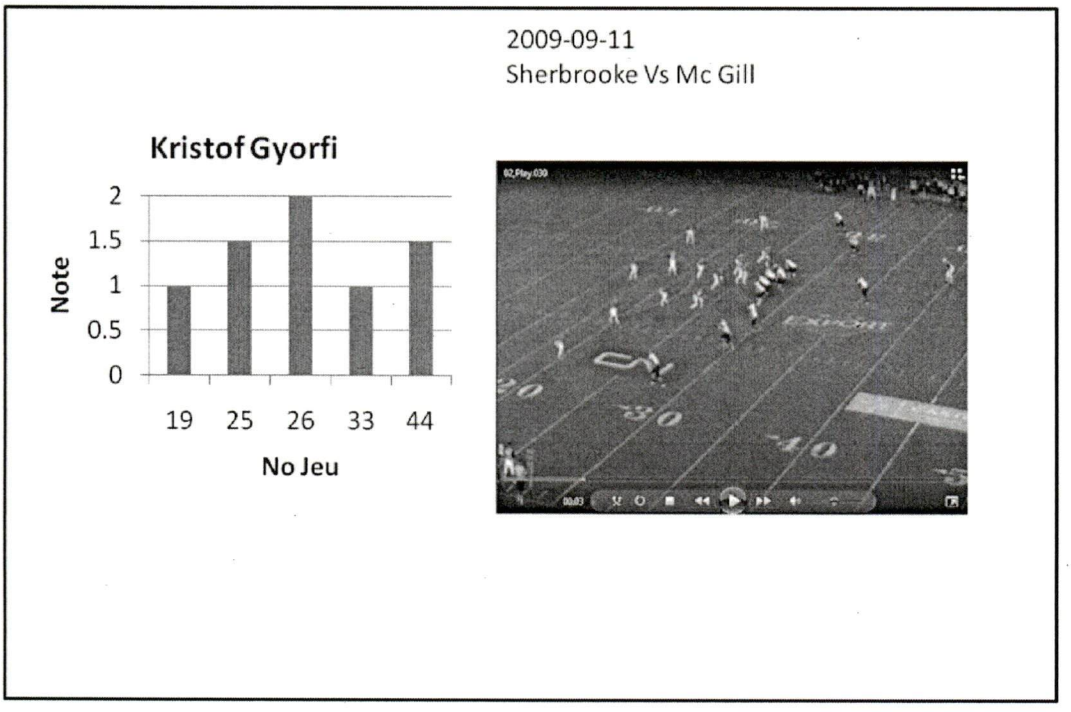


Figure 17 Maquette d'un tableau de bord

Exemple :

La Figure 17 représente un exemple d'analyse fait pour un joueur de la ligne offensive permettant de déterminer la note qu'il a reçue pour les jeux où il était sur le terrain et de lui offrir la possibilité de revoir les erreurs qu'il a commises. L'Annexe 5 présente le code SQL nécessaire à cette requête.

4.5 Description de l'entrepôt de données

Dans le cadre de la recherche, nous avons utilisé seulement un échantillon de données de jeu et vidéo de trois matchs de la saison 2009. Nous communiquons dans cette section diverses statistiques descriptives de l'EDD que nous avons construit.

Premièrement, notre modèle est composé de 26 tables, dont 11 tables de faits et 14 tables de dimensions. En moyenne les tables de faits et tables de dimensions possèdent cinq attributs, en ne comptant pas les clés (Table 4).

Deuxièmement, dans les trois parties de données que nous avons utilisées, on compte en moyenne 126 jeux par parties. Ce qui veut dire que nous devons stocker en moyenne 126 séquences vidéo par partie. Si nous ne comptons pas les séries de fins de saison, le Vert et Or joue huit parties par saison, alors une saison compte en moyenne 1008 vidéos. Chaque vidéo fait en moyenne 99MB, donc une saison requiert en moyenne 97.45 GB (Table 5).

Nous voulions démontrer en quelques chiffres l'EDD que nous avons construit dans le cadre de la recherche. On peut remarquer que la vidéo prend un très grand volume au niveau de l'espace disque. Il faut que le Vert et Or possède assez d'espace disque pour stocker l'ensemble des vidéos.

Description	Nombre de tables	Nombre moyen d'attributs	Nombre maximum d'attributs	Nombre minimum d'attributs
Tables de faits	11	5	15	1
Tables de dimensions	14	5	9	2
Total	25			

Table 4 Nombre de tables dans l'EDD

Nombre de vidéo moyen d'une partie	126
Volume moyen d'une vidéo	99MB
Nombre moyen de vidéos dans une saison	1008
Volume moyen d'une saison en vidéo	97.45GB

Table 5 Détail en nombre et en volume des données vidéo

4.6 Conclusion des Résultats

Ce chapitre a présenté les résultats de la recherche en répondant à notre question de recherche et par le fait même, à la problématique initiale du Vert et Or.

La création et le chargement du modèle de données démontre la pertinence de notre projet. En effet, notre modèle de données intègre des données non structurées, la vidéo, dans un environnement structuré. De plus, comme mentionné dans la Revue de littérature, par définition, les faits que l'on retrouve dans l'EDD, sont de type additif ou semi additif. Par contre, dans notre modèle de données, nous intégrons un fait vidéo qui ne correspond pas à cette règle.

Ainsi, nous pensons que le prototype aurait un impact sur les processus d'évaluation du Vert et Or. En transférant les données de jeu et les vidéos dans un EDD vidéo, nous améliorons la rapidité des analyses. De plus, l'aspect temporel de l'EDD facilitera les analyses d'évolution des joueurs. Finalement, nous pensons que notre prototype valorise davantage la vidéo, qui est importante pour l'équipe, puisque les données de jeu viennent compléter et détailler l'information incluse dans la vidéo. Cependant, la création du tableau fonctionnel de bord ne fait pas partie de l'entendu de notre recherche en raison de la complexité informatique du processus. De plus, nos recherches ne nous ont pas permis de trouver un outil de rapport permettant l'intégration de la vidéo, c'est pourquoi nous présentons une maquette papier.

Chapitre 5 Discussion et conclusion

Ce chapitre vise à faire le point sur notre recherche. Premièrement, nous faisons un retour sur l'objectif de la recherche. Deuxièmement, nous traitons des différentes avenues de recherche possibles. Troisièmement, nous explorons différents milieux où il est possible d'appliquer notre recherche. Quatrièmement, nous proposons des recommandations à l'équipe du Vert et Or. Cinquièmement, nous concluons cette étude avec les limites et des suggestions de recherche future.

5.1 Retour sur l'objectif de recherche

Le but principal de la recherche était de comprendre comment il est possible d'intégrer des données multimédias et des données de jeu dans un modèle de données multidimensionnel. Découlant de cet objectif, nous avons proposé la conception d'un prototype d'un modèle de données multidimensionnel fonctionnel incluant les données de trois parties de la saison 2009. Celui-ci offre plusieurs avantages analytiques à l'équipe du Vert et Or.

Premièrement, l'intégration de la dimension temporelle facilite et accélère les analyses axées sur l'évolution de la performance de l'équipe et permet de

voir l'état d'une situation à un moment figé dans le temps. La dimension temporelle est représentée sous différentes granularités. Dans notre modèle de données, le niveau le plus élevé représente une saison, ce qui permet de suivre l'évolution au travers de celles-ci. Ensuite, on retrouve le niveau des matchs qui permet de faire entre autres des analyses comparatives entre deux matchs ou voir l'évolution de la ligne offensive sur plusieurs matchs. Par la suite, on retrouve les données de jeu par jeu, qui se trouvent au niveau de l'évaluation des tactiques de jeu.

Deuxièmement, afin de construire notre entrepôt de données, nous devions imposer une structure multidimensionnelle en étoile aux données. Comme mentionné dans le Cadre théorique, cette structure facilite la compréhension des données pour les utilisateurs finaux en plus d'accélérer le temps de réponse des requêtes faites par ceux-ci. L'utilisation d'une structure multidimensionnelle permet de cumuler des faits (données numériques) et de les contextualiser en les liant à des dimensions (données descriptives). De cette façon, les données sont déjà structurées dans une perspective d'affaires et sont facilement accessibles.

Troisièmement, l'avantage principal du prototype se situe au niveau de l'intégration de la vidéo et des données de jeu. Comme mentionné précédemment, la dimension et le fait vidéo permettent d'obtenir une vidéo du jeu lorsqu'une requête est faite au niveau d'agrégation jeu par jeu. Lorsqu'une

requête est faite sur une autre table de faits que celle de jeu par jeu, il n'est présentement pas possible de la lier à des vidéos puisque ces données se trouvent seulement au niveau de la partie. Pour lier la vidéo aux autres tables de faits, il faudrait avoir l'ensemble de l'information détaillé au niveau de jeu par jeu.

Quatrièmement, la création du prototype n'a pas pu être complétée lors de la rédaction de ce mémoire. Seule la partie infrastructure est terminée. À la base, notre prototype devait inclure une dimension tableau de bord bâtit à l'aide de l'outil JasperSoft. Cependant, aucun outil de tableau de bord n'intègre la vidéo liée à des données. Alors, nous aurions été obligés de faire des modifications sur un logiciel libre comme JasperSoft qui offre cette flexibilité. Bien qu'il soit facile d'imbriquer une vidéo dans une page web, il est toute fois plus complexe de le faire lorsqu'il faut lier celle-ci à des données. Actuellement, nous rencontrons un problème au niveau de la modification du code XML (Extensible Markup Language) puisqu'il s'agit d'un XML propre à JasperSoft. À ce stade, il est impossible pour nous de remplir complètement cet objectif. Cette partie sera reléguée à un autre projet en raison de la complexité technique qu'il comporte.

En somme, nous pensons que notre prototype a le potentiel d'impacter positivement le processus d'évaluation de la performance des joueurs. Il est impossible présentement de le valider avec les entraineurs du Vert et Or

puisque celui-ci n'est pas complet. Nous pensons qu'il est plus facile de faire des analyses pour les entraineurs et les joueurs puisque le couplage des données de jeu à la vidéo valorise cette dernière. L'intégration des deux types de données fournies une information plus complète que chacune prise séparément. Notre prototype apporte une plus grande flexibilité au processus d'évaluation actuel puisqu'il est possible de filtrer une vidéo selon des données de jeu. Par exemple, un joueur de la ligne offensive qui veut revoir les jeux sur lesquels il a obtenu une note inférieure à 1 peut facilement le faire contrairement au système en place où il faut chercher manuellement le bon jeu. En plus, le prototype s'aligne parfaitement avec les processus d'analyse de performance actuels, puisque ceux-ci sont majoritairement basés sur l'analyse des clips vidéo.

5.2 Avenues de recherche et évolution

Dans cette section, nous voyons comment il est possible de faire évoluer le prototype que nous avons développé pour l'organisation du Vert et Or et proposons par le fait même des avenues de recherche. Tout d'abord, nous abordons l'intégration possible des différents processus d'affaires de l'équipe. En suite, nous proposons l'ajout d'une nouvelle fonctionnalité au système.

5.2.1 Intégration des autres processus d'affaires

Comme mentionné dans le Chapitre quatre, le suivi de la performance des joueurs ne se fait pas simplement sur le terrain. En effet, les entrainements physiques en salle, la nutrition et la réussite scolaire sont au cœur de la performance de l'équipe.

Entrainement physique

Au niveau de l'entrainement physique, l'équipe a un processus de programme d'entrainement en salle pour chaque joueur. Ceux-ci ont une routine d'entrainement personnalisée. Chaque joueur note le nombre de répétitions faites et le poids utilisé pour chacun des exercices. L'entraineur fait une évaluation de l'évolution des routines pour chacun des joueurs et apporte des modifications au besoin. Le processus actuel permet difficilement de suivre l'évolution dans le temps puisque l'information est notée sur papier. L'entraineur base ses décisions sur son expérience et son intuition pour faire évoluer les routines d'entrainement. L'intégration des données d'entrainement en salle permettrait aux entraineurs et aux joueurs de mieux suivre l'évolution de leur condition physique. En outre, il serait aussi possible de faire des analyses entre la performance sur le terrain et l'évolution de la condition physique des joueurs.

Il serait possible de créer un processus formalisé permettant de saisir l'information dans une base de données plutôt que sur papier. Cette base de données pourrait même être accessible en ligne pour que les joueurs puissent entrer eux-mêmes les données d'entrainement à distance. De plus, l'utilisation d'une base de données pour stocker les données d'entrainement faciliterait la création du processus ETC puisque l'information serait déjà stockée de manière structurée.

Pour ce faire, nous devons apporter des modifications au modèle de données. En effet, l'ajout d'au moins une table de fait pour compiler les données d'entrainement est essentiel. De plus, l'ajout de tables de dimensions pour les exercices et les programmes d'entrainement est nécessaire. Les dimensions « joueurs » et « date » sont déjà existantes dans le modèle, alors nous n'avons pas besoin de les recréer. L'intégration des données d'entrainement au modèle de données actuel permettrait premièrement de centraliser l'information en une seule place. Deuxièmement, elle permettrait aussi aux utilisateurs de faire des analyses croisées entre la performance sur le jeu et l'entrainement physique à l'aide de la fonctionnalité « drill across ».

Nutrition

Présentement, il n'existe aucun processus de suivi de la nutrition des joueurs. Cependant, lors de nos rencontres avec l'entraineur de la ligne

offensive, il nous a mentionné qu'il aimerait instaurer un programme de nutrition dans les prochaines années. La nutrition est un aspect qui est très important aux yeux de l'entraineur puisqu'il considère qu'elle a un impact direct sur la forme physique et la performance sur le terrain.

L'ajout de ces données dans l'EDD viendrait compléter les données de jeu et d'entrainement physique en apportant une nouvelle perspective d'analyse sur l'évolution de la performance des joueurs. Pour ce faire, il faudrait ajouter de nouvelles tables de dimensions et tables de faits au modèle de données actuel. En effet, il faudrait ajouter les dimensions telles que aliments et programmes de nutrition. Du côté des faits, il faudra au moins une table de faits pour cumuler les portions consommées.

Performance scolaire

Pour ce qui est de la performance scolaire, il n'existe encore aucun processus formel pour faire le suivi de la réussite scolaire des joueurs de l'équipe. De plus, l'Université de Sherbrooke n'impose aucun règlement clair à ce niveau. Par contre, l'entraineur nous dit qu'il est possible que l'Université prenne position et impose un processus de suivi de la performance scolaire dans les années à venir puisque le but premier est que les joueurs obtiennent leur diplôme. Dans le cas où il devient obligatoire pour l'équipe de se doter d'un processus d'évaluation de la performance scolaire, il sera possible d'incorporer

ces données à l'EDD. En plus de permettre le suivi des joueurs dans leur réussite scolaire, il sera possible de faire le parallèle entre les performances scolaires et celles sur le terrain afin de déceler des tendances dans l'attitude des joueurs et mieux les encadrer.

Pour faire l'ajout de la performance scolaire dans l'EDD, il faudrait faire plusieurs modifications. En effet, il faudrait ajouter des dimensions comme cours, programme et concentration, en plus d'au moins une table de faits pour cumuler les notes. De plus, la sensibilité des données pourrait poser problème, puisqu'il s'agit de données confidentielles. Il est possible de sécuriser les données et de limiter leur accès à certains usagers. Ceci dit, il y a toujours un risque lorsque l'on joue avec des données confidentielles. Aussi, le sort réservé à ce type de données lorsqu'un joueur quitte l'équipe est un enjeu sur lequel l'équipe devra se positionner. Si l'université et l'équipe prennent cette route dans les prochaines années, il faudra qu'ils implémentent un processus sécuritaire.

5.2.2 Superposition de texte à la vidéo

Dans cette section, nous élaborons sur une avenue de recherche traitant d'une fonctionnalité de superposition de texte dynamique qui pourrait être ajoutée au prototype. Nous allons d'abord expliquer cette fonctionnalité. Ensuite, nous expliquons de quelles façons elle peut être mise en œuvre.

La fonctionnalité de superposition de texte dynamique à la vidéo est une fonctionnalité qui pourrait enrichir l'expérience de visionnement de la vidéo en superposant du texte descriptif provenant des dimensions à la vidéo. La Figure 18 montre un exemple de superposition d'un texte sur la vidéo, soit le « down and distance » dans le coin supérieur gauche de la vidéo. Ceci permet à la personne qui regarde une vidéo en format plein écran de voir certaines informations lier à celle-ci. Lors de nos rencontres avec le Vert et Or, nous avons remarqué que l'application XOs permet de faire une fonctionnalité similaire lors du visionnement de vidéo et l'entraineur l'apprécie beaucoup. Actuellement, il utilise cette fonctionnalité pour mettre l'information du « down and distance » sur la vidéo puisque cette information est importante pour lui et les joueurs.

Figure 18 Exemple de superposition de texte à la vidéo

Il est possible de faire cette superposition de deux façons. La première technique est de préalablement définir l'information que l'on voudrait voir et de la figer sur la vidéo. Ceci peut être fait avec des logiciels spécialisés de création et de montage vidéo. Cette technique a comme avantage d'être simple, mais elle n'est pas flexible puisque l'information n'est pas modifiable. La deuxième technique, en est une plus flexible et permet de choisir l'information que l'on veut superposer à la vidéo. L'idée derrière cette technique est d'utiliser les attributs des dimensions pour générer l'information à superposer. Ceci apporte une grande flexibilité, puisque l'on peut choisir l'information que l'on veut voir sur la vidéo selon le but de l'analyse. Que ce soit le « down and distance », le stade où se déroule le match ou toute autre dimension liée à la vidéo dans le

modèle de données, il est possible d'utiliser cette information pour la superposition. Comme l'information à superposer est stockée présentement dans l'EDD, il est possible d'intégrer cette fonctionnalité sans avoir de perte au niveau de l'information. Le langage SMIL[8] (Synchronized Multimedia Integration Language), qui est une forme de XML, peut être utilisé pour faire de la superposition de texte sur la vidéo. Cependant, il faudrait faire des recherches supplémentaires afin de voir comment il est possible d'intégrer le SMIL dans la création de tableaux de bord.

5.2.3 Gestion des angles de vues

Dans le cadre de notre recherche, nous n'avons pas considéré la gestion des angles de vues, puisque dans le cas du Vert et Or, les séquences vidéo sont créées de façon à montrer chaque jeu avec les deux angles de vue un à la suite de l'autre. Par contre, les équipes professionnelles ont accès à plusieurs angles de vue pour un même jeu. Ceci devient très intéressant puisque les multiples angles de vues permettent aux entraineurs d'analyser le jeu de façon plus complète. En plus, de pouvoir filtrer par angle de vue, il serait profitable de pouvoir choisir plusieurs angles de vues pour un visionnement en parallèle à l'aide d'une fenêtre fractionnée. Selon l'entraineur du Vert et Or, il est très difficile d'analyser l'ensemble du jeu avec un seul angle de vue. L'ajout d'une dimension d'angle de vue permettrait aux utilisateurs de filtrer les vidéos selon

[8] http://www.w3.org/AudioVideo/

l'angle de vue désiré et offrirait une plus grande flexibilité lors des analyses. Cependant, il serait intéressant d'effectuer une recherche sur le sujet afin de déterminer les enjeux réels d'une telle fonctionnalité. Cette recherche devrait être faite dans un département d'informatique puisqu'il s'agit d'un sujet très technique.

5.2.4 Conclusion avenues de recherche et évolution

Bien que les différentes évolutions proposées dans cette section soient difficiles d'accès pour des raisons de manque d'expertise et de ressource financière pour équipe de football universitaire, elle pourrait procurer un avantage compétitif face aux autres équipes de la ligue. Pour sa part, une équipe sportive professionnelle aurait les moyens de le faire. Cependant, il faudrait faire davantage de recherche au niveau de l'interaction et de l'intégration des technologies. Ceci pourrait être fait sous un type de recherche en « design science » afin d'aller chercher une problématique réelle du monde sportif et d'obtenir les besoins et enjeux directement à la source.

5.3 Application des connaissances

Dans cette section, nous abordons les différentes applications pratiques de notre recherche. Tout d'abord, nous présentons les applications possibles de notre modèle dans d'autres ligues de football d'Amérique du Nord. Ensuite,

nous voyons comment il est possible d'exporter notre modèle dans d'autres sports.

5.3.1 Football

Nous avons construit notre modèle de données pour qu'il convienne aux besoins du Vert et Or et aux spécificités du football universitaire canadien. Cependant, il est possible de transposer celui-ci dans d'autres ligues de football comme la ligue canadienne de football (LCF), la ligue nationale de football (LNF) et la « National Collegiate Athletic Association » (NCAA).

D'entrée de jeu, quelques modifications s'imposent au niveau des règlements qui sont légèrement différents d'une ligue à l'autre. Par exemple, dans la LNF, il y a quatre essais plutôt que trois dans le cas du Vert et Or et de la LCF. Ceci implique qu'il faudrait modifier les dimensions impactées par les changements de règlement.

Ensuite, il est nécessaire d'ajouter l'aspect financier afin de transposer notre modèle à des ligues professionnelles. En effet, contrairement aux ligues universitaires, les équipes professionnelles doivent gérer un aspect supplémentaire, la masse salariale. Il est donc primordial de considérer le salaire et le bonus des joueurs dans l'évaluation de la performance de ceux-ci

puisque les directeurs généraux veulent obtenir le meilleur joueur possible au prix le plus bas.

Par la suite, il est certain que les ligues et les équipes professionnelles cumulent beaucoup plus de données que la ligue de football universitaire canadienne. En effet, si l'on regarde sur les sites web des ligues professionnelles, on remarque qu'elles cumulent une très grande quantité de données statistiques sur les équipes et les joueurs. De plus, on peut aisément supposer que chaque équipe cumule des données supplémentaires à l'interne. En effet, les équipes professionnelles ont un plus grand nombre d'entraineurs, de recruteurs, de coordonnateurs et de gestionnaires. Ceux-ci sont tous des consommateurs de données et chacun a besoin de données spécifiques à sa tâche. Afin de pouvoir intégrer l'ensemble des données, il faut créer de nouvelles mesures et possiblement de nouvelles tables de faits. Une analyse des données est tout de même essentielle pour être certain de ne rien oublier.

Finalement, il est possible de transposer le modèle de données que nous avons développé pour l'équipe du Vert et Or à d'autres ligues de football. Un nombre minimal de modifications est requis au modèle de données pour permettre les mêmes analyses que celles pour le Vert et Or. Dans le cas où l'on voudrait incorporer la réalité financière des équipes professionnelles, un grand nombre d'ajouts s'impose puisque présentement rien n'est fait pour gérer ce sujet. Une analyse plus approfondie nous donnerait plus de détails.

5.3.2 Autres sports

Comme mentionné dans le Chapitre un, plusieurs équipes de différents sports tentent de trouver une façon de mesurer objectivement la performance de l'équipe et des joueurs. Il est intéressant de voir comment nous pouvons faire évoluer notre modèle de données dans d'autres sports que le football. Les sports professionnels populaires en Amérique du Nord sont le baseball, le basket-ball et le hockey. Dans cette section, nous survolons l'application possible de notre modèle de données et de l'impact de la différence entre les sports sur notre modèle, en se concentrant sur le hockey à titre d'exemple.

Tout d'abord, les sports d'équipe comme le football, le hockey et le basketball sont considérés comme des sports dynamiques. C'est l'ensemble des actions des joueurs qui fait avancer le jeu. Par exemple, au hockey, la passe d'un joueur permet à celui-ci de faire une action qui elle-même permet à un autre joueur d'en faire une à son tour. Chaque sport a ses spécificités. Au football, le dynamisme ne se passe pas qu'au porteur de ballon, mais beaucoup au niveau des blocs. D'autre part, au hockey, le dynamisme s'illustre beaucoup plus avec les passes, donc aux porteurs de la rondelle. Il est évident que le sport a un impact sur le modèle de données. Par contre, dans tous les cas, la vidéo permet l'analyse globale et dynamique du jeu en lien avec les statistiques des joueurs, ce qui rend le cœur de notre projet transportable d'un sport à un autre. Il est évident que les règlements du sport exigent d'apporter des

modifications au niveau des dimensions telles que temps, périodes et actions. En plus, certaines dimensions comme « Ball on » et « Down and Distance » sont spécifiques au football et devront être remplacées par des dimensions propres au sport.

5.3.3 Conclusion de l'application des connaissances

Finalement, bien que notre modèle de données ait été construit en respectant les spécificités du football canadien universitaire et plus particulièrement ceux de l'équipe du Vert et Or, nous voyons qu'il est possible de le transposer vers d'autres ligues de football et d'autres sports. Nous avons identifié certaines modifications qui nous apparaissent essentielles dans la transposition de notre modèle de données. Nous avons aussi établi que le cœur de notre projet, soit l'utilisation de la vidéo dans l'évaluation de la performance dans le sport, est aussi transposable à d'autres sports dynamiques comme le hockey.

5.4 Limites de la recherche et recommandations

Dans cette section, nous présentons les différentes limites de notre recherche et par le fait même, nous mettons en perspective les résultats communiqués dans le Chapitre quatre. Ensuite, nous traitons des recommandations que nous faisons à l'équipe du Vert et Or et à la ligue

canadienne de football universitaire au niveau de la collecte et de la diffusion des données.

5.4.1 Limites de qualité des données

Premièrement, la qualité des données est un élément qui limite la recherche. Comme mentionné dans le chapitre Résultats, nous avons rencontré plusieurs problèmes au niveau de la qualité des données. Les données entrées en texte libre et le nom des joueurs en sont de bons exemples. Afin que l'on puisse quand même compléter notre projet, nous avons corrigé manuellement les données, ce qui ouvre la porte à des erreurs humaines.

5.4.2 Recommandations sur la qualité des données

Le RSEQ est chargé de collecter et de diffuser les données de jeu sur son site Web. Par contre, la façon dont les données sont diffusées n'encourage en aucun point leur exploitation et favorise plutôt la création d'un énorme cimetière de données. En effet, les données diffusées sur le Web sont dans un format de fichier positionnel ou de texte libre, très difficile à exploiter. De plus, on y retrouve de nombreuses erreurs. Les noms ne sont pas toujours écrits de la même façon et souvent en format abrégé, par exemple J-P Shoiry plutôt que Jean-Philippe Shoiry (voir Chapitre quatre). Tout ça ajoute une grande complexité à l'automatisation des routines ETL et par le fait même érige une

barrière pour les équipes qui n'ont pas de ressources expérimentées en informatique et qui voudraient exploiter ces données

5.4.3 Limite de la précision des données

En cours de projet, nous avons fait face à des situations où nous voulions construire le modèle de données d'une certaine façon et ne pouvions pas puisque les données n'étaient pas disponibles. Par exemple, avec les données mises à notre disposition, nous ne pouvons pas savoir quels joueurs sont présents sur le terrain à chaque jeu, sauf s'il touche au ballon. Dans le cadre de notre projet, nous les avons simplement laissées de côté puisqu'il est impossible de les utiliser. Ceci restreint énormément les analyses possibles lorsque l'on veut déterminer l'efficacité d'un joueur sur un jeu.

5.4.4 Recommandations sur la précision des données

Il est difficile de déterminer qui entre les deux acteurs au niveau des données, le RSEQ ou le Vert et Or, devrait cumuler la donnée de présence des joueurs sur le jeu. Ce qui est certain, c'est qu'il y a une grande valeur ajoutée à le faire. Que cela soit fait au cours de la partie ou à postériori en regardant la vidéo, comme les données de la ligne offensive, cette information est d'une grande importance si l'on veut mesurer de façon plus précise l'apport de chaque joueur, même ceux qui ne font pas l'action principale du jeu. Dans l'éventualité où l'information sera disponible, il faudrait modifier la table de faits

« play by play » pour ajouter la clé primaire de tous les joueurs présents sur le terrain à chaque jeu.

5.4.5 Limite de la granularité des données

Comme abordé dans les Résultats, le manque de données à la granularité jeu par jeu vient réduire le nombre possible d'analyses liées à des séquences vidéo de jeu. En effet, la majorité des données d'un match sont calculées au niveau de granularité du match, ce qui nous empêche de pouvoir les lier à la vidéo. Par exemple, l'ensemble des statistiques défensives disponibles sur le site du RSEQ sont cumulées par partie ce qui rend impossible le croisement des données défensives avec la vidéo.

5.4.6 Recommandations sur la granularité des données

Encore une fois, il est difficile de déterminer quel acteur devrait être responsable de cette collecte de données supplémentaires. Il est évident que l'équipe du Vert et Or peut se créer un avantage compétitif face aux autres équipes si elle décide de prendre en charge la collecte des données, puisqu'elle serait la seule équipe à posséder ces données. Bien sûr, il ne faut pas seulement détenir les données, il faut aussi s'en servir pour faire des analyses si l'on veut obtenir l'avantage sur les autres équipes. Nous pensons qu'il est possible de dédier la tâche à des membres du personnel ou à des joueurs qui ne sont pas en uniforme durant la partie. D'un autre côté, le RSEQ

pourrait se charger de cette tâche puisqu'il cumule déjà des données. Il faudrait juste les cumuler avec plus de détails. Par la suite, il faudra ajouter plusieurs tables de faits à une granularité jeu par jeu tel que « defense ».

5.4.7 Limite performance des joueurs

Bien que notre modèle de données puisse améliorer le processus d'évaluation de la performance des joueurs, il est impossible de garantir qu'à lui seul il engendrera le succès de l'équipe. En effet, la technologie peut venir en aide à la prise de décision pour le choix d'une stratégie de jeu, mais au final, c'est un humain qui exécute l'action et celui-ci n'est pas infaillible à l'erreur. Cet outil pourra certainement leur apporter un avantage compétitif, mais l'équipe du Vert et Or ne doit pas surestimer la technologie et plutôt l'imbriquer à une gestion saine et complète de l'équipe.

5.5 Conclusion du mémoire

En somme, cette étude a permis de comprendre comment il est possible d'intégrer des données multimédias et des indicateurs de performance dans un modèle de données multidimensionnel afin de soutenir des analyses de performance liées à des clips vidéo. Pour ce faire, nous avons collaboré avec l'équipe de football de l'Université de Sherbrooke, le Vert et Or, dans le but de fonder notre recherche sur un cas et une problématique réels. Nous avons créé un modèle de données multidimensionnel vidéo à l'aide des connaissances que

nous avons puisées dans le cadre théorique. Finalement, avec les résultats obtenus, nous sommes en mesure de contribuer à notre tour aux connaissances techniques et scientifiques sur lesquelles d'autres recherches pourront être fondées.

Références

Arigon, A. M, Tchounikine A., and Miquel M (2006), *Handling multiple points of view in a multimedia data warehouse*, ACM Trans. Multimedia Comput. Commun. *Appl.* 2, n°. 3 : 199-218.

Ballard, C., Farrel, D.M., Gupta, A., Mazuela, C., Vohnik, S. (2008) *Dimensional Modeling: In a Business Intelligence Environment*, IBM Redbooks

Bhandari, I., et Parker, J. (1997), *Advanced Scout: Data Mining and Knowledge Discovery in NBA Data*, Data Mining and Knowledge Discovery 1, n°. 1 : 121-125.

Bonifati, A., Cattaneo, F., Ceri, S., Fuggetta, A., and Paraboschi, S. (2001), *Designing data marts for data warehouses*, ACM Trans. Softw. Eng. Methodol. 10, n°. 4 : 452-483.

Cabibbo, L., and Torlone, R. (1998) A logical approach to multidimensional databases, Proceedings of the 6th International Conference on Extending Database Technology: Advances in Database Technology

Canadian Football League www.CFL.com

Carrer, M., Ligresti, L., Ahanger, G., Little, T. D. C (1997)., *An Annotation Engine for Supporting Video Database Population*, Multimedia Tools Appl. 5, n°. 3 : 233-258.

Chaudhuri, S. And, Dayal, I. (1997), *An overview of data warehousing and OLAP technology*, SIGMOD *Rec.* 26, n°. 1 : 65-74.

Davenport, T., Harris, J., and Morison, R., (2010) *Analytics at Work* , Harvard Business Press.

Davenport, T., Harris, J (2007) *Competing on analytics,* Harvard Business Press.

Decleir, C., Kouloumdjian,J., and Hacid, S.(1999) A Database Approach for Modeling and Querying Video Data, *Proceedings of the 15th International Conference on Data Engineering.*

Giorgini, P., Rizzi, S., Garzetti, M. (2005) *Goal-Oriented Requirement Analysis for Data Warehouse Design*, Proceedings of the 8th ACM international workshop on Data warehousing and OLAP

Han, J.,Kamber, M. (2000), *Data Mining: Concepts and Techniques*, 1er edition.

Hevner, A. R., March, S.T., Park, J. (2004) *Design Science in Information Systems Research*, MIS Quaterly 75-105

Hevner, A.R (2007) *A Three Cycle View of Design Science Reasearch*, Scandinavian Journal of Information Systems

Inmon, W. (2005) *Building the data warehouse*, Wiley.

Inmon, W., Strauss,D., Neushloss,D.(2008) *DW 2.0,* Morgan Kaufmann.

Kemper. H (2000), Conceptual Architecture of DataWarehouses - A Transformation-Oriented View, *AMCIS 2000 Proceedings.* Paper 180.

Kim, H. H., and Park,S. S (2003) *Building a web-enabled multimedia data warehouse*, Proceedings of the 2nd international conference on Human.society@internet.

Kimball, R., and Caserta,J. (2004) *The data warehouse ETL toolkit,* Wiley.

Kimball, R., and Ross,M. (2002) *The data warehouse toolkit,* Wiley.

Klein, J., and Reif, K. (2001) *The hockey compendium* ,M&S.

Lewis, M. (2003) *Moneyball : The Art of Winning an Unfair Game*, W. W. Norton & Company; 1st edition

National Football League (NFL) www.NFL.com

Negash, S., Gray,P. (2004) *Business Intelligence*, Communication of the Association for Information Systems, Volume 13, 177-195.

Nielsen, J. (1993). Usability engineering. Boston, MA, Academic Press Harcourt Brace & Company.

Réseau du Sport Étudiant du Québec www.sportetudiant.com

Romero, O., Abello, A. (2009) *A survey of Multidemensional Modeling Methodologies*, International Journal of Data Warehousing and Minning.

Schneider,S. (2008) *A general model for the design of data warehouses*, International Journal of Production Economics 112, n°. 1 : 309-325.

SMIL http://www.w3.org/AudioVideo/

Smith, T., Davenport, G. (1993) *The stratification system a design environment for random access video*, *Network and Operating System Support for Digital Audio and Video*, 250-261.

Sport Motion www.sportmotion.com

Turban. E., Aronson, J. E., Sharda, R., King, D. (2007) *Business intelligence* Pearson Prentice Hall.

Udayan, B.(2011) *Is cloud computing losing its value proposition?,* cloud computing jounal

Vert et Or Football www.udes.ca/football

Wookey, L., Yongkyu, K., Yunsun, L., and Jinho, K. (1999) *Developing multimedia data warehouse of education on-demand systems*, Proceedings of the IEEE Region 10 Conference, 945 - 945.

XoS digital www.xosdigital.com

You, J., Dillon, T., Liu, J., and Pissaloux, E. (2001) *On hierarchical multimedia information retrieval, Proceedings of the 2001 international Conference on Image Processing* 729-732.

Zaïane, O.R., Han, J., Li, Z-H, Hou, J (1998) *Mining multimedia data*, Proceedings of the 1998 conference of the Centre for Advanced Studies on Collaborative research

Annexe 1 Liste des tables et champs

d_actions	d_ball_on	d_clock	d_clock_offset	d_dates	d_down_distance	d_games
id_action	id_ball_on	id_clock	id_clock_offset	id_date	id_down_distance	id_game
action	side	clock	offset	date	down	home_team
offensive_formation	side_full	minutes	minutes	day_of_week	distance	home_team_short
defensive_formation	yards	secondes	secondes	day		away_team
				month_full		away_team_short
				weekend_mark		Stade
				season_day		game_number
				last_day_of_month		

d_players	d_referees	d_seasons	d_stadiums	d_teams	d_times	d_videos
id_player	id_referees	id_season	id_stadium	id_teams	id_time	id_video
no_player	name_full	season_start_date	name	name	time_of_day	video
first_name	first_name	season_end_date	address	team_number	time_of_day_12	time
last_name	last_name	season_name	city	short_name	hour24	name
position	date_of_dirth	season_year	province	division	hour12	both_angle
heigth	in_league_since		postal_code	university	minutes	time_in_game
weight				address	seconds	Time_offset_in_game
cegep				city	halfday_mark	
feild_of_study				province	halfday_mark_short	
birthplace				postal_code		
country				stade		

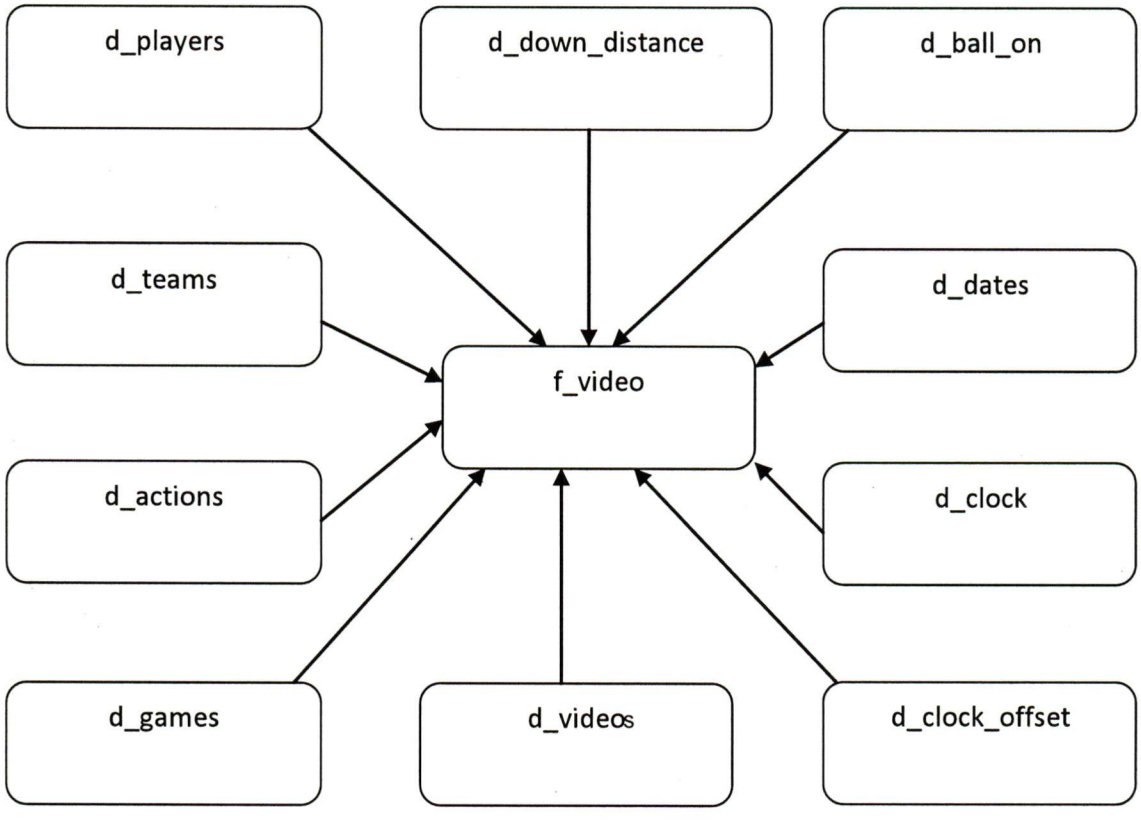
f_penalities	f_defense_game	f_game_score	f_offensive_line_play	f_passing_individual_game	f_play_by_play
id_home_teams	id_games	id_dates	id_players	id_games	id_home_teams
id_away_teams	id_teams	id_time	id_video	id_teams	id_away_teams
id_dates	id_players	id_stadium	id_game	id_players	id_dates
id_down_distance	solo_tackle	id_home_team	id_date	id_dates	id_down_distance
id_video	assit_taclke	id_away_team	dd_jeux	attempt	id_video
id_ball_on	total_tackle	home_score	score	completion	id_ball_on
id_players	tackle_for_loss	away_score		interception	id_players
id_actions	tackle_for_loss_yards			yards	id_actions
id_clock_offset	forced_fumble			touchdown	id_clock_offset
id_clock	fumble_recovered			long	id_clock
id_referees	fumble_recovery_yards				sack
dd_penality_type	interception				yard
yards	interception_yards				quarters
	break_up				
	block_punt				
	sack				
	sack_yards				
	quarterback_hurry				

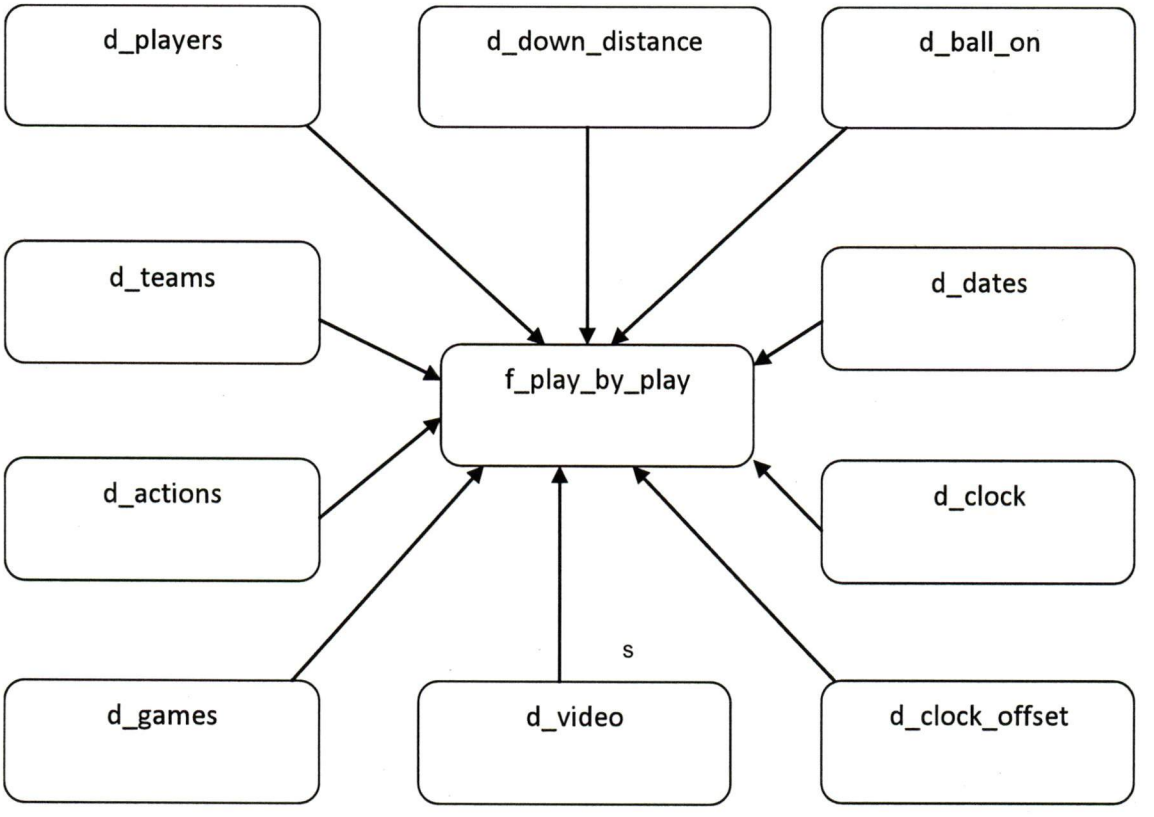
f_punting_individual_game	f_receiving_individual_game	f_return_individual_game	f_rushing_individual_game	f_standing	f_video
id_game	id_game	id_player_short	id_game	id_teams	id_home_teams
id_teams	id_teams	id_teams	id_teams	id_season	id_away_teams
id_player	id_player	id_game	id_player_short	id_date	id_date
id_date	Id_date	id_date	id_date	id_game	id_down_distance
number	receiving_number	dd_return_type	number	game_played	id_video
yards	yards	return_number	gain	win	id_ball_on
average	touchdown	return_yards	loss	loss	id_players
long	longess_receiving	return_long	net	goals_win	id_actions
in_20			touchdown	goals_loss	id_clock_offset
touchback			long	points	id_clock
			average		Video

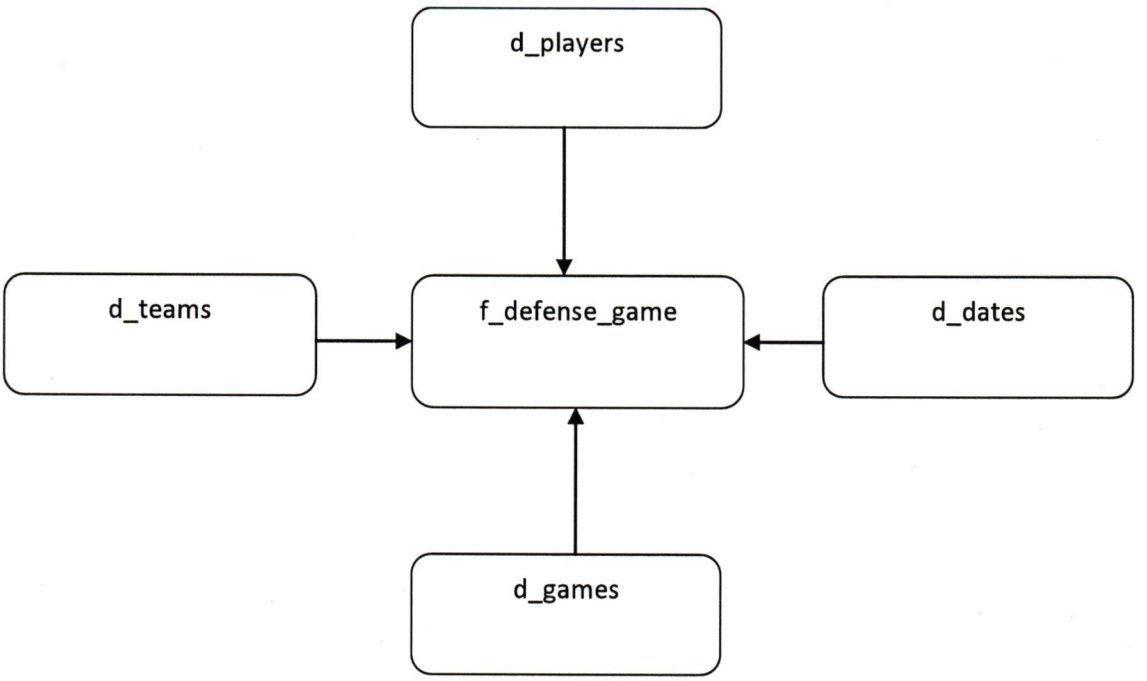
Table A1 - 1 Liste des attributs par table

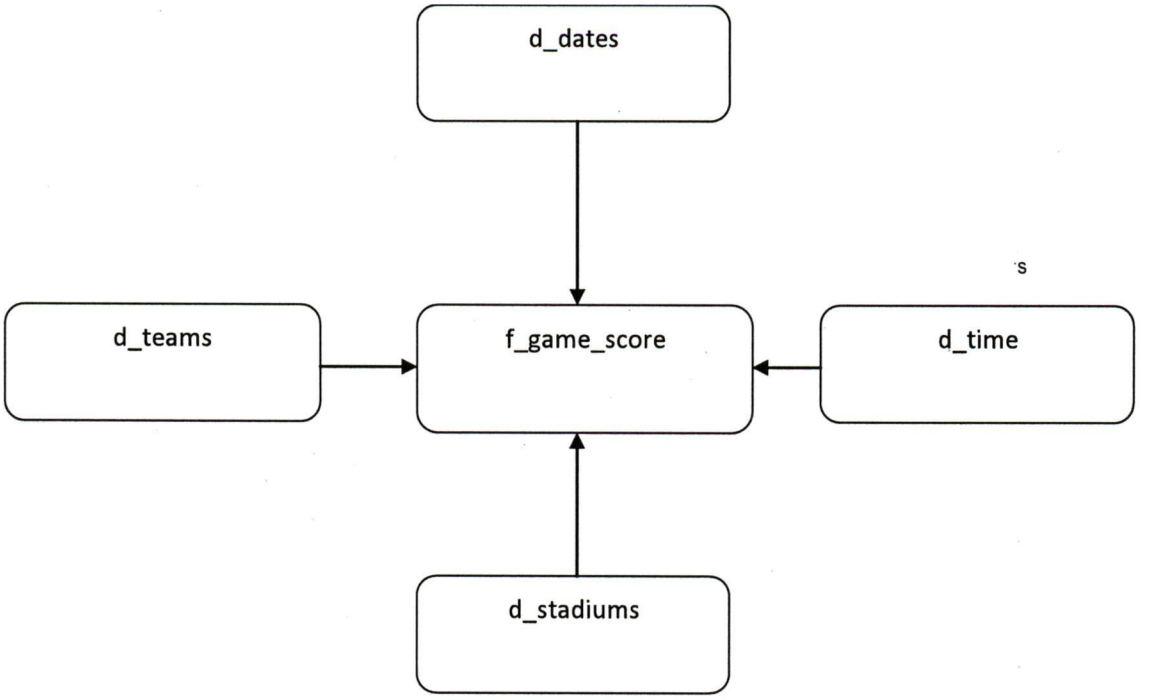
Annexe 2 Modèle de données

Comme il est impossible de montrer l'ensemble du modèle en une page, nous avons décidé de montrer chaque table de faits et ses dimensions séparément.

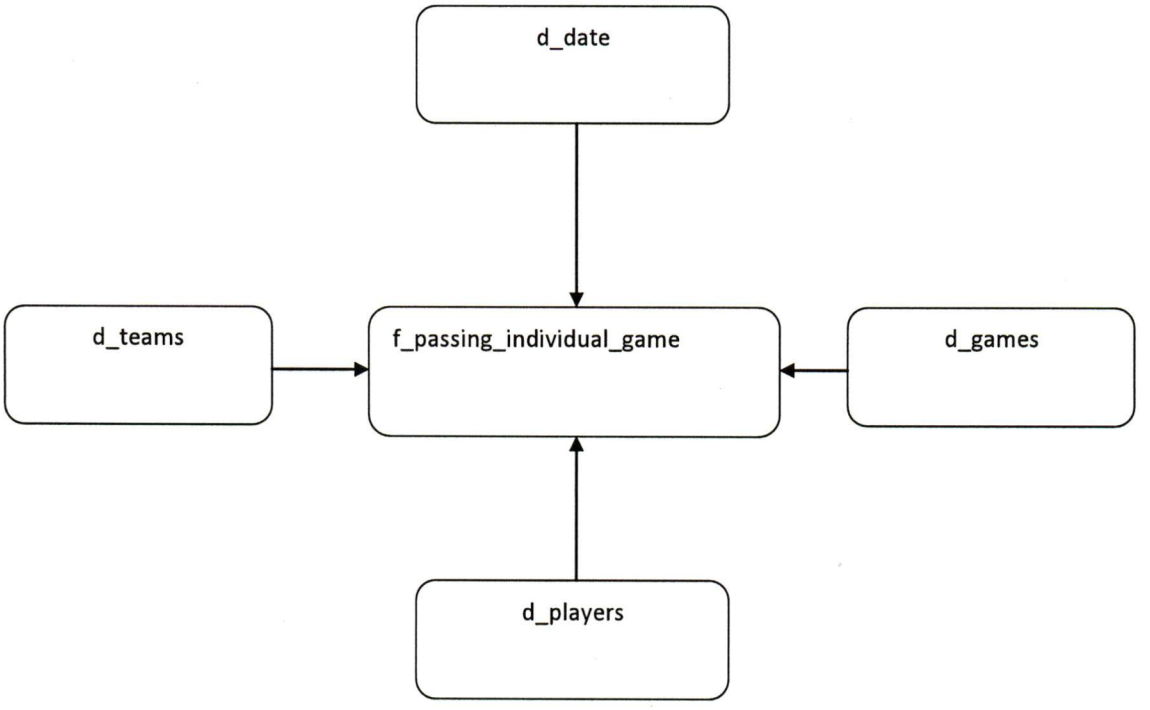



d_players
d_down_distance
d_ball_on
d_teams
d_dates
f_play_by_play
d_actions
d_clock
s
d_games
d_video
d_clock_offset

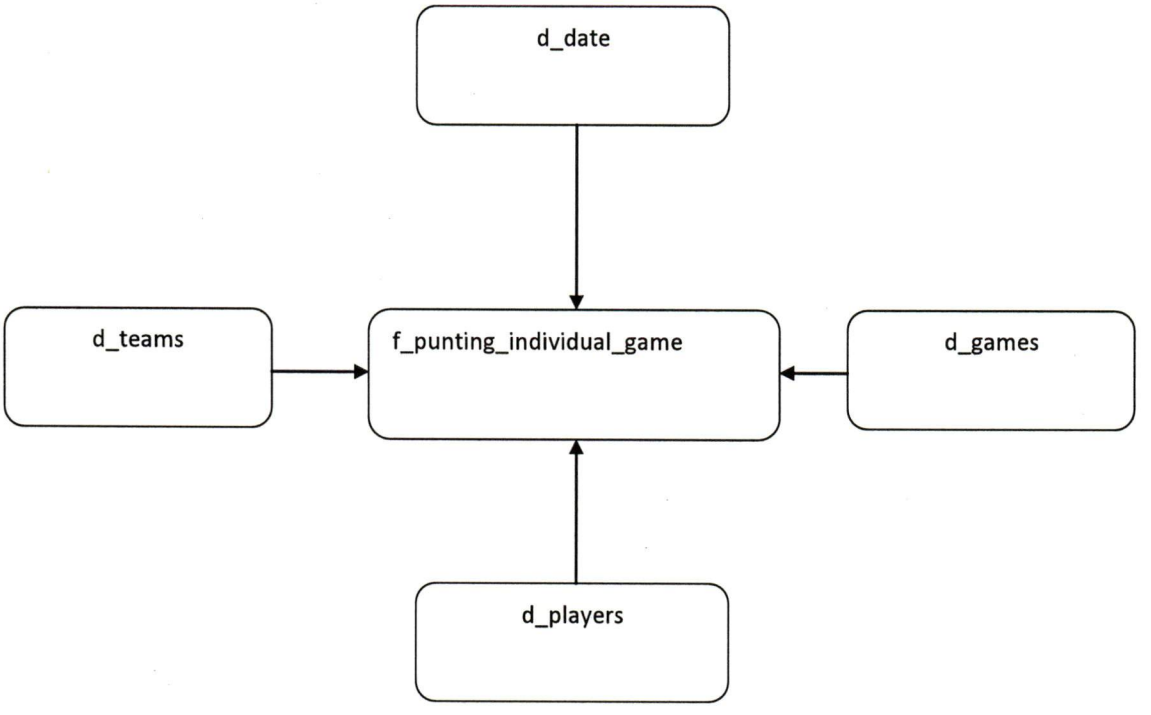

d_players
d_teams
f_defense_game
d_dates
d_games


d_dates
d_teams
f_game_score
d_time
d_stadiums

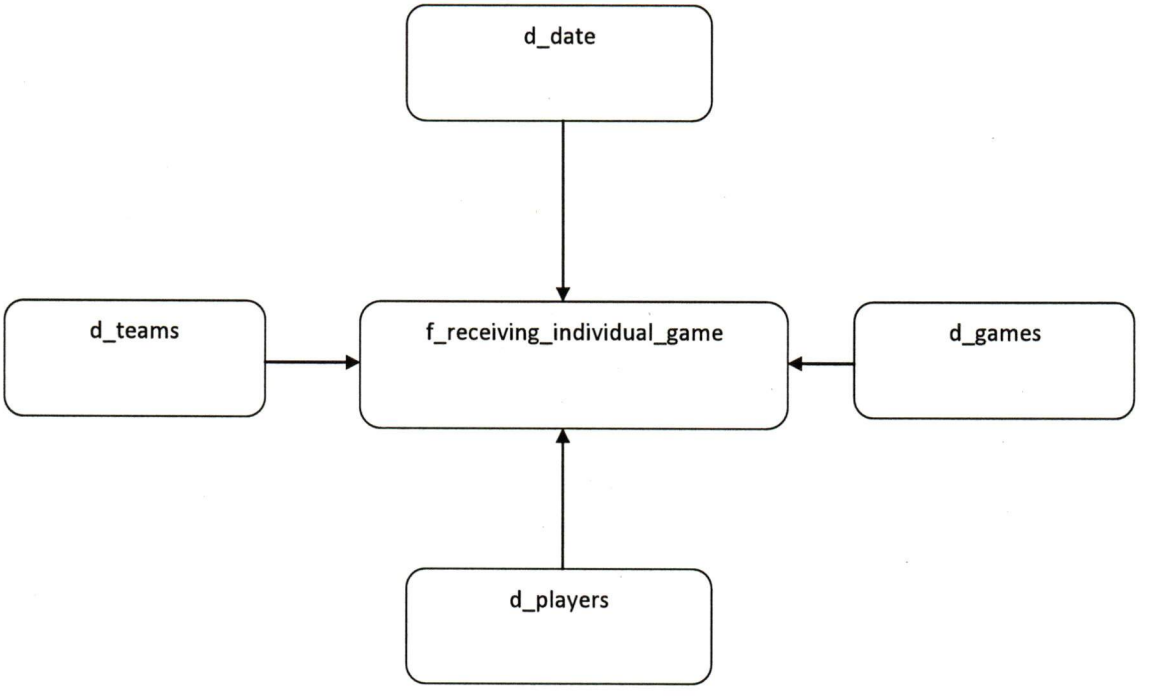
s

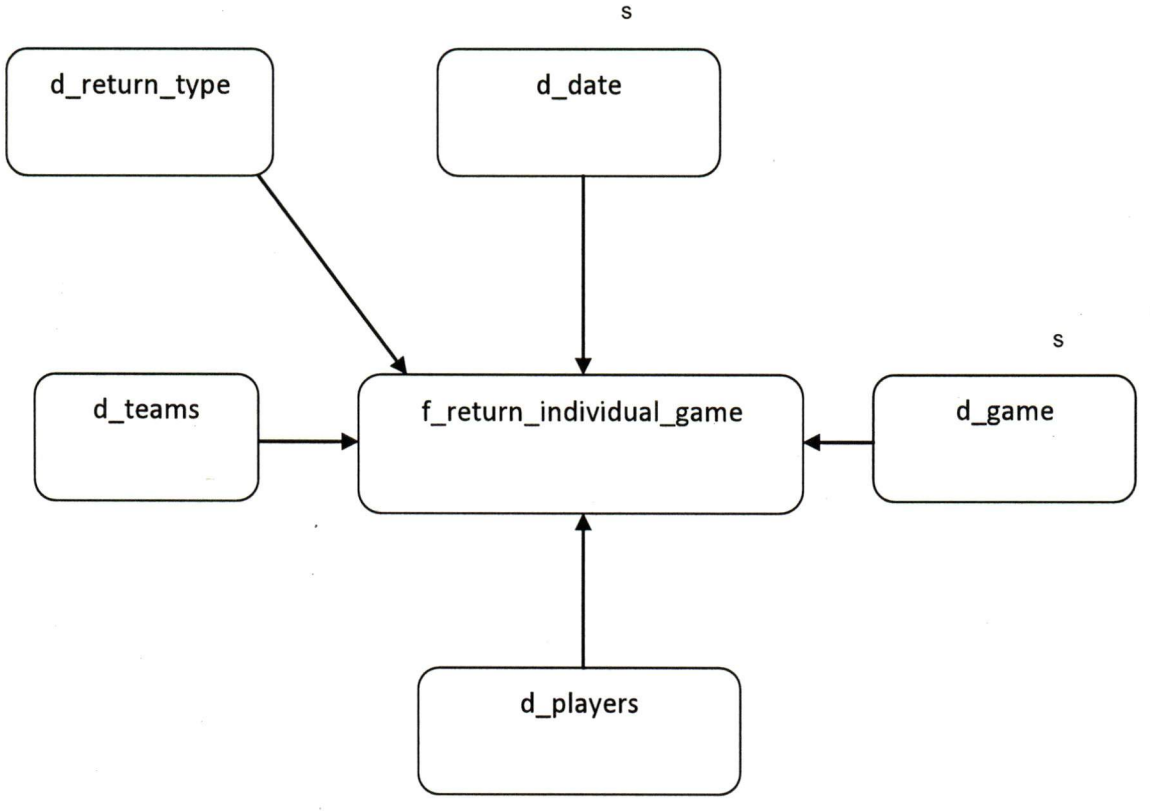


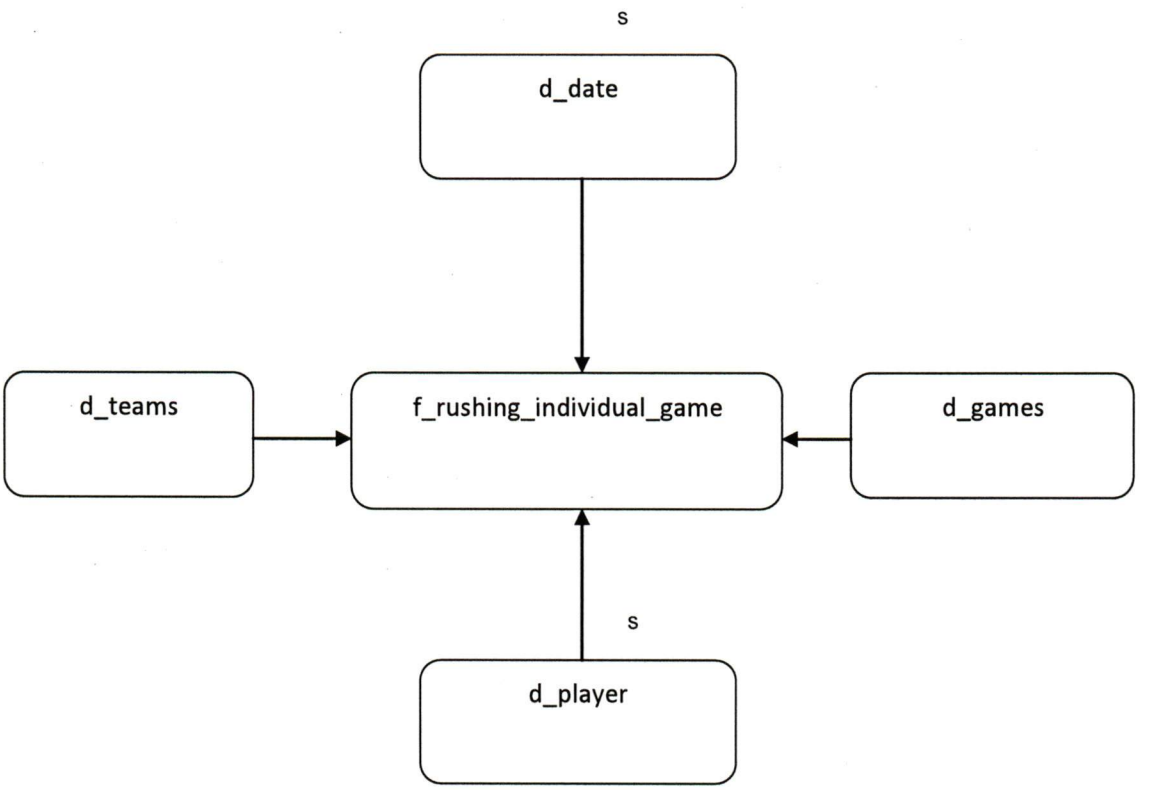
s

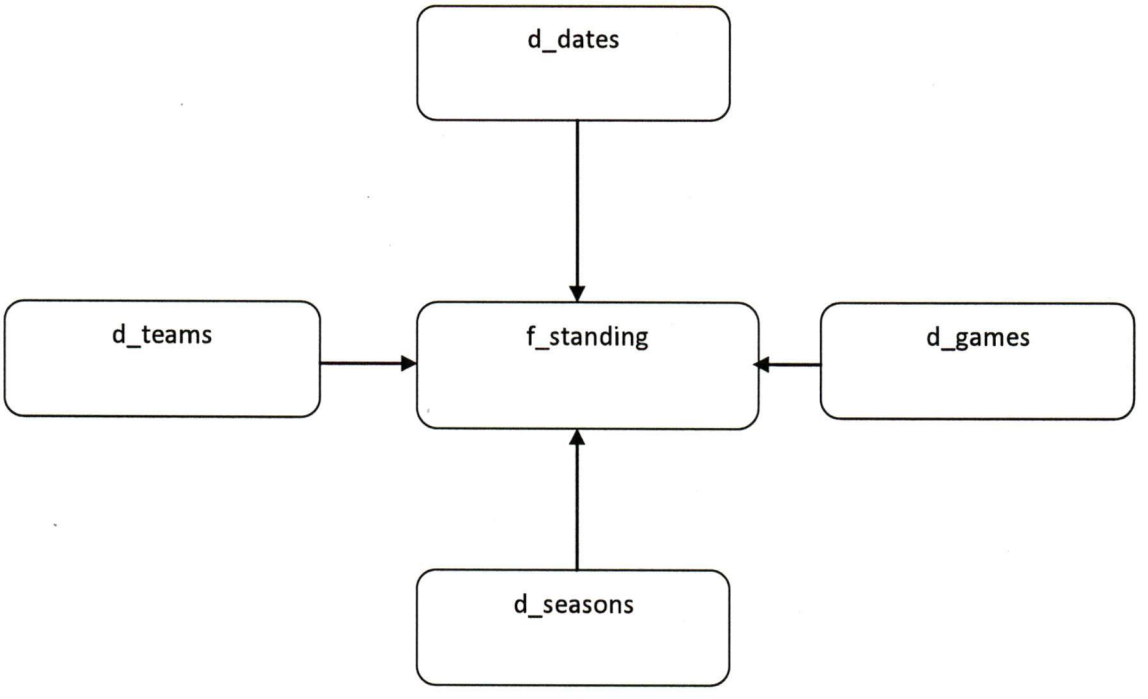


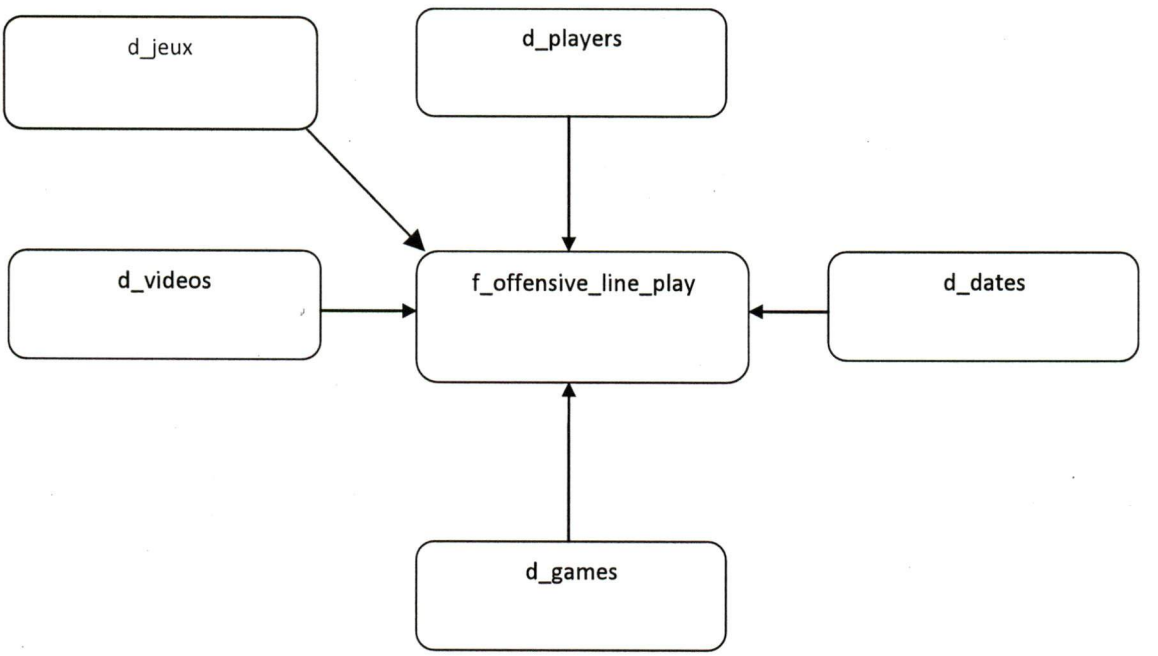
s




s
d_return_type
d_date
s
d_teams
f_return_individual_game
d_game
d_players


s
d_date
d_teams
f_rushing_individual_game
d_games
s
d_player


d_dates
d_teams
f_standing
d_games
d_seasons


d_jeux
d_players
d_videos
f_offensive_line_play
d_dates
d_games

Annexe 3 Détails de la base de données

Outil d'administration	Phpmyadmin
Base de données	MySQL 5.0
Moteur de stockage MySQL	InnoDB

Table A3 - 1 Détails de la base de données

Annexe 4 Routines ETC

Nous avons créé 26 routines ETC, soit une pour chaque table de dimension et de fait. La figure ci-dessous présente la routine ETC qui charge les données de la table de faits f_return_invividual_game.

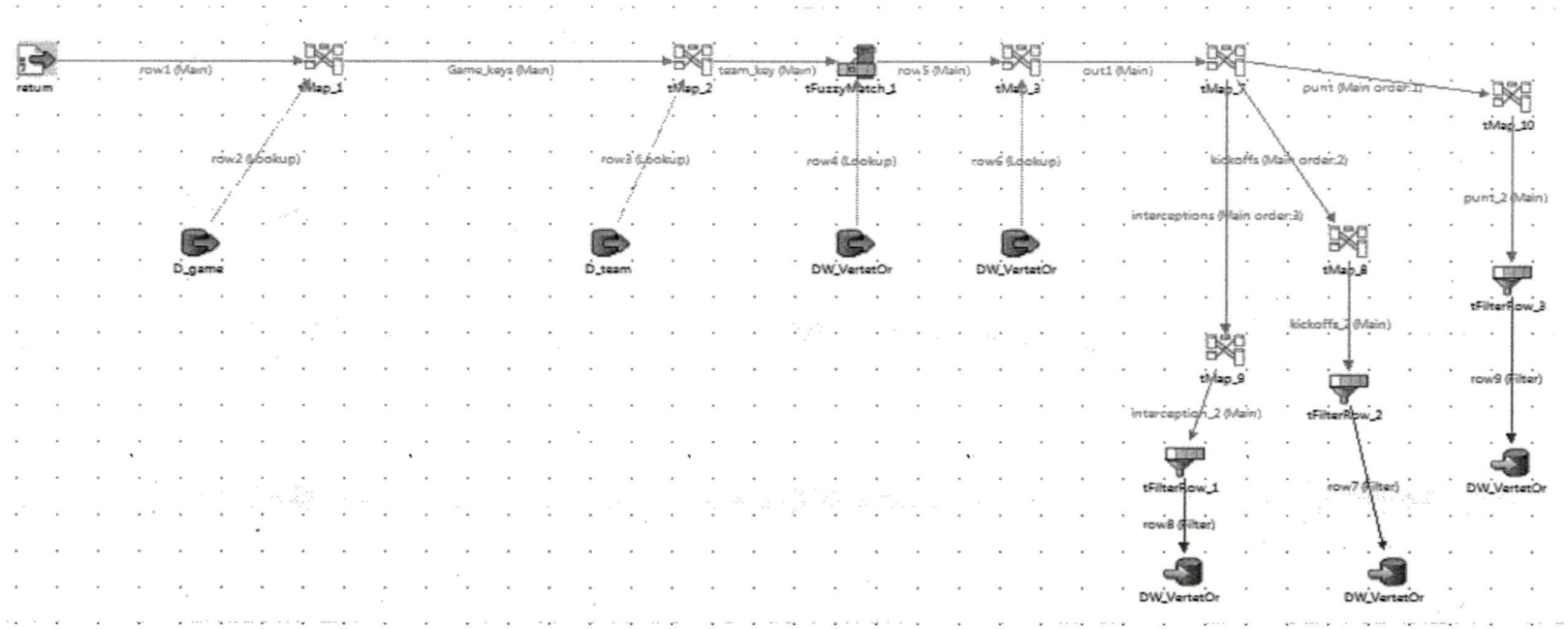


Figure A4 - 1 Routine ETC pour le chargement du fait f_return_individual_game

Annexe 5 Exemple d'une requête SQL

```
Select  f_offensive_line_play .score, f_offensive_line_play.dd_jeu,
(d_players.first_name + d_players.last_name) as player_name, d_videos.video,
d_games.home_team, d_games.away_team, d_dates.date

from f_offensive_line_play

inner join d_players on d_players.id_player = f_offensive_line_play.id_player

inner join d_videos on d_videos.id_video = f_offensive_line_play.id_video

inner join d_games on d_games.id_game = f_offensive_line_play.id_game

inner join d_dates on d_dates.id_date = f_offensive_line_play.id_date

where d_players.first_name = 'Kristof' and d_players.last_name ='Gyorfi' and
d_games.game_number = 3
```