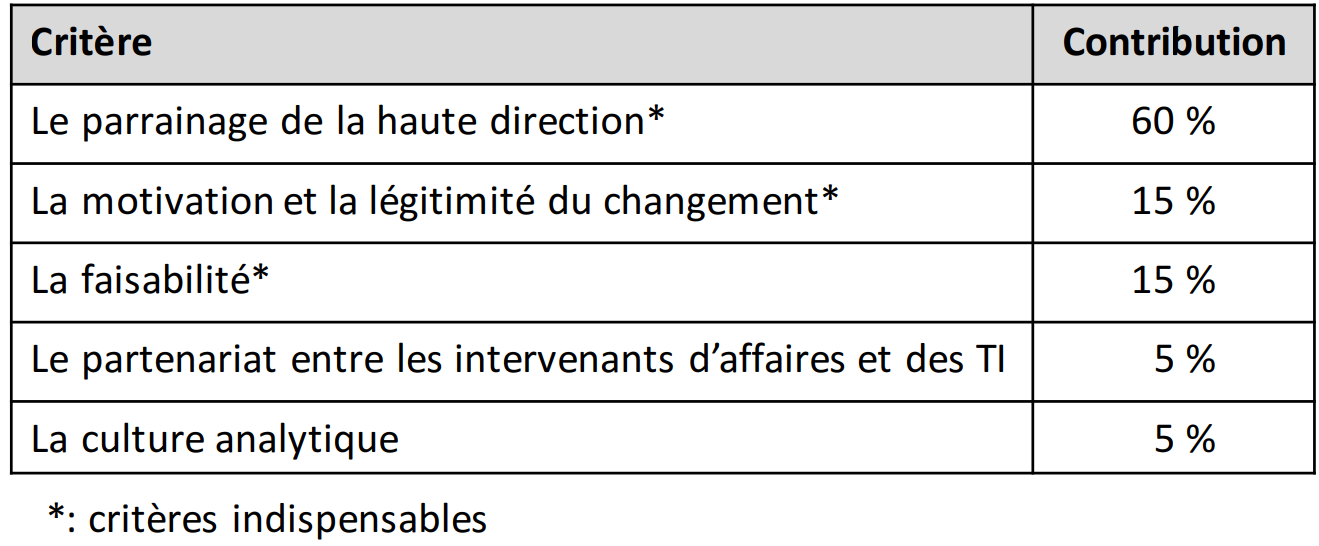
1. Les facteurs les plus déterminants pour la réussite du projet BI

Étant donné notre expérience, nous constatons généralement que la première intention d’un organisme souhaitant réaliser un système décisionnel est de partir des données à leur disposition. Si ce réflexe est parfaitement compréhensible, il est l’opposé de ce qu’il faut faire si l’on souhaite mener à bien un projet et obtenir des résultats.

Un projet décisionnel nécessite, tout d’abord et forcément, d’être mené en trio : l’IT représenté par l’équipe BI, un membre du service fonctionnel concerné par le périmètre traité et enfin un membre du contrôle de gestion. Ce projet doit si possible s’inscrire dans un projet d’entreprise disposant du sponsoring ou du moins du soutien d’un membre influent de la direction.

Ce mode de fonctionnement est essentiel, car il est indispensable de mener chaque étape du projet du haut vers le bas, c’est-à-dire du besoin global, le plus stratégique, vers la donnée la plus détaillée.

Les 5 critères de succès :

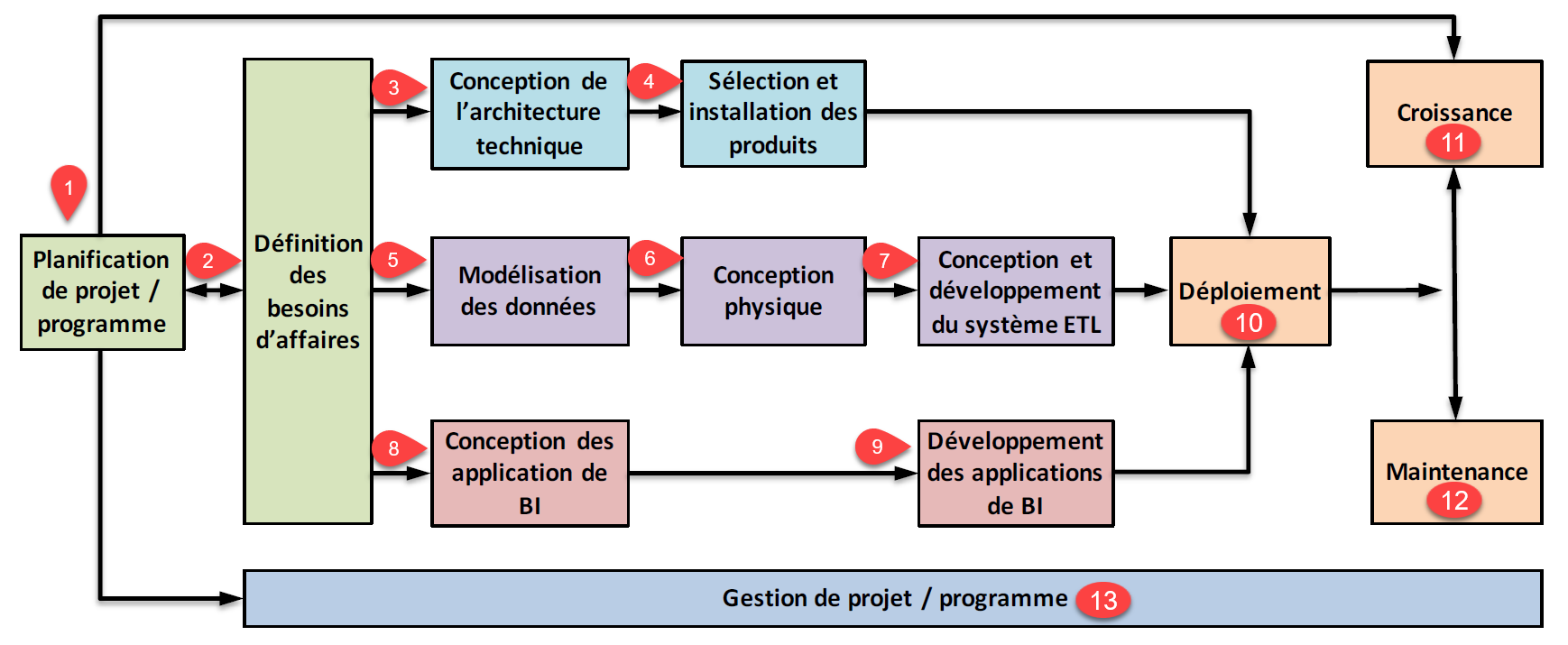


Pour débuter, on commence par s’assurer le périmètre projet auprès des décideurs stratégiques. Le périmètre projet initial est alors fractionné en lots que l’on va prioriser et étaler dans le temps. Un lot est généralement un sujet fonctionnel (pas trop grand) dont l’ordre de priorité est défini suivant deux critères : sa facilité de réalisation et sa valeur ajoutée.

Dans les faits, nous commençons toujours si possible par le périmètre le plus simple à réaliser (disponibilité de l’information, disponibilité du service, réflexion déjà bien avancée…) et par le périmètre qui apporte le plus de valeur ajoutée (criticité de la qualité de l’information, transversalité fonctionnelle du lot, orientation stratégique majeure…).

Le projet sera mené par itération successive en respectant le **cycle de vie du projet du projet BI**

cycle de vie du projet projet BI



1.Planification

Comprend:

• Définition de la portée;

• Mis en place de l’équipe du projet de BI

• Identification, estimation et affectation des tâches.

Définition de la portée du projet

**Périmètre du projet :**

Dans le cas de SHD, une première étude a été menée pour définir les contours du projet. Le périmètre projet a été défini. Il s’agit dans l’ordre des priorités :

1. **De la facturation (facture + commande + budget)**
2. **Des stocks :**
3. **Des achats (commandes fournisseurs + commandes en attente + facturation fournisseurs)**
4. **Des Formations (RH)**
5. **Des Salaires (RH)**
6. **Du service après-vente**
7. **Du marketing**

* XXX années de données historiques sur XXXXX ;
* Maximum de XX utilisateurs initialement, avec le plan d’augmenter la capacité à XXX utilisateurs;
* L’architecture technique sera basée sur SQL SERVER 2019 et Power BI.

**Exclusions:**

• Données externes, telles que XXXXXXXXXXXXXXXXXXXX

• Former les utilisateurs sur la création et partage de dashboard Power BI.

•XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

**Critères de succès du projet:**

Réaliser

- l'architecture des serveurs et le choix des licences  
- la modélisation de l'entrepôt de données  
- la conception du cube Analysis Services  
- la réalisation des différents types de flux d'alimentation ETL avec Integration Services  
- la mise en place de référentiels de données avec Master Data Services  
- l'utilisation d'Excel pour exploiter les données décisionnelles  
- la réalisation de rapports opérationnels et décisionnels avec Reporting ServicesWxw  
- et la création d'un dashboard interactif avec **Power BI**.

**Risques et stratégies de gestion des risques:**

- Ne pas pouvoir recruter un sponsor influent et visionnaire provenant de la haute direction;   
- En faire trop en même temps, au lieu d’employer une approche de développement itérative centrée sur un sujet à la fois;

- Être séduit par la technologie, au lieu de se concentrer sur les objectifs et les besoins de l’entreprise;   
- Supposer qu’il est possible de développer le projet sans l’apport des membres d’affaires;

- Aller de l’avant avec le projet, même si les données source sont de piètre qualité ou n’ont pas été validées;

- Faire la conception d’un entrepôt de données comme s’il s’agissait d’une BD transactionnelle (ex: normalisation);   
- Sous-estimer la quantité de travail reliée au nettoyage et la consolidation des données;

- Accorder trop d’importance au système ETL, au détriment de la performance et la qualité des applications de BI (et viceversa).

L’Équipe du projet BI et leur rôle

•**Sponsor :**

– Membre de la haute direction;

– Approuve les budgets;

– Défend le projet en cas de besoin.

• **Directeur de projet :**

– Sert de liaison entre les sponsors et les autres membres de l’équipe;

– Responsable du leadership et de la direction du projet;

– Influence les décisions stratégiques et architecturelles

• **Chef de projet:**

– Gère le projet au jour le jour:

• Coordination des ressources et des activités;

• Planification des tâches et des budgets;

• Suivi de l’état et de l’avancement du projet.

– Doit pouvoir réagir rapidement aux problèmes rencontrés, avant que ceux-ci s’intensifient.

• **Analyste d’affaires:**

– Gère la collecte et la définition des besoins;

– Assure la juste représentation des besoins dans les autres étapes du projet;

– Doit très bien connaître les processus d’affaires de l’entreprise.

• **Analyste en QA / intendant des données:**

– Responsable d’établir une vue consolidée des définitions et règles des données, à l’échelle de l’entreprise;

– S’assure que les données insérées dans l’entrepôt sont valides et complètes;

– Peut également être responsable de vérifier l’intégrité des applications de BI.

**• Architecte de données: Mohamed DIBASSI**

– En charge de définir une architecture facilitant la réutilisabilité, l’intégration et l’optimisation des données;

– Conçoit et développe le modèle dimensionnel des données;

– Participe parfois à la collecte des besoins.

**• Administrateur de BD (DBA): Mohamed DIBASSI**

– Responsable de convertir le modèle des données en structures physiques de l’entrepôt (tables, colonnes, etc.);

– Choisit des paramètres physiques maximisant la performance de l’entrepôt (ex: disposition des disques, partitionnement, indexes, etc.)

– Gère l’intégrité, la disponibilité et la performance de l’entrepôt au quotidien.

**• Architecte/développeur ETL: Mohamed DIBASSI**

– Responsable de la conception et développement du système ETL;

– Architecte est souvent impliqué dans la collecte des besoins;

**• Architecte/développeur d’application de BI: Mohamed DIBASSI**

– En charge de concevoir la couche d’interaction aux données;

– Conçoit et développe les applications de BI, souvent à l’aide de produits commerciaux;

– Configure la couche sémantique des outils de BI.

• Note:

– Un projet de BI comporte normalement tous ces rôles;

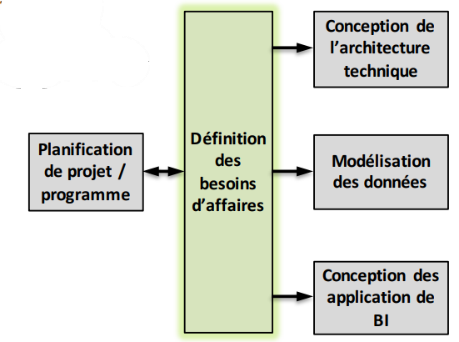
– Grand projets: plusieurs personnes peuvent avoir le même rôle;

– Petits projets: une personne peut jouer plusieurs rôles.

Identification, estimation et affectation des tâches :

**Voir Tableau kaban sur devops**

**2.Définition des besoins**



• Étape critique à la réussite du projet; Se concentre sur les utilisateurs, et non les données; Identifie les besoins les plus prioritaires;

• Comprend:

– Préparation du questionnaire et entrevue des utilisateurs/membres de l’équipe ;

– Identification des processus d’affaires;

– Préparation du document de description des besoins.

3.Conception de l’architecture technique

•Définit la vision d’ensemble de la solution et l’intégration des technologies du projet;

•Se concentre sur les besoins, pas les aspects techniques;

•Doit considérer :

– Environnement technique actuel;

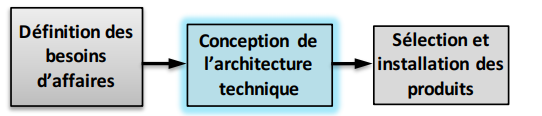
– Directions stratégiques prévues.

**• Comprend :**

–Choix des licences;

– Identification des besoins techniques;

– Création du plan d’architecture.



**3.1 Choix des licences**

Sur le fond, le besoin de SHD sur la solution Microsoft BI est structurée autour des trois promesses du décisionnel :

* Améliorer l’accès et la qualité des données :
* Gagner en finesse d’analyse et de compréhension de données
* Gérer les performances de l’organisation et de ses politiques :

Sur la forme, l’offre Microsoft BI est structurée au sein de trois licences. Chacune de ces licences contient de nombreux outils à usage décisionnel correspondant aux tâches énoncées plus haut :

* SQL Server 2019 : Integration Services, Master Data Services, Analysis Services, Reporting Services, Data Quality Services.
* Office 365 : Excel, SharePoint et Teams.
* Power Platform : Power BI, Power App, Power Automate.

Si les outils de la gamme SQL Server 2019 sont plutôt des produits techniques destinés, à l’usage et au service informatique, les outils Office 2019/Office 365 sont plus spécifiquement destinés aux utilisateurs finaux et les produits de la gamme Power Platform à des utilisateurs d’un niveau intermédiaire.

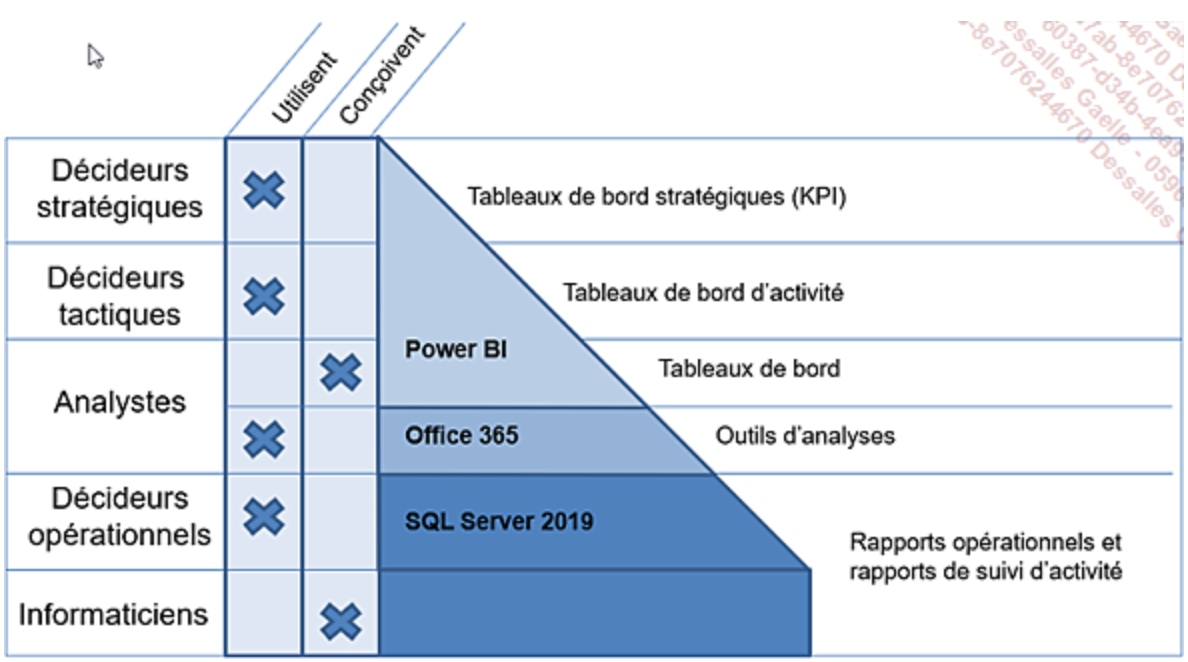
**La mise en place globale de la solution décisionnelle Microsoft nécessitera trois licences distinctes. Dans leurs versions les plus récentes à ce jour, ces licences sont :**

* **SQL Server 2019**
* **Office 2019 / Office 365**
* **Power BI**

**La licence SQL Server 2019 permet de détenir et d’installer le Power BI Report Server. La licence Power BI Premium est uniquement requise dans le cas de l’utilisation cloud des services Power BI.**

**3.2Critères sélection des outils**

Dans la suite de cette partie, nous allons étudier le contenu, l’utilisation et le positionnement de chacune de ces solutions.



L’utilisation des licences Microsoft suivant les profils utilisateurs

**2.2 SQL Server 2019**

Nous avons vu précédemment que la construction d’un système décisionnel ne se limite pas à l’utilisation d’une base de données relationnelle. Avec la licence SQL Server, Microsoft nous livre toute la panoplie d’outils dont un service informatique a besoin pour bâtir un système décisionnel dans les règles de l’art.

Pour bâtir notre système d’aide à la décision, nous aurons besoin de :

* SQL Server Integration Services : l’ETL.
* SQL Server Master Data Services : le gestionnaire des données de référence.
* SQL Server Analysis Services : la base de données multidimensionnelle (OLAP) et le métamodèle.
* SQL Server Data Quality Services : la base de connaissances permettant de qualifier et nettoyer certaines données constitutives de vos référentiels de données ou des bases marketing.
* SQL Server Reporting Services : l’outil de reporting opérationnel et de reporting de masse.

Dans les sections qui suivent, nous allons présenter plus en détail chacun de ces outils.

* + - * **2.2.1 SQL Server Integration Services**

La constitution des données de référence et l’alimentation d’un entrepôt de données représentent généralement **près de 80 % de la charge du projet.** De prime à bord, beaucoup de services informatiques ont tendance à sous-estimer le temps nécessaire à récupérer l’information, mais aussi à la croiser. Pour nous aider dans cette tâche, Microsoft met à notre disposition, dans la licence SQL Server 2019, un outil d’ETL (Extract Transform Load) : SQL Server Integration Services (SSIS).

* + - * 2.2.2 SQL Server Master Data Services

SQL Server Master Data Services (MDS) est une solution de MDM (Master Data Management) ou Management des données de référence.

**2.2.3 SQL Server Data Quality Services**

Cette solution permet de contrôler et d’améliorer la qualité des données qui circulent au sein de l’organisation à l’aide de bases de connaissances liées au contenu et à la structure de vos données.

Ici, nous allons contrôler la qualité en utilisant le langage regex

2.2.4 SQL Server Analysis Services

SQL Server Analysis Services regroupe en fait plusieurs services destinés à faciliter l’exploration d’une quantité importante de données.

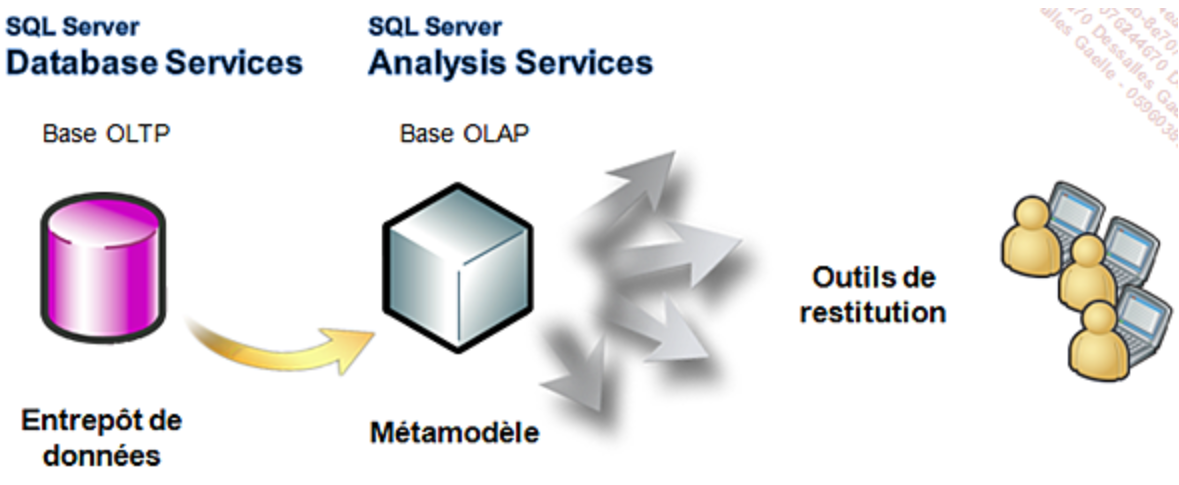
Les trois facettes de la solution sont :

* Le mode multidimensionnel
* Le mode exploration de données ou data mining
* Le mode tabulaire

**Le mode multidimensionnel**

Nous ne saurions que trop insister tout au long sur l’importance d’une modélisation appropriée, car la performance de son moteur OLAP est directement liée à la modélisation de la base de données sous-jacente.

Dans les faits, les utilisateurs n’accèdent pas directement à Analysis Services. Ce sont les outils de restitution qui proposent et consomment les données du cube. Il est donc indispensable, lorsque vous mettez en place SSAS, de proposer parallèlement aux utilisateurs, des outils de restitution adaptés (par exemple Excel).



Analysis Services est l’interface entre l’entrepôt de données et les outils de restitution.

**Le mode tabulaire**

Tout comme le mode multidimensionnel, le mode tabulaire offre un modèle sémantique qui se veut ouvert à tous les types de restitution : reporting, analyse, tableau de bord

Le mode tabulaire propose une autre façon de concevoir des modèles et de les stocker. Contrairement au mode multidimensionnel qui nécessite une modélisation dimensionnelle et un stockage OLAP, le mode tabulaire utilise une modélisation relationnelle et un moteur d’analyse en mémoire vive pour le stockage et le calcul.

De ce fait, le mode tabulaire va être plus proche des modèles relationnels classiques, tout en permettant une vitesse de restitution de données extrêmement rapide, de la donnée détaillée comme de la donnée agrégée.

**Pour utiliser le mode tabulaire, il vous faudra toutefois créer une instance dédiée et nous vous conseillons de dédier une machine physique et puissante à cette instance.**

**C’est pourquoi nous vous suggérons de réaliser la nouvelle solution en mode multidimensionnel.**

**2.2.5 SQL Server Reporting Services**

Enfin, la suite SQL Server dispose d’un serveur de rapports permettant d’afficher et de diffuser des informations.

Reporting Services est un excellent outil de reporting opérationnel. Elle est parfaite pour mettre en page une facture, un bon de livraison, un suivi de commande, un inventaire, un catalogue produit, la liste des clients à relancer… Tous les états dont une application de gestion a besoin sont parfaitement réalisables avec Reporting Services. Ce n’est pas un outil d’analyse destiné à des utilisateurs finaux. En revanche, il a totalement sa place dans la diffusion de rapports de masse, c’est-à-dire dans les rapports décisionnels destinés à de nombreux décideurs opérationnels.

Par exemple, la diffusion par mail au format PDF du rapport mensuel de suivi des ventes à tous les commerciaux de l’entreprise.

* + - 2.3 Office 2019/Office 365

Au sein de la solution BI Microsoft, Excel est l’outil d’analyse des utilisateurs finaux. Toutefois, Excel n’est utilisé que pour accéder, manipuler et naviguer dans les données d’Analysis Services. Les données ne sont plus contenues directement dans le fichier, mais sur un serveur. Et ces données ne sont plus traitées directement par les utilisateurs fonctionnels mais par le service informatique par le biais de l’ETL. Excel conserve toutefois toutes ses capacités de représentation graphique, de mise en page et de personnalisation à l’aide de formules.

2.4 Power BI

Power BI permet de fournir des interfaces interactives de visualisation des données. Il est tourné vers des utilisateurs non informaticiens, leur permettant à l’aide de l’outil Power BI desktop de concevoir des dashboards interactifs très riches. Contrairement à Excel, qui n’est pas conçu pour partager des visualisations à une équipe ou à son organisation, Power BI est vraiment pensé pour cela.

Nous restituerons les données décisionnelles - Power BI en déployant un serveur de rapports Power BI sans en passer par le cloud.

Cependant, Power BI n’est pas un outil d’analyse. Il est très efficace pour présenter l’information, mais très peu performant pour l’explorer. Excel reste en cela incontournable.

Il serait intéressant de prévoir une formation sur la création, et le partager au sein de votre organisation d’un dashboard Power BI.

* + **NB Microsoft BI et Azure :**

La conception avec succès des projets décisionnels dans le cloud est aujourd’hui possible. Il y a toutefois des problématiques autour du transfert de données et de l’authentification qui s’ajoutent à ceux d’un projet On-Premise. Bien entendu, une fois ces problèmes surmontés, le système décisionnel bénéficie de toute la puissance du cloud en termes d’élasticité, de disponibilité et de sécurité.

**Nous suggérons de nous consacrer à la mise en place d’un système décisionnel entièrement On-premise.**

**3.2– Identification des besoins techniques;**

La mise en place d’un projet décisionnel nécessite en fait plusieurs environnements de travail. On dénombre généralement :

* Un environnement de développement
* Un environnement de recette
* Un environnement d’intégration
* Un environnement de production

En début de projet décisionnel, deux environnements peuvent suffire : celui de développement et celui de production, l’environnement de production jouant dans un premier temps aussi le rôle d’environnement de recette et d’intégration.

Toutefois, une fois que le projet est lancé et que le premier lot projet va être en production et bientôt mis à disposition des utilisateurs, il devient indispensable de disposer a minima d’un environnement de recette, pouvant jouer aussi pendant un temps le rôle d’intégration.

**NB :** Dans la solution Microsoft, le serveur et l’outil de contrôle de code source sont inclus dans l’offre Azure Devops Services pour sa version cloud et Azure Devops Server dans sa version On-premise.

Ensuite, quand le projet arrive à une certaine maturité et que le système décisionnel devient incontournable et donc critique, il convient alors de distinguer l’environnement de recette de l’environnement d’intégration.

# **3.3 Création du plan d’architecture**

## 3.3.1 L’environnement de production

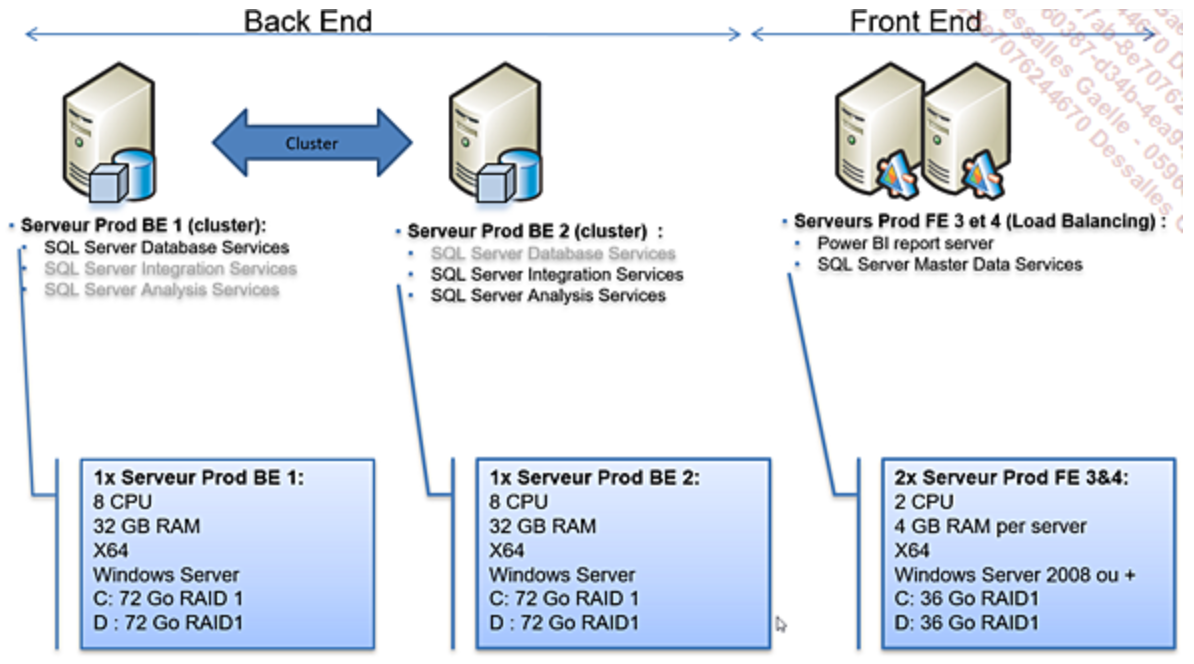
La performance d’un environnement de production est évaluée suivant deux axes :

* Sa capacité à monter en charge pour répondre efficacement à toutes les demandes utilisateurs.
* Sa haute disponibilité. Par haute disponibilité, on entend sa capacité à offrir un service chaque fois qu’un utilisateur en a besoin.

Les critères entrant en jeu lors de la définition de l’environnement de production sont :

* La volumétrie
* Le nombre d’utilisateurs au total et le nombre d’analystes
* Les prévisions d’évolution et de montée en charge du système
* La haute disponibilité

**La configuration de production intégrant la haute disponibilité :**



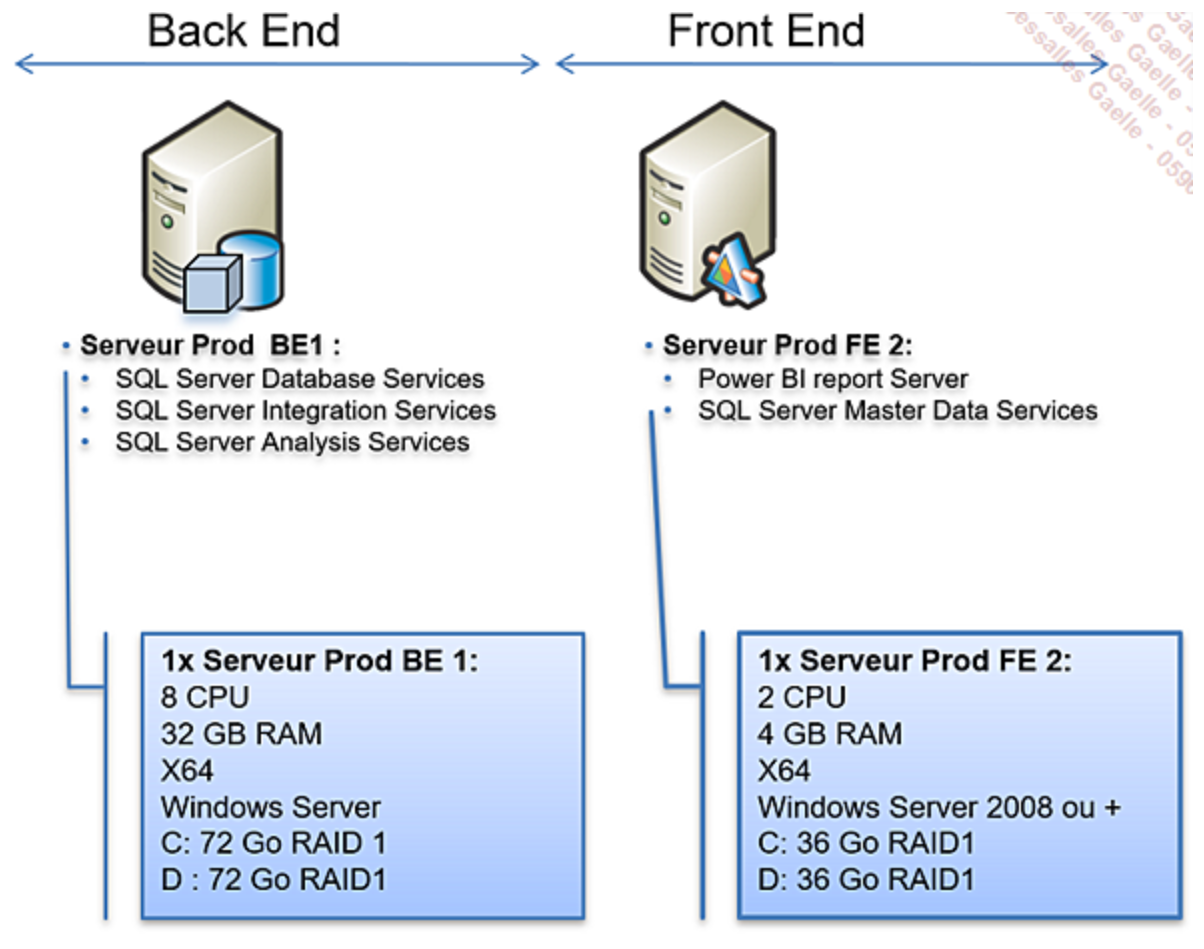
Typologie de l’architecture de production en haute disponibilité

Cette architecture nécessite :

* 2 serveurs back-end en cluster, qui devront être deux grosses machines physiques.
* 2 serveurs front-end en load balancing (ou équilibrage de charge), qui peuvent être deux petites machines virtuelles.
* 1 serveur Active Directory : il s’agira vraisemblablement d’un serveur existant. Sachez cependant que le service Active Directory sera nécessaire pour installer et configurer correctement la solution BI Microsoft.
* Une baie SAN : l’optimisation d’une architecture serveur passe nécessairement par un plan de découpage et d’isolement des bases et des partitions sur des disques physiques distincts :
  + Il s’agit tout d’abord de séparer les fichiers Data, Logs et Temp du service SQL Server Database Services sur des disques distincts.
  + Il s’agit ensuite qu’au sein d’un même flux le disque contenant la base ou la table source soit distinct de celui contenant la base ou la table de destination. Et ce, afin d’éviter qu’un disque physique travaille à la fois en lecture et en écriture au cours de l’exécution d’un flux.

Si vous ne disposez pas de SAN et que vous ne souhaitez pas investir, prévoyez un maximum de disques durs.

La configuration de production sans haute disponibilité :



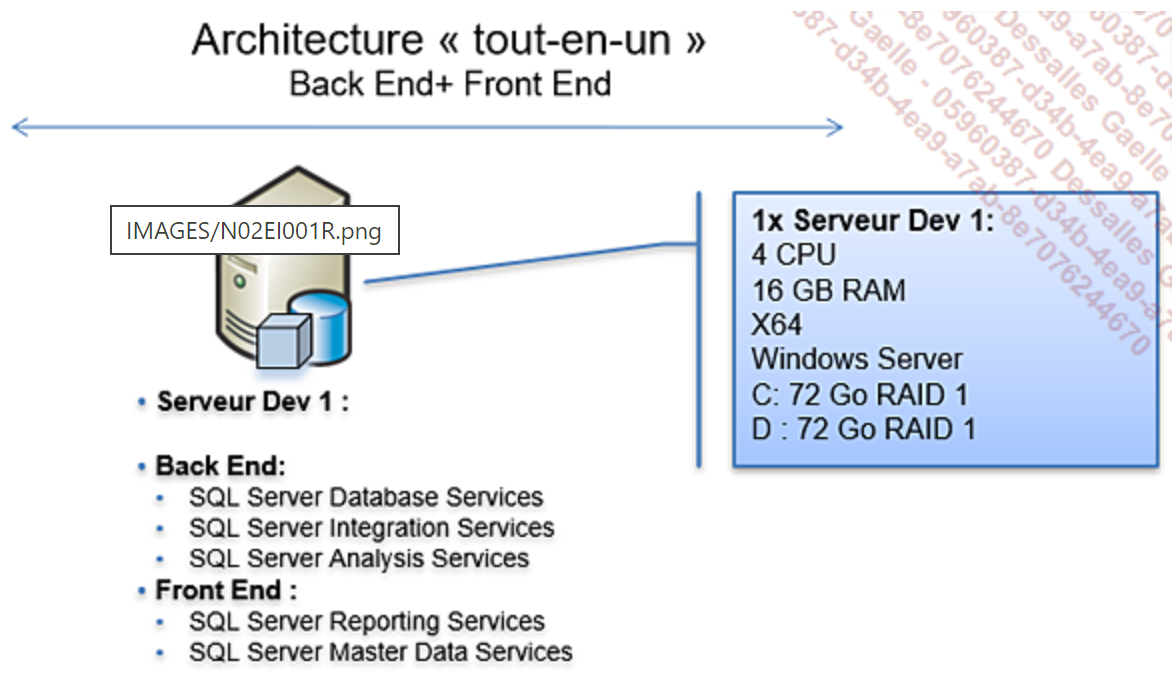
## 3.3.2. L’environnement de développement

L’environnement de développement n’est pas soumis aux contraintes de haute disponibilité. Il s’agit toutefois de réaliser un environnement robuste et performant afin que l’équipe de développement puisse travailler avec tout le confort nécessaire.

Afin d’améliorer la productivité de vos équipes, nous vous engageons à investir dans un outil de contrôle de code source. Cet outil vous permettra de partager vos projets, historiser les changements, tracer les modifications et permettre de réaliser une sauvegarde régulière et quotidienne de l’ensemble.

En termes d’architecture serveur, une architecture tout-en-un comprenant l’ensemble des serveurs présents en production est suffisante. Prévoyez tout simplement une bonne machine.

À titre d’exemple, nous vous proposons cette configuration, qui devra elle aussi être adaptée à votre contexte



Exemple de configuration pour un environnement de développement

## 3.3.3. Les environnements de recette et d’intégration

En termes de configuration, l’environnement de recette n’a pas de contrainte particulière et peut ressembler par exemple à l’environnement de développement.

L’environnement d’intégration quant à lui devra si possible ressembler autant que possible à l’environnement de production pour y intégrer les mêmes contraintes.

En termes de budget, cela peut revenir bien évidemment très cher. Un compromis moyen peut être trouvé en considérant au moins un environnement de recette/intégration avec une architecture de production sans la haute disponibilité.

3.Modélisation des données

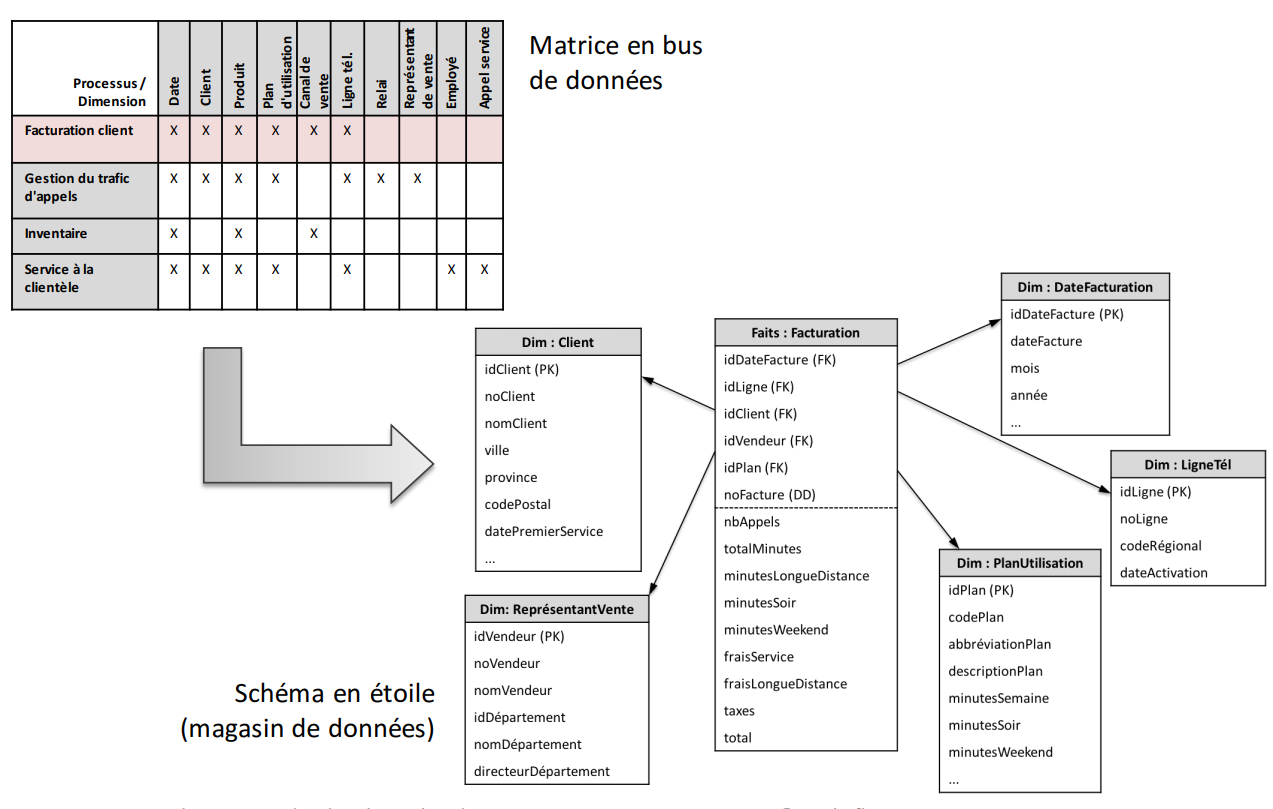
• Se base sur le document de description des besoins; • Processus hautement itératif et dynamique; Vise la simplicité et l’efficacité;

• Comprend:

– Identification des faits (mésures) et leur granularité;

– Identification des dimensions et leur hiérarchie;

– Stratégies : Gestion des changements (upsert, scd3, etc), dénormalisation etc.



4.Conception physique des données

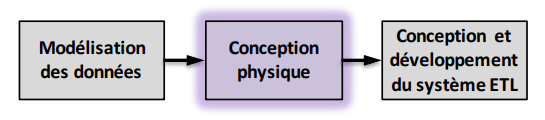
• Utilise le modèle de données;

• Comprend:

– Détail du schéma relationnel (ex: clés, types, contraintes, etc.);

– Optimisation de la performance (ex: indexes, partitionnement, agrégation, etc.);

– Gestion de la sécurité.



6.Conception et développement du système ETL

• Représente environ 70% des efforts et risques du projet;

• Doit considérer:

– Nombre et type des sources de données;

– Outils disponibles.

• Comprend:

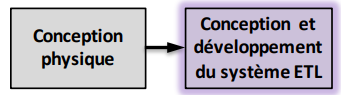
– Identification et analyse des sources de données;

– Développement des méthodes d’extraction, de nettoyage et de consolidation des données (code maison ou outils commerciaux);

– Développement des méthodes d’insertion de données (ex: scripts, outils complexes);

– Validation de la qualité des données

– Définition de la Matrice de transformation



7.Conception et développement des applications BI

• Se fait en parallèle avec la modélisation des données et la conception du plan d’architecture;

• Nécessite une interaction importante avec les utilisateurs;

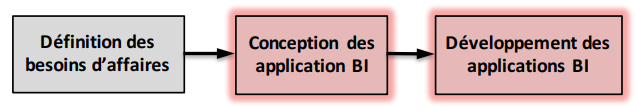
• Comprend:

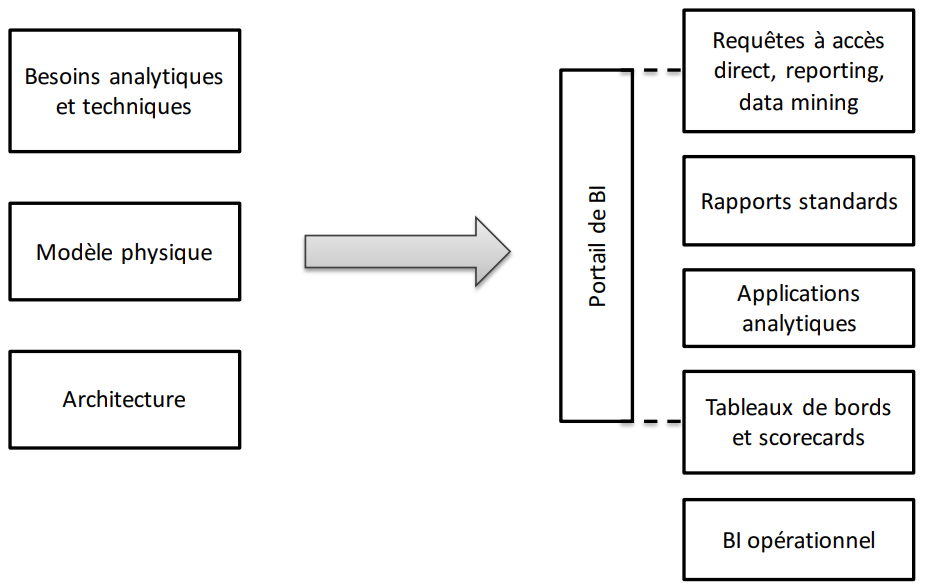
– Modélisation des tableaux de bord, rapports, indices de performances (KPI) adaptés aux utilisateurs;

– Définition des modèles de prédiction, classification et clustering; (si possible)

– Configuration des outils et des métadonnées;

– Implémentation du portail de navigation; – Validation des applications.





8.Gestion du projet

• Assure la coordination des activités du projet;

• Comprend:

– Suivi de l’avancement et de l’état du projet;

– Suivi des problèmes;

– Contrôle des changements (ex: portée du projet);

– Développement du plan de communication (affaires et TI).

9.Plan de projet

•Définit pour chaque activité:

– Ressources responsables;

– Les dates de début et de fin estimées (ne doivent pas être modifiées);

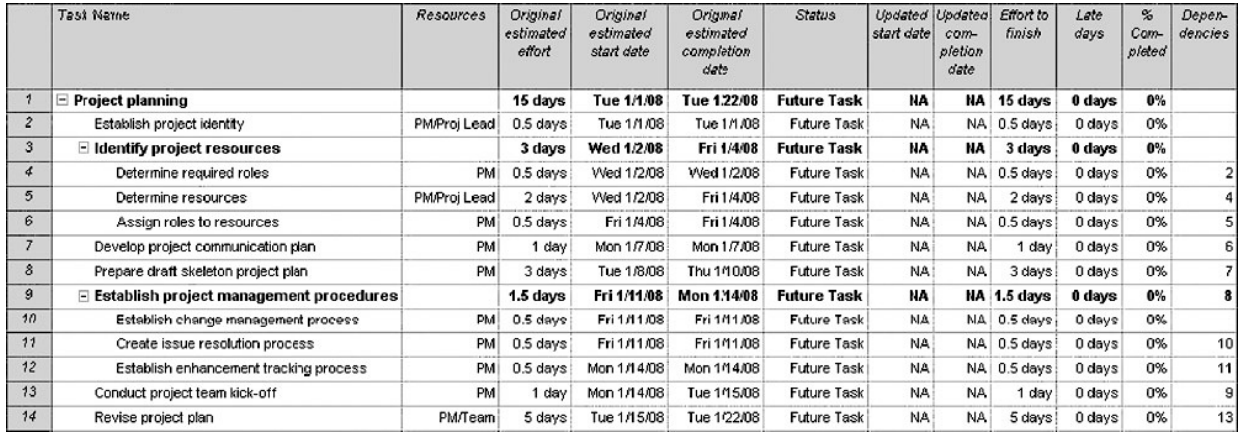
– L’état courant (ex: future, en cours ou complétée);

– Les dates de début et de fin mises à jour;

– Le nombre de jours de retard;

– Le % complété;

– Les dépendances (autres tâches).



10.Déploiement

• Point de convergence des activités de développement;

• Ne doit pas être fait avant d’avoir:

– Matériel de formation des utilisateurs;

– Mécanismes de gestion et de suivi d’erreurs;

– Documentation complète;

– Validation des données et outils;

– Processus efficace de communication.

**Utiliser Devops si possible**

11. Maintenance et croissance

• Assure le fonctionnement optimal du système et prévoit l’ajout de nouvelles fonctionnalités;

• Comprend:

– Suivi de l’utilisation et réglages de performance (tuning);

– Sauvegarde et récupération des données;

– Support aux utilisateurs;

– Préparation de nouveaux cycles de développement.