

# Projet 1 IDL 1

## Avant propos

*Il y a maintenant un peu plus de vingt ans, je séjournais à Sant Cugat Del Valles, en Catalogne espagnole, pour parfaire la réalisation d'une solution industrialisée de commerce électronique reposant sur l'offre Intershop ePages, qui, à l'époque, était la première plateforme scalable pouvant permettre de mettre en ligne des milliers de sites marchands disposant de toutes les fonctionnalités usuelles ... catalogue, panier, solution de paiement, front et back office personnalisable ...*

*Ce projet était partie intégrante d'un programme de développement régional, financé par des structures institutionnelles, et contenait la mise aux meilleures normes de réseau en fibre optique, la mise en place d'un centre technique d'hébergement de ressources d'infrastructures, et des "accessoires" comme la mise à disposition, pour la commune de Sant Cugat, municipalité regroupant à l'époque 50.000 habitants en banlieue de Barcelone, d'une capacité d'un site personnel pour chaque habitant qui en ferait la demande, et d'un site de commerce électronique pour chaque commerçant, artisan, ou entreprise de plus grande ampleur souhaitant présenter son ouvrage et commercialiser ses produits ou services.*

*Ce projet tenait lieu, au plan local et régional, d'évènement significatif car il allumait un projecteur appréciable sur la modernité de l'agglomération de Barcelone et de la province.*

*La livraison du projet, au sens qui me concernait, consistait à avoir réalisé le déploiement des serveurs, assuré la connectivité des applications et le premier paramétrage, de telle sorte que, les tests passés, l'alcade de Sant Cugat pourrait faire une inauguration devant les médias, en montrant qu'il était en mesure de se connecter avec un PC à Internet pour passer commande à un commerçant, d'un produit qu'on lui livrerait sur le champ.*

*La petite équipe que j'encadrais, fine fleur du département ingénierie basé à Iéna en Allemagne, avait fait son œuvre, non sans mal, et même parfois avec âpreté, car ces moments intenses étaient faits autant d'une certaine science, d'une réelle foi, et d'un peu ou beaucoup de sens de l'initiative.*

*A l'approche de la date prévue de l'inauguration, la chaîne de production complète semblait fonctionner.*

*La veille de l'inauguration, tout ronronnait.*

*Le jour de l'inauguration, il ne restait plus qu'à décider de quel commerçant, que nous avions déjà installé dans le système, il allait quérir un produit quelconque.*

*La décision ne fut pas simple, car les commerçants ciblés n'avaient pas tous été bien préparés à recevoir une commande par Internet avec un PC disponible sur leur comptoir ... il peut sembler étonnant qu'il en soit ainsi ... mais !*

*A bien voir et revoir la liste, les personnes, il fut décidé que le boulanger installé près de la grande place du vieux Sant Cugat était la personne la plus adaptée. Il avait un PC, et semblait pouvoir s'en servir sans surprise, y compris pour se connecter à son backoffice de commerçant Internet, c'est-à-dire au backend de sa boutique sur le serveur Intershop.*

*Que pouvait-on lui commander ... un boulanger par Internet, aujourd'hui, on trouverait peut-être plus facilement des idées, en 1999 il fut décidé qu'on choisirait la baguette, qu'on montrerait comment accéder à la page de la baguette, au maire de la ville, qui devait ... faire en sorte d'être seul à la manœuvre devant les caméras.*

*L'heure venue ... vous devinez que si je vous raconte tout ceci, ce n'est pas pour un bout de pain ... l'heure venue, donc, l'alcade se rapprocha du petit kiosque sur lequel trônait un PC, allumé et opérationnel. On le vit ouvrir Netscape, eh oui, cela remonte à un certain nombre d'années, et il put se connecter, grâce à un raccourci habilement prévu pour lui faire gagner du temps, à la boutique du boulanger.*

*Honnêtement j'ai complètement zappé le prénom du boulanger. Mais il était connecté de son côté, et pu montrer à d'autres caméras qui étaient près de lui, que la commande venait de tomber ... aussi vite qu'elle venait de partir, non pas parce que la distance était courte entre le milieu de la grande place et le début de la rue, mais parce que la commande avait quand même transité par Internet, entre postes clients et serveurs, et, ce qui advint de cette démonstration est, je crois, mémorable et extraordinaire !*

*Le boulanger, après avoir reçu la commande payée, prit la baguette, après avoir bien mis en évidence l'endroit où il devait la livrer, et l'apportât à son maire, assez complice, qui la prit en contrepartie de sa commande payée.*

*Je peux vous assurer qu'à la même époque un dénommé Jeff Bezos était encore en train de tenter de convaincre des investisseurs pour financer un e-commerce d'autre part balbutiant, et que lui était loin du compte pour imaginer pouvoir prendre en charge une logistique mondiale avec livreurs en livrée. J'ai eu l'opportunité, aussi, d'accompagner quelques travaux pour Amazon, qui avait encore sa coquille d'œuf sur le sommet du crâne.*

*Nous avons réalisé ce projet de serveur ePages à Sant Cugat del Vallès en préparant les opérations depuis les services installés à Léna, et formant une petite équipe variant de trois à six personnes dépêchées sur place, en nous débrouillant pour que ça fonctionne, et en sachant que les lauriers resteraient longtemps posés sur la tête du maire qui visait la députation.*

*Ce qui l'intéressait, lui, était de pouvoir montrer à des milliers de personnes qu'il savait mettre en place l'infrastructure du vingt et unième siècle dans sa ville et dans les environs.*

*Ce qui nous a occupés, nous, c'était de le faire fonctionner.*

*Vingt et quelques années plus tard, il vous appartient de prendre votre tour, de le faire dans un cadre pas tellement différent de celui que je décrivais juste avant. Vous êtes une promotion d'apprentis, vous visez un titre qui n'a rien d'honorifique, il faut le mériter et faire preuve de ce qu'on annonce sur un curriculum vitae. Vous allez maintenant rentrer en phase projet de fin d'études.*

*Ensemble, vous, toutes et tous ensemble, il vous appartient de vous organiser de manière autonome pour réaliser quelque chose à partir d'une description qui retient l'intérêt d'un certain nombre de gens.*

*Je vous parlais de pouvoir commander une baguette sur Internet, on a fait des progrès. J'y ai un tout petit peu contribué, et ce sera bientôt votre tour de faire les choses. Non, ce ne sera pas un morceau de pain ou de gâteau, vous allez devoir faire en sorte que la machine aide les humains à reconnaître un peu les choses qui les entourent.*

*Sans doute, dans une prochaine génération, les machines auront encore gagné en fonctionnalités, peut-être seront-elles vraiment plus intelligentes, ou aussi intelligentes que le voudront les programmes qu'on aura mis dedans ... à ce jour, nos machines, nos infrastructures, peuvent réaliser des trucs un peu avancés quand même ... comme permettre de catégoriser un fonds d'image, d'extraire des données d'archives numérisées, ou dans un autre domaine de reconnaître des êtres vivants, ou des objets, et de leur associer des informations provenant de sources multiples et variées.*

*Vous le devinez, il s'agit de bâtir une solution qui vous mène au carrefour de ces choses dont on entend tellement parler et que si peu de gens savent expliquer. Encore moins savent comment faire pour structurer des solutions, des services basés sur de l'intelligence artificielle, des mégadonnées ou de l'apprentissage profond ... oui, là ça ne brille pas du même éclat que si je dis AI, big data, ... enfin bref, vous savez, ces mots qu'on prononce volontiers dans les salons ou aux fins des repas ... et bien c'est à votre tour de pouvoir, non pas les prononcer sauf si cela vous amuse, mais les faire vivre concrètement.*

*Voici la demande que nous recevons, et dont nous vous confions la mission de satisfaire la livraison.*

*Le travail est attendu comme le fruit d'une coopération entre tous les membres de la promotion, chacune et chacun devant y apporter sa libre et profitable contribution au bénéfice de tous les autres.*

# Projet Reconnaissance

## Demande

Le client ПолиРешения ООО installé à Пролетарский dans l'oblast de Пенза, souhaite créer un service de gestion documentaire permettant de réaliser différents travaux possibles de stockage de documents, d'analyse de contenu, aux fins de permettre de disposer d'une infrastructure commune pour des applications différentes :

- disposer des outils nécessaires pour traiter des images, captures diverses sous forme de fichiers graphiques, en faire le classement, et permettre d'analyser ces documents pour en reconnaître des objets, personnes, et des éléments de texte figurant sur ces documents.
- mettre cette infrastructure à disposition d'utilisateurs pouvant gérer leurs documents, avec la possibilité d'échanger ou partager les documents et informations entre utilisateurs d'un même groupe projet.

Le principe général de la solution doit répondre aux attentes d'utilisateurs travaillant dans un service d'archives documentaires, d'une part, mais également à ceux qui ont la mission d'identifier et gérer la circulation des personnels de l'organisation. Dans cette seconde perspective, la reconnaissance de contenu doit pouvoir répondre aux exigences d'une réponse sans délai de la plateforme.

## Synopsis

Proposer une solution d'infrastructure complète permettant d'offrir à des utilisateurs gérés de

- alimenter des données documentaires pouvant être liées à des objets ou personnes identifiées dans des images. Les données documentaires consistent à reconnaître les sujets selon des critères prédéfinis, ces critères pouvant évoluer par un module d'apprentissage. Les images peuvent être soumises à la solution par envoi dans un formulaire, dans le cadre d'une session sécurisée, chaque utilisateur pouvant disposer d'un stockage personnalisé, partageable.
- la plateforme doit pouvoir être activée comme un service en ligne, répondre à tous les standards adaptés de sécurité, permettre de traiter les besoins d'un nombre d'utilisateurs inscrits, et d'utilisateurs simultanés dont la démographie permet de définir la taille d'infrastructure à mettre en œuvre. La plateforme répond également aux exigences adaptées de maintien en condition opérationnelles et de continuité et reprise d'activité en cas de défaillance
- les éléments d'infrastructure doivent donner toutes les garanties de sécurité et de performances, adaptables selon les besoins en ressources.
- le redéploiement de l'infrastructure entière pour clonage complet ou remise en conditions d'exploitation sur un hébergement de secours doit être possible avec un temps de MCO inférieur à une heure.

## En détail

Plateforme proposant des outils de reconnaissance de textes et d'images, basée sur les technologies Tesseract et Tensorflow.

La solution intègre les outils fondamentaux et propose un service via le Web, de stockage et de reconnaissance d'images et de textes.

La partie Web présente les interfaces en front et backend. Les utilisateurs sont gérés par une gestion de compte, et une gestion des accès, avec tous les niveaux de sécurité adaptés.

Les utilisateurs peuvent exploiter un espace de stockage individualisé, avec ou quotas en fonction du type de souscription, et peuvent :

- récupérer les textes contenus dans des documents bitmaps de leur stock ou importés,
- faire des recherches sur des textes enregistrés au format texte,
- demander un classement d'images en fonction des contenus à partir de la reconnaissance de formes proposée avec l'environnement Tensorflow.

L'infrastructure proposée dispose d'un « plan données », un « plan calcul » et un « plan gestion ». Il est prévisible que les plans de travail informatique soient préférentiellement séparés, ou que l'architecture permette de les isoler en fonction de la dimension de l'infrastructure en production.

Les périmètres de données des utilisateurs sont cloisonnés, protégés, mais les utilisateurs peuvent affecter des droits d'accès par utilisateur et groupe d'appartenance dans l'annuaire centralisé.

La reconnaissance d'images peut utiliser les capacités intégrées de base avec les composants Tensorflow, mais un module apprentissage permet à chaque utilisateur autorisé d'enrichir la capacité de reconnaissance de la plateforme.

La reconnaissance d'images peut utiliser l'amélioration intelligente des images par rééchantillonnage automatique pour mieux catégoriser les contenus.

La reconnaissance de textes "imprimés" permet de travailler avec les capacités intégrées à Tesseract et peut bénéficier de l'amélioration intelligente pour un rendement approprié. Pour les textes de qualité manuscrite, les outils d'apprentissage peuvent contribuer à un rendement significatif de reconnaissance des textes.

## Vision produit

L'offre technique de la plateforme permet d'envisager des utilisations diverses et variées. Des profils d'utilisation peuvent alors être proposés pour rendre la plateforme commerciale.

Un profil, parmi d'autres, peut permettre de gérer une bibliothèque d'images présentant des animaux, ou des véhicules, et des documents classés en fonction de leur lien avec des

animaux, ou des véhicules, et de leur contenu. Cette catégorisation mentionnée ici est un exemple.

Un autre scénario d'utilisation pourrait être la reconnaissance de documents, d'empreintes, d'images faciales, ouvrant la possibilité de gérer des accès à des périmètres d'activité par reconnaissance face à des données répertoriées, dans le cadre d'une gestion rationalisée des personnes, ou d'une identification des personnes à la volée.

On parle également d'un scénario ayant une autre approche sur la reconnaissance de personnes, permettant d'établir un connecteur "plug-in" sur une solution commerciale, et qui permettrait d'identifier des clients référencés pour réaliser des opérations promotionnelles, du push de produits, à partir de règles métier, ou de processus programmables.

Ces scénarios sont des propositions ou suggestions permettant de démontrer la polyvalence de la solution autour de la reconnaissance d'images, de formes et de textes.

Dans une première version de la solution, qui doit montrer un périmètre opérationnel suffisant, il n'est pas demandé de prouver en détail toutes les fonctionnalités, et de démontrer la réalité, en exploitation, des scénarios mentionnés ci-dessus. L'infrastructure doit arriver au niveau exploitable, l'aboutissement de scénarios précis peut se satisfaire d'un niveau POC de tests et de démonstration.

Le livrable doit permettre, outre la démonstration technique et fonctionnelle, de diriger le travail restant à faire par une équipe à venir et qui réalisera une version 1 du produit.

## **Quelques explications techniques**

L'infrastructure doit respecter les standards habituels de sécurité. Elle doit être construite en suivant le plus possible la règle de "un serveur d'application pour un système d'exploitation" et contenir les machines sous forme virtuelle. En phase de prototypage l'infrastructure peut être montée sur la base d'un hyperviseur quelconque, dans une architecture réseau aboutie.

On peut imaginer, une fois déterminées les premières étapes et périmètres fonctionnels, la mise au point d'un démonstrateur simplifié sur la base de quelques VMs, qui devra toutefois permettre de valider les choix d'architecture retenus au final.

La plateforme finale de production qui tient lieu de livrable devrait être déployée sur une solution d'hébergement sous hyperviseur industriel, comme vSphere servi par des lames ESXi, en hébergement accessible depuis le Web. Cette partie peut être réalisée en s'appuyant sur l'infrastructure mise à disposition par l'AFORP.

On peut penser que l'architecture devrait permettre d'identifier les parties fonctionnelles essentielles que représentent :

- les moyens d'identification / authentification et gestion des accès
- le stockage des données

- les éléments de traitement et de calcul

Il n'est pas prédit que ce sera sous cette forme que vous choisirez de dessiner l'architecture, et de fabriquer la solution. Le choix vous en appartient.

Mais votre choix devra tenir compte des problématiques de sécurité, et des questions de performances. Ayez toujours à l'esprit ces deux aspects, dès la préparation de l'architecture, en retenant que la plateforme proposée in fine doit être modulable en fonction des performances supposées nécessaires.

Pour le stockage des données, outre les questions de volumétrie, peut se poser la question de savoir s'il faut retenir une solution ouvertement NoSQL, ou inversement full SQL ... ou les deux. La réponse, là encore, vous appartient, en fonction de ce que vous prévoyez de gérer comme données. Ce qui veut dire que vous devez imaginer un modèle de données avant toute chose ? C'est à vous d'en décider.

La partie de contrôle des accès est une partie, peut-être, la moins sujette à controverse : il y a juste à décider s'il est préférable de prévoir un environnement LDAP ou assimilable, ou pas. C'est-à-dire Active Directory ou une autre solution de gestion des accès et des droits d'accès.

Vous n'avez pour contrainte de choix de distribution d'exploitation de vos machines que la recherche des meilleures solutions. Visez la performance finale, et pas nécessairement l'optimisation irraisonnée des ressources consommées.

Ce qui veut dire que vous devrez avoir validé les différentes parties de votre architecture, validé et donc testé, et être capable de dire les ressources nécessaires en fonction de la charge.

Concernant l'infrastructure, pour finir, et c'est probablement la porte d'entrée pour la partie réalisation que vous allez entreprendre, vous devez penser une infrastructure construite sur, ou autour, d'un plan de sécurité, sur ou autour d'un modèle de données, sur ou autour d'un plan de flux de données et de code.

## **Quelques explications non techniques**

En conclusion de cette présentation, vous vous trouvez devant un projet qui peut vous paraître ressembler à une montagne. C'est normal, cela fait toujours cet effet.

Vous avez un calendrier de 6 semaines-projet prévisibles, qui se clôturent par une soutenance projet collective. 6 semaines, plus, selon les vocations, quelques minutes, heures par-ci par-là, mais vous êtes, le reste du temps, en activité, et avez une vie personnelle ... donc vous avez trente jours de production.

Ce sont, en pratique, un peu plus de deux cent heures ... ou encore douze mille minutes ... choisissez l'unité de temps qui vous convient pour la mettre à profit !

Vous avez été formés pour travailler en mode Agile, pour coopérer sur des travaux en commun, pour documenter votre travail, et vous êtes techniquement, avec des profils qui peuvent différer d'une personne à l'autre, techniquement assez bien préparés.

S'il vous reste des points de difficultés, des arbitrages pour lesquels vous souhaitez de l'aide, je suis là pour vous satisfaire autant que possible.

S'il s'avérait nécessaire de trouver un expert avec lequel vous passeriez du temps pour combler quelques lacunes, c'est aussi une chose qui peut être envisagée.

Vous êtes devant quelque chose qui va demander des efforts, réels, de la constance et une disponibilité accrue. Il a pourtant été convenu, sur la foi de l'expérience, que le temps dont vous disposez est largement suffisant pour faire bien le travail que nous pouvons espérer vous voir réaliser.

Vous devez donc aborder votre travail en commençant par son découpage, l'affectation des tâches, les missions ou fonctions à remplir pour chacune et chacun ... et toute cette organisation est confiée ... à vous-mêmes, ensemble. Ne pensez pas qu'il peut vous être utile de définir une hiérarchie. Il est utile de définir des compétences ... acquises pour les uns, à absorber pour les autres ... et celles et ceux qui font une chose particulière doivent le documenter, le commenter en réunion, de telle sorte que vous puissiez, au terme de cette période projet, être vous-mêmes assurés et rassurés de vos aptitudes et compétences.

Le projet est d'une dimension industrielle, à vocation commerciale, certaine. Il est d'une portée qui est telle que vous n'en ferez assurément pas le tour en six semaines. Il vous appartient de choisir ce qui sera réalisé, ce qui sera démontré, et ce qui ne sera pas montré, mis en œuvre, mais qui devra être correctement documenté pour que d'éventuels repreneurs du projet puissent le faire vivre et l'augmenter de fonctionnalités nouvelles.

Montrez-moi ce que vous savez faire.

## **A propos de méthode**

Il est recommandé de mettre en œuvre le projet dans un contexte pleinement Agile, en s'appuyant sur les concepts et préceptes régulièrement admis en environnement Scrum / Kanban, ou XP.

L'application suivie d'une approche ne signifie pas de maintenir précisément un rythme de sprint constant, c'est un point à arbitrer selon les besoins. Mais la planification, le découpage du travail, en phases, en features, la mise en place d'un backlog, les revues, réunions, sont des aspects qui doivent être visibles, quand bien même le travail se fait de manière adaptée par rapport à une méthode précise ou une autre.

Documenter le travail à faire, en cours et réalisé, en tous points est un élément clef de la réalisation.

Les outils de documentation et de suivi des activités doivent être présents, utilisés, et accessibles aussi bien à tous les membres de l'équipe qu'aux observateurs, parties prenantes.



Cette recommandation est suggérée. L'équipe est libre de s'organiser selon ses propres choix, l'attente « client » est de voir le travail avancer, dans les détails, de pouvoir consulter ce travail, de pouvoir intervenir en conseil et assistance, de pouvoir éventuellement tester les livrables aux étapes de rendu ou d'avoir, à minima des présentations et démonstrations éloquentes.

Les livrables attendus sont en principe :

- cahiers des charges
- maquettes d'architecture et prototypes fonctionnels
- version opérationnelle 0 de démonstration partielle en quatrième semaine projet
- version opérationnelle 1 de livraison en sixième semaine projet

Les maquettes et prototypes peuvent être présentés selon les ressources accessibles, sans documentation particulière, les versions opérationnelles doivent être présentées avec documentation générale, livret d'installation déploiement, manuel de l'utilisateur, manuel de l'administrateur, documents adaptés pour la programmation ou la personnalisation si approprié.

Les travaux doivent faire l'objet de comptes-rendus quotidiens et hebdomadaires qui sont partie intégrante du travail et du projet.

Le projet doit être documenté de manière à conserver en permanence la vision claire de ce qui est attendu, en cours de réalisation et réalisé, de telle sorte qu'il soit possible de redéployer l'ensemble du travail obtenu à date. Ce résultat doit permettre le transfert du projet à une équipe de reprise du projet dans le futur, pour finalisation ou extensions possibles.

Les documents techniques d'avancement doivent donc permettre de comprendre et reconnaître ce qui est réalisé et ce qui ne l'est pas en termes fonctionnels, et si possible les plans d'exécution à tenir pour compléter la réalisation.