**SAE501.VCOD - Analyse et conception d'un outil décisionnel**

Séquence 2 – Mener une veille technique

Thibault VIVIER

Table des matières

[Recherches Vidéos IoT : 3](#_Toc180751686)

[1- Kit de survie pour l'IoT façon DIY 3](#_Toc180751687)

[2- Premiers pas avec un microcontrôleur et Google Cloud IoT Core (Gautier MECHLING) 3](#_Toc180751688)

[3- La sécurité dans l'IoT difficultés, failles et contre-mesures (A. Duque) 4](#_Toc180751689)

[Vidéos Data : 4](#_Toc180751690)

[1- IOT, timeseries and prediction with Android, Cassandra and Spark by Amira Lakhal 4](#_Toc180751691)

[2- Le Big Data, de la récolte jusqu'à la production à grande échelle ! (Alexis Slawny & Mathias Kluba) 5](#_Toc180751692)

[3- Géopolitique de la data (Benjamin BAYART) 5](#_Toc180751693)

[Visionnages critiques 7](#_Toc180751694)

[Partie 1 : Vidéo IoT 7](#_Toc180751695)

[Les microcontrollers 7](#_Toc180751696)

[Le circuit 7](#_Toc180751697)

[Google Cloud IoT Core 7](#_Toc180751698)

[Partie 2 : Vidéo DATA 8](#_Toc180751699)

[Introduction : 8](#_Toc180751700)

[Cassandra 8](#_Toc180751701)

[Spark : 9](#_Toc180751702)

[Analyse des données : 9](#_Toc180751703)

[Bibliographie associé à la vidéo : 9](#_Toc180751704)

# Recherches Vidéos IoT :

### 1- Kit de survie pour l'IoT façon DIY

<https://www.youtube.com/watch?v=IMzgrpLqZmg>

#### Résumé

La présentation a pour but de sensibiliser le public sur la facilité de construction d’un réseau IoT. Le présentateur, Laurent HUET, aborde les différents points clé d’un réseau électronique en commençant notamment par la RasberryPI et d’autre microcontrollers équivalents. Il parle ensuite des coûts des différents composants et de leur rôle électronique (ex : les résistances permettent d’éviter une sur tension des autres composants). Il insiste sur la facilité de mise en place et montre ainsi que l’IoT, loin des laboratoires ou de demander des gros moyens financier peut être réalisé très facilement à une échelle locale sans un gros budget.

#### Justification du choix

J’ai choisis cette vidéo car il me semblait intéressant de découvrir les divers composant d’un réseau électronique basé sur RasberryPI (ou équivalent) car cela s’inscrit dans le cadre du projet. De plus, la vidéo m’a permis de comprendre que la mise en place de ces circuits est loin d’être inaccessible et que nous pouvions tout à fait être en mesure d’en mettre un en place après quelques heures de travail. Enfin cette vidéo est une bonne introduction à l’aspect pratique et technique de l’IoT.

### 2- Premiers pas avec un microcontrôleur et Google Cloud IoT Core (Gautier MECHLING) - visionnage critique

<https://www.youtube.com/watch?v=eYDb0X70HGM>

#### Résumé

Cette présentation Devoxx Fr nous permet d’avoir un aperçu des bases sur les microcontrollers et la récupération des données des capteurs. Gauthier MECHLING commence sa présentation en introduisant les microcontrollers, leurs différents composant et en expliquant leur principale fonction : faire tourner du code en continu sur un processeur indépendant. Après cela, il montre comment « flasher » un microcontroller (= intégration de code) en live. Il explique ensuite une partie de connectique et d’électronique en créant un capteur à son circuit et récupérant les données du capteur à l’aide du microcontroller. Enfin il envoie ces données sur un Cloud et s’en sert pour créer des visualisations en temps réel.

#### Justification du choix

J’ai décidé de choisir cette vidéo car elle me semblait essentielle pour comprendre les bases de l’IoT et des microcontroller. En effet, notre projet concerne la récupération de données à partir de capteurs et de microcontroller, cette vidéo semblait tout à fait adaptée à cette problématique. De plus, le présentateur appuie tous ses points théoriques avec un exemple pratique bien expliqué. Ce qui m’a le plus attiré dans cette vidéo était la mise en place des différentes connections (physiques ou numériques) entre les différents composant (microcontroller, capteur, Cloud) ainsi que le cheminement de la données dans ce circuit.

### 3- La sécurité dans l'IoT difficultés, failles et contre-mesures (A. Duque)

#### <https://www.youtube.com/watch?v=ThdFcdyqMik>

#### Résumé

Cette présentation Devoxx nous permet d’avoir un aperçu des concepts de la sécurité informatique appliquée à l’IoT. La présentation débute par une brève introduction de l’IoT et la sécurité, notamment en caractérisant des données sécurisées comme étant disponibles, intègre, et avec une confidentialité garantie. Alexis Duque introduit ensuite les dangers auquel est confronté l’IoT, comme l’environnement des objets connecté, la grande variété des appareils dans un réseau IoT et l’étendue du champ d’attaque de ces appareils (physique et numérique) entre autres. Il insiste sur le fait que les utilisateurs et les entreprises sont encore trop peu sensibilisés due à la nouveauté de l’IoT. Il détaille ensuite toutes les formes les plus courantes d’attaque informatique sur des objets connecté en donnant des exemples. Il cite notamment des attaques d’escalade des privilèges, d’injection de code et d’attaque par Bluetooth (Blue Born). Il insiste notamment sur le fait que c’est le rôle des développeurs d’être conscient des problématiques de sécurité dans l’IoT et que parfois, il suffirait d’être aussi prudent que sur le développement des applications web pour éviter certaine vulnérabilité.

#### Justification du choix

J’ai décidé de choisir cette vidéo car les problématiques de sécurité me semblaient pertinentes bien que cela ne soit pas forcement notre rôle. Etre sensible aux problématiques de sécurités est quelque chose d’important quel que soit le domaine d’application pour éviter les problèmes parfois mineurs mais souvent avec de grosses conséquences pour les individus et/ou les entreprises. De plus, nous vivons dans un monde où l’IoT est omniprésent, connaître les failles ou des moyens de se prémunir des attaques est toujours utiles que cela soit en tant que développeurs ou en tant qu’utilisateur.

# Vidéos Data :

### 1- IOT, timeseries and prediction with Android, Cassandra and Spark by Amira Lakhal – visionnage critique

<https://www.youtube.com/watch?v=1DV9Kdec0-A>

#### Résumé

Présentée par Amira Lakhal, cette présentation aborde divers sujets et exemples d’utilisation d’objet connectés notamment pour la collecte de données dans la course à pied. Elle présentera rapidement l’IoT et enchainera ensuite par la présentation de divers outils utilisé dans le traitement et l’analyse des données utilisés sur les données provenant de l’IoT. Elle s’attardera ensuite sur des méthodes d’analyse utilisées dans un cas concret : la classification des activités d’un utilisateur à partir des données obtenues depuis un smartphone.

#### Justification du choix

J’ai choisi cette vidéo car il me semblait intéressant d’aborder la question de l’intégration, du traitement et de l’analyse des données venant de l’IoT. Il est d’autant plus intéressant de découvrir les outils utilisés (qui sont pour la plupart des outils de BigData), et les méthodes d’analyse appliquées sur ces données – qui sont des méthodes vues en cours (datamining). Les aspects technique et pratique de la présentation sont d’autant plus intéressants car ils permettent de comprendre correctement les points théoriques détaillés dans le talk.

### 2- Le Big Data, de la récolte jusqu'à la production à grande échelle ! (Alexis Slawny & Mathias Kluba)

<https://www.youtube.com/watch?v=uqajB-7ghVM>

#### Résumé

Cette présentation sur le BigData commence par une brève introduction au concept et à ses domaines d’application (marketing, surveillance, santé, aéronautique, etc…). Ensuite les présentateurs détaillent comment fonctionne le BigData ou plutôt ses problématiques fondamentales à savoir le traitement d’un grand volume de données. Il explique les différentes solutions existantes : scalabilité verticale (augmentation des performances d’une machine) et horizontale (division des traitements de données sur plusieurs machines – cf Hyper convergence). Ils donnent ensuite un exemple pour expliquer ces concepts fondamentaux. Après cela, ils présentent les différentes technologies utilisées dans le BigData : **Hadoop, Spark, NoSQL**, et leur fonctionnement basique respectif. Ils insistent beaucoup sur la scalabilité horizontale et sur la répartition de l’exécution des traitements sur plusieurs machines (ressemble au multi-processing python mais sur plusieurs machines). Enfin, les présentateurs appuient leur propos avec divers exemples de traitement de données en Open Source (meilleure bière de Paris et algorithme de choix de la meilleure route commerciale dans un jeu vidéo).

#### Justification du choix

Cette vidéo m’a grandement intéressé car elle permet de revoir les fondamentaux de BigData et elle fait écho à des concepts vus en cours (multi-processing). De plus, elle est aussi intéressante dans le cadre de ce projet car au vu des données à traiter (très variées et volumineuses), avoir une connaissance basique du BigData semble essentiel. Enfin le dernier aspect qui m’a orienté est le fait que les applications BigData présentées ne soient pas restreintes à un seul domaine d’application mais semblent bel et bien utilisable dans de nombreux projets.

### 3- Géopolitique de la data (Benjamin BAYART)

#### <https://www.youtube.com/watch?v=EOOhYaGGArc>

#### Résumé

Ce *talk* Devoxx Fr introduit la notion de géopolitique des données et Benjamin BAYART explique comment la donnée occupe déjà une place centrale dans la géopolitique moderne. Il commence par expliquer les différents problèmes qui accompagnent l’automatisation et la digitalisation du monde moderne en s’appuyant principalement sur le concept de fatalité des décisions prise par la machine – quand un programme prend une décision, elle est irrévocable contrairement à celle d’un humain – et de maltraitance des données humaines. En effet, les données humaines, en étant soumises constamment à des traitements divers et variés (et fatals), sont maltraitées au point parfois qu’on oublie qu’il s’agisse d’humain. Il explique ensuite le concept de géopolitique, à savoir l’étude des rapports de force et domination entre les puissances. Enfin, il détaille la place des données en géopolitique. Benjamin BAYART introduit les différentes mesures politique et juridiques prises ces dernières années en Europe et explique leur importance. Le cadre juridique et politique qui encadre l’utilisation de ces données est donc d’une importance capitale car non seulement elle protège les individus mais ce cadre octroie un pouvoir politique et juridique des personnes/institutions qu’il protège. Les données deviennent alors un enjeu géopolitique et un levier de contrainte quand il s’agit d’exercer son pouvoir.

#### Justification du choix

J’ai choisis cette vidéo pour le thème DATA car elle me semblait importante et intéressante. Il s’agit plus ici d’une vidéo de veille que véritablement associée au projet. Mais en la mettant en parallèle avec toutes les autres vidéo, on se rend compte de l’importance capitale qu’occupe la donnée dans le monde moderne et à quel point être conscient des risques et enjeux liés au milieu de la données est d’une importance capitale en tant que professionnel ou simplement en tant que citoyen. Bien que peu d’aspect technique soit détaillé, cette vidéo reste tout à fait pertinente et elle ouvre des pistes de réflexions sur l’avenir.

# Visionnages critiques

## Partie 1 : Vidéo IoT

Premiers pas avec un microcontrôleur et Google Cloud IoT Core (Gautier MECHLING)

<https://www.youtube.com/watch?v=eYDb0X70HGM>

### Les microcontrollers

Un microcontroller est un circuit intégré qui intègre tous les éléments essentiels au fonctionnement d’un ordinateur, à savoir un processeur, une mémoire vive et des entrée IO (input – output) pour communiquer avec des composants externes. Il existe une multitude de microcontroller. Les plus récents ont des antennes qui permettent de communiquer des données via wifi et/ou Bluetooth. Certains microcontroller ont des ports USB qui permettent de « **flasher** » les microcontroller. « Flasher » un microcontroller est le fait d’injecter du code dans le microcontroller via un ordinateur. Le prix de ces microcontroller est en moyenne légèrement inférieur à 5€. D’autres fonctionnalités peuvent être intégrées au microcontroller comme des composants de chiffrement par exemple.

Il est possible de coder dans une multitude de langage de programmation sur les microcontroller comme le C, C++, JavaScript, Python, etc… Dans la vidéo, Gautier MECHLING va utiliser le langage Arduino. Le code des microcontrollers se compose d’une **méthode d’initialisation** suivi de l’appel d’une méthode de « **loop** » qui tourne en boucle sur le microcontroller. Il s’agit d’une structure de code relativement simple qui offre beaucoup de possibilité, notamment avec les capteurs.

### Le circuit

Le présentateur introduit ensuite un capteur qui permet de récupérer des données atmosphériques. Il montre le capteur mais relève les problèmes de connexion entre les capteurs et le microcontroller. Il va ensuite logiquement présenter divers outils pour réaliser des schémas et des plaques physiques afin d’avoir une meilleure visualisation de la connectique. Il prend un exemple simple dans sa présentation mais l’utilisation de ces outils peut devenir essentielle lorsque les circuits deviennent complexes. Ces divers outils sont :

* **Fritzing** -> réalisations des schémas
* Aizler -> création des cartes physique (3€ l’unité) pour la mise en place du circuit. Ces cartes sont créées à partir des fichiers zip exportés depuis Fritzing.

La réalisation des circuits peut tout à fait se faire avec une connectique plus basique (plaques et fils).

Après avoir connecté le capteur au microcontroller dans son circuit, Gauthier MECHLING récupère les données du capteur dans le microcontroller avec du code Arduino. Ces données brutes récupérées peuvent être intéressantes a exploitées, c’est pourquoi il va ensuite présenter un service Cloud qui permet de récupérer et d’exploiter ces données. Ce service Cloud se nomme Google IoT Core.

### Google Cloud IoT Core

Il s’agit d’un service Cloud qui facilite la connexion d’appareil disposant d’une connexion internet au Cloud Google. Cet accès au Cloud Google permet ensuite d’accéder à d’autre service (ex : **PubSub**). Les protocoles d’échange de données sont des protocoles classique – HTTP et **MQTT**. Après une brève présentation de Google Cloud IoT Core, Gauthier MECHLING connecte son circuit (via le microcontroller) au service Cloud. Pour connecter un microcontroller à Google IoT Core, il faut créer dans ce service un « **registery** » (= un ensemble d’appareil connectés à Google IoT Core). Il recommande d’utiliser le protocole MQTT plutôt que HTTP. Il faut ensuite créer l’appareil utilisé dans Google IoT Core pour pouvoir s’en servir. On connecte ensuite l’appareil physique et l’appareil virtuel avec **un combo clé privée – clé publique** qui permettent d’avoir un niveau optimal de sécurité. La création de ces clés et d’un **JSON web token** (standard d’échange de données), se font en ligne de commandes depuis le service Cloud. Il faut bien faire correspondre les clés publiques et le JSON web token pour pouvoir envoyer les données du microcontrollers au service Cloud. Après avoir connecté le microcontroller à Google Cloud IoT Core, il est possible de récupérer les données dans le client PubSub. Le présentateur récupère ensuite les données du client PubSub depuis un autre client nommé **FireBase**. Il exploite ainsi différents client du service Cloud pour offrir un graphique retranscrivant les données atmosphériques en temps réel. Enfin, il montre encore diverse possibilité en rajoutant des lunettes connectée au réseau pour afficher la température en temps réel sur ses lunettes. Enfin, il montre un dernier projet qui consiste a récupérer ses données biométrique via un capteur et un RasberryPI pour les envoyer sur le Cloud et les exploiter en temps réel.

## Partie 2 : Vidéo DATA

Vidéo : IOT, timeseries and prediction with Android, Cassandra and Spark by Amira Lakhal

<https://www.youtube.com/watch?v=1DV9Kdec0-A>

### Introduction :

Amira Lakhal est une développeuse française de l’entreprise Valtech. Elle présente une conférence à propos de l’IoT, des différents outils de collecte et d’analyse de données liés à ce sujet.

Elle commence par se présenter et présenter l’IoT en lui-même. L’internet des objets est un terme qui peut être associé au dépassement du nombre d’humain sur terre par le nombre d’appareil connecté. D’après un rapport scientifique (CISCO), ce phénomène a eu lieu vers 2008-2009. Depuis le nombre d’appareil connecté par personne n’a cessé d’augmenter. Aujourd’hui, il y a près de 50 milliard d’appareils connectés dans le monde. Tous ces appareils produisent ou traitent un grand volume de données quotidiennement. C’est d’ailleurs pour cela que nous sommes actuellement dans l’ère du BigData. En prenant comme exemple des données issues de la course à pied (distance, coordonnées, accélération, …), elle va présenter divers outils, leurs utilisation et leurs utilité dans le cycle de la data appliqué à l’IoT. Il s’agit principalement d’outil de BigData.

Cassandra :   
Créer par Facebook, c’est un SGBD NoSQL qui propose des BD vectorielles. Le langage SQL est utilisé pour créer les tables et intégrer les données dans la base.

Elle explique les différents problèmes auxquels il est possible d’être confronté dans le stockage des données dans les « **SSTable** » Cassandra (Sorted String Table – cf bibliographie). Elle évoque notamment les problématiques de compaction (perte de données historisée), l’utilisation des « **Time Series** » dans Cassandra et l’usage des « **partition key** » dans les clés primaires de ses tables de données. L’objectif de tous ces concept présentés est de nous sensibiliser à l’importance d’historiser les données, pour pouvoir les analyser ensuite, mais surtout nous sensibiliser aux problématiques de stockage d’un grand volume de données et aux bonne pratiques pour résoudre ces problème et éviter d’avoir de mauvaises surprises.

### Spark :

Outil Open Source programmé en Scala. C’est un Framework BigData qui facilite le **clustering**. Spark propose une diversité d’outil pour une partie du cycle de la donnée (streaming, traitement, machine learning et création de graphique par exemple). L’outil principal de Spark est **Spark Core API** qui permet le traitement de données via de multiples langages (R, Python, Java et Scala entre autres). Spark peut être utilisé avec diverses sources de données (MySQL, PostgreSQL, Cassandra, JSON, …). Amira Lakhal explique ensuite brièvement comment les données sont stockées dans Spark via les RDD (**Resilient Distributed Dataset** – cf bibliographie). Elle évoque aussi les principes d’exécutions basiques de Spark (cluster). Après cela, la présentatrice explique comment connecter les deux outils (Cassandra et Spark). Avant de présenter le streaming de données dans Spark.

### Analyse des données :

La dernière partie de la vidéo est consacrée à l’analyse des données collectées via les capteurs utilisés dans la course à pied (montre & smartphone). Amira Lakhal présente diverses analyses utilisée sur les données avec Spark :

* Régression logistique
* Arbre décisionnel

Puis elle présente l’algorithme de machine learning qu’elle utilise pour classifier ses données pour faire de la classification sur les différentes activités enregistrées par les capteurs. Elle présente Random Forest. Pour chacune de ces méthodes et algorithme, elle présente le code Spark associé (JAVA) et les résultats graphiques qu’elle a pu obtenir avec les Time Series construite plus tôt (Cassandra). Elle utilise son algorithme de classification pour comparer les données obtenues entre les différentes activités et extraire des comportements significatifs de chacune de ces activités. Elle s’attardera particulièrement sur **l’algorithme Random Forest** pour la classification et le machine learning en l’expliquant et en démontrant son implémentation.

### Bibliographie associé à la vidéo :

SSTable – Cassandra :

<https://www.youtube.com/watch?v=gSggVewYwoU>

Spark : RDD + Fondamentaux

<https://www.youtube.com/watch?v=nH6C9vqtyYU>