

Rapport d'analyse stratégique :

Introduction générale



Projet : Amazon Review Analysis

Auteur : Ismaël SYLLA

A propos de ce document

L'objectif de ce document est de comprendre le contexte stratégique, organisationnel et data d'Amazon, afin d'identifier ses leviers clés.

Cette analyse fournit les bases nécessaires à la conception de modèles analytiques et de contrôle de qualité des données pour les étapes suivantes.

Table des matières

1.	Introduction	4
1.1.	Présentation d'Amazon	4
2.	Analyse SWOT d'Amazon.....	5
3.	Activités et organisation d'Amazon.....	6
3.1.	Structure.....	6
3.2.	Services et catégories de produits	6
3.2.1.	E-commerce et marketplace	6
3.2.2.	Cloud Computing — AWS (Amazon Web Services)	6
3.2.3.	Streaming et divertissement	6
3.2.4.	Objets connectés et domotique	6
3.2.5.	Logistique et transport.....	7
3.2.6.	Publicité et services aux entreprises	7
3.2.7.	Santé et autres initiatives émergentes	7
4.	Flux de données chez Amazon.....	7
5.	Matrice de flux de données	10
5.1.	Matrice interface de design.....	10
5.2.	Matrice vue carrée	11
5.3.	Explication des interactions :	12
6.	Conclusion.....	12
7.	Références	13

1. Introduction

1.1. Présentation d'Amazon

Amazon est aujourd'hui l'un des acteurs les plus influents au monde dans les domaines du commerce en ligne, du cloud computing et plus largement de la technologie numérique. Fondée en 1994 à Seattle par Jeff Bezos, l'entreprise a débuté comme une librairie en ligne avant de se diversifier progressivement vers une multitude de produits et de services.

Grâce à une stratégie d'expansion continue, Amazon est devenue un écosystème global qui couvre désormais la vente de produits physiques et numériques, les services de diffusion en ligne (Prime Video, Amazon Music), le cloud computing (AWS), les objets connectés (Alexa, Ring), ainsi qu'un réseau logistique mondial de pointe.

En 2024, Amazon se classe 3^e au classement Interbrand des marques mondiales les plus valorisées, avec une valeur estimée à 298,1 milliards de dollars, en hausse de 8 % par rapport à l'année précédente. Cette performance illustre sa capacité à innover, à s'adapter aux évolutions du marché et à maintenir un haut niveau de confiance auprès des consommateurs et des entreprises.

2. Analyse SWOT d'Amazon

Forces	Faiblesses
<ul style="list-style-type: none">- 3^e marque mondiale selon Interbrand (2024).- Réseau logistique extrêmement performant permettant des livraisons rapides, même dans certaines zones rurales.- Présence dominante sur les marchés du e-commerce, du transport et du cloud computing avec AWS.- Capacités technologiques avancées (IA, NLP, big data) au service de la personnalisation et de l'efficacité opérationnelle.- Large base de clients fidèles et plateforme incontournable pour les achats en ligne.	<ul style="list-style-type: none">- Forte dépendance aux vendeurs tiers, ce qui peut affecter la qualité et la fiabilité des produits.- Problèmes de transparence perçus par les commerçants concernant les algorithmes de recommandation.- Risques juridiques et réglementaires liés à des condamnations et enquêtes antitrust.- Insatisfaction persistante sur certains délais de livraison, notamment en zones rurales.- Vulnérabilité aux cyberattaques et difficulté à maintenir une sécurité parfaite sur une infrastructure mondiale aussi vaste.- Complexité organisationnelle croissante, rendant la gestion globale plus difficile.
Opportunités	Menaces
<ul style="list-style-type: none">- Croissance continue dans les marchés émergents, où le e-commerce est en forte expansion.- Utilisation stratégique de l'IA et des LLM (via AWS).- Valorisation et analyse des retours clients pour améliorer l'expérience utilisateur et la qualité produit.- Partenariats marketing avec influenceurs et intégration aux réseaux sociaux pour toucher les jeunes générations.- Développement durable et innovation logistique (livraison verte, automatisation des entrepôts).- Fidélisation client et accroissement du panier moyen.	<ul style="list-style-type: none">- Concurrence accrue dans le retail et la tech (Alibaba, Walmart, Shein, etc.).- Pression réglementaire liée au respect du RGPD, du CCPA et aux procédures antitrust (FTC, Commission européenne).- Multiplication des cyberattaques et risques de fuite de données sensibles.- Risque de perte de confiance des consommateurs en cas de scandales ou de saturation du marché.- Perte de crédibilité des évaluations produits.- Marché très concurrentiel et en mutation rapide, nécessitant une innovation continue pour maintenir sa position dominante.

3. Activités et organisation d'Amazon

3.1. Structure

Amazon fonctionne à travers plusieurs entités interconnectées :

- **Amazon Marketplace** : plateforme où Amazon et les vendeurs tiers proposent des produits.
- **Amazon Prime** : service d'abonnement offrant la livraison rapide, le streaming vidéo et des offres exclusives.
- **Amazon Web Services (AWS)** : division cloud d'Amazon, supportant à la fois les services internes et externes.
- **Amazon Logistics** : réseau logistique interne gérant la préparation et la livraison des commandes.
- **Service client** : gestion des demandes, retours et réclamations des clients.

3.2. Services et catégories de produits

Amazon a diversifié ses activités bien au-delà du e-commerce. Ses services couvrent aujourd'hui plusieurs pôles stratégiques :

3.2.1. E-commerce et marketplace

- Vente de produits neufs et d'occasion via Amazon.com et les marketplaces locales (France, Allemagne, Japon, Inde, etc.).
- Gestion des vendeurs tiers via le programme FBA (Fulfillment by Amazon), qui prend en charge le stockage, l'expédition et le service client.

3.2.2. Cloud Computing — AWS (Amazon Web Services)

- Service leader mondial dans le cloud, proposant des solutions de stockage, calcul, bases de données, IA/ML, analyse de données, sécurité, et déploiement applicatif. AWS représente une part majeure de la rentabilité du groupe.

3.2.3. Streaming et divertissement

- Prime Video (films, séries, productions originales).
- Amazon Music (streaming musical).
- Twitch (plateforme de streaming en direct, principalement pour le gaming).
- Audible (livres audio et podcasts).

3.2.4. Objets connectés et domotique

- Écosystème Alexa (assistant vocal) et appareils connectés Echo, Fire TV, Kindle, Ring (vidéosurveillance et sécurité domestique), et Blink (caméras intelligentes).
- Développement de la maison connectée via des partenariats avec d'autres marques compatibles Alexa.

3.2.5. Logistique et transport

- Amazon Logistics : réseau mondial de livraison intégrée (camions, avions, entrepôts automatisés, drones expérimentaux).
- Services spécialisés : Amazon Prime Now, Amazon Fresh, Amazon Relay (transport B2B).

3.2.6. Publicité et services aux entreprises

- Amazon Ads : plateforme publicitaire interne pour les marques et vendeurs.
- Amazon Business : solutions B2B pour les entreprises et administrations.
- Amazon Pay : solution de paiement en ligne sécurisée intégrée à la plateforme.

3.2.7. Santé et autres initiatives émergentes

- Amazon Pharmacy et Amazon Clinic : vente et livraison de médicaments, téléconsultations.
- Amazon One : service d'identification biométrique (empreinte de la paume).
- Investissements dans la robotique, la recherche en IA, et les services de livraison autonome.

4. Flux de données chez Amazon

Amazon s'appuie sur une architecture data à grande échelle permettant de collecter, stocker, traiter et exploiter des volumes massifs de données provenant de multiples sources. Ce flux de données est conçu pour être scalable, résilient et sécurisé, afin de supporter aussi bien des usages opérationnels en temps réel que des analyses avancées et des applications d'intelligence artificielle.

Le schéma ci-dessous illustre les principales étapes du flux de données chez Amazon.

Sources de données

Les données proviennent de sources hétérogènes internes et externes à l'écosystème Amazon. Il s'agit notamment :

- Des données de ventes et de transactions (commandes, paiements, livraisons),
- Des logs techniques et événements système générés par les infrastructures et services,
- Des applications web et mobiles, incluant les interactions utilisateurs, la navigation ou les fonctionnalités applicatives.

Ces sources produisent des données structurées, semi-structurées ou non structurées, avec des volumes et des fréquences variables.

Collecte et ingestion des données

Les données sont ensuite collectées via des mécanismes d'ingestion en streaming ou en batch, selon les besoins métiers et techniques. Amazon utilise différents services et patterns d'ingestion tels que :

- Des plateformes de streaming comme AWS Kinesis ou Amazon MSK (Kafka),
- Des intégrations applicatives via API (pull/push),
- Des connecteurs de données et services SaaS via AWS AppFlow.

Cette couche assure la fiabilité, la scalabilité et la gestion des flux entrants.

Zone de stockage initial – Data Lake

Les données brutes sont stockées dans un data lake centralisé, basé principalement sur Amazon S3.

Cette zone conserve les données dans leur format d'origine, sans transformation majeure, afin de :

- Garantir la traçabilité,
- Permettre des re-traitements ultérieurs,
- Supporter différents cas d'usage analytiques et exploratoires.

Le data lake constitue le socle de l'architecture data d'Amazon.

Traitement et enrichissement des données

Les données stockées sont ensuite traitées et enrichies à l'aide de pipelines ETL / ELT. Cette étape comprend :

- Le nettoyage et la normalisation des données,
- L'agrégation et la consolidation de plusieurs sources,
- L'enrichissement par des données de référence ou des modèles analytiques.

Des services tels que AWS Glue, AWS Lambda, Apache Spark ou Flink sont utilisés pour le traitement distribué. Des modèles de Machine Learning, développés et déployés via Amazon SageMaker, permettent d'effectuer des analyses avancées, des prédictions ou des classifications.

Stockage transformé et optimisé

Les données transformées sont ensuite stockées dans des systèmes optimisés pour l'analyse, notamment :

- Des data warehouses comme Amazon Redshift,
- Des bases analytiques structurées et indexées.

Cette couche permet des requêtes analytiques performantes (OLAP) et sert de base aux usages décisionnels et applicatifs.

Consommation et exploitation des données

Les données sont enfin exploitées par différents consommateurs :

- Des outils de Business Intelligence et de reporting (par exemple Amazon QuickSight),
- Des dashboards et applications internes,

- Des services exposés via API, alimentant des systèmes de recommandation, d'optimisation ou d'aide à la décision.

Cette étape transforme les données en valeur métier concrète.

Gouvernance et sécurité

L'ensemble du flux de données est encadré par des mécanismes stricts de gouvernance, de sécurité et de conformité. Cela inclut :

- La gestion des accès et des rôles (IAM / RBAC),
- La gestion des métadonnées et la traçabilité des transformations,
- Le respect des réglementations (RGPD, audits, conformité interne).

Ces contrôles garantissent l'intégrité, la confidentialité et la fiabilité des données à chaque étape du cycle de vie.

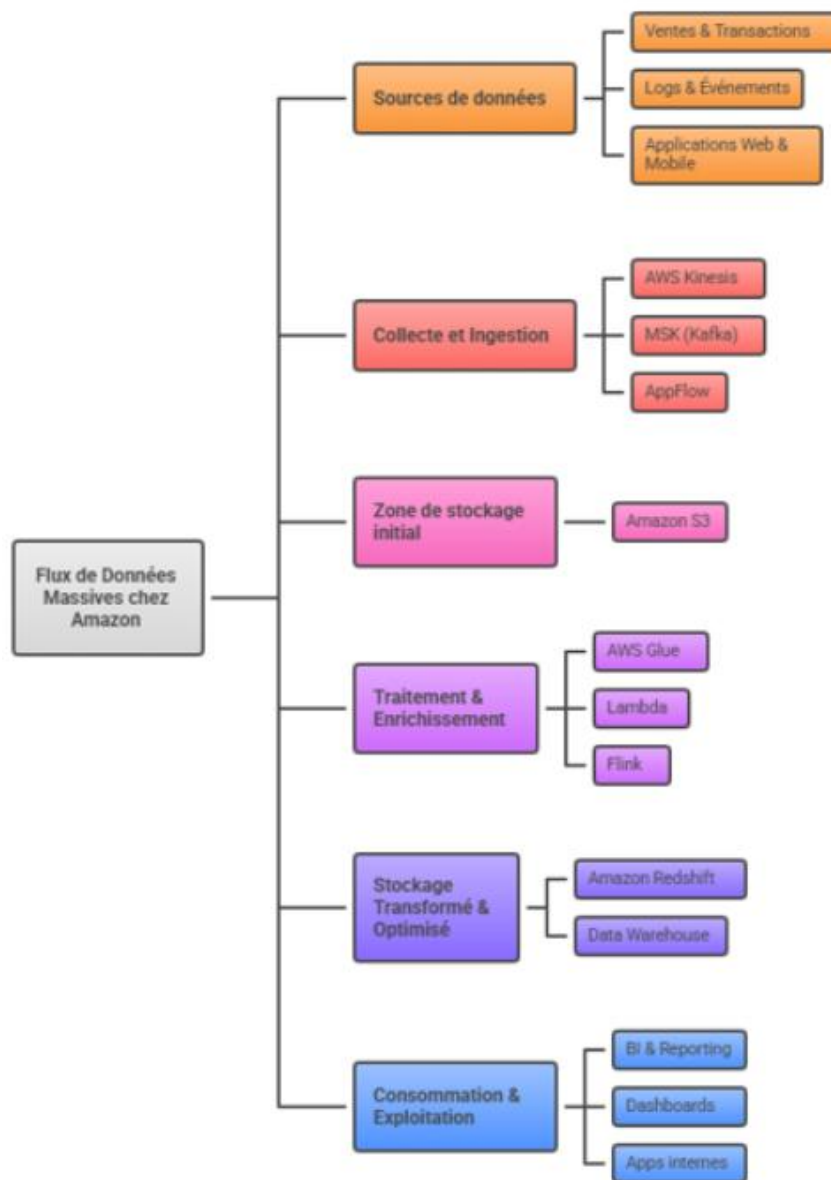


Illustration 1 : Cartographie des flux de données chez Amazon

5. Matrice de flux de données

5.1. Matrice interface de design

Étape	Source	Type de données	Destination	Objectif ou traitement effectué
Génération des données	Applications Web et Mobile, systèmes métiers, plateformes internes	Transactions, événements, logs, interactions utilisateurs	Systèmes sources Amazon	Création et émission des données opérationnelles
Collecte et ingestion	APIs, services applicatifs, flux événementiels	Données brutes (streaming / batch)	Couche d'ingestion (Kinesis, MSK, AppFlow)	Centralisation, fiabilisation et acheminement des flux de données
Stockage initial	Couche d'ingestion	Données brutes avec métadonnées	Data Lake (Amazon S3)	Stockage, historisation et conservation des données dans leur format d'origine
Traitement et enrichissement	Data Lake	Données structurées, semi-structurées, non structurées	Pipelines ETL / ELT (Glue, Spark, Lambda, Flink)	Nettoyage, normalisation, agrégation et enrichissement des données
Analyse avancée et ML	Données transformées	Features analytiques, indicateurs, prédictions	Plateformes ML et analytiques (SageMaker, moteurs analytiques)	Modélisation, prédiction, classification et analyses avancées
Stockage optimisé	Résultats des traitements	Données analytiques optimisées	Data Warehouse (Redshift), bases analytiques	Support des requêtes analytiques et décisionnelles
Consommation et exploitation	Données agrégées et optimisées	KPIs, rapports, flux de données	Outils BI, dashboards, applications internes, APIs	Aide à la décision, pilotage métier et services data-driven
Gouvernance et sécurité	Ensemble du cycle de vie des données	Métadonnées, logs d'accès, règles	Outils de gouvernance et de sécurité	Contrôle des accès, traçabilité, conformité réglementaire

5.2. Matrice vue carrée

Flux de données	Sources de données	Couche d'ingestion (APIs / Streaming)	Raw Data Lake	ETL / Qualité des données	Data Warehouse (stockage analytique)	Bases NoSQL	Modèles analytiques et ML	BI / Applications et Services
Sources de données		1						
Couche d'ingestion (APIs / Streaming)	1		2					
Raw Data Lake		2		3				
ETL / Qualité des données			3		4	5		
Data Warehouse				4			6	8
Bases NoSQL				5				
Modèles analytiques et ML					6			7
BI / Applications et Services					8		7	

	→ Pas de flux (même système)
	→ flux séquentiel (ex. Data Lake → NLP → Modèle)
	→ flux parallèle (ex. modèle alimente à la fois Data Warehouse et BI/application et services)

5.3. Explication des interactions :

1. Sources de données → Couche d'ingestion (APIs / Streaming) :
Les données sont générées par les applications web et mobiles, les systèmes métiers et les plateformes internes. Elles sont transmises via des APIs, des flux événementiels ou des mécanismes de streaming vers la couche d'ingestion.
2. Couche d'ingestion → Raw Data Lake :
Les données collectées sont ingérées et stockées dans le data lake au format brut. Cette étape garantit la conservation, l'historisation et la traçabilité des données, indépendamment de leur structure.
3. Raw Data Lake → ETL / Qualité des données :
Les données brutes sont extraites du data lake et soumises à des traitements de nettoyage, de normalisation et de contrôle qualité afin de les rendre exploitables pour les analyses.
4. ETL / Qualité des données → Data Warehouse :
Les données transformées et consolidées sont chargées dans le data warehouse afin de supporter les requêtes analytiques, les agrégations et les usages décisionnels.
5. ETL / Qualité des données → Bases NoSQL :
Les données non conformes, incomplètes ou nécessitant un traitement spécifique sont stockées dans des bases NoSQL, permettant la supervision, l'audit et la gestion des exceptions.
6. Data Warehouse → Modèles analytiques et Machine Learning
Les données stockées dans le data warehouse servent de base à l'exécution de modèles analytiques et de Machine Learning, destinés à produire des indicateurs avancés, des prédictions ou des classifications.
7. Modèles analytiques et Machine Learning → BI / Applications et Services
Les résultats produits par les modèles (scores, tendances, recommandations) alimentent les applications internes, les services exposés via API et les systèmes orientés utilisateur.
8. Data Warehouse → BI / Applications et Services
En parallèle, les données enrichies et agrégées du data warehouse sont directement consommées par les outils de Business Intelligence et les tableaux de bord pour l'analyse globale et le pilotage métier.

6. Conclusion

A l'issue de cette première étape, nous disposons désormais d'une vision claire des enjeux stratégiques, organisationnels et data d'Amazon.

Cette analyse, permet de mieux comprendre le rôle central des avis clients dans l'écosystème du géant du e-commerce, ainsi que la complexité du cycle de vie de ces données, depuis leur création par les consommateurs jusqu'à leur exploitation à des fins analytiques et décisionnelles.

7. Références

- Interbrand (May 9, 2024) [Best Global Brands](#)
- Department of Justice (May 9, 2023) [Amazon Agrees to Injunctive Relief and \\$25 Million Penalty](#)
- European Commission (2022) [Commission Accepts Commitments by Amazon](#)
- Federal Trade Commission (2023) [FTC Sues Amazon for Illegally Maintaining Monopoly Power](#)
- Federal Trade Commission (2023) [FTC Says Ring Employees Illegally Surveilled Customers](#)
- Financial Times (2020) [Amazon Deletes 20,000 Reviews After Evidence of Profits for Posts](#)
- Ashmore Grou (2018) [MC_10 May18_2.pdf \(ashmoregroup.com\)](#)
- Euro Monitor International (2023). [Significance of E-Commerce Investment in Emerging Markets – Euromonitor.com](#)
- [OECD business insights on emerging markets](#)
- [Strategies for Expanding into Emerging Markets with E-Commerce](#)
- Forbes Advison (2024) [35 E-Commerce Statistics of 2024 – Forbes Advisor](#)
- Statista (2024) [Chart: Amazon Maintains Cloud Lead as Microsoft Edges Closer | Statista](#)
- [Amazon Sharholder information](#)
- [Amazon's wikipedia page](#)
- [Overview of Amazon's IT archistecture](#)
- [Simplified data flow documentation](#)
- [Dataflow Matrix Tutorial](#)