De L'Informatique Décisionnelle à la décision "Immédiate"

Charly Clairmont • CTO, Synaltic

Agenda

De L'Informatique Décisionnelle à la décision "Immédiate"

Charly Clairmont

1. Comment nous avons démarré avec la modélisation dimensionnelle
2. Open Source & Business Intelligence
3. Agile & Business Intelligence
4. Rupture technologique
5. Data Lakehouse, flexibilité, simplicité
6. Décider "immédiatement"

Charly Clairmont

Synaltic / CTO

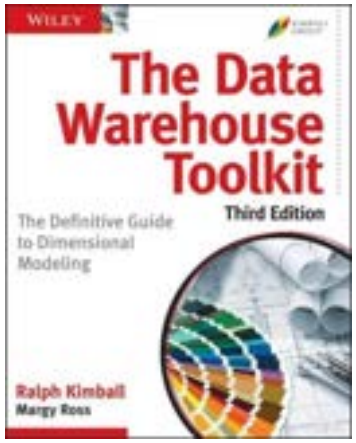
Charly Clairmont promeut la culture de la donnée, l'open source depuis 2004 tant sur le plan professionnel avec Synaltic qu'au travers des communautés Hadoop User Group France, Paris Spark Meetup, et désormais Modern Data Stack France.

Je contribue à Data-Asso principalement en tant qu'architecte.

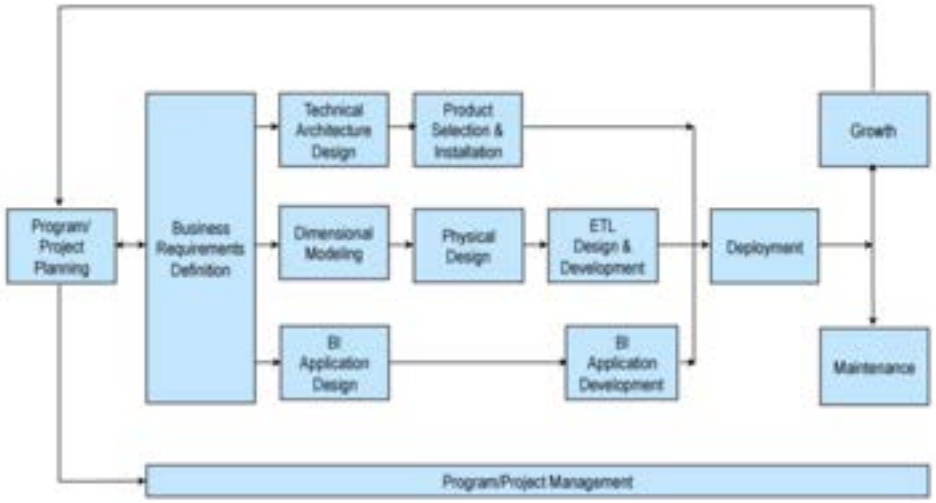
Comment nous avons démarré !

J'ai lu ces livres.

Et j'ai pratiqué.



Synaltic



2004 : Des Spécialistes de l'Innovation nous disent que l'**#InformatiqueDécisionnelle** en **#OpenSource** n'existe pas et n'existera pas !

Pourquoi l'Informatique Décisionnelle en Open Source

Synaltic

Donner aux utilisateurs métier l'opportunité de gérer les données par eux-mêmes.

Dans les années 2000 peu d'entreprises peuvent acheter des data warehouses et des plateformes analytiques comme Exadata, Teradata, Netezza…

Engineering Informatica 2004 :

- Le projet est plus important que le produit, mais les produits coûtent plus cher que les projets
- Les produits BI ne réduisent pas les coûts des projets
- Les produits BI ne sont généralement pas complètement exploités
- OSBI permet « d'expérimenter » avant de décider de mettre en place un projet BI parce qu'il n'implique pas de coût initial
- OSBI évite les problèmes de verrouillage
- OSBI s'intègre facilement dans la solution existante

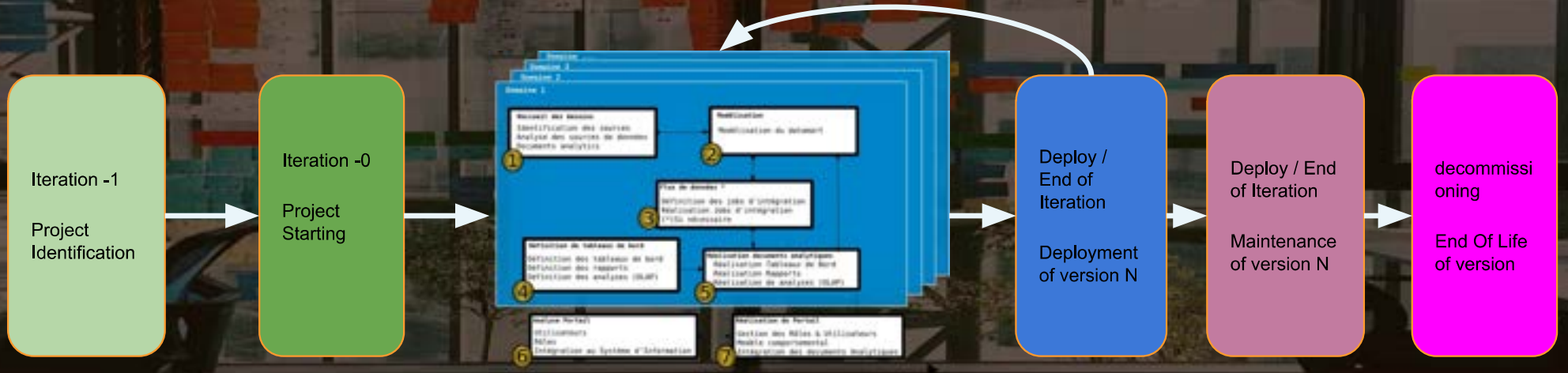
L'écosystème Open Source en Business Intelligence en 2007

Synaltic

Information Delivery				
ExoPortal, Liferay, JetSpeed, JackRabbit, XWiki, Lucene				
BI Applications & Tools : SpagoBI, JasperSoft, Penthao				
Dashboard	**Reporting**	**OLAP**	**GIS**	**Data Mining**
FreeChart	JasperReports, BIRT	Mondrian, Palo, JRubik	PostGIS, GeoKeetle, Talend Spatial, GeoMondrian	Weka, R, RapidMiner, Orange
Data Integration		**Data Management**		
Kettle, Talend, KETL, CloverETL, Octopus		MySQL, PostgreSQL, Ingres Icebreaker, BizGres		
System Software, and Hardware				
Quarz, Jbpm, Bonita, Spagic, Spago, Struts, Spring, Eclipse RCP Linux				

Embrace Agile Data

Le maître mot : s'adapter



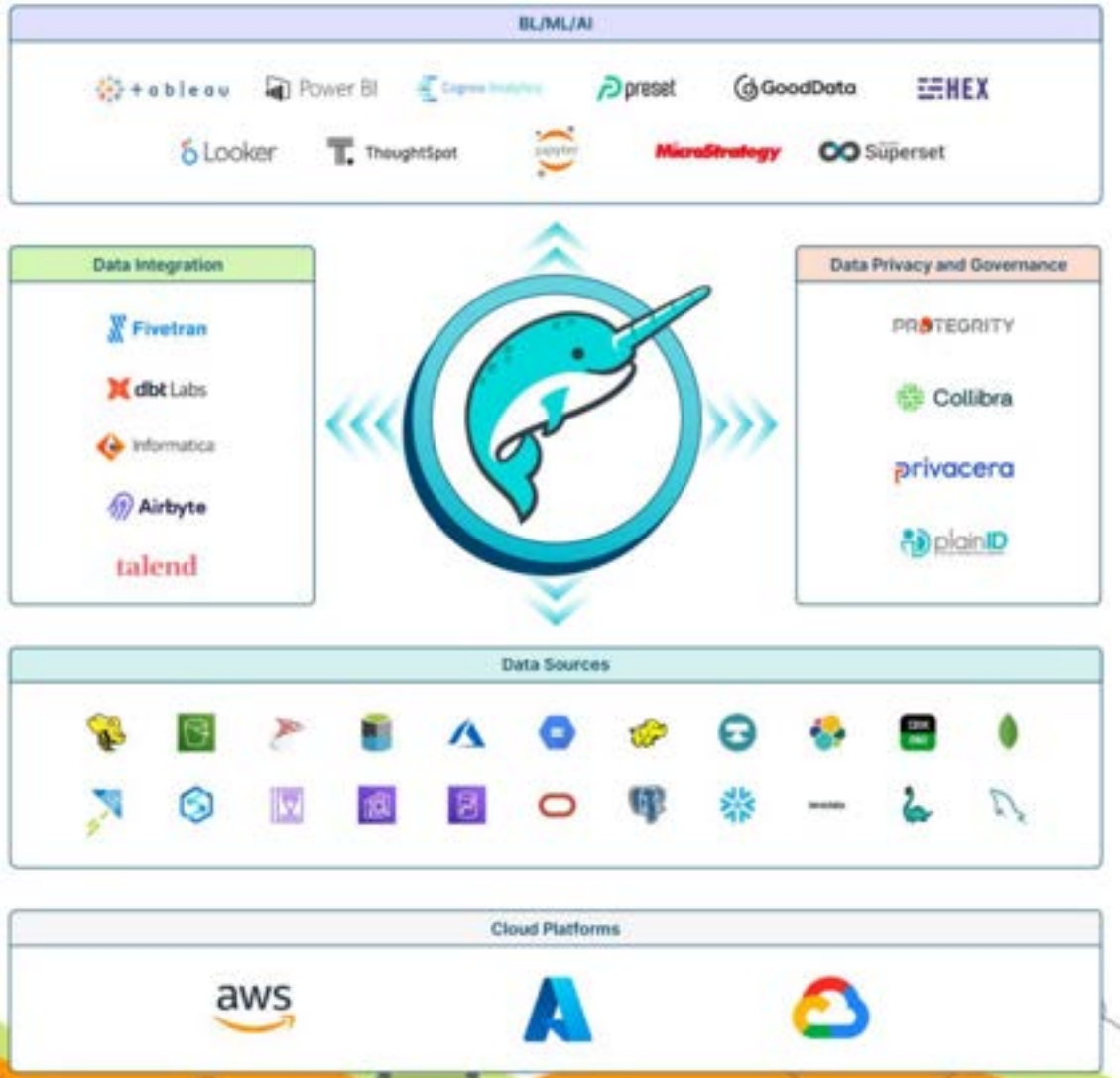
Garder le rythme

Big Data : Séparation du stockage, calcul, données
Déconstruction du Data Warehouse, Architecture Kappa, Passage à l'échelle

Cloud : Stockage flexible et peu cher

Data Driven : Data littératie, autonomisation (Data Mesh)

Que devient la modélisation dimensionnelle avec tout ça ?



KISSO

Kipt It Stupid Simple & Open

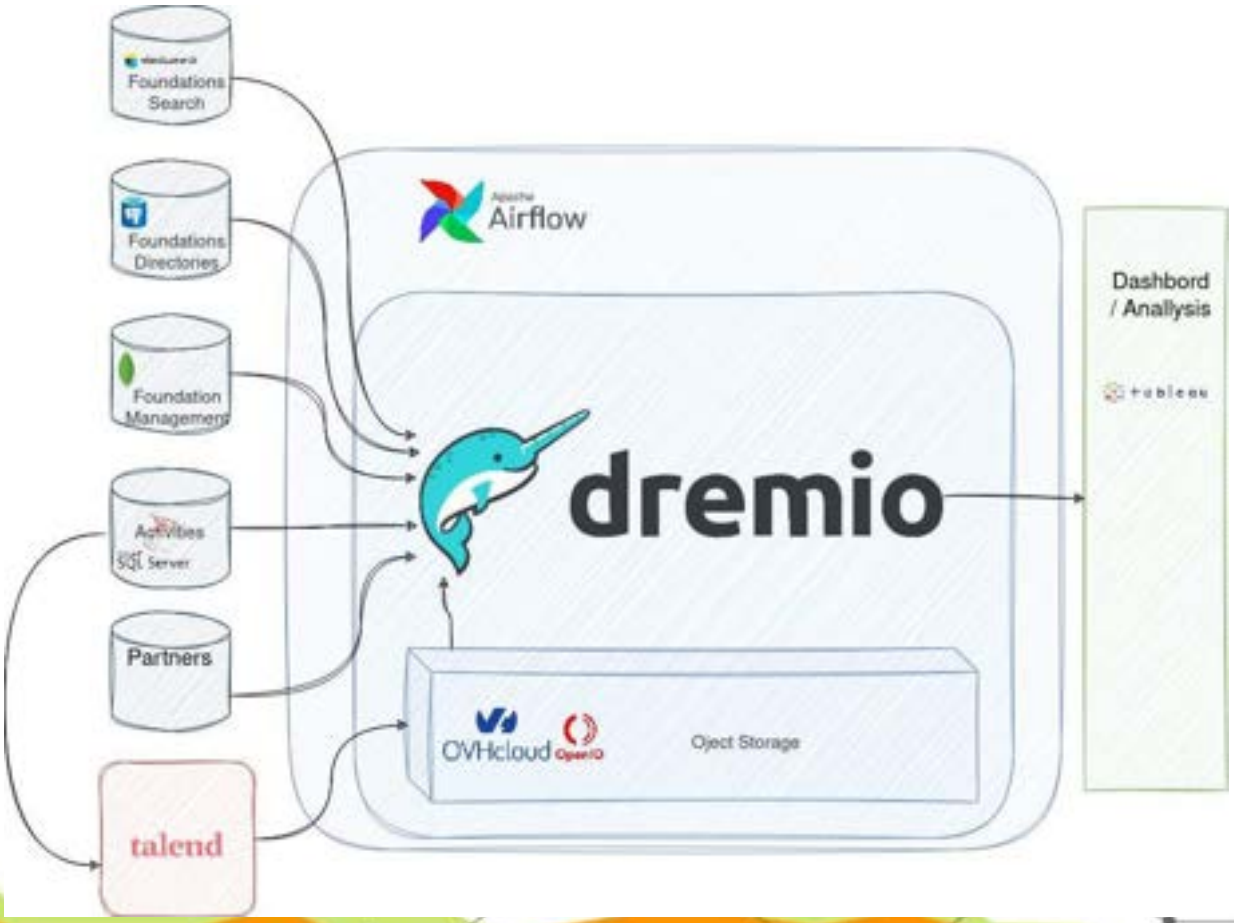
1. Connectez votre **Stockage Objet**,
2. Charger les données au **format ouvert**,
3. Apportez votre propre infrastructure,
4. Ajoutez vos **Couches Sémantiques**,
5. Montrez et partagez vos résultats.

Data Stack 1#

Un domaine par mois

1. Connectez SGBD PostgreSQL,
2. Ajoutez vos couches sémantiques,
3. Montrez et partagez vos résultats.
4. Former les utilisateurs clés
5. Partager des tableaux de bord
6. Démarrer un nouveau domaine

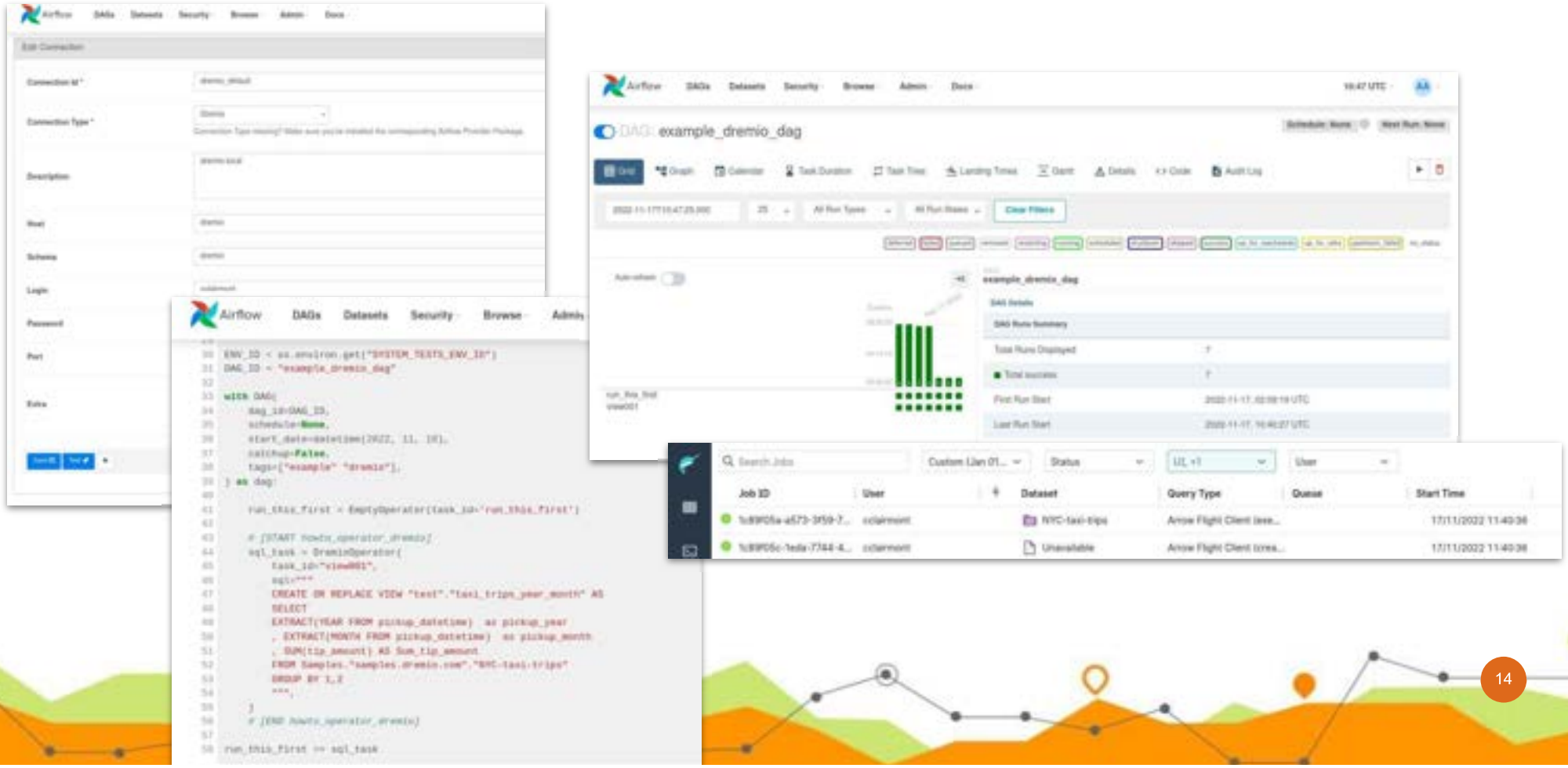
Data Stack 2#



Traiter l'historisation !

1. Toutes les sources de données connectées sur Dremio
2. Requêtes **partitionnées** CTAS pour gérer les données historiques
3. Pipelines gérés par Talend, Airflow

Data Stack 2# / Apache Airflow : Dremio Operator



Data Stack 3#

Décider Immédiatement

1. Collectez des données **en temps quasi réel CDC** avec **Debezium**
2. Transporter des données avec **Apache Kafka**
3. Partagez-le en tant que table **Apache Iceberg** **Kafka-Connect-Iceberg**
4. Interroger la table Iceberg avec **Dremio**

Voilà !

Merci !

Charly Clairmont
@egwada
www.synaltic.fr