

Archi Data @  lesfurets



Google Cloud Platform

Chafik NOUIRA, LesFurets.com  
tp-bigdata@lesfurets.com

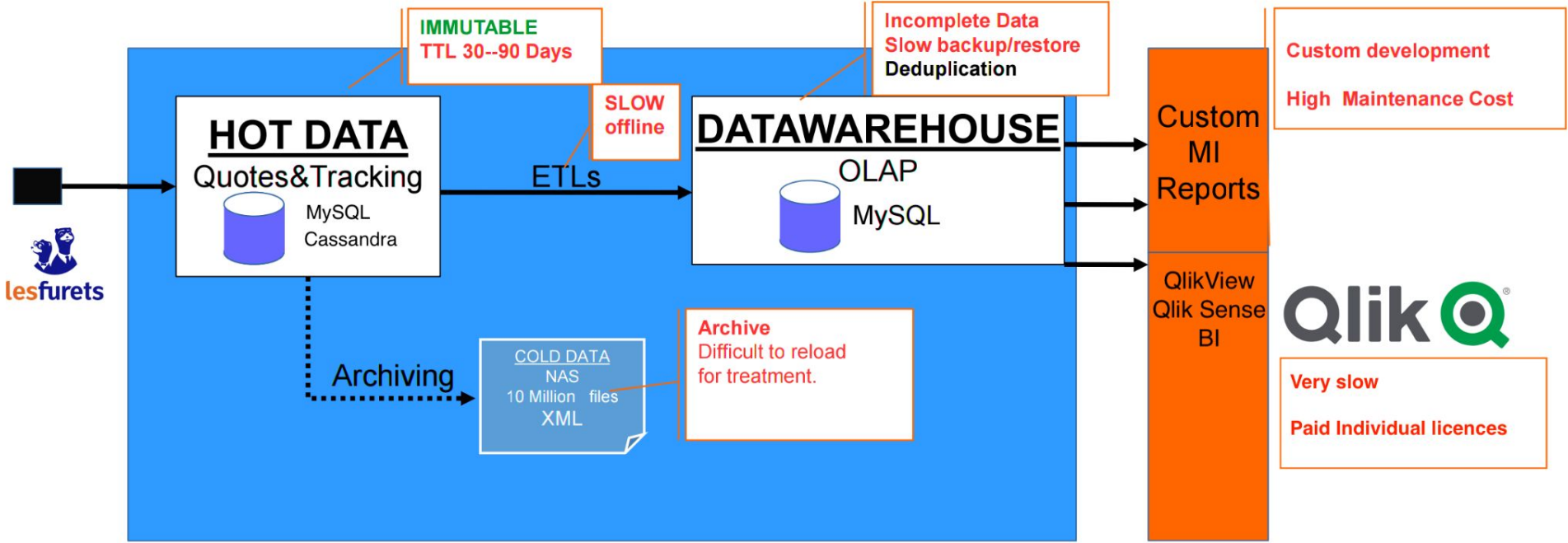
# Plan

1. Intro
2. Data @ lesfurets
  - 2.1. Forms
  - 2.2. Tracking
  - 2.3. Reports
3. GCP - Data Services
4. Data Platform @ lesfurets
  - 4.1. Live Data
  - 4.2. Data Lake
  - 4.3. Data Warehouse
  - 4.4. Analytics & BI
  - 4.5. Global Data Architecture
  - 4.6. Limits

# Intro

- Before Cloud Migration
  - Private Cloud on OVH : MariaDB, Cassandra
  - On premise : Jenkins, Jira, Confluence, Gitlab, Qlik Sense (BI)
  - Monolithic application : B2C, BackOffice, Data...
  - One release per day (but not on fridays)

# Data Architecture before Cloud Migration



# Intro


- Before Cloud Migration
  - Private Cloud on OVH : MariaDB, Cassandra
  - On premise : Jenkins, Jira, Confluence, Gitlab, Qlik Sense (BI)
  - Monolithic application : B2C, BackOffice, Data...
  - One release per day (but not on fridays)
  
- After Cloud Migration (May 2021)
  - Public Cloud on GCP
  - Jira Cloud, Confluence Cloud, GitlabCI
  - Less monolithic application : Data Platform separated from the rest
  - Several releases per day (but still not on fridays)
  - Data releases isolated from other releases

# Plan

1. Intro
- 2. Data @ lesfurets**
  - 2.1. Forms**
  - 2.2. Tracking
  - 2.3. Reports
3. GCP - Data Services
4. Data Platform @ lesfurets
  - 4.1. Live Data
  - 4.2. Data Lake
  - 4.3. Data Warehouse
  - 4.4. Analytics & BI
  - 4.5. Limits

# Forms

Votre voiture



**Bonjour Ahmed, content de vous revoir !**  
Votre formulaire est pré-complété suite à votre dernière visite.

**Quelle voiture souhaitez-vous assurer ?**

Ma voiture actuelle

Une voiture que je souhaite acheter

**Recherchez votre voiture...**


Recherchez par plaque d'immatriculation pour gagner du temps et répondre plus facilement.

Par plaque d'immatriculation

Par marque/modèle

Vous avez une voiture sans permis ? [Cliquez ici](#)

**Votre véhicule** ?



**PEUGEOT 208 II**  
1.2 PURETECH 100 ACTIVE  
Février 2020  
Essence / Manuelle / Berline 5 portes / 5 cv

[Modifier votre voiture](#)

**Quand l'avez-vous achetée ?** ?

Juin ▼ 2020 ▼

[Continuer](#)

**Quel est votre prénom ?**

Hervé

**Quel est votre nom ?**

Dupont

**Quelle est votre adresse email ?**

Votre adresse électronique nous permettra de vous envoyer le récapitulatif de votre comparaison par email et sera communiquée uniquement aux assureurs avec lesquels vous souhaitez être mis en relation.

herve.dupont@lesfurets.com

**Quel est votre numéro de téléphone ?**

Votre numéro de téléphone sera communiqué uniquement aux assureurs avec lesquels vous souhaitez être mis en relation.

06 12 34 56 78

**Accepteriez-vous d'être contacté par l'un de nos partenaires proposant le meilleur rapport qualité-prix en lien avec vos besoins ?**

Oui  Non

J'accepte les **Conditions Générales d'Utilisation** et d'être contacté par nos partenaires assurance auto.

J'accepte de recevoir des offres personnalisées de lesfurets.

lesfurets traite les données personnelles recueillies dans ce questionnaire pour proposer des offres correspondant au profil du navigateur, réaliser des opérations d'analyse et de statistiques, pour assurer la relation commerciale avec les clients et réaliser de la prospection commerciale par courrier électronique. Vous disposez de différents droits (notamment le droit d'opposition, d'accès, de rectification, d'effacement). Pour en savoir plus, consultez notre [Charte de protection des données personnelles](#).

[Découvrir mes offres](#)




# Forms

Hervé, voici les 30 meilleures offres pour votre PEUGEOT 208 II, classées par prix croissants et de façon impartiale.

Affichage      Formule      Prix

Simple     Détaillé     Tiers     Vol & incendie     Tous risques     Par an     Par mois    [Plus de filtres](#)

↩️ Nouveauté ! Vous pouvez comparer jusqu'à 3 offres entre elles

 <input type="checkbox"/> Comparer les offres	<b>398 € /an</b> Frais de dossier gratuits <a href="#">Plus de détails</a>	<a href="#">Poursuivre sur le site de l'assureur</a> <a href="#">Être rappelé par l'assureur →</a> <a href="#">Recevoir un devis par mail →</a>
 <input type="checkbox"/> Comparer les offres	<b>408 € /an</b> Frais de dossier gratuits <a href="#">Plus de détails</a>	<a href="#">Poursuivre sur le site de l'assureur</a> <a href="#">Être rappelé par l'assureur →</a> <a href="#">Recevoir un devis par mail →</a>
 <a href="#">Réseau d'agences</a> <input type="checkbox"/> Comparer les offres	<b>419 € /an</b> Frais de dossier gratuits <a href="#">Plus de détails</a>	<a href="#">Poursuivre sur le site de l'assureur</a> <a href="#">Être rappelé par l'assureur →</a>



# Forms

- Insurance products forms Data (car, bike, home, health...)
- Personal information
- Quotations
- Leads

# Plan

1. Intro
- 2. Data @ lesfurets**
  - 2.1. Forms
  - 2.2. Tracking**
  - 2.3. Reports
3. GCP - Data Services
4. Data Platform @ lesfurets
  - 4.1. Live Data
  - 4.2. Data Lake
  - 4.3. Data Warehouse
  - 4.4. Analytics & BI
  - 4.5. Limits

# Tracking

- User Uid (expires after 90 days of inactivity)
- Visit Uid (expires after 2 hours of inactivity)
- IP Address (helps to detect some robots)
- User Agent (helps to detect some robots)
- Forms logs (help to debug forms)

# Plan

1. Intro
- 2. Data @ lesfurets**
  - 2.1. Forms
  - 2.2. Tracking
  - 2.3. Reports**
3. GCP - Data Services
4. Data Platform @ lesfurets
  - 4.1. Live Data
  - 4.2. Data Lake
  - 4.3. Data Warehouse
  - 4.4. Analytics & BI
  - 4.5. Global Data Architecture
  - 4.6. Limits

# Reports

- Business reports
- Partners reports
- XLSX, CSV and Dashboards (Qlik Sense)

# Plan

1. Intro
2. Data @ lesfurets
  - 2.1. Forms
  - 2.2. Tracking
  - 2.3. Reports
- 3. GCP - Data Services**
4. Data Platform @ lesfurets
  - 4.1. Live Data
  - 4.2. Data Lake
  - 4.3. Data Warehouse
  - 4.4. Analytics & BI
  - 4.5. Limits

# GCP - Data Services



Cloud SQL

- Relational database service for MySQL, PostgreSQL, and SQL Server
- Fully managed
- Automation of backups, replication, encryption patches and storage capacity increases
- Secure & Scalable

# GCP - Data Services



**Cloud Firestore**

- Serverless NoSQL document database
- Fully managed
- Automatic multi-region replication
- Integration with Google Cloud services like Cloud Functions and BigQuery
- Customizable security and data validation rules to ensure the data is always protected



# GCP - Data Services



**Cloud Storage**

- Object storage (files and their metadata)
- Object lifecycle management
- Object versioning
- Turbo replication (15 minutes)
- Encryption
- Object and bucket-level permissions

# GCP - Data Services



**BigQuery**

- Serverless data warehouse
- Fully managed
- Highly scalable (petabytes)
- Built-in ML and AI integrations
- Standard SQL
- Integration with Cloud Storage, Cloud SQL, Bigtable, Google Drive

# GCP - Data Services



BigQuery

## Best practices

- Avoid SELECT \*
- Use Preview to explore data
- Use the query validator
- Use partitioning and clustering
- Do NOT use LIMIT
- Use nested/repeated fields and arrays
- [bigquery/docs/best-practices-costs](https://cloud.google.com/bigquery/docs/best-practices-costs)

# GCP - Data Services



**Pub/Sub**

- Fully managed real-time messaging service
- Real-time event distribution
- Parallel processing
- Integration with Dataflow, Cloud Composer, Cloud Functions and other services
- Access control using IAM

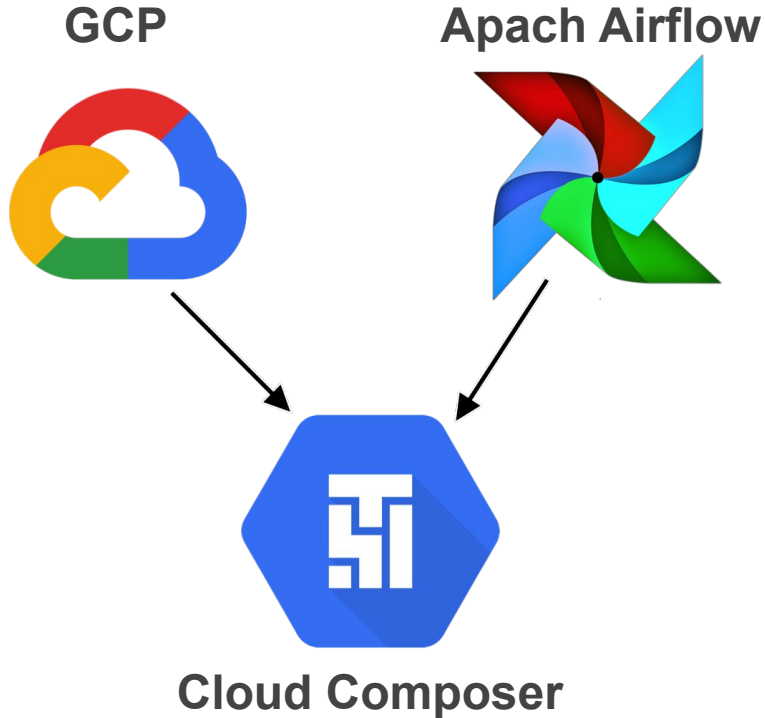
# GCP - Data Services



**Dataflow**

- Serverless stream and batch data processing
- Vertical and horizontal autoscaling
- Dataflow templates
- Inline monitoring

# GCP - Data Services



- Workflow orchestration service
- Built on Apache Airflow (management platform for data pipelines created by Airbnb)
- Integrates with other GCP products (BigQuery, Dataflow, GCS, Pub/Sub...)
- Supports hybrid and multi-cloud
- Uses Python programming language


# GCP - Data Services



**Data Studio**

- Free interactive data visualization tool
- Creating/Sharing/Scheduling reports (various charts and graphs)
- Integrates many data sources (BigQuery, Google Sheets, Cloud SQL, GCS...)
- BI Engine (used with BigQuery)
- Easier to use than other tools like Google Analytics

# GCP - Other services

- **Terraform (not GCP)** - Infrastructure as code tool
- **Cloud Functions** - Functions as a service (event-driven and autoscaling)
- **Compute Engine** - Virtual machines
- **Identity and Access Management (IAM)** - Permissions management system
- **Cloud GPUs** - Useful to mine Bitcoins 
- Many other GCP services here <https://cloud.google.com/products>
- Third-party services in the Marketplace <https://cloud.google.com/marketplace>



# Plan

1. Intro
2. Data @ lesfurets
  - 2.1. Forms
  - 2.2. Tracking
  - 2.3. Reports
3. GCP - Data Services
- 4. Data Platform @ lesfurets**
  - 4.1. Live Data**
  - 4.2. Data Lake
  - 4.3. Data Warehouse
  - 4.4. Analytics & BI
  - 4.5. Limits

# Live Data

Compute Engine



Application Web

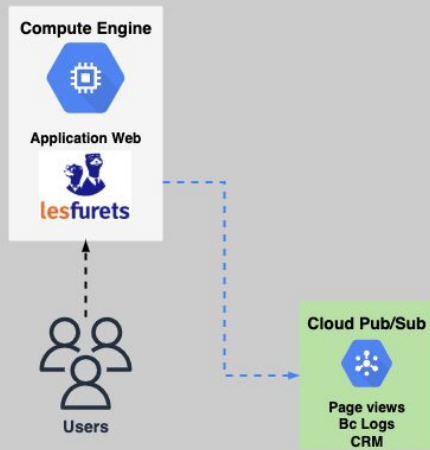


lesfurets

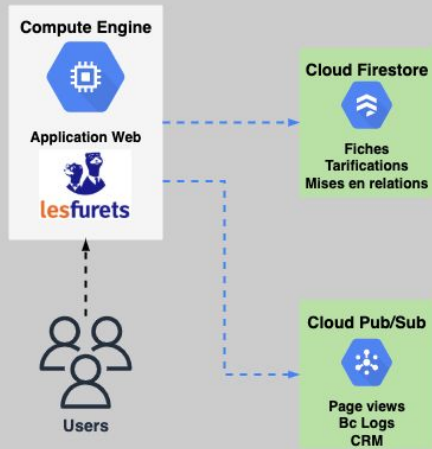
# Live Data



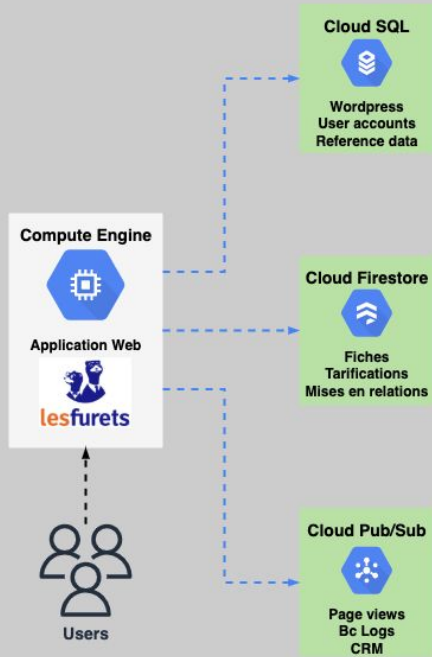
# Live Data



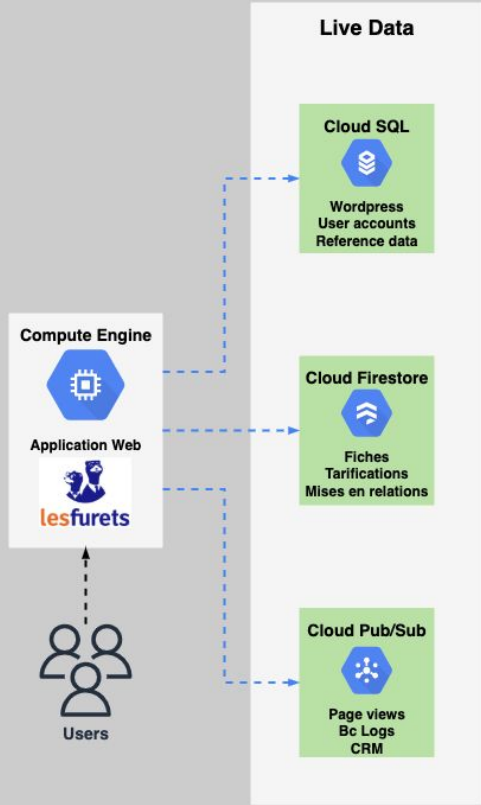
# Live Data



# Live Data



# Live Data



# Plan

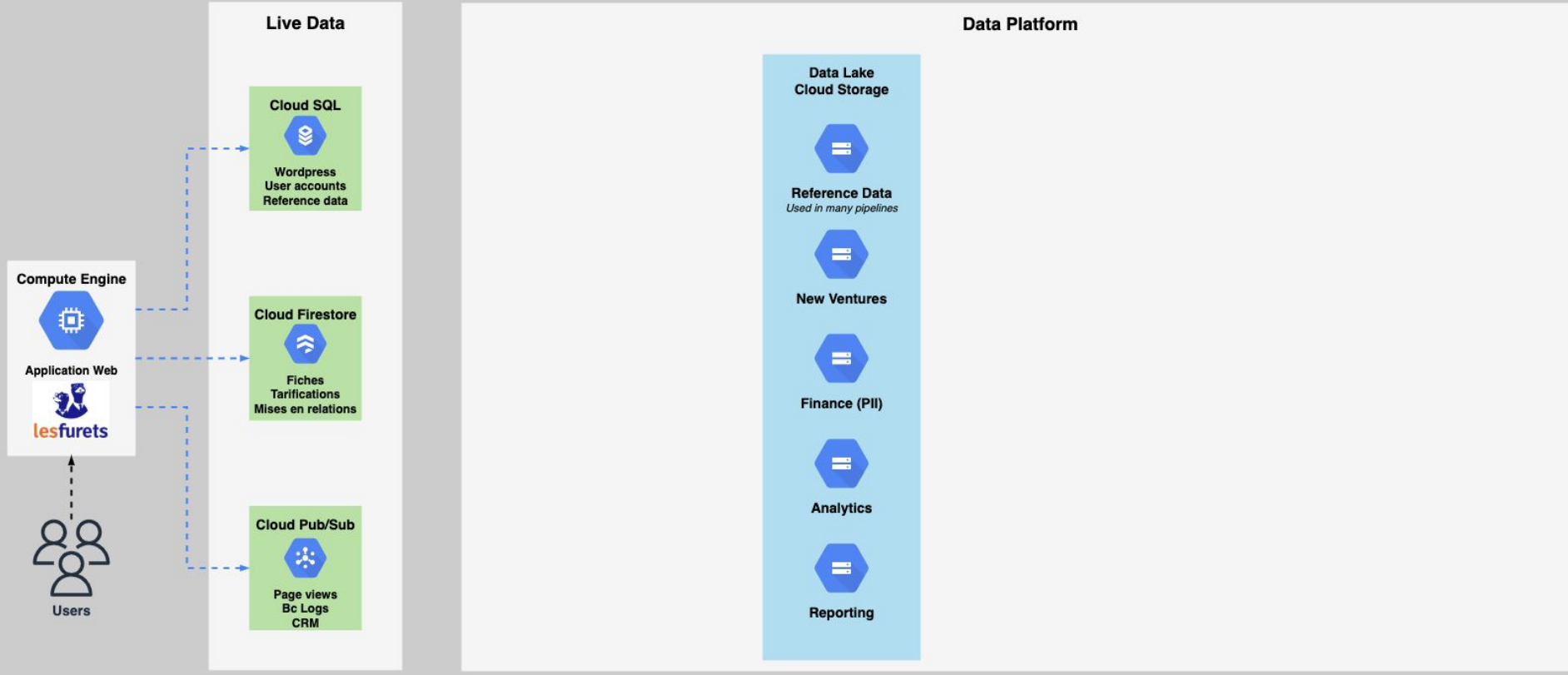
1. Intro
2. Data @ lesfurets
  - 2.1. Forms
  - 2.2. Tracking
  - 2.3. Reports
3. GCP - Data Services
- 4. Data Platform @ lesfurets**
  - 4.1. Live Data
  - 4.2. Data Lake**
  - 4.3. Data Warehouse
  - 4.4. Analytics & BI
  - 4.5. Limits



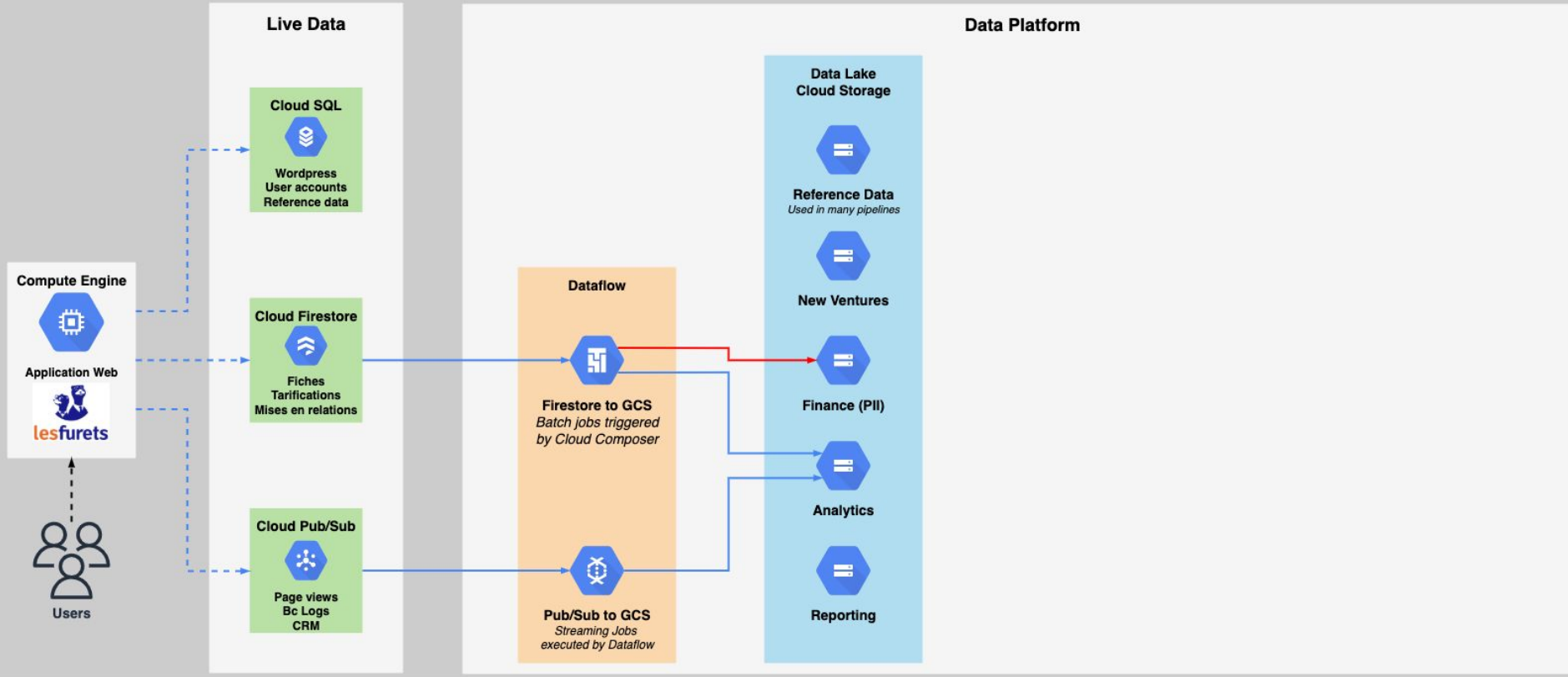
# Data Lake



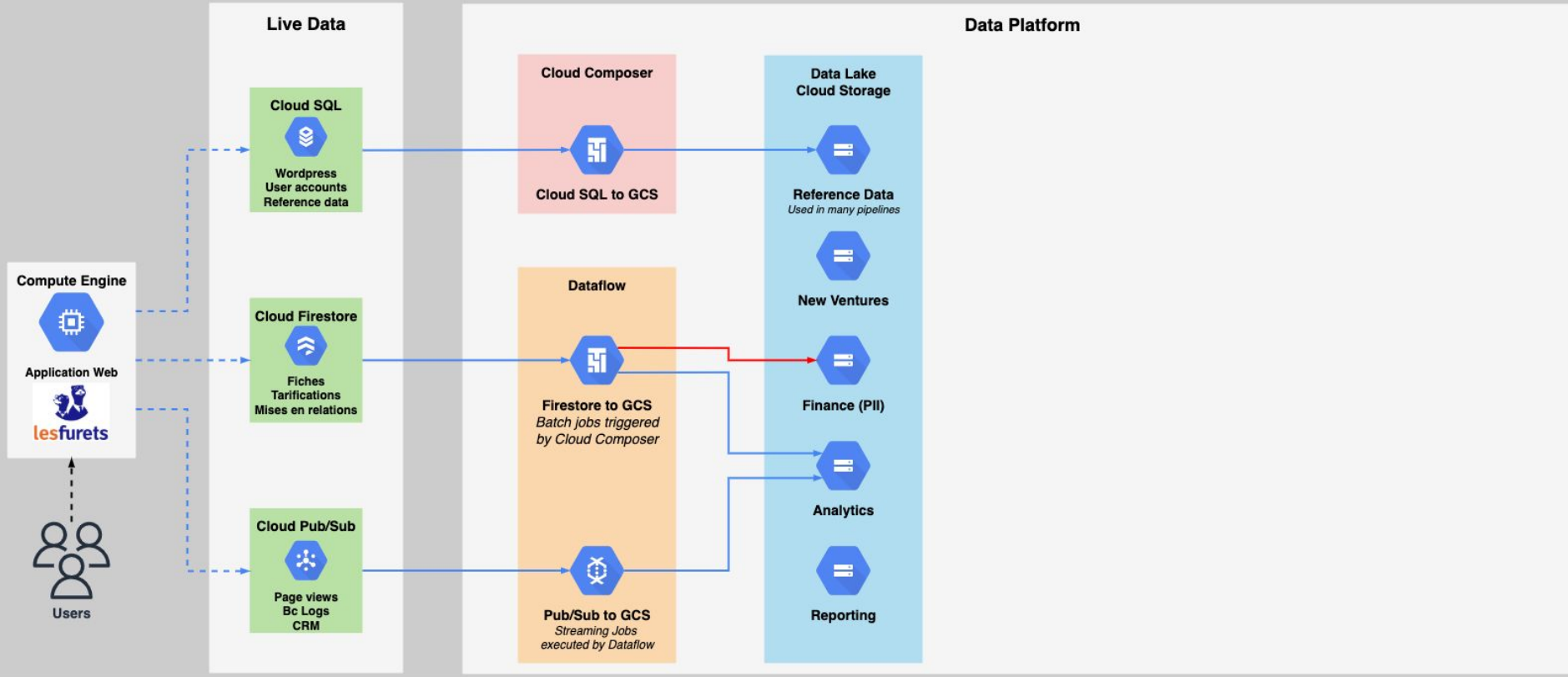
# Data Lake



# Data Lake



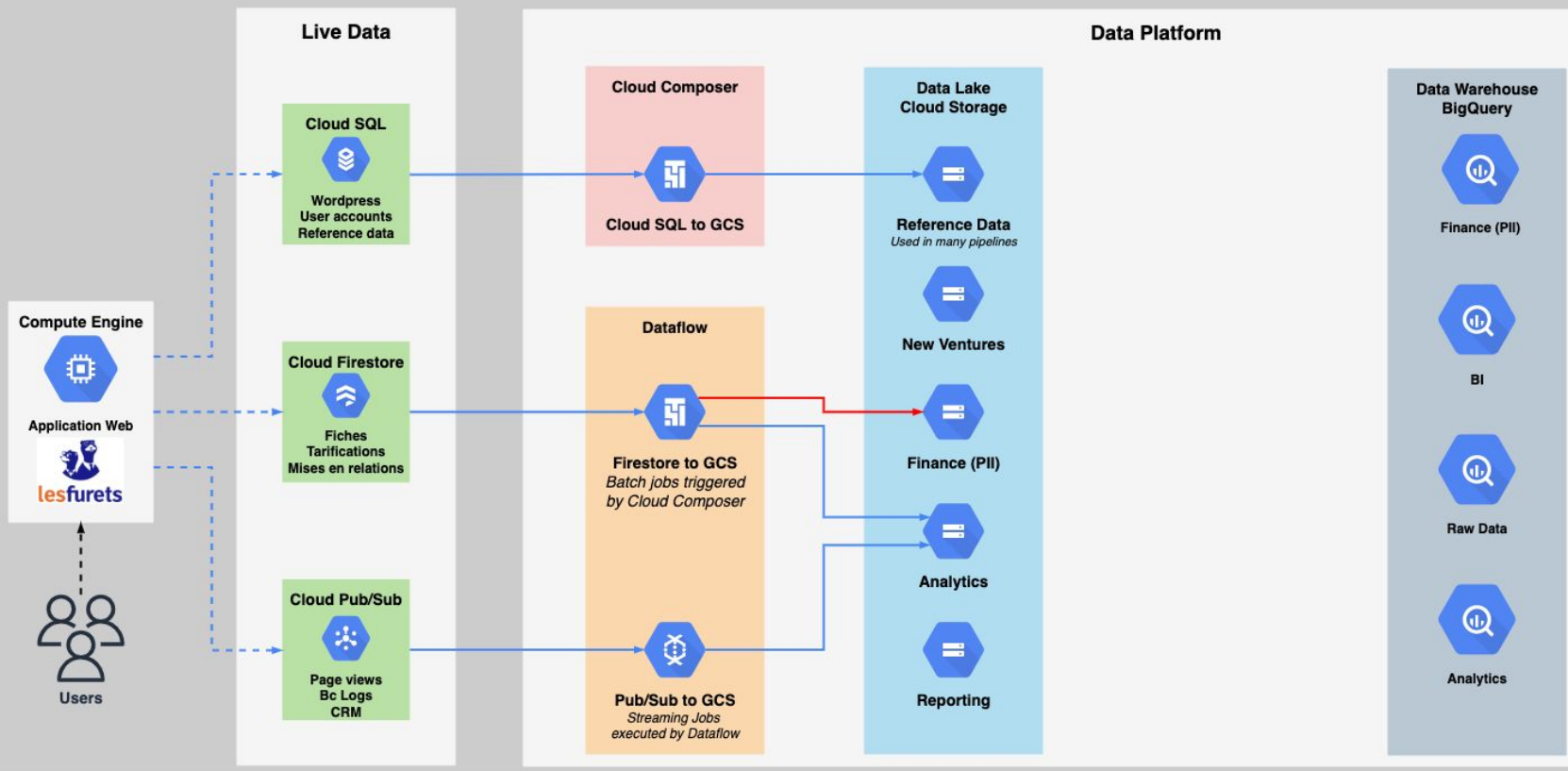
# Data Lake



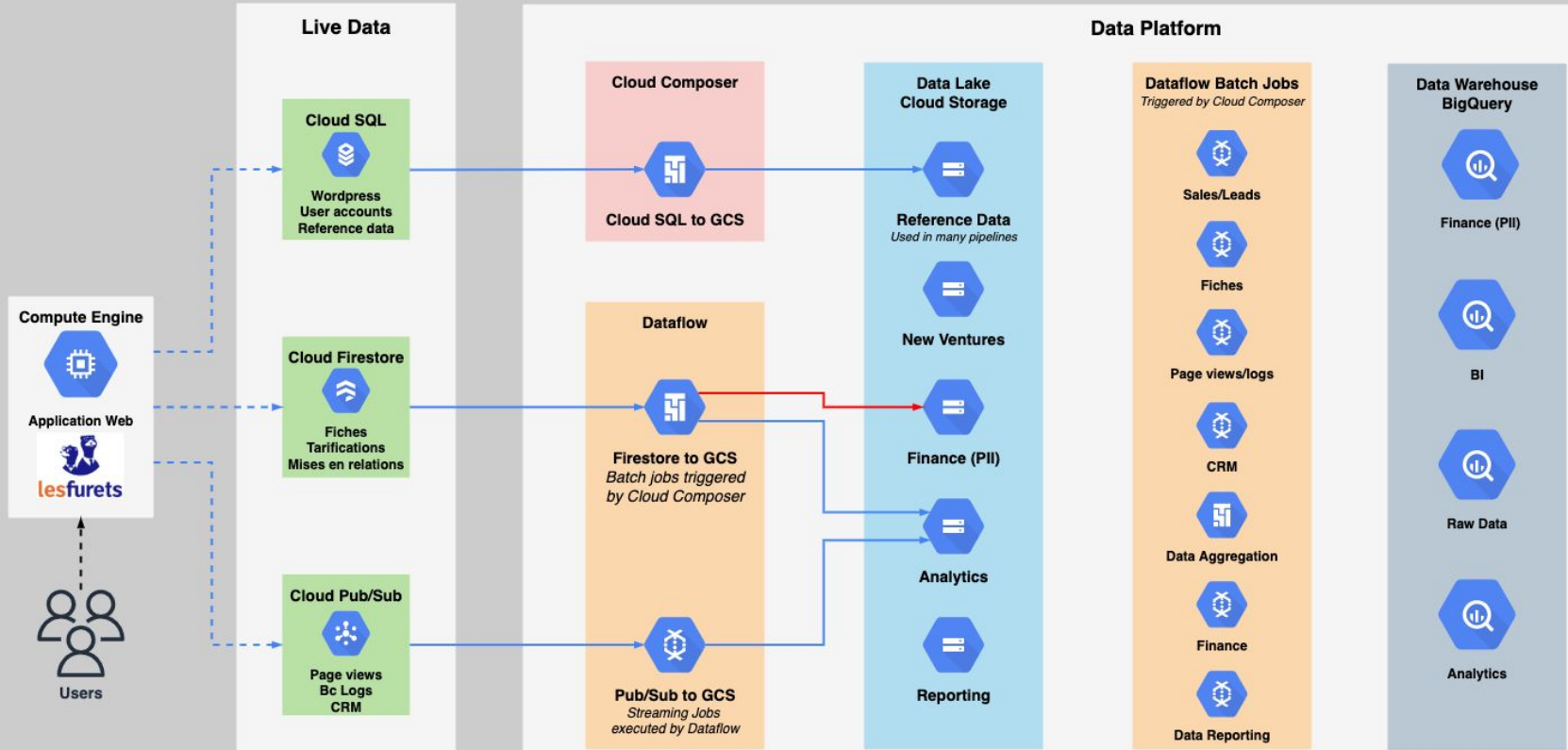
# Plan

1. Intro
2. Data @ lesfurets
  - 2.1. Forms
  - 2.2. Tracking
  - 2.3. Reports
3. GCP - Data Services
- 4. Data Platform @ lesfurets**
  - 4.1. Live Data
  - 4.2. Data Lake
  - 4.3. Data Warehouse**
  - 4.4. Analytics & BI
  - 4.5. Limits

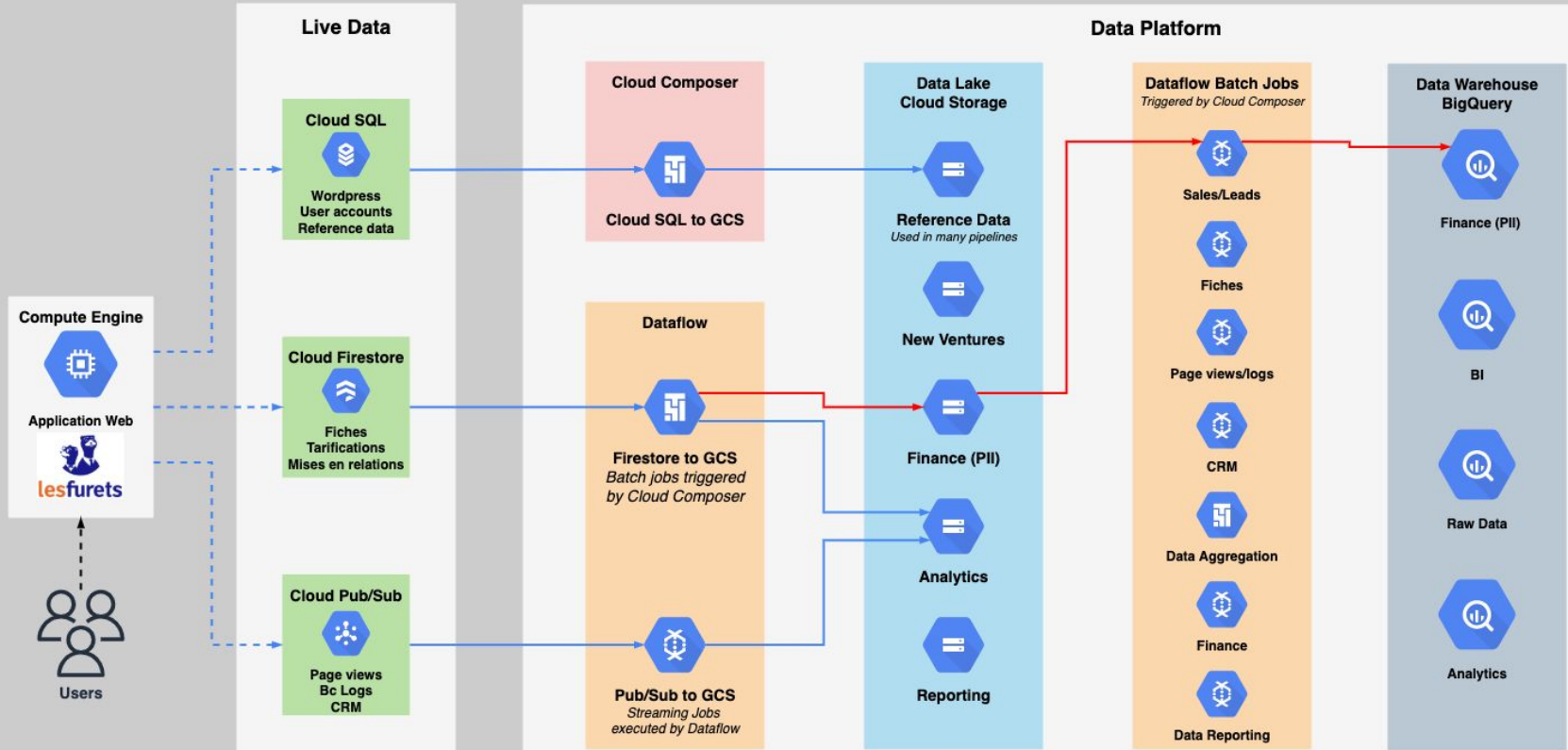
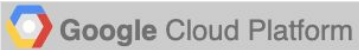
# Data Warehouse



# Data Warehouse

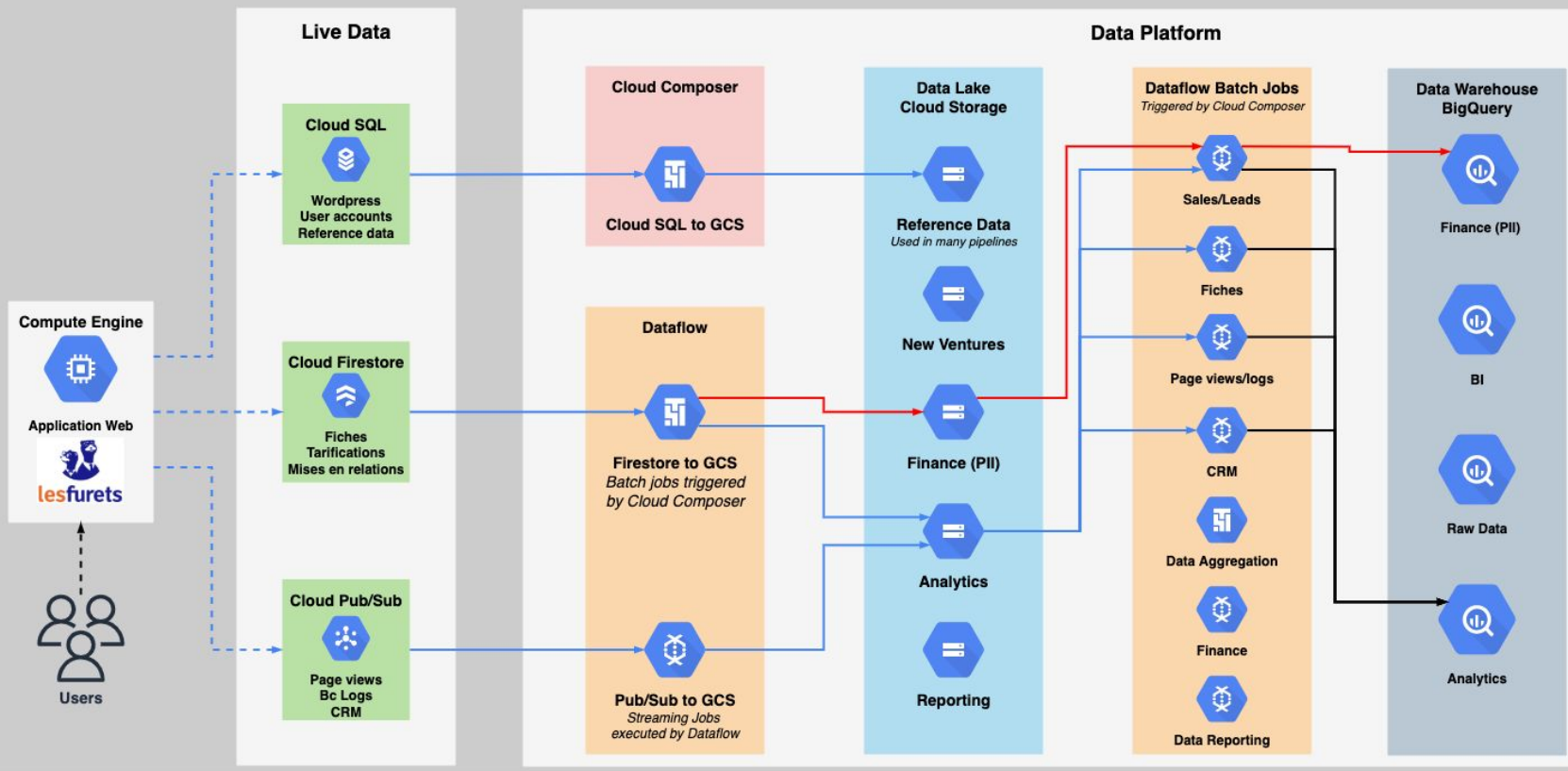


# Data Warehouse

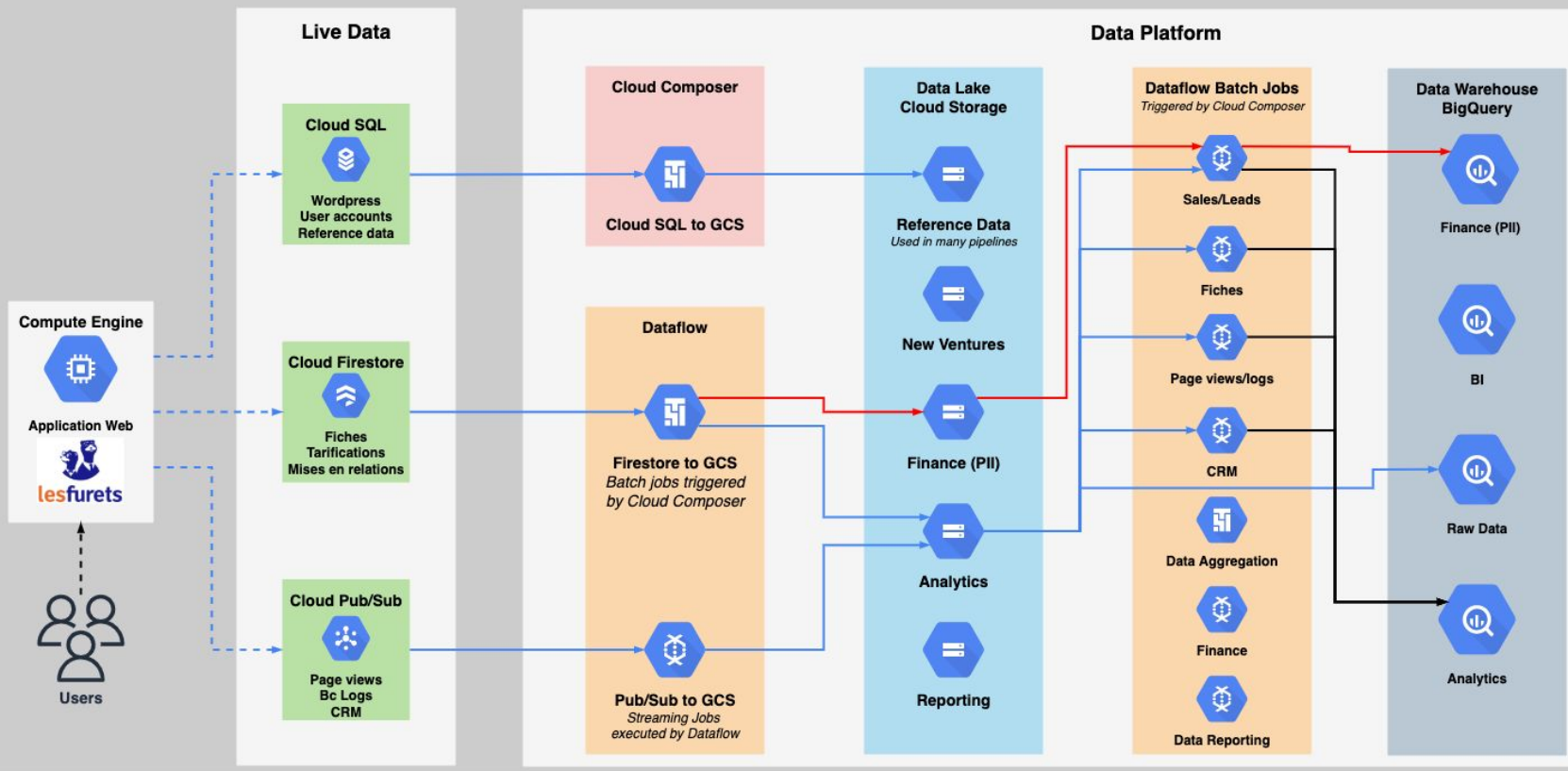




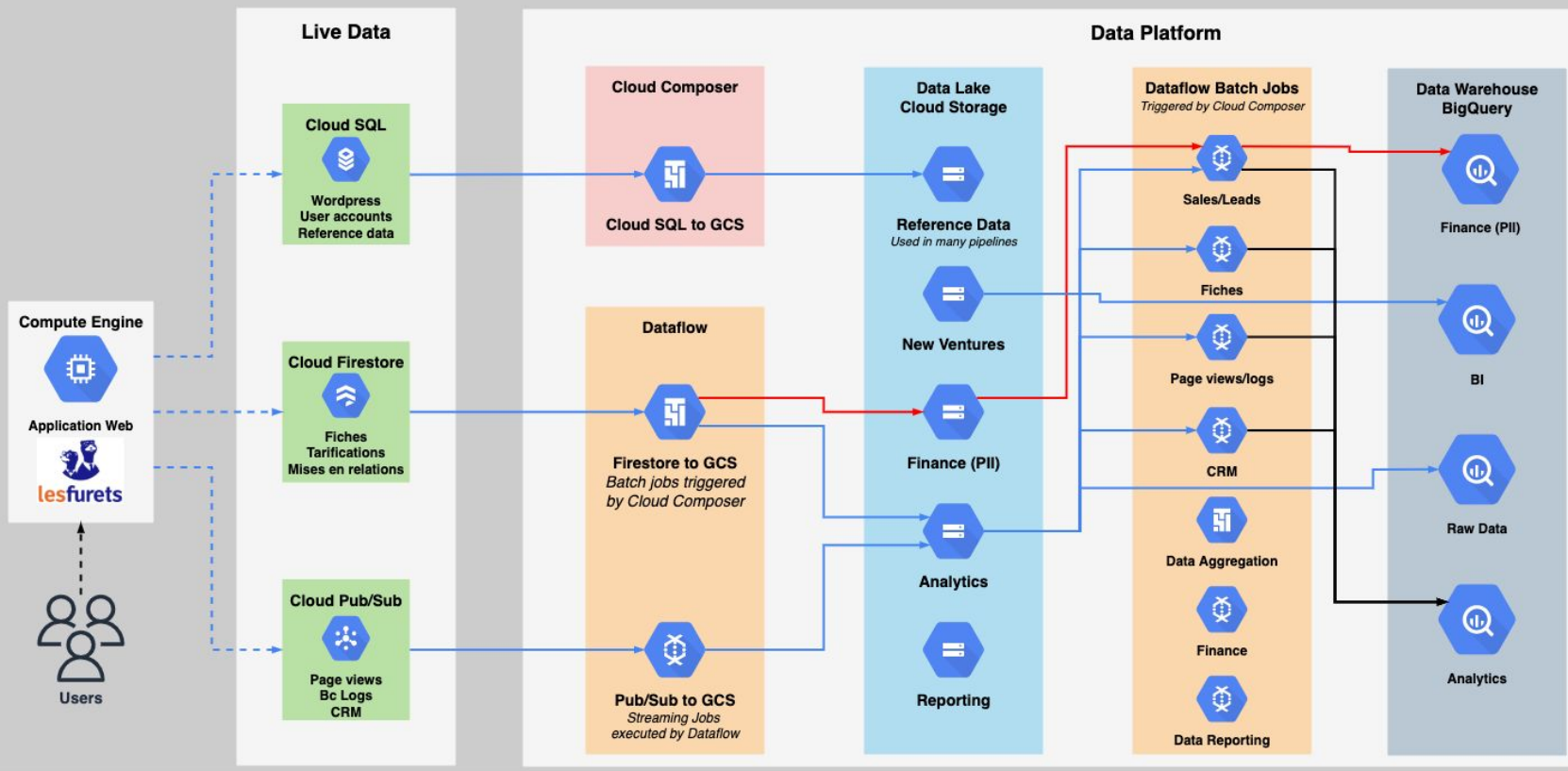
# Data Warehouse



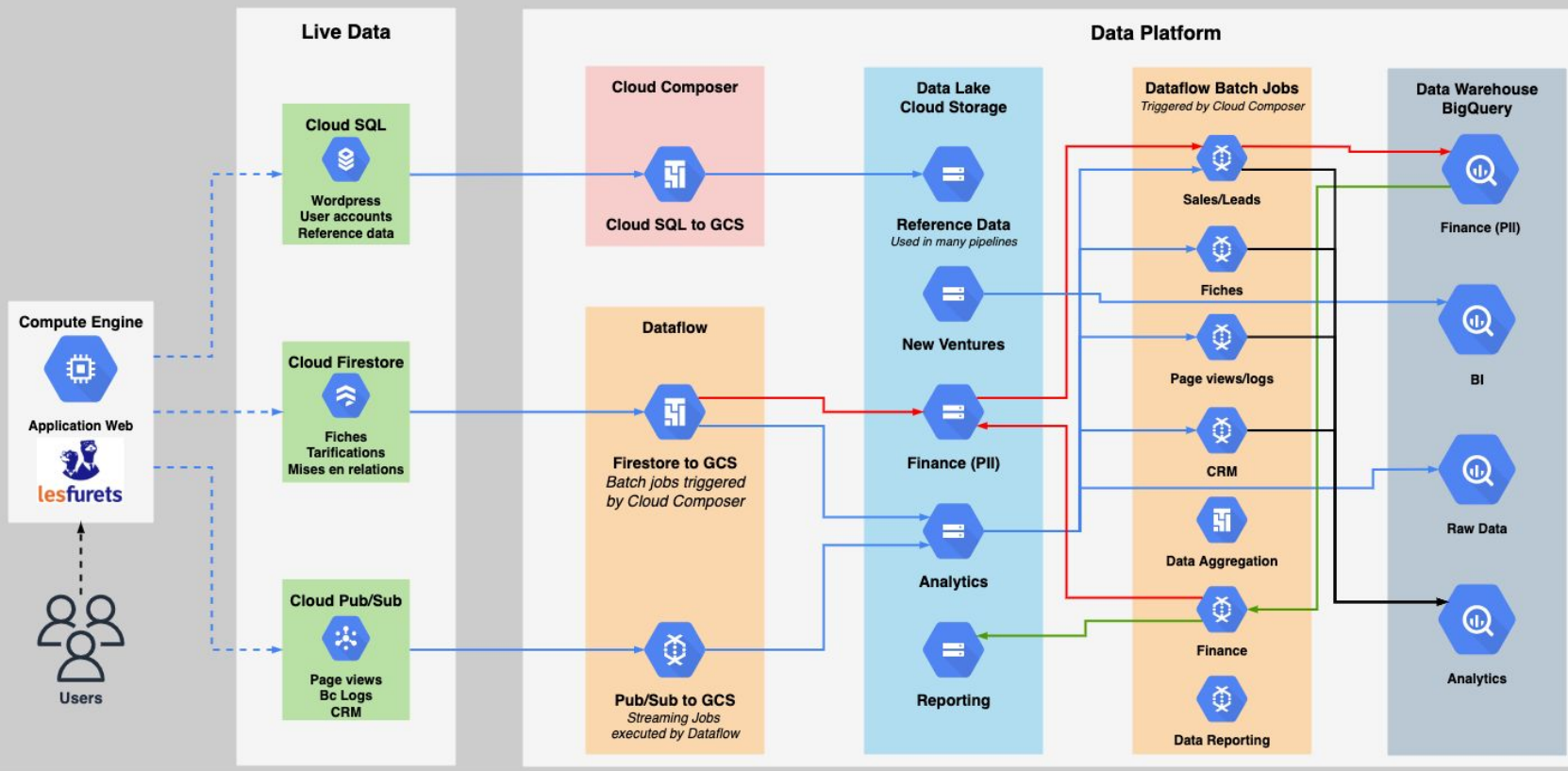
# Data Warehouse



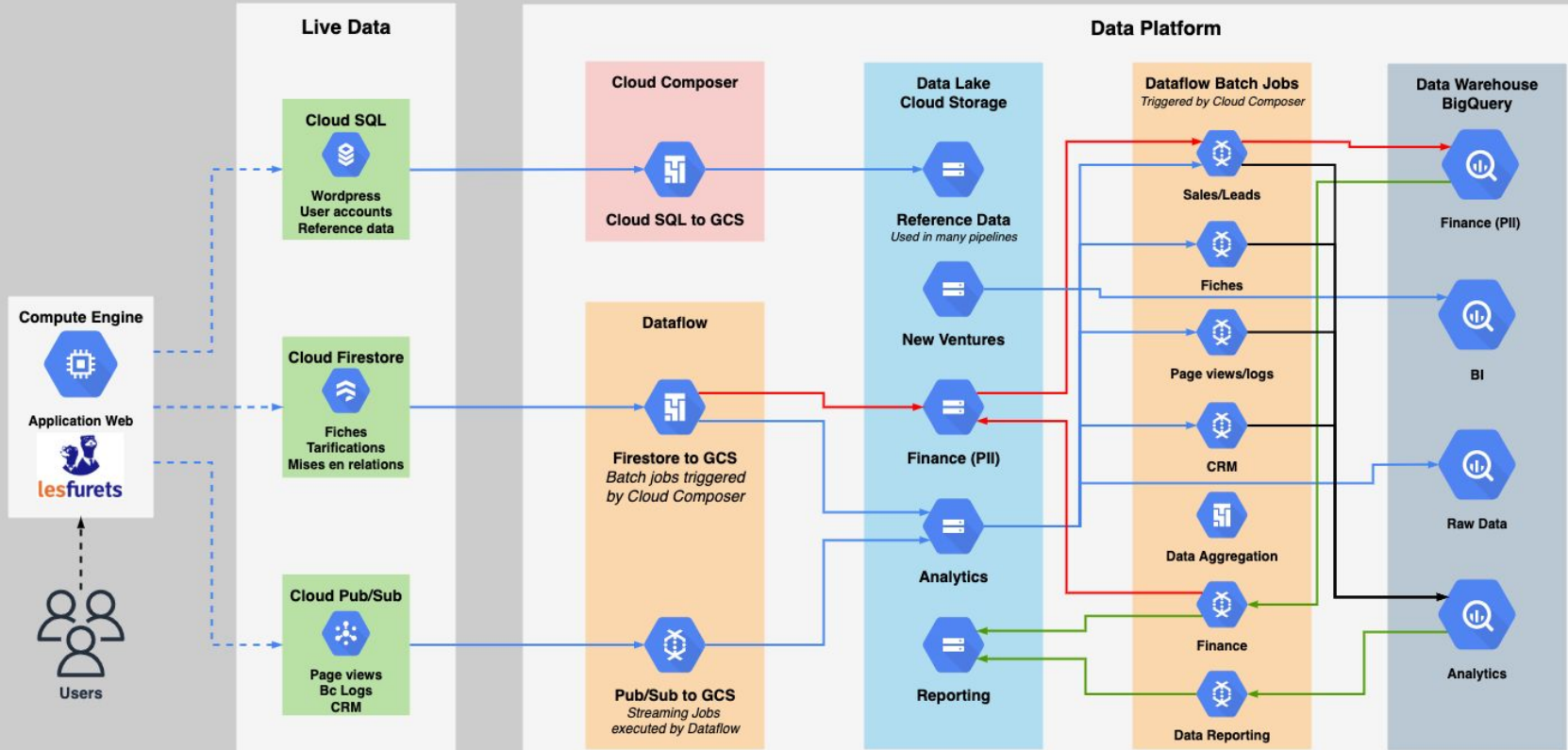
# Data Warehouse



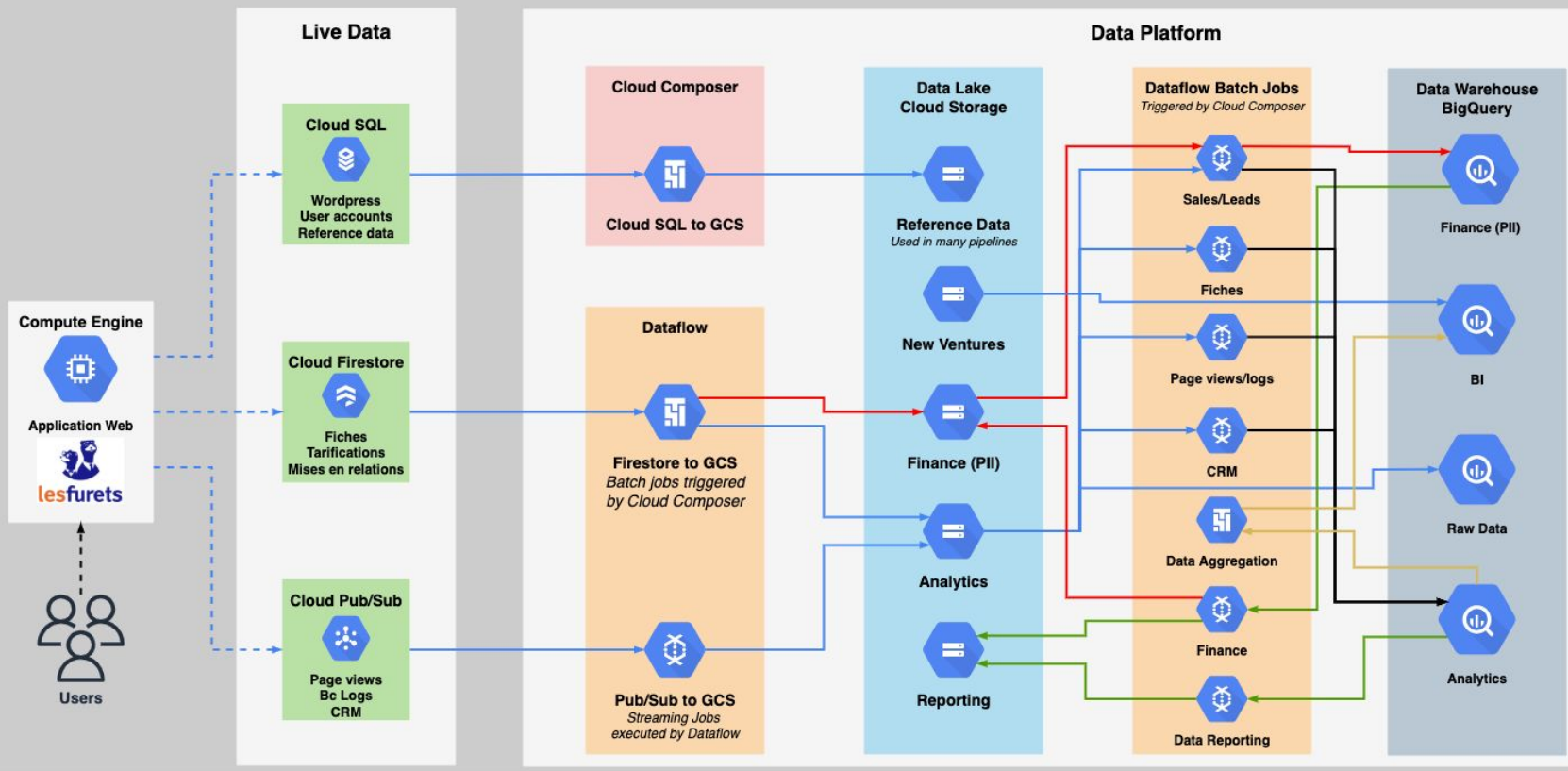
# Data Warehouse



# Data Warehouse



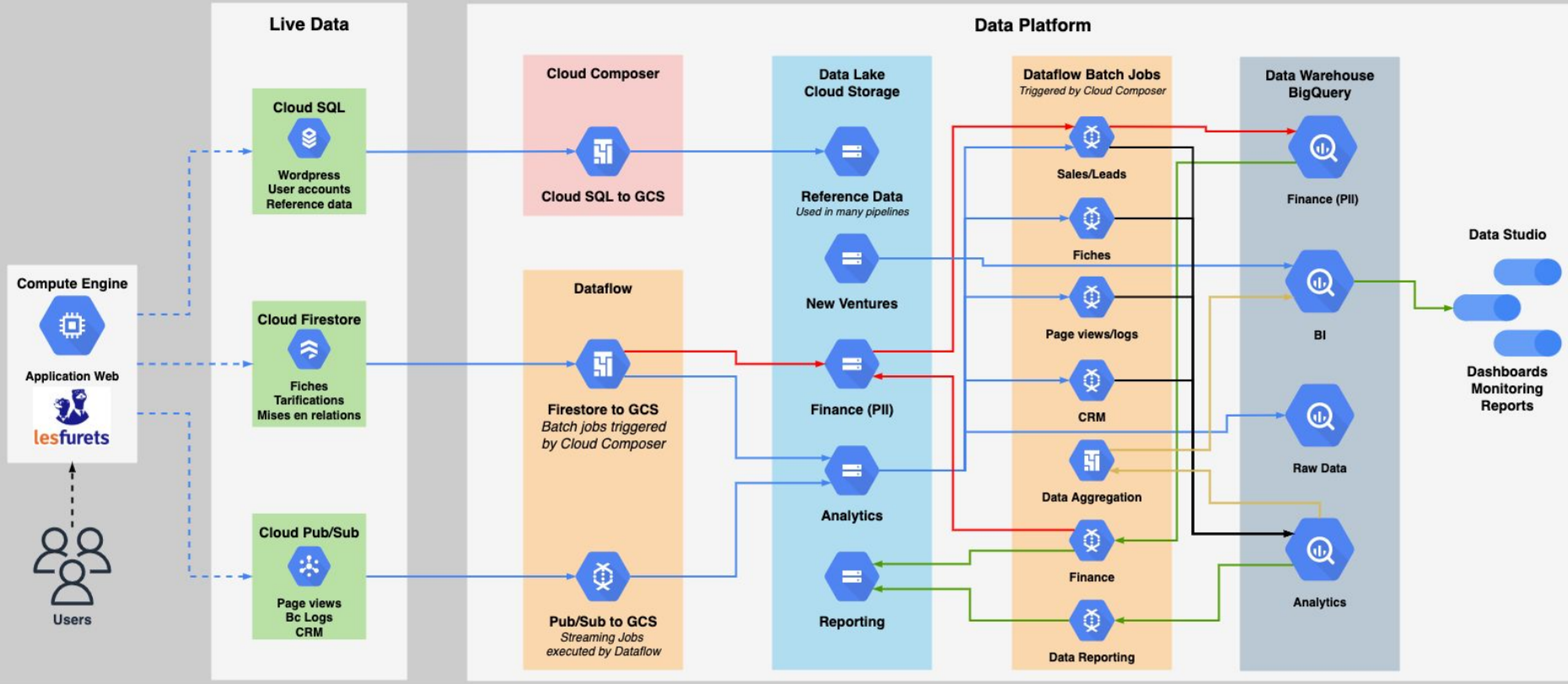
# Data Warehouse



# Plan

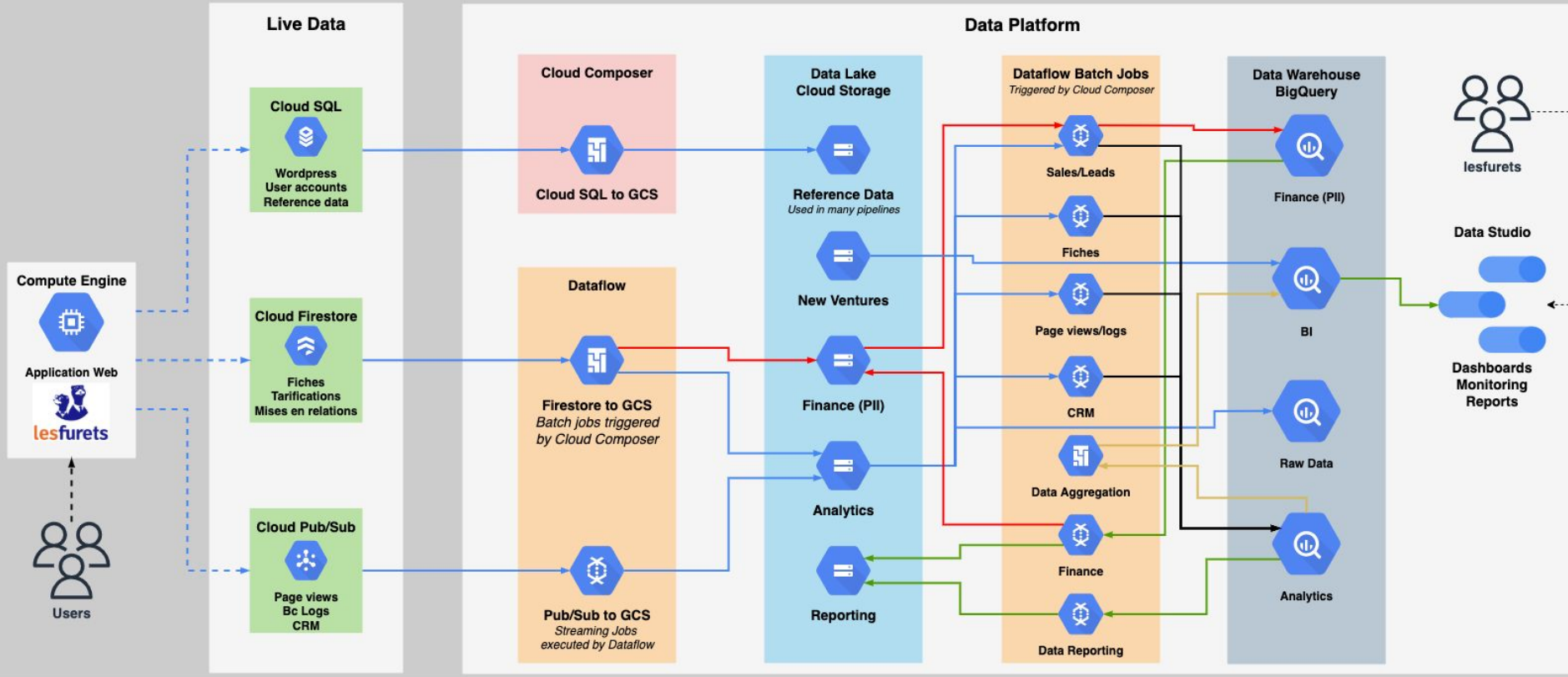
1. Intro
2. Data @ lesfurets
  - 2.1. Forms
  - 2.2. Tracking
  - 2.3. Reports
3. GCP - Data Services
- 4. Data Platform @ lesfurets**
  - 4.1. Live Data
  - 4.2. Data Lake
  - 4.3. Data Warehouse
  - 4.4. Analytics & BI**
  - 4.5. Limits

# Analytics & BI

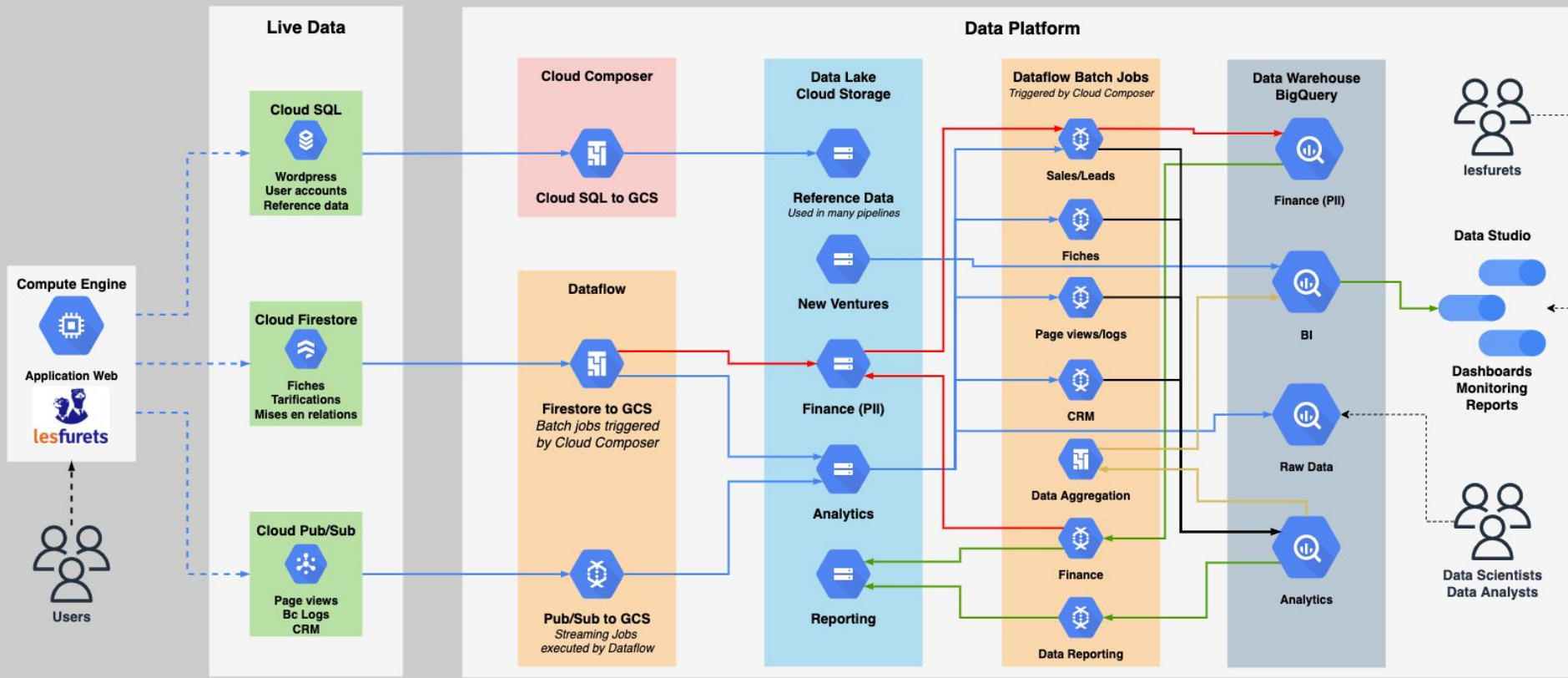
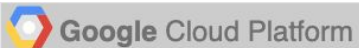




# Analytics & BI



# GCP Data Architecture



# Plan

1. Intro
2. Data @ lesfurets
  - 2.1. Forms
  - 2.2. Tracking
  - 2.3. Reports
3. GCP - Data Services
- 4. Data Platform @ lesfurets**
  - 4.1. Live Data
  - 4.2. Data Lake
  - 4.3. Data Warehouse
  - 4.4. Analytics & BI
  - 4.5. Limits**

# Limits

- Data Studio is slow, and lacks many features
- New dashboards take some time
- Data models versions
- Tests
- 