The new modern data stack

Powered by Airbyte, Airflow, dbt
Hello!

I am Michel Tricot
co-founder & CEO of Airbyte

@MichelTricot
michel-tricot
/in/micheltricot

Airbyte
Open Data Integration (ELT)

3,000 users
1,900 Slack members
3.5k GitHub stars
Extract - Load - Transform

A new paradigm for modern data teams
Extract

Source-specific routines to pull selected data from an external system.

Transform

Business logic specific to your organization to serve an analytics or operational use case.

Load

Destination specific routines to push data where it is going to be consumed.
ETL doesn’t work in today’s world

**Inflexible**
- Friction when changing an existing pipeline.
- Hard to add new data.
- Most issues force data to be re-extracted.

**Lack of Autonomy**
- Warehouses made data consumers more autonomous.
- Changes require engineering involvement.

**Complex**
- Custom DSL.
- Force adoption of a data stack.
- Address 70% of the needs, 30% still built and maintained in-house.
**Extract**

General-purpose routines to pull selected data from a source.

**Load**

General-purpose routines to push raw data where it is going to be consumed.

**Transform**

Business logic specific to your organization to serve an analytics or operational use case with SQL / dbt / ...
ELT fixes the ETL-related issues

**Flexibility**
- All the data available on the destination.
- Data consumers are free to use what they need for the insights they want.

**Autonomy**
- Data consumers can leverage SQL queries to transform the data the way they want.
- No need to involve the engineering team.

**Future proof**
- Issues during transformation don’t prevent access to the data.
- Easy to update transformation schemas.
2

The open modern-stack

Compose with best of breed
The world of data is changing

All companies have custom data needs

- Types
- Skills
- Scale

- Budget
- Security
- Privacy

- Turnaround
- Compliance
- ...

We are moving away from horizontal, end-to-end solutions
Select the best building blocks

**Extract & Load**
- Airbyte
- Fivetran

**Transformation**
- dbt
- SQL

**Data warehouse**
- Snowflake
- Amazon Redshift
- Databricks

**Quality**
- Great Expectations

**BI / Visualization**
- Superset
- Looker

**Observability**
- Monte Carlo
- Bigeye

**Infrastructure**
- Docker
- Azure
- Amazon Web Services
Composed with the best glue
3

Airbyte-dbt-Airflow

To rule them all
Airbyte: Solving Data Integration

- **Open-Source** data integration platform
- **80 high-quality data connectors** (targeting 200 by EOY)
- **Connector Development Kit** to build your own data connector
- Active **Slack community**
- Super fun mascot: **Octavia Squidington III**
An open data platform for modern teams

- Monitoring
  - Incremental updates

- Connector Development Kit
  - Monitoring

- Data Exchange Protocol
  - Source
  - Source
  - Destination
  - Destination
  - Source
  - Source
  - Destination
  - Destination

- Platform
  - UI
  - CLI
  - API
  - Config & State
  - Scheduler

- Connectors
  - Amazon Web Services
  - Docker
  - Azure
dbt: Solving Data Transformation

- **Open-Source** data transformation application
  - Transform SQL to tables, views
  - Cascades transformations
- Supports most SQL **Databases & Warehouses**
- Brings software engineering **best practices** to Analytics Engineers
  - Version control
  - Testing
  - Packaging
Airflow: Solving Scheduling

Do I really need to tell you what Airflow is?
Reverse ETL
Data Warehouse

Extract  Load  Transform  Activate

BI/Visualization
- Superset
- Looker

Reverse ETL
- high touch
- Census
- Airbyte
Demo

All source code & configurations available on github
Thanks!

Any questions?

@MichelTricot
@michel@airbyte.io
slack.airbyte.io(@Michel)
airbytehq/airbyte