



Yellowbrick

Yellowbrick: The Perfect
Data Warehouse
for Databricks



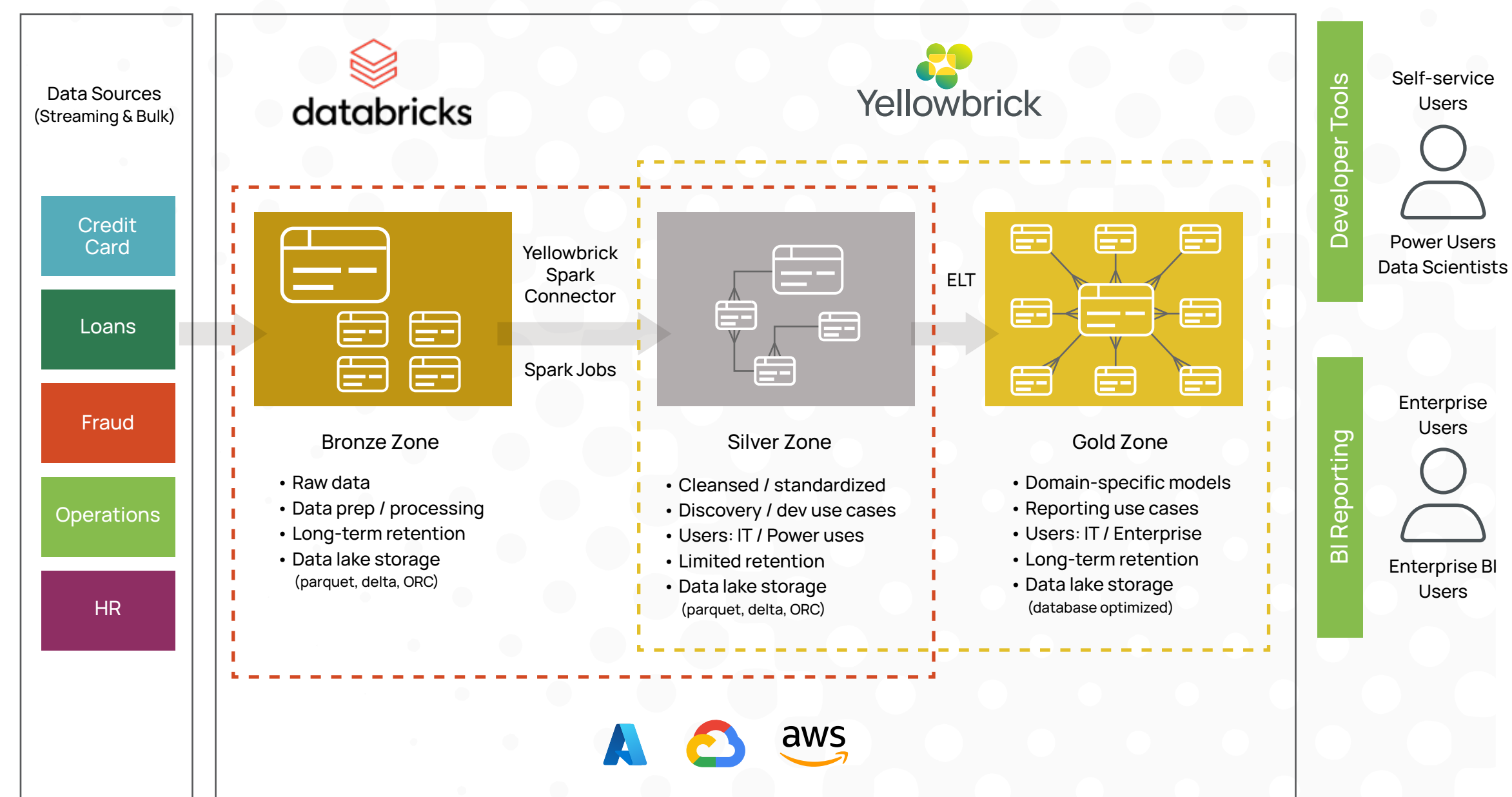
Yellowbrick and Databricks are Perfect Partners.

It's commonplace for all large enterprises to deploy a "data science" platform alongside a "data warehouse" platform because the two have different strengths and weaknesses. Both Yellowbrick and Databricks can run together inside customers' own VPCs, minimizing cost and data movement which also eases security concerns.

Since its inception, Yellowbrick has been built as a high-quality, enterprise-grade database supporting highly concurrent, ad-hoc queries by thousands of users across complex schemas and changing data. Yellowbrick supports such mixed workloads with strong transactional consistency and the high availability required for Tier 1 business applications. It's common to find Yellowbrick backing business-critical websites and applications in the world's largest telcos, hospitality businesses, insurers payment processors, and credit card companies. Yellowbrick has been running such complex, business-critical workloads in production for seven years, taking advantage of built-in asynchronous replication for disaster recovery to ensure business continuity.

Real-time, streaming inserts of data is supported, unlike with other cloud data warehouse platforms, enabling up-to-the-second reporting. Yellowbrick integrates transparently with industry-standard ETL and data movement tools

Yellowbrick and Databricks Lakehouse Architecture



Databricks strengths:

- SparkSQL data processing pipelines
- Job orchestration
- Support for diverse data sources and types
- Developer focus

Yellowbrick strengths:

- High concurrency / mixed SQL workloads
- Built-in Spark & Kafka connectors
- Optimized for processing relational data
- Business, SQL & Analyst focus



from vendors such as Informatica and Oracle, as well as all widely available BI and analytics tools. Support for rapid movement of data from Spark and Kafka is real-time and built-in, enabling trivial integration into modern data platforms, and connectivity to Python and R is provided through standard PostgreSQL packages. The data warehouse is fully elastic, with separate storage and compute managed through SQL, and requires little to no management or fine-tuning whatsoever. Automated tooling allows assessment of the cost and timeframe for data warehouse migrations and typically automates >95% of the porting effort from legacy platforms such as Teradata, even including BTEQ scripts.

Being a PostgreSQL-compatible database like Greenplum, Netezza, Redshift, and Vertica, migration to Yellowbrick from these platforms can be completed quickly and easily, resulting in improved performance, higher uptimes, reduced cloud infrastructure, and lower costs. Yellowbrick supports stored procedures and ANSI-standard SQL with extensions for compatibility with other enterprise databases like Oracle, SQL Server, and Teradata to ease migration.

Databricks started as a processing engine – a managed version of Apache Spark – and is well known to offer the best platform for data science, machine learning, and data engineering across structured and unstructured data. It has since been

extended to include a data lake and a SQL engine but was never conceived as a database and thus cannot offer the concurrency, uptime, interoperability, or availability guarantees of a hard-core enterprise database. It's designed to be used by specialists who have experience fine-tuning Spark. Key features are reserved for their commercial products. For this reason, businesses from financial services, to telcos, to telemetry and hospitality vendors will always deploy Databricks alongside a data warehouse such as Yellowbrick or Snowflake: One platform excels at serving data through SQL to users, and the other excels at providing the tools that data scientists expect.

Given the intense multi-year focus required to build a solid enterprise database, and the intense, multi-year focus needed to build and maintain an ever-evolving data science platform, it is unlikely that Databricks will become a great data warehouse vendor, or the database vendors become great data science platforms, any time soon. Customers should continue to choose best-of-breed tools: A database such as Yellowbrick for highly concurrent, highly available, complex ad-hoc queries on structured data, as well as supporting sub-second interactive queries with strict SLAs, and a data science platform such as Databricks for programmers and data scientists to handle machine learning and data engineering.

Book a Demo Intrigued? Find your biggest, most complex data set and most important business challenge and invite us in for a Yellowbrick Demo. We'll show you how high-speed data analytics can redefine your business opportunities.

