

Anzo Smart Data Lake[®]

& AnzoGraph

Arthur Keen, Managing Director
Financial Services

About

History

2007 – Spun out of IBM's Advanced Internet Technology Group
2015 – Acquired MPP database Netezza & Paraccel

Leading customers:



Strategic partnerships:

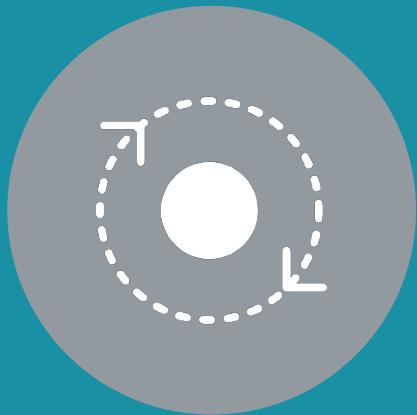


Industry recognition



Anzo Smart Data Lake®

The industry leading platform for building a
Semantic Layer for the Enterprise



End-To-End

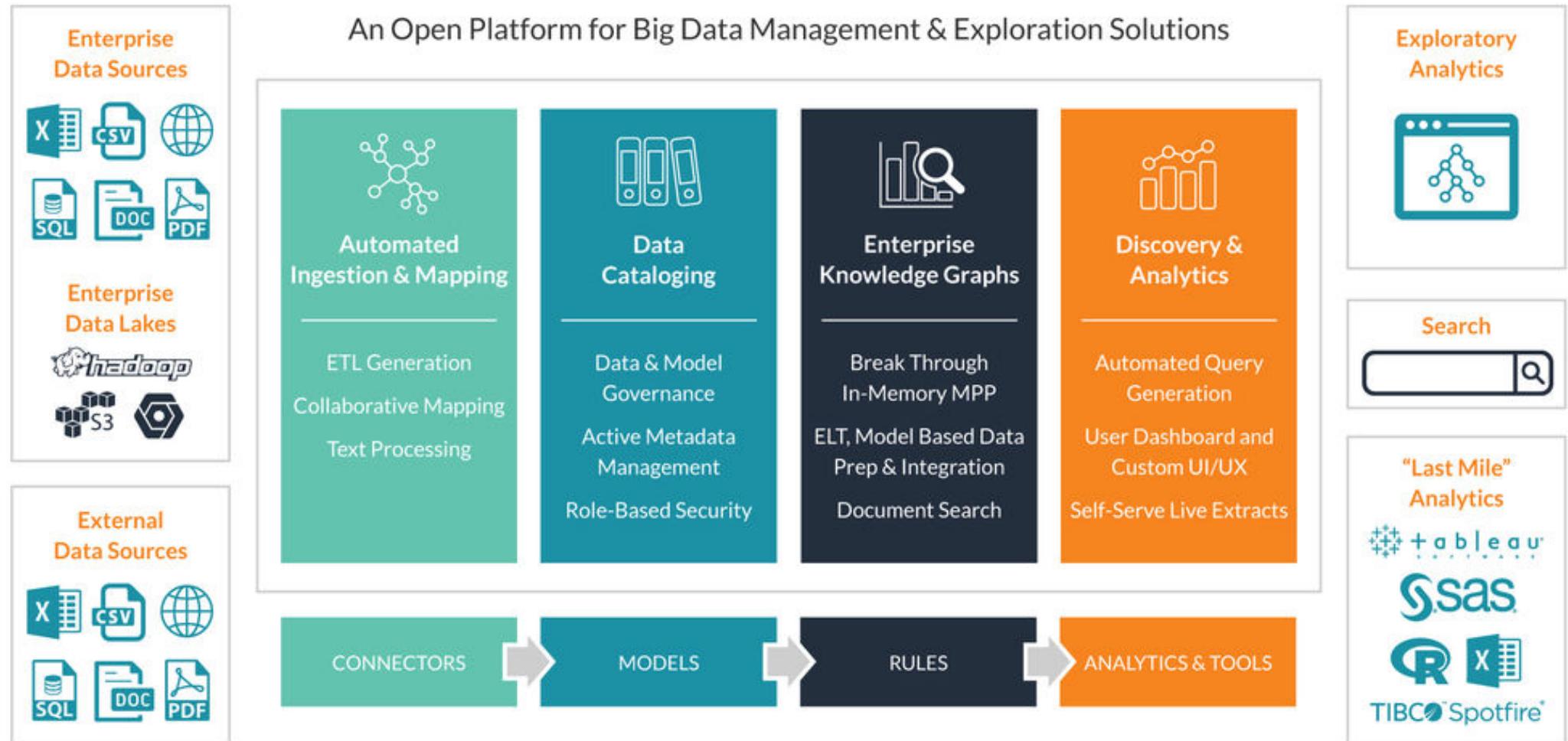


Open Standards



Enterprise Scale

Anzo Smart Data Lake 4.0





THREE RING CIRCUS
MASTER CLOWNS SHOW

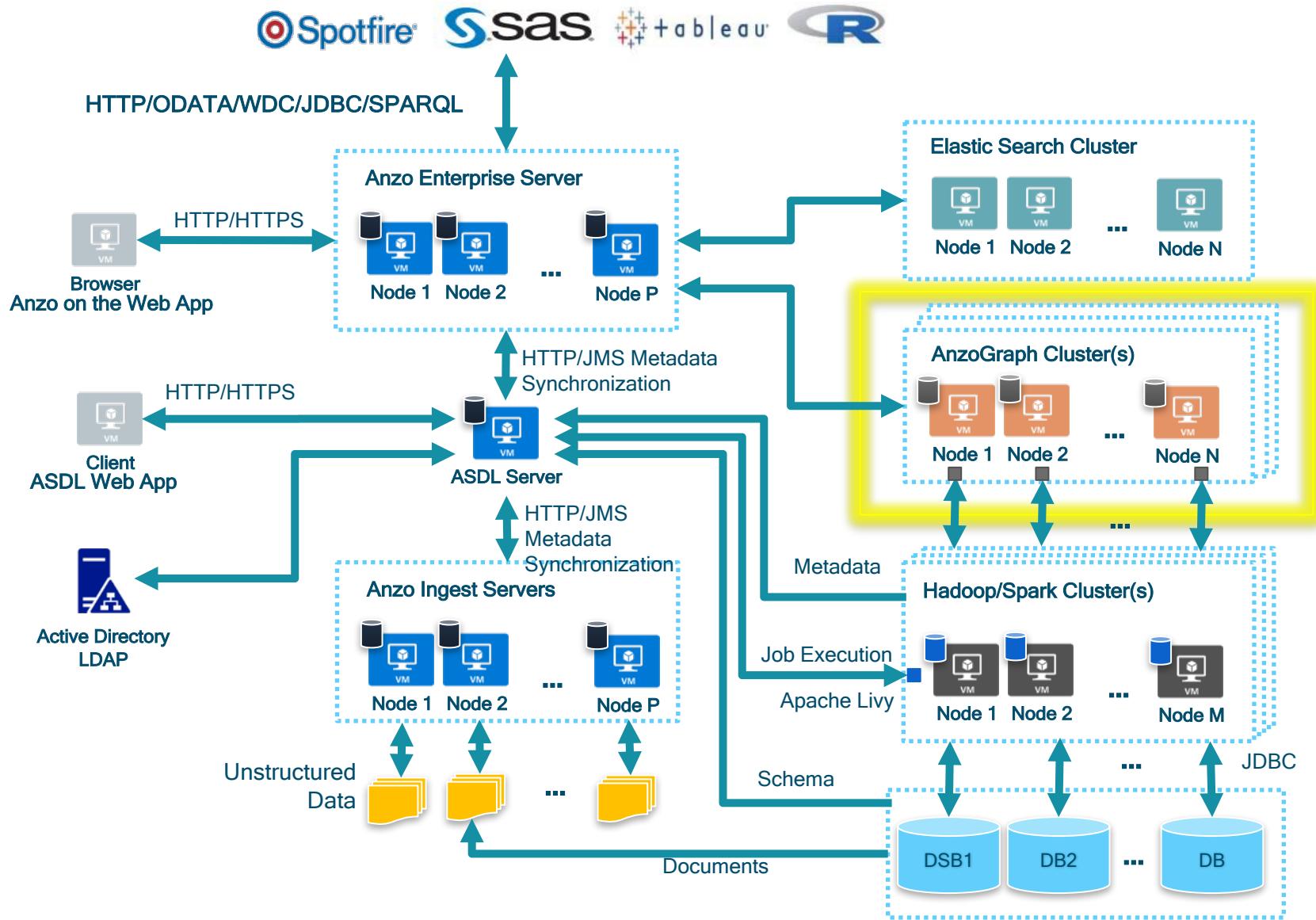
CIRCUS

BALL

3.

Richard Capes

Reference Architecture



AnzoGraph



AnzoGraph

- Massively Parallel Processing (MPP)
 - Designed for Analytics (join vs. traverse)
 - Shared Nothing
 - Each core contributes to each query
- Ancestry Paraccel, Redshift, Neteza
 - Same lead developers from Paraccel
 - Perpetual license to some of structural code

Graph Database	Load Time	Inference Time	Query Time
Oracle Database 12c [2014]	115.2 hours	86.5 hours	22.5 hours
Cambridge Semantics ANZO Graph Query Engine [2016]	1,760 seconds	4,574 seconds	840 seconds
Cray Graph Engine [2018]	4,124 seconds	535 seconds	96 seconds

AnzoGraph Strengths

- Very High Speed Load
 - 250 GB/hour/node (CSV on 32vCPU nodes)
 - Fully parallel
 - Cores over-subscribed in pipeline
- Complex Queries
 - Cost/Rules hybrid query planner
- Extended SPARQL/RDF
 - Window functions
 - 70+ added “Excel” builtins, including aggregates
 - Views
- Auto-index & compression
 - Background Vacuum & Index optimization

Highlights

- Generated code (C++)
 - Runs “close to the silicon”
 - Eliminates interpreter overhead
 - Re-uses previous code snippets
 - Educated by Netezza, Paraccel
- Long pipelined flows
 - Non-materializing across (multiple) network hops
 - Each query uses multiple threads on every core
- Network Model
 - Throughput, not latency (dataset “push”)
- Dictionary-based compression
 - Late decompression

Current Projects

- Peta-Scale Scanning
 - HDFS, Local drives, S3
- RDF* “RDR - Reification Done Right”
- Property Graphs
- Tinkerpop Gremlin Support

Questions ?

Try out AnzoGraph:
<http://www.anzograph.com>