

AWS SEMICONDUCTOR & HITECH SOLUTIONS

Modernizing Semiconductor Manufacturing Analytics Platform using AWS Data Analytics Services

Upasana Pandya

Nov 7th 2024



Agenda

- Semiconductor Industry Trends, Challenges
- Semiconductor Smart Manufacturing Use Cases
- Case Study

Amazon and Semiconductors

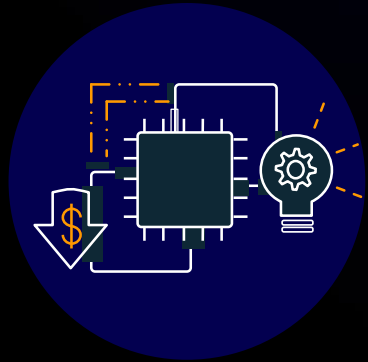
Amazon develops and uses semiconductor devices for:

- AWS data center infrastructure
- Amazon fulfillment centers
- Consumer devices
- Robotics and AI
- Space/satellite infrastructure
- Autonomous vehicles
- And more

We value our semiconductor industry partnerships



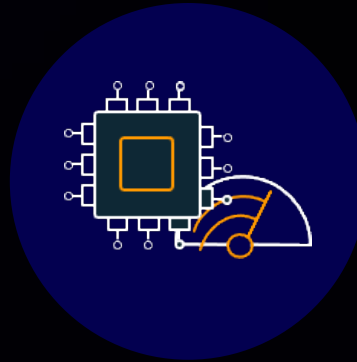
Industry Trends and Challenges



Supply Chain and
Manufacturing
Capacity Constraints



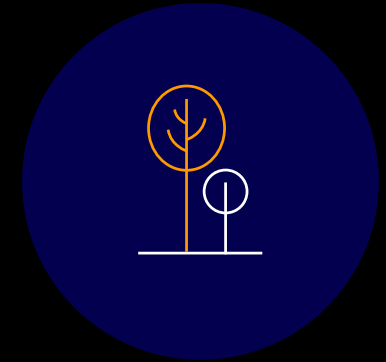
Increased Multi-Party
ASIC/SoC
Development



Balancing Power,
Performance, Area,
Cost (PPAC)



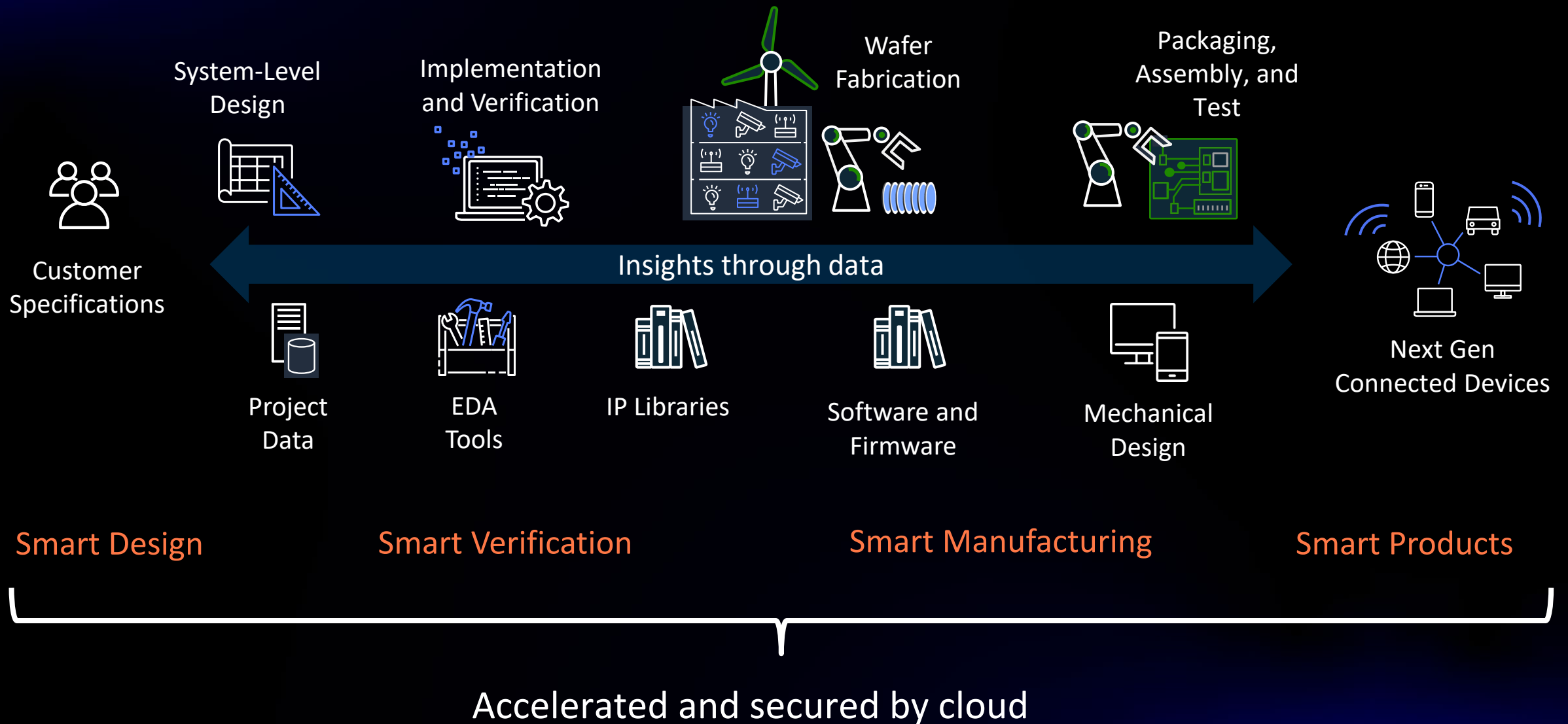
Engineering Staffing
Shortages



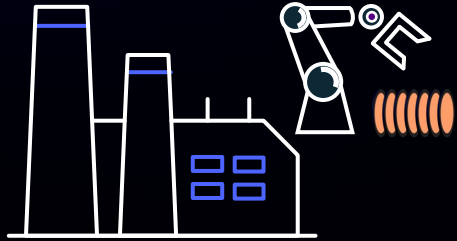
Sustainability
and
ESH/ESG

Semiconductor design, verification, and smart manufacturing, accelerated with AWS Cloud, help to address these challenges across the entire supply chain

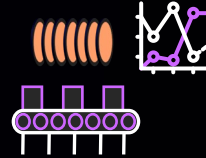
Bridging the semiconductor supply chain with data



Semiconductor Smart Manufacturing use cases



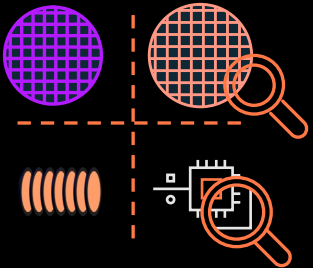
Yield and Failure Analysis



Preventive/predictive maintenance for machines



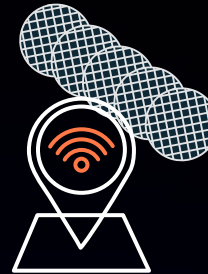
Machine as a service



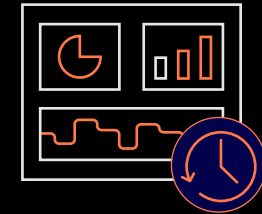
Computer vision for quality control



Automated material management



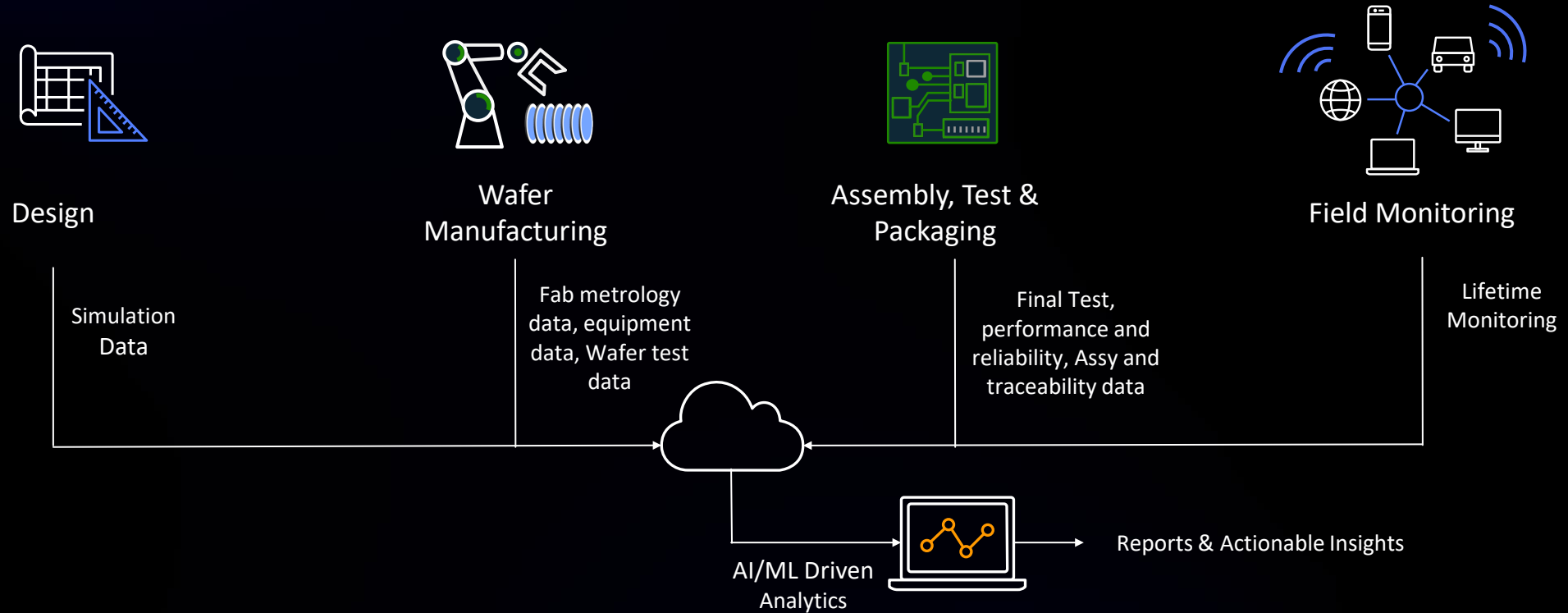
Track & Trace



OEE visualization

What use case does this solution address?

Yield and Failure Analysis



Data across the value chain can be connected to provide end-to-end visibility, monitoring and analytics to improve product quality and achieve faster yield improvement to shorten silicon time to market

How to build such an ecosystem:

Build Data Lakes quickly

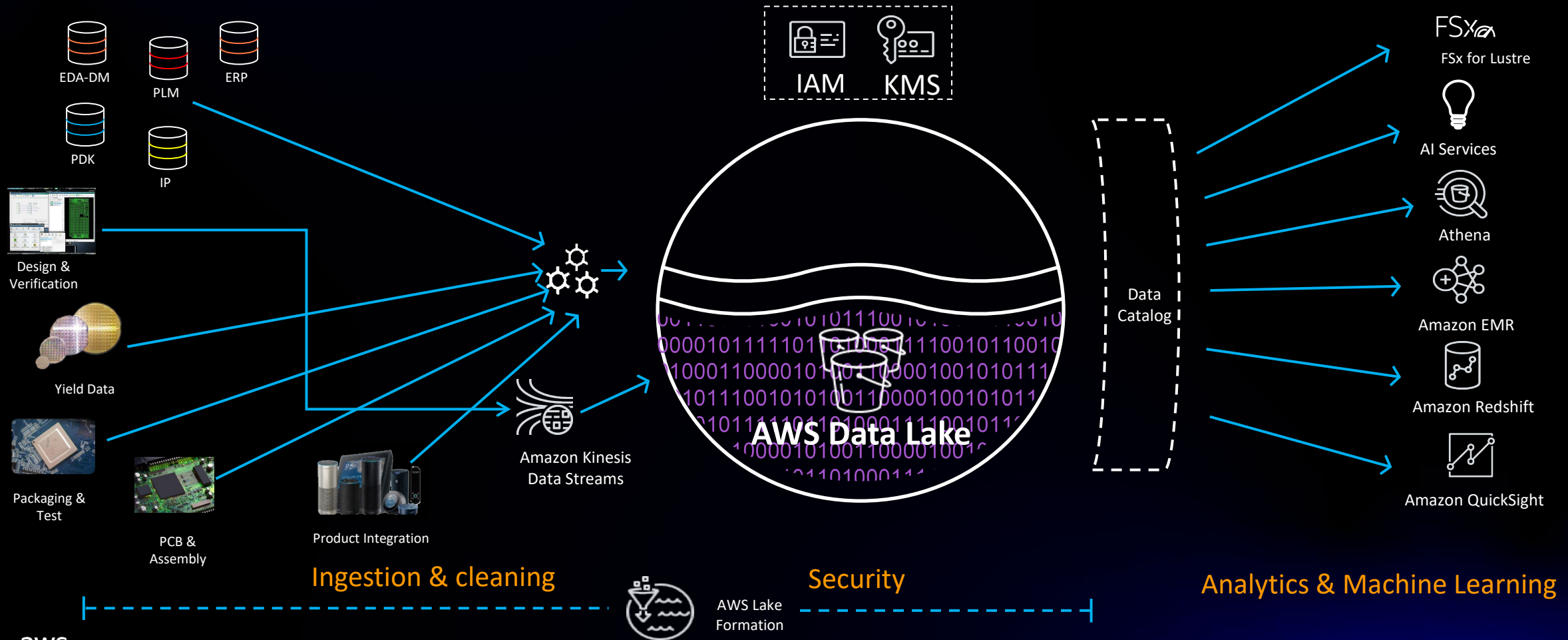
- Identify, crawl, and catalog sources
- Ingest and clean data
- Transform into optimal formats

Simplify security management

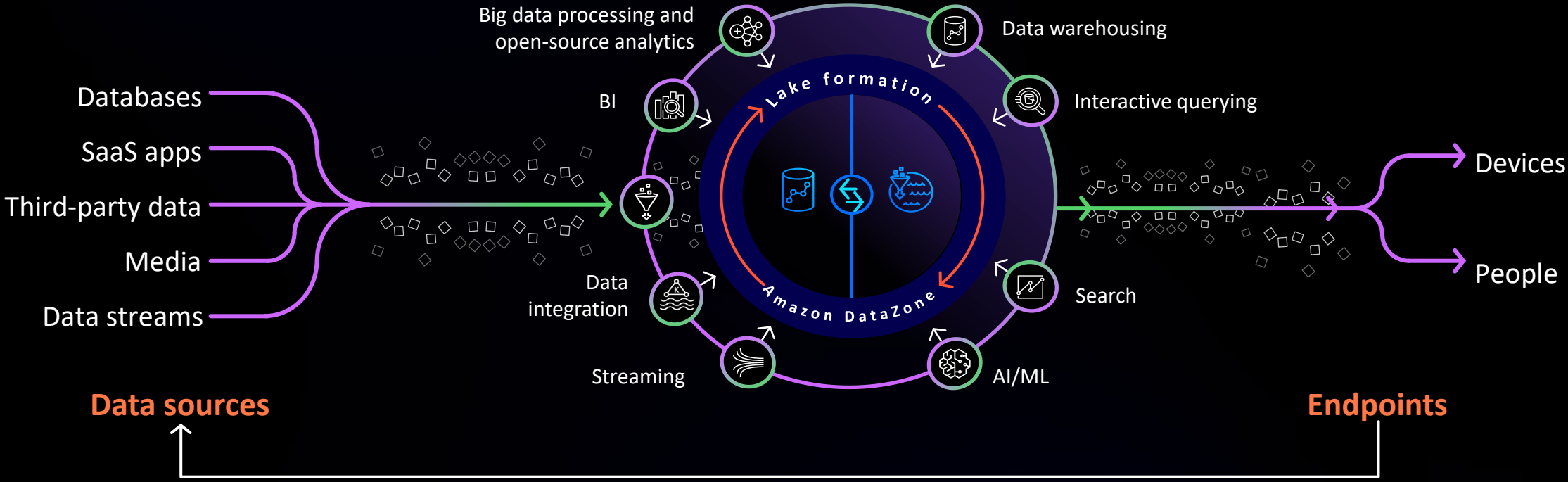
- Enforce encryption
- Define access policies
- Implement audit login

Enable self-service and combined analytics

- Data from single data catalog
- Choice of analytic tools



Build an end-to-end data strategy



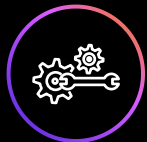
Unparalleled innovation, high-performance analytics, democratizing insights



Modern architecture: Powering trusted analytics



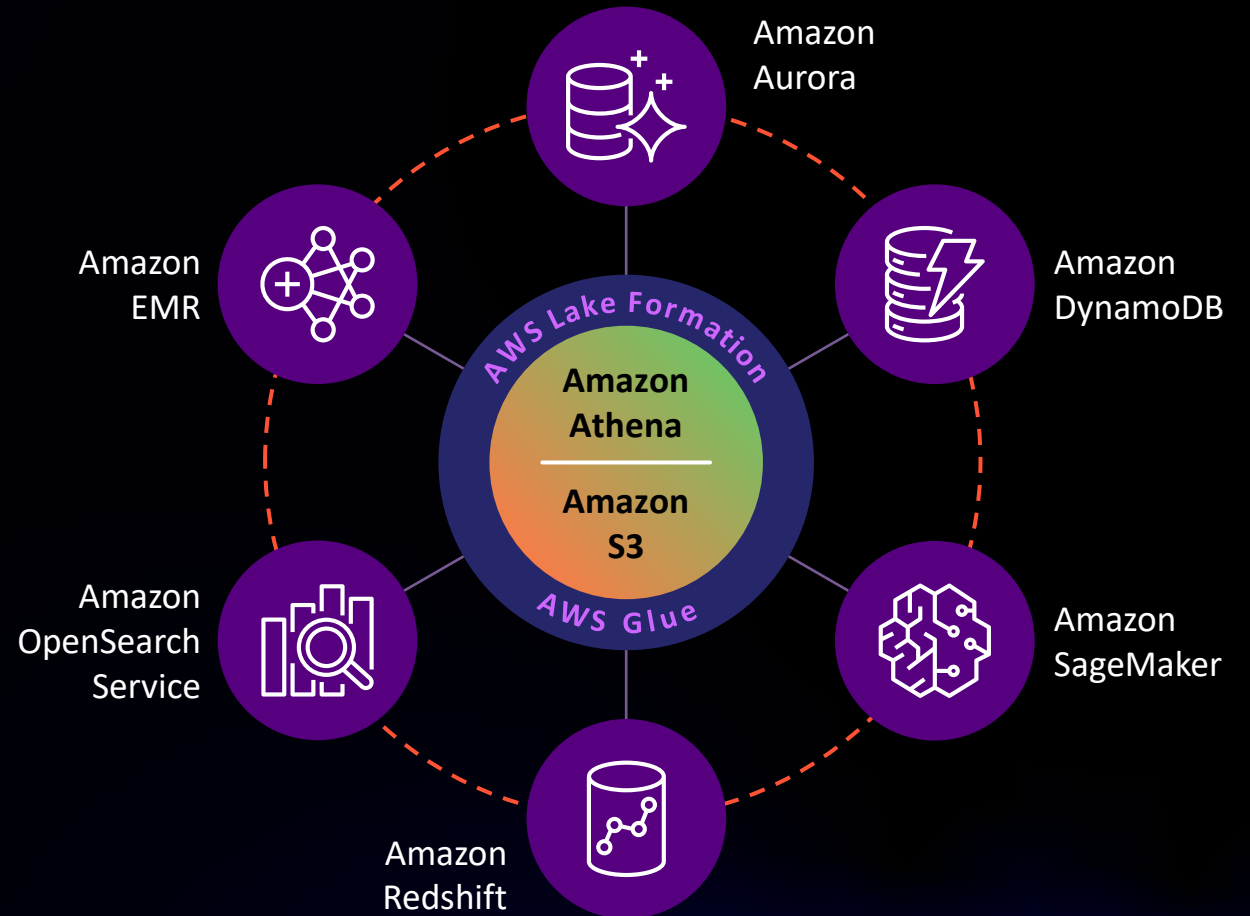
A unified ecosystem means fewer siloes and more collaboration



Embracing "produce once, share many" is critical for trust



The design scales well with an evolving enterprise



A world map with a dark blue and purple color scheme. Three circular icons, each containing a white database symbol (three stacked cylinders), are placed on the map: one in North America, one in Europe, and one in Asia. A horizontal line with a purple-to-orange gradient is positioned below the title.

Beginning of our Data Transformation Journey

Disparate siloed data sources

Legacy, non-scalable infrastructure

High maintenance cost

Significant capex needs

Solution transfer across sites not viable

400+ Mfg application and 300+ Enterprise applications feeding data in non-unified manner across sites.

Vision

Unleash a Galaxy of Analytical Possibilities with the Power of Unified Data to improve Manufacturing Processes – Optimize Cost, Efficiency and Quality

Architect to **Modernize** our Data and Analytics and warehouse environments

Establish a **One** data lake for both Process and Business data types

Define Common Schema for data types across sites with **High Quality** and Short Latency for ML and analytics

Select the **Right Partners**

Upskill our internal resources

Build **Center of Excellence**

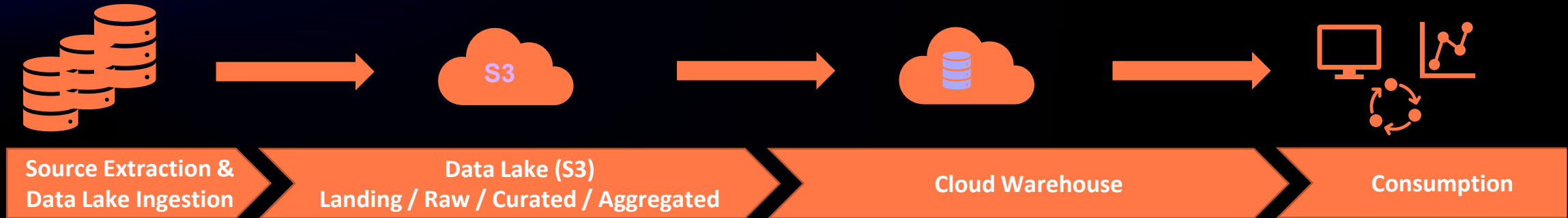
Establish **Governance** Model

Implement **Target Operating Model**

Application and user migration to cloud-native data lake solutions

Sunset legacy, on-prem, discrete regional solutions

2-Year journey to the cloud



Vision

Enable power of cloud on **One** data: Scalability / Uptime / Security / ML

Mission

Modernize our Data & Analytics and warehouse environments through an AWS partnership

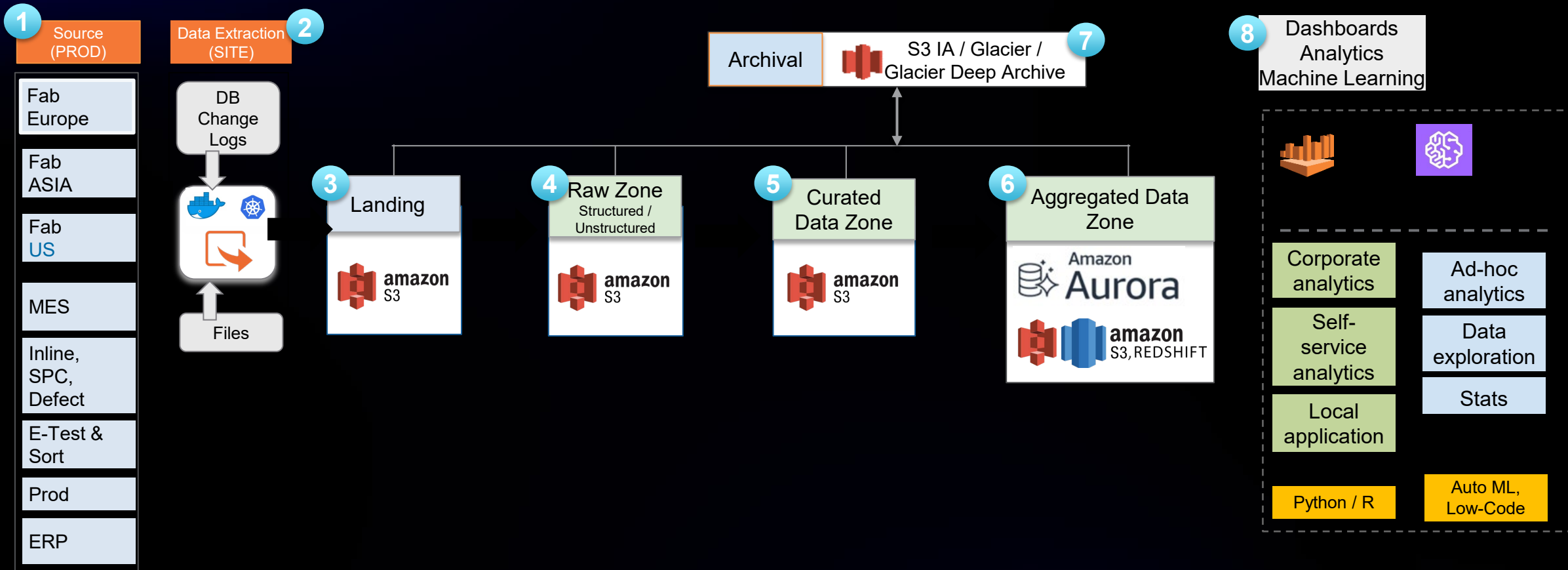
Establish a **One** data lake for both Process & Business data types

Define Common Schema for data types across sites with high quality & short latency for ML & analytics

Application and user migration for existing data and applications to cloud-native data lake solutions

Sunset legacy, on-prem, discrete regional solutions

Cloud Data Lake Architecture



Landing: Data Ingestion / Staging Area

Raw-Unstructured: Golden Copy / Long term retention / Reprocessing

Raw-Structured: Standard file formats (AVRO, Parquet).

Curated: OneGF schema

Aggregated: Merged / Final Copy of Data before loading into live DBs (Postgres/Redshift)

Results

100 Billion

rows inserting to
Cloud data lake / day

3x Faster

Performance Querying
data Vs On-Prem

10 TB

Volume of data
uploaded to AWS / day

~20 Min

To load 1-day volume of
data to the Cloud (6x Faster)

5 PB+

Volume of data
uploaded to AWS as of today

10 Gbps

Speed to access the
Cloud data from On-Prem

ONE

Schema for All the FABs across the
Globe Supporting 3 Geos & 5 Sites

10K + Table

1M+ Columns

Compare high volume tool data set across all Mfg units to benchmark and drive **throughput increase**.

Image classification quality inspection leveraging ML capability driving significant productivity gain | 12hrs to <3mins turn-around time **240x improvement**.

Virtual metrology AI initiative driving **capex reduction** and tool thruput increase.

Cloud ML predictive maintenance solution helped extend maintenance from **50k wafers to > 100k wafers**.

Lead time reduction for NPI on RF tech ramp from 21days to 9days (57% reduction) enabling **speed to market**.

What's Next



EXTEND

ML Ops framework
to support Gen AI
use cases



SCALE

ML Use cases across
all functions



BUILD

data products and operating
model to monetize data

Appendix

Architecture & Services

Fab 8 / Fab 9 – USA

Fab 7 / Giga – Singapore

Fab 1 – Germany

On Premise Sources

Manufacturing

MES

Equipment

Inline

E-Test and SORT

Defect

FDC

SPC

Enterprise

Finance

Supply Chain

Sales

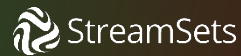
Procurement

HR

Ingestion Pipelines



CDC Change Logs



AWS DataSync



Messages Queues, File shares, offline data

Store, Process, and Analyze

ETL Processing and Orchestration



Amazon EMR



Amazon DynamoDB



MWWA

Monitoring and Validation



Amazon OpenSearch



Amazon CloudWatch

Data Storage



Landing Zone



Raw Zone



Curated Apache Iceberg Tables



Archive

Aggregated Data



Amazon Redshift (Provisioned)



Data sharing



Amazon Redshift (Serverless)



Amazon Aurora

Catalog and Discovery



AWS Glue



AWS Glue Data Catalog



Business data catalog



Amazon DataZone

Governance



AWS Lake Formation



AWS Control Tower

Analytics and Consumption



Amazon SageMaker



Amazon Athena

BI Platform

Analytic Platform

ML/AI Platform