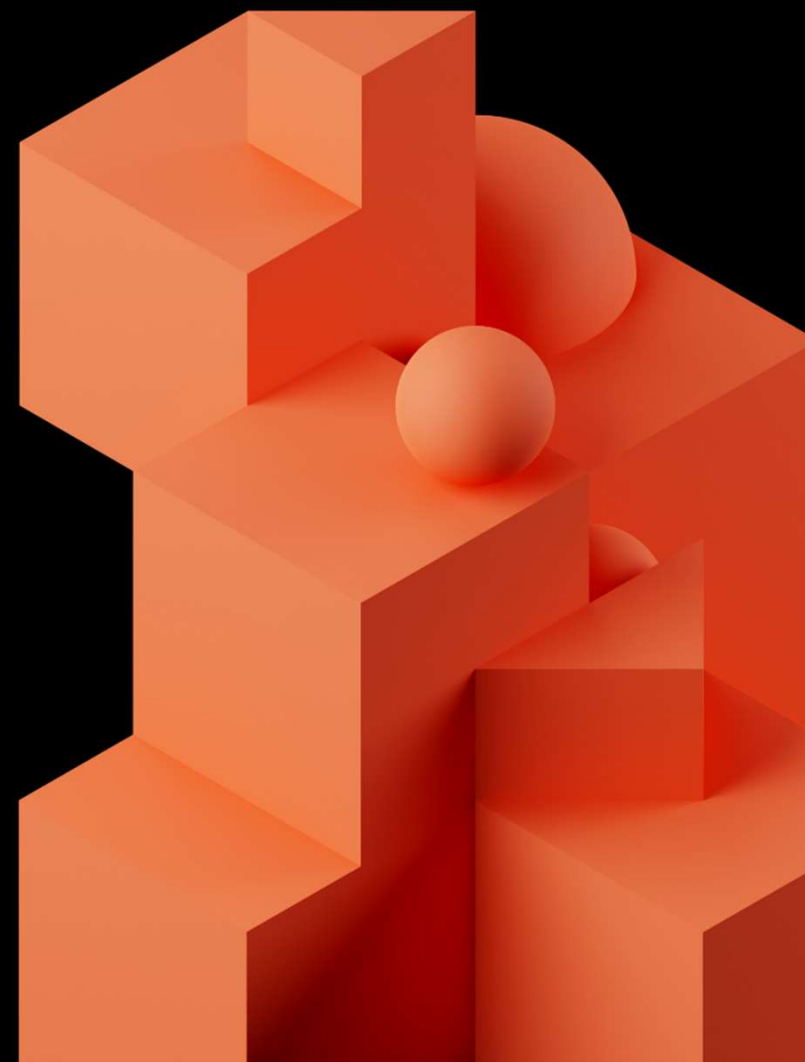


**DATA+AI
SUMMIT**
BY  databricks

Self-Service Analytics and Governance

at Enterprise Scale with Unity Catalog

Databricks
2023



Pascal van Bellen

Senior Consultant | ORAYLIS Germany

- › 7 years of experience
- › Certified Databricks Champion
- › Specialist for Cloud & Big Data



Artem Meshcheryakov

Consultant | ORAYLIS Germany

- › 5+ years of experience
- › Certified Databricks Champion
- › Specialist for Cloud & Big Data



We are your guide on the path to a data-driven company



24 years

of experience in projects and technology



18 years

Microsoft Managed Partner, more than 100 certified consultants



145 colleagues

with burning passion for your data solution



820 projects

in BI, Big Data, IoT, Data Analytics & AI



100% agile

due to team approach in projects and flat hierarchies



15%+

organic growth per year



5 awards

from Great Place to Work in their annual comparison of Germany's best employers



100%

focus on our team

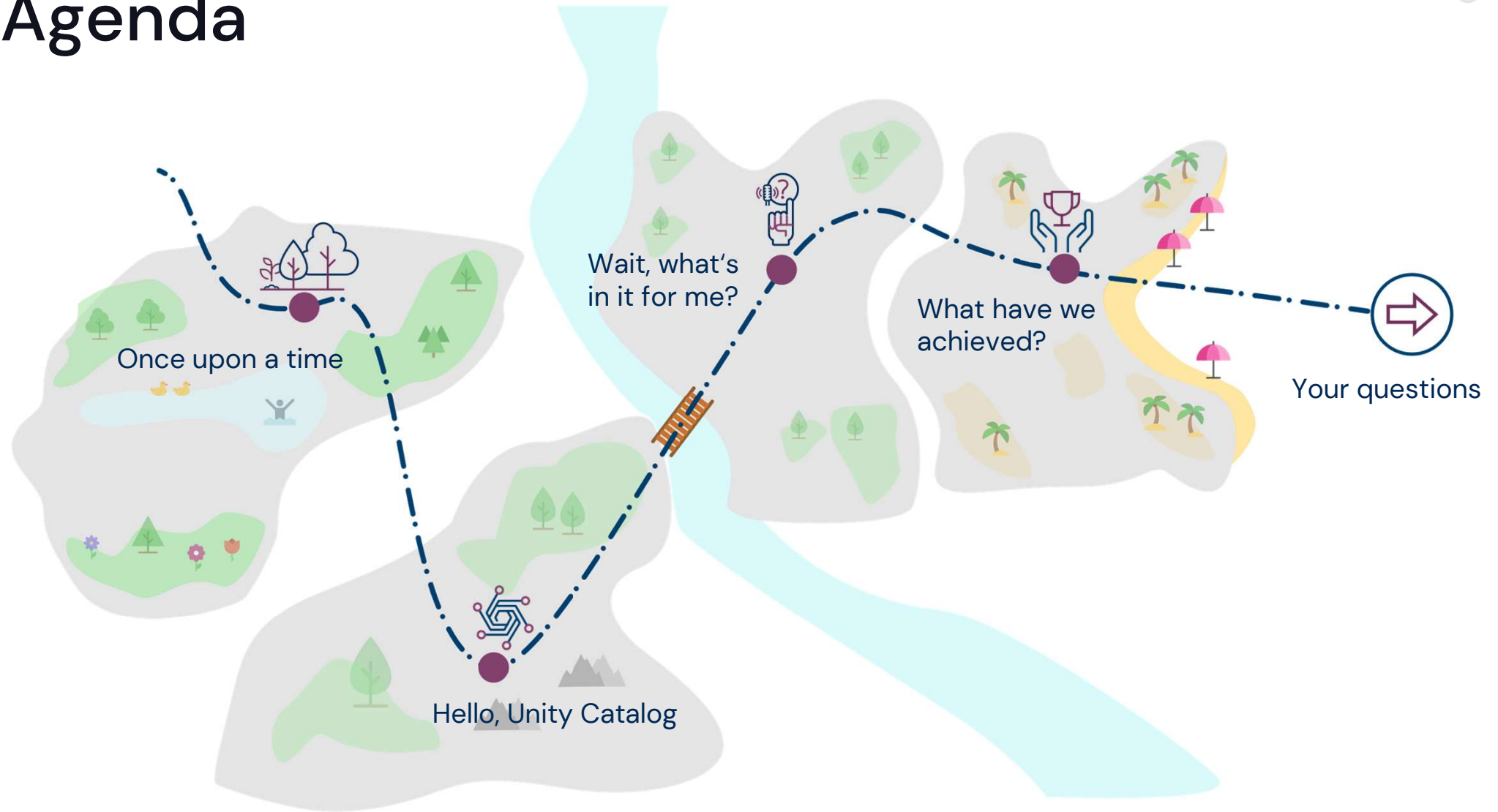
Self-Service Analytics and Governance

Your key takeaways

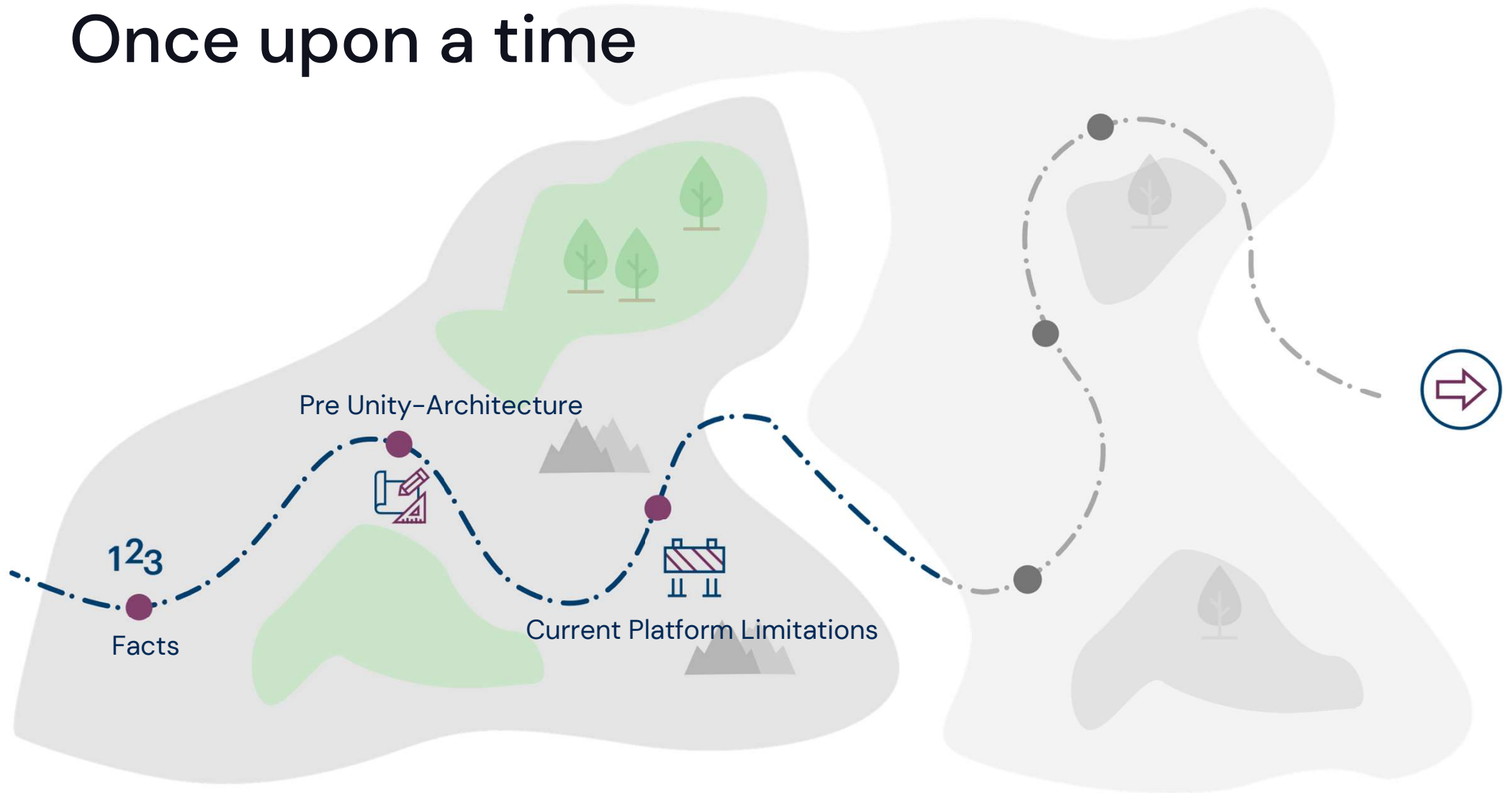
- Governance is for all
 - is an enabler → not a blocker
 - is essential → not optional
 - is recurring → not a one time task
- Your stakeholders are your key
 - as requestor & beneficiary
- Unity Catalog unifies your Lakehouse governance
 - enables accessibility & security & maintenance



Agenda



Once upon a time



Once upon a time

Start of our journey

- Historically grown
- Complex dependencies
- A number of active use cases & large user group
- Tool & architecture limitations
- High security expectations



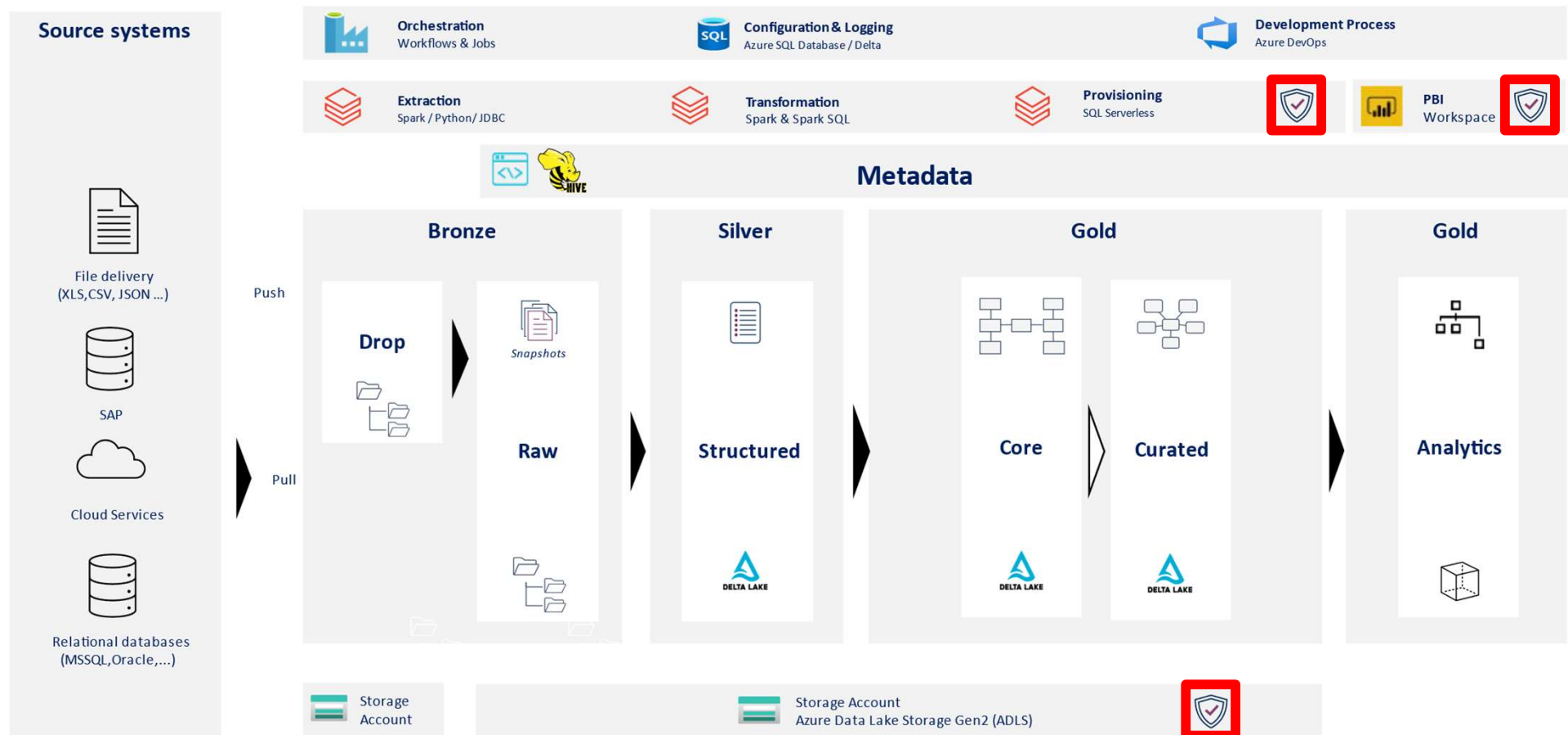
Once upon a time

Facts 1²3



Once upon a time

Pre Unity-Architecture



Once upon a time

Current platform limitations:



1. Limited access to global data



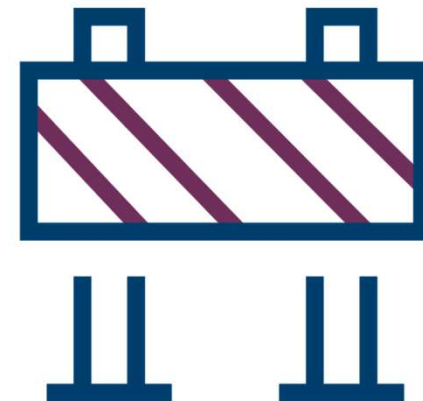
2. Limited global collaboration



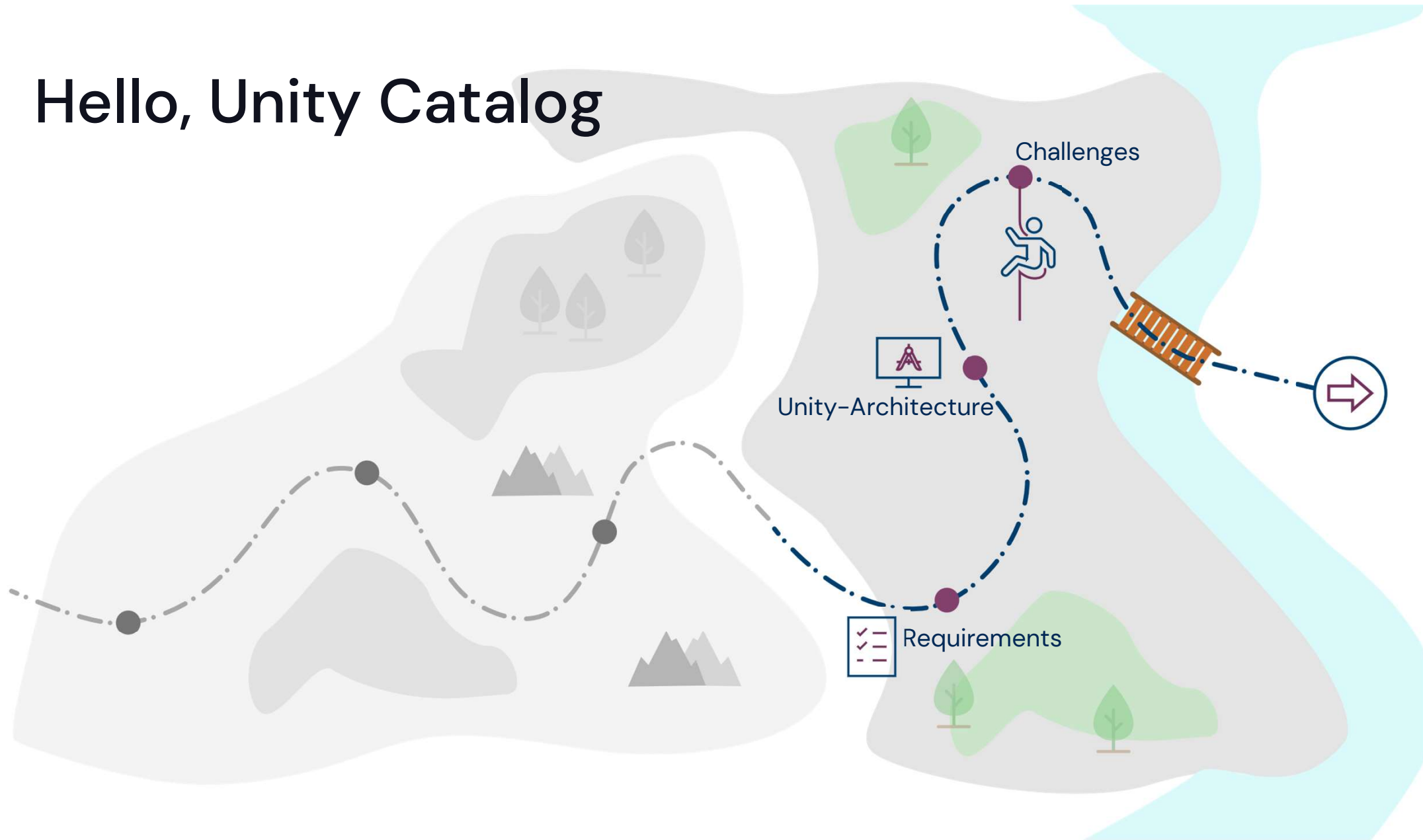
3. Missing unified governance



4. Missing compliance capabilities



Hello, Unity Catalog



Hello, Unity Catalog

Requirements



1. All Pre-Unity platform requirements



2. Enable global data access, data visibility and sharing



3. Enable global collaboration



4. Support sophisticated permission concept

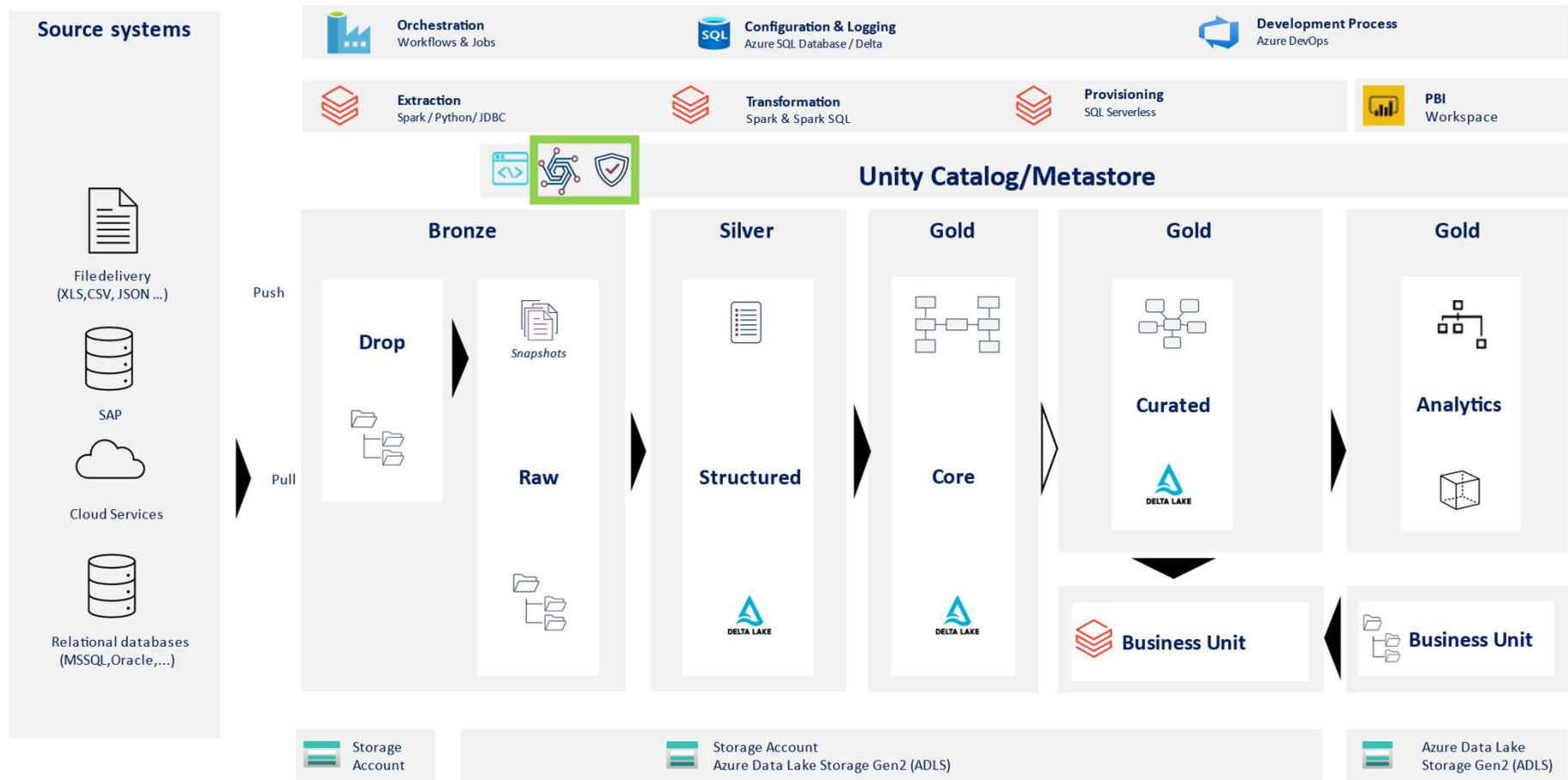


5. Fulfill Governance, security and compliance expectations



Hello, Unity Catalog

Unity-Architecture



Hello, Unity Catalog

Challenges

- Perform upgrades over multiple Runtime versions
- Some limitations for shared clusters (now documented)
- A super tight project deadline
- An incomplete migration path
- Struggle with SCIM Connector collaboration in AAD
- To set up a complex access concept



Wait, what's in it for me?



Wait, what's in it for me?

Considerations & Actions



Our questions:

- XLarge platform
- Distributed storages
- Big Bang or rolling wave ?



Challenges to tackle:

- Libraries (Pandas, Excel, ...)
- Limit Python UDF's
- No Scala, no RDD



How to design for:

- Cross unit
- How to use global data

Wait, what's in it for me?

Considerations & Actions



Our experience:

- Big Bang → failed
- Rolling wave → worked
- Segmentation is easier



Our solution:

- Get rid of technical debt
- Review & replace
- Simplify & use Spark



Our solution:

- Layer concept

Wait, what's in it for me?

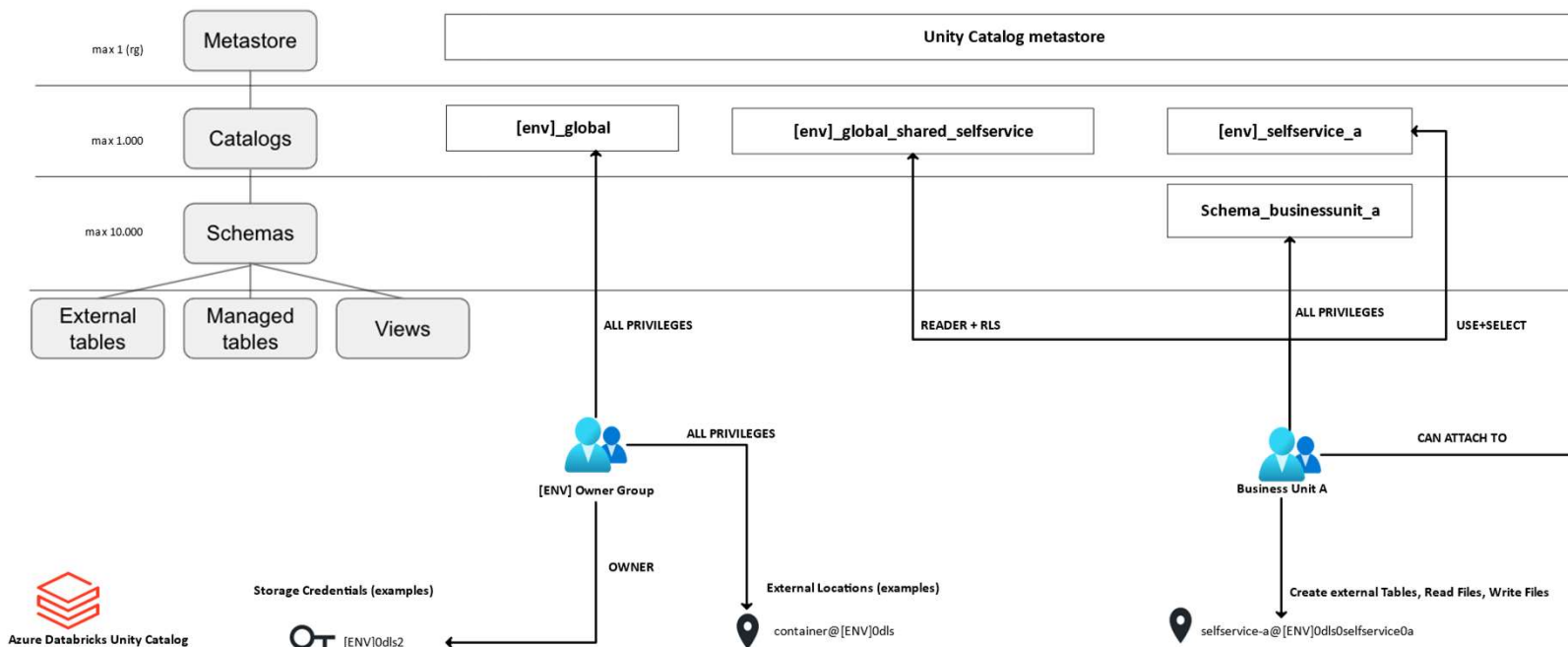
Layer concept



Subscription



Azure Databricks Unity Catalog



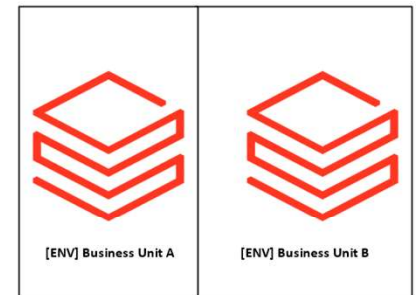
Azure Databricks Unity Catalog



[ENV]0dls2



Self-Service Subscription-[ENV]



Wait, what's in it for me?

Considerations & Actions



Unity consideration

Things to keep in mind:

- Early adopter challenges
- Hard & soft limits
- 3 Tier architecture ?



Cluster consideration

Challenges to tackle:

- Libraries & Init scripts
- Code languages



Objects to migrate

What to migrate and how:

- Tables & Views
- Functions
- Referenced tables
- Metastore database

Wait, what's in it for me?

Considerations & Actions



Unity consideration

Solution:

- Architecture adaption
- Unity Catalog workspace binding
- Close exchange with vendor
- Previews (talk with vendor)



Cluster consideration

Solution:

- Divide & conquer
- Processing architecture
- Capsulation



Objects to migrate

Solution:

- On-board resources (sync)
- Scripts
- APIs

Wait, what's in it for me?

Lessons Learned



Organisational

- Always talk with your stakeholders
- Define operating processes
- Work out a role concept



Technical

- Global AAD Admin
- Object types
- Runtime changes
- Clean up your code
- Representative test
- Try & catch
- SCIM connector



Architectural

- Migration plan
- Multilayer architecture

Wait, what's in it for me?

Dos

- Upgrade to the latest LTS version
- Check infrastructure
- Choose the right access mode
- Set up logging
- Clean your code & test it
- Review & simplify code



Wait, what's in it for me?

Don'ts

- Upgrade multiple DBRs at once
- Skip runtime LTS versions
- Migrate in Big Bang
- Think "I don't need logging"
- Think "I will fix it later"
- Do it via GUI only



What have we achieved?



What have we achieved?

Business Value



Enablement

- Add governance path
- New user groups
- Data democratization
- Data Sharing
- Sophisticated security & permission



Added Value

- Self service capabilities
- Data federation
- Accelerate growth
- Improved access mgmt.
- Lineage capabilities
- Higher user orientation
- Operational cost reduction



Unlocked capabilities

- Foundation for new Business Cases
- Data Mesh architecture
- Comprehensive data governance
- Usage monitoring
- Query Tuning



What have we achieved?

Project Outlook

- Use of Query Federation
- Attribute based access control (“Tagging”)
- Classification Tagging
- Global rollout Unity Catalog data lineage
- Use of Volumes



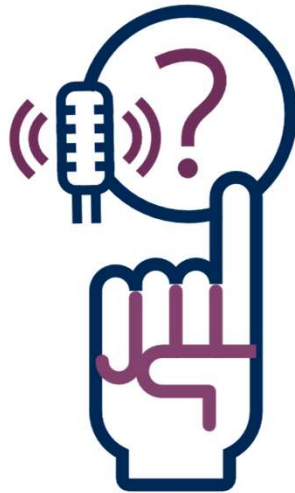
What have we achieved?

Your key takeaways

- Governance is for all
 - is an enabler → not a blocker
 - is essential → not optional
 - is recurring → not a one time task
- Your stakeholders are your key
 - as requestor & beneficiary
- Unity Catalog unifies your Lakehouse governance
 - enables accessibility & security & maintenance



Q&A



ORAYLIS



turn your data into value

powered by



databricks

