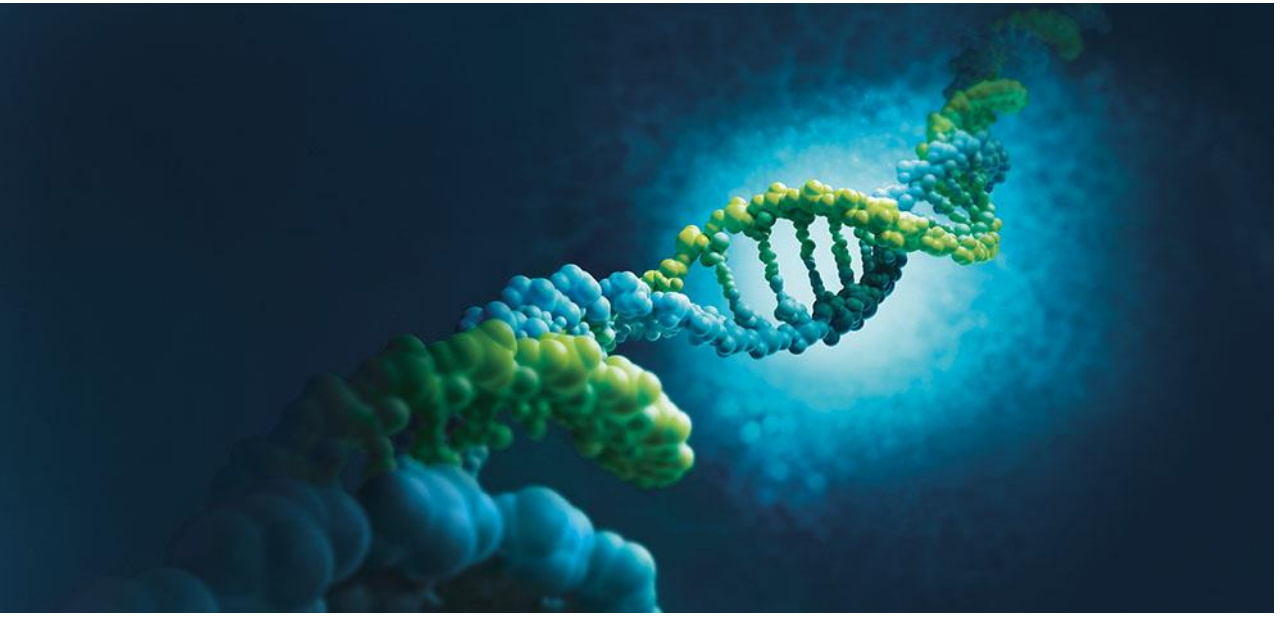
Science For A Better Life



iRODS and Objectstorage

UGM 2016, Chapel Hill

2016-06-08 / Othmar Weber, Bayer Business Services / v0.2

Agenda

- iRODS at Bayer
- Situation and call for action
- Object Storage PoC
- Pillow talks

Overview

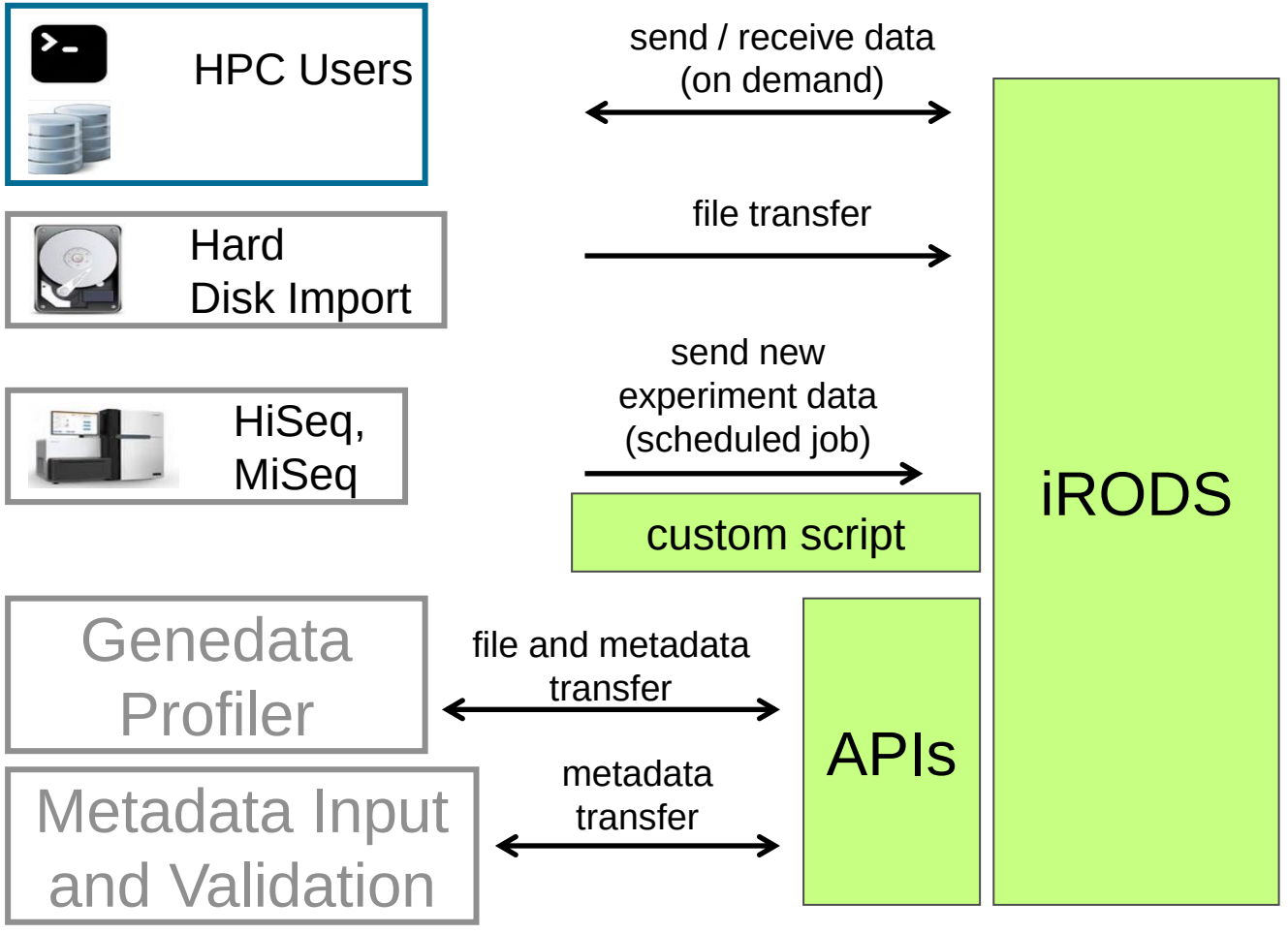
iRODS@Bayer

- Introduced in 2014 by an IT research project for LifeSciences
- Developed as internal data platform to securely store and share genomic data
- Integrated into the Bayer specific environment
- Identified as key enabler for future research data strategy
- Provided as a solution by a joint operations team (DevOps style)
- 2016: Bayer joined the iRODS consortium

- Used by
  - HC Bioinformatics (Research) for data management
  - HC Biomarker as an archival system
  - CS Digital Farming
  - … in consideration by other projects

iRODS integration example

HealthCare Bioinformatics



transparent storage access

file system, object storage

metadata storage and query

user + access rights management

PRIMA 4 PWP
IAM IDENTITY & ACCESS MANAGEMENT

audit trail

iRODS setup 1.0

- Traditional deployment like any other server
  - iCAT: VM with local database (local=SAN)
  - iRES: VM with local filesystem (local=SAN)

It works. Why should I change something?

Ever heard of NGS?

- NGS means ~ 300 GB per individual
- How do we handle the data of a study with 1000 individuals?

iRODS setup 1.0

- Deployment like any other server
  - icat: VM with local database (local=SAN)
  - ires: VM with local filesystem (local=SAN)

How does this work at Petabyte scale?

It simply does not!

- Cost: SAN storage + Backup
- DR: Data mirror in 2nd site needed
- Recovery: How do I revert a user or system error?

How should a storage solution in PB scale look like? (Whishes are for free …)

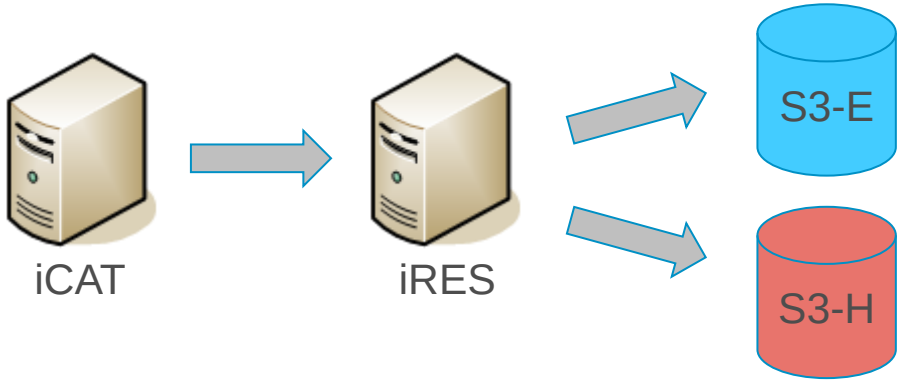
- highly scalable (from TB to PB)
- low price tag
- fully automated
- low administrative effort (self service?)
- backup included by design (replication + versioning)
- DR included by design (replication)
- data security features (fast rebuilt, check summing, self healing, access control, encryption)
- appliance with vendor support (no D.I.Y.)
- S3 interface support (de facto standard with iRODS plugin available)
- performant

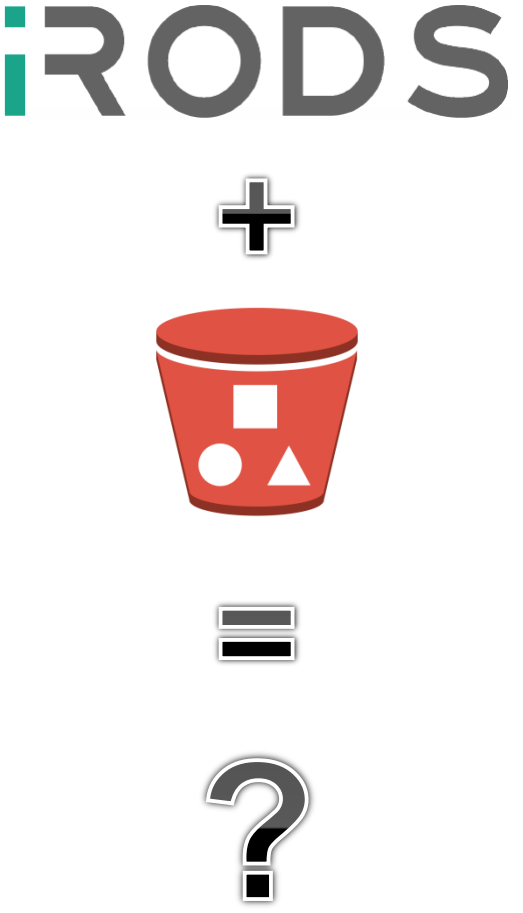
➢ **Object Storage!**

The Proof of Concept

Setup

- Setup:
  - two different storage solutions
  - S3 interface
  - dedicated resource server with local file systems
  - both storage solutions connected to the same resource server
  - iRODS version 4.0.3 (start), 4.1.8 (later)

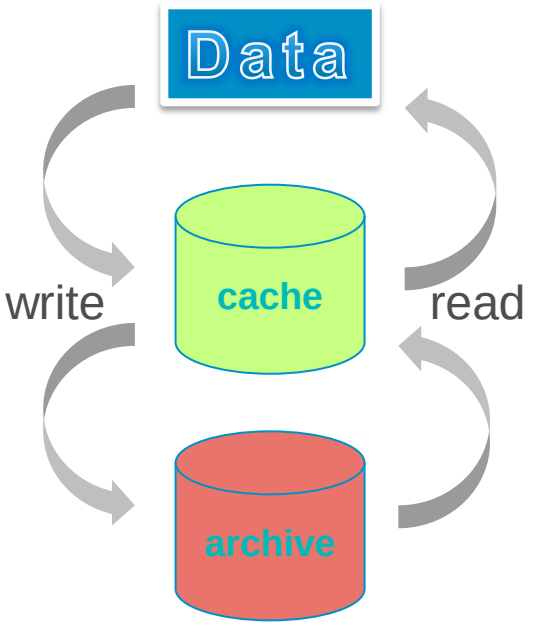




The Proof of Concept

iRODS integration (Compound resource)

- S3 resource is implemented as compound resource
  - Cache and Archive resource
- No direct storage access
  - additional cache file system needed
  - dedicated resource server with local file system
  - both storage solutions connected to the same resource server
- Observations/Findings
  - additional caching tier is not for free (housekeeping, performance)
  - consistency monitoring between cache and archive necessary
  - iput/iget provide purge cache option to delete cached objects
    - not available for all commands



The Proof of Concept

iRODS S3 resource plugin

- current available plugin is at version v1.3
  - compatible with iRODS v4.1.8 only
- Support of multiple connection parameters (->)
- Relies on libs3 & libcurl

- Object versioning was enabled on bucket level
  - restore script to revert old versions

Parameter	Meaning
S3_DEFAULT_HOSTNAME	S3 endpoint address (e.g. s3.acme.com:80)
S3_AUTH_FILE	File with Access key and secret key
S3_KEY_ID	Access key (instead of S3_AUTH_FILE)
S3_ACCESS_KEY	Secret key (instead of S3_AUTH_FILE)
S3_RETRY_COUNT	Number of retries on connection error
S3_WAIT_TIME_SEC	Time between retries
S3_PROTO	HTTP or HTTPS
…	More to come …

```
Example:
iadmin mkresc S3archive s3 server:/bucket/irods/vault
"S3_DEFAULT_HOSTNAME=s3.acme.com:80;S3_AUTH_FILE=/var/lib/irods/.S3_Auth;S3_RETRY_COUNT=3;S3_WAIT_TIME_SEC=3;S3_PROTO=HTTP"
```

The Proof of Concept

iRODS S3 resource plugin observations and findings

- Load balancing somehow tricky
  - S3 endpoints are usually hidden behind a load balancer
  - iput with bulk option was only using one S3 connection
    - multiple iput requests were faster because of load balancing
- Configuring multiple endpoints on same iRES did not work (issue #~~1826~~)
- AWS v4 signatures currently not supported (regions eu-central-1 and ap-northeast-2, issue #1831)
- Renaming issues for files with special characters (issue #1829)
- Renaming issues for large files (no multipart copy, #1830)
  - rename is a two step procedure (copy & delete) -> processing time related to object size!
  - consistency issues between iRODS catalog and S3 bucket
- Issue to delete orphan catalog entries (issue #~~3154~~)
- ichksum did not work with archive resources (hierarchy issue, #3127)

Pillow talks (1/2)

How to improve the relationship of iRODS and object storage

- iRODS currently relies on a POSIX file system
  - implemented as compound resource
    - Native object implementation possible?
- ✓ S3 plugin uses catalog entries as object keys
  - leads to issues (rename timeout, …)
  - archive nature of object storage should be considered
    - decouple object and catalog entries
- Utilize object storage checksum capabilities (iscan, ichksum)
- Consider increasing parallel sessions on bulk transfer to better utilize load balancing
  - multiple single files in bulk upload
  - multipart upload (~~#1824,#1827~~)

Pillow talks (2/2)

How to improve the relationship of iRODS and object storage

- Error handling for object storage access
  - object storage needs to be handled different than POSIX files
    - see: S3 ErrorBestPractices.html
- Think about metadata usage in object store
  - metadata is a feature of object storage
  - why not use this feature for recovery purposes?

❖ Things get better with plugin version 1.4:

- V4 signature, decoupling object names, renaming, multipart, new parameters, …

Any questions?

OTHMAR WEBER
Scientific Computing Competence Center

Phone: +49 – 214 – 30 56972

E-mail: othmar.weber@bayer.com

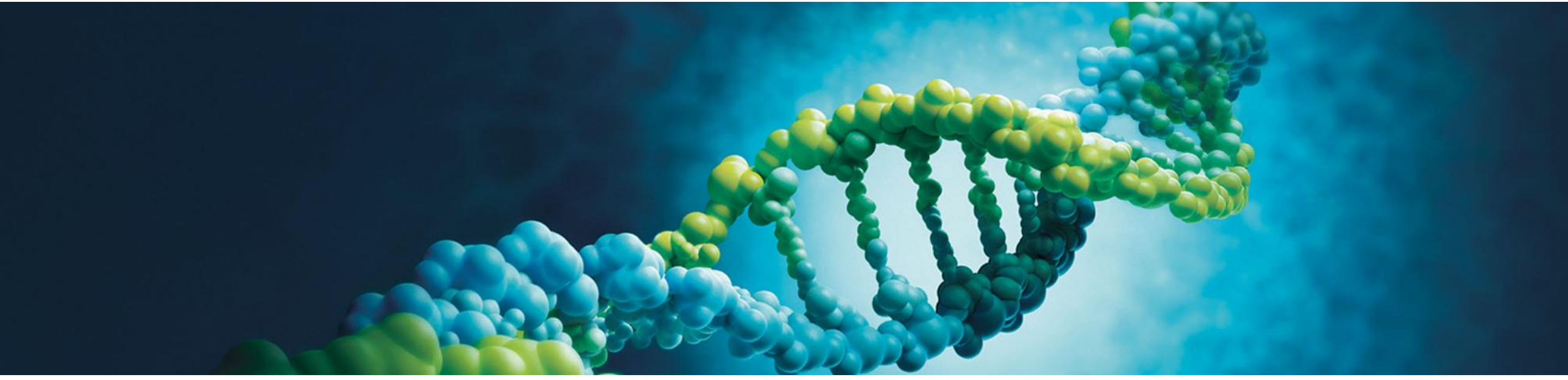
Forward-Looking Statements

This website/release/presentation may contain forward-looking statements based on current assumptions and forecasts made by Bayer management.

Various known and unknown risks, uncertainties and other factors could lead to material differences between the actual future results, financial situation, development or performance of the company and the estimates given here. These factors include those discussed in Bayer's public reports which are available on the Bayer website at http://www.bayer.com/.

The company assumes no liability whatsoever to update these forward-looking statements or to conform them to future events or developments.

Science For A Better Life



Thank you!

Image credits

- Slide 6, DNA Sequencer CC BY 2.0, link to picture
- Slide 8, Think out of the box CC0, link to picture
- Slide 9, S3 Bucket, CC BY-SA 3.0, link to picture
- Slide 13/14, Pillow talk CC BY 2.0, link to picture