IBM Spectrum Scale Strategy Days 2021

—

HDFS Update and Customer Insights from HUK-COBURG

Dr. Qais Noorshams, IBM DE
Renar Grunenberg, HUK-COBURG

IBM Spectrum Scale HDFS Transparency

HDFS

NFS

SMB

Object

Other

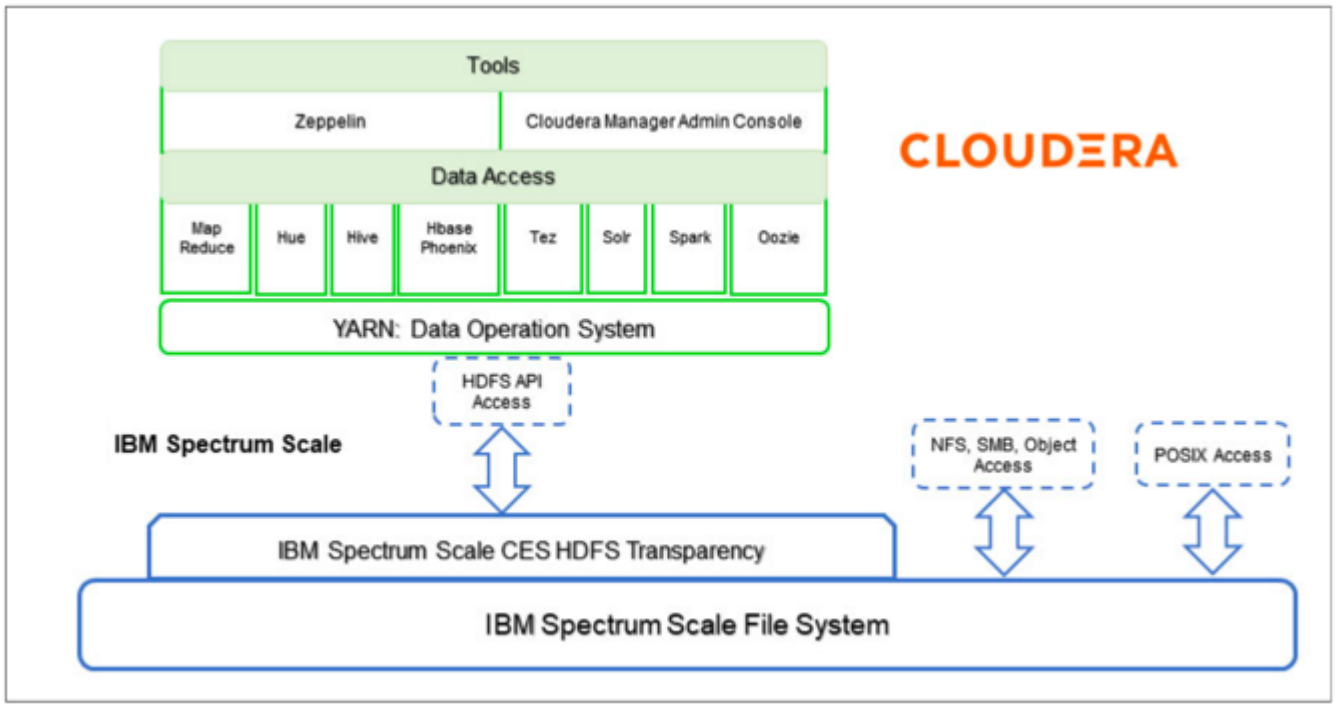
IBM Elastic Storage Servers

- Analytics and data insights:
  Run AI & ML workloads seemlessly over HDFS on IBM Spectrum Scale
- Best of both worlds:
  Exploit the open source and the IBM Spectrum Scale ecosystem
- Save space, time, and money:
  No data copying, no redundant data replication, built-in data protection

CDP Private Cloud Base certification with IBM Spectrum Scale

Official support:

- CDP Private Cloud Base is the on-premises version of Cloudera Data Platform (CDP)
- CDP Private Cloud Base is certified on CM 7.2.3 and CDH 7.1.4 with IBM Spectrum Scale 5.1.0.1 (incl. CES HDFS Transparency 3.1.1-3)
- The installation toolkit can install HDFS Transparency as part of the CES protocol stack

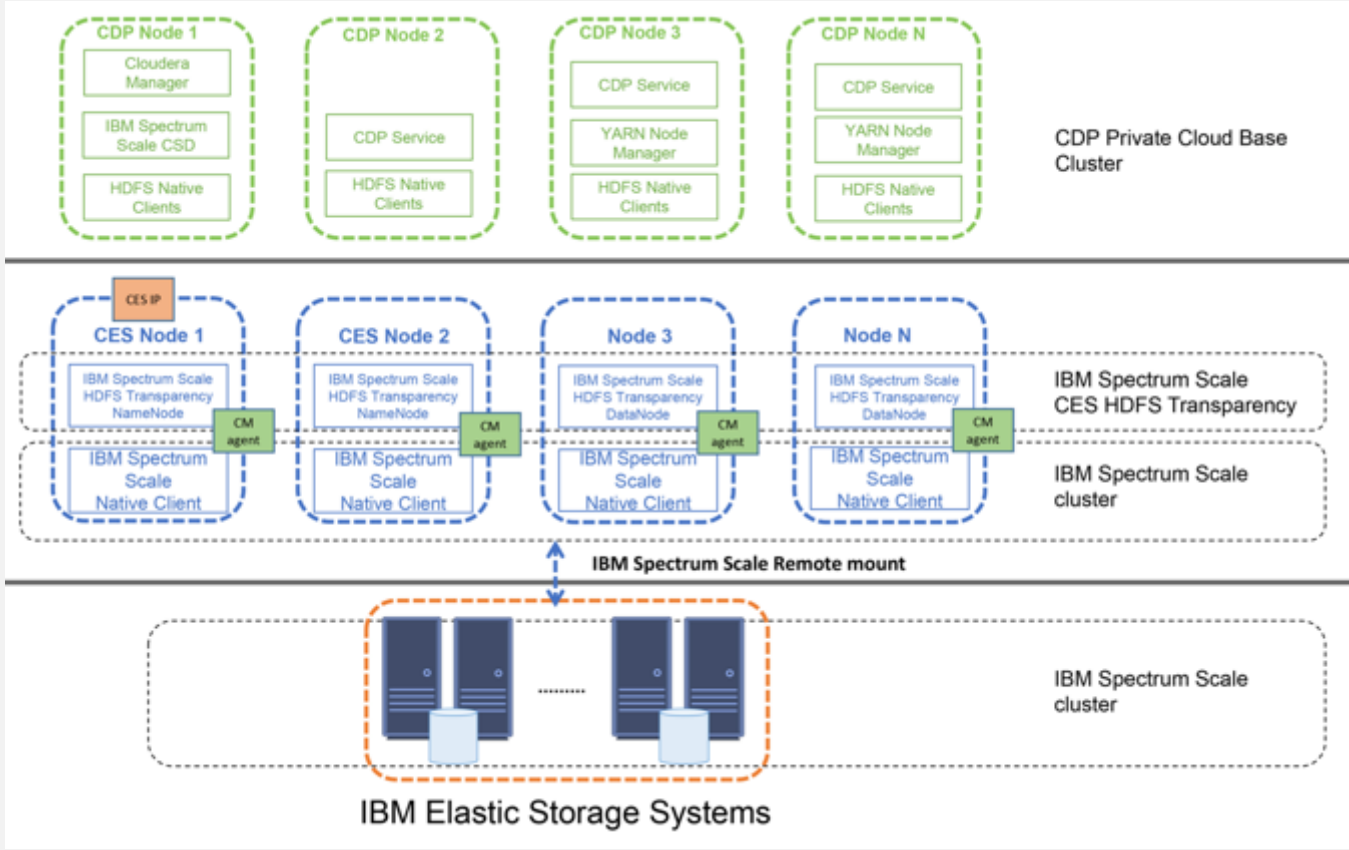


http://www.redbooks.ibm.com/abstracts/redp5608.html
https://www.ibm.com/support/knowledgecenter/de/STXKQY_BDA_SHR/bl1bda_cdpsupportmatrix.htm

CDP Private Cloud Base certification with IBM Spectrum Scale

Technical highlights:

- Loosely coupled architecture between Cloudera and IBM Spectrum Scale
- NameNode HA by CES HA support
- Administration and metrics integration of HDFS Transparency with Cloudera Manager (CM)
- Health checks and monitoring available using the IBM Spectrum Scale management GUI

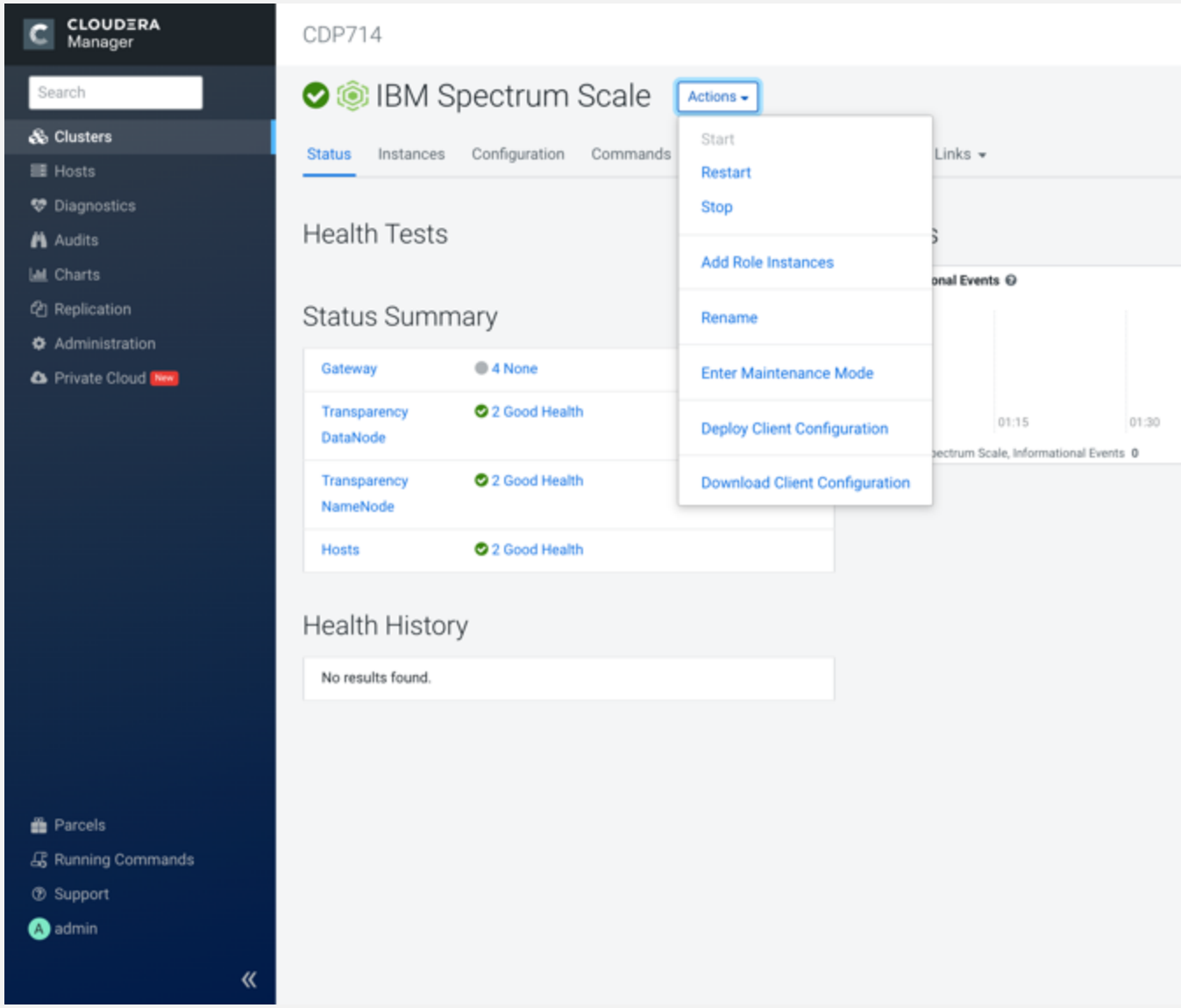


http://www.redbooks.ibm.com/abstracts/redp5608.html

CDP Private Cloud Base certification with IBM Spectrum Scale

Technical highlights:

- Loosely coupled architecture between Cloudera and IBM Spectrum Scale
- NameNode HA by CES HA support
- Administration and metrics integration of HDFS Transparency with Cloudera Manager (CM)
- Health checks and monitoring available using the IBM Spectrum Scale management GUI

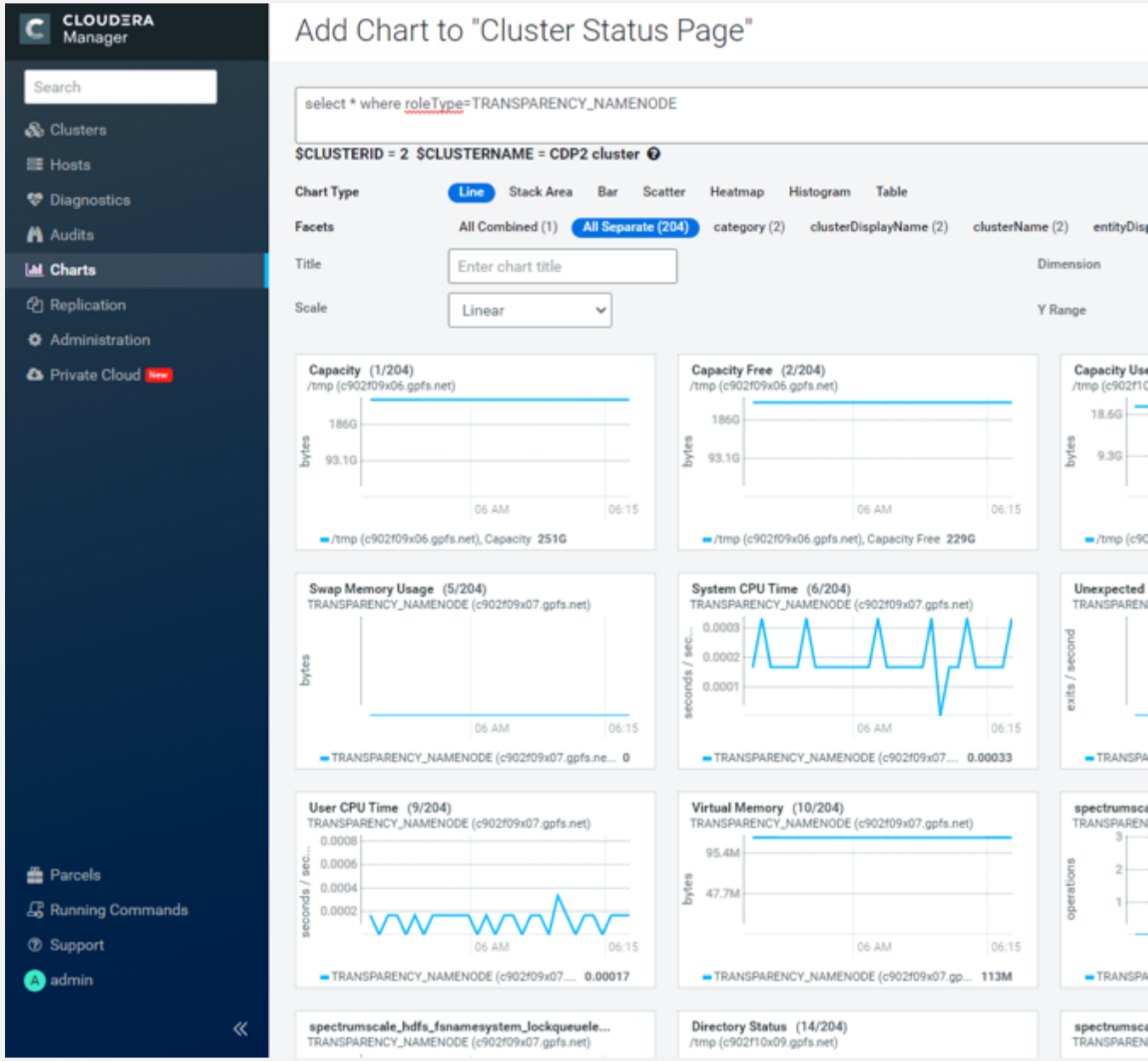


http://www.redbooks.ibm.com/abstracts/redp5608.html

CDP Private Cloud Base certification with IBM Spectrum Scale

Technical highlights:

- Loosely coupled architecture between Cloudera and IBM Spectrum Scale
- NameNode HA by CES HA support
- Administration and metrics integration of HDFS Transparency with Cloudera Manager (CM)
- Health checks and monitoring available using the IBM Spectrum Scale management GUI

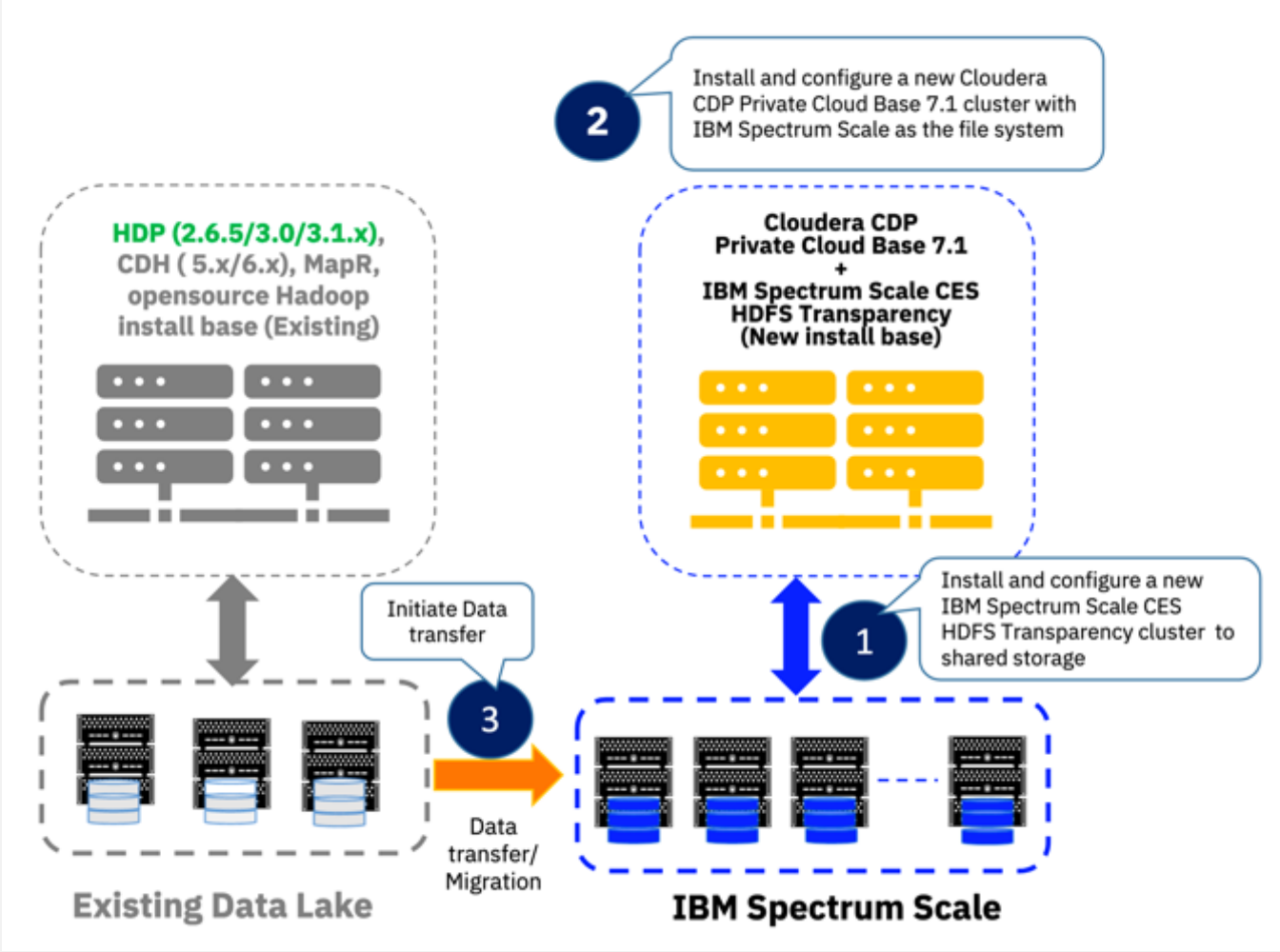


http://www.redbooks.ibm.com/abstracts/redp5608.html

Migration to CDP Private Cloud Base with IBM Spectrum Scale

Side-by-side migration:

- Instantiate a new CDP Private Cloud Base with IBM Spectrum Scale cluster and move the data to the new environment
- You can test your applications gradually and move workloads while retaining your production environment
- Contact IBM or Cloudera professional services to support your migration



http://www.redbooks.ibm.com/abstracts/redp5608.html

HDFS Transparency Releases

Performance improvements with new architecture

Transparency 3.1.0-6 and 3.1.1-3:

- New file system locking architecture with a fine-grained locking approach for increased parallelism
- Set *gpfs.ranger.enabled* to *scale* (replaces *true/false* options) in gpfs-site.xml to exploit the new architecture
- Parameter *gpfs.ranger.enabled* only controls Apache Ranger support in HDFS Transparency – you need to setup Ranger service and policy in order to use Ranger

https://www.ibm.com/support/knowledgecenter/STXKQY_BDA_SHR/bl1bda_soc.htm

Performance improvements with parallelized APIs

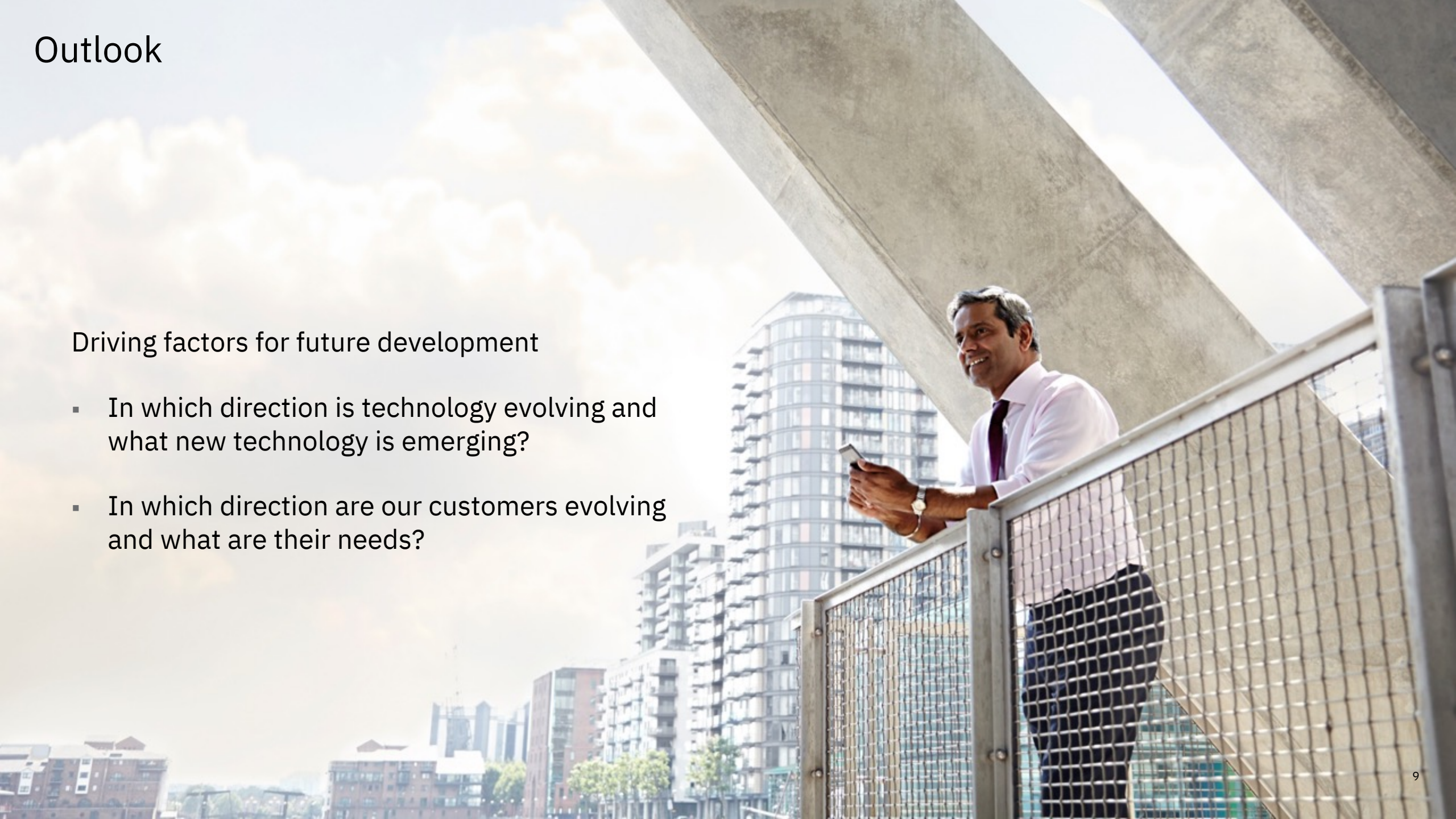
Transparency 3.1.0-5/6 and 3.1.1-2:

- Multi-threaded RPCs (*getListing, getContentSummary, delete*) for improved file system operations (*hdfs dfs -ls, hdfs dfs -du, hdfs dfs -rm, ...*)
- The improved file system operations can be controlled with parameters based on the command

General improvements in Transparency 3.1.0-7

https://www.ibm.com/support/knowledgecenter/STXKQY_BDA_SHR/bl1adv_diff.htm

Outlook



Driving factors for future development

- In which direction is technology evolving and what new technology is emerging?
- In which direction are our customers evolving and what are their needs?

Links

- Cloudera Data Platform (CDP) Private Cloud Base with IBM Spectrum Scale documentation
  https://www.ibm.com/support/knowledgecenter/STXKQY_BDA_SHR/bl1ba_cdp.htm
- CES HDFS documentation
  https://www.ibm.com/support/knowledgecenter/STXKQY_BDA_SHR/bl1bda_ceshdfs.htm
- Redpaper: Cloudera Data Platform Private Cloud Base with IBM Spectrum Scale
  http://www.redbooks.ibm.com/abstracts/redp5608.html