OpenWorld 2016
Querying Hadoop/HDFS from PL/SQL
CON6359

September 18–22, 2016
San Francisco

Nicholas Van Wyen
MTI
September 18, 2016

ORACLE

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

Agenda

1. The Real-World
2. The Problem
3. The Solution
4. Considerations
5. Questions

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

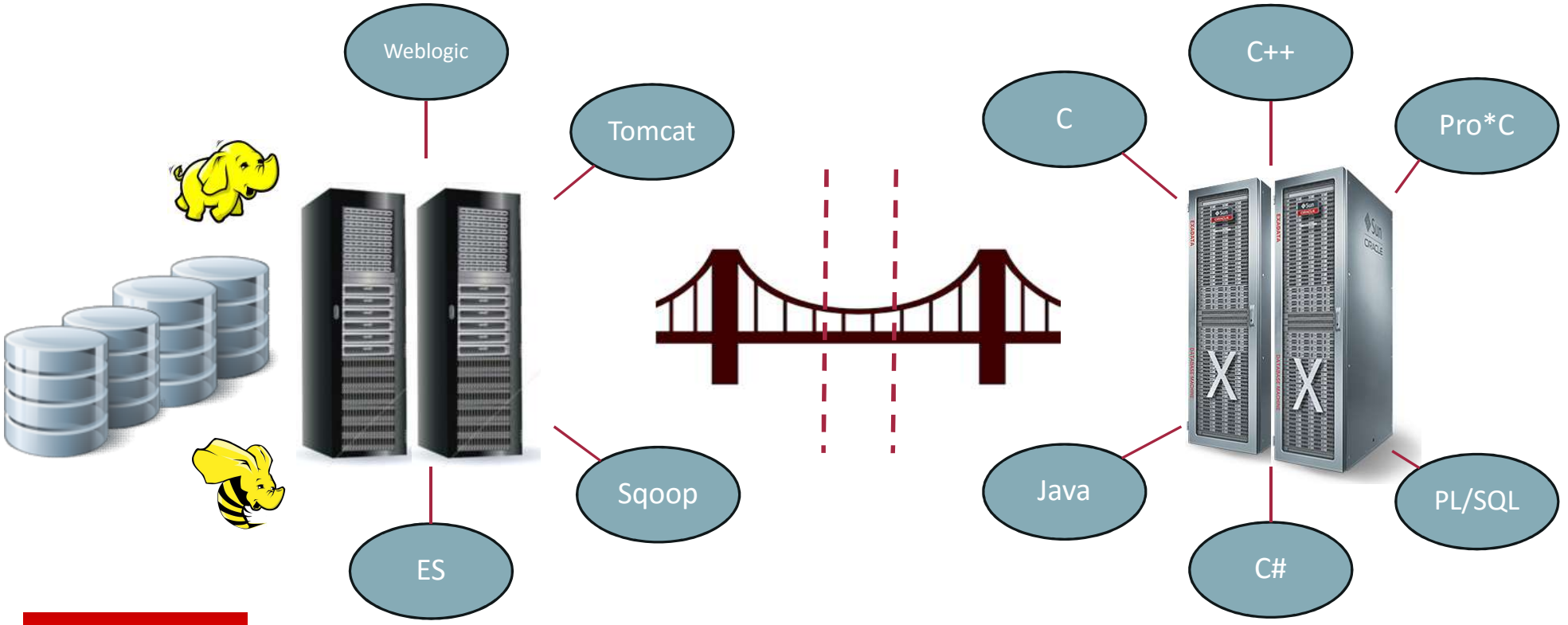
Let’s get started

1. The Real-World
2. The Problem
3. The Solution
4. Considerations
5. Questions

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

The Real-World

Different solutions, for different requirements



Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

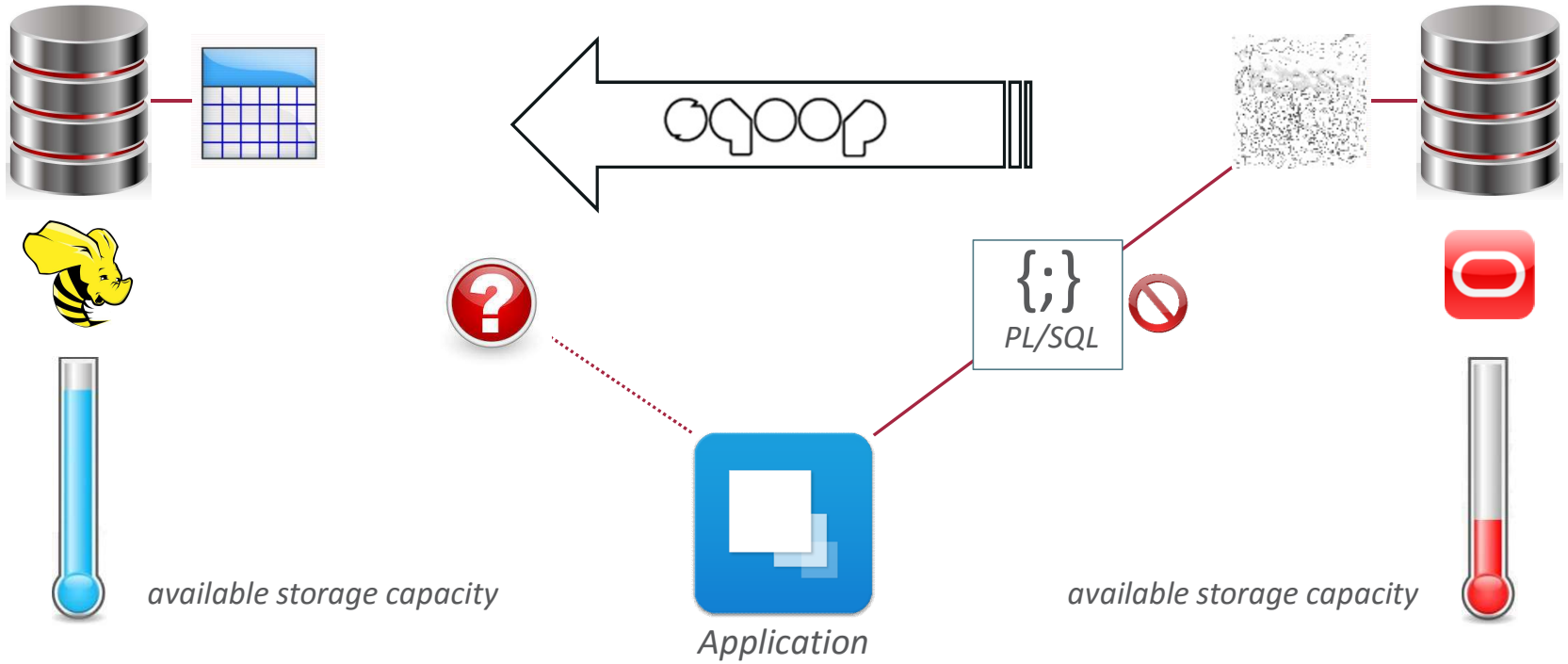
Moving on

1. The Real-World
2. The Problem
3. The Solution
4. Considerations
5. Questions

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

The Problem

Changes

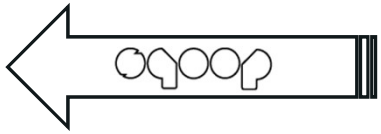


Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

The Problem

Example

```
$ beeline -u jdbc:hive2://hive.corp.com:10000 \
          -n oracle -w welcome1.passwd

  0: jdbc:hive2://localhost:10000> desc user_log;
  +------------------+------------+---------+
  |     col_name     | data_type  | comment |
  +------------------+------------+---------+
  | stamp            | date       |         |
  | account          | string     |         |
  | message          | string     |         |
  +------------------+------------+---------+
```



```
SQL> desc SCOTT.USER_LOG

     Name        Null?    Type
     ----------- -------- ---------------
     STAMP       NOT NULL DATE
     ACCOUNT              VARCHAR2(30)
     MESSAGE              VARCHAR2(4000)
```

```
create view user_log_monthly
as
select stamp,
       account,
       message
  from scott.user_log
 where stamp between sysdate - 30
                 and sysdate;
```

```
procedure user_report( p out xmltype ) is
begin

    for rec in ( select account,
                        message
                   from scott.user_log
                  order by account ) loop
     ...

end user_report;
```

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

Next

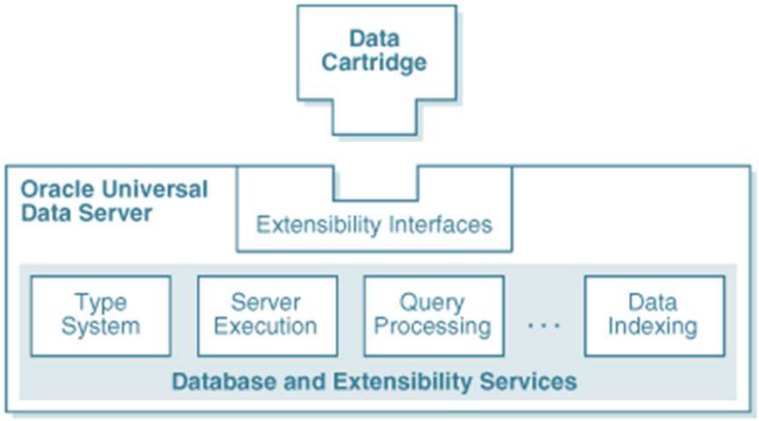
1. The Real-World
2. The Problem
3. The Solution
4. Considerations
5. Questions

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

The Solution

Introduction

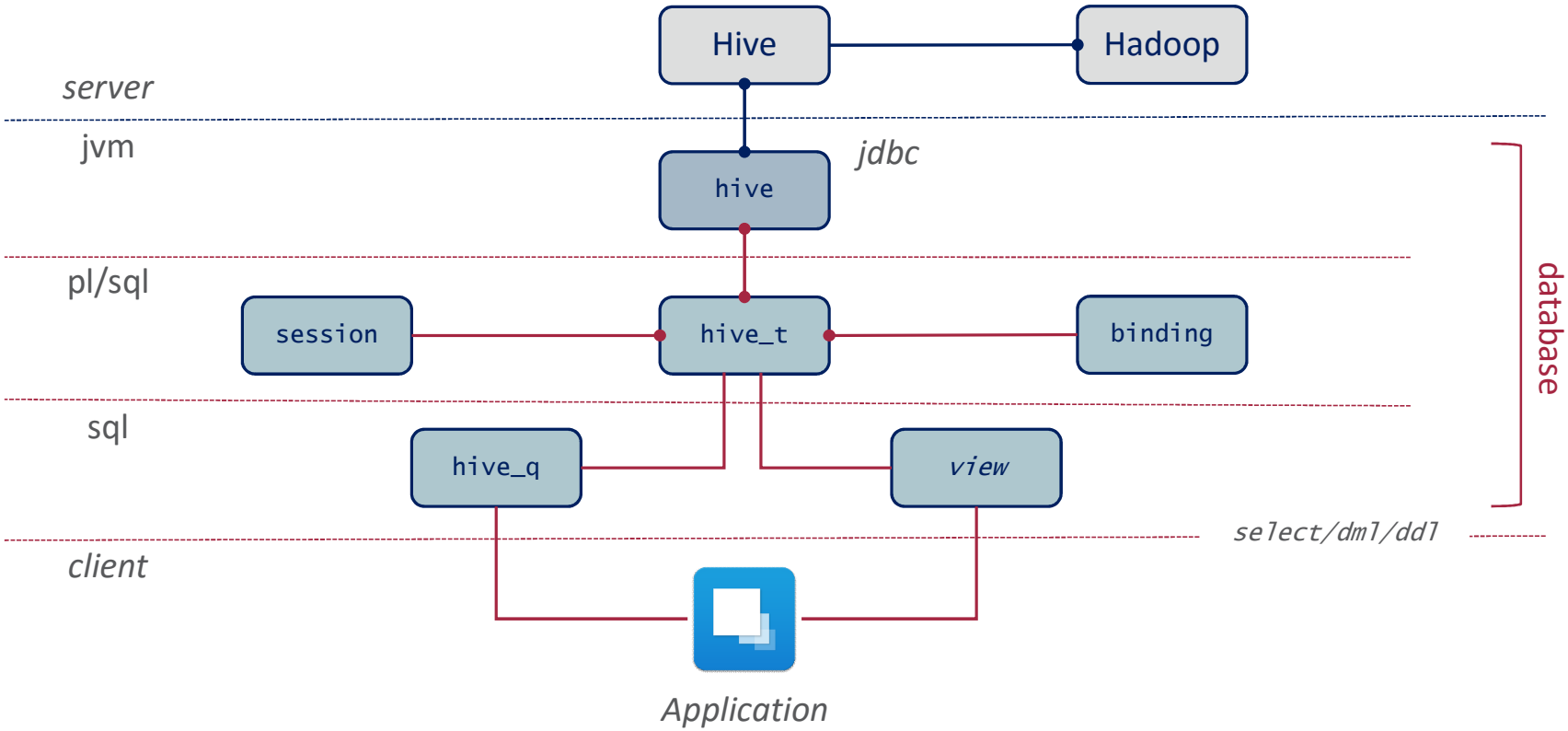
- Presenting Hive-ODCI
  - Built on Oracle Data Cartridge Interface
  - Dynamically accessing Hadoop/Hive within an Oracle 12c RDBMS
    - A JDBC Hadoop/Hive connector
  - Allow for First-Class Oracle objects
  - Leveraging existing RBAC
  - Support for active Bind variables
    - User defined, Static or Saved
  - Native support for SQL and PL/SQL
  - Easy to use and integrate



Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

The Solution

Overview

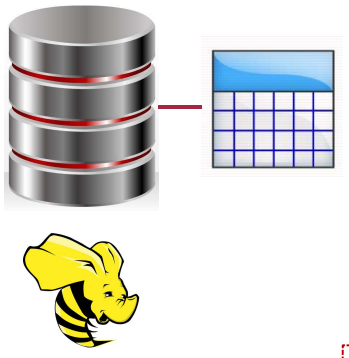


Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

The Solution

Example

```
param( 'hive_jdbc_url',   'jdbc:hive2://hive.corp.com:10000' );
param( 'hive_jdbc_url.1', 'user=oracle' );
param( 'hive_jdbc_url.2', 'password=welcome1' );
```



hive_t

```
create or replace view scott.user_log
    (
        stamp,
        account,
        message
    )
as
select *
  from table( hive_q( q'[ select stamp,
                                 account,
                                 message
                            from user_log
                           order by stamp ]' ) )
```

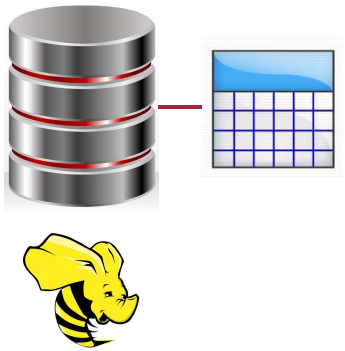
```
SQL> alter procedure scott.user_log_report compile;

   Procedure altered.
```

```
procedure user_report( p out xmltype ) is
begin

    for rec in ( select account,
                        message
                   from scott.user_log
                  order by account ) loop
      ...

end user_report;
```

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

The Solution

Example



```
create or replace view scott.user_log_monthly
       (
           stamp,
           account,
           message
       )
   as
   select *
     from table( hive_q( q'[ select stamp,
                                    account,
                                    message
                               from user_log
                              where stamp between ? and ? ]',
                 hive_binds( hive_bind( to_char( sysdate - 30,
                                                 'yyyy-mm-dd' ),
                                        1 /* type_date */,
                                        1 /* ref_in */ ),
                             hive_bind( to_char( sysdate, ,
                                                 'yyyy-mm-dd' ),
                                        1 /* type_date */,
                                        1 /* ref_in */ ) ) )
```

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

Wrapping it up

1. The Real-World
2. The Problem
3. The Solution
4. Considerations
5. Questions

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

Considerations

In Oracle

- Become familiar with the Hive-ODCI API
  - Read the documentation
  - Ask questions and test, test, test
- Use session isolation whenever possible
  - Particularly authentication, set at the session not the system
- Lean on your experience and your DBA Team
  - Keep signatures consistent
  - Performance tune
  - Change code if necessary

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

Considerations

In Hive

- Analytics over in-line views
  - Review queries and use common sense
- Leverage the CBO and gather statistics
- Use best practices
  - ORCFile
  - Apache Tez
  - Vectorized queries
- Lean on your experience and your BDS Team

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

Considerations

Reach out

- If you have questions, concerns or comments
  - Feel free to contact me
- Available on Github
  - https://github.com/nvanwyen/hive-odci
  - https://github.com/nvanwyen/hive-odci/releases/latest
- Contact
  - nvanwyen@mtihq.com

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

That’s it

1. The Real-World
2. The Problem
3. The Solution
4. Considerations
5. Questions

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. | Confidential – Oracle Internal/Restricted/Highly Restricted

ORACLE®