



# Whats new in Apache Hive 0.9.0

Ashutosh Chauhan  
hashutosh@apache.org



# Compatible with

---

- Hadoop 0.20.2
- Hadoop 0.20.yyy
- Hadoop 1.x.y
- Hbase 0.92
- Hcatalog 0.4
- Zookeeper 3.4.3
- Hadoop 0.23.x (mostly)
- Hadoop 2.0.x (mostly)

# New Operators

---

- **between:**

```
select * from src where key between 200 and 300 limit 1;
```

- **Null-safe equality operator:**

```
select true<=>NULL, NULL<=>NULL limit 1;  
false      true
```

# New Features

---

- Pre-event listeners to metastore
- Insert overwrite table x partition (a=b,c-d) IF NOT EXISTS
- Optionally, get output of ddl commands in json format.
- Null-safe equi-joins
  - Select \* from a join b on a.key <=> b.key

# New Features

---

- Hive hbase integration now works in secure mode
- Can specify a timestamp for your batch of writes in hbase table
- Better interoperability of hive and hbase clients
- Hive 0.8:
  - Hive can only read data written by hive into hbase tables.
  - Hbase client cannot read data written by hive.
- Hive 0.9:
  - Hive and hbase clients can read data written by each other.

# New built-in Functions

---

- **printf():**  
select printf("Hello World %d %s", 100, "days") from src limit 1;  
Hello World 100 days
- **Sort\_array():**  
select sort\_array( array (2, 9, 6, 3, 5, 1)) from src limit 1;  
1 2 3 5 6 9
- **Concat\_ws():**  
select concat\_ws('.', array('www','apache','org')) from src limit 1;  
www.apache.org

# New Optimizations

- Coalesce multiple union-all in one MR job.
- ```
select t.a, t.b from (  
    select a,b from t1 where b < 10    union all  
    select a,b from t2 where b < 10    union all  
    select a,b from t3 where b < 10    union all  
    select a, count(1) as b from t4 group by a) t;
```

Hive 0.8 : Number of MR jobs = 5

Hive 0.9 : Number of MR jobs = 2

# New Optimizations

---

- Coalesce multiple group-bys on same data with same keys in one MR job.

- from t2

insert overwrite table t5 select a, sum (substr(a,5)) group by a

insert overwrite table t4 select a, sum (substr(a,5)) group by a;

Hive 0.8 : Number of MR jobs = 2

Hive 0.9 : Number of MR jobs = 1



# New Optimizations

---

- Filter push-down from hive into hbase for key column

```
Select * from hbase_table where key < 200 and key > 100;
```

# Join us!

<http://hive.apache.org>  
[user@hive.apache.org](mailto:user@hive.apache.org)

