

Hadoop Inspector

Ken Farmer
William Farmer

August 14, 2015

Abstract

1 Background

Data quality problems have plagued analytical systems for twenty years: continually appearing in the top four reasons for project failure. In this space data quality problems loom large - a small defect that could be safely ignored or forgotten in the transactional world hamper queries and cause users to question our credibility for months.

The advent and innovation in Big Data and Data Science has not diminished this challenge. On Hadoop specifically:

- Our data generally lacks any enforced constraints to ensure data validity
- We are adding data faster than ever, with less time to research upstream and ETL pipeline issues
- We are building vast systems, sometimes with hundreds of thousands of tasks being defined
- We often have democratized access to our clusters - with dozens of different people adding data.

Additionally, in these large clusters most teams struggle to comply with policies and other requirements, whether regulatory, corporate or defined by their own teams. These might define general data retention requirements, or specific requirements for individual tables. They might define table naming conventions, security requirements, or stats aging & collection requirements.

2 Objective

Hadoop-Inspector is being built because we believe that the complexity of a large, constantly loaded cluster defies an unmanaged approach or QA testing in the development process. It requires something more like an automobile assembly line: continuous quality control (QC) that can take into account undocumented changes from upstream sources, accidental changes to production, changes that bypass QA, etc. And it shouldn't be limited to traditional quality tests, but should be able to test for compliance with policies as well.

3 Technical Details

Users determine the schedule on which they wish to test their cluster, and every allotted time increment, every check is run. These results are kept in a local database which is then referenced when users examine the frontend.

Again, given the intricacies of a hadoop cluster, most checks should be written by the cluster maintainers and operators. Every environment and system is different, and every cluster will have different requirements. We've strived to provide some basic checks, as well as suggestions for checks that may apply to cluster-specific requirements, however for the most part, these checks are the user's responsibility.

3.1 Checks

Every check is simply a script that lives in a type-specific folder. Rules and Warnings exist separately. The directory structure is up to the user, however we recommend using our standard layout:

```
checks
  |--Rules
  |--Warnings
```

The final directory structure is configurable through either JSON config, or command-line arguments.

Each script returns a JSON-encoded object to `stdout` that meets the following specification. Every check that is added to the system must abide by this specification in order to be added to the record. If a check does not meet this specification its output will be ignored.

```
{
  "name": "Name of test",
  "violations": 9000,
  "Output": "Test Specific output. Either JSON or String"
}
```

3.1.1 Rules

Rules are strict rules about the cluster. They should never be ignored, or disobeyed, and any violation of a set rule results in a violation. Ideally, a “healthy” cluster should have no rule violations.

An example rule would be that a specific column only contains integers.

3.1.2 Warnings

Checks are suggestions about specific tables, or databases. These are a lot more fluid, and a warning from a check does not necessarily indicate an issue with the database, but rather that something new may have happened.

For example, a check that examines the average number of a column may throw a warning if an entry is too far from the mean. This isn’t an indication that the data is bad, just that it *may* be bad.

3.2 Viewing Results

All results can be viewed through the front end. This provides a way to analyze your cluster’s health, as well as its current status.

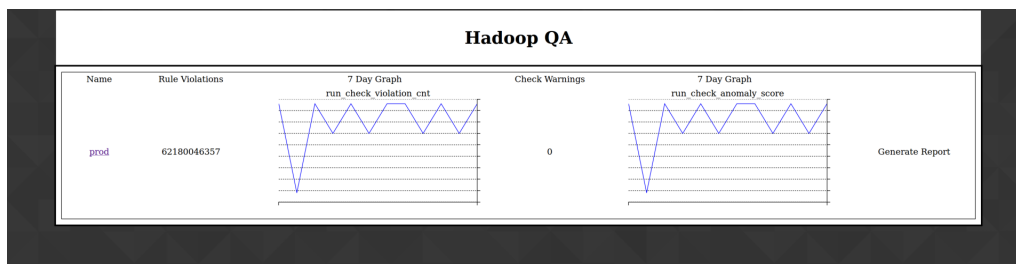


Figure 1: Index

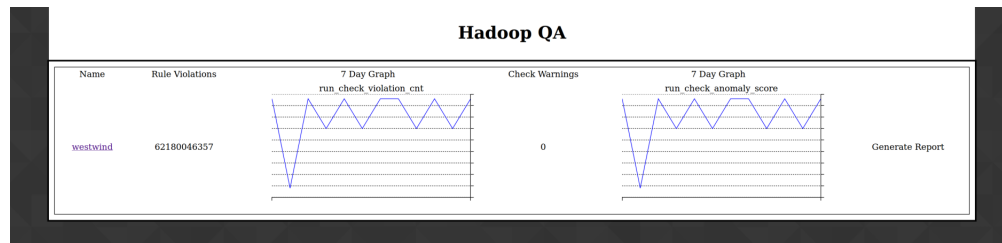


Figure 2: Environment View

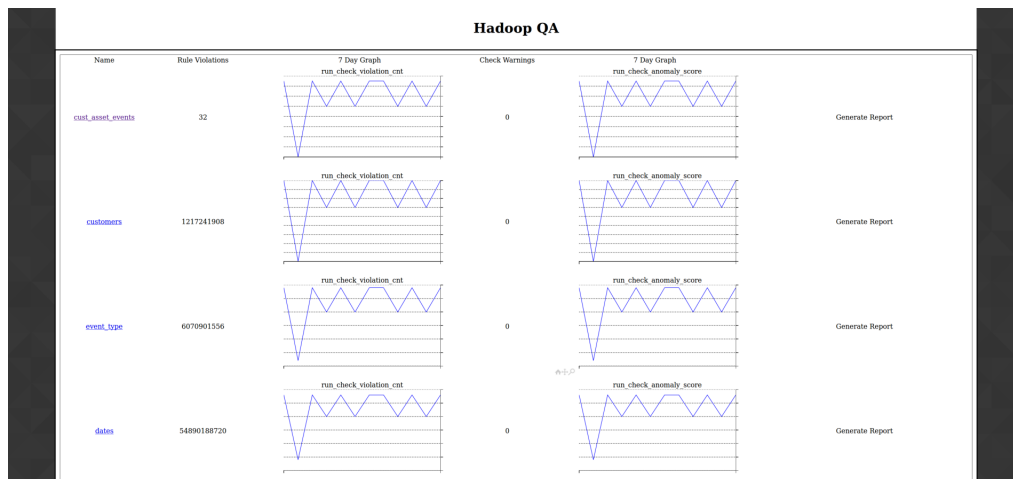


Figure 3: Database View

3.3 Configuration

Most aspects of this framework can be configured using the JSON files located in `/config`. As of August 14, 2015, the files are located in the appendix.

4 Current Status

Our initial focus has been on building a demo to help us validate ideas, and build some of our UIs. This includes:

- **hadoopinspector-demogen** - Can generate 50,000+ check results against a hypothetical user hadoop environment
- **server** - Runs a website that allows the user to analyze these demo results
- **report** - Produces a pdf check result summary report

1 Config Files

1.1 Server Config

```
{
  "logfile": "./hadoopqa.log",
  "testdir": "./tests",
  "db": "hadoopqa.db",
  "port": 9997
}
```

1.2 Demo Data Config

1.2.1 /data/instanceprodfailover_dbAssetUserEvents_config.json

```
{
  "customers": {
    "partition_row_cnt_avg": 10000,
    "partition_key": null,
    "mode": "full",
    "cols": [
      "cust_id",
      "cust_name",
      "cust_status",
      "cust_type_id"
    ],
    "checks": {
      "replication_consistency_peer_row_cnt": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "medium",
        "policy_type": "consistency",
        "mode": "full"
      },
      "replication_consistency_peer_field_spot_check": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "medium",
        "policy_type": "consistency",
        "mode": "full"
      },
      "stats_not_stale": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.1,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
      }
    }
  },
}
```

```
        "stats_exist": {
            "severity": "low",
            "violation_unit": "tables",
            "failure_rate": 0.001,
            "mode": "full",
            "policy_type": "data-management",
            "check_type": "rule"
        }
    },
    "event_type": {
        "partition_row_cnt_avg": 5,
        "partition_key": null,
        "mode": "full",
        "cols": [
            "event_type_id",
            "event_type_name"
        ],
        "checks": {
            "replication_consistency_peer_row_cnt": {
                "violation_unit": "rows",
                "check_type": "rule",
                "severity": "medium",
                "policy_type": "consistency",
                "mode": "full"
            },
            "replication_consistency_peer_field_spot_check": {
                "violation_unit": "rows",
                "check_type": "rule",
                "severity": "medium",
                "policy_type": "consistency",
                "mode": "full"
            },
            "stats_not_stale": {
                "severity": "low",
                "violation_unit": "tables",
                "failure_rate": 0.1,
                "mode": "full",
                "policy_type": "data-management",
                "check_type": "rule"
            },
            "stats_exist": {
                "severity": "low",
                "violation_unit": "tables",
                "failure_rate": 0.001,
                "mode": "full",
                "policy_type": "data-management",
                "check_type": "rule"
            }
        }
    }
}
```

```
    }  
  }  
},  
"dates": {  
  "partition_row_cnt_avg": 3650,  
  "partition_key": null,  
  "mode": "full",  
  "cols": [  
    "date_id",  
    "date_name",  
    "year",  
    "quarter_of_year",  
    "month_of_year",  
    "month_name",  
    "week_of_year_iso",  
    "day_of_month",  
    "day_of_year",  
    "day_of_week",  
    "day_name",  
    "day_name_abbrev"  
  ],  
  "checks": {  
    "replication_consistency_peer_row_cnt": {  
      "violation_unit": "rows",  
      "check_type": "rule",  
      "severity": "medium",  
      "policy_type": "consistency",  
      "mode": "full"  
    },  
    "replication_consistency_peer_field_spot_check": {  
      "violation_unit": "rows",  
      "check_type": "rule",  
      "severity": "medium",  
      "policy_type": "consistency",  
      "mode": "full"  
    },  
    "stats_not_stale": {  
      "severity": "low",  
      "violation_unit": "tables",  
      "failure_rate": 0.1,  
      "mode": "full",  
      "policy_type": "data-management",  
      "check_type": "rule"  
    },  
    "stats_exist": {  
      "severity": "low",  
      "violation_unit": "tables",  
      "failure_rate": 0.001,  
    }  
  }  
}
```



```
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    }
}
},
"cust_asset_event_months": {
    "partition_row_cnt_avg": 40000000,
    "partition_key": "date_id",
    "mode": "incremental",
    "cols": [
        "cust_id",
        "asset_id",
        "month_id",
        "event_type_id",
        "event_count"
    ],
    "checks": {
        "replication_consistency_peer_row_cnt": {
            "violation_unit": "rows",
            "check_type": "rule",
            "severity": "medium",
            "policy_type": "consistency",
            "mode": "full"
        },
        "replication_consistency_peer_field_spot_check": {
            "violation_unit": "rows",
            "check_type": "rule",
            "severity": "medium",
            "policy_type": "consistency",
            "mode": "full"
        },
        "stats_not_stale": {
            "severity": "low",
            "violation_unit": "tables",
            "failure_rate": 0.1,
            "mode": "full",
            "policy_type": "data-management",
            "check_type": "rule"
        },
        "stats_exist": {
            "severity": "low",
            "violation_unit": "tables",
            "failure_rate": 0.001,
            "mode": "full",
            "policy_type": "data-management",
            "check_type": "rule"
        }
    }
}
```

```
    }
  },
  "asset_type": {
    "partition_row_cnt_avg": 5,
    "partition_key": null,
    "mode": "full",
    "cols": [
      "asset_type_id",
      "asset_type_name"
    ],
    "checks": {
      "replication_consistency_peer_row_cnt": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "medium",
        "policy_type": "consistency",
        "mode": "full"
      },
      "replication_consistency_peer_field_spot_check": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "medium",
        "policy_type": "consistency",
        "mode": "full"
      },
      "stats_not_stale": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.1,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
      },
      "stats_exist": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
      }
    }
  },
  "cust_asset_events": {
    "partition_row_cnt_avg": 400000000,
    "partition_key": "date_id",
    "mode": "incremental",
    "cols": [
```

```
        "cust_id",
        "asset_id",
        "event_id",
        "date_id",
        "event_type_id"
    ],
    "checks": {
        "replication_consistency_peer_row_cnt": {
            "violation_unit": "rows",
            "check_type": "rule",
            "severity": "medium",
            "policy_type": "consistency",
            "mode": "full"
        },
        "replication_consistency_peer_field_spot_check": {
            "violation_unit": "rows",
            "check_type": "rule",
            "severity": "medium",
            "policy_type": "consistency",
            "mode": "full"
        },
        "stats_not_stale": {
            "severity": "low",
            "violation_unit": "tables",
            "failure_rate": 0.1,
            "mode": "full",
            "policy_type": "data-management",
            "check_type": "rule"
        },
        "stats_exist": {
            "severity": "low",
            "violation_unit": "tables",
            "failure_rate": 0.001,
            "mode": "full",
            "policy_type": "data-management",
            "check_type": "rule"
        }
    }
},
"cust_type": {
    "partition_row_cnt_avg": 5,
    "partition_key": null,
    "mode": "full",
    "cols": [
        "cust_type_id",
        "cust_type_name"
    ],
    "checks": {
```

```
    "replication_consistency_peer_row_cnt": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "medium",
      "policy_type": "consistency",
      "mode": "full"
    },
    "replication_consistency_peer_field_spot_check": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "medium",
      "policy_type": "consistency",
      "mode": "full"
    },
    "stats_not_stale": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.001,
      "mode": "full",
      "policy_type": "data-management",
      "check_type": "rule"
    },
    "stats_exist": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.001,
      "mode": "full",
      "policy_type": "data-management",
      "check_type": "rule"
    }
  },
  "assets": {
    "partition_row_cnt_avg": 10000000,
    "partition_key": null,
    "mode": "full",
    "cols": [
      "cust_id",
      "asset_id",
      "asset_name",
      "asset_status",
      "asset_type_id"
    ],
    "checks": {
      "replication_consistency_peer_row_cnt": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "medium",
```

```

        "policy_type": "consistency",
        "mode": "full"
    },
    "replication_consistency_peer_field_spot_check": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "medium",
        "policy_type": "consistency",
        "mode": "full"
    },
    "stats_not_stale": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.1,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "stats_exist": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    }
}
}
}
}

```

1.2.2 /data/instancetypeaging_dbAssetUserEvents_config.json

```

{
    "customers": {
        "partition_row_cnt_avg": 10000,
        "partition_key": null,
        "cols": [
            "cust_id",
            "cust_name",
            "cust_status",
            "cust_type_id"
        ],
        "checks": {
            "cust_typeid_fk": {
                "violation_unit": "rows",
                "check_type": "rule",
                "severity": "high",
                "policy_type": "quality",

```

```
        "mode": "full"
    },
    "cust_name_uk": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "full"
    },
    "cust_id_uk": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "full"
    },
    "cust_status_ck": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "full"
    },
    "stats_not_stale": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.1,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "stats_exist": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    }
},
"mode": "full"
},
"event_type": {
    "partition_row_cnt_avg": 5,
    "partition_key": null,
    "cols": [
        "event_type_id",
        "event_type_name"
    ]
}
```

```
],
  "checks": {
    "event_not_empty": {
      "severity": "high",
      "violation_unit": "tables",
      "failure_rate": 0.01,
      "mode": "full",
      "policy_type": "quality",
      "check_type": "rule"
    },
    "event_name_uk": {
      "severity": "high",
      "violation_unit": "rows",
      "failure_rate": 0.01,
      "mode": "full",
      "policy_type": "quality",
      "check_type": "rule"
    },
    "stats_not_stale": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.001,
      "mode": "full",
      "policy_type": "data-management",
      "check_type": "rule"
    },
    "stats_exist": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.001,
      "mode": "full",
      "policy_type": "data-management",
      "check_type": "rule"
    },
    "event_type_uk": {
      "severity": "high",
      "violation_unit": "rows",
      "failure_rate": 0.01,
      "mode": "full",
      "policy_type": "quality",
      "check_type": "rule"
    }
  },
  "mode": "full"
},
"dates": {
  "partition_row_cnt_avg": 3650,
  "partition_key": null,
```

```
"cols": [  
  "date_id",  
  "date_name",  
  "year",  
  "quarter_of_year",  
  "month_of_year",  
  "month_name",  
  "week_of_year_iso",  
  "day_of_month",  
  "day_of_year",  
  "day_of_week",  
  "day_name",  
  "day_name_abbrev"  
],  
"checks": {  
  "date_name_uk": {  
    "severity": "high",  
    "violation_unit": "rows",  
    "failure_rate": 0.01,  
    "mode": "full",  
    "policy_type": "quality",  
    "check_type": "rule"  
  },  
  "dates_not_empty": {  
    "severity": "high",  
    "violation_unit": "tables",  
    "failure_rate": 0.01,  
    "mode": "full",  
    "policy_type": "quality",  
    "check_type": "rule"  
  },  
  "stats_not_stale": {  
    "severity": "low",  
    "violation_unit": "tables",  
    "failure_rate": 0.001,  
    "mode": "full",  
    "policy_type": "data-management",  
    "check_type": "rule"  
  },  
  "monthofyear_range_ck": {  
    "severity": "high",  
    "violation_unit": "rows",  
    "failure_rate": 0.01,  
    "mode": "full",  
    "policy_type": "quality",  
    "check_type": "rule"  
  },  
  "weekofyear_range_ck": {
```



```
        "severity": "high",
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    },
    "date_id_uk": {
        "severity": "high",
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    },
    "stats_exist": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "dayofmonth_range_ck": {
        "severity": "high",
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    }
},
"mode": "full"
},
"cust_asset_event_months": {
    "partition_row_cnt_avg": 40000000,
    "partition_key": "date_id",
    "cols": [
        "cust_id",
        "asset_id",
        "month_id",
        "event_type_id",
        "event_count"
    ],
    "checks": {
        "event_count_consistency_ck": {
            "violation_unit": "rows",
            "check_type": "rule",
```

```
    "severity": "high",
    "policy_type": "consistency",
    "mode": "incremental"
  },
  "cols_not_null": {
    "severity": "medium",
    "violation_unit": "rows",
    "failure_rate": 0.1,
    "mode": "incremental",
    "policy_type": "quality",
    "check_type": "rule"
  },
  "event_type_id_fk": {
    "violation_unit": "rows",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "quality",
    "mode": "incremental"
  },
  "stats_not_stale": {
    "severity": "low",
    "violation_unit": "tables",
    "failure_rate": 0.1,
    "mode": "full",
    "policy_type": "data-management",
    "check_type": "rule"
  },
  "table_not_empty": {
    "violation_unit": "tables",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "quality",
    "mode": "incremental"
  },
  "asset_id_fk": {
    "violation_unit": "rows",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "quality",
    "mode": "incremental"
  },
  "cust_id_fk": {
    "violation_unit": "rows",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "quality",
    "mode": "incremental"
  },
}
```

```
    "stats_exist": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.001,
      "mode": "full",
      "policy_type": "data-management",
      "check_type": "rule"
    },
    "month_id_fk": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "incremental"
    }
  },
  "mode": "incremental"
},
"asset_type": {
  "partition_row_cnt_avg": 5,
  "partition_key": null,
  "cols": [
    "asset_type_id",
    "asset_type_name"
  ],
  "checks": {
    "asset_not_empty": {
      "severity": "high",
      "violation_unit": "tables",
      "failure_rate": 0.01,
      "mode": "full",
      "policy_type": "quality",
      "check_type": "rule"
    },
    "asset_name_uk": {
      "severity": "high",
      "violation_unit": "rows",
      "failure_rate": 0.01,
      "mode": "full",
      "policy_type": "quality",
      "check_type": "rule"
    },
    "stats_not_stale": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.001,
      "mode": "full",
      "policy_type": "data-management",
```

```
        "check_type": "rule"
    },
    "stats_exist": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "asset_type_uk": {
        "severity": "high",
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    }
},
"mode": "full"
},
"cust_asset_events": {
    "partition_row_cnt_avg": 400000000,
    "partition_key": "date_id",
    "cols": [
        "cust_id",
        "asset_id",
        "event_id",
        "date_id",
        "event_type_id"
    ],
    "checks": {
        "date_id_fk": {
            "violation_unit": "rows",
            "check_type": "rule",
            "severity": "high",
            "policy_type": "quality",
            "mode": "incremental"
        },
        "event_type_id_fk": {
            "violation_unit": "rows",
            "check_type": "rule",
            "severity": "high",
            "policy_type": "quality",
            "mode": "incremental"
        },
        "stats_not_stale": {
            "severity": "low",
```

```
        "violation_unit": "tables",
        "failure_rate": 0.1,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "table_not_empty": {
        "violation_unit": "tables",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "incremental"
    },
    "asset_id_fk": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "incremental"
    },
    "cust_id_fk": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "incremental"
    },
    "stats_exist": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "event_id_not_null": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "incremental"
    }
},
"mode": "incremental"
},
"cust_type": {
    "partition_row_cnt_avg": 5,
    "partition_key": null,
```

```
"cols": [  
  "cust_type_id",  
  "cust_type_name"  
],  
"checks": {  
  "cust_name_uk": {  
    "severity": "high",  
    "violation_unit": "rows",  
    "failure_rate": 0.01,  
    "mode": "full",  
    "policy_type": "quality",  
    "check_type": "rule"  
  },  
  "cust_type_uk": {  
    "severity": "high",  
    "violation_unit": "rows",  
    "failure_rate": 0.01,  
    "mode": "full",  
    "policy_type": "quality",  
    "check_type": "rule"  
  },  
  "stats_not_stale": {  
    "severity": "low",  
    "violation_unit": "tables",  
    "failure_rate": 0.001,  
    "mode": "full",  
    "policy_type": "data-management",  
    "check_type": "rule"  
  },  
  "stats_exist": {  
    "severity": "low",  
    "violation_unit": "tables",  
    "failure_rate": 0.001,  
    "mode": "full",  
    "policy_type": "data-management",  
    "check_type": "rule"  
  },  
  "cust_not_empty": {  
    "severity": "high",  
    "violation_unit": "tables",  
    "failure_rate": 0.01,  
    "mode": "full",  
    "policy_type": "quality",  
    "check_type": "rule"  
  }  
},  
"mode": "full"  
},
```

```
"assets": {
  "partition_row_cnt_avg": 10000000,
  "partition_key": null,
  "cols": [
    "cust_id",
    "asset_id",
    "asset_name",
    "asset_status",
    "asset_type_id"
  ],
  "checks": {
    "asset_id_uk": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "full"
    },
    "asset_status_ck": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "full"
    },
    "asset_name_uk": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "full"
    },
    "asset_typeid_fk": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "full"
    },
    "stats_not_stale": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.1,
      "mode": "full",
      "policy_type": "data-management",
      "check_type": "rule"
    },
    "stats_exist": {
```

```

        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    }
},
    "mode": "full"
}
}

```

1.2.3 /data/instanceprod_dbAssetUserEvents_config.json

```

{
    "customers": {
        "partition_row_cnt_avg": 10000,
        "partition_key": null,
        "cols": [
            "cust_id",
            "cust_name",
            "cust_status",
            "cust_type_id"
        ],
        "checks": {
            "cust_typeid_fk": {
                "violation_unit": "rows",
                "check_type": "rule",
                "severity": "high",
                "policy_type": "quality",
                "mode": "full"
            },
            "cust_name_uk": {
                "violation_unit": "rows",
                "check_type": "rule",
                "severity": "high",
                "policy_type": "quality",
                "mode": "full"
            },
            "cust_id_uk": {
                "violation_unit": "rows",
                "check_type": "rule",
                "severity": "high",
                "policy_type": "quality",
                "mode": "full"
            },
            "cust_status_ck": {
                "violation_unit": "rows",

```



```
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "full"
    },
    "stats_not_stale": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.1,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "stats_exist": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    }
},
"mode": "full"
},
"event_type": {
    "partition_row_cnt_avg": 5,
    "partition_key": null,
    "cols": [
        "event_type_id",
        "event_type_name"
    ],
    "checks": {
        "event_not_empty": {
            "severity": "high",
            "violation_unit": "tables",
            "failure_rate": 0.01,
            "mode": "full",
            "policy_type": "quality",
            "check_type": "rule"
        },
        "event_name_uk": {
            "severity": "high",
            "violation_unit": "rows",
            "failure_rate": 0.01,
            "mode": "full",
            "policy_type": "quality",
            "check_type": "rule"
        }
    }
},
```

```
    "stats_not_stale": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.001,
      "mode": "full",
      "policy_type": "data-management",
      "check_type": "rule"
    },
    "stats_exist": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.001,
      "mode": "full",
      "policy_type": "data-management",
      "check_type": "rule"
    },
    "event_type_uk": {
      "severity": "high",
      "violation_unit": "rows",
      "failure_rate": 0.01,
      "mode": "full",
      "policy_type": "quality",
      "check_type": "rule"
    }
  },
  "mode": "full"
},
"dates": {
  "partition_row_cnt_avg": 3650,
  "partition_key": null,
  "cols": [
    "date_id",
    "date_name",
    "year",
    "quarter_of_year",
    "month_of_year",
    "month_name",
    "week_of_year_iso",
    "day_of_month",
    "day_of_year",
    "day_of_week",
    "day_name",
    "day_name_abbrev"
  ],
  "checks": {
    "date_name_uk": {
      "severity": "high",
      "violation_unit": "rows",
```

```
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    },
    "dates_not_empty": {
        "severity": "high",
        "violation_unit": "tables",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    },
    "stats_not_stale": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "monthofyear_range_ck": {
        "severity": "high",
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    },
    "weekofyear_range_ck": {
        "severity": "high",
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    },
    "date_id_uk": {
        "severity": "high",
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    },
    "stats_exist": {
        "severity": "low",
        "violation_unit": "tables",
```

```
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "dayofmonth_range_ck": {
        "severity": "high",
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    }
},
"mode": "full"
},
"cust_asset_event_months": {
    "partition_row_cnt_avg": 40000000,
    "partition_key": "date_id",
    "cols": [
        "cust_id",
        "asset_id",
        "month_id",
        "event_type_id",
        "event_count"
    ],
    "checks": {
        "event_count_consistency_ck": {
            "violation_unit": "rows",
            "check_type": "rule",
            "severity": "high",
            "policy_type": "consistency",
            "mode": "incremental"
        },
        "cols_not_null": {
            "severity": "medium",
            "violation_unit": "rows",
            "failure_rate": 0.1,
            "mode": "incremental",
            "policy_type": "quality",
            "check_type": "rule"
        },
        "event_type_id_fk": {
            "violation_unit": "rows",
            "check_type": "rule",
            "severity": "high",
            "policy_type": "quality",
            "mode": "incremental"
        }
    }
}
```

```
    },
    "stats_not_stale": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.1,
      "mode": "full",
      "policy_type": "data-management",
      "check_type": "rule"
    },
    "table_not_empty": {
      "violation_unit": "tables",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "incremental"
    },
    "asset_id_fk": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "incremental"
    },
    "cust_id_fk": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "incremental"
    },
    "stats_exist": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.001,
      "mode": "full",
      "policy_type": "data-management",
      "check_type": "rule"
    },
    "month_id_fk": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "incremental"
    }
  },
  "mode": "incremental"
},
```

```
"asset_type": {
  "partition_row_cnt_avg": 5,
  "partition_key": null,
  "cols": [
    "asset_type_id",
    "asset_type_name"
  ],
  "checks": {
    "asset_not_empty": {
      "severity": "high",
      "violation_unit": "tables",
      "failure_rate": 0.01,
      "mode": "full",
      "policy_type": "quality",
      "check_type": "rule"
    },
    "asset_name_uk": {
      "severity": "high",
      "violation_unit": "rows",
      "failure_rate": 0.01,
      "mode": "full",
      "policy_type": "quality",
      "check_type": "rule"
    },
    "stats_not_stale": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.001,
      "mode": "full",
      "policy_type": "data-management",
      "check_type": "rule"
    },
    "stats_exist": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.001,
      "mode": "full",
      "policy_type": "data-management",
      "check_type": "rule"
    },
    "asset_type_uk": {
      "severity": "high",
      "violation_unit": "rows",
      "failure_rate": 0.01,
      "mode": "full",
      "policy_type": "quality",
      "check_type": "rule"
    }
  }
}
```

```
    },
    "mode": "full"
  },
  "cust_asset_events": {
    "partition_row_cnt_avg": 400000000,
    "partition_key": "date_id",
    "cols": [
      "cust_id",
      "asset_id",
      "event_id",
      "date_id",
      "event_type_id"
    ],
    "checks": {
      "date_id_fk": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "incremental"
      },
      "event_type_id_fk": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "incremental"
      },
      "stats_not_stale": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.1,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
      },
      "table_not_empty": {
        "violation_unit": "tables",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "incremental"
      },
      "asset_id_fk": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
```

```
        "mode": "incremental"
    },
    "cust_id_fk": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "incremental"
    },
    "stats_exist": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "event_id_not_null": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "incremental"
    }
},
"mode": "incremental"
},
"cust_type": {
    "partition_row_cnt_avg": 5,
    "partition_key": null,
    "cols": [
        "cust_type_id",
        "cust_type_name"
    ],
    "checks": {
        "cust_name_uk": {
            "severity": "high",
            "violation_unit": "rows",
            "failure_rate": 0.01,
            "mode": "full",
            "policy_type": "quality",
            "check_type": "rule"
        },
        "cust_type_uk": {
            "severity": "high",
            "violation_unit": "rows",
            "failure_rate": 0.01,
            "mode": "full",
```



```
        "policy_type": "quality",
        "check_type": "rule"
    },
    "stats_not_stale": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "stats_exist": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "cust_not_empty": {
        "severity": "high",
        "violation_unit": "tables",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    }
},
"mode": "full"
},
"assets": {
    "partition_row_cnt_avg": 10000000,
    "partition_key": null,
    "cols": [
        "cust_id",
        "asset_id",
        "asset_name",
        "asset_status",
        "asset_type_id"
    ],
    "checks": {
        "asset_id_uk": {
            "violation_unit": "rows",
            "check_type": "rule",
            "severity": "high",
            "policy_type": "quality",
            "mode": "full"
        }
    }
},
```

```

    "asset_status_ck": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "full"
    },
    "asset_name_uk": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "full"
    },
    "asset_typeid_fk": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "full"
    },
    "stats_not_stale": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.1,
      "mode": "full",
      "policy_type": "data-management",
      "check_type": "rule"
    },
    "stats_exist": {
      "severity": "low",
      "violation_unit": "tables",
      "failure_rate": 0.001,
      "mode": "full",
      "policy_type": "data-management",
      "check_type": "rule"
    }
  },
  "mode": "full"
}

```

1.2.4 /data/instancedev_dbAssetUserEvents_config.json

```

{
  "customers": {
    "partition_row_cnt_avg": 10000,
    "partition_key": null,

```

```
"cols": [
  "cust_id",
  "cust_name",
  "cust_status",
  "cust_type_id"
],
"checks": {
  "cust_typeid_fk": {
    "violation_unit": "rows",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "quality",
    "mode": "full"
  },
  "cust_name_uk": {
    "violation_unit": "rows",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "quality",
    "mode": "full"
  },
  "cust_id_uk": {
    "violation_unit": "rows",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "quality",
    "mode": "full"
  },
  "cust_status_ck": {
    "violation_unit": "rows",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "quality",
    "mode": "full"
  },
  "stats_not_stale": {
    "severity": "low",
    "violation_unit": "tables",
    "failure_rate": 0.1,
    "mode": "full",
    "policy_type": "data-management",
    "check_type": "rule"
  },
  "stats_exist": {
    "severity": "low",
    "violation_unit": "tables",
    "failure_rate": 0.001,
    "mode": "full",
```

```
        "policy_type": "data-management",
        "check_type": "rule"
    },
    },
    "mode": "full"
},
"event_type": {
    "partition_row_cnt_avg": 5,
    "partition_key": null,
    "cols": [
        "event_type_id",
        "event_type_name"
    ],
    "checks": {
        "event_not_empty": {
            "severity": "high",
            "violation_unit": "tables",
            "failure_rate": 0.01,
            "mode": "full",
            "policy_type": "quality",
            "check_type": "rule"
        },
        "event_name_uk": {
            "severity": "high",
            "violation_unit": "rows",
            "failure_rate": 0.01,
            "mode": "full",
            "policy_type": "quality",
            "check_type": "rule"
        },
        "stats_not_stale": {
            "severity": "low",
            "violation_unit": "tables",
            "failure_rate": 0.001,
            "mode": "full",
            "policy_type": "data-management",
            "check_type": "rule"
        },
        "stats_exist": {
            "severity": "low",
            "violation_unit": "tables",
            "failure_rate": 0.001,
            "mode": "full",
            "policy_type": "data-management",
            "check_type": "rule"
        },
        "event_type_uk": {
            "severity": "high",
```

```
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    },
    "mode": "full"
},
"dates": {
    "partition_row_cnt_avg": 3650,
    "partition_key": null,
    "cols": [
        "date_id",
        "date_name",
        "year",
        "quarter_of_year",
        "month_of_year",
        "month_name",
        "week_of_year_iso",
        "day_of_month",
        "day_of_year",
        "day_of_week",
        "day_name",
        "day_name_abbrev"
    ],
    "checks": {
        "date_name_uk": {
            "severity": "high",
            "violation_unit": "rows",
            "failure_rate": 0.01,
            "mode": "full",
            "policy_type": "quality",
            "check_type": "rule"
        },
        "dates_not_empty": {
            "severity": "high",
            "violation_unit": "tables",
            "failure_rate": 0.01,
            "mode": "full",
            "policy_type": "quality",
            "check_type": "rule"
        },
        "stats_not_stale": {
            "severity": "low",
            "violation_unit": "tables",
            "failure_rate": 0.001,
            "mode": "full",
```

```
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "monthofyear_range_ck": {
        "severity": "high",
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    },
    "weekofyear_range_ck": {
        "severity": "high",
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    },
    "date_id_uk": {
        "severity": "high",
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    },
    "stats_exist": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "dayofmonth_range_ck": {
        "severity": "high",
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    }
},
"mode": "full"
},
"cust_asset_event_months": {
    "partition_row_cnt_avg": 40000000,
```

```
"partition_key": "date_id",
"cols": [
  "cust_id",
  "asset_id",
  "month_id",
  "event_type_id",
  "event_count"
],
"checks": {
  "event_count_consistency_ck": {
    "violation_unit": "rows",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "consistency",
    "mode": "incremental"
  },
  "cols_not_null": {
    "severity": "medium",
    "violation_unit": "rows",
    "failure_rate": 0.1,
    "mode": "incremental",
    "policy_type": "quality",
    "check_type": "rule"
  },
  "event_type_id_fk": {
    "violation_unit": "rows",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "quality",
    "mode": "incremental"
  },
  "stats_not_stale": {
    "severity": "low",
    "violation_unit": "tables",
    "failure_rate": 0.1,
    "mode": "full",
    "policy_type": "data-management",
    "check_type": "rule"
  },
  "table_not_empty": {
    "violation_unit": "tables",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "quality",
    "mode": "incremental"
  },
  "asset_id_fk": {
    "violation_unit": "rows",
```

```
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "incremental"
    },
    "cust_id_fk": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "incremental"
    },
    "stats_exist": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "month_id_fk": {
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "incremental"
    }
},
"mode": "incremental"
},
"asset_type": {
    "partition_row_cnt_avg": 5,
    "partition_key": null,
    "cols": [
        "asset_type_id",
        "asset_type_name"
    ],
    "checks": {
        "asset_not_empty": {
            "severity": "high",
            "violation_unit": "tables",
            "failure_rate": 0.01,
            "mode": "full",
            "policy_type": "quality",
            "check_type": "rule"
        },
        "asset_name_uk": {
            "severity": "high",
```



```
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    },
    "stats_not_stale": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "stats_exist": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "asset_type_uk": {
        "severity": "high",
        "violation_unit": "rows",
        "failure_rate": 0.01,
        "mode": "full",
        "policy_type": "quality",
        "check_type": "rule"
    }
},
"mode": "full"
},
"cust_asset_events": {
    "partition_row_cnt_avg": 400000000,
    "partition_key": "date_id",
    "cols": [
        "cust_id",
        "asset_id",
        "event_id",
        "date_id",
        "event_type_id"
    ],
    "checks": {
        "date_id_fk": {
            "violation_unit": "rows",
            "check_type": "rule",
            "severity": "high",
```

```
    "policy_type": "quality",
    "mode": "incremental"
  },
  "event_type_id_fk": {
    "violation_unit": "rows",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "quality",
    "mode": "incremental"
  },
  "stats_not_stale": {
    "severity": "low",
    "violation_unit": "tables",
    "failure_rate": 0.1,
    "mode": "full",
    "policy_type": "data-management",
    "check_type": "rule"
  },
  "table_not_empty": {
    "violation_unit": "tables",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "quality",
    "mode": "incremental"
  },
  "asset_id_fk": {
    "violation_unit": "rows",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "quality",
    "mode": "incremental"
  },
  "cust_id_fk": {
    "violation_unit": "rows",
    "check_type": "rule",
    "severity": "high",
    "policy_type": "quality",
    "mode": "incremental"
  },
  "stats_exist": {
    "severity": "low",
    "violation_unit": "tables",
    "failure_rate": 0.001,
    "mode": "full",
    "policy_type": "data-management",
    "check_type": "rule"
  },
  "event_id_not_null": {
```

```
        "violation_unit": "rows",
        "check_type": "rule",
        "severity": "high",
        "policy_type": "quality",
        "mode": "incremental"
    },
    "mode": "incremental"
},
"cust_type": {
    "partition_row_cnt_avg": 5,
    "partition_key": null,
    "cols": [
        "cust_type_id",
        "cust_type_name"
    ],
    "checks": {
        "cust_name_uk": {
            "severity": "high",
            "violation_unit": "rows",
            "failure_rate": 0.01,
            "mode": "full",
            "policy_type": "quality",
            "check_type": "rule"
        },
        "cust_type_uk": {
            "severity": "high",
            "violation_unit": "rows",
            "failure_rate": 0.01,
            "mode": "full",
            "policy_type": "quality",
            "check_type": "rule"
        },
        "stats_not_stale": {
            "severity": "low",
            "violation_unit": "tables",
            "failure_rate": 0.001,
            "mode": "full",
            "policy_type": "data-management",
            "check_type": "rule"
        },
        "stats_exist": {
            "severity": "low",
            "violation_unit": "tables",
            "failure_rate": 0.001,
            "mode": "full",
            "policy_type": "data-management",
            "check_type": "rule"
        }
    }
}
```

```
    },
    "cust_not_empty": {
      "severity": "high",
      "violation_unit": "tables",
      "failure_rate": 0.01,
      "mode": "full",
      "policy_type": "quality",
      "check_type": "rule"
    }
  },
  "mode": "full"
},
"assets": {
  "partition_row_cnt_avg": 10000000,
  "partition_key": null,
  "cols": [
    "cust_id",
    "asset_id",
    "asset_name",
    "asset_status",
    "asset_type_id"
  ],
  "checks": {
    "asset_id_uk": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "full"
    },
    "asset_status_ck": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "full"
    },
    "asset_name_uk": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
      "policy_type": "quality",
      "mode": "full"
    },
    "asset_typeid_fk": {
      "violation_unit": "rows",
      "check_type": "rule",
      "severity": "high",
```

```
        "policy_type": "quality",
        "mode": "full"
    },
    "stats_not_stale": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.1,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    },
    "stats_exist": {
        "severity": "low",
        "violation_unit": "tables",
        "failure_rate": 0.001,
        "mode": "full",
        "policy_type": "data-management",
        "check_type": "rule"
    }
},
"mode": "full"
}
```