

What's new on Apache Zeppelin

...

Jongyoul Lee, ZEPL

Ahyoung Ryu, ZEPL

Hoon Park, ZEPL

Abstract

- Multi-user
 - Per note/per user
 - Impersonation
 - Personalized mode
- Interpreters
 - Python interpreter
 - Spark interpreter
 - New interpreters
 - Interpreter-setting.json
- Helium
- Spell

Multi-user

- 노트와 사용자 단위로 인터프리터를 할당할 수 있음
 - 리소스를 예측할 수 있게 됨
 - 다른 사용자의 인터프리터에 문제가 생겼을 경우, 그 인터프리터만 재시작하면 됨
- 인터프리터 실행 시, `ssh/sudo` 등을 이용하여, 실행계정을 지정할 수 있음
 - 잘 활용하면, 다중 사용자 환경을 완벽히 구현할 수 있음
 - 보안설정에 더욱더 신경을 써야함
- 개인작업 모드가 추가됨
 - 데이터 탐색 중 다른 사용자가 재실행한 결과가 반영되는 것을 방지할 수 있음
 - 변경된 스크립트는 반영되고, 결과만 반영이 안됨
 - 개인작업설정을 종료하면, 최신 결과가 모두에게 반영 됨

Per note/per user

	Per note/per user
Shared	1 process 1 classloader
Scoped	1 process N classloaders
Isolated	N processes 1 classloader

- Spark interpreter
 - Per note: Isolated
 - Per user: Isolated
- JDBC interpreter
 - Per note: scoped
 - Per user: scoped
- Markdown
 - Shared

Impersonation

```
ZEPPELIN_IMPERSONATE_USER="${OPTARG}"  
if [[ -z "$ZEPPELIN_IMPERSONATE_CMD" ]]; then  
    ZEPPELIN_IMPERSONATE_RUN_CMD=`echo "ssh ${ZEPPELIN_IMPERSONATE_USER}@localhost" `  
else  
    ZEPPELIN_IMPERSONATE_RUN_CMD=$(eval "echo ${ZEPPELIN_IMPERSONATE_CMD} ")  
fi
```

- 별도의 커멘드를 설정하지 않으면, SSH를 이용하여, 사용자 계정으로 접근하여, 인터프리터를 실행함
- Sudo 등의 명령어를 활용하여, 사용자를 변경할 수도 있음
- 잘 활용하면 docker등을 띄우는데도 활용할 수 있을 것으로 보임
- InterpreterRunner를 이용해서 아예 다른 명령어를 사용하여 인터프리터를 실행할 수 있음

Interpreters

- Python interpreter
 - 파이썬 인터프리터가 추가되고, 파이썬 사용자를 위한 다양한 편의 기능을 제공함(conda, matplotlib...)
 - 앞으로도 파이썬 지원에 우선순위를 둘 예정
- Spark interpreter
 - 0.7.0에서는 “공식적으로” 1.4.x부터 2.1까지 스팩까지 모두 지원함
- New interpreters
 - Apache Beam, Apache Pig, scio가 추가되었음. 다만, 버전 이슈로 직접 빌드해야 함
 - 새로운 인터프리터의 지속적인 유입으로 사용자 확대 기여
- Interpreter-setting.json
 - 인터프리터 기본설정을 변경할 수 있음

Interpreter-setting.json

```
{
  {
    "group": "spark",
    "name": "spark",
    "className": "org.apache.zeppelin.spark.SparkInterpreter",
    "defaultInterpreter": true,
    "properties": {
      "spark.executor.memory": {
        "envName": null,
        "propertyName": "spark.executor.memory",
        "defaultValue": "",
        "description": "Executor memory per worker instance. ex) 512m, 32g"
      },
      "args": {"envName": null...},
      "zeppelin.spark.useHiveContext": {"envName": "ZEPPELIN_SPARK_USEHIVECONTEXT"...},
      "spark.app.name": {"envName": "SPARK_APP_NAME"...},
      "zeppelin.spark.printREPLOutput": {"envName": null...},
      "spark.cores.max": {"envName": null...},
      "zeppelin.spark.maxResult": {"envName": "ZEPPELIN_SPARK_MAXRESULT"...},
      "master": {"envName": "MASTER"...}
    },
    "editor": {
      "language": "scala",
      "editOnDbClick": false
    }
  },
  {
    "group": "spark",
    "name": "sql",
    "className": "org.apache.zeppelin.spark.SparkSqlInterpreter",
    "properties": {...},
    "editor": {"language": "sql"...}
  },
  {
    "group": "spark",
    "name": "dep",
    "className": "org.apache.zeppelin.spark.DepInterpreter",
    "properties": {...},
    "editor": {"language": "scala"...}
  },
  {
    "group": "spark",
    "name": "pyspark",
    "className": "org.apache.zeppelin.spark.PySparkInterpreter",
    "properties": {...},
    "editor": {"language": "python"...}
  }
}
```

- 인터프리터 그룹 및 이름 변경 가능
- 프로퍼티 기본값 변경 가능
- interpreter/spark/interpreter-setting.json

Demo