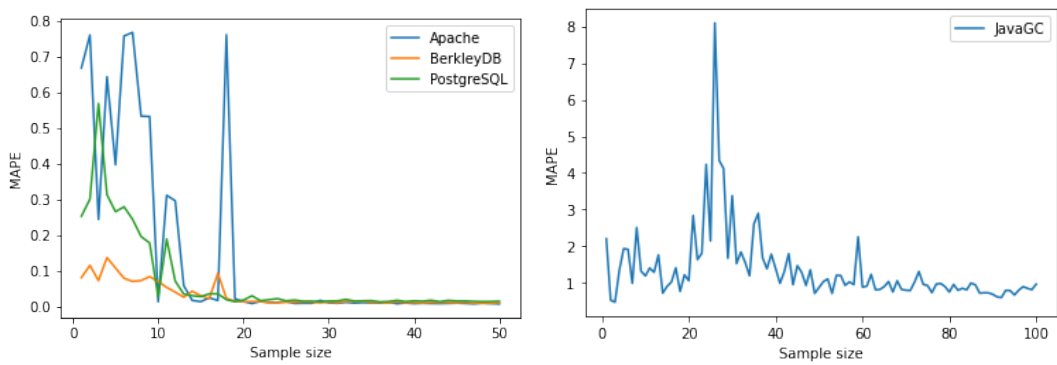


---

## 0.1 Results

### 0.1.1 Random sampling

Although true random sampling is hard to achieve in a sensible manner, it allows for a good baseline comparison of other sampling algorithms which may be domain specific. In our case, since all possible configurations of a given system are already known, we can perform true random sampling of configurations by randomly picking a configuration out of the known configurations. With this we can fit a linear regression and see, how effective a linear regression model is trained with randomly sampled data.



**Figure 1:** The MAPE value on the Apache, BerkleyDB and PostgreSQL datasets in correlation to the sample size

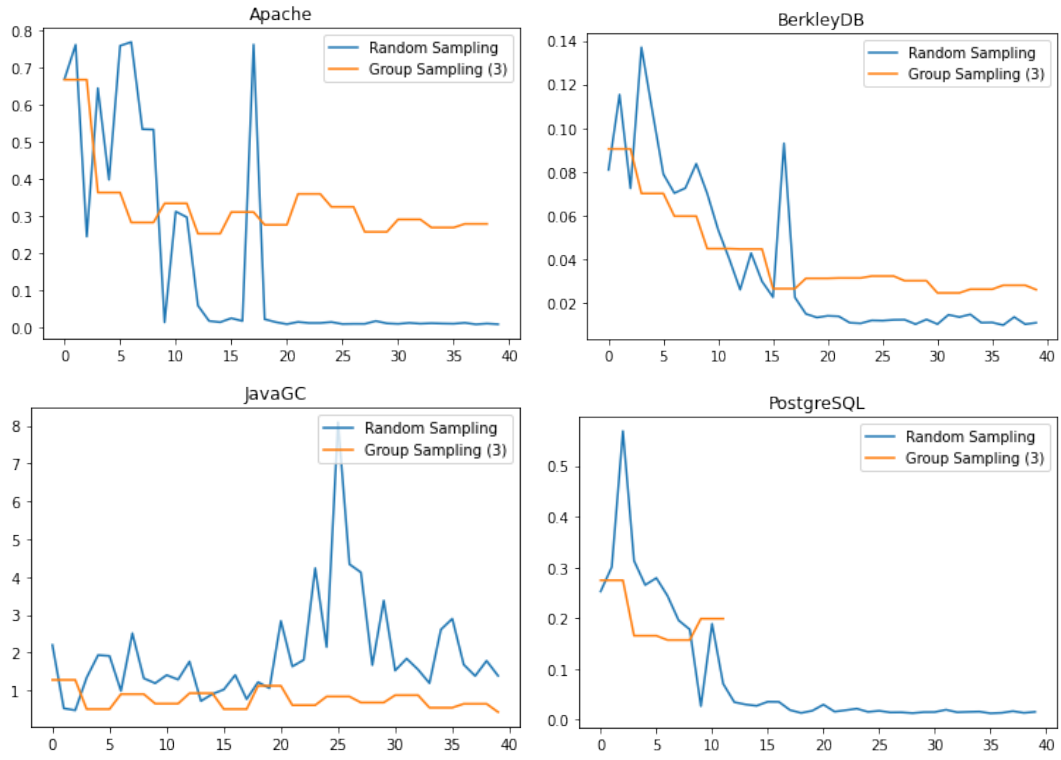
With our real world examples we achieve

### 0.1.2 Group Sampling

Real world datasets

Large datasets

Feature interactions



**Figure 2:** Group Sampling compared to random sampling with linear regression