1. Are we supposed to know the algorithms inside the virus detector and spam filter in an interview?

A general idea is good to have. A lot of malware detection algorithms use machine learning. You can talk about what you would do to find malware (monitor a program for resource consumption etc...), in case the interviewer wants to go in detail.

2. I want to know more about the search engine.

You can have a look at Lucene to understand how text search engines work. Since we also need to search on metadata like date and receiver, we will need indexes on those fields.

3. Why are we using a separate search engine? Why don't we ask the email service to provide search APIs for the emails?

The email service can provide some basic search APIs like search by receiver, date etc... However, complex search queries will require a join across tables, which is expensive.

If we use a NoSQL database and dump all the denormalised email data into an email table, we can avoid joins. Even then, we would like to have an offline process to reverse index the emails. We would also like to separate the complex search operations from the CRUD operations in the email service.

Hence, it would be better to separate the two services. A drawback of this could be inconsistencies in the email service and search engine records (The search engine gets the email event after some time).

4. What about other features like auto-reply, attachment previews, etc...?

In a 45 minute interview, it would be hard to talk about the whole system. Amongst the features you list, you will have to choose those you want to expand upon (choose the most critical features).

You can get more details through the links in the video descriptions.